Applications of Lattice Codes in Communication Systems

by

Amin Mobasher

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2007

DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

In the last decade, there has been an explosive growth in different applications of wireless technology, due to users' increasing expectations for multi-media services. With the current trend, the present systems will not be able to handle the required data traffic. Lattice codes have attracted considerable attention in recent years, because they provide high data rate constellations. In this thesis, the applications of implementing lattice codes in different communication systems are investigated. The thesis is divided into two major parts. Focus of the first part is on constellation shaping and the problem of lattice labeling. The second part is devoted to the lattice decoding problem.

In constellation shaping technique, conventional constellations are replaced by lattice codes that satisfy some geometrical properties. However, a simple algorithm, called lattice labeling, is required to map the input data to the lattice code points. In the first part of this thesis, the application of lattice codes for constellation shaping in Orthogonal Frequency Division Multiplexing (OFDM) and Multi-Input Multi-Output (MIMO) broadcast systems are considered. In an OFDM system a lattice code with low Peak to Average Power Ratio (PAPR) is desired. Here, a new lattice code with considerable PAPR reduction for OFDM systems is proposed. Due to the recursive structure of this lattice code, a simple lattice labeling method based on Smith normal decomposition of an integer matrix is obtained. A selective mapping method in conjunction with the proposed lattice code is also presented to further reduce the PAPR. MIMO broadcast systems are also considered in the thesis. In a multiple antenna broadcast system, the lattice labeling algorithm should be such that different users can decode their data independently. Moreover, the implemented lattice code should result in a low average transmit energy. Here, a selective mapping technique provides such a lattice code.

Lattice decoding is the focus of the second part of the thesis, which concerns the operation of finding the closest point of the lattice code to any point in N-dimensional real space. In digital communication applications, this problem is known as the integer least-square problem, which can be seen in many areas, e.g. the detection of symbols transmitted over the multiple antenna wireless channel, the multiuser detection problem in Code Division Multiple Access (CDMA) systems, and the simultaneous detection of multiple users in a Digital Subscriber Line (DSL) system affected by crosstalk. Here, an efficient lattice decoding algorithm based on using Semi-Definite Programming (SDP) is introduced. The proposed algorithm is capable of handling any form of lattice constellation for an arbitrary labeling of points. In the proposed methods, the distance minimization problem is expressed in terms of a binary quadratic minimization problem, which is solved by introducing several matrix and vector lifting SDP relaxation models. The new SDP models provide a wealth of trade-off between the complexity and the performance of the decoding problem.

ACKNOWLEDGEMENTS

I wish to take this opportunity to express my appreciation for numerous individuals who made my years at Waterloo a rewarding experience.

I am very grateful to my thesis supervisor, Prof. Amir K. Khandani, for giving me the opportunity and freedom to pursue my research under his supervision. He has been instrumental in bringing me to Waterloo and has provided support during my stay here. I wish to thank him for his constant encouragement and patience. His profound knowledge and scientific curiosity have set high standards and are a constant source of inspiration. Interacting with him as his student and advisee I benefited tremendously from his numerous qualities of a mentor and collaborator. His clarity of thought and keen insight have greatly influenced my research. His advice, both scholarly and non-academic, and most of all his friendship, leave me greatly indebted.

I also wish to thank the other members of my dissertation committee: Professors Babak Hassibi, Arne Storjohann, Anthony Vannelli, Safieddin Safavi-Naeini, and Murat Uysal for having accepted to take the time out of their busy schedules to read my thesis and to provide me with their comments and suggestions.

I have been fortunate to associate with many remarkable people during the time spent at Waterloo. My gratitude goes to numerous friends and colleagues at the University of Waterloo especially in the Coding and Signal Transmission (CST) laboratory for providing such a creative, intellectually stimulating and fun environment. There are numerous members of the CST lab who contribute to the excellent working environment and creative atmosphere. In fact, there are too many to list in the acknowledgements but I am nevertheless truly grateful. However, in particular, I would like to thank Mahmoud Taherzadeh and Mohammad-Ali Maddah-Ali for many lively discussions ranging over a rather wide variety of topics. Masoud Ebrahimi and Mohammad Fakharzadeh, my friends and my previous roommates here in Waterloo, made my graduate life tolerable and kept me sane. I will always remember their support, friendship, and love.

I am indebted to Dr. Renata Sotirov and Prof. H. Wolkowicz in helping me in many inspiration discussions, tremendous help, and support throughout the optimization work in my research. Many people have commented on different aspects of this thesis while work was in progress. In particular, I would like to thank Professors Andrea Goldsmith, Zhi-Quan Luo, Nikos Sidiropoulos, and Tim Davidson for numerous insightful remarks and/or questions

relating to the topics discussed in the thesis. Also, I would like to thank Steve Benson and Futakata Yoshiaki in providing help and support in using the DSDP and SDPA softwares.

During the past two years, I have been in constant support of my wife and love of my life, Mona Mojdeh. Her love, understanding, and support have been phenomenal. You are the reason for the smile on my face, because you make me feel all my dreams can be real. You are always my pillar, joy, and guiding light.

Last, and most importantly, I am forever indebted to my parents, Aman Mobasher and Ashraf Khalilpour, and my sisters, Azin, Arezou, and Azadeh. Their invaluable and relentless support, encouragement, and love are without doubt the most important reasons for my success. I could not have achieved this without their unlimited sacrifice. My parents bore me, raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

To my parents

Aman Mobasher & Ashraf Khalilpour

with all my love

TABLE OF CONTENTS

CHAPTERS

PART I LATTICE LABELING AND CONSTELLATION SHAPING

PART II LATTICE DECODING

APPENDICES

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AWGN	Additive White Gaussian Noise.
BFGS	Broyden-Fletcher-Goldfarb-Shanno.
BPSK	Binary Phase Shift Keying.
CCDF	Complementary Cumulative Density Function.
CDMA	Code Division Multiple Access.
DFD	Decision Feedback Detector.
DSDP	Dual scaling algorithm for SDP.
DSL	Digital Subscriber Line.
FFT	Fast Fourier Transform.
IFFT	Inverse Fast Fourier Transform.
i.i.d.	independent identically distributed.
IPM	Interior Point Method.
KZ	Korkin-Zolotarev.
LLL	Lenstra-Lentsra-Lovasz.
LLR	Log-Likelihood Ratios.
MAP	Maximum A Posteriori.
MIMO	Multiple-Input Multiple-Output.
ML	Maximum-Likelihood.
MLSDP	Matrix Lifting Semi-Definite Programming.
MMSED	Minimum Mean Squared Error Detector.
MSB	Most Significant Bit.
NP-hard	nondeterministic polynomial-time hard.
OFDM	Orthogonal Frequency Division Multiplexing.
PAPR	Peak to Average Power Ratio.
PSK	Phase Shift Keying.

PTS	Partial Transmit Sequence.
QAM	Quadratic Amplitude Modulation.
QPSK	Quadratic Phase Shift Keying.
SD	Sphere Decoder.
SDP	Semi-Definite Programming.
SDPA	Semi-Definite Programming Algorithm.
SeDuMi	Self Dual Minimization.
SER	Symbol Error Rate.
SLM	Selective Mapping.
SNF	Smith Normal Form.
SNR	Signal to Noise Ratio.
VBLAST	Vertical Bell Laboratories Layered Space-Time.
VLSDP	Vector Lifting Semi-Definite Programming.
ZFD	Zero Forcing Detector.

NOTATIONS

Following notations are used throughout the thesis. Plain letters x and X are used for scalars. Boldface letters are used for vectors (lower case) or matrices (upper case), i.e. $\mathbf{x}$ or $\mathbf{X}$. For a $K \times N$ matrix $\mathbf{X}$ the $(i, j)^{th}$ element is represented by x_{ij}, where $1 \leq i \leq K,\ 1 \leq j \leq N$, i.e. $\mathbf{X} = [x_{ij}]$. Also, x_i denotes the i^{th} element of the vector $\mathbf{x}$. For $N \times K$ matrix $\mathbf{X}$, the notation $\mathbf{X}(1 : i, 1 : j)$, $i < N$ and $j < K$ denotes the sub-matrix of $\mathbf{X}$ containing the first i rows and the first j columns.

$\mathbb{R}^N$	N-dimensional real space
$\mathbb{Z}^N$	N-dimensional integer space
$\mathcal{M}_{K\times N}$	The space of $K \times N$ real matrices
$\mathcal{M}_N$	The space of $N \times N$ real matrices
$\mathcal{S}_N$	The space of $N \times N$ symmetric matrices
$\mathbf{e}_N$ $(\mathbf{0}_N)$	The $N \times 1$ vector of all ones (zeros)
$\mathbf{E}_{N\times K}$	The $N \times K$ matrix of all ones
$\mathbf{I}_N$	The $N \times N$ Identity matrix
$\mathbf{x}^T, \mathbf{X}^T$	Transpose of a vector $\mathbf{x}$ or matrix $\mathbf{X}$
$\mathbf{x}^*, \mathbf{X}^*$	Hermitian transpose of a vector $\mathbf{x}$ or matrix $\mathbf{X}$
$\mathbf{X}^{-1}$	Inverse of a square matrix
$\mathbf{X}^\dagger$	Pseudo inverse (Moore-Penrose inverse [78]) of a matrix $\mathbf{X}$
$\|\|\mathbf{x}\|\|$	The Euclidean norm of the vector $\mathbf{x}$
$\|\|\mathbf{X}\|\|_{\mathbb{F}}$	The Frobenius norm of the matrix $\mathbf{X}$ ($\|\|\mathbf{X}\|\|_{\mathbb{F}}^2 = \text{trace}(\mathbf{X}\mathbf{X}^T)$)
$\mathbf{X} \succeq 0$	Positive semi-definiteness matrix $\mathbf{X} \in \mathcal{S}_N$
$\mathbf{X} \succ 0$	Positive definite matrix $\mathbf{X} \in \mathcal{S}_N$

$\mathbf{X} \geq \mathbf{Y}$	For matrices $\mathbf{X}, \mathbf{Y} \succ 0$, $\mathbf{X} - \mathbf{Y} \succeq 0$
$\mathbf{X} \geq \mathbf{Y}$	For matrices $\mathbf{X}, \mathbf{Y} \in \mathcal{M}_{K \times N}$, $x_{ij} \geq y_{ij}$
$\mathbf{X} > \mathbf{Y}$	For matrices $\mathbf{X}, \mathbf{Y} \in \mathcal{M}_{K \times N}$, $x_{ij} > y_{ij}$
$\mathbf{X} \otimes \mathbf{Y}$	The Kronecker product of two matrices $\mathbf{X}$ and $\mathbf{Y}$ (see [55])
$\text{trace}(\mathbf{X})$	Trace of a square matrix $\mathbf{X}$
$\langle \mathbf{X}, \mathbf{Y} \rangle$	The trace inner product $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{trace}(\mathbf{X}\mathbf{Y})$
$\text{vec}(\mathbf{X})$	For $\mathbf{X} \in \mathcal{M}_{N \times K}$, the NK-D vector, formed by columns of $\mathbf{X}$
$\text{diag}(\mathbf{X})$	For $\mathbf{X} \in \mathcal{M}_N$, the vector of the diagonal elements of $\mathbf{X}$
$\text{diag}(\mathbf{x})$	A diagonal matrix with elements of $\mathbf{x}$
$\text{rank}(\mathbf{X})$	The rank of matrix $\mathbf{X}$
$\det(\mathbf{X})$	The determinant of matrix $\mathbf{X}$
$\Re(.)$	The real part of a matrix or vector
$\Im(.)$	The imaginary part of a matrix or vector
$\{.\}$	The nearest integer vector or matrix
$\lfloor . \rfloor$	The nearest integer vector or matrix towards $-\infty$
$\lvert . \rvert$	The cardinality of a set
$\text{conv}(.)$	The convex hull of a set
$E\{\mathbf{u}\}$	Average of a random vector $\mathbf{u}$
$\mathcal{H}(\mathbf{u})$	The *entropy* of a random vector $\mathbf{u}$ with probability density $f(\mathbf{u})$,
	$\mathcal{H}(\mathbf{u}) = -\int f(\mathbf{u}) \log f(\mathbf{u}) d\mathbf{u}$ [39]
$\log(.)$	Logarithm in base 2
$\ln(.)$	Logarithm in base e
$\mathcal{C}_M(0, 2A)$	The M-D cube with side length $2A$, centered at the origin
$\mathcal{B}_M(0, r)$	The M-D Ball (sphere) with radius $\mathbf{r}$, centered at the origin

$\mathcal{O}_M(0, r_i)$	The M-D Oval centered at the origin with radius $\mathbf{r}$, where r_i is the radius in the i^{th} dimension
$N(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$U(\mathcal{R})$	Uniform distribution over region $\mathcal{R}$
$\mathbb{P}\{.\}$	Probability distribution of an event
$F_\gamma(x)$	The probability distribution of γ

CHAPTER 1

INTRODUCTION

Due to the recent developments and available processing power in digital receivers, there is a lot of interest in the design of dense signal constellations. They have been used extensively in different applications such as code design for single antenna Rayleigh fading channel [110], design of dense lattices for Gaussian channel [10], and design of space-time block codes for coherent multiple antenna channels [12, 13]. The theory of Euclidean lattices is shown to be a very powerful tool in the design of such constellations. These constellations are constructed as multidimensional lattice signal sets having some desired geometric properties. An attractive feature of these signal sets is that a significant improvement in error performance is obtained without requiring the use of any conventional channel coding.

Research on coded modulation schemes obtained from lattice constellations began more than twenty years ago, and extensive work has been done to improve the performance of these lattice constellations in different applications. A basic schematic block diagram of a system employing lattice constellations is shown in Figure 1. The main part of this system is a lattice constellation which is defined as a finite subset of points of an N-Dimensional lattice bounded within a support region in $\mathbb{R}^N$. This collection of points is also called a *lattice code*. There are two important tasks in any such system employing a lattice code:

1. Enumeration of lattice points within a lattice constellation, and

2. Search for the nearest neighbor point in a bounded lattice constellation.

Enumerating lattice code points is called *lattice labeling*. The main challenge in lattice labeling is to find a simple algorithm to map the input information bits to lattice constellation points such that the mapping (performed at the transmitter side) and its inverse (performed at

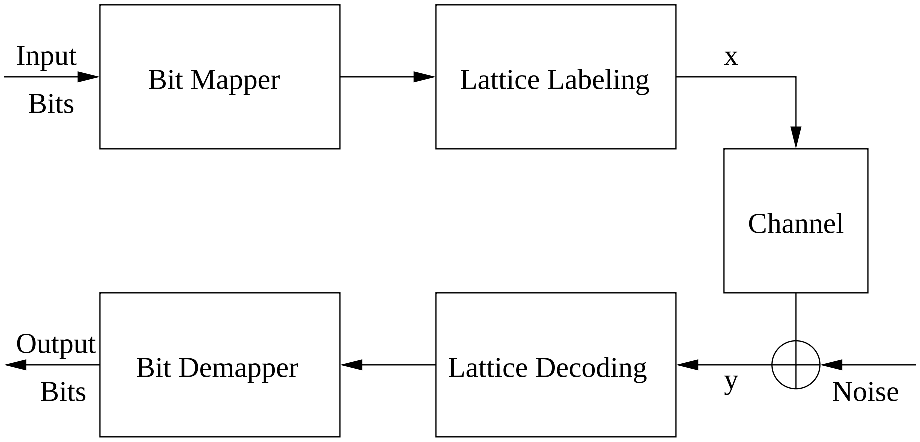

Figure 1: Basic Schematic of a System Employing Lattice Codes

the receiver side) can be implemented with a reasonable complexity. This operation can be potentially of enormous complexity because the lattice constellation typically possesses a huge number of points. Moreover, the lattice codes are designed to satisfy some geometric properties as well, e.g. a reduced peak to average power. This may be very critical for the complexity of practical implementations, as labeling such a lattice code may not always be an easy task. Therefore, an important feature in designing lattice codes is to avoid the use of a huge look-up table to perform lattice labeling while satisfying the desired geometrical property.

The second important feature to consider when utilizing a lattice code is its decoding. In the receiver, a corrupted version of a lattice constellation point affected by the channel and noise is received (see Figure 1). Decoding concerns the operation of recovering the transmitted point form the designed lattice code from the received signal. This problem is known as *lattice decoding*, which is defined as finding the nearest point (in the sense of minimum Euclidean distance) of a lattice code to any point in $\mathbb{R}^N$. Due to the huge number of points in a lattice code, this operation can also be potentially of enormous complexity. Therefore, in any system employing lattice codes a lattice decoding algorithm is desired to avoid the exhaustive enumeration of all points of the lattice constellation. Moreover, due to extensive applications of arbitrary lattice codes in different systems, a universal lattice decoding algorithm is desired that can be used in any such system employing lattice codes.

In this thesis, the applications of implementing lattice codes in different communication systems are investigated. The thesis is divided to two major parts. The focus of the first part is on constellation shaping and the problem of lattice labeling and the second part is devoted to the lattice decoding problem.

Constellation shaping is one of the main reasons that lattice codes are used in different applications. The aim in constellation shaping is to replace the conventional constellations by lattice codes that satisfy some geometrical properties, e.g. reducing the peak to average power ratio or reducing average energy. In the first part of this thesis, the application of lattice codes for constellation shaping in two different scenarios are considered, while having a simple lattice labeling algorithm.

- In Orthogonal Frequency Division Multiplexing (OFDM) systems, a lattice code with low Peak to Average Power Ratio (PAPR) is desired. Here, a lattice code based on Hadamard matrix is designed which results in a low PAPR. Moreover, the lattice labeling algorithm for this lattice constellation (lattice code) is a simple algorithm based on using Smith Normal Form (SNF) of an integer matrix.

- In multiple antenna broadcast systems, a lattice code resulting in low average transmit energy is required. However, design of such a constellation shaping is not an easy task since the corresponding lattice labeling algorithm should be such that different users can decode their data independent of each other. Here, a Selective Mapping (SLM) technique is introduced to provide a lattice code resulting in a low average transmit energy, while users can decode their data independent of each other.

Lattice decoding problem is the focus of the second part of the thesis. In mathematics, lattice decoding algorithm is known as a universal decoding algorithm for any arbitrary lattice code. This problem is equivalent to finding the closest point of the lattice code to any point in $\mathbb{R}^N$. In digital communication applications, this problem is known as the integer least-square problem, which can be seen in many areas, e.g. the detection of symbols transmitted over the multiple antenna wireless channel [134], the multiuser detection problem in Code Division Multiple Access (CDMA) systems [41], and the simultaneous detection of multiple users in a DSL system affected by crosstalk [50]. In this part of thesis, a lattice decoding algorithm based on Semi-Definite Programming (SDP) is proposed. Several relaxation models based on vector lifting and matrix lifting SDP are introduced. It should be mentioned that although the lattice decoding algorithms and analysis are general in nature, and applicable in a wide variety of scenarios in communication systems, in this thesis, the problem of lattice decoding is presented in the context of detection in multiple antenna systems.

1.1 Contributions and Outline

The aim of this thesis is to investigate the applications of lattice codes in different communication systems and their corresponding problems. The thesis is divided into two parts. The first part (Part I including Chapters 2 and 3) is focused on constellation shaping and lattice labeling algorithms. The second part (Part II including Chapters 4, 5, and 6) is devoted to the lattice decoding problem in communication systems.

The focus of Part I and Part II are different in nature. In Part I, the main focus is finding a proper lattice code whose labeling algorithm is simple and it satisfies some geometric property. In an OFDM system a lattice code resulting in a low PAPR is desired. Multiple antenna broadcast systems require a lattice code resulting in a low average transmit energy, while keeping an independent decoding for users. In Part II, dealing with lattice decoding problem, the focus is finding a decoding algorithm with low complexity based on SDP.

1.1.1 Part I - Lattice Labeling and Constellation Shaping

The focus of the first part is on constellation shaping and the corresponding problem of lattice labeling. Constellation shaping is a method to replace the conventional constellations by lattice codes that satisfy some geometrical properties. In this part, the application of lattice codes for constellation shaping in OFDM systems and MIMO broadcast systems are considered.

In Chapter 2, a constellation shaping method with a considerable PAPR reduction is proposed for an OFDM system. The boundary of this cubic constellation, called the Hadamard constellation, is along the bases defined by the Hadamard matrix in the transform domain. In addition, this constellation can be employed in conjunction with another PAPR reduction method. Here, an SLM method is applied in conjunction with the proposed Hadamard constellation to further reduce the PAPR. The encoding method for this shaping technique is derived from the Smith Normal Form (SNF) decomposition, and has a minimal complexity. This new technique offers a PAPR that is significantly lower than that of the best known techniques reported in the literature without any loss in terms of the energy and/or spectral efficiency and without any side information being transmitted. The material in this chapter have been previously published in the works listed below.

- [96] Amin Mobasher and Amir K. Khandani, “PAPR Reduction in OFDM Systems Using Constellation Shaping", In *Proc.* 22nd *Biennial Symposium on Comm.*, Kingston, ON, Canada, May 31 - June 3, 2004

- [97] Amin Mobasher and Amir K. Khandani, “PAPR Reduction Using Integer Structures in OFDM Systems", In *Proc. IEEE Vehicular Technology Conference*, VTC, LA, CA, USA, Sept. 26 - Sept. 29, 2004

- [98] Amin Mobasher and Amir K. Khandani, “Integer-Based Constellation Shaping Method for PAPR Reduction in OFDM Systems," *IEEE Trans. on Comm.*, Vol. 54, Issue 1, pp 119-127, Jan. 2006

Chapter 3 is devoted to finding a lattice code with a low average transmit energy in multiple antenna broadcast systems with channel inversion techniques. By using channel inversion, the decoding at the receiver side is independent of channel matrix. However, because the channel is not orthogonal, the energy of the transmit signal can be very high. Therefore, a lattice code with low average transmit energy is desired in this application. Here, a Selective Mapping (SLM) technique is introduced to provide a lattice code with a low average energy for the transmitted signal in a broadcast system. Using the strong literature on quantization [54, and ref. therein], the gain that the SLM technique can provide is derived. In order to implement the SLM method effectively, using lattice decomposition techniques is proposed. The material in this chapter have been recently submitted for publication.

- [101] Amin Mobasher and Amir. K. Khandani, “Precoding in Multiple-Antenna Broadcast Systems with a Probabilistic Viewpoint", In *Proc. the* 10th *Canadian Workshop on Information Theory*, CWIT’07, Edmonton, Alberta, Canada, June 6 - June 8, 2007

- [102] Amin Mobasher, and Amir. K. Khandani, “Probabilistic Behavior of Average Transmit Energy in Broadcast Systems with Precoding", To be shortly Submitted to *IEEE Trans. on Info. Theory*, 2007

1.1.2 Part II - Lattice Decoding

Lattice decoding is the focus of the second part of the thesis, which concerns the operation of finding the closest point of the lattice code to any point in N-dimensional real space. In Chapter 4, this problem is expressed with the terminology and assumptions of detection in multiple antenna systems.

Chapter 5 develops an efficient approximate Maximum Likelihood (ML) decoder for Multiple Input Multiple Output (MIMO) systems based on Vector Lifting Semi-Definite Programming (VLSDP). In the proposed method, the transmitted vector is expanded as a linear combination (with zero-one coefficients) of all the possible constellation points in each dimension. Using this formulation, the distance minimization in Euclidean space is expressed in terms of a binary quadratic minimization problem. The minimization of this problem is over the set of all binary rank-one matrices with row sums equal to one. In order to solve this minimization problem, two relaxation models (Model III and IV) are presented, providing a trade-off between the computational complexity and the performance (both models can be solved with polynomial-time complexity). The decoding algorithms built on these models have a near-ML performance with polynomial computational complexity. The material in this chapter have been previously published in the works listed below.

- [93] Amin Mobasher, Mahmoud Taherzadeh, Renata Sotirov, and Amir K. Khandani, “An Efficient Quasi-Maximum Likelihood Decoding for Finite Constellations", In *Proc. the* 39*th* *Conference on Information Sciences and Systems*, CISS’05, Baltimore, MD, USA, Mar. 16 - Mar. 18, 2005

- [92] Amin Mobasher, Mahmoud Taherzadeh, Renata Sotirov, and Amir K. Khandani, “A Randomization Method for Quasi-Maximum Likelihood Decoding", In *Proc. the* 9*th* *Canadian Workshop on Information Theory*, CWIT’05, Montréal, Québec, Canada, June 5 - June 8, 2005

- [91] Amin Mobasher, Mahmoud Taherzadeh, Renata Sotirov, and Amir K. Khandani, “A Near Maximum Likelihood Decoding Algorithm for MIMO Systems Based on Graph Partitioning", In *Proc IEEE International Symposium on Information Theory*, ISIT’05, Adelaide, Australia, Sept. 4 - Sept. 9, 2005

- [94] Amin Mobasher, Mahmoud Taherzadeh, Renata Sotirov, and Amir K. Khandani, “A Near Maximum Likelihood Decoding Algorithm for MIMO Systems Based on Semi-Definite Programming," to appear, *IEEE Trans. on Info. Theory*, Vol. 53, No. 11, Nov. 2007

In Chapter 6, an algorithm based on Matrix Lifting Semi-Definite Programming (MLSDP) [40, 11] is introduced for any constellation (QAM or PSK) and any labeling method. This algorithm is inspired by the method proposed in Chapter 5 with an efficient implementation resulting in a better performance and lower computational complexity. In SDP optimization problems, the computational complexity is a polynomial function of the number of variables. Using the proposed method, the number of variables in Chapter 5 is decreased from $(NK+1)^2$ to $(2N+K)^2$, where N is the number of antennas and K is the number of constellation points in each real dimension. Since the computational complexity of solving MLSDP is a polynomial function of the number of variables, a significant complexity reduction is achieved. The material in this chapter has been recently published in the following papers.

- [99] Amin Mobasher and Amir. K. Khandani, “Matrix-Lifting Semi-Definite Programming for Decoding in Multiple Antenna Systems", In *Proc. the* 10*th* *Canadian Workshop on Information Theory*, CWIT’07, Edmonton, Alberta, Canada, June 6 - June 8, 2007

- [100] Amin Mobasher, and Amir. K. Khandani, “Matrix-Lifting Semi-Definite Programming for Decoding in Multiple Antenna Systems", Submitted to *IEEE Trans. on Info. Theory*, Aug. 2007

1.1.3 Some Other Contributions by the Author During his PhD Studies

In this section, some contributions relating to the field of digital communications which does not fall within the scope of the thesis are briefly presented.

Lattice Basis Reduction in Communication in Multiple Antennas Systems

In the recent years, communications over multiple-antenna fading channels has attracted the attention of many researchers. Multiple-antenna systems are the only solution to realize the capacity increase required in the next generation of wireless networks. It has been shown that multi-user systems can exploit most of the advantages of multiple-antenna systems. This section considers a multiple-antenna broadcast system. In a broadcast system, different users should be able to decode the transmitted data independently. Channel inversion at the transmitter is a technique that can simply separate the data for different users. However, this method may result in a very high transmitting power. In [115], the authors have introduced a vector perturbation technique which has a good performance in terms of symbol error rate. In this paper, a new viewpoint for the MIMO broadcast channel based on the lattice-basis reduction ie presented. Instead of approximating the closest lattice point in the perturbation problem, the lattice-basis reduction is applied to reduce the average transmitted energy by reducing the second moment of the fundamental region generated by the lattice basis. The theoretical aspects of this method have also been proven in the following papers. Moreover, the introduced theoretical methods have been applied to the lattice basis reduction technique used in decoding of multiple-antenna systems.

- [132] Mahmoud Taherzadeh, Amin Mobasher, and Amir K. Khandani, "Lattice-Basis Reduction Achieves the Precoding Diversity in MIMO Broadcast Systems", In *Proc. the* 39th *Conference on Information Sciences and Systems*, CISS'05, Baltimore, MD, USA, Mar. 16 - Mar. 18, 2005

- [133] Mahmoud Taherzadeh, Amin Mobasher, and Amir K. Khandani, "LLL Lattice-Basis Reduction Achieves Maximum Diversity In MIMO Systems", In *Proc. IEEE International Symposium on Information Theory* ISIT'05, Adelaide, Australia, Sept 4. - Sept. 9, 2005

- [131] Mahmoud Taherzadeh, Amin Mobasher, and Amir K. Khandani, "Communication Over MIMO Broadcast Channels Using Lattice-Basis Reduction," In *Proc. the 42nd Annual Allerton Conference on Communication, Control, and Computing*, Allerton, Monticello, IL, USA, Sept. 29 - Oct. 1, 2004

- [129] Mahmoud Taherzadeh, Amin Mobasher, and Amir K. Khandani, “Communication over MIMO Broadcast Channels Using Lattice-Basis Reduction," to appear, *IEEE Trans. on Info. Theory*, Vol. 53, No. 12, Dec. 2007
- [130] Mahmoud Taherzadeh, Amin Mobasher and Amir K. Khandani, “LLL Reduction Achieves the Receive Diversity in MIMO Decoding," to appear, *IEEE Trans. on Info. Theory*, Vol. 53, No. 12, Dec. 2007

Fairness in Multiuser Systems

In multi-user systems, multiple transmitters and/or receivers share a common communication medium, and therefore, users should compete for available resources. Different information theoretic criteria lead to several alternatives for distributing the limited resources among users. For example, achieving a high spectral efficiency results in assigning a higher portion of the resources to the users with stronger channels. However, this criterion ignores fairness among the users. Usually, there is a trade-off between global performance and the system fairness. Therefore, providing fairness, while achieving a high performance, is a desired solution. A lot of research has addressed this problem and suggested different criteria to design a fair system. However, the computational complexity of such algorithms is high. The main purpose of this paper is to find a point on the sum-capacity facet which satisfies a notion of fairness among active users. This problem is addressed in two cases: (i) where the complexity of achieving interior points is not feasible, and (ii) where the complexity of achieving interior points is feasible. For the first case, the corner point for which the minimum rate of the active users is maximized (max-min corner point) is desired for signaling. A simple greedy algorithm is introduced to find the optimum max-min corner point. For the second case, the polymatroid properties are exploited to locate a rate-vector on the sum-capacity facet which is optimally fair in the sense that the minimum rate among all users is maximized (max-min rate). In the case that the rate of some users can not increase further (attain the max-min value), the algorithm recursively maximizes the minimum rate among the rest of the users. This work has been published in the following papers:

- [87] Mohammad A. Maddah-Ali, Amin Mobasher, and Amir. K. Khandani, “On the Fairest Corner Point of the MIMO-BC Capacity Region," In *Proc. the 43rd Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Sept. 28 - Sept. 30, 2005
- [88] Mohammad A. Maddah-Ali, Amin Mobasher, and Amir. K. Khandani, “Using Polymatroids to provide Fairness in Multi-User Systems", In *Proc. IEEE International*

Symposium on Information Theory, ISIT'06, Seattle, WA, USA, July 9 - July 14, 2006

- [89] Mohammad Ali Maddah-Ali, Amin Mobasher, and Amir K. Khandani, "Fairness in Multiuser Systems with Polymatroid Capacity Region", *IEEE Trans. on Info. Theory*, Revised, Expected publication, Jul. 2007

1.2 Background on Lattice Theory

Since some basic knowledge of lattices is required for this thesis, this section recalls elementary definitions and properties of lattices. For more details, the reader is referred to [28].

Defenition 1 *A Group G is a set that is closed under an operation $*$, and has an identity element $e \in G$ and an inverse g^{-1} for each $g \in G$.*

The identity element $e \in G$ is defined as an element such that $e * g = g * e = g \forall g \in G$. The inverse of each element g is an element g^{-1} such that $g * g^{-1} = g^{-1} * g = e$. A group G is called *Abelian* if the operation $*$ is cummulative, i.e., if $g * h = h * g, \forall g, h \in G$. the *order* of a group G is the number of its elements, $|G|$.

Defenition 2 *A subgroup G_1 of G is a subset of G that is a group under the operation of G.*

A *coset* of a subgroup G_1 is defined as a subset $g * G_1 = \{g * g_1 : g_1 \in G_1\}$ of G, where $g \in G$. Two cosets of G_1 are either equal or disjoint. Every element of G belongs to one of these cosets. Hence, the set G/G_1 of the cosets of G_1 in G is a partition of G. All the cosets of G_1 have the same size of $|G_1|$. Thus, the number of elements of G/G_1, called the *index* of G_1 in G, is equal to $|G|/|G_1|$. It can be seen that G/G_1 forms a group under the operation $\bullet$ defined by $(g * G_1) \bullet (g' * G_1) = (g * g') * G_1$. This group is called the *quotient group* (of G modulo G_1).

Defenition 3 *Let G and G' be groups under the operations $*$ and $\circ$, respectively. A group homomorphism $\Psi : G \longrightarrow G'$ is a mapping such that $\Psi(g * h) = \Psi(g) \circ \Psi(h), \forall g, h \in G$. G and G' are called isomorphic if there exists a group homomorphism $\Psi : G \longrightarrow G'$ that is both one-to-one and onto.*

In the M-dimensional (M-D) real vector space $\mathbb{R}^M$ the Euclidean distance is defined as $||\mathbf{x}|| =< \mathbf{x}, \mathbf{x} >^{\frac{1}{2}}$. A set of vectors V is called *discrete* if there exists a positive number ρ such that any two vectors of V have distance greater than or equal to ρ.

Defenition 4 *A real lattice Λ is a discrete set of M-D vectors in the real Euclidean space $\mathbb{R}^M$ that forms a group under ordinary vector addition. The set of M-D integer vectors, $\mathbb{Z}^M$, is called integer lattice.*

Every lattice Λ is generated as an Abelian group by the integer linear combinations of a set of linearly independent vectors $\mathbf{b}_1, \cdots, \mathbf{b}_N \in \Lambda$, where the integer $N(\leq M)$ is called the dimension of the lattice Λ.On the other hand,

$$\Lambda = \{\mathbf{y} = \sum_{i=1}^{N} x_i \mathbf{b}_i, x_i \in \mathbb{Z}\}.$$

The set of vectors $\mathbf{b}_1, \cdots, \mathbf{b}_N$ is called a basis of Λ, and the $M \times N$ matrix $B = [\mathbf{b}_1, \cdots, \mathbf{b}_N]$ which has the basis vectors as its columns is called the basis matrix (or generator matrix) of Λ. Any point $\mathbf{y}$ in the lattice can be represented by $\mathbf{y} = \mathbf{B}\mathbf{x}$, where $\mathbf{x} \in \mathbb{Z}^N$. The matrix $\mathbf{G} = \mathbf{B}^T\mathbf{B}$ is called the *Gram matrix* for the lattice Λ. The *determinant of the lattice* is defined as $\det(\Lambda) = \det(\mathbf{G})$.

Defenition 5 *A coset of a lattice Λ, denoted by $\Lambda + \mathbf{c}$, is the set of all N-dimensional vectors of the form $\lambda + \mathbf{c}$, where λ is any point in Λ and $\mathbf{c}$ is some constant vector, known as coset leader, that specifies the coset.*

Geometrically, the coset $\Lambda + \mathbf{c}$ is a translate of Λ by $\mathbf{c}$. Two N-tuples are *equivalent modulo* Λ if their difference is a point in Λ. Thus, the coset $\Lambda + \mathbf{c}$ is the set of all points equivalent to $\mathbf{c}$ modulo Λ.

Defenition 6 *A subset of the elements of Λ, e.g. Λ_s, that is itself a lattice is called sublattice.*

This sublattice induces a partition of Λ into equivalent classes modulo Λ_s. The order of this partition is shown by $|\Lambda/\Lambda_s|$.

Defenition 7 *Fundamental region of a lattice is a building block which when repeated many times fills the whole space with just one lattice point in each copy.*

The fundamental parallelotope is an example of a fundamental region for the lattice Λ. It is defined as the parallelotope consisting of the points $x_i\mathbf{b}_1 + \cdots + x_N\mathbf{b}_N$, for $0 \leq x_i < 1$.

Voronoi region is another example of a fundamental region. Around each lattice point $\mathbf{y}_i \in \Lambda$ is its Voronoi region $\mathcal{V}(\mathbf{y}_i)$ consisting of all points of the underlying space which are closer to that point than to any other point of lattice Λ. More precisely,

$$\mathcal{V}(\mathbf{y}_i) = \left\{\mathbf{x} \in \mathbb{R}^N : ||\mathbf{x} - \mathbf{y}_i|| \leq ||\mathbf{x} - \mathbf{y}_j|| \text{ for all } j \neq i\right\} \tag{1}$$

Any translate $\Lambda + \mathbf{c}$ of Λ is the union of $|\Lambda/\Lambda_s|$ cosets of Λ_s. A *Voronoi Constellation* is the set of these coset leaders, which fall within the Voronoi region around the origin of Λ_s. More generally, a *Lattice constellation (lattice code)* is a finite set of points from an N dimensional lattice $\boldsymbol{\Lambda}$ that lies within a finite region $\mathcal{R} \subset \mathbb{R}^M$. This constellation is known as a lattice code

Defenition 8 *The basis for representing a lattice is not unique. The procedure of finding a basis for a lattice, which is composed of relatively short and nearly orthogonal vectors, is called Lattice Basis Reduction.*

Several distinct notions of reduction have been studied, including those associated to the names Minkowski, Korkin-Zolotarev (KZ), and more recently LLL reduced basis, which can be computed in polynomial time.

A basis $\{\mathbf{b}_1, \cdots, \mathbf{b}_N\}$ is *Minkowski-Reduced Basis* [61] if

- $\mathbf{b}_1$ is the shortest nonzero vector in the lattice Λ, and
- For each $k = 2, ..., N$, $\mathbf{b}_k$ is the shortest nonzero vector in Λ such that $\{\mathbf{b}_1, \cdots, \mathbf{b}_k\}$ may be extended to a basis of Λ.

Finding Minkowski reduced basis is equivalent to finding the shortest vector in the lattice and this problem by itself is NP-hard. Thus, there is no polynomial time algorithm known for this reduction method.

A basis $\{\mathbf{b}_1, \cdots, \mathbf{b}_N\}$ is *KZ Reduced basis* [71] if for its equivalent lower triangular form in equation (2)

$$\hat{\mathbf{B}} = \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \vdots \\ \hat{\mathbf{b}}_N \end{bmatrix} = \begin{bmatrix} \hat{b}_{11} & 0 & \cdots & 0 \\ \hat{b}_{21} & \hat{b}_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{b}_{N1} & \hat{b}_{N2} & \cdots & \hat{b}_{NN} \end{bmatrix}, \tag{2}$$

either $N = 1$ or

- $\hat{\mathbf{b}}_1$ is the shortest nonzero vector in lattice generated by $\hat{\mathbf{B}}$, and
- $|\hat{b}_{i1}| \leq |\hat{b}_{11}|/2$ for $i = 2, \cdots, N$, and
- the submatrix

$$\begin{bmatrix} \hat{b}_{22} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \hat{b}_{N2} & \cdots & \hat{b}_{NN} \end{bmatrix} \tag{3}$$

is KZ reduced.

It can be shown that for each lattice there is at least one KZ basis [119]. There is no polynomial time algorithm known for KZ reduction; however, the fastest known algorithm for this reduction is due to Schnorr [119]. This algorithm is an improved version of Kannan's shortest lattice vector algorithm [68].

A basis $\{\mathbf{b}_1, \cdots, \mathbf{b}_N\}$ is *LLL Reduced Basis* [80] if for its equivalent lower triangular form in equation (2) either $N = 1$ or

- $\|\hat{\mathbf{b}}_1\| \leq (2/\sqrt{3})\|\hat{\mathbf{b}}_2\|$, and
- $|\hat{b}_{i1}| \leq |\hat{b}_{11}|/2$ for $i = 2, \cdots, N$, and
- the submatrix in equation (3) is LLL reduced.

LLL reduced basis has extended applications in several contexts due to the polynomial time algorithm for computing this basis.

PART I

Lattice Labeling and Constellation Shaping

CHAPTER 2

PAPR REDUCTION IN OFDM SYSTEMS

Abstract **–** The problem of reducing the Peak to Average Power Ratio (PAPR) in an Orthogonal Frequency Division Multiplexing (OFDM) system is considered. A cubic constellation, called the Hadamard constellation, is designed whose boundary is along the bases defined by the Hadamard matrix in the transform domain. Then, the PAPR is further reduced by applying the Selective Mapping technique. The encoding method, following the method introduced in [75], is derived from a decomposition known as the Smith Normal Form (SNF). This new technique offers a PAPR that is significantly lower than that of the best known techniques without any loss in terms of energy and/or spectral efficiency and without any side information being transmitted. Moreover, it has a low computational complexity.

2.1 *Introduction*

Orthogonal Frequency Division Multiplexing (OFDM) is a multi-carrier transmission technique which is widely adopted in different communication applications. OFDM prevents inter symbol interference by inserting a guard interval and mitigates the frequency selectivity of a multi-path channel by using a simple equalizer. This simplifies the design of the receiver and leads to inexpensive hardware implementations. Also, OFDM offers some advantages in higher order modulations and in the networking operations. These advantages position OFDM as the technique of choice for the next generation of wireless networks. However, OFDM systems suffer from a large Peak to Average Power Ratio (PAPR) of the transmitted signals, requiring power amplifiers with a large linear range.

Figure 2 shows a basic block diagram of an OFDM transmitter and its receiver. Let $\mathbf{x} = [x_0, x_1, \cdots, x_{N-1}]^T$ denote a vector of $2N$ Dimensional ($2N$-D) constellation points. This vector is selected from a set of N identical 2-D sub-constellations, $\{s_1, \cdots, s_K\}$, and it is transmitted by using one OFDM vector of size N; namely, $\mathbf{y}$.

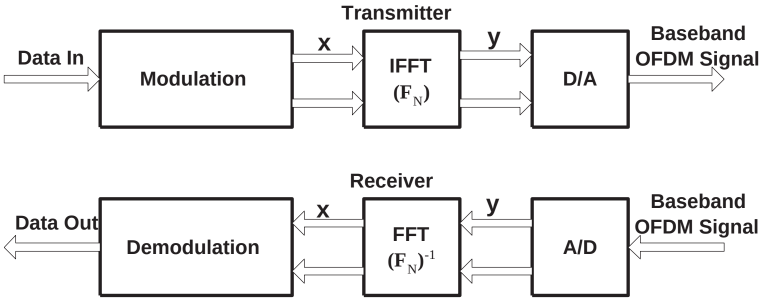

Figure 2: Basic OFDM transmitter and receiver

The discrete time samples of the OFDM signal can be expressed as

$$y_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} x_k e^{j2\pi \frac{nk}{N}}. \tag{4}$$

The matrix representation of this signal is

$$\mathbf{y} = \mathbf{F}_N \mathbf{x}, \tag{5}$$

where $\mathbf{y} = [y_0 \cdots y_{N-1}]^T$, $\mathbf{x} = [x_0 \cdots x_{N-1}]^T$, and $\mathbf{F}_N$ is the **IFFT** matrix,

$$\mathbf{F}_N = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & \cdots & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & \cdots & e^{j2\pi \frac{nk}{N}} & \cdots & e^{j2\pi \frac{n(N-1)}{N}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & \cdots & e^{j2\pi \frac{k(N-1)}{N}} & \cdots & e^{j2\pi \frac{(N-1)^2}{N}} \end{bmatrix}. \tag{6}$$

The 2-D constellation points, $\{x_0, x_1, \cdots, x_{N-1}\}$, may add constructively and produce a time domain signal with a large amplitude. Thus, the output signal $\mathbf{y}$ may have high output levels, which leads to the requirement of an expensive analog front end.

Usually, the level of the amplitude fluctuation of the discrete time OFDM signal is measured in terms of the ratio of the peak power to the average envelope power of the signal as

$$\text{PAPR}(\mathbf{y}) = \frac{||\mathbf{y}||_\infty^2}{E_y\left[\frac{1}{N}||\mathbf{y}||^2\right]}. \tag{7}$$

The continuous time PAPR is typically estimated by the discrete time PAPR by employing the IFFT of length LN for the zero padded sequence of length LN derived from the sequence

$\{x_0, x_1, \cdots, x_{N-1}\}$ in (4) [136, 153, 109]. Therefore,

$$y_n = \frac{\sqrt{L}}{\sqrt{LN}} \sum_{k=0}^{LN-1} x'_k e^{j2\pi \frac{nk}{LN}}, \tag{8}$$

where

$$x'_k = \begin{cases} x_k, & \text{for } k < N, \\ 0, & \text{for } k \geq N, \end{cases} \tag{9}$$

and L is the oversampling factor.

In the sequel, the concentration is on matrices and equations with real entries and complex equations like (5) are represented by real matrices as

$$\begin{bmatrix} \Re(\mathbf{y}) \\ \Im(\mathbf{y}) \end{bmatrix} = \begin{bmatrix} \Re(\mathbf{F}_N) & -\Im(\mathbf{F}_N) \\ \Im(\mathbf{F}_N) & \Re(\mathbf{F}_N) \end{bmatrix} \begin{bmatrix} \Re(\mathbf{x}) \\ \Im(\mathbf{x}) \end{bmatrix}, \tag{10}$$

In [75], this model is used for representing the OFDM signal by real matrices.

A large number of methods for the PAPR reduction has been proposed [36, 113, 116, 81, 120, 108, 109, 104, 17, 149, 21, 26, 72, 62, 107, 75, 76, 77]. In [36, 113], coding techniques are used for PAPR reduction; however, codes offering a low PAPR can be constructed only at the cost of sacrificing the data rate. Clipping the OFDM signal before amplification is a simple and typical method for the PAPR reduction [116, 81, 120]. The effects of over-sampling and clipping for an OFDM signal are analyzed in [116, 120, 109]. The authors in [26] propose a new lattice-based multicarrier modulation technique for Digital Subscriber Line (DSL) applications with a low PAPR; however, this technique is not based on a sinusoidal modulation that is usually employed for OFDM systems.

Another types of PAPR reduction methods are the probabilistic schemes. These schemes are classified in two known groups. One is the Partial Transmit Sequence (PTS) [104] in which each sub-block of subcarriers is multiplied by a constant phase factor, and these phase factors are optimized to minimize the PAPR. The other scheme is Selective Mapping (SLM) in which multiple sequences are generated from the same information, and the sequence with the lowest PAPR is transmitted [17, 149, 21]. Typically, the receiver needs to know which sequence is selected in order to recover the data. However, the methods introduced in [104, 17, 149, 21] eliminate the need for this explicit side information.

Constellation shaping is another important technique in PAPR reduction. In the method proposed in [72], the outer constellation points are extended to minimize the PAPR of the OFDM symbol. The idea of applying the trellis shaping technique to reduce PAPR in OFDM systems is introduced in [62]. This line of research is further investigated in [107] by exploiting the property that the autocorrelation of the data sequence in the frequency domain and the power spectrum in the time domain form a Fourier transform pair. Therefore, minimizing the sidelobe of the autocorrelation of the data sequence is equivalent to reducing the PAPR of the OFDM signal. A comparison with [107] will be provided later. In [75, 76, 77], another constellation shaping technique is proposed to reduce the PAPR of the OFDM signals. The encoding and decoding algorithms of this method are based on the relations and generators in a free Abelian group. Due to the large complexity of this algorithm, its practical implementation, in the case of Fourier transformation in OFDM systems, is very challenging.

In this chapter, a constellation shaping method in an OFDM system with a considerable PAPR reduction is proposed. The boundary of this cubic constellation, called the Hadamard constellation, is along the bases defined by the Hadamard matrix in the transform domain. In addition, this constellation can be employed in conjunction with another PAPR reduction method. Here, an SLM method is applied in conjunction with the proposed Hadamard constellation to further reduce the PAPR. The encoding method for this shaping technique, following the method introduced in [75], is derived from the Smith Normal Form (SNF) decomposition, and has a minimal complexity. This new technique offers a PAPR that is significantly lower than that of the best known techniques reported in the literature without any loss in terms of the energy and/or spectral efficiency and without any side information being transmitted.

2.2 Constellation Shaping

In the constellation shaping technique, a constellation in the frequency domain must be found such that the resulting shaping region in the time domain has a low PAPR. A new constellation shaping method is introduced in [75, 76, 77] by Kwok and Jones. Based on the encoding algorithm introduced in [75, 76, 77], a cubic constellation, along with an SLM method is proposed to reduce the PAPR in an OFDM system.

In a PAPR reduction problem, the peak value of the signal vector is bounded by a specified value $\|\mathbf{y}\|_\infty \leq \beta$ (without loss of generality, assume $\beta = 1$). If the time domain signal is related to the frequency domain constellation point by $\mathbf{y} = \mathbf{A}\mathbf{x}$, this inequality on the time domain boundary translates to a parallelotope[1] in the frequency domain, defined by $\mathbf{A}^{-1}$. Indeed, the constellation boundary is a parallelotope, defined by $\mathbf{Q}_N = \left[\alpha \mathbf{A}^{-1}\right]$. The parameter α is the smallest value that guarantees the number of points in the shaped constellation is the same as the number of points in unshaped constellation. The rounding operation is required to impose the constraint that the parallelotope corners lie in an integer lattice. The main challenge in constellation shaping is to find a unique way to map the input data to the constellation points such that the mapping (encoding) and its inverse (decoding) can be implemented by a reasonable

[1] The parallelotope bases are defined along the columns of $\mathbf{A}^{-1}$.

complexity. Kwok in [75] proved that the shaped constellation for an OFDM system is the points inside the quotient group $\mathbb{Z}^N/\Lambda(\mathbf{Q}_N)$, where $\Lambda(\mathbf{Q}_N)$ is the lattice defined by $\mathbf{Q}_N$, which is based on rounding off the scaled version of the inverse of the IFFT matrix. The points inside this parallelotope (lattice code) are used as the constellation points in transmitting the OFDM signals. Using the relations and generators in a free Abelian group, the points inside this constellation are encoded (labeled) in [75]. The following theorem provides the mathematical tool for the encoding procedure of these points:

Theorem 1 ([75]) *Any relation matrix $\mathbf{Q}_N$ can be decomposed into $\mathbf{Q}_N = \mathbf{UDV}$, where $\mathbf{D}$ is diagonal with the entries $\{\sigma_i\}_{i=1}^N$ such that $\sigma_1 \mid \sigma_2 \mid \cdots \mid \sigma_N$, and $\mathbf{U}$ and $\mathbf{V}$ are unimodular matrices*[2].

The decomposition of the relation matrix $\mathbf{Q}_N$ is performed via column and row operations [75], which is impractical for an OFDM system.

It can be observed that this decomposition is known as the Smith Normal Form (SNF) decomposition of an integer matrix [24] in the mathematical literature, and the matrix $\mathbf{D}$ is called the SNF of the matrix $\mathbf{Q}_N$. The SNF decomposition is a diagonalization of a matrix in the integer domain. Introduced by Smith [122], this concept has been used in many applications such as solving linear diophantine equations, finding the permutation equivalence and similarity of matrices, determining the canonical decomposition of the finitely generated Abelian groups, integer programming, computing additional normal forms, including Frobenius and Jordan normal forms, and separable computing of the discrete Fourier transform. For more historical remarks and applications of the SNF, see [105, 126, 16].

The major contributions to the computational complexity in [75] are the decomposition of the matrix $\mathbf{Q}_N$, the *off-line* procedure, and the encoding algorithm for this constellation, the *online* procedure. The interpretation of the column and row operations as SNF of an integer matrix links the problem to a rich body of knowledge developed in the context of SNF decomposition. Unfortunately, computing the SNF decomposition for an OFDM system is impractical due to the rapid growth in the size of the intermediate integer values. Moreover, in [75], it is shown that the complexity of the encoding procedure is $O(N^2)$, i.e. for a realistic OFDM system the complexity of the online procedure remains very high.

If the SNF decomposition of the matrix $\mathbf{Q}_N$ is given, the encoding algorithm for the shaped

[2] *The condition $\sigma_1 \mid \sigma_2 \mid \cdots \mid \sigma_N$ in Theorem 1 is defined for finding a unique decomposition and can be ignored in the encoding procedure.*

constellation can be represented by [75]

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{U}\boldsymbol{\lambda} \\ \boldsymbol{\gamma} &= \left\lfloor \mathbf{Q}_N^{-1}\hat{\mathbf{x}} \right\rfloor \\ \mathbf{x} &= \hat{\mathbf{x}} - \mathbf{Q}_N\boldsymbol{\gamma}, \end{aligned} \tag{11}$$

where $N = 2^n$, $\boldsymbol{\lambda}$ is the canonical representation of an integer I which represents the data to be sent, and $\mathbf{x}$ is the constellation point corresponding to I. The time domain signal is computed using the IFFT operation. The canonical representation of an integer I can be calculated by the recursive modulo operation; namely,

$$\begin{aligned} \lambda_1 &= I \bmod \sigma_1 \\ I_1 &= \frac{I - \lambda_1}{\sigma_1} \\ \lambda_i &= I_{i-1} \bmod \sigma_i \\ I_i &= \frac{I_{i-1} - \lambda_i}{\sigma_i}, \end{aligned} \tag{12}$$

where $1 \leq i \leq N$.
Also, the reverse operation for finding I from the N-D vector $\mathbf{x}$ is [75]

$$\begin{aligned} &\boldsymbol{\lambda} = \mathbf{U}^{-1}\mathbf{x} = (\lambda_1, \lambda_2, \cdots, \lambda_N)^T, \\ &\tilde{\lambda}_i = \lambda_i \bmod \sigma_i\ , \\ &I = \tilde{\lambda}_1 + \sigma_1(\tilde{\lambda}_2 + \sigma_2(\cdots(\tilde{\lambda}_{N-1} + \sigma_{N-1}\tilde{\lambda}_N)\cdots)). \end{aligned} \tag{13}$$

In [76], it is shown that if the matrix $\mathbf{Q_N}$ is replaced by the Hadamard matrix, $\mathbf{H}_{2^n}$, the corresponding encoding and decoding algorithms for the constellation can be implemented by a butterfly structure that uses only bit shifting and logical AND. This simplicity is due to the following recursive formula for the Hadamard matrix:

$$\mathbf{H}_{2^n} = \begin{bmatrix} \mathbf{H}_{2^{n-1}} & \mathbf{H}_{2^{n-1}} \\ \mathbf{H}_{2^{n-1}} & -\mathbf{H}_{2^{n-1}} \end{bmatrix}, \text{ where } \mathbf{H}_1 = [1]\,. \tag{14}$$

The SNF decomposition of (14) can be easily computed as $\mathbf{H}_{2^n} = \mathbf{U}_{2^n}\mathbf{D}_{2^n}\mathbf{V}_{2^n}$, where

$$\begin{aligned} \mathbf{U}_{2^n} &= \begin{bmatrix} \mathbf{U}_{2^{n-1}} & 0 \\ \mathbf{U}_{2^{n-1}} & \mathbf{U}_{2^{n-1}} \end{bmatrix} \mathbf{D}_{2^n} = \begin{bmatrix} \mathbf{D}_{2^{n-1}} & 0 \\ 0 & 2\mathbf{D}_{2^{n-1}} \end{bmatrix} \\ \mathbf{V}_{2^n} &= \begin{bmatrix} \mathbf{V}_{2^{n-1}} & \mathbf{V}_{2^{n-1}} \\ 0 & -\mathbf{V}_{2^{n-1}} \end{bmatrix} \mathbf{U}_{2^n}^{-1} = \begin{bmatrix} \mathbf{U}_{2^{n-1}} & 0 \\ -\mathbf{U}_{2^{n-1}} & \mathbf{U}_{2^{n-1}} \end{bmatrix}, \end{aligned} \tag{15}$$

and $\mathbf{U}_1 = \mathbf{U}_1^{-1} = \mathbf{D}_1 = \mathbf{V}_1 = [1]$.

2.3 Hadamard Constellation in OFDM Systems

As mentioned in Section 2.2, in OFDM systems, the boundary of the constellation that leads to a low PAPR is along the bases of the IFFT matrix. However, the corresponding SNF decomposition required in the encoding procedure cannot be computed. If the IFFT operation is replaced by the Hadamard operation, a simple encoding algorithm results. However, this type of multicarrier modulation is not very popular because it does not offer the advantages of the conventional OFDM [23].

In this part, the conventional constellation in OFDM systems is replaced by a cubic constellation, called the Hadamard constellation, whose boundary is along the bases defined by the Hadamard matrix in the transform domain. Figure 3 shows the boundaries of these two constellations. The solid line represents the boundary of the constellation which is based on the IFFT matrix. The dashed line shows the boundary of the Hadamard constellation. The IFFT and the Hadamard are both orthogonal matrices, and; therefore, the constellation boundaries along these orthogonal bases are a rotated version of each other. As a result, it is expected that a large number of points within these boundaries will be the same, as shown in Figure 3. Therefore, by substituting the constellation along the IFFT matrix with a constellation along the Hadamard matrix, the resulting PAPR is reduced. Moreover, the encoding of this new constellation, based on the SNF decomposition of the Hadamard matrix, is simple and practical.

Note that in this work, the time domain signal, $\mathbf{y}$, is obtained by the IFFT transformation of the constellation point, $\mathbf{x}$. This results in a traditional OFDM signal based on IFFT/FFT operation. In other words, only the constellation boundary is determined using the Hadamard matrix, i.e. $\mathbf{Q}_N = \mathbf{H}_{2^n}$ in (11).

To further reduce the PAPR, the Hadamard constellation can be concatenated with other methods for PAPR reduction. This motivates us to apply a Selective Mapping (SLM) technique [9, 42] to the Hadamard constellation. In typical SLM methods [9, 42], the major PAPR reduction is achieved by the first few redundant bits. Employing more redundant bits necessitates a high level of complexity to obtain modest improvements in the PAPR. However, in the proposed SLM method, employing the Hadamard constellation causes a considerable PAPR reduction by itself. As a result, by using just one or two redundant bits in SLM, this method significantly outperforms the other PAPR reduction techniques reported in the literature. Note that, it is also possible to apply a PTS method [17] to the Hadamard constellation.

2.3.1 Complex Representation

As stated in Section 2.1, (10) can be applied to change the complex equations of an OFDM system to real equations. This leads to the change of the constellation boundary. Generally,

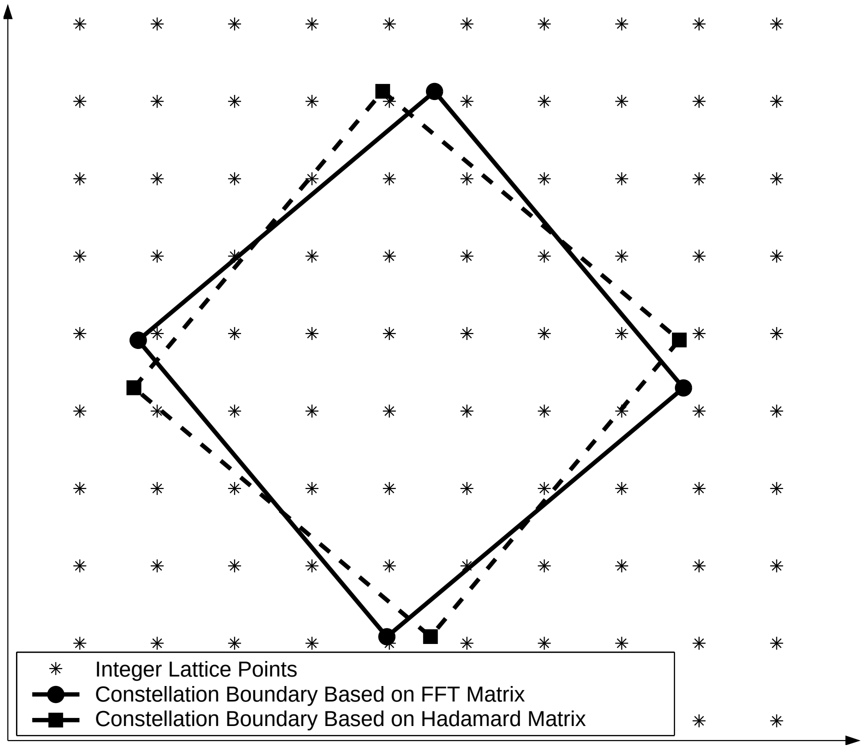
Integer Lattice Points
Constellation Boundary Based on FFT Matrix
Constellation Boundary Based on Hadamard Matrix

Figure 3: N-D signal constellation for IFFT and Hadamard matrix.

two classes of boundaries [106, 73] can be distinguished: 1) the Cartesian boundary that results by viewing the real and imaginary parts of the signal as two separate real signals, and 2) the Polar boundary that considers the envelope and phase of the OFDM signal in a complex plane. The Cartesian boundary limits each component of the complex signal within a square, while the Polar boundary limits this component within a circle. In this part, the complex representation of the OFDM signal is avoided by treating the real and the imaginary parts of the signal separately, which is equivalent to using a Cartesian boundary.

2.3.2 Encoding Procedure

The points inside the Hadamard constellation are mapped to the input data by the encoding procedure, introduced in (11)–(13). The number of these points inside the shaped constellation is determined by the determinant of the Hadamard matrix, $\det(\mathbf{H}_{2^n})$ [47].

Theorem 2 *The size of the shaped constellation defined by a $2^n \times 2^n$ Hadamard matrix is* $\det(\mathbf{H}_{2^n}) = 2^{n2^{n-1}}$.

Proof: see Appendix A. ■

According to the large Hadamard constellation size, in (12), the canonical representation of the large numbers should be computed. The canonical representation of the integer numbers can be simplified based on the fact that digital communication systems deal with binary input streams. Based on (13), an integer I can be represented by

$$I = \lambda_1 + \sigma_1\lambda_2 + \sigma_1\sigma_2\lambda_3 + \cdots + \sigma_1 \ldots \sigma_{N-1}\lambda_N, \tag{16}$$

where $N = 2^n$, and $\{\lambda_i\}_{i=1}^N$ is the canonical representation of I, given in (12), with $\lambda_1 = 0$. According to (15), for a $2^n \times 2^n$ Hadamard matrix, all $\{\sigma_i\}_{i=1}^N$ are powers of 2, i.e.,

$$\{\sigma_i\}_{i=1}^N = \{1, 2, 2, 4, 2, 4, 4, 8, \cdots, 2^n\}. \tag{17}$$

Let $k_i = \log_2 \sigma_i$; therefore,

$$\begin{aligned} I &= 2^{k_1}\lambda_2 + 2^{k_1+k_2}\lambda_3 + 2^{k_1+k_2+k_3}\lambda_4 \\ &\quad + \cdots + 2^{k_1+\cdots+k_{N-1}}\lambda_N \\ &= \lambda_2 + 2\lambda_3 + 2^2\lambda_4 + 2^4\lambda_5 + \cdots + 2^{n2^{n-1}-n}\lambda_N. \end{aligned} \tag{18}$$

The representation of $d = 2^{n2^{n-1}}$ integer numbers corresponding to the Hadamard constellation points necessitate that $n_b = \log_2(d) = n2^{n-1}$ bits represent these numbers. Thus, the binary representation of I is expressed as

$$\begin{aligned} I = {} & b_0 + 2b_1 + 2^2b_2 + 2^3b_3 + \cdots + 2^{n_b-1}b_{n_b-1} \\ = {} & b_0 + 2b_1 + 2^2(b_2 + 2b_3) + 2^4b_4 + 2^5(b_5 + 2b_6) \\ & + \cdots + 2^{n2^{n-1}-n}(b_{n_b-n} + \cdots + 2^{n-1}b_{n_b-1}) \end{aligned} \tag{19}$$

A comparison of (18) and (19) is depicted in Figure 4. Each λ_i consists of $k_i = \log_2 \sigma_i$ bits of the input binary data. This representation will simplify the encoding algorithm. Moreover, the problem of using large numbers in the encoding procedure will be avoided.

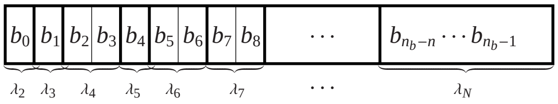

Figure 4: Mapping between binary representation of the information and $\{\lambda_i\}$.

Theorem 2 shows that the size of the Hadamard constellation for a $2^n \times 2^n$ Hadamard matrix is $2^{n2^{n-1}}$. Therefore, the transmission rate is related to $N = 2^n$, the number of subcarriers, in the OFDM system[3]. This rate is unacceptable not only because it depends on N, but also because it is usually higher than the required value. Therefore, a subset of the points inside the shaped constellation are selected for transmission such that they form a constellation with the desired rate. Also, the selected points should be uniformly distributed in the original Hadamard

[3] For $N = 2^n$, the rate for each real component is $\log_2(2^{n2^{n-1}})/N = \dfrac{n}{2}$.

constellation in order to maintain the same peak as well as average energy values (assuming continuous approximation). Note that the Hadamard constellation is called the root constellation for the aforementioned set of the uniformly distributed points in the sequel.

Noting (11) and (12), there is an isomorphism between the integer set

$$\mathcal{S} = \left\{0, 1, \cdots, 2^{n2^{n-1}} - 1\right\} \tag{20}$$

and the set of the points within the Hadamard constellation. Equivalently, the set $\mathcal{S}$ can be considered as a label group for the constellation points (refer to [48] for the definition). A subgroup of the constellation points results in a uniformly distributed subset of the Hadamard constellation points. Consequently, this subgroup of constellation points is isomorphic to a subgroup in the label group $\mathcal{S}$. This subgroup can be selected such that its elements are congruent to zero modulo c, namely

$$\mathcal{P} = \{I \in \mathcal{S} \mid I = 0 \mod c\}, \tag{21}$$

where c is determined by the ratio of the size of the Hadamard constellation, $2^{n2^{n-1}}$, and the size of the constellation, 2^{rN}, with the desired rate, r. Employing (11) and (12), the labels in the subgroup $\mathcal{P}$ determine the set of uniformly distributed points in the Hadamard constellation. By relying on the continuous approximation, such a uniform distribution affects neither the probabilistic behavior of the PAPR nor the average energy of the constellation points.

The Hadamard constellation has almost the same average energy as the constellation resulted by employing QAM signalling in an OFDM system. It can be easily seen that the Hadamard constellation points in (11) can be represented by $\mathbf{x} = \mathbf{H}_N\mathbf{c}$, where $-\frac{1}{2} \leq \mathbf{c} < \frac{1}{2}$. Therefore, the Hadamard constellation contains all the integer points inside a hyper-cube whose boundary is along the columns of the Hadamard matrix. By considering $\mathbf{H}_N\mathbf{H}'_N = N\mathbf{I}_N$, while $\mathbf{F}_N\mathbf{F}'_N = \mathbf{I}_N$, the Hadamard constellation is N times smaller than a cubic constellation whose sides are the columns of the Hadamard matrix. Then, it is straightforward that the average energy per each dimension of the Hadamard constellation is

$$E_{ave}(\frac{n}{2}) = \frac{1}{12}\frac{2^{2n} - 1}{2^n}. \tag{22}$$

Note that (22) shows the average energy per dimension for the root constellation, i.e. the transmission rate is $\frac{n}{2}$. This energy is $\frac{2^n+1}{2^n}$ times the average energy of the equivalent constellation in an OFDM system employing QAM signalling with the same transmission rate.

In the case that the transmission rate is r, as mentioned in (21), the constellation points form a subgroup of the Hadamard constellation points (uniformly distributed subset). Therefore, the constellation has the same energy as in (22); however, the distance among the points is increased by a factor of 2^{n-2r}. Therefore,

$$E_{ave}(r) = \frac{1}{2^{n-2r}}E_{ave}(\frac{n}{2}). \tag{23}$$

Note that the average energy in (23) is $\frac{2^{2n}-1}{2^{2n}} \times \frac{2^{2r}}{2^{2r}-1}$ times the average energy of the equivalent constellation in an OFDM system employing QAM signalling with the same transmission rate. This justifies the earlier claim that the average energy remains almost constant.

2.3.3 Decoding Procedure

At the receiver end, the time domain signal is filtered by a low pass filter and sampled at the Nyquist rate. The samples are processed by an FFT to recover the constellation point in the frequency domain. For an Additive White Gaussian Noise (AWGN) channel, the received vector is given by

$$\mathbf{z} = \mathbf{y} + \mathbf{n}, \tag{24}$$

where $\mathbf{y}$ is the transmitted time domain signal in (11) and $\mathbf{n}$ is a zero-mean complex AWGN. The approximated constellation point is written as

$$\hat{\mathbf{x}} = \mathrm{FFT}(\mathbf{z}) = \mathbf{x} + \mathrm{FFT}(\mathbf{n}) = \mathbf{x} + \mathbf{n}', \tag{25}$$

where $\mathbf{x}$ is the transmitted constellation point, and $\mathbf{n}'$ is a zero-mean complex AWGN. Since the FFT matrix is an orthonormal matrix, the resulting noise is still AWGN. The maximum likelihood decoder simply rounds off the received constellation point $\hat{\mathbf{x}}$ in the integer domain. Then, the resulting constellation point is replaced in (13) to decode the transmitted signal.

2.3.4 Example

To further clarify the algorithm, the constellation points in an OFDM system with 16 subcarriers are computed. Theorem 2 states that there are 2^{32} points inside the Hadamard constellation, i.e. the real and imaginary parts of the signal can be one of these points (equivalent to using a 16-QAM in the OFDM system). The Hadamard matrix and its SNF decomposition are calculated by (14) and (15). Then, the input data is encoded using (11). In Table 1, some of the constellation points are computed.

The SNF decomposition of the matrix $\mathbf{Q}_N$ based on the IFFT matrix, even for this small case, is difficult.

2.4 *Selective Mapping*

SLM is a method to reduce the PAPR in an OFDM system, which involves generating a large set of data vectors that represent the same information, where the data vector with the lowest PAPR is used for the transmission. Here, a method to apply the SLM technique is presented to further reduce the PAPR in the constellation developed earlier.

data	$b_0b_1b_2\cdots b_{31}$	$\lambda_1\lambda_2\cdots\lambda_{16}$	$x_0x_1\cdots x_{15}$
0	00000000000000000000000000000000	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1	10000000000000000000000000000000	0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0	-1 0 -1 0 -1 0 -1 0 -1 0 -1 0 -1 0 -1 0
2	01000000000000000000000000000000	0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0	-1 -1 0 0 -1 -1 0 0 -1 -1 0 0 -1 -1 0 0
3	11000000000000000000000000000000	0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0	-1 0 0 1 -1 0 0 1 -1 0 0 1 -1 0 0 1
4	00100000000000000000000000000000	0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1
5	10100000000000000000000000000000	0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0	0 -1 0 0 0 -1 0 0 0 -1 0 0 0 -1 0 0
6	01100000000000000000000000000000	0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0	0 0 -1 0 0 0 -1 0 0 0 -1 0 0 0 -1 0
7	11100000000000000000000000000000	0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0	1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0
8	00010000000000000000000000000000	0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0	-2 0 0 0 -2 0 0 0 -2 0 0 0 -2 0 0 0
9	10010000000000000000000000000000	0 1 0 2 0 0 0 0 0 0 0 0 0 0 0 0	-1 0 1 0 -1 0 1 0 -1 0 1 0 -1 0 1 0
10	01010000000000000000000000000000	0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0	-1 1 0 0 -1 1 0 0 -1 1 0 0 -1 1 0 0
$\vdots$	$\ddots$	$\ddots$	$\vdots$
10^3	00010111110000000000000000000000	0 0 0 2 0 3 3 1 0 0 0 0 0 0 0 0	-1 1 1 1 -1 0 0 0 -1 1 1 1 -1 0 0 0
$\vdots$	$\ddots$	$\ddots$	$\vdots$
10^6	00000010010000101111000000000000	0 0 0 0 0 2 0 1 0 2 2 7 0 0 0 0	0 0 0 0 0 -2 0 -1 -2 0 0 1 -2 -2 0 0
$\vdots$	$\ddots$	$\ddots$	$\vdots$
$2^{32}-1$	11111111111111111111111111111111	0 1 1 3 1 3 3 7 1 3 3 7 3 7 7 15	2 -1 -1 -1 -1 -1 -1 1 -1 -1 -1 1 -1 1 1 3

Table 1: Example of the encoding procedure for the constellation points in a Hadamard constellation in an OFDM system with 16 sub-carriers employing 16-QAM

Assume that the data rate is r bits per block of length-N FFT symbols. Let r_s denote the number of redundant bits specified for SLM ($r_s \ll r$ and $r = \log_2$(constellation size)). Therefore, there are $N_s = 2^{r_s}$ constellation points representing the same information for transmission in SLM. In the proposed SLM method, the input integers, I, are mapped to the Hadamard constellation points, and the constellation points corresponding to the integers with the same r_s Most Significant Bits (MSBs) are classified in the same subset. Note that the constellation points in each subset represent the same information. The time domain signals corresponding to the frequency domain constellation points are computed by the IFFT transformation, and the constellation point with the lowest PAPR is transmitted.

The details of this scheme are described in the following. In the first step, the input binary sequence is divided into blocks of $r - r_s$ bits. Then, r_s bits of zeros are added to each information block, and these blocks are divided into subblocks of lengths $\log_2 \sigma_i$, $i = 1, \cdots, N$, bits (refer to Figure 4). The binary representations of these subblocks form the vector $\boldsymbol{\lambda}$ which leads to the calculation of the constellation point using (11). The other data vectors are obtained by changing the r_s MSBs of the binary information sequence. Therefore, N_s Hadamard constellation points with different values for the PAPR is calculated. Finally, the constellation point with the lowest PAPR is selected for the transmission.

The different constellation points that represent the same information have the same $r - r_s$ bits. Thus, at the receiver end, the constellation point is decoded by (13), and the r_s extra bits are discarded. Therefore, this method can be expressed as a variant of SLM in which no side information on the choice of the transmit signal needs to be transmitted. The degradation in the data rate can be ignored, since a significant PAPR reduction is obtained by using only one or two redundant bits. To be fair in viewing the potential loss in the data rate, the impact of using the SLM method on the average energy of the constellation should be considered as well. The Hadamard constellation has a zero shaping gain[4] due to its cubic boundary (shaping gain is computed using continuous approximation [49]). Numerical results show that applying the SLM method to the resulting cubic constellation results in a reduction in the average energy, reflected in a small, but positive shaping gain. This justifies the earlier claim that the reduction in the PAPR is achieved at no extra cost in terms of a reduction in the spectral efficiency and/or an increase in the average energy of the constellation.

2.5 Simulation Results

In this section, the simulation results for a complex baseband OFDM system with $N = 128$ subchannels employing 16-QAM are presented by using 10^7 randomly generated OFDM symbols. First, the PAPR performance of the Hadamard constellation is showed. The next step is then to show the capability of the SLM technique, when it is applied to the Hadamard constellation to achieve further PAPR reduction. The simulation results are presented as the Complementary Cumulative Density Function (CCDF) of the PAPR of the OFDM signals, expressed as follows:

$$\mathrm{CCDF}\left\{\mathrm{PAPR}(\mathbf{y})\right\} = \mathbb{P}\left\{\mathrm{PAPR}(\mathbf{y}) > \gamma\right\}. \tag{26}$$

This equation can be interpreted as the probability that the PAPR of a symbol block exceeds some clip level γ (it is referred to symbol clip probability [72]).

According to (8) and (9), the continuous PAPR can be estimated by the IFFT of the zero padded sequence of length LN. Results for the oversampling to $L = 1, 2,$ and 4 are shown in Figure 5. The continuous PAPR can be approximated by an oversampling factor of $L = 4$. As mentioned in [136, 109, 153], further oversampling will result in minor changes. The PAPR reduction of more than 4dB at 10^{-5} symbol clip probability is achieved.

Figure 6 shows the PAPR of an OFDM signal using the Hadamard constellation with different numbers of block length N. The effect of the constellation size is also investigated. It is observed that the achieved PAPR is rather insensitive to the constellation size, see Figure 7. The Symbol Error Rate (SER) of the proposed method and that of a conventional OFDM system are compared. As shown in Figure 8, the gap is minimal.

Figure 9 shows the simulation results of applying SLM technique to the Hadamard constellation. As illustrated in Figure 9, using only one bit of redundancy in 4×128 bits per block of

[4] Shaping gain is defined as the relative reduction in the required average energy for a given number of constellation points with respect to a cubic constellation [49].

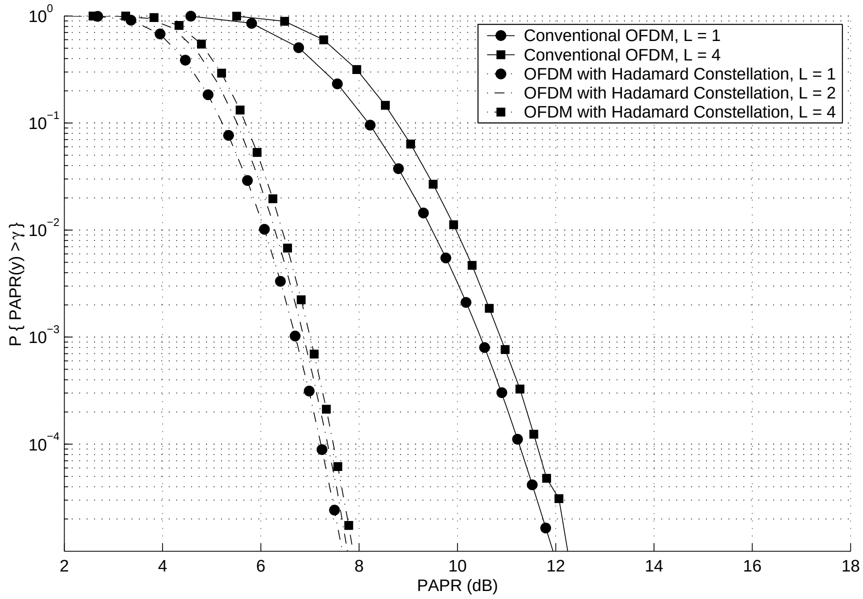

Figure 5: CCDF of PAPR for a Hadamard constellation with different over-sampling factors (128 channel OFDM system with 16-QAM constellation).

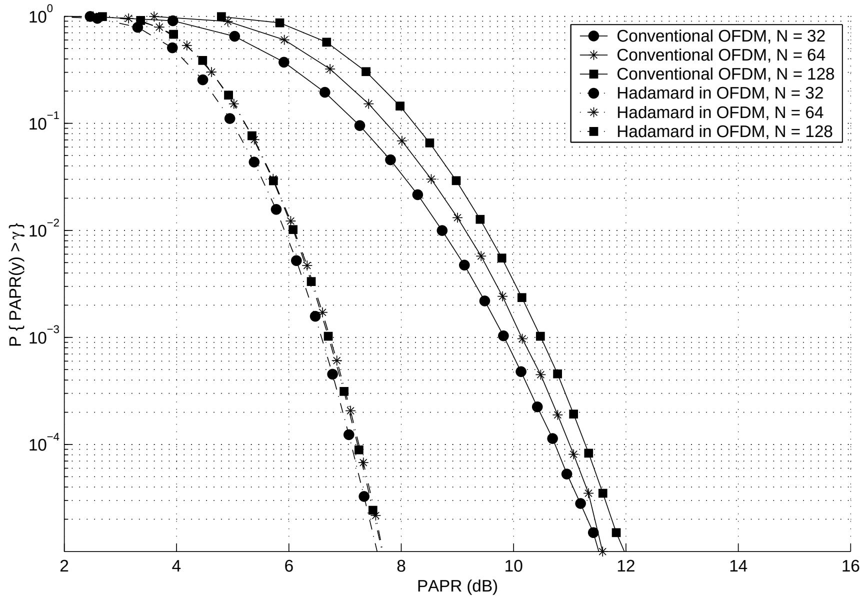

Figure 6: CCDF of PAPR for a Hadamard constellation in an N channel OFDM system employing 16-QAM constellation and $L = 1$.

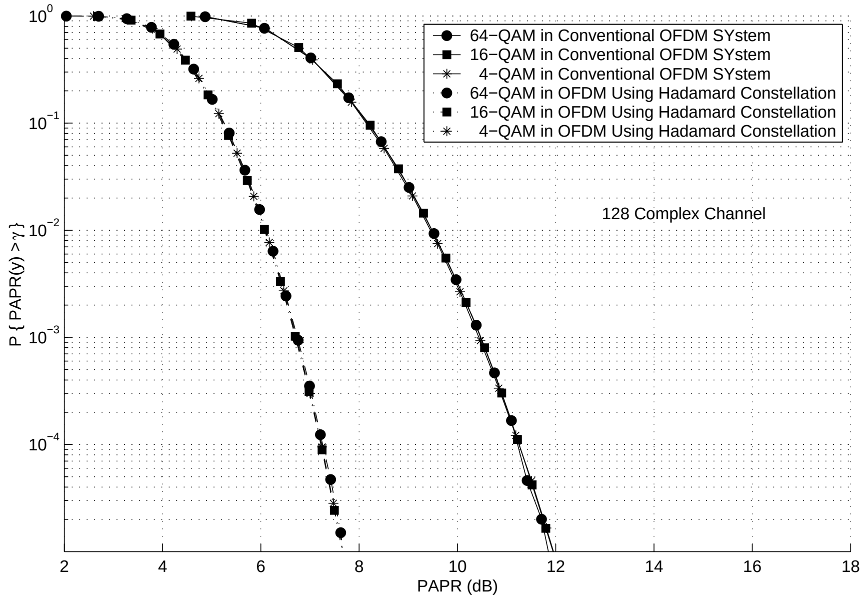

Figure 7: CCDF of PAPR for a Hadamard constellation in a 128 channel OFDM system employing different QAM constellations and $L = 1$.

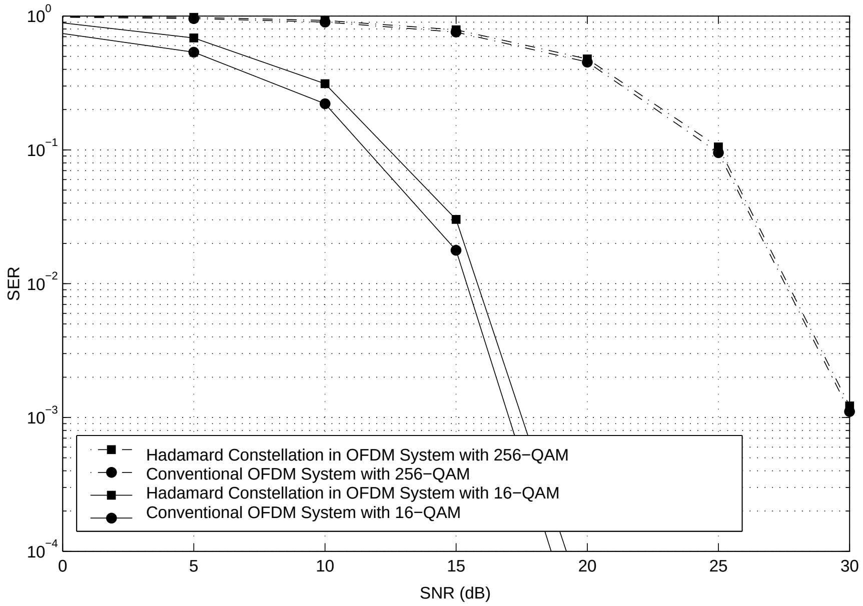

Figure 8: SER comparison for the proposed method in a 256 channel OFDM system employing 256-QAM and 16-QAM constellation.

a 128 FFT symbol[5] results in a 5.6dB reduction in the PAPR. Simulation results show that by employing more redundant bits the PAPR approaches its optimal value for a cubic constellation[6], namely $10\log_{10}(3)$.

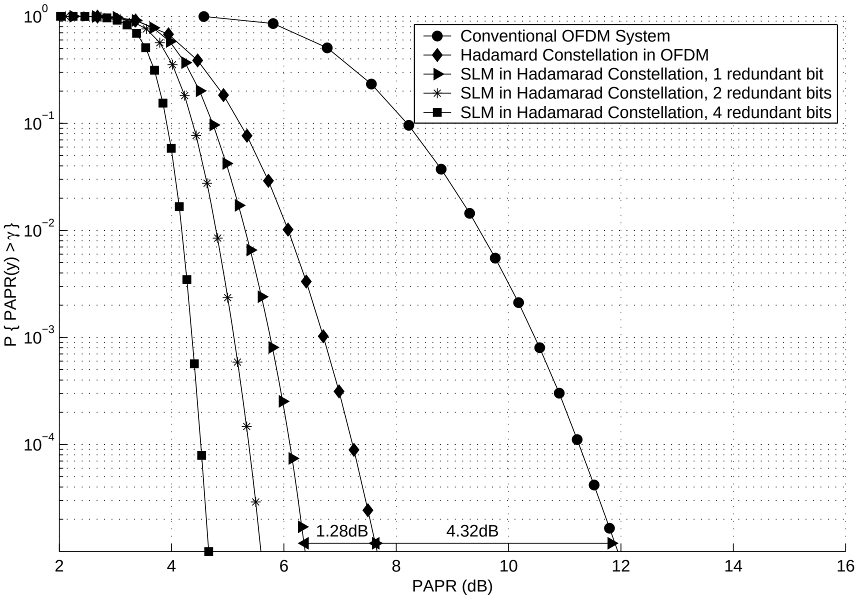

Figure 9: CCDF of PAPR by SLM method based on Hadamard constellation in a 128 channel OFDM system employing 16-QAM constellation.

2.5.1 Some Insight to the Achieved Performance

In a conventional OFDM system with N different subcarriers, the time domain samples can be approximated by zero mean Gaussian random variables, based on adopting the central limit theorem. Therefore, the amplitude of these samples has a Rayleigh distribution, and the CCDF of the PAPR of the OFDM signal can be approximated as follows [79]:

$$\mathbb{P}\{\mathrm{PAPR}(\mathbf{y}) > \gamma\} = 1 - (1 - e^{-\gamma})^{N}. \tag{27}$$

The use of N_s statistically independent vectors that have the same information for transmission in the SLM method changes the CCDF of the PAPR of the OFDM signal such that

$$\mathbb{P}\{\mathrm{PAPR}(\mathbf{y}) > \gamma\} = \left(1 - (1 - e^{-\gamma})^{N}\right)^{N_s}. \tag{28}$$

Therefore, in the logarithmic CCDF vs. PAPR graph, the slope of the curve is proportional to N_s (see Figure 10). By increasing the number of vectors with the same information, the

[5] By using 16-QAM in a 128 channel OFDM system, there are $16^{128} = 2^{4\times128}$ constellation points.

[6] The PAPR of a cubic constellation is computed using continuous approximation.

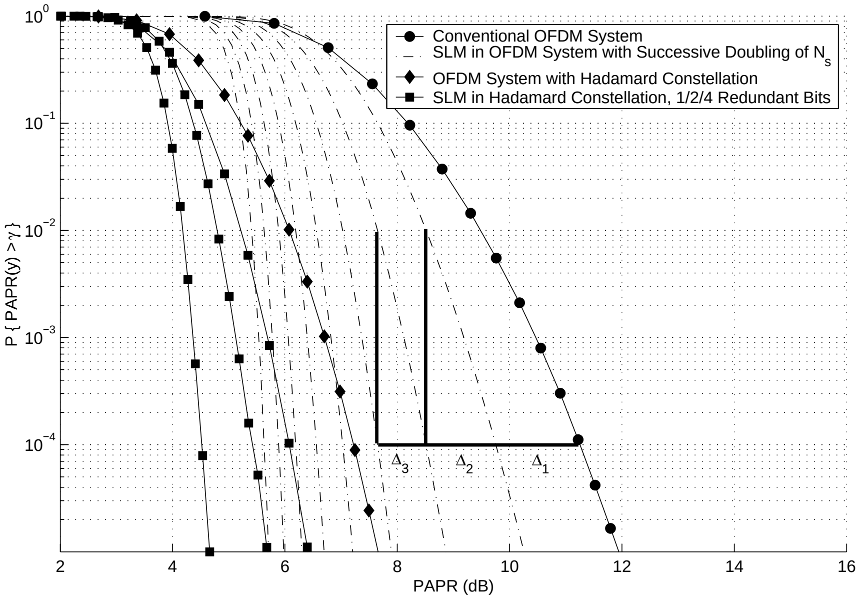

Figure 10: CCDF of PAPR in a 128 channel OFDM system with SLM method using different number of redundant bits, $L = 1$.

corresponding slope increases. Thus, the major PAPR reduction is achieved by the first few redundant bits, as shown in Figure 10 ($\Delta_1 > \Delta_2 > \cdots$). In other words, a saturation effect on the PAPR reduction is resulted by increasing r_s. This is the reason that the SLM technique have been applied to the Hadamard constellation. As mentioned in Section 2.3, the method employing only the Hadamard constellation considerably reduces the PAPR. By adopting the Hadamard constellation in the proposed SLM method, not only the PAPR can be lowered considerably, but also the slope of the CCDF vs. PAPR curve can be approximately maintained. This results in a considerably lower PAPR by using a small number of redundant bits before reaching the saturation.

2.5.2 Comparison

In numerical simulations, the system parameters have been selected to be compatible with some recent works on PAPR reduction reported in [17, 104, 135, 73, 79]. As a complexity measurement, the main complexity of the proposed method is due to the encoding algorithm and the multi-IFFT computations in the SLM technique. The complexity of the encoding algorithm is in the matrix multiplications of (11). As mentioned in Section 2.2, all the elements of the Hadamard matrix and its SNF decomposition matrices are $+1, -1$, or 0, and consequently, these operations can be easily implemented using a butterfly structure. Note that in the SLM technique, for each of the N_s time domain signals, one IFFT should be computed.

In [17], an SLM method based on multiplying the constellation point by N_s different pseudo-random but fixed vectors is introduced. For the same system as ours, with $N_s = 4$ different vectors, a PAPR reduction of 3dB is gained at the symbol clip probability close to 10^{-5}. However, for the same symbol clip rate and $N_s = 4$, a 6dB reduction is achieved by using the proposed SLM method. Also, the complexity of this algorithm is comparable with the method in [17]. Note that in [17] some side information (with high sensitivity to channel error) needs to be transmitted.

Another approach, similar to [17], is introduced for the SLM in [79]. The authors have introduced this method for MIMO[7]-OFDM systems. The simulation results in [79] are similar to [17] (the relative comparison between the proposed method and the one in [17] is explained earlier).

The tone reservation [135] is a well known method for PAPR reduction in multicarrier systems, provided that it can quickly converge to a good solution. An efficient approximation for the tone reservation approach with a faster convergence is developed in [73]. The complexity of [73] is comparable with ours; however, about 3dB lower PAPR than that in [135] or [73] is achieved for similar system parameters. Note that in the tone reservation method, some tones are reserved for the PAPR reduction and some of the tones are not used for data transmission, implying a loss in the data rate. Note that [73] reduces the PAPR by solving a min-max problem. This problem is solved by an interior-point method which requires a descent direction and a constraint to find the solution recursively.

In [107], for a 256 complex channel OFDM system employing 256-QAM, a 4.5dB reduction in the PAPR is obtained using a trellis shaping technique. In the proposed method, for a 128 complex channel OFDM system employing a 128-QAM, a 6dB reduction is gained. In [107], the main complexity is in finding the path with minimum cost through a trellis diagram (this complexity is considerably higher than that of a Viterbi decoder). However, the author investigates methods to reduce this complexity by window truncation and sacrificing PAPR reduction, but still the overall complexity in [107] is significantly higher as compared to the method proposed here.

There is no comparison with [75], as the method in [75] relies on using the SNF of the IFFT matrix which is not known. Indeed, computing this SNF decomposition would be an interesting open problem. If this matrix were available, the resulting PAPR reduction in [75] would be asymptotically equal to the optimum value of $10\log_{10}(3)$. Also, as mentioned in Section 2.2, the computational complexity of the encoding algorithm of the constellation based on the IFFT matrix is $O(N^2)$, while the complexity for the encoding of the Hadamard constellation in the butterfly structure is $O(\frac{3}{2}\log_2(N))$ [75].

[7] Multi-Input Multi-Output

2.6 Conclusion

A constellation shaping method is proposed that achieves a substantial reduction in the PAPR in an OFDM system with a low complexity. The boundary of the proposed constellation is along the basis defined by the Hadamard matrix in the transform domain. An SLM technique is applied to this constellation to further reduce the PAPR of the OFDM signal. The proposed scheme significantly outperforms other PAPR reduction techniques reported in the literature, without any loss in terms of the energy and/or spectral efficiency.

CHAPTER 3

PRECODING IN MULTIPLE-ANTENNA BROADCAST SYSTEMS

***Abstract* –** In this chapter, the average transmit energy of a multiple antenna broadcast system with channel inversion is investigated. It is shown that the reduction in average transmit energy, due to shaping, can be significantly higher than the conventional gains in reduction that are mentioned in the literature. In order to approach this gain, a Selective Mapping (SLM) technique is introduced. Using the strong literature in quantization, the gain that the SLM technique can provide is derived. The proposed SLM method can be implemented by lattice decomposition techniques.

3.1 *Introduction*

Multiuser Multi-Input Multi-Output (MIMO) antenna systems have received much attention due to achieving a very high data rate. In a MIMO broadcast system, the sum-capacity grows linearly with the minimum number of the transmit and receive antennas [140]. However, achieving the sum-capacity in a MIMO broadcast system is more complicated than that in a MIMO point-to-point system, since each user must decode its signal independently from the other users.

Generally, there have been two different approaches in the literature for implementing a broadcast system. To achieve the promised sum capacity while having independent decoding ability, some information theoretic schemes provide a precoding method based on interference cancelation methods, e.g. [154] or [155]. The alternative precoding approach is based on using channel inversion technique, e.g. [114] or [129].

In precoding methods based on interference cancelation, the channel is first transformed into a series of parallel sub-channels with non-causally known interference. Then, having the interference knowledge, a signal is transmitted over these sub-channels by using a trellis precoder [154] or nested lattice codes [155]. These methods are mainly based on the idea of dirty-paper-coding theorem [29].

Channel inversion technique separates the data for different users at the transmitter side. By assuming the complete knowledge of the channel, the data that is supposed to be received at the receiver side is multiplied by the inverse of the channel and is transmitted. This operation guarantees the independent decoding at the receiver side. However, the average transmit energy of channel inversion technique is high, as the channel matrix is nor orthogonal. In [115], the authors have introduced a vector perturbation technique based on channel inversion which has a good performance in terms of symbol error rate. Nonetheless, this technique requires a lattice decoder which is an NP-hard problem. The lattice decoder is replaced by a lattice-reduction-aided decoding method in [129], resulting in reducing the average transmitted energy, with a reasonable complexity.

In this chapter, the average transmit energy of a multiple antenna broadcast system with channel inversion is investigated. By using channel inversion, the decoding at the transmitter side is independent of the channel matrix. Due to the fact that the channel is not orthogonal, the reduction in the average transmit energy due to constellation shaping is significantly higher than the regular shaping gains reported in literature, e.g. [115].

In the sequel, a Selective Mapping (SLM) technique is introduced to reduce the average energy of the transmitted signal in a broadcast system. The average transmit energy of the proposed SLM technique is also derived. Since finding the optimum average transmit energy is difficult, a lower bound on the average transmit energy is calculated. This lower bound is based on the assumption that the users at the receiver side can cooperate with each other. Simulation results show that the proposed SLM method has an average transmit energy which is close to this lower bound. Moreover, the proposed method outperforms the other known techniques in the literature. In order to implement the SLM method effectively, lattice decomposition techniques are proposed.

3.2 System Model

A multiple antenna broadcast system with $\tilde{N}$ transmit antennas and K users, where user $i = 1, \cdots, K$ is equipped with n_i antennas ($\tilde{M} = \sum_{i=1}^{K} n_i$), is modeled as

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{h}}_i \tilde{\mathbf{x}} + \tilde{\mathbf{w}}_i, \quad i = 1, \cdots, K \tag{29}$$

where $\tilde{\mathbf{y}}_i$ is the received vector of dimension n_i by user $i = 1, \cdots, K$, $\tilde{\mathbf{x}}$ is an $\tilde{N} \times 1$ data vector with $E\{||\tilde{\mathbf{x}}||\} = 1$, $\tilde{\mathbf{w}}_i$ is an $\tilde{n}_i \times 1$ complex additive white Gaussian noise vector with zero mean. The equations in (29) can be written as

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\tilde{\mathbf{x}} + \tilde{\mathbf{w}}, \tag{30}$$

where $\tilde{\mathbf{y}} = [\tilde{\mathbf{y}}_1^T, \cdots, \tilde{\mathbf{y}}_K^T]^T$, $\tilde{\mathbf{w}} = [\tilde{\mathbf{w}}_1^T, \cdots, \tilde{\mathbf{w}}_K^T]^T$, and $\tilde{\mathbf{H}} = \begin{bmatrix} \tilde{\mathbf{h}}_1^T & \vdots & \tilde{\mathbf{h}}_K^T \end{bmatrix}^T$. $\tilde{\mathbf{H}}$ is the $\tilde{M} \times \tilde{N}$ channel matrix composed of independent, identically distributed (i.i.d.) complex Gaussian random elements with zero mean and unit variance and $E\{\tilde{\mathbf{w}}\tilde{\mathbf{w}}^\star\} = \dfrac{1}{\rho}\mathbf{I}_{\tilde{M}}$, where the parameter ρ is the SNR per receive antenna.

To avoid using complex matrices, the system model (30) is represented by real matrices in (31).

$$\begin{aligned} \begin{bmatrix} \Re(\tilde{\mathbf{y}}) \\ \Im(\tilde{\mathbf{y}}) \end{bmatrix} &= \begin{bmatrix} \Re\left(\tilde{\mathbf{H}}\right) & \Im\left(\tilde{\mathbf{H}}\right) \\ -\Im\left(\tilde{\mathbf{H}}\right) & \Re\left(\tilde{\mathbf{H}}\right) \end{bmatrix} \begin{bmatrix} \Re(\tilde{\mathbf{x}}) \\ \Im(\tilde{\mathbf{x}}) \end{bmatrix} + \begin{bmatrix} \Re(\tilde{\mathbf{w}}) \\ \Im(\tilde{\mathbf{w}}) \end{bmatrix} \\ \Rightarrow \mathbf{y} &= \mathbf{H}\mathbf{x} + \mathbf{w}, \end{aligned} \tag{31}$$

where $\mathbf{y}$ is the *received vector*, $\mathbf{x}$ is the *transmit vector*, $\bar{M} = 2\tilde{M}$, and $\bar{N} = 2\tilde{N}$.

3.2.1 Channel Inversion

It is assumed that the channel state information is available at the transmitter side and the channel inversion precoding is performed on the input data. In this case, the decoding at the receive side is independent of the channel matrix.

By assuming channel inversion, the transmitter sends

$$\mathbf{s} = \mathbf{H}^{+}\mathbf{u}, \tag{32}$$

where $\mathbf{H}^{+} = \mathbf{H}^{*}(\mathbf{H}\mathbf{H}^{*})^{-1}$, $\mathbf{H}^{*}$ is the Hermitian of $\mathbf{H}$, $\mathbf{u} = \left[\mathbf{u}^{\mathbf{1T}}; \cdots; \mathbf{u}^{\mathbf{KT}}\right]^{\mathbf{T}}$ is the data vector, i.e. $\mathbf{u}^i$ is the data for the i^{th} user, and $\mathbf{s}$ is the transmitted signal before the normalization. For the simplicity, it is assumed $\bar{M} = \bar{N} := M$, i.e. the transmitted signal is

$$\mathbf{s} = \mathbf{H}^{-1}\mathbf{u}. \tag{33}$$

As in [114], the normalized transmitted signal would be $\mathbf{x} = \dfrac{\mathbf{s}}{\sqrt{E\{\gamma\}}}$, where $\gamma = ||\mathbf{s}||^2$ is the energy of the transmitted signal, called *transmit energy*. Note that by using channel inversion, the decoding at the transmitter side is independent of the channel matrix. Substituting (33) in (31) results in

$$\mathbf{y} = \frac{1}{\sqrt{E\{\gamma\}}}\mathbf{u} + \mathbf{n}. \tag{34}$$

The problem arises when $\mathbf{H}$ is poorly conditioned and γ becomes very large, resulting in a high power consumption. This situation occurs when at least one of the eigenvalues of $\mathbf{H}$ is very small which results in vectors with large norms as the columns of $\mathbf{H^{-1}}$.

3.3 Average Transmit Energy of Probabilistic Constellations

In a multiple antenna broadcast system, a proper input constellation should be designed such that given a minimum distance two conditions are satisfied: (i) data can be decoded independently at the receivers (*independency condition*), and (ii) the average transmit energy is as low as possible (*energy condition*). The design of the constellation is known as the shaping.

Lattice constellation shaping technique deals with the problem of finding a finite set of points from an M dimensional lattice $\mathbf{\Lambda}$ that lies within a finite region $\mathcal{R} \subset \mathbb{R}^M$. This constellation is known as a lattice code. If $\mathbb{C}$ is a lattice code of reasonably large size, then the distribution of its points in M dimensional space is well approximated by a uniform continuous distribution over the region $\mathcal{R}$ (the continuous approximation) [49].

Having a uniform distribution over region $\mathcal{R}$ does not guarantee a uniform distribution over each dimension. In other words, if $\mathbf{u}$ is selected uniformly over $\mathcal{R}$, each element of $\mathbf{u}$ may have a nonuniform distribution. Throughout this chapter, the probability distribution of the elements of $\mathbf{u}$ is called *marginal probability distribution* of $\mathbf{u}$.

Let $\mathbf{Q} := \left(\mathbf{H^{-1}}\right)^T \mathbf{H^{-1}} = \mathbf{U\Lambda U^T}$, where $\mathbf{U}$ is the unitary matrix of eigenvectors of $\mathbf{Q}$ and $\mathbf{\Lambda}$ is the diagonal matrix of the corresponding eigenvalues, $\lambda_i,\ i = 1, \cdots, M$. Assume $\mathbf{u} \in \mathcal{R}$ be a random vector with mean $E\{\mathbf{u}\} = \boldsymbol{\mu}$ and the correlation matrix $E\{(\mathbf{u}-\boldsymbol{\mu})(\mathbf{u}-\boldsymbol{\mu})^T\} = \mathbf{\Sigma} > 0$. In general, the average transmit energy of the signal $\mathbf{s} = \mathbf{H}^{-1}\mathbf{u}$ can be written as

$$\begin{aligned} E\{\gamma\} &= E\left\{\mathbf{u}^T \left(\mathbf{H^{-1}}\right)^T \mathbf{H^{-1}u}\right\} = E\left\{\mathbf{u}^T\mathbf{Qu}\right\} \\ &= \mathrm{trace}\left(E\left\{\mathbf{u}^T\mathbf{Qu}\right\}\right) = E\left\{\mathrm{trace}\left(\mathbf{u}^T\mathbf{Qu}\right)\right\} \\ &= E\left\{\mathrm{trace}\left(\mathbf{Quu}^T\right)\right\} = \mathrm{trace}\left(\mathbf{Q}E\left\{\mathbf{uu}^T\right\}\right) \\ &= \mathrm{trace}\left(\mathbf{Q}(\Sigma + \boldsymbol{\mu\mu}^T)\right) \\ &= \mathrm{trace}(\mathbf{Q\Sigma}) + \boldsymbol{\mu}^T\mathbf{Q}\boldsymbol{\mu}. \end{aligned} \tag{35}$$

3.3.1 Average Transmit Energy in Broadcast Systems

In the literature, there have been different precoding schemes for broadcast systems, e.g. [60, 114, 115, 144, 129, 20]. Different approaches aim to reduce the average transmit energy, while introducing minimum interference among users. In order to achieve this goal, the focus is on choosing different shaping regions $\mathcal{R}$ and their corresponding marginal probability distributions for the vector $\mathbf{u}$ in (31).

Assume that the vector $\mathbf{u}$ is selected from a lattice code $\mathbb{C}$ with shaping region $\mathcal{R}$, $\mathbf{u} \in \mathbb{C}$. Therefore, the transmitting vector $\mathbf{s}$ is selected from a lattice code $\mathbb{C}'$ with shaping region $\mathcal{R}'$, where $\mathcal{R}' = \{\mathbf{s}|\mathbf{s} = \mathbf{H^{-1}u}, \forall \mathbf{u} \in \mathcal{R}\}$. For large enough constellation sizes, continuous approximation implies that the vectors $\mathbf{u}$ and $\mathbf{s}$ are selected uniformly over regions $\mathcal{R}$ and $\mathcal{R}'$, respectively. It is assumed that the region $\mathcal{R}$ has a volume $\text{Vol}(\mathcal{R}) = \mathbb{V}$, resulting in the entropy of $\log \mathbb{V}$. Note that when there are M independent dimensions, the entropy per each dimension is $\mathcal{H} = \frac{1}{M} \log \mathbb{V}$.

In the Appendix B, the average transmit energy for different known methods in the literature is calculated, based on the probabilistic view point of (35). In all these cases, it is assumed that the vector $\mathbf{u}$ is distributed uniformly over a hypercube centered at the origin with side length of $2A$, $\mathcal{R} = C_M(0, 2A)$. By using simple channel inversion, the region $\mathcal{R}'$ is an orthotope centered at the origin and along the eigenvectors of $\mathbf{Q}$. In the regularization method [114], before the channel inversion, the vector $\mathbf{u}$ is multiplied by a linear transformation. This reduces the average transmit energy at the cost of introducing interference among users (see Appendix B). In the perturbation technique [115], by increasing the constellation points and using a non-linear modulo operation, the region $\mathcal{R}'$ is the Voronoi region of the matrix $\tau\mathbf{H}^{-1}$, where τ is the constellation length in each real dimension. In summary, different methods try to change the region $\mathcal{R}'$ such that, by keeping the independency condition, the average transmit energy is reduced.

3.3.2 Lower Bound on the Average Transmit Energy

It is shown that different precoding methods change the region $\mathcal{R}'$ (and its corresponding marginal probability distribution) in order to reduce the average transmit energy. Theoretically, it is well known that by conventional shaping methods (case IV in Appendix B - vector $\mathbf{u}$ with Gaussian random elements), there is at most $1.53dB$ more reduction in the average transmit energy compared to the simple channel inversion. The question is how much further the average energy can be reduced.

As finding the optimum average transmit energy is difficult, a lower bound for the average transmit energy is desirable. A broadcast system can be considered as a MIMO point-to-point system with an extra independency condition, see [117]. Therefore, in order to find a lower bound, the energy condition is considered and the independency at the receivers is ignored. Following theorem finds the region $\mathcal{R}$ (equivalently $\mathcal{R}'$) in such a system by minimization of the average transmit energy in (35) over different marginal probability distributions.

Theorem 3 *Let $\mathbf{u} = [u_1, u_2, \cdots, u_M] \in \mathbb{R}^M$ be a random vector with probability distribution $f(u_1, u_2, \cdots, u_M)$, mean $E\{\mathbf{u}\} = \boldsymbol{\mu}$, and the correlation matrix $E\{(\mathbf{u} - \boldsymbol{\mu})(\mathbf{u} - \boldsymbol{\mu})^T\} = \boldsymbol{\Sigma} > 0$, in* *the system model introduced in (31). Let $\mathcal{H}(\mathbf{u})$ denote the entropy of the data vector $\mathbf{u}$. Then, a multivariate Gaussian random vector $\mathbf{u}$ with $\boldsymbol{\mu} = \mathbf{0}$ and the covariance matrix*

$$\boldsymbol{\Sigma} = \sqrt[M]{\Pi \lambda_i} \sigma^2 \mathbf{Q}^{-1} \tag{36}$$

minimizes the energy of the transmit signal given a fixed entropy $\mathcal{H}(\mathbf{u}) = \log(\mathbb{V})$, where σ^2 is the variance of a Gaussian random variable with entropy $\mathcal{H} = \frac{1}{M}\log(\mathbb{V})$.

Proof: See Appendix A. ■

This choice of $\boldsymbol{\Sigma}$ suggests that the lower bound for the average transmit energy among signals with different probability distributions is

$$E_{bound} = E\{\gamma\} = M \sqrt[M]{\Pi \lambda_i} \sigma^2 \tag{37}$$

The average transmit energy in (37) corresponds to the average energy of an M-dimensional Gaussian random vector with i.i.d. elements with zero mean and variance $\mathcal{R}_{eq}^2 = \sqrt[M]{\Pi \lambda_i} \sigma^2$ (each element of vector $\mathbf{v} = \sqrt{\boldsymbol{\Lambda}} \mathbf{U}\mathbf{u}$ has a Gaussian distribution with zero mean and variance $\mathcal{R}_{eq}^2$). Here, the shaping problem is to select the region $\mathcal{R}$ or $\mathcal{R}'$ to minimize the average transmit energy, while considering the independency condition.

3.4 Selective Mapping Precoding

In this section, a Selective Mapping (SLM) technique is introduced to reduce the average transmit energy. In the following, we first briefly review the principals of SLM technique.

3.4.1 An Overview on SLM Technique

The SLM technique generates a large set of data vectors that represent the same information, where the data vector resulting in the lowest energy is selected for transmission. It is assumed that originally the vector $\mathbf{s}$ is selected uniformly over region $\mathcal{R}$, where the volume of this region is fixed, $\text{Vol}(\mathcal{R}) = \mathbb{V}$. In order to provide multiple choices for the SLM method, the region is changed to $\overline{\mathcal{R}}$ such that the volume is increased to $\overline{\mathbb{V}}$. It is assumed that for each data vector there are N points, where

$$N = \frac{\overline{\mathbb{V}}}{\mathbb{V}}. \tag{38}$$

In other words, N i.i.d. samples of the vector $\mathbf{s}$ are generated in $\overline{\mathcal{R}}$, i.e. $\{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \cdots, \mathbf{s}^{(N)}\} \in \overline{\mathcal{R}}$, and $\mathbf{s}^l$ with the lowest energy is selected for transmission. In other words,

$$\gamma^{(l)} = \min\{\gamma^{(1)}, \gamma^{(2)}, \cdots, \gamma^{(N)}\},$$

where $\gamma^{(i)} = \|\mathbf{s}^{(i)}\|^2$. Consider the case that the samples of $\mathbf{s}$ are Gaussian random vectors, $N(0, \sigma^2\mathbf{I}_M)$. By increasing the volume (entropy)[1] from $\mathbb{V}$ to $\bar{\mathbb{V}}$ the variance would increase from σ^2 to σ'^2, where

$$\frac{\bar{\mathbb{V}}}{\mathbb{V}} = \left(\frac{\sigma'}{\sigma}\right)^M = N. \tag{39}$$

The energy $\gamma = \mathbf{u}^T\mathbf{Q}\mathbf{u}$ is a quadratic expression of random vector $\mathbf{u}$. There is a lot of research on the probabilistic relations between γ and $\mathbf{u}$ [90, and ref. therein]. In order statistic references [35, and ref. therein], the relations between the probability behavior of $\gamma^{(l)}$ and γ are discussed.

3.4.2 SLM with Gaussian Variables in Time Domain

In this section, the aim is to approach the lower bound on the average transmit energy in a broadcast system with a theoretical point of view. The first step is to provide a high dimensional constellation for each user. It is known that when a random vector is uniformly selected in a D-dimensional sphere centered at the origin with radius $\sqrt{D}\sigma$, i.e. $\mathcal{B}_D(0, \sqrt{D}\sigma)$, in the limit of $D \longrightarrow \infty$, each dimension has a Gaussian distribution with zero mean and variance σ^2.

Consider a time frame of L consecutive symbols for each user. If the i^{th} user has n_i antennas, the L symbols corresponding to user i can be considered to be selected in a $D_i = 2Ln_i$ dimensional space. In other words, the D_i-dimensional vector $\mathbf{u}_{1:L,i} = [\mathbf{u}_{1,i}^T, \mathbf{u}_{2,i}^T, \cdots, \mathbf{u}_{L,i}^T]^T$ is selected from the region $\mathcal{R}_i$ and it corresponds to L vectors $\mathbf{u}_{t,i}$ of dimension $2n_i$ which are selected at time slots $t = 1, \cdots, L$ for user i. Provided that the vector $\mathbf{u}_{1:L,i}$ is selected uniformly over a D_i-dimensional sphere $\mathcal{R}_i = \mathcal{B}_{D_i}(0, \sqrt{D_i}\sigma_i)$, the elements of this vector are i.i.d. Gaussian elements with variance σ_i^2 (in the limit of $L \longrightarrow \infty$). At each time slot $t = 1, \cdots, L$, the M-dimensional vector $\mathbf{u}_t = [\mathbf{u}_{t,1}^T, \mathbf{u}_{t,2}^T, \cdots, \mathbf{u}_{t,K}^T]^T$, consisting the data vectors for different users, is selected and the vector $\mathbf{s}_t = \mathbf{H}^{-1}\mathbf{u}_t$ is transmitted.

The next step is to apply Selective Mapping (SLM) to provide multiple choices for transmitted information. In D_i dimensional space of each user, the constellation is expanded to $\bar{\mathcal{R}}_i$ such that the volume (also the variance) of the new constellation follows (39), i.e.

$$\frac{\mathrm{Vol}(\bar{\mathcal{R}}_i)}{\mathrm{Vol}(\mathcal{R}_i)} = \left(\frac{\sigma_i'}{\sigma_i}\right)^{D_i} = N_i. \tag{40}$$

For transmitting any D_i dimensional data vector to each user, N_i i.i.d. random vectors are selected uniformly over the D_i-dimensional region of $\bar{\mathcal{R}}_i$, that is $\mathcal{U}_{1:L}^{(i)} = \{\mathbf{u}_{1:L,i}^{(1)}, \mathbf{u}_{1:L,i}^{(2)}, \cdots, \mathbf{u}_{1:L,i}^{(N_i)}\} \in$

[1] The discussion in this section and the following sections are based on the increase in the entropy, $\mathcal{H}(\mathbf{u}) = \log(\mathbb{V})$ and $\mathcal{H} = \frac{1}{2}\log(2\pi e\sigma^2)$. In this approach, the minimum distance of the constellation is fixed. An alternative approach can be considered based on the capacity viewpoint, where $C = \frac{1}{2}\log(1 + SNR)$ and SNR is defined based on σ^2 and noise power. This guarantees a free error transmission, as long as the size of the increased constellation follows the capacity formula.

$\bar{\mathcal{R}}_i$, where $\mathbf{u}_{1:L,i}^{(j)} = [\mathbf{u}_{1,i}^{(j)T}, \mathbf{u}_{2,i}^{(j)T}, \cdots, \mathbf{u}_{L,i}^{(j)T}]^T$ corresponds to data vectors for the i^{th} user at time slot $t = 1 \cdots, L$ for $j = 1, \cdots, N_i$. These N_i data vectors are equivalent and they represent the same information.

Considering all K users, $\{\mathcal{U}_{1:L}^{(1)}, \mathcal{U}_{1:L}^{(2)}, \cdots, \mathcal{U}_{1:L}^{(K)}\}$, is equivalent to the M-dimensional data vectors of users in a time frame of length L. For any data vector in the space of each user, there are N_i equivalent data vectors (representing the same information). Therefore, in the ML dimensional space of all users, there are $N = \prod N_i$ equivalent data vectors. The corresponding transmitted signals are $\{\mathbf{s}_{1:L}^{(1)}, \mathbf{s}_{1:L}^{(2)}, \cdots, \mathbf{s}_{1:L}^{(N)}\}$, where $\mathbf{s}_{1:L}^{(j)} = \{\mathbf{s}_1^{(j)}, \mathbf{s}_2^{(j)}, \cdots, \mathbf{s}_L^{(j)}\}$, $\mathbf{s}_t^{(j)} = [\mathbf{s}_{t,1}^{(j)T}, \mathbf{s}_{t,2}^{(j)T}, \cdots, \mathbf{s}_{t,K}^{(j)\,T}]$, $\mathbf{s}_t^{(j)} = \mathbf{H}^{-1}[\mathbf{u}_{t,1}^T, \mathbf{u}_{t,2}^T, \cdots, \mathbf{u}_{t,K}^T]^T$ and $\mathbf{u}_{t,i} \in \mathcal{U}_t^{(i)}$, where $\mathcal{U}_t^{(i)}$ is the set of symbols at time slot t for the user i in $\mathcal{U}_{1:L}^{(i)}$, i.e. $\mathcal{U}_t^{(i)} = \{\mathbf{u}_{t,i}^{(1)}, \mathbf{u}_{t,i}^{(2)}, \cdots, \mathbf{u}_{t,i}^{(N_i)}\}$. The symbol energy for each of N equivalent ML-dimensional data vectors are defined as

$$\gamma^{(j)} \triangleq \frac{1}{L}\sum_{t=1}^{L} \|\mathbf{s}_t^{(j)}\|^2, \tag{41}$$

for $j = 1, \cdots, N$. The data vector with the lowest energy among the N equivalent data vectors is selected for transmission, i.e.

$$\gamma^{(l)} = \min_{j=1,\cdots,N} \gamma^{(j)}.$$

The probabilistic behavior of $\gamma^{(l)}$ is desired. In the next section, the average transmit energy resulted by using the proposed SLM method is analyzed.

3.5 Average Transmit Energy Analysis in SLM

In this section, the effect of applying the proposed SLM technique in broadcast systems is analyzed. First, applying the proposed SLM method in the space of each user is considered. Then, the results is generalized for a broadcast system with multiple users.

3.5.1 Effect of SLM for a Single User

In order to analyze this technique, first, the effect of the SLM method is considered in the m-dimensional space of a given user. Assume that the relation between the m-dimensional vector $\mathbf{u}$ and the transmit vector $\mathbf{s}$ is defined as $\mathbf{s} = \mathbf{G}\mathbf{u}$, where $\mathbf{G}$ is a known $n \times m$ matrix. Considering a time frame of L consecutive symbols, the vector $\mathbf{u}_{1:L} = [\mathbf{u}_1^T, \mathbf{u}_2^T, \cdots, \mathbf{u}_L^T]^T$ is selected over the Lm-dimensional space of $\mathcal{R}$.

For implementing an SLM method, N i.i.d. Lm-dimensional symbols are generated uniformly over the expanded region $\bar{\mathcal{R}}$, i.e. $\{\mathbf{u}_{1:L}^{(1)}, \mathbf{u}_{1:L}^{(2)}, \cdots, \mathbf{u}_{1:L}^{(N)}\} \in \bar{\mathcal{R}}$, where $\mathrm{Vol}(\bar{\mathcal{R}}) = N\mathrm{Vol}(\mathcal{R})$. Each vector $\mathbf{u}_{1:L}^{(i)} = [\mathbf{u}_1^{(i)T}, \mathbf{u}_2^{(i)T}, \cdots, \mathbf{u}_L^{(i)T}]^T$ corresponds to L vectors $\mathbf{u}_j^{(i)}$ of dimension m which will be sent at time slots $j = 1, \cdots, L$. At each time slot, the transmit vector is defined as

$\mathbf{s}_j^{(i)} = \mathbf{G}\mathbf{u}_j^{(i)}$, resulting in D-dimensional vector $\mathbf{s}_{1:L}^{(i)} = (\mathbf{I}_L \otimes \mathbf{G})\mathbf{u}_{1:L}^{(i)}$, where $D = Ln$. Among the corresponding N transmit vectors $\mathbf{s}_{1:L}^{(j)}$, the vector $\mathbf{s}_{1:L}^{(l)}$ with the lowest transmit energy is selected for transmission. In other words,

$$l = \arg \min_{1 \le i \le N} \|\mathbf{s}_{1:L}^{(i)}\|^2. \tag{42}$$

In order to simplify the notations in this subsection, $\mathbf{s}_{1:L}^{(i)}$ and $\mathbf{u}_{1:L}^{(i)}$ are denoted by $\mathbf{s}^{(i)}$ and $\mathbf{u}^{(i)}$, respectively. Let $\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \cdots, \mathbf{s}^{(N)}$, be i.i.d. $\mathbb{R}^D$-valued random variables with distribution F, i.e.

$$\begin{aligned} F(\mathbf{v}) &= \mathbb{P}\{\mathbf{s}^{(i)} \le \mathbf{v}\} \\ &= \mathbb{P}\{s_1^{(i)} \le v_1, \cdots, s_D^{(i)} \le v_D\} \quad i = 1, \cdots, N, \end{aligned} \tag{43}$$

where $\mathbf{v} = (v_1, \cdots, v_D) \in \mathbb{R}^D$. For any region $\mathcal{R}$, the probability $F(\mathcal{R})$ is the probability that there is at least one code point in the region $\mathcal{R}$, i.e.

$$F(\mathcal{R}) = \int_{\mathcal{R}} F(d\mathbf{y}).$$

Define the r^{th} *moment of the transmitted signal* as

$$\gamma_{r,N}^F = \min_{1 \le i \le N} \|\mathbf{s}^{(i)}\|^r, \tag{44}$$

where based on the previous notation $\gamma_l = \gamma_{2,N}^F$. In this part, the asymptotic probabilistic behavior of $\gamma_{r,N}^F$, when $L \longrightarrow \infty$ (or equivalently $D \longrightarrow \infty$) and $N \longrightarrow \infty$, is investigated. Specifically, the average transmit energy in the SLM technique is calculated.

Theorem 4 *Let $\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \cdots, \mathbf{s}^{(N)}$, be i.i.d. $\mathbb{R}^D$-valued random variables with distribution F. Then,*

$$\lim_{N \to \infty} E\left\{N^{\frac{r}{D}} \gamma_{r,N}^F\right\} = B_D^{-\frac{r}{D}} \Gamma(1 + \frac{r}{D}) g_\rho^{-\frac{r}{D}} \tag{45}$$

where $B_1 = 2$, $B_D = Vol\,(\mathcal{B}_D(0,1)) = \pi^{D/2}/\Gamma(1 + D/2)$ for $D = 2, \cdots$, and g_ρ is defined for any $\rho > 0$ as

$$g_\rho := \inf_{\delta \in (0,\rho]} \frac{F\,(\mathcal{B}_D(0,\delta))}{Vol\,(\mathcal{B}_D(0,\delta))}.$$

Proof: See Appendix A. ■

Now, consider the special case of uniform distribution. When there is a large lattice code, it can be assumed that there is a uniform distribution over the region where the lattice code is defined. Applying SLM technique, over a region with uniform distribution results in the following average for the r^{th} moment of the transmitted signal.

Theorem 5 *Let $\bar{\mathcal{R}}' \subset \mathbb{R}^D$ be a compact set with volume Vol($\bar{\mathcal{R}}'$) and let $\mathbf{s}^{(1)}, \cdots, \mathbf{s}^{(N)}$ be i.i.d. random variables with uniform distribution over $\bar{\mathcal{R}}'$. Then,*

$$\lim_{N\to\infty} E\left\{N^{\frac{r}{D}}\gamma^{F}_{r,N}\right\} = B_D^{-\frac{r}{D}}\Gamma(1+\frac{r}{D})Vol(\bar{\mathcal{R}}')^{\frac{r}{D}}. \tag{46}$$

Proof: Let F be a uniform distribution over $\bar{\mathcal{R}}'$, i.e. $F = U(\bar{\mathcal{R}}')$. Therefore,

$$F\left(\mathcal{B}_D(0, \frac{\nu^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right) = \frac{\mathrm{Vol}\left(\mathcal{B}_D(0, \frac{\nu^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)}{\mathrm{Vol}(\bar{\mathcal{R}}')}, \tag{47}$$

and

$$g_\rho = \inf_{\delta\in(0,\rho]} \frac{F\left(\mathcal{B}_D(0,\delta)\right)}{\mathrm{Vol}\left(\mathcal{B}_D(0,\delta)\right)} = \frac{1}{\mathrm{Vol}(\bar{\mathcal{R}}')}. \tag{48}$$

Substituting (48) in (45) completes the proof. ■

Note that the data vectors $\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(N)}$ are i.i.d. random variables which are selected uniformly over a region $\bar{\mathcal{R}}$. Accordingly, the vectors $\mathbf{s}^{(i)} = (\mathbf{I}_L \otimes \mathbf{G})\mathbf{u}^{(i)}$ are selected uniformly over the region $\bar{\mathcal{R}}'$, where it can be defined as $\bar{\mathcal{R}}' = (\mathbf{I}_L \otimes \mathbf{G})\bar{\mathcal{R}}$. The volume of this region is $\mathrm{Vol}(\bar{\mathcal{R}}') = \left(\sqrt{\det(\mathbf{Q})}\right)^L \mathrm{Vol}(\bar{\mathcal{R}})$, where $\mathbf{Q} = \mathbf{G}^T\mathbf{G}$.

In order to find the asymptotic average transmit energy of the SLM method, set $r = 2$ and $F = U(\bar{\mathcal{R}}')$, where $\bar{\mathcal{R}}'$ is the region for the transmit vector $\mathbf{s}$. Therefore, according to the expression in (46), the average transmit energy for large window size L and large enough N can be approximated by

$$E_{SLM} = \frac{1}{L}E\left\{\gamma_{2,N}^{U(\bar{\mathcal{R}}')}\right\} = \frac{1}{L}B_D^{-\frac{2}{D}}\Gamma(1+\frac{2}{D})N^{-\frac{2}{D}}\mathrm{Vol}(\bar{\mathcal{R}}')^{\frac{2}{D}} \tag{49}$$

The volume of the region $\bar{\mathcal{R}}'$ is fixed, $\mathrm{Vol}(\bar{\mathcal{R}}') = \left(\sqrt{\det(\mathbf{Q})}\right)^L \mathrm{Vol}(\bar{\mathcal{R}}) = (\sqrt{\Pi\lambda_j})^L N\mathbb{V}^L$. Therefore, the average transmit energy of the SLM method can be represented by

$$E_{SLM} = \frac{1}{L}\Gamma(1+\frac{2}{D})\sqrt[n]{\Pi\lambda_j}\left(\frac{\mathbb{V}^L}{B_D}\right)^{\frac{2}{D}}. \tag{50}$$

For a spherical region, $\mathbb{V}^L = \mathrm{Vol}(\mathcal{B}_D(0, \sqrt{D}\sigma))$, where σ^2 is the variance of a Gaussian random variable with entropy $\mathcal{H} = \frac{1}{D}\log(\mathbb{V}^L)$. Considering the fact that B_D is the volume of a D-dimensional sphere with unit radius and $\Gamma(1+\frac{2}{D}) \longrightarrow \Gamma(1) = 1$ for large enough D, the average transmit energy for the SLM technique is

$$E_{SLM} = n\mathcal{R}_{eq}^2 \qquad (L, N \longrightarrow \infty), \tag{51}$$

where $\mathcal{R}_{eq}^2 = \sqrt[n]{\Pi\lambda_j}\sigma^2$.

3.5.2 Broadcast System

Now, consider the original broadcast system introduced in (31) in a time frame of length L. At each time slot, user $i = 1, \cdots, K$ has access to only $2n_i$ elements of each M-dimensional vector. Therefore, considering the L time slots together, each user selects its data uniformly over a $D_i = 2n_iL$-dimensional space $\mathcal{R}_i$. In the following analysis, it is assumed that each region $\mathcal{R}_i$ is a spherical region $\mathcal{B}_{D_i}(0, \sqrt{D_i}\sigma_i)$. In order to implement the proposed SLM method, each region $\mathcal{R}_i$ is expanded to $\overline{\mathcal{R}}_i$ such that $\text{Vol}(\overline{\mathcal{R}}_i) = N_i\text{Vol}(\mathcal{R}_i)$. Then, for each user, N_i i.i.d. random vectors are selected uniformly over the region $\overline{\mathcal{R}}_i$. Therefore, in the broadcast system (31), $N = N_1 \cdots N_K$ data vectors $\mathbf{u}^{(j)}_{1:L} = [\mathbf{u}^{(j)}_{1:L,1}, \cdots, \mathbf{u}^{(j)}_{1:L,K}]$ (for $j = 1, \cdots, N$) are selected over the region $\overline{\mathcal{R}} = \overline{\mathcal{R}}_1 \times \overline{\mathcal{R}}_2 \times \cdots \times \overline{\mathcal{R}}_K$. Note that $\mathbf{u}^{(j)}_{1:L,i}$ is selected in D_i-dimensional space $\overline{\mathcal{R}}_i$, independent of other users' data. Therefore, the data vectors $\mathbf{u}^{(j)}_{1:L}$ are uniformly distributed over $\overline{\mathcal{R}}$. Therefore, according to Theorem 5, the average energy of the SLM method in broadcast system (31) is given by the following theorem.

Theorem 6 *In a SLM technique applied to a broadcast system defined in (31), the average energy is*

$$\lim_{L,N\longrightarrow\infty} E_{Broadcast} = M\sqrt[M]{\Pi\lambda_i}\sigma^2, \tag{52}$$

where σ^2 *is the variance of a Gaussian random variable with entropy* $\mathcal{H} = \frac{1}{M}\log(\mathbb{V})$ *and* $\sigma^M = \sigma_1^{2n_i} \cdots \sigma_K^{2n_K}$.

Proof: See Appendix A. ∎

This implies that the average transmit energy in the proposed SLM technique can approach the average transmit energy in the lower bound in (37).

3.5.3 Some Implementation Issues

In any practical SLM method, a lattice code $\mathbb{C}$ (equivalently region $\mathcal{R}_{\mathbb{C}}$) is expanded such that the number of constellation points are multiplied by N, resulting in a new lattice code $\mathbb{C}'$ (equivalently region $\mathcal{R}_{\mathbb{C}'}$). The new set of constellation points are grouped in $|\mathbb{C}|$ sets containing N points. Transmitting any point in each set transfer the same information. The unique specification of each set is that the N equivalent points are selected uniformly over $\mathcal{R}_{\mathbb{C}'}$. Moreover, in implementing the SLM method, the shaping concept should be used in each user's space. It must be emphasized that the conventional precoding schemes in broadcast systems, such as [115] and [154], or even those in MIMO point to point systems [44], treat multiple antennas of users as different virtual users.

In order to implement an SLM method with these properties, the best approach is applying *lattice decomposition* [47]. Assume that in the D_i-dimensional space of each user, there is a lattice partition $\mathbf{\Lambda}/\mathbf{\Lambda}'$, where $|\mathbf{\Lambda}/\mathbf{\Lambda}'| = N_i$. The points in $\mathbf{\Lambda}'$ are known as a coset code [46]. For transmitting any point in $\mathbf{\Lambda}'$, N_i equivalent points $\{\mathbf{u}^{(1)}_{1:L,i}, \mathbf{u}^{(2)}_{1:L,i}, \cdots, \mathbf{u}^{(N_i)}_{1:L,i}\}$ are selected in $\mathbf{\Lambda}$. Having N_i points in the space of each user, $N = \prod N_i$ equivalent transmitted signals $\{\mathbf{s}^{(1)}_{1:L}, \mathbf{s}^{(2)}_{1:L}, \cdots, \mathbf{s}^{(N)}_{1:L}\}$ can be formed in the ML-dimensional space of all users, where $\mathbf{s}^{(j)}_t = [\mathbf{s}^{(j)}_{t,1}, \mathbf{s}^{(j)}_{t,2}, \cdots, \mathbf{s}^{(j)}_{t,K}]$ and $\mathbf{s}^{(j)}_t = \mathbf{H}^{-1}[\mathbf{u}_{t,1}, \mathbf{u}_{t,2}, \cdots, \mathbf{u}_{t,K}]$. Then among these signals the one with the lowest average energy is transmitted.

3.6 Simulation Results

In this section, the proposed SLM method is simulated in a MIMO broadcast system. In this system, $K = 2$ users with two antennas are considered, i.e. $n_1 = n_2 = 2$, resulting in $M = 4$. Figure 11 shows the average transmit energy for the simple channel inversion method [114], applying i.i.d. Gaussian random vector $\mathbf{u}$ (Case IV in Appendix B – conventional shaping gain), Perturbation technique [115], the proposed SLM method, and the lower bound for MIMO system in (37). For this simulation, a quasi static channel is assumed. Specifically, 10000 random channel matrices are generated and for each sample 1000 uniform random input vector over the corresponding continuous constellation are generated. It is assumed that the volume of the constellations are equal and the number of redundant symbols (if needed) is equal in all different cases. Note that for the proposed SLM method $N_1 = N_2 = 3^4$ and $L = 100$.

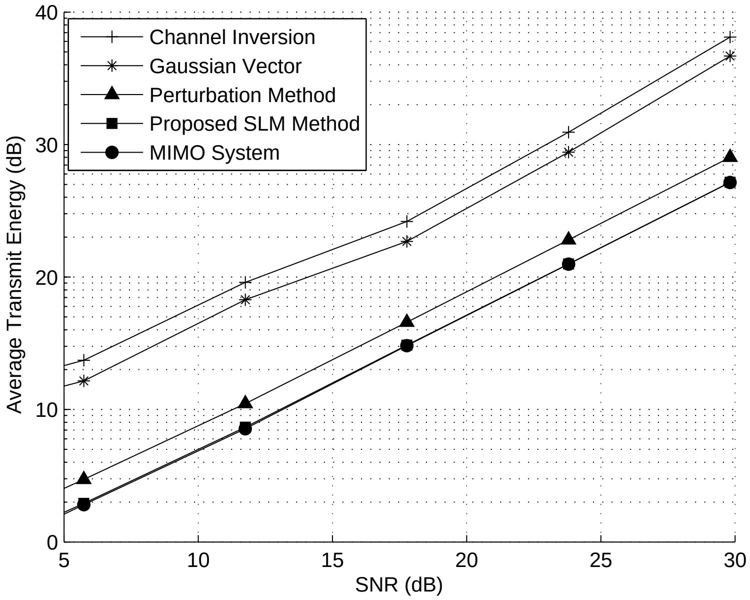

Figure 11: Average transmit energy for different methods in a quasi static channel

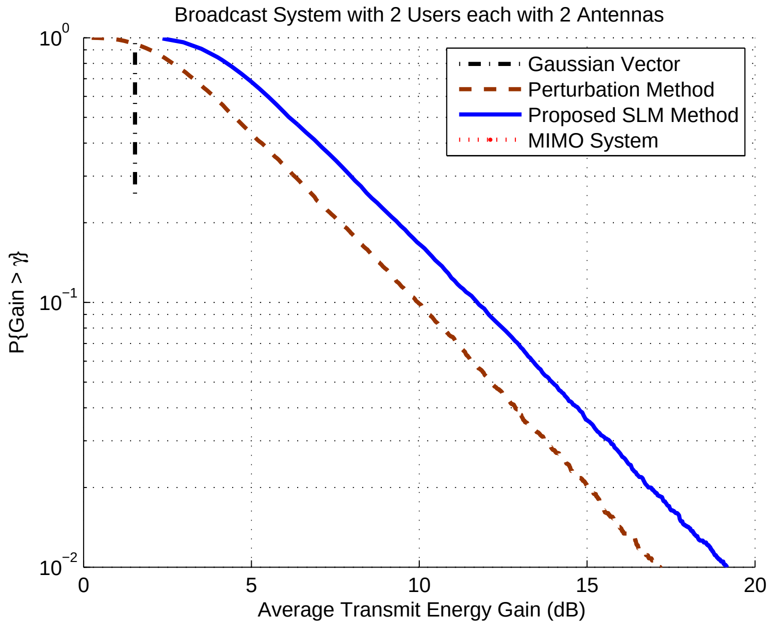

Figure 12: CCDF for the gain in average transmit energy of different methods compared to channel inversion in a quasi static channel

In the simulation, there are 10000 random samples of channel matrix. For each sample of the channel, the gain of each method is defined as the ratio of the average transmit energy of the simple channel inversion to the that method's average transmit energy. Figure 12 shows the Complementary Cumulative Density Function (CCDF) of these gains in average transmit energy compared to the channel inversion method (at $SNR = 11.7577$).

3.7 Conclusion

In this chapter, a lattice code with a low average transmit energy in multiple antenna broadcast systems employing channel inversion technique is designed. By using channel inversion, the decoding at the receiver side is independent of the channel matrix. Here, an SLM technique is introduced to provide the lattice code required for the constellation shaping. The average transmit energy that the SLM technique can provide is also derived. In order to implement the SLM method effectively, using lattice decomposition techniques is proposed. The capacity analysis of the proposed SLM technique can be considered as a direction for the future work.

PART II

Lattice Decoding

CHAPTER 4

LATTICE DECODING IN MIMO SYSTEMS

Abstract – In Multi-Input Multi-Output (MIMO) systems, Maximum-Likelihood (ML) decoding is equivalent to finding the closest lattice point in an N-dimensional complex space. In general, this problem is known to be NP-hard. In this part, several quasi-maximum likelihood algorithms based on using Semi-Definite Programming (SDP) are proposed. The SDP relaxation models are based on vector lifting and matrix lifting SDP. The computational complexity of matrix lifting SDP models are low. However, a near-ML performance with higher complexity with vector lifting SDP models can be achieved.

4.1 Introduction

Recently, there has been a considerable interest in Multi-Input Multi-Output (MIMO) antenna systems due to achieving a very high capacity as compared to single-antenna systems [134,45]. In MIMO systems, a vector is transmitted by the transmit antennas. In the receiver, a corrupted version of this vector affected by the channel noise and fading is received. Decoding concerns the operation of recovering the transmitted vector from the received signal. This problem is usually expressed in terms of “lattice decoding" which is known to be NP-hard.

To overcome the complexity issue, a variety of sub-optimum polynomial time algorithms are suggested in the literature for lattice decoding. However, unfortunately, these algorithms usually result in a noticeable degradation of performance. Examples of such polynomial time algorithms include: Zero Forcing Detector (ZFD) [118,69], Minimum Mean Squared Error Detector (MMSED) [65, 148], Decision Feedback Detector (DFD) and Vertical Bell Laboratories Layered Space-Time Nulling and Cancellation Detector (VBLAST Detector) [52, 38].

Lattice basis reduction has been applied as a pre-processing step in sub-optimum decoding algorithms to reduce the complexity and achieve a better performance. Minkowski reduction [61], Korkin-Zolotarev reduction [71] and LLL reduction [80] have been successfully used for this purpose in [1, 4, 6, 7, 103, 145].

In the last decade, Sphere Decoder (SD)[1] is introduced as a Maximum Likelihood (ML) decoding method for MIMO systems with near-optimal performance [31]. In the SD method, the lattice points inside a hyper-sphere are generated and the closest lattice point to the received signal is determined. In [66], an exponential lower bound is derived on the average complexity of SD, and it is shown that the worst case complexity is exponential [1, 58]. However, it is experienced that over certain ranges of rate, Signal to Noise Ratio (SNR) and dimension the average complexity is polynomial [58].

Recently, a variety of sub-optimum polynomial time algorithms based on Semi-Definite Programming (SDP) are suggested for lattice decoding [124, 125, 85, 86, 143, 121, 150, 95, 91]. In all of these methods, the detection problem is lifted into a higher dimension and the discrete set of possible vectors is replaced by a convex (and therefore connected) set. By using the convex optimization techniques, an approximate solution to lattice decoding problem is obtained. Overall, this procedure yields a polynomial time approximation of the difficult optimization problem present in the ML detection problem.

This part of the thesis presents several relaxation models based on SDP for the lattice decoding problem. These relaxation models can be categorized in (i) *Vector Lifting Semi-Definite Programming* (VLSDP) models and (ii) *Matrix Lifting Semi-Definite Programming* (MLSDP) models [40, 11], where they result in different performance and computational complexity trade-offs. In the proposed relaxation models, the transmitted vector is expanded as a linear combination (with zero-one coefficients) of all the possible constellation points in each dimension. Using this formulation, the distance minimization in Euclidean space is expressed in terms of a binary quadratic minimization problem. The minimization of this problem is over the set of all binary rank-one matrices with column sums equal to one.

In order to solve this minimization problem, by using VLSDP, two relaxation models are presented, providing a trade-off between the computational complexity and the performance (both models can be solved with polynomial-time computational complexity). Simulation results show that the performance of the last model is near optimal for M-ary QAM or PSK constellation (with an arbitrary binary labeling, say Gray labeling). Therefore, the decoding algorithm built on the proposed model using VLSDP has a near-ML performance with polynomial computational complexity. In the relaxation models based on MLSDP, a large reduction

[1] This technique is introduced in the mathematical literature several years ago [19, 43].

in the computational complexity is achieved as compared to that in VLSDP models (the number of variables is decreased from $O(N^2K^2)$ to $O((N+K)^2)$), where N is the number of antennas and K is the number of constellation points in each real dimension.

4.2 Lattice Decoding Problem Formulation

A MIMO system with $\tilde{N}$ transmit antennas and $\tilde{M}$ receive antennas ($\tilde{M} \times \tilde{N}$ MIMO system) is modeled as

$$\tilde{\mathbf{y}} = \sqrt{\frac{SNR}{\tilde{M}\tilde{E}_{s_{av}}}}\tilde{\mathbf{H}}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \tag{53}$$

where $\tilde{\mathbf{H}} = \left[\tilde{h}_{ij}\right]$ is the $\tilde{M} \times \tilde{N}$ channel matrix composed of independent, identically distributed complex Gaussian random elements with zero mean and unit variance, $\tilde{\mathbf{n}}$ is an $\tilde{M} \times 1$ complex AWGN vector with zero mean and unit variance, and $\tilde{\mathbf{x}}$ is an $\tilde{N} \times 1$ data vector whose components are selected from a complex set $\{\tilde{s}_1, \tilde{s}_2, \cdots, \tilde{s}_K\}$ with an average energy of $\tilde{E}_{s_{av}}$. The parameter SNR in (53) is the SNR per receive antenna.

To avoid using complex matrices, the system model (53) is represented by real matrices in (54).

$$\begin{aligned}
\begin{bmatrix} \Re(\tilde{\mathbf{y}}) \\ \Im(\tilde{\mathbf{y}}) \end{bmatrix} &= \sqrt{\frac{SNR}{\tilde{M}\tilde{E}_{s_{av}}}} \begin{bmatrix} \Re\left(\tilde{\mathbf{H}}\right) & \Im\left(\tilde{\mathbf{H}}\right) \\ -\Im\left(\tilde{\mathbf{H}}\right) & \Re\left(\tilde{\mathbf{H}}\right) \end{bmatrix} \begin{bmatrix} \Re(\tilde{\mathbf{x}}) \\ \Im(\tilde{\mathbf{x}}) \end{bmatrix} \\
&+ \begin{bmatrix} \Re(\tilde{\mathbf{n}}) \\ \Im(\tilde{\mathbf{n}}) \end{bmatrix} \\
\Rightarrow \mathbf{y} &= \mathbf{H}\mathbf{x} + \mathbf{n},
\end{aligned} \tag{54}$$

where $\mathbf{y}$ is the *received vector*, and $\mathbf{x}$ is the *input vector*. Decoding concerns the operation of recovering the transmitted vector $\mathbf{x}$ from the received signal $\mathbf{y}$, based upon the knowledge[2] of $\mathbf{y}$ and $\mathbf{H}$ and produce an estimated transmitted vector $\hat{\mathbf{x}}$. Note that the representation in (54) is known as the lattice representation of the MIMO system [30, 31, 32]. The matrix $\mathbf{H}$ is the lattice generator/basis of the MIMO system in (53).

The important performance measure of the decoding algorithms is the *probability of error* which is defined as the probability that the estimated transmitted vector, $\hat{\mathbf{x}}$, is not the input vector, $\mathbf{x}$, i.e.

$$\mathbb{P}\{\hat{\mathbf{x}} \neq \mathbf{x}\}. \tag{55}$$

[2] Note that, throughout the paper, it is assumed that the detector, or receiver, has access to not only the received vector, $\mathbf{y}$, but also the channel matrix, $\mathbf{H}$. The problem of accessing to these quantities is not in the scope of this thesis.

In the sequel, different methods are compared based on the performance which is defined as the probability of error. The method that minimizes this performance measure is known as ML method. At the receiver, the ML decoding rule is given by [138]

$$\hat{\mathbf{x}} = \arg\min_{x_i \in \mathcal{S}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2. \tag{56}$$

In the context of mathematics, the problem in (56) is known as the *lattice decoding problem* which encounters finding the closest point of the lattice defined by **H** to the received vector **y**. In digital communication applications, this problem is known as the integer least-square problem, which can be seen in many areas, e.g. the detection of symbols transmitted over the multiple antenna wireless channel [134], the multiuser detection problem in Code Division Multiple Access CDMA systems [41], the simultaneous detection of multiple users in a DSL system affected by crosstalk [50], and cryptography. The results, here, can be applied to all these applications. However, the problem of lattice decoding in MIMO systems is the focus of this thesis.

4.3 Lattice Decoding Methods

Existing methods for solving the integer least-square problem are classified in four categories:

1. Heuristic methods
2. Lattice Basis Reduction methods
3. Sphere Decoding methods
4. Semi-Definite Programming methods

4.3.1 Heuristic Methods

ZFD [118, 69] and MMSED [65, 148] are the well known methods in heuristic methods. In ZFD, the vector resulted by multiplying the received vector by pseudo inverse (Moore-Penrose inverse [78]) of matrix **H**, denoted by $\mathbf{H}^{\dagger}$, is rounded off to the closest integer.

$$\hat{\mathbf{x}} = \left\lceil \mathbf{H}^{\dagger}\mathbf{y} \right\rfloor, \tag{57}$$

where this point is known as Babai point [4].

MMSED has the same principal as ZFD, but this detector considers the effect of the noise variance. By using inter-antenna interference in ZFD, the $\mathbf{H}^{\dagger}$ in (57) is replaced by

$$\left(\mathbf{H}^{*}\mathbf{H} + \sigma^{2}\mathbf{I}\right)^{-1}\mathbf{H}^{*}, \tag{58}$$

where σ^2 is the noise variance.

DFD is another type of heuristic method in which the components of the signal point are estimated recursively by one of the aforementioned answers (ZFD or MMSED) and the effect of each detected component is cancelled on the next ones. This procedure is also called *Nulling and Cancellation* method. Nulling and cancellation method is suffering from error propagation. If the first component is estimated wrong, it has an adverse effect on the estimation of the other components.

In [52, 38], VBLAST detection algorithm based on DFD algorithm and an appropriate ordering for detecting components is introduced. The detection order is accordance with the descending order of SNR of different elements in the received point. This algorithm has the same answer with the *Nearest Plane* algorithm that Babai presented in [4] using lattice reduction concept (Babai point).

VBLAST can handle high data rates with reasonable complexity; however, the loss in performance as compared to ML decoding is usually significant. In addition: (i) VBLAST transmits independent data streams on its antennas, so there is no built-in spatial or temporal coding, and (ii) the decoding scheme does not work with fewer receive than transmit antennas.

4.3.2 Lattice Basis Reduction Methods

The algorithm for finding the closest point of $\mathbb{Z}^N$ to the received point in N-D real space is just to round off the point components in real space. However, the rounding off procedure in ZFD or MMSED can not always provide the right answer. It can be shown that the aforementioned suboptimal procedures are guarantees to find the nearest point in the lattice if and only if the basis vectors of $\mathbf{H}$ are mutually orthogonal. Unfortunately, such condition cannot be satisfied for every generator matrix $\mathbf{H}$.

The solution for this problem may be found in *lattice basis reduction* context. Lattice basis reduction is transforming a given lattice basis into a basis consisting vectors which are *short* and fairly *orthogonal*. In other words, the channel matrix $\mathbf{H}$ is decomposed to $\mathbf{H} = \mathbf{QH}'$ such that $\mathbf{Q}$ is unimodular[3] and the matrix $\mathbf{H}'$ consists of vectors which are shorter and more orthogonal compared to the vectors defined by $\mathbf{H}$.

General known procedures for lattice decoding using lattice basis reduction find the solution in three steps (see Chapter 5.5):

1. Reduce lattice basis ($\mathbf{H} = \mathbf{H}'\mathbf{Q}$).

2. Perform finding the point in reduced lattice (finding the closest lattice point to $\mathbf{y}$ in the

[3] A unimodular matrix is an integer square matrix whose determinant is one.

lattice defined by $\mathbf{H}'$).

3. Transform result to the original lattice (multiplying $\mathbf{Q}^{-1}$ to the answer).

It can be shown that for the decoding of a lattice, the KZ reduced basis is a more powerful tool than the LLL reduced basis [6, 1] for small dimensions ($N < 15$) [145]. However, for large dimensions LLL reduced basis speeds up the decoding algorithm up to two orders of magnitude. Moreover, in [130], it is shown that the decoding algorithm using LLL basis reduction achieves the same receive diversity as the ML decoding algorithm (which is equal to the number of receive antennas). In [103], LLL reduced basis is used with suboptimal lattice decoders, such as ZFD and VBLAST. Moreover, an algorithm for changing LLL reduction for complex matrix is proposed. In [4], Babai showed the LLL assumption for the goodness of the Babai point is essential.

4.3.3 Sphere Decoding Methods

ML decoder finds the optimal solution by searching over all lattice points in the constellation (lattice code). This decoder is practically infeasible due to its exponential complexity [139]; however, its improvement in the bit error rate performance cannot be neglected[4].

A number of algorithms can be found in mathematical contexts for general lattice decoding. Kannan's algorithm [67] recursively searches all the lattice points inside an N-D rectangular parallelepiped (cube), centered at $\mathbf{y}$ with its edges along the Gram-Schmidt vectors of a proper basis of the lattice. [7], [68] and [61] are different variants of Kannan's algorithm. In Fincke and Pohst algorithm [43] (called as *sphere decoding* algorithm), lattice points inside an N-D hypersphere centered at $\mathbf{y}$ are searched. By finding a lattice point inside the sphere, the radius of the sphere is updated. Later, Schnorr and Euchner [19] introduced an improved version of Fincke and Pohst's algorithm [43] . They suggested to enumerate the lattice points inside the N-D hypersphere in an order which enumerates the values of components in the order of increasing distance from the components of integer point corresponding to $\mathbf{y}$.

Several applications in communications have used these algorithms. Viterbo and Boutros [141] used an algorithm based on Fincke and Pohst's [43] algorithm for decoding in fading channels. Damen et. al. used this idea for decoding in MIMO channels [30, 31, 32]. They have explored the lattice representation of a multi antenna system and the algebraic space-time codes for any number of transmit and receive antennas. Based on this representation, they have applied the sphere decoding algorithm of [141].

Hassibi and Vikalo introduced a *sphere decoding* algorithm [58, 59] for MIMO systems

[4] More comparisons between ML decoder and suboptimal decoders are in [57].

based on Fincke and Pohst's algorithm. Instead of determining the lattice points in an N-D space, this algorithm recursively determines components of the lattice in each dimension. The only problem here is the choice of a proper radius. Radius in [58, 59] is a scaled version of the noise variance such that with a high probability a lattice point inside the sphere can be found. In *Closest Lattice Point Search Algorithm* [1], Agrell et. al. have generalized Schnorr and Euchner's algorithm for decoding of any MIMO system. In this method, the radius is ignored in the search algorithm due to the special ordering. There are several variants for this algorithm. Chan and Lee [22] have used the same ordering as Schnorr-Euchner ordering in Fincke and Pohst's sphere decoding algorithm. In [103], an algorithm based on using LLL reduction and Schnorr-Euchner ordering in Fincke and Pohst algorithm is proposed.

In [33], Finite signal sets and a fresh look at the class of decoding algorithms as stack algorithms are considered. Based on Fincke and Pohst's algorithm and Schnorr and Euchner's algorithm two reduced complexity algorithms for general lattice codes are proposed. By using efficient pre-processing stage a near ML decoding algorithm that uniformly outperforms all known sphere decoders in terms of receiver complexity is introduced.

It has been shown that sphere decoding gains huge improvement over VBLAST decoding method, full diversity of coded multi antenna systems, high spectral efficiency, independency of the constellation size, and maximum likelihood performance. There have been several attempts for improving the sphere decoding algorithm specially because of ML decoder performance. However, in [66], an exponential lower bound is derived on the average complexity of sphere decoding, and it is shown that the worst case complexity is exponential [1, 58]. However, it is experienced that over certain ranges of rate, SNR and dimension the average complexity is polynomial [58].

4.3.4 Semi-Definite Programming methods

Recently, due to exponential complexity of sphere decoding algorithms, the attention is moved to a variety of sub-optimum polynomial time algorithms based on Semi-Definite Programming (SDP), [124, 125, 85, 86, 143, 121, 150, 95, 91].

In [124, 86], a quasi-ML method for lattice decoding is introduced. Each signal constellation is expressed by its binary representation and the decoding is transformed into a quadratic minimization problem [124]. Then, the resulting problem is solved using a relaxation for rank-one matrices in SDP context. It is shown that this method has a near optimum performance and a polynomial time worst case complexity. However, the method proposed in [124] is limited to scenarios that the constellation points are expressed as a linear combination of bit labels. A typical example is the case of natural labeling in conjunction with PSK constellation [125].

Another quasi-maximum likelihood decoding method is introduced in [85] for larger PSK constellations with near ML performance and low complexity.

Another quasi-ML decoding method is introduced in [143] for the MIMO systems employing 16-QAM, where the structure of constellation is captured by a polynomial constraint. Then, by introducing some slack variables, the constraints are expressed in terms of quadratic polynomials. This method can be generalized for larger constellations at the cost of defining more slack variables, increasing the complexity, and significantly decreasing the performance. The method proposed in [121] is a further relaxation of [143], only utilizing upper and lower bounds on the symbol energy in the relaxation step. There is a very slight degradation in performance compared to [143]; however, its computational complexity is independent of the constellation size for any uniform QAM (order of complexity is cubic). The method in [150] is a further tightening of [121] by appending some inequality conditions that are implicit in the alphabet constraint. Its computational complexity is still less than that in [143].

This part presents several quasi-ML algorithms based on using SDP. The proposed SDP relaxation models are based on vector lifting and matrix lifting SDP relaxations. The computational complexity of matrix lifting SDP models are low. However, a near-ML performance with higher complexity with vector lifting SDP models can be achieved. In the next two chapters, these relaxation models are presented in details.

CHAPTER 5

VECTOR LIFTING SEMI-DEFINITE PROGRAMMING

Abstract **–** In this chapter, a quasi-maximum likelihood algorithm based on Semi-Definite Programming (SDP) is proposed. Several SDP relaxation models for Multiple-Input Multiple-Output (MIMO) systems, with increasing complexity, are introduced. Interior-point methods are used to solve the models and obtain a near-ML performance with polynomial computational complexity. Lattice basis reduction is applied to further reduce the computational complexity of solving these models.

5.1 Introduction

Semi-Definite Programming (SDP) is a powerful tool for bounding the optimal value of a combinatorial optimization problem. Its first application in communications was to bound the Shannon capacity of a graph [84]. The successful implementation of a sub-optimal solution for MAX-CUT problem in [51] based on SDP opened a new approach on using SDP in different applications. The application of SDP to the detection problem considered herein has been studied previously in the communications literature [124, 125, 85, 86, 143, 121, 150, 95, 91].

This chapter develops an efficient approximate Maximum Likelihood (ML) decoder for Multiple-Input Multiple-Output (MIMO) systems based on Vector Lifting Semi-Definite Programming (VLSDP). In the proposed method, the transmitted vector is expanded as a linear combination (with zero-one coefficients) of all the possible constellation points in each dimension. Using this formulation, the distance minimization in Euclidean space is expressed in terms of a binary quadratic minimization problem. The minimization of this problem is over the set of all binary rank-one matrices with row sums equal to one. In order to solve this minimization problem, two relaxation models (Model III and IV) are presented, providing a trade-off between the computational complexity and the performance (both models can be solved with polynomial-time complexity). Two additional relaxation models (Model I and II) are presented as intermediate steps in the derivations of Model III and IV.

Model I: A preliminary SDP relaxation of the minimization problem is obtained by removing the rank-one constraint in the problem and using Lagrangian duality [146]. This relaxation has many redundant constraints and no strict interior point in the feasible set (there are numerical difficulties in computing the solution for a problem without an interior point).

Model II: To overcome this drawback, the feasible set is projected onto a face of the semidefinite cone. Then, based on the identified redundant constraints, another form of the relaxation is obtained, which can be solved using interior-point methods.

Model III: The relaxation Model II results in a weak lower bound. To strengthen this relaxation model, the structure of the feasible set is investigated. An interesting property of the feasible set imposes a zero pattern for the solution. Adding this pattern as an extra constraint to the previous relaxation model results in a stronger model.

Model IV: Finally, the strongest relaxation model in this chapter is introduced by adding some additional non-negative constraints. The number of non-negative constraints can be adjusted to provide a trade-off between the performance and complexity of the resulting method.

Simulation results show that the performance of the last model is near optimal for M-ary QAM or PSK constellation (with an arbitrary binary labeling, say Gray labeling[1]). Therefore, the decoding algorithm built on this model has a near-ML performance with polynomial computational complexity.

The proposed models result in a solution that is not necessarily a binary rank-one matrix. This solution is changed to a binary rank-one matrix through a randomization algorithm. As the first step, the conventional randomization algorithms are modified to adopt to the SDP problem here. Moreover, a new randomization procedure is introduced which finds the optimal binary rank-one solution in a smaller number of iterations than the conventional ones. Finally, using a lattice basis reduction method to further reduce the computational complexity of the proposed relaxation models is discussed. The extension of the decoding technique for soft output decoding is also investigated.

5.2 *Problem Formulation*

Consider the MIMO system defined in (54). Noting $x_i \in \{s_1, \cdots, s_K\}$, for $i = 1, \cdots, N$,

$$x_i = u_{i,1}s_1 + u_{i,2}s_2 + \cdots + u_{i,K}s_K, \tag{59}$$

[1] It is shown that Gray labeling, among all possible constellation labeling methods, offers the lowest possible average probability of bit errors [2].

where

$$u_{i,j} \in \{0,1\} \quad \text{and} \quad \sum_{j=1}^{K} u_{i,j} = 1, \quad \forall\, i = 1,\cdots,N. \tag{60}$$

Let

$$\mathbf{U} = \begin{bmatrix} u_{1,1} & \cdots & u_{1,K} \\ u_{2,1} & \cdots & u_{2,K} \\ \vdots & \ddots & \vdots \\ u_{N,1} & \cdots & u_{N,K} \end{bmatrix} \quad \text{and} \quad \mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix}.$$

Therefore, the transmitted vector is $\mathbf{x} = \mathbf{U}\mathbf{s}$ with the constraint $\mathbf{U}\mathbf{e}_K = \mathbf{e}_N$. This constraint represents a constraint on the binary matrix $\mathbf{U}$ with row sums equal to one.

Remark 1 *Note that for the case of complex numbers, e.g. PSK constellations, the same approach can be applied, resulting in $\tilde{\mathbf{x}} = \mathbf{U}\tilde{\mathbf{s}}$ for the MIMO system in (53), where $\tilde{\mathbf{s}}$ is the vector of PSK constellation points.*

Considering the new notation for the input vector $\mathbf{x}$, the ML decoding rule in (56) is equivalent to

$$\begin{aligned} &\min_{\mathbf{U}\mathbf{e}_K=\mathbf{e}_N} ||\mathbf{y} - \mathbf{H}\mathbf{U}\mathbf{s}||^2 \equiv \\ &\min_{\mathbf{U}\mathbf{e}_K=\mathbf{e}_N} \mathbf{s}^T\mathbf{U}^T\mathbf{H}^T\mathbf{H}\mathbf{U}\mathbf{s} - 2\mathbf{y}^T\mathbf{H}\mathbf{U}\mathbf{s}. \end{aligned} \tag{61}$$

Therefore, the decoding problem can be formulated as

$$\begin{aligned} \min \quad & \mathbf{s}^T\mathbf{U}^T\mathbf{H}^T\mathbf{H}\mathbf{U}\mathbf{s} - 2\mathbf{y}^T\mathbf{H}\mathbf{U}\mathbf{s} \\ s.t. \quad & \mathbf{U}\mathbf{e}_K = \mathbf{e}_N \\ & u_{i,j} \in \{0,1\}. \end{aligned} \tag{62}$$

Let $\mathbf{Q} = \mathbf{H}^T\mathbf{H}$, $\mathbf{S} = \mathbf{s}\mathbf{s}^T$, $\mathbf{C} = -\mathbf{s}\mathbf{y}^T\mathbf{H}$, and let $\mathcal{E}_{N\times K}$ denote the set of all binary matrices in $\mathcal{M}_{N\times K}$ with row sums equal to one, i.e.

$$\mathcal{E}_{N\times K} = \left\{\mathbf{U} \in \mathcal{M}_{N\times K} : \mathbf{U}\mathbf{e}_K = \mathbf{e}_N, u_{ij} \in \{0,1\}\right\}. \tag{63}$$

Therefore, the minimization problem (62) is

$$\begin{aligned} \min \quad & \operatorname{trace}\left(\mathbf{S}\mathbf{U}^T\mathbf{Q}\mathbf{U} + 2\mathbf{C}\mathbf{U}\right) \\ s.t. \quad & \mathbf{U} \in \mathcal{E}_{N\times K} \end{aligned} \tag{64}$$

5.3 Vector-Lifting Semi-Definite Programming Solution

A *quadratic vector optimization* solution of (64) can be obtained by defining $\mathbf{u} = \text{vec}(\mathbf{U}^T), \mathbf{U} \in \mathcal{E}_{N\times K}$. By using this notation, the objective function is replaced by $\mathbf{u}^T(\mathbf{Q}\otimes\mathbf{S})\mathbf{u} + 2\text{vec}(\mathbf{C})^T\mathbf{u}$,i.e. the minimization problem (64) is

$$\begin{aligned} \min \quad & \mathbf{u}^T(\mathbf{Q}\otimes\mathbf{S})\mathbf{u} + 2\text{vec}(\mathbf{C})^T\mathbf{u}. \\ s.t. \quad & \mathbf{u} = \text{vec}(\mathbf{U}^T), \mathbf{U} \in \mathcal{E}_{N\times K} \end{aligned} \tag{65}$$

This is a quadratic minimization problem with binary variables [146]. Some recent studies on solving binary quadratic minimization problems such as Graph Partitioning [147] and Quadratic Assignment Problem [156, 123] show that SDP is a very promising approach to provide tight relaxations for such problems. In the following, several SDP relaxation models for the minimization problem in (65) are derived. Appendix C provides the mathematical framework for these models using the Lagrangian duality [146].

Consider the minimization problem in (65). Since $\mathbf{u}$ is a binary vector, the objective function can be represented by

$$\begin{aligned} \mathbf{u}^T(\mathbf{Q}\otimes\mathbf{S})\mathbf{u} + 2\text{vec}(\mathbf{C})^T\mathbf{u} &= \text{trace}\left(\begin{bmatrix} 1 & \mathbf{u}^T \end{bmatrix}\mathcal{L}_{\mathbf{Q}}\begin{bmatrix} 1 \\ \mathbf{u} \end{bmatrix}\right) \\ &= \text{trace}\left(\mathcal{L}_{\mathbf{Q}}\begin{bmatrix} 1 \\ \mathbf{u} \end{bmatrix}\begin{bmatrix} 1 & \mathbf{u}^T \end{bmatrix}\right) \\ &= \text{trace}\left(\mathcal{L}_{\mathbf{Q}}\left[\begin{array}{c|c} 1 & \mathbf{u}^T \\ \hline \mathbf{u} & \mathbf{u}\mathbf{u}^T \end{array}\right]\right), \end{aligned} \tag{66}$$

where $\mathcal{L}_{\mathbf{Q}} := \begin{bmatrix} 0 & \text{vec}(\mathbf{C})^T \\ \text{vec}(\mathbf{C}) & \mathbf{Q}\otimes\mathbf{S} \end{bmatrix}$. Therefore, the minimization problem (65) using VLSDP can be written as

$$\begin{aligned} \min \quad & \text{trace } \mathcal{L}_Q \left[\begin{array}{c|c} 1 & \mathbf{u}^T \\ \hline \mathbf{u} & \mathbf{u}\mathbf{u}^T \end{array}\right] \\ \text{s.t.} \quad & \mathbf{u} = \text{vec}(\mathbf{U}^T),\ \mathbf{U} \in \mathcal{E}_{N\times K}. \end{aligned} \tag{67}$$

Note that in the VLSDP optimization problem (65), $\mathbf{u}$ is an $n = NK$-D vector, and hence, the optimization parameter is a matrix in $\mathcal{S}_{NK+1}$, which has $(NK+1)^2$ variables.

To derive the first semi-definite relaxation model, a direct approach based on the well known lifting process [5] is selected. In accordance to (67), for any $\mathbf{U} \in \mathcal{E}_{N\times K}$, $\mathbf{u} = \text{vec}(\mathbf{U}^T)$, the feasible points of (67) are expressed by

$$\mathbf{Y}_{\mathbf{u}} = \begin{bmatrix} 1 \\ \mathbf{u} \end{bmatrix}\begin{bmatrix} 1 & \mathbf{u}^T \end{bmatrix} = \left[\begin{array}{c|c} 1 & \mathbf{u}^T \\ \hline \mathbf{u} & \mathbf{u}\mathbf{u}^T \end{array}\right]. \tag{68}$$

The matrix $\mathbf{Y_u}$ is a rank-one and positive semi-definite matrix. Also,

$$\operatorname{diag}(\mathbf{Y_u}) = \mathbf{Y}^T_{\mathbf{u}_{0,:}} = \mathbf{Y}_{\mathbf{u}_{:,0}},$$

where $\mathbf{Y}_{\mathbf{u}_{0,:}}$ (resp. $\mathbf{Y}_{\mathbf{u}_{:,0}}$) denotes the first row (resp. the first column)[2] of $\mathbf{Y_u}$ (Note that $\mathbf{u}$ is a binary vector, and consequently, $\operatorname{diag}(\mathbf{uu}^T) = \mathbf{u}$).

In order to obtain a tractable SDP relaxation of (67), the rank-one restriction is removed from the feasible set. In fact, the feasible set is approximated by another larger set $\mathcal{F}$, defined as

$$\mathcal{F} := \operatorname{conv}\left\{\mathbf{Y_u} : \mathbf{u} = \operatorname{vec}(\mathbf{U}^T),\ \mathbf{U} \in \mathcal{E}_{N\times K}\right\}. \tag{69}$$

This results in the first relaxation model (**Model I**) for the original problem given in (65):

$$\begin{array}{ll} \min & \operatorname{trace}\mathcal{L}_{\mathbf{Q}}\mathbf{Y} \\ \text{s.t.} & \mathbf{Y} \in \mathcal{F} \end{array} \tag{70}$$

It is clear that the matrices

$$\mathbf{Y_u} \text{ for } \mathbf{u} = \operatorname{vec}(\mathbf{U}^T),\ \mathbf{U} \in \mathcal{E}_{N\times K}$$

are the feasible points of $\mathcal{F}$. Moreover, since these points are rank-one matrices, they are contained in the set of extreme points of $\mathcal{F}$, see e.g. [111]. In other words, if the matrix $\mathbf{Y}$ is restricted to be rank-one in (70), i.e. $\mathbf{Y} = \begin{bmatrix} 1 \\ \mathbf{u} \end{bmatrix}\begin{bmatrix} 1 & \mathbf{u}^T \end{bmatrix}$, for some $\mathbf{u} \in \mathbb{R}^n$, then the optimal solution of (70) provides the optimal solution of (65).

The SDP relaxation problem (70) is not solvable in polynomial time and $\mathcal{F}$ has no interior points. Therefore, the goal is to approximate the set $\mathcal{F}$ by a larger set containing $\mathcal{F}$. In the following, it is shown that $\mathcal{F}$ actually lies in a smaller dimensional subspace. Moreover, relative to this subspace, $\mathcal{F}$ will have interior points.

5.3.1 Geometry of the Relaxation

In order to approximate the feasible set $\mathcal{F}$ for solving the problem, the geometrical structure of this set is elaborated. In other words, the constraints defining $\mathbf{U}\mathbf{e}_K = \mathbf{e}_N$ is eliminated by providing a tractable representation of the linear manifold spanned by this constraint. This method is called *gradient projection* or *reduced gradient method* [56]. The following lemma is on the representation of matrices having sum of the elements in each row equal to one.

Lemma 7 *Let*

$$\mathbf{G} = \left[\begin{array}{c|c} \mathbf{I}_{K-1} & -\mathbf{e}_{K-1} \end{array}\right] \in \mathcal{M}_{(K-1)\times K} \tag{71}$$

[2] Matrix $\mathbf{Y_u}$ is indexed from zero.

and

$$\mathbf{F} = \frac{1}{K}\left(\mathbf{E}_{N\times K} - \mathbf{E}_{N\times(K-1)}\mathbf{G}\right) \in \mathcal{M}_{N\times K}. \tag{72}$$

A matrix $\mathbf{U} \in \mathcal{M}_{N\times K}$ *with the property that the summation of its elements in each row is equal to one, i.e.* $\mathbf{U}\mathbf{e}_K = \mathbf{e}_N$*, can be written as*

$$\mathbf{U} = \mathbf{F} + \hat{\mathbf{U}}\mathbf{G}, \tag{73}$$

where $\hat{\mathbf{U}} = \mathbf{U}(1 : N, 1 : (K-1))$.

Proof: see Appendix A. ■

Corollary 2 $\forall \mathbf{U} \in \mathcal{E}_{N\times K}$, $\exists \hat{\mathbf{U}} \in \mathcal{M}_{N\times(K-1)}$, $\hat{u}_{ij} \in \{0, 1\}$ *s.t.* $\mathbf{U} = \mathbf{F} + \hat{\mathbf{U}}\mathbf{G}$*, where* $\hat{\mathbf{U}} = \mathbf{U}(1 : N, 1 : (K-1))$.

Using Lemma 7, the following theorem can be proved which provides the structure of the elements in the set $\mathcal{F}$.

Theorem 8 *Define the vector* $\mathbf{b_V}$ *as*

$$\mathbf{b_V} = \frac{1}{K}(\mathbf{e}_{NK} - (\mathbf{I}_N \otimes \mathbf{G}^T)\mathbf{e}_{(K-1)N})$$

and let

$$\hat{\mathbf{V}} = \left[\begin{array}{c|c} 1 & \mathbf{0}^T_{N(K-1)} \\ \hline \mathbf{b_V} & \mathbf{I}_N \otimes \mathbf{G}^T \end{array}\right], \tag{74}$$

where $\hat{\mathbf{V}} \in \mathcal{M}_{(NK+1)\times((K-1)N+1)}$*. For any* $\mathbf{Y} \in \mathcal{F}$*, there exists a symmetric matrix* $\mathbf{R}$ *of order* $N(K-1)+1$*, indexed from 0 to* $N(K-1)$*, such that*

$$\mathbf{Y} = \hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T,\ \mathbf{R} \geq 0,\ \ \textit{and}\ \ r_{00} = 1,\ r_{ii} = r_{0i},\ \forall i. \tag{75}$$

Also, if $\mathbf{Y}$ *is an extreme point of* $\mathcal{F}$*, then* $r_{ij} \in \{0, 1\}$*, otherwise* $r_{ij} \in [0, 1]$ *for* $i, j \in \{0, \ldots, N(K-1)\}$.

Proof: see Appendix A. ■

Using Theorem 8, it is easy to show that the set $\mathcal{F}_r$ contains $\mathcal{F}$:

$$\begin{aligned} \mathcal{F}_r = \big\{\mathbf{Y} \in \mathcal{S}_{NK+1} : \exists\,\mathbf{R} \in \mathcal{S}_{(K-1)N+1},\ \mathbf{R} \geq \mathbf{0}, \\ r_{00} = 1, \mathbf{Y} = \hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T,\ \mathrm{diag}(\mathbf{Y}) = \mathbf{Y}_{0,:}\big\}. \end{aligned} \tag{76}$$

Therefore, the feasible set in (70) is approximated by $\mathcal{F}_r$. This results in the second relaxation model (**Model II**) of the original problem given in (65):

$$\begin{aligned} \min\ \ & \mathrm{trace}\ (\hat{\mathbf{V}}^T\mathcal{L}_{\mathbf{Q}}\hat{\mathbf{V}})\mathbf{R} \\ \text{s.t.}\ \ & \mathrm{diag}(\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T) = (1, (\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T)_{0,1:n})^T \\ & \mathbf{R} \geq 0. \end{aligned} \tag{77}$$

Note that the matrices $\mathbf{Y}_\mathbf{u}$ are contained in the set of extreme points of $\mathcal{F}$. Only those faces of $\mathcal{F}$ which contain all of the extreme points are required to be considered. Therefore, only the *minimal face* (the intersection of all these faces) is desirable. It will be shown that the SDP relaxation (77) is the projection of the SDP relaxation (70) onto the minimal face of $\mathcal{F}$.

Solving the relaxation model in (70) over $\mathcal{F}$ results in the optimal solution of the original problem in (67), but this problem is NP-hard. Solving the relaxation model in (77) over $\mathcal{F}_r$ results in a weaker bound for the optimal solution. In order to improve this bound, the relaxation is strengthen by adding an interesting property of the matrix $\mathbf{Y}_\mathbf{u}$. This results in the next relaxation model.

5.3.2 Tightening the Relaxation by Gangster Operator

The feasible set of the minimization problem (77) is convex. It contains the set of matrices of the form $\mathbf{Y}_\mathbf{u}$ corresponding to different vectors $\mathbf{u}$. However, the SDP relaxations may contain many points that are not in the affine hull of these $\mathbf{Y}_\mathbf{u}$. In the following, a condition which is implicit in the matrix $\mathbf{Y}_\mathbf{u}$ is extracted and explicitly is added to the relaxation model (77). Subsequently, some redundant constraint are removed and this results in an improved relaxation (relaxation *Model III*).

Theorem 9 *Let $\mathcal{U}$ denote the set of all binary vectors $\mathbf{u} = \mathrm{vec}(\mathbf{U}^T)$, $\mathbf{U} \in \mathcal{E}_{N\times K}$. Define the* barycenter point, $\hat{\mathbf{Y}}$, *as the arithmetic mean of all the feasible points in the minimization problem (67); therefore,*

$$\hat{\mathbf{Y}} = \frac{1}{K^N} \sum_{\mathbf{u}\in\mathcal{U}} \mathbf{Y}_\mathbf{u} = \frac{1}{K^N} \sum_{\mathbf{u}\in\mathcal{U}} \left[\begin{array}{c|c} 1 & \mathbf{u}^T \\ \hline \mathbf{u} & \mathbf{u}\mathbf{u}^T \end{array}\right]. \tag{78}$$

Then:

i) $\hat{\mathbf{Y}}$ has (a) the value of 1 as its $(0,0)$ element, (b) N blocks of dimension $K \times K$ on its diagonal which are diagonal matrices with elements $1/K$, and (c) the first row and first column equal to the vector of its diagonal elements. The rest of the matrix is composed *of $K \times K$ blocks with all elements equal to $1/K^2$:*

$$
\begin{aligned}
\hat{\mathbf{Y}} &= \left[\begin{array}{c|cccc}
1 & \multicolumn{4}{c}{\frac{1}{K}\mathbf{e}_n^T} \\ \hline
 & \frac{1}{K}\mathbf{I}_K & \frac{1}{K^2}\mathbf{E}_K & \cdots & \frac{1}{K^2}\mathbf{E}_K \\
\frac{1}{K}\mathbf{e}_n & \vdots & \vdots & \ddots & \vdots \\
 & \vdots & \vdots & \ddots & \vdots \\
 & \frac{1}{K^2}\mathbf{E}_K & \cdots & \frac{1}{K^2}\mathbf{E}_K & \frac{1}{K}\mathbf{I}_K
\end{array}\right] \\
&= \begin{bmatrix} 1 \\ \frac{1}{K}\mathbf{e}_n \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{K}\mathbf{e}_n^T \end{bmatrix} \\
&+ \left[\begin{array}{c|c}
0 & \mathbf{0}_n^T \\ \hline
\mathbf{0}_n & \frac{1}{K^2}\mathbf{I}_N \otimes (K\mathbf{I}_K - \mathbf{E}_K)
\end{array}\right]; \qquad (79)
\end{aligned}
$$

ii) *rank*$(\hat{\mathbf{Y}}) = N(K-1)+1$;

iii) *The $NK+1$ eigenvalues of $\hat{\mathbf{Y}}$ are given in the vector*

$$\left(\frac{K+N}{K}, \frac{1}{K}\mathbf{e}_{N(K-1)}^T, \mathbf{0}_N^T\right)^T;$$

iv) *The null space of $\hat{\mathbf{Y}}$ can be expressed by $\mathcal{N}(\hat{\mathbf{Y}}) = \left\{u : u \in \mathcal{R}(\mathbf{T}^T)\right\}$, where the constraint matrix $\mathbf{T}$ is the following $N \times (NK+1)$ matrix*

$$\mathbf{T} = \left[\begin{array}{c|c} -\mathbf{e}_N & \mathbf{I}_N \otimes \mathbf{e}_K \end{array}\right];$$

v) *the range of $\hat{\mathbf{Y}}$ can be expressed by the columns of the $(NK+1) \times (N(K-1)+1)$ matrix $\hat{\mathbf{V}}$. Furthermore, $\mathbf{T}\hat{\mathbf{V}} = 0$.*

Proof: see Appendix A. ■

Remark 3 *The faces of the positive semi-definite cone are characterized by the null space of the points in their relative interior. The minimal face of the SDP problem contains matrices $\mathbf{Y}_\mathbf{u}$ and can be expressed as $\hat{\mathbf{V}}\mathcal{S}_{N(K-1)+1}\hat{\mathbf{V}}^T$. Thus, the SDP relaxation (77) is a projected relaxation onto the minimal face of the feasible set $\mathcal{F}$.*

Theorem 9 suggests a zero pattern for the elements of $\mathcal{F}$. Therefore, *Gangster Operator* [156] can be used to represent these constraints more efficiently. Let J be a set of indices, then this operator is defined as

$$(\mathcal{G}_J(\mathbf{Y}))_{ij} = \begin{cases} y_{ij} & \text{if } (i,j) \text{ or } (j,i) \in J \\ 0 & \text{otherwise.} \end{cases} \tag{80}$$

Considering the barycenter point, $\mathcal{G}_J(\hat{\mathbf{Y}}) = 0$ for

$$\begin{aligned} J = \{(i,j) : \ & i = K(p-1)+q, \ \ j = K(p-1)+r, \\ & q < r, q, r \in \{1, \cdots, K\}, p \in \{1, \cdots, N\}\}. \end{aligned} \tag{81}$$

Since $\hat{\mathbf{Y}}$ is a convex combination of all matrices in $\mathcal{U}$ with entries either 0 or 1; hence, from (81), it is clear that $\mathcal{G}_J(\mathbf{Y_u}) = 0$. Also, all the points from the feasible set $\mathcal{F}$ are the convex combination of $\mathbf{Y_u}$. Therefore,

$$\mathcal{G}_J(\mathbf{Y}) = 0, \qquad \forall \mathbf{Y} \in \mathcal{F}. \tag{82}$$

The feasible set of the projected SDP in (77) is tightened by adding the constraints $\mathcal{G}_J(\mathbf{Y}) = 0$. By combining these constraints and (77), there are some redundant constraints that can be removed to enhance the relaxation model. This is expressed in the following lemma.

Lemma 10 *Let* $\mathbf{R}$ *be an arbitrary* $(N(K-1)+1) \times (N(K-1)+1)$ *symmetric matrix with*

$$\mathbf{R} = \left[\begin{array}{c|ccc} r_{00} & \mathbf{R}_{01} & \cdots & \mathbf{R}_{0N} \\ \hline \mathbf{R}_{10} & \mathbf{R}_{11} & \cdots & \mathbf{R}_{1N} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{R}_{N0} & \mathbf{R}_{N1} & \cdots & \mathbf{R}_{NN} \end{array}\right], \tag{83}$$

where r_{00} *is a scalar,* $\mathbf{R}_{i0}$*, for* $i = 1, \cdots, N$ *are* $(K-1) \times 1$ *vectors and* $\mathbf{R}_{ij}$*, for* $i, j = 1, \cdots, N$*, are* $(K-1) \times (K-1)$ *blocks of* $\mathbf{R}$*. Theorem 8 states that* $\mathbf{Y} = \hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T$*. The matrix* $\mathbf{Y}$ *can also be partitioned as*

$$\mathbf{Y} = \left[\begin{array}{c|ccc} y_{00} & \mathbf{Y}_{01} & \cdots & \mathbf{Y}_{0N} \\ \hline \mathbf{Y}_{10} & \mathbf{Y}_{11} & \cdots & \mathbf{Y}_{1N} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{Y}_{N0} & \mathbf{Y}_{N1} & \cdots & \mathbf{Y}_{NN} \end{array}\right], \tag{84}$$

where y_{00} *is a scalar,* $\mathbf{Y}_{i0}$*, for* $i = 1, \cdots, N$ *are* $K \times 1$ *vectors and* $\mathbf{Y}_{ij}$*, for* $i, j = 1, \cdots, N$*, are* $K \times K$ *blocks of* $\mathbf{Y}$*. Then,*

1. $y_{00} = r_{00}$ *and* $\mathbf{Y}_{0i}\mathbf{e}_K = r_{00}$*, for* $i = 1, \cdots, N$.

2. $\mathbf{Y}_{0j} = \mathbf{e}_K^T \mathbf{Y}_{ij}$ *for* $i, j = 1, \cdots, N$.

Proof: Noting $\mathbf{TY} = \mathbf{0}$ (see Theorem 9), the proof follows. ■

If the Gangster operator is applied to (77), the following constraint would be redundant:

$$\operatorname{diag}(\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T) = (1, (\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T)_{0,1:n})^T. \tag{85}$$

Note that using Lemma 10, $\mathbf{Y}_{0j} = \mathbf{e}_K^T \mathbf{Y}_{jj}$ for $j = 1, \cdots, N$ and the off-diagonal entries of each $\mathbf{Y}_{jj}$ are zero. Therefore, by defining a new set $\bar{J} = J \cup \{0, 0\}$ and eliminating the redundant constraints, a new SDP relaxation model (**Model III**) is obtained:

$$\begin{aligned} \min \quad & \operatorname{trace}(\hat{\mathbf{V}}^T \mathcal{L}_{\mathbf{Q}} \hat{\mathbf{V}})\mathbf{R} \\ \text{s.t.} \quad & \mathcal{G}_{\bar{J}}(\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^{\mathbf{T}}) = \mathbf{E}_{00} \\ & \mathbf{R} \succeq 0, \end{aligned} \tag{86}$$

where $\mathbf{R}$ is an $(N(K-1)+1) \times (N(K-1)+1)$ matrix and $\mathbf{E}_{00}$ is an $(NK+1) \times (NK+1)$ all zero matrix except for a single element equal to 1 in its $(0,0)$th entry. With this new index set $\bar{J}$, all the redundant constraints can be removed while maintaining the SDP relaxation. The relaxation model in (86) corresponds to a tighter lower bound and has an interior point in its feasible set as shown in the following theorem.

Theorem 11 *The* $(N(K-1)+1) \times (N(K-1)+1)$ *matrix* $\hat{\mathbf{R}}$ *defined as*

$$\left[\begin{array}{c|c} 1 & \frac{1}{K}\mathbf{e}_{N(K-1)}^T \\ \hline \frac{1}{K}\mathbf{e}_{N(K-1)} & \frac{1}{K^2}(\mathbf{E}_{N(K-1)} + \mathbf{I}_N \otimes (K\mathbf{I}_{K-1} - \mathbf{E}_{K-1})) \end{array}\right] \tag{87}$$

is a strictly interior point of the feasible set for the relaxation problem (86).

Proof: The matrix $\hat{\mathbf{R}}$ is positive definite. The rest of the proof follows by showing $\hat{\mathbf{V}}\hat{\mathbf{R}}\hat{\mathbf{V}}^{\mathbf{T}} = \hat{\mathbf{Y}}$. ■

The relaxation in (86) is further tightened by considering the *non-negativity constraints* [123]. All the elements of the matrix $\mathbf{Y}$ which are not covered by the Gangster operator are greater than or equal to zero. These inequalities can be added to the set of constraints in (86), resulting in a stronger relaxation model (**Model IV**):

$$\begin{aligned} \min \quad & \operatorname{trace}(\hat{\mathbf{V}}^T \mathcal{L}_{\mathbf{Q}} \hat{\mathbf{V}})\mathbf{R} \\ \text{s.t.} \quad & \mathcal{G}_{\bar{J}}(\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^{\mathbf{T}}) = \mathbf{E}_{00} \\ & \mathcal{G}_{\hat{J}}(\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^{\mathbf{T}}) \geq 0 \\ & \mathbf{R} \succeq 0, \end{aligned} \tag{88}$$

where the set $\hat{J}$ indicates those indices which are not covered by $\bar{J}$.

Note that this model is considerably stronger than model (86) because non-negative constraints are also imposed in the model. The advantage of this formulation is that the number of inequalities can be adjusted to provide a trade-off between the strength of the bounds and the complexity of the problem. The larger number of the constraints in the model is, the better it approximates the optimization problem (67) (with an increase in the complexity).

The most common methods for solving SDP problems of moderate sizes (with dimensions on the order of hundreds) are Interior Point Methods (IPMs), whose computational complexities are polynomial, see e.g. [3]. There are a large number of IPM-based solvers to handle SDP problems, e.g., DSDP [15], SeDuMi [127], SDPA [70], etc. In the numerical experiments, DSDP and SDPA are used for solving (86), and SeDuMi is implemented for solving (88). Note that adding the non-negativity constraints increases the computational complexity of the model. Since the problem sizes of interest are moderate, the complexity of solving (88) with IPM solvers is tractable.

5.4 Randomization Method

Solving the SDP relaxation models (86) and (88) results in a matrix $\mathbf{R}$. This matrix is transformed to $\mathbf{Y}$ using $\mathbf{Y} = \hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T$, whose elements are between 0 and 1. This matrix has to be converted to a binary rank-one solution of (67), i.e. $\mathbf{Y}_\mathbf{u}$, or equivalently, a binary vector $\mathbf{u}$ as a solution for (65).

For any feasible point of (67), i.e. $\mathbf{Y}_\mathbf{u}$, the first row, the first column, and the vector of the diagonal elements of this symmetric matrix are equal to a binary solution for (65). For any matrix $\mathbf{Y}$ resulting from the relaxation problems (86) or (88), its first row, its first column, and the vector of its diagonal elements are equal. Therefore, the vector $\mathbf{u}$ is approximated by rounding off the elements of the first column of the matrix $\mathbf{Y}$. However, this transformation results in a loose upper bound on the performance. In order to improve the performance, $\mathbf{Y}$ is transformed to a binary rank-one matrix through a randomization procedure. An intuitive explanation of the randomization procedure is presented in [124]. In the following, two randomization algorithms are presented to transform $\mathbf{Y}$ to a binary rank-one matrix.

5.4.1 Algorithm I

Goemans and Williamson [51] introduced an algorithm that randomly transforms an SDP relaxation solution to a rank-one solution. This approach is used in [124] for the quasi-ML decoding of a PSK signalling. This technique is based on expressing the BPSK symbols by $\{-1, 1\}$ elements. After solving the relaxation problem in [124], the Cholesky factorization is applied to the $n \times n$ matrix $\mathbf{Y}$ and the Cholesky factor $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is computed, i.e.

$\mathbf{Y} = \mathbf{V}\mathbf{V}^T$. In [124], it is observed that one can approximate the solution of the distance minimization problem, $\mathbf{u}$, using $\mathbf{V}$, i.e. u_i is approximated using $\mathbf{v}_i$. Thus, the assignment of -1 or 1 to the vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is equivalent to specifying the elements of $\mathbf{u}$.

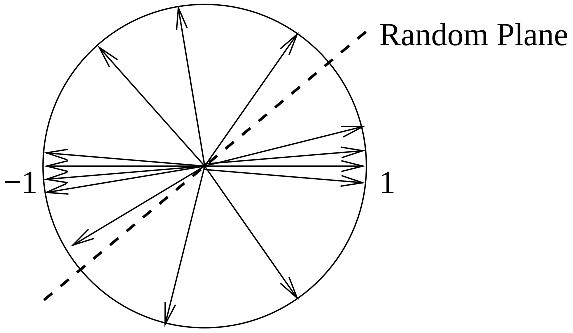

Figure 13: Representation of the randomization algorithm in [51]

It is shown that norms of the vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ are one, and they are inside an n-D unit sphere [124], see Figure 13. These vectors should be classified in two different groups corresponding to 1 and -1. In order to assign -1 or 1 to these vectors, the randomization procedure generates a random vector uniformly distributed in the sphere. This vector defines a plane crossing the origin. Among given vectors $\mathbf{v}_i$, $i = 1, \ldots n$, all the vectors at one side of the plane are assigned to 1 and the rest are assigned to -1, as shown in Figure 13. This procedure is repeated several times and the vector $\mathbf{u}$ resulting in the lowest objective function is selected as the answer.

In the proposed approach, the variables are binary numbers. In order to implement the randomization procedure of [51], the computed solution of the $\{0, 1\}$ SDP formulation is bijectively mapped to the solution of the corresponding $\{-1, 1\}$ SDP formulation. More precisely, the following mapping is used:

$$\mathbf{M} = \left[\begin{array}{c|cccc} 1 & 0 & \cdots & 0 \\ \hline -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & \cdots & 2 \end{array}\right],$$

$$\mathbf{Y}_{\{-1,1\}} = \mathbf{M}\mathbf{Y}_{\{0,1\}}\mathbf{M}^T, \tag{89}$$

where $\mathbf{Y}_{\{0,1\}}$ is the resulting matrix from the relaxation model (86) or (88) and $\mathbf{Y}_{\{-1,1\}}$ is its corresponding matrix with $\{-1, 1\}$ elements. Using (89), the solution for (65) can be computed using a similar randomization method as in [124]. The computational complexity of this randomization algorithm is polynomial [124].

Considering zero-one elements in the proposed SDP models, a new randomization procedure inspired by [51] is proposed. This algorithm can be applied to $\{0, 1\}$ formulation directly.

Therefore, the complexity of the whole randomization procedure is reduced, since the preprocessing step, i.e. bijective mapping in (89), is omitted.

5.4.2 Algorithm II

After solving the relaxation model (86) or (88), the Cholesky factorization of $\mathbf{Y}$ results in a matrix $\mathbf{V} = [\mathbf{v_1}, \ldots, \mathbf{v_n}]$ such that $\mathbf{Y} = \mathbf{V}\mathbf{V}^T$. The matrix $\mathbf{Y}$ is neither binary nor rank-one. Therefore, norms of the resulting vectors $\mathbf{v}_i$ are between zero and one. These vectors are depicted in Figure 14. Intuitively, a sphere with a random radius uniformly distributed between zero and one has the same functionality as the random plane in Figure 13.

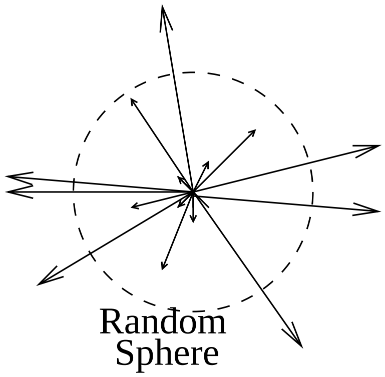

Figure 14: Graphic representation for the proposed randomization algorithm

In order to assign 0 or 1 to these vectors, the randomization procedure generates a random number, uniformly distributed between 0 and 1, as the radius of the sphere. Among given vectors $\mathbf{v}_i$, $i = 1, \ldots, n$, all the vectors whose norms are larger than this number are assigned to 1 and the rest are assigned to 0. In another variation of this algorithm, the radius of the sphere can be fixed, and norms of these vectors are multiplied by a random number. This procedure is repeated several times and the vector $\mathbf{u}$ resulting in the smallest objective function value in (65) is selected as the solution. Simulation results confirm that the proposed method results in a slightly better performance for the lattice decoding problem compared to the first algorithm. Also, the computational complexity of the randomization algorithm is decreased, due to the removal of the preprocessing step in (89). It is worth mentioning that the randomization procedure can also be implemented for the matrix $\mathbf{R}$, which results in further reduction in the computational complexity.

5.5 Complexity Reduction Using Lattice Basis Reduction

Assume that an initial solution for the lattice decoding problem is computed using one of the simple sub-optimal algorithms such as ZFD or channel inversion, e.g. $\mathbf{s}' = \left[\mathbf{H}^{-1}\mathbf{y}\right]$. If the channel is not ill-conditioned, i.e. the columns of the channel matrix are nearly orthogonal and short, it is most likely that the ML solution of the lattice decoding problem is around $\mathbf{s}'$. Therefore, using a reduced basis for the lattice, each x_i in (59) can be expressed by a few points around s'_i, not all the points in the set $\{s_1, \cdots, s_K\}$. In general, this results in a sub-optimal algorithm. However, for the special case of a MIMO system with two antennas (with real coefficients), it has been shown that by using the LLL approximation and considering two points per dimension the ML decoding performance is achieved [151].

Let $\mathbf{L} = \mathbf{HQ}$ be the LLL reduced basis for the channel matrix $\mathbf{H}$, where $\mathbf{Q}$ is a unimodular matrix. The MIMO system model in (54) can be written as

$$\mathbf{y} = \mathbf{L}\mathbf{Q}^{-1}\mathbf{x} + \mathbf{n}. \tag{90}$$

Consider the QAM signaling. Without loss of generality, it can be assumed that coordinates of $\mathbf{x}$ are in the integer grid. Since $\mathbf{Q}$ is a unimodular matrix, the coordinates of a new variable defined as $\mathbf{x}' = \mathbf{Q}^{-1}\mathbf{x}$ are also in the integer grid. Therefore, the system in (90) is modelled by $\mathbf{y} = \mathbf{L}\mathbf{x}' + \mathbf{n}$. Note that by multiplying $\mathbf{x}$ by $\mathbf{Q}^{-1}$ the constellation boundary will change. However, it is shown that in the lattice decoding problem with finite constellations the best approach is to ignore the boundary and compute the solution [34]. If the solution is outside the region, it is considered as an error. This change of boundary will result in some performance degradation. The performance degradation for some scenarios are depicted in Figure 19 and Figure 20.

In order to implement the proposed method using LLL basis reduction, each component of $\mathbf{x}'$ is expressed by a linear combination (with zero-one coefficients) of L (usually much smaller than K) integers around s'_i, where $\mathbf{s}' = \left[\mathbf{L}^{-1}\mathbf{y}\right]$. Then, the proposed algorithm can be applied to this new model. Due to the change of constellation boundary, there is a degradation in the performance. However, the complexity reduction is large. The trade-off between performance degradation and complexity reduction can be controlled by the choice of L (see simulation results). The reduction in the complexity is more pronounced for larger constellations. Note that the dimension of the semi-definite matrix $\mathbf{Y}$ is $N * (K - 1) + 1$. Therefore, the LLL reduction decreases the dimension of the matrix $\mathbf{Y}$ to $N * (L - 1) + 1$ (where usually $L \ll K$), and consequently, decreases the computational complexity of the proposed algorithm. The performance of this method is shown in the simulation results.

5.6 *Extension for Soft Decoding*

In this section, the proposed SDP relaxation decoding method is extended for soft decoding in MIMO systems. The SDP soft decoder is derived as an efficient solution of the Max-Log approximated soft ML decoder. The complexity of this method is much less than that in the soft ML decoder. Moreover, the performance of the proposed method is comparable with that in the ML decoder. Also, the proposed method can be applied to any arbitrary constellation and labeling method, say Gray labeling.

In the MIMO system defined in (54), any transmit data $\mathbf{x}$ is represented by $N_b = \log_2 K$ bits ($\mathbf{x} = \text{map}(\mathbf{b})$, where $\mathbf{b}$ is the corresponding binary input). Given a received vector $\mathbf{y}$, the soft decoder returns the soft information about the likelihood of $b_j = 0$ or $1,\ j = 1, \cdots, NN_b$. The likelihoods are calculated by Log-Likelihood Ratios (LLR) in a Maximum A Posteriori (MAP) decoder by

$$\mathcal{L}(b_j|\mathbf{y}) = \log\left(\frac{P(b_j = 1|\mathbf{y})}{P(b_j = 0|\mathbf{y})}\right). \tag{91}$$

Define

$$\mathcal{L}_A(b_j|\mathbf{y}) = \log\frac{P(b_j = 1)}{P(b_j = 0)}. \tag{92}$$

It is shown that the LLR values are formulated by [63]

$$\begin{aligned}\mathcal{L}(b_j|\mathbf{y}) \quad &= \quad \underbrace{\log\frac{\sum_{\mathbf{b}\in\mathbb{B}_{k,1}} p(\mathbf{y}|\mathbf{b}).\exp\left(\frac{1}{2}\mathbf{b}_{[k]}^T.\mathbf{L}_{A,[k]}\right)}{\sum_{\mathbf{b}\in\mathbb{B}_{k,0}} p(\mathbf{y}|\mathbf{b}).\exp\left(\frac{1}{2}\mathbf{b}_{[k]}^T.\mathbf{L}_{A,[k]}\right)}}_{\mathcal{L}_E(b_k|\mathbf{y})} \\ &+ \quad \mathcal{L}_A(b_j|\mathbf{y}),\end{aligned} \tag{93}$$

where $\mathbf{b}_{[k]}$ denotes the sub-vector of $\mathbf{b}$ obtained by omitting its kth element b_k, $\mathbf{L}_{A,[k]}$ denotes the vector of all $\mathcal{L}_A$ values, also omitting b_k, and $\mathbb{B}_{k,1}$ (resp. $\mathbb{B}_{k,0}$) denotes the set of all input vectors, $\mathbf{b}$, such that $b_k = 1$ (resp. $b_k = 0$). Note that there is an isomorphism between $\mathbb{B}_{k,1}$ (resp. $\mathbb{B}_{k,0}$) and $\mathbb{X}_{k,1}$ (resp. $\mathbb{X}_{k,0}$), where $\mathbb{X}_{k,1}$ (resp. $\mathbb{X}_{k,0}$) denotes the set of all corresponding constellation symbols, $\mathbb{X}_{k,1} = \{\mathbf{x} : \mathbf{x} = \text{map}(\mathbf{b}), \mathbf{b} \in \mathbb{B}_{k,1}\}$ (resp. $\mathbb{X}_{k,0} = \{\mathbf{x} : \mathbf{x} = \text{map}(\mathbf{b}), \mathbf{b} \in \mathbb{B}_{k,0}\}$).

As shown in [63], the computation of the LLR values in (93) requires computing the likelihood function $p(\mathbf{y}|\mathbf{b})$, i.e.

$$p(\mathbf{y}|\mathbf{x} = \text{map}(\mathbf{b})) = \frac{\exp\left[-\frac{1}{2\sigma^2}.\parallel \mathbf{y} - \mathbf{Hx} \parallel^2\right]}{(2\pi\sigma^2)^N}, \tag{94}$$

where $\sigma^2 = \frac{1}{SNR}$.

By having the likelihood functions, these LLR values are approximated efficiently using the Max-Log approximation [63]

$$\begin{aligned}\mathcal{L}_E(b_k|\mathbf{y}) \approx &+\frac{1}{2}\max_{\mathbf{b}\in\mathbb{B}_{k,1}}\left\{-\frac{1}{\sigma^2}\parallel \mathbf{y}-\mathbf{Hx} \parallel^2 + \mathbf{b}_{[k]}^T.\mathbf{L}_{A,[k]}\right\} \\ &-\frac{1}{2}\max_{\mathbf{b}\in\mathbb{B}_{k,0}}\left\{+\frac{1}{\sigma^2}\parallel \mathbf{y}-\mathbf{Hx} \parallel^2 + \mathbf{b}_{[k]}^T.\mathbf{L}_{A,[k]}\right\}.\end{aligned} \tag{95}$$

Without loss of generality, it can be assumed that all components, x_i, of an input vector are equiprobable[3]; therefore, the second term in each maximization in (95) will be removed. Hence, computing the LLR values requires to solve problems of the form

$$\min_{\mathbf{x}\in\mathbb{X}_{k,\zeta}} \parallel \mathbf{y} - \mathbf{Hx} \parallel^2, \tag{96}$$

where $k = 1, \cdots, NN_b$ and $\zeta = 0$ or 1. Note that, as mentioned in [142], only $NN_b + 1$ problems among $2NN_b$ problems of the form (96) are considered.

The quasi-ML decoding method proposed in this paper can be applied to the problem (96). However, $\mathbb{X}_{k,\zeta}$ must be defined in implementing the algorithm. This set includes all the input vectors, $\mathbf{x} \in \mathcal{S}_N$, such that $b_k = \zeta$. Assigning 0 or 1 to one of the bits in $\mathbf{b}$ removes half of the points in $\mathcal{S}_N$. In other words, when $b_k = \zeta$, one of the components of the input vector $\mathbf{x}$, say x_p, can only select half of the points in the set $\{s_1, \cdots, s_K\}$, say $\left\{s_{p_1}, \cdots, s_{p_{\frac{K}{2}}}\right\}$. Therefore, the pth component of $\mathbf{x}$ is represented by

$$x_p = u_p(1)s_{p_1} + \cdots + u_p(\frac{K}{2})s_{p_{\frac{K}{2}}}. \tag{97}$$

As a result, the same matrix expression as (59) is obtained, except that the length of the vector $\mathbf{u}$ is $(N-1) * K + \frac{K}{2}$. Now, the proposed method can be applied to the new equation based on the new vector $\mathbf{u}$.

5.7 Simulation Results

5.7.1 Performance Analysis

The two proposed Models III and IV are simulated for decoding in MIMO systems with QAM and PSK constellations. Figure 15 demonstrates that the proposed quasi-ML method using Model III and the randomization procedure achieves near ML performance in an un-coded 2×2 MIMO system with QPSK constellation. Figure 16 shows the performance in a 4×4 MIMO system with 16-QAM. The performance analysis of a MIMO system with different number of antennas employing 8-PSK is shown in Figure 17. In Figures 15, 16, and 17, the curved lines with the stars represent the performance of the system using relaxation Model III, while a simple rounding algorithm, as described in Section 5.4, transforms matrix $\mathbf{Y}$ to the binary vector $\mathbf{u}$. The ML decoding performance is also denoted by a curved line with circles. By increasing the dimension, the resulting gap between the relaxation Model III and the ML decoding increases. However, using the randomization Algorithm I with $M_{rand} = 30$ to 50

[3] In order to consider the effects of non-equiprobable symbols, both approaches presented in [125] and [142] can be applied.

significantly decreases this gap (curved line with diamonds). The curved lines with squares show the performance of the relaxation Model IV with a simple rounding, in which all the non-negative constraints are included. This curve is close to ML performance. It is clear that the relaxation Model IV is much stronger than the relaxation Model III. Note that adopting different number of non-negative constraints will change the performance of the system between the two curves with diamonds and squares. In other words, the trade-off between complexity and performance relies on the number of extra non-negative constraints.

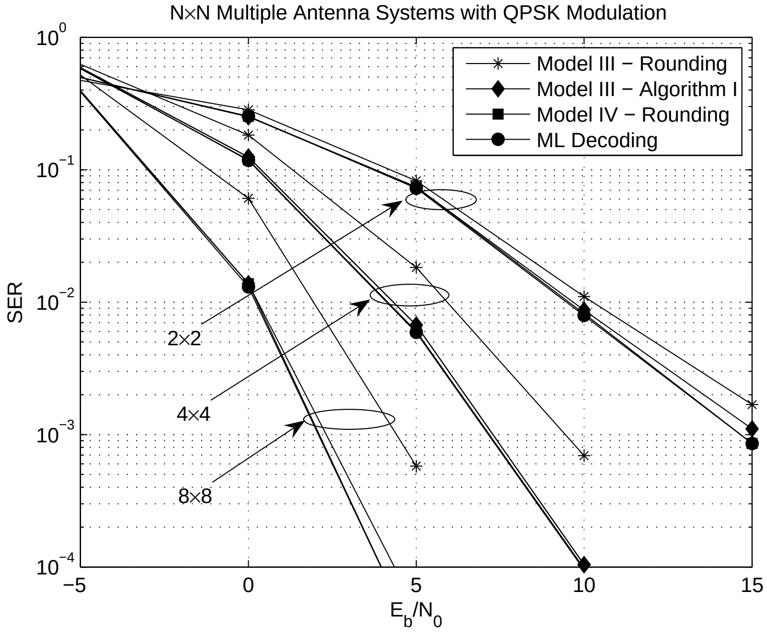

Figure 15: Performance of the proposed Models III and IV in a $N \times N$ MIMO system employing QPSK

Figure 18 compares the two proposed Randomization procedure for the relaxation Models III and IV. The effect of the randomization methods, Algorithm I and II, for the relaxation Model III is shown. As expected, Algorithm II performs slightly better, while its computational complexity is lower. The solution of the relaxation model in (88), in most cases, corresponds to the optimal solution of the original problem (65). In the other words, because the model in (88) is strong enough, there is no need for the randomization algorithm. Several compromises for improving the performance can be done, e.g. including only some of the non-negative constraints in (88) and/or using a randomization procedure with a fewer number of iterations.

In order to reduce the computational complexity of the proposed method, the LLL lattice basis reduction is implemented as a pre-processing step for the relaxation Model IV. Figure 19 and Figure 20 show the effect of using the LLL lattice basis reduction in 2×2 and 4×4 multiple antenna systems with 64-QAM and 256-QAM. In a system with 64-QAM and 256-QAM,

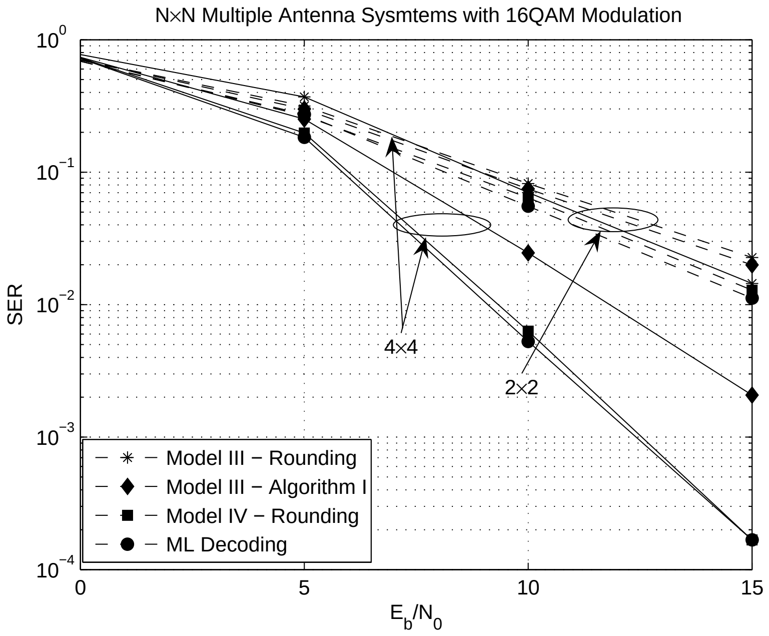

Figure 16: Performance of the proposed Models III and IV in a $N \times N$ MIMO system employing 16-QAM

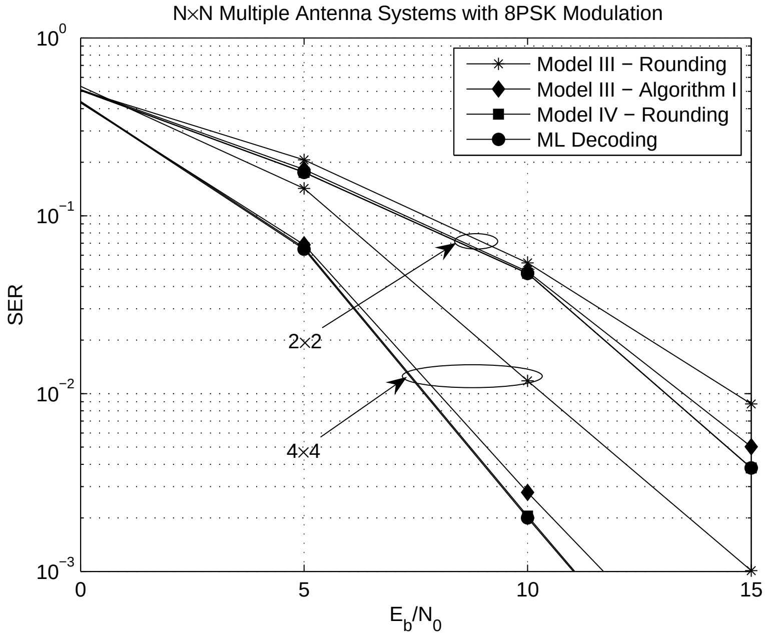

Figure 17: Performance of the proposed Models III and IV in a $N \times N$ MIMO system employing 8-PSK

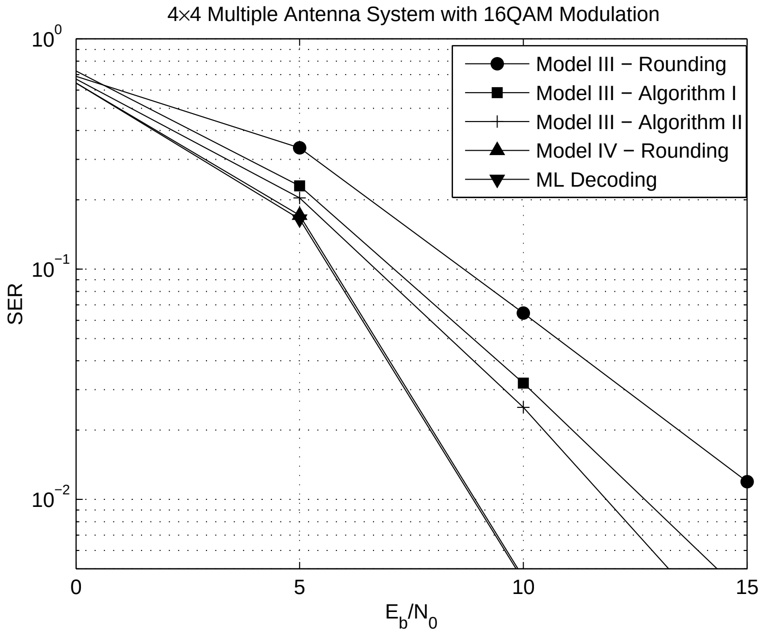

Figure 18: Different randomization algorithms in a 4 × 4 MIMO system employing 16-QAM

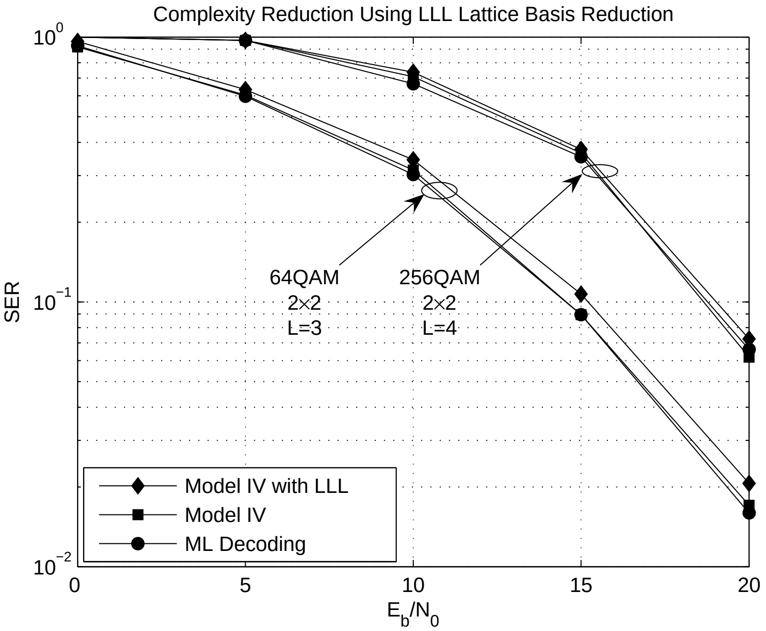

Figure 19: Performance of using LLL lattice basis reduction for relaxation Model IV in a 2×2 MIMO system with $L = \log 2(K)$

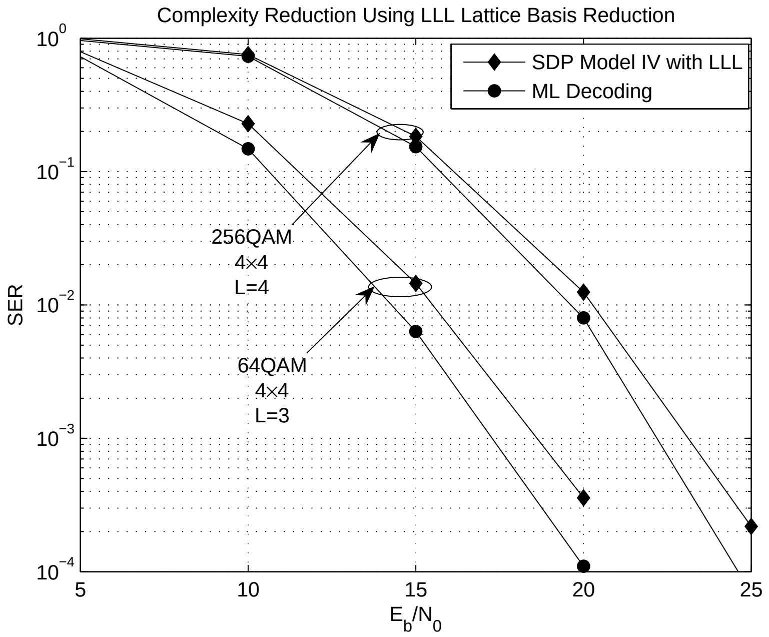

Figure 20: Performance of using LLL lattice basis reduction for relaxation Model IV in a 4×4 MIMO system with $L = \log 2(K)$

the performance of the relaxation Model IV is close to the ML performance with $K = 8$ and $K = 16$, respectively. By using LLL reduction and considering $L = \log_2(K)$ symbols around the initial point, the performance degradation is acceptable, see Figure 19 and Figure 20. Note that the resulting gap in the performance is small, while the reduction in computational complexity is substantial.

5.7.2 Complexity Analysis

Semi-definite programs of reasonable size can be solved in polynomial time within any specified accuracy by IPMs. IPMs are iterative algorithms which use a Newton-like method to generate search directions to find an approximate solution to the nonlinear system. The IPMs converge vary fast and an approximately optimal solution is obtained within a polynomial number of iterations. For a survey on IPMs see [37, 152]. In the sequel, an analysis for the worst case complexity of solving Models III and IV by IPMs is provided.

It is known (see e.g. [137]) that a SDP with rational data can be solved, within a tolerance ϵ, in $O(\sqrt{m}\log(1/\epsilon))$ iterations, where m is the dimension of the matrix variable. Note that for the SDP problems in (86) and (88), $m = N(K-1)+1$.

The computational complexity for one interior-point iteration depends on several factors. The main computational task in each iteration is solving a linear system of order determined by the number of constraints, p. This task requires $O(p^3)$ operations. The remaining computational tasks involved in one interior-point iteration include forming system matrix whose total construction requires $O(pm^3 + p^2m^2)$ arithmetic operations. Thus, the complexity per iteration of the IPM for solving SDP problem whose matrix variable is of dimension m and number of equality constraints p, is $O(pm^3 + p^2m^2 + p^3)$. This means for a given accuracy ϵ, an interior-point method in total requires at most $O(p(m^3 + pm^2 + p^2)\sqrt{m}\log(1/\epsilon))$ arithmetic operations.

Since the SDP relaxation Model III contains $O(K^2N)$ equality constraints, it follows that a solution to (86) can be found in at most $O(N^{4.5}K^{6.5}\log(1/\epsilon))$ arithmetic operations. SDP relaxation Model IV contains $O(K^2N)$ equations and $O(K^2N^2)$ sign constraints. In order to solve relaxation (88), the SDP model is formulated as a standard linear cone program (see e.g. [128]) by adding some slack variables. The additional inequality constraints make the model in (88) considerably stronger than the model in (86) (see numerical results), but also more difficult to solve. An IPM for solving SDP Model IV within a tolerance ϵ requires at most $O(N^{6.5}K^{6.5}\log(1/\epsilon))$ arithmetic operations. Since the problem sizes of interest are moderate, the problem in (88) is tractable. However, there exist a trade-off between the strength of the bounds and the computational complexity for solving these two models (see Section 5.3).

The complexity of the randomization procedure applied to the model (86) is negligible compared to that of solving the problem itself. Namely, if the number of randomization iterations is denoted by N_{rand}, then the worst case complexity of the randomization procedure is $O(NKN_{rand})$.

The optimization problems (86) and (88) are polynomially solvable. These problems have many variables; however, they contain sparse low-rank (rank-one) constraint matrices. Exploiting the structure and sparsity characteristic of semi-definite programs is crucial to reduce the complexity. In [14], it is shown that rank-one constraint matrices (similar to the proposed models) reduce the complexity of the interior-point algorithm for positive semi-definite programming by a factor of NK. In other words, the complexities of the SDP relaxation problems (86) and (88) are decreased to $O(N^{3.5}K^{5.5}\log(1/\epsilon))$ and $O(N^{5.5}K^{5.5}\log(1/\epsilon))$, respectively. Also, implementing the rank-one constraint matrices results in a faster convergence and a saving in the computation time and memory requirements.

Remark 4 *When the LLL lattice basis reduction is used in conjunction with Model III and IV, the value of K is replaced with L in the aforementioned analysis. As mentioned before, this value is much smaller than K, e.g. in the simulation results $L = \log_2(K)$, which results in reducing the computational complexity.*

5.7.3 Comparison

The *worst-case complexity* of the SD method [58, 1] is known to be an exponential function of dimension M over all ranges of rate and SNR [58]. The complexity analysis shows that the proposed SDP algorithms possess a polynomial-time worst case complexity. It should be emphasized that in real time problems, the time spent for decoding the received vector is important and it can be considered as a measure of the complexity.

In the following, the worst case complexities of the algorithm based on Model III, the method proposed in [143], the method in [124], and the SD algorithm [1] are compared with different random values of input vector, channel matrix, and noise for

$$E_b/N_0 = \{-5, 0, 5, 10, 15\}dB.$$

For each value of E_b/N_0, the algorithms are performed for 10^5 times and the maximum time spent for the decoding procedure is saved in $MaxTime$. The average time spent for decoding each case is stored in $AveTime$ (all provided numbers for $AveTime$ and $MaxTime$ are in seconds).

It should be emphasized that the $MaxTime$ for each case depends on how the algorithm is implemented. There are numerous variants for SD algorithm. In the following, the SD algorithm is implemented based on the Schnorr-Euchner strategy proposed in [1]. Moreover, the simulations of the proposed algorithms are implemented by one of the simplest available packages, the SDPA package [70]. However, by utilizing the sparsity of the constraint matrices as suggested in [14] and using the DSDP package, the computed $AveTime$ and $MaxTime$ can be reduced dramatically (a factor of NK in the analysis), without any performance degradation.

Table 2: Comparison of $MaxTime$ for different methods in a 4×4 MIMO system employing 16-QAM

E_b/N_0	-5	0	5	10	15
Model III	0.1037	0.1095	0.1108	0.1178	0.1196
Method [143]	0.0685	0.0640	0.0697	0.0735	0.0624
Method [124]	0.0580	0.0633	0.0536	0.0646	0.0596
SD Method	61.8835	47.0480	28.0347	4.3848	2.2477

Table 2 shows the simulation results for a MIMO system with $\tilde{M} = \tilde{N} = 4$ employing 16-QAM. The maximum time for decoding a symbol using SD algorithm is much longer than the corresponding time in the proposed SDP relaxation method. The other three methods have comparable $MaxTime$. As it is also shown in the analysis, the proposed Model III is more

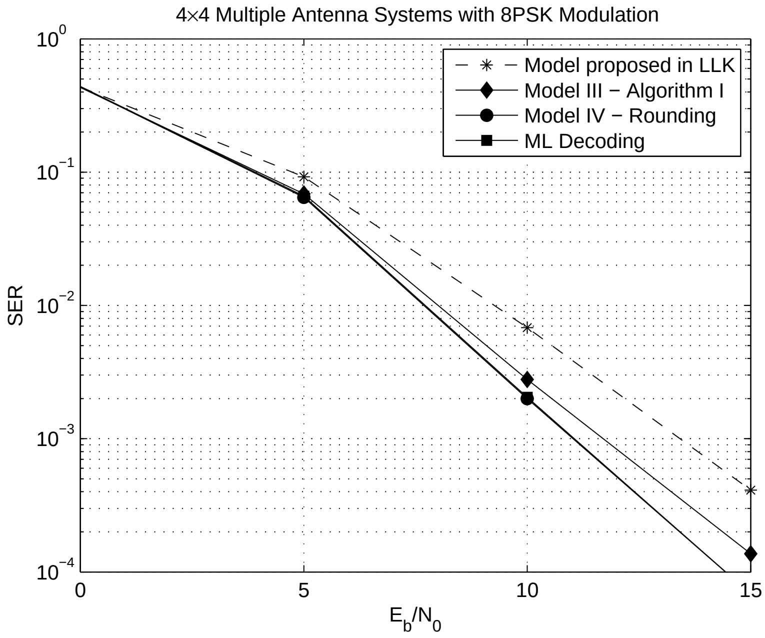

Figure 21: Comparison of the relaxation model proposed in LLK [85] and that in the proposed method in a 4×4 MIMO system employing 8-PSK modulation

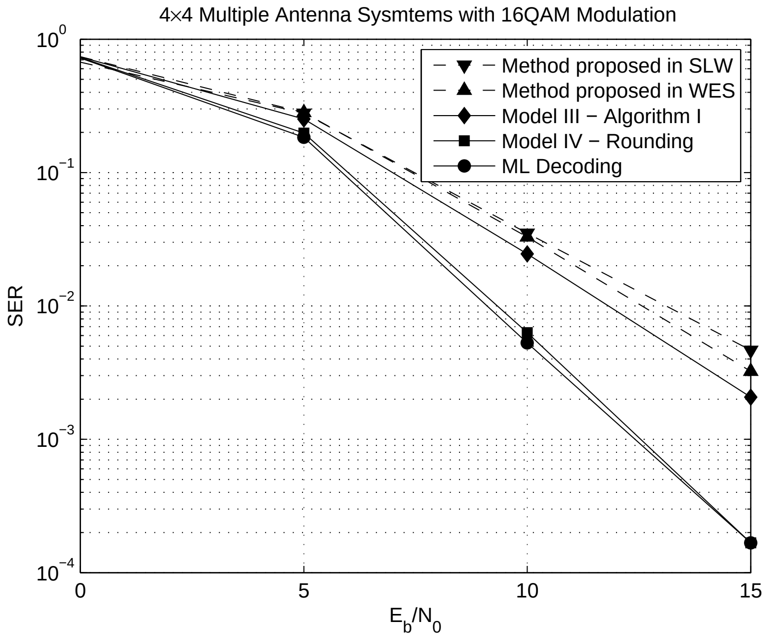

Figure 22: Comparison of the relaxation model proposed in SLW [124], WES [143] and that in our proposed method in a 4×4 MIMO system employing 16-QAM modulation

complex compared to the two other SDP methods. However, this method outperforms the other SDP methods in [124] and [143].

The relaxation Model III outperforms the SDP methods proposed in papers [124] and [85]. Figure 21 compares the performance of [85] and the relaxation Model III and the performance of the method proposed in [124] is shown in Figure 22 in a 4×4 MIMO system. The order of the complexity of [124] is comparable to the proposed Model III and the order of the complexity of [85] is less than that of the Model III ($O(N^2)$ vs. $O(N^{3.5})$). The method in [124] can handle QAM constellations; however, it achieves near ML performance only in the case of BPSK and QPSK constellations. Also, the method in [85] is limited to PSK constellations. Note that the proposed models can be used for any arbitrary constellation and labeling.

The comparison of the performance of the relaxation model in [143] and that in the proposed method is shown in Figure 22 (4×4 antenna system employing 16-QAM). It is observed that the SDP relaxation Models III and IV perform better than [143]. The order of the complexity of [143] is the same as that of the model (77), while the Model IV is more complex ($O(N^{5.5})$ vs. $O(N^{3.5})$).

Although the worst case complexities of the SD algorithm [1, 58] and the other variants are exponential, in several papers, the average complexity of these algorithms are investigated. In [66], it is shown that generally, there is an exponential lower bound on the *average complexity* of the SD algorithm. However, it is shown that for large values of E_b/N_0 and small values of dimension M, the average complexity can be approximated by a polynomial function of dimension M.

Table 3: Comparison of $AveTime$ for different methods in a 4×4 MIMO system employing 16QAM

E_b/N_0	-5	0	5	10	15
Model III	0.0372	0.0377	0.0394	0.0428	0.0417
Method [143]	0.0130	0.0134	0.0142	0.0156	0.0156
Method [124]	0.0116	0.0118	0.0126	0.0141	0.0141
SD Method	0.0449	0.0139	0.0060	0.0026	0.0016

In Table 3, the average time $AveTime$ spent for decoding the received vectors in the previous scenario is shown. As it can be seen, the average complexities of all SDP methods are gradually increasing with E_b/N_0 while the average complexity of SD method is decreasing exponentially. This suggests that for different dimensions M and values of E_b/N_0, there is a threshold that the proposed SDP methods perform better than SD algorithm even in terms of the average complexity. However, Table 2 shows that how inefficient the SD algorithm performs in terms of the worst-case complexity.

Table 4: Decoding Time in a 4 × 4 MIMO system employing QPSK

	E_b/N_0	-5	0	5	10	15
$AveTime$	Model III	0.0154	0.0156	0.0238	0.0278	0.0236
	SD Method	0.0199	0.0074	0.0046	0.0028	0.0020
$MaxTime$	Model III	0.4271	0.4251	0.4765	0.7572	0.8417
	SD Method	28.326	26.3109	25.4260	2.2232	0.9663

The performance of the proposed algorithm based on Model III and SD algorithm are shown in terms of $AveTime$ and $MaxTime$ in Tables 4 and 5, for different number of antennas and constellations. It can be seen that, in terms of the worst-case complexity the proposed algorithm based on Model III always outperforms SD algorithm. Generally, it can be concluded that by increasing the dimension and rate, the range of E_b/N_0 that the proposed model outperforms the SD algorithm increases. In order to show that the $MaxTime$ values are not sporadic, the values of $AveMaxTime$ is also provided in Table 5. This number is the average of the largest 100 decoding times in each case.

Table 5: Decoding Time in a 8 × 8 MIMO system

		E_b/N_0	-5	0	5	10	15
QPSK	$AveTime$	Model III	0.0152	0.0152	0.0174	0.0224	0.0306
		SD Method	0.6005	0.1061	0.0319	0.0149	0.0052
	$MaxTime$	Model III	0.0965	0.0655	0.1666	0.6586	0.6959
		SD Method	433.3972	179.0310	19.7889	16.7787	7.7819
	$AveMaxTime$	Model III	0.0658	0.0587	0.0642	0.1492	0.2109
		SD Method	73.4830	16.3274	6.1652	5.9249	1.8074
16-QAM	$AveTime$	Model III	0.0936	0.0948	0.0984	0.1050	0.1059
		SD Method	42.2894	1.6575	0.4762	0.2955	0.1080
	$MaxTime$	Model III	0.2867	0.2974	0.2772	0.2916	0.3273
		SD Method	8633.8	383.40	290.89	121.92	91.032
	$AveMaxTime$	Model III	0.1574	0.1580	0.1633	0.1682	0.1712
		SD Method	411.6743	15.3724	4.5987	2.9336	1.0162

The performance of the proposed SDP relaxation model (88), Model IV, is close to the ML performance. Similar to the SDP relaxation model (86), the algorithm based on Model IV outperforms the SD algorithm in terms of the worst case complexity (polynomial vs. exponential). Furthermore, by using the LLL lattice basis reduction before the proposed SDP model, the complexity is reduced, with an acceptable degradation in the performance (as shown in simulation results).

As a final note, it must be emphasized that in the complexity analysis for Model IV, all the non-negative constraints are considered. This suggests that the complexity of this model is not tractable. However, it is not required to consider all the non-negative constraints. In order to implement this model more efficiently, the SDP relaxation (88) can be solved with only the *most violated constraints*. These constraints correspond to those positions in matrix $\mathbf{Y}$ where their values are the minimum negative numbers. Implementing Model IV based on the most violated constraints reduces the complexity to almost a number of times more complex compared to the Model III.

5.8 Conclusion

A method for quasi-ML decoding based on two semi-definite relaxation models is introduced. The proposed semi-definite relaxation models provide a wealth of trade-off between the complexity and the performance. The strongest model provides a near-ML performance with polynomial-time worst-case complexity (unlike the SD that has exponential-time complexity). Moreover, the soft decoding method based on the proposed models is investigated. By using lattice basis reduction the complexity of the proposed SDP decoding methods is reduced.

CHAPTER 6

MATRIX LIFTING SEMI-DEFINITE PROGRAMMING

Abstract – This chapter presents a computationally efficient decoder for multiple antenna systems. The decoder is based on Matrix Lifting Semi-Definite Programming (MLSDP). The strength of the proposed method lies in a new relaxation algorithm applied to the methods proposed in Chapter 5. This results in a reduction of the number of variables from $(NK + 1)^2$ to $(2N + K)^2$, where N is the number of antennas and K is the number of constellation points in each real dimension. Moreover, the proposed method offers a better performance as compared to the best quasi maximum likelihood decoding methods reported in the literature.

6.1 Introduction

In this chapter, a new algorithm based on Matrix Lifting Semi-Definite Programming (MLSDP) [40, 11] is introduced for any constellation (QAM or PSK) and any labeling method. This algorithm is inspired by the method proposed in Chapter 5 with an efficient implementation resulting in a better performance and lower computational complexity. In SDP optimization problems, the computational complexity is a polynomial function of the number of variables. Using the proposed method, the number of variables in Chapter 5 is decreased from $(NK + 1)^2$ to $(2N + K)^2$, where N is the number of antennas and K is the number of constellation points in each real dimension. Since the computational complexity of solving MLSDP is a polynomial function of the number of variables, a significant complexity reduction is achieved. In addition to this large reduction in the complexity, simulation results show that the proposed algorithm also outperforms all other known convex quasi-ML decoding methods, e.g. [143, 121, 150].

6.2 Matrix-Lifting Semi-Definite Programming

The SDP relaxation models proposed in Chapter 5 offer a large computational complexity. In order to solve the optimization problem in (65) based on VLSDP, the optimization parameter $\mathbf{U}$ is lifted to the vector $\mathbf{u}$, which results in the matrix $\mathbf{Y}$ with large dimensions. However, if the relaxation models are defined based on the matrix $\mathbf{U}$, a large reduction in the dimension of the variable parameters and computational complexity of the optimization problem can be achieved.

To keep the matrix $\mathbf{U}$ in its original form in (64), the idea is to use the constraint $\mathbf{X} = \mathbf{U}^T\mathbf{U}$. As a result, the relaxation is $\mathbf{X} \succeq \mathbf{U}^T\mathbf{U}$, or equivalently, by the Schur complement, $\begin{bmatrix} \mathbf{I}_N & \mathbf{U} \\ \mathbf{U}^T & \mathbf{X} \end{bmatrix} \succeq 0$. This is known as matrix-lifting semi-definite programming. Define the new variable $\mathbf{V} = \mathbf{US}$. Since the matrix $\mathbf{S}$ is symmetric, the objective function in (64) can be represented as the Quadratic Matrix Program [11]

$$\begin{aligned} & \text{trace} \left(\begin{bmatrix} \mathbf{U}^T & \mathbf{V}^T \end{bmatrix} \begin{bmatrix} \mathbf{0} & \frac{1}{2}\mathbf{Q} \\ \frac{1}{2}\mathbf{Q} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} + 2\mathbf{CU} \right) \\ = \; & \text{trace} \left(\begin{bmatrix} \mathbf{0} & \frac{1}{2}\mathbf{Q} \\ \frac{1}{2}\mathbf{Q} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{U}^T & \mathbf{V}^T \end{bmatrix} + 2\mathbf{CU} \right) \\ = \; & \text{trace} \left(\mathcal{L}_{QC} \mathbf{W}_{\mathbf{U}} \right), \end{aligned} \tag{98}$$

where

$$\mathcal{L}_{QC} = \begin{bmatrix} \mathbf{0} & \mathbf{C} & \mathbf{0} \\ \mathbf{C}^T & \mathbf{0} & \frac{1}{2}\mathbf{Q} \\ \mathbf{0} & \frac{1}{2}\mathbf{Q} & \mathbf{0} \end{bmatrix} \tag{99}$$

and

$$\mathbf{W}_{\mathbf{U}} = \begin{bmatrix} \mathbf{I} & \mathbf{U}^T & \mathbf{V}^T \\ \mathbf{U} & \mathbf{UU}^T & \mathbf{UV}^T \\ \mathbf{V} & \mathbf{VU}^T & \mathbf{VV}^T \end{bmatrix}. \tag{100}$$

To linearize $\mathbf{W}_{\mathbf{U}}$, consider the matrix

$$\begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{U}^T & \mathbf{V}^T \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}, \tag{101}$$

where $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathcal{S}_N$. This equality can be relaxed to

$$\begin{bmatrix} \mathbf{UU}^T & \mathbf{UV}^T \\ \mathbf{VU}^T & \mathbf{VV}^T \end{bmatrix} - \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix} \preceq 0. \tag{102}$$

It can be shown that this relaxation is convex in the Löwner partial order and it is equivalent to the linear constraint [40]

$$\mathbf{W} \triangleq \begin{bmatrix} \mathbf{I} & \mathbf{U}^T & \mathbf{V}^T \\ \mathbf{U} & \mathbf{X} & \mathbf{Y} \\ \mathbf{V} & \mathbf{Y} & \mathbf{Z} \end{bmatrix} \succeq 0. \tag{103}$$

On the other hand, the feasible set in (64) is the set of binary matrices in $\mathcal{M}_{N\times K}$ with row sum equal to one, the set $\mathcal{E}_{N\times K}$ in (63). By relaxing the rank-one constraint for the matrix variable in (98), a tractable SDP problem is obtained. The feasible set for the objective function in (98) is approximated by

$$\begin{aligned} \mathcal{F}_{\mathcal{M}} = \text{conv}\,\{\mathbf{W}_{\mathbf{U}} \ \mid \mathbf{U} \in \mathcal{M}_{N\times K} : \mathbf{U}\mathbf{e}_K = \mathbf{e}_N, \\ u_{ij} \in \{0,1\}, \forall i, j; \mathbf{V} = \mathbf{U}\mathbf{S}\Big\} \end{aligned} \tag{104}$$

Therefore, the decoding problem can be represented by

$$\begin{aligned} \min \quad & \text{trace}\,(\mathcal{L}_{QC}\mathbf{W}) \\ s.t. \quad & \mathbf{W} \in \mathcal{F}_{\mathcal{M}}. \end{aligned} \tag{105}$$

Note that the size of matrix $\mathbf{W}$ is $(2N + K) \times (2N + K)$, compared to $(NK + 1) \times (NK + 1)$ in Chapter 5. In SDP optimization problems, the computational complexity is a polynomial function of the number of variables (elements of $\mathbf{W}$). By the new implementation of (105), the number of variables in Chapter 5 is decreased from $(NK + 1)^2$ to $(2N + K)^2$, resulting in a large reduction in the complexity.

Although the rank constraint in (101) is relaxed, some more additional linear constraints can be considered to further improve the quality of the solution. These constraints are valid for the non-convex rank-constrained decoding problem. However, the SDP problem is forced to satisfy these constraints. Consider the auxiliary matrix $\mathbf{V}$ and the symmetric matrices $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ in matrix $\mathbf{W}$. Since $\mathbf{U} \in \mathcal{E}_{N\times K}$ and $\sum_{j=1}^{N} u_{ij}^2 = 1$, it is clear that $\text{diag}(X) = \mathbf{e}_N$. Also, $\mathbf{Y}$ represents $\mathbf{U}\mathbf{S}\mathbf{U}^T$ and $\mathbf{Z}$ represents $\mathbf{U}\mathbf{S}^2\mathbf{U}^T$. It is easy to show that

$$\text{diag}(\mathbf{Y}) = \mathbf{U}\text{diag}(\mathbf{S}) \quad \text{and} \quad \text{diag}(\mathbf{Z}) = \mathbf{U}\text{diag}(\mathbf{S^2}). \tag{106}$$

Moreover, $\mathbf{S} = \mathbf{s}\mathbf{s}^T$ (rank-one matrix) and $\mathbf{S}^2 = (\sum_{1=i}^{K} s_i^2)\mathbf{S}$. Therefore, instead of $\text{diag}(\mathbf{Z}) =$

$\mathbf{U}\mathrm{diag}(\mathbf{S^2})$, a stronger result for $\mathbf{Z}$ is $\mathbf{Z} = (\sum_{1=i}^{K} s_i^2)\mathbf{Y}$. Therefore,

$$
\begin{aligned}
\min \quad & \mathrm{trace}\left(\mathcal{L}_{QC}\begin{bmatrix} \mathbf{I} & \mathbf{U}^T & \mathbf{V}^T \\ \mathbf{U} & \mathbf{X} & \mathbf{Y} \\ \mathbf{V} & \mathbf{Y} & \mathbf{Z} \end{bmatrix}\right) \\
s.t. \quad & \mathbf{U}\mathbf{e}_K = \mathbf{e}_N \;\; ; \;\; \mathbf{U} \geq 0 \\
& \mathbf{V} = \mathbf{U}\mathbf{S} \\
& \mathrm{diag}(\mathbf{X}) = \mathbf{e}_N \\
& \mathrm{diag}(\mathbf{Y}) = \mathbf{U}\mathrm{diag}(\mathbf{S}) \\
& \mathbf{Z} = (\sum_{1=i}^{K} s_i^2)\mathbf{Y} \\
& \begin{bmatrix} \mathbf{I} & \mathbf{U}^T & \mathbf{V}^T \\ \mathbf{U} & \mathbf{X} & \mathbf{Y} \\ \mathbf{V} & \mathbf{Y} & \mathbf{Z} \end{bmatrix} \succeq 0 \\
& \mathbf{U}, \mathbf{V} \in \mathcal{M}_{N\times K}, \mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathcal{S}_N
\end{aligned} \tag{107}
$$

The equation in (106) determines the diagonal elements of $\mathbf{Y}$. This property is hidden in the special structure of $\mathbf{U}$, i.e. $\mathbf{U} \in \mathcal{E}_{N\times K}$. By using this property, more constraints can be added. The equation $\mathbf{Y} = \mathbf{U}\mathbf{S}\mathbf{U}^T$ implies that $y_{ij} = s_{kl}$ for some k and l. Therefore, the value of y_{ij} is between the minimum and the maximum elements of $\mathbf{S}$. In addition, it can be easily shown that in communication applications, $\mathbf{S}$, $\mathbf{Y}$, and $\mathbf{Z}$ are diagonal dominant matrices (since $\mathbf{s}^T\mathbf{e}_K = 0$). This property can be also used to add more constraints to improve the quality of the solution. Our studies show that the improvements due to including the above constraints are marginal. Therefore, in the sequel, the focus is on the form given in (107).

The objective function in (64) is $\mathrm{trace}\left(\mathbf{S}\mathbf{U}^T\mathbf{Q}\mathbf{U} + 2\mathbf{C}\mathbf{U}\right)$ which can be written as

$$
\mathrm{trace}\left(\mathbf{S}\mathbf{U}^T\mathbf{Q}\mathbf{U} + 2\mathbf{C}\mathbf{U}\right) = \mathrm{trace}\left(\mathbf{Q}\mathbf{U}\mathbf{S}\mathbf{U}^T + 2\mathbf{U}\mathbf{C}\right). \tag{108}
$$

Exchanging the role of $\mathbf{Q}$ and $\mathbf{S}$ in the proposed method results in two different formulations. Here, the auxiliary variable $\mathbf{V}$ is defined as $\mathbf{Q}\mathbf{U}$. Similarly, the auxiliary variables $\mathbf{X}, \mathbf{Y}$, and $\mathbf{Z}$ represents $\mathbf{U}^T\mathbf{U}, \mathbf{U}^T\mathbf{Q}\mathbf{U}$, and $\mathbf{U}^T\mathbf{Q}^2\mathbf{U}$, respectively. Therefore, it is easy to show that the equivalent minimization problem is

$$\begin{array}{ll} \min & \operatorname{trace}\left(\begin{bmatrix} \mathbf{0} & \mathbf{C} & \mathbf{0} \\ \mathbf{C}^T & \mathbf{0} & \frac{1}{2}\mathbf{S} \\ \mathbf{0} & \frac{1}{2}\mathbf{S} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{U} & \mathbf{V} \\ \mathbf{U}^T & \mathbf{X} & \mathbf{Y} \\ \mathbf{V}^T & \mathbf{Y} & \mathbf{Z} \end{bmatrix}\right) \\ s.t. & \mathbf{U}\mathbf{e}_K = \mathbf{e}_N \;\;;\;\; \mathbf{U} \geq 0 \\ & \mathbf{V} = \mathbf{Q}\mathbf{U} \\ & \operatorname{diag}(\mathbf{X}) = \mathbf{U}^T\mathbf{e}_N \;\;;\;\; X_{ij} = 0 \; i \neq j \\ & \mathbf{Y}\mathbf{e}_K = \mathbf{U}^T\mathbf{Q}\mathbf{e}_N \;\;;\;\; \operatorname{trace}(\mathbf{Y}\mathbf{E}_K) = \operatorname{trace}(\mathbf{Q}\mathbf{E}_N) \\ & \mathbf{Z}\mathbf{e}_K = \mathbf{U}^T\mathbf{Q}^2\mathbf{e}_N \;\;;\;\; \operatorname{trace}(\mathbf{Z}\mathbf{E}_K) = \operatorname{trace}(\mathbf{Q}^2\mathbf{E}_N) \\ & \begin{bmatrix} \mathbf{I} & \mathbf{U} & \mathbf{V} \\ \mathbf{U}^T & \mathbf{X} & \mathbf{Y} \\ \mathbf{V}^T & \mathbf{Y} & \mathbf{Z} \end{bmatrix} \succeq 0 \\ & \mathbf{U}, \mathbf{V} \in \mathcal{M}_{N\times K}, \mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathcal{S}^K, \end{array} \tag{109}$$

where the size of the variable matrix is $(2K+N)$. Note that both (107) and (109) are equivalent, however, depending on the structure of the system (values of N and K), the one which offers a smaller number of variables can be used. In the following, the focus is on (107), which is a better choice for $N \leq K$.

6.3 Geometry of the Relaxation

In this section, similar to Chapter 5, the constraints defining $\mathbf{U}\mathbf{e}_K = \mathbf{e}_N$ are eliminated by providing a tractable representation of the linear manifold spanned by this constraint. Consider the minimization problem (64). By substituting (73), the objective function is

$$\begin{aligned} & \operatorname{trace}\left(\mathbf{S}\mathbf{U}^T\mathbf{Q}\mathbf{U} + 2\mathbf{C}\mathbf{U}\right) \\ = \; & \operatorname{trace}\left(\mathbf{S}(\mathbf{F} + \hat{\mathbf{U}}\mathbf{G})^T\mathbf{Q}(\mathbf{F} + \hat{\mathbf{U}}\mathbf{G}) + 2\mathbf{C}(\mathbf{F} + \hat{\mathbf{U}}\mathbf{G})\right) \\ = \; & \operatorname{trace}\Big(\mathbf{G}\mathbf{S}\mathbf{G}^T\hat{\mathbf{U}}^{\mathbf{T}}\mathbf{Q}\hat{\mathbf{U}} + \mathbf{G}\mathbf{S}\mathbf{F}^{\mathbf{T}}\mathbf{Q}\hat{\mathbf{U}} + \mathbf{Q}\mathbf{F}\mathbf{S}\mathbf{G}^{\mathbf{T}}\hat{\mathbf{U}}^{\mathbf{T}} \\ & \qquad + \mathbf{G}\mathbf{C}\hat{\mathbf{U}} + \mathbf{C}^{\mathbf{T}}\mathbf{G}^{\mathbf{T}}\hat{\mathbf{U}}^{\mathbf{T}} + \mathbf{2CF} + \mathbf{S}\mathbf{F}^{\mathbf{T}}\mathbf{Q}\mathbf{F}\Big) \\ = \; & \operatorname{trace}\left(\hat{\mathcal{L}}\mathbf{W}_{\hat{\mathbf{U}}} + \mathbf{2CF} + \mathbf{S}\mathbf{F}^{\mathbf{T}}\mathbf{Q}\mathbf{F}\right), \end{aligned} \tag{110}$$

where

$$\begin{aligned}
\hat{\mathcal{L}} &= \begin{bmatrix} \mathbf{0} & \mathbf{GSF}^T\mathbf{Q}+\mathbf{GC} & \mathbf{0} \\ \mathbf{QFSG}^T+\mathbf{C}^T\mathbf{G}^T & \mathbf{0} & \frac{1}{2}\mathbf{Q} \\ \mathbf{0} & \frac{1}{2}\mathbf{Q} & \mathbf{0} \end{bmatrix}, \\
\mathbf{W}_{\hat{\mathbf{U}}} &= \begin{bmatrix} \mathbf{I} & \hat{\mathbf{U}}^{\mathbf{T}} & \hat{\mathbf{V}}^{\mathbf{T}} \\ \hat{\mathbf{U}} & \hat{\mathbf{U}}\hat{\mathbf{U}}^{\mathbf{T}} & \hat{\mathbf{U}}\hat{\mathbf{V}}^{\mathbf{T}} \\ \hat{\mathbf{V}} & \hat{\mathbf{V}}\hat{\mathbf{U}}^{\mathbf{T}} & \hat{\mathbf{V}}\hat{\mathbf{V}}^{\mathbf{T}} \end{bmatrix}, \\
\hat{\mathbf{V}} &= \hat{\mathbf{U}}\mathbf{GSG}^{\mathbf{T}}.
\end{aligned} \tag{111}$$

Therefore, the optimization problem (64) can be written as

$$\begin{array}{ll}
\min & \text{trace}\left(\hat{\mathcal{L}}\mathbf{W}_{\hat{\mathbf{U}}}\right) \\
s.t. & \hat{\mathbf{U}} = \mathbf{U}(\mathbf{1}:\mathbf{N},\mathbf{1}:(\mathbf{K}-\mathbf{1})); \quad \mathbf{U}\in\mathcal{E}_{\mathbf{N}\times\mathbf{K}} \\
 & \hat{\mathbf{V}} = \hat{\mathbf{U}}\left(\mathbf{GSG}^{\mathbf{T}}\right)
\end{array} \tag{112}$$

Using a similar procedure, it can be shown that the optimization problem (112) is equivalent to the following reduced MLSDP problem:

$$\begin{array}{ll}
\min & \text{trace}\left(\hat{\mathcal{L}}\begin{bmatrix} \mathbf{I} & \hat{\mathbf{U}}^{\mathbf{T}} & \hat{\mathbf{V}}^{\mathbf{T}} \\ \hat{\mathbf{U}} & \hat{\mathbf{X}} & \hat{\mathbf{Y}} \\ \hat{\mathbf{V}} & \hat{\mathbf{Y}} & \hat{\mathbf{Z}} \end{bmatrix}\right) \\
s.t. & \hat{\mathbf{U}}\mathbf{e}_{\mathbf{K}-\mathbf{1}} \leq \mathbf{e}_{\mathbf{N}} \;\; ; \;\; \hat{\mathbf{U}} \geq \mathbf{0} \\
 & \hat{\mathbf{V}} = \hat{\mathbf{U}}\left(\mathbf{GSG}^{\mathbf{T}}\right) \\
 & \text{diag}(\hat{\mathbf{X}}) = \hat{\mathbf{U}}\mathbf{e}_{\mathbf{K}-\mathbf{1}} \\
 & \text{diag}(\hat{\mathbf{Y}}) = \hat{\mathbf{U}}\text{diag}\left(\mathbf{GSG}^{\mathbf{T}}\right) \\
 & \hat{\mathbf{Z}} = \left(\sum_{\mathbf{1}=\mathbf{i}}^{\mathbf{K}-\mathbf{1}}(\mathbf{s}_{\mathbf{i}}-\mathbf{s}_{\mathbf{K}})^2\right)\hat{\mathbf{Y}} \\
 & \begin{bmatrix} \mathbf{I} & \hat{\mathbf{U}}^{\mathbf{T}} & \hat{\mathbf{V}}^{\mathbf{T}} \\ \hat{\mathbf{U}} & \hat{\mathbf{X}} & \hat{\mathbf{Y}} \\ \hat{\mathbf{V}} & \hat{\mathbf{Y}} & \hat{\mathbf{Z}} \end{bmatrix} \succeq 0 \\
 & \hat{\mathbf{U}}, \hat{\mathbf{V}} \in \mathcal{M}_{\mathbf{N}\times(\mathbf{K}-\mathbf{1})}, \hat{\mathbf{X}}, \hat{\mathbf{Y}}, \hat{\mathbf{Z}} \in \mathcal{S}_{\mathbf{N}}
\end{array} \tag{113}$$

Note that this method can also be applied to the equivalent formulation in (109).

6.4 Solving the SDP Problem

The relaxed decoding problems can be solved using common Interior-Point Methods (IPMs), such as DSDP [15], SeDuMi [127], SDPA [70], etc.

In the MLSDP optimization problem (107), the rank-constrained matrix $\mathbf{W_u}$ is relaxed to the positive semi-definite matrix $\mathbf{W}$. Utilizing the rank-constrained property of the variable parameter, the relaxed problem (107) can be solved using a non-linear method, known as the *augmented Lagrangian algorithm*. This approach offers a significant complexity reduction for “large" problem sizes as compared to common IPM-based methods, while the performance degradation is negligible. For a comparison on the computational complexity and performance of IPMs and augmented Lagrangian algorithm, the reader is referred to [18]. In continue, a brief review of this method is presented. By using this method, an algorithm to solve the MLSDP relaxed problems is proposed. However, in the numerical experiments, the SDPA package is used for solving the MLSDP models.

6.4.1 The Augmented Lagrangian Algorithm

Recently, Burer and Monteiro [18] proposed a new method for solving a full-rank SDP problem

$$\begin{aligned} \min \quad & \mathrm{trace}(\mathcal{L}_{QC}\mathbf{W}) \\ s.t. \quad & \mathrm{trace}(\mathbf{A}_i\mathbf{W}) = b_i \ \text{ for } \ i = 1, \cdots, m \\ & \mathbf{W} \succeq 0. \end{aligned} \tag{114}$$

The distinguishing feature of the algorithm is a change of variables that replaces the symmetric, positive semi-definite variable $\mathbf{W} \in \mathcal{M}^n$ of (114) with a rectangular variable $\mathbf{R}$ according to the factorization $\mathbf{W} = \mathbf{R}\mathbf{R}^T$. In [8, 112], it is shown that, for an SDP problem (114) with m constraints, there exists an optimal solution with rank r such that $r(r+2) \leq m$. In [18], $\mathbf{R}$ is chosen in $\mathcal{M}_{n\times r}$. By using this formulation, the positive semi-definite constraint is removed since $\mathbf{W} = \mathbf{R}\mathbf{R}^T$ automatically enforces the constraint. Now the problem (114) can be reformulated as

$$\begin{aligned} \min_{\mathbf{R}\in\mathcal{M}_{n\times r}} \quad & \mathrm{trace}(\mathbf{R}^T\mathcal{L}_{QC}\mathbf{R}) \\ s.t. \quad & \mathrm{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R}) = b_i \ \text{ for } \ 0 \leq i \leq m. \end{aligned} \tag{115}$$

In order to solve this non-linear problem, an *augmented Lagrangian method* is introduced in [18]. The augmented Lagrangian is defined as

$$\begin{aligned} L(\mathbf{R},\boldsymbol{\lambda},\sigma) &= \operatorname{trace}(\mathbf{R}^T\mathcal{L}_{QC}\mathbf{R}) \\ &- \sum_{i=1}^{m}\lambda_i\left(\operatorname{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R})-b_i\right) \\ &+ \frac{\sigma}{2}\sum_{i=1}^{m}\left(\operatorname{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R})-b_i\right)^2, \end{aligned} \tag{116}$$

where $\mathbf{R} \in \mathcal{M}_{n\times r}$, $\boldsymbol{\lambda} \in \mathbb{R}^m$, and $\sigma \in \mathbb{R}^+$. The last term is the penalty term indicating the Euclidean norm of the infeasibility of $\mathbf{R}$ with respect to r.

To minimize the augmented Lagrangian (116), this function is alternatively minimized with respect to $\mathbf{R}$ and with respect to $\boldsymbol{\lambda}$ and σ. The optimization of (116) with respect to $\mathbf{R}$ can be achieved by a limited-memory BFGS algorithm, which uses the gradient of L:

$$\begin{aligned} \nabla_{\mathbf{R}}L(\mathbf{R},\boldsymbol{\lambda},\sigma) = &-2\sum_{i=1}^{m}\left(\lambda_i-\sigma\left(\operatorname{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R})-b_i\right)\right)\mathbf{A}_i\mathbf{R} \\ &+2\mathcal{L}_{QC}\mathbf{R}. \end{aligned} \tag{117}$$

This algorithm has the advantage of maintaining $O(nr)$ memory overhead, but also has the speed of a quasi-Newton method.

In order to optimize (116) with respect to $\boldsymbol{\lambda}$ and σ, the authors in [18] updates λ_i by $\lambda_i - \sigma\left(\operatorname{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R}) - b_i\right)$ and σ by a multiplicative factor. They have shown that this technique always empirically converges to the globally optimal solution, although the quadratic programming problem is non-convex. Moreover, they have shown that this method is significantly faster than existing SDP solvers.

6.4.2 Decoding Problem as a Quadratic Non-Linear Problem

Define $\mathbf{R} = [\ \mathbf{I}\quad \mathbf{U}^T\quad \mathbf{V}^T\ ]^T$. Therefore, the MLSDP optimization problem (107) can be reformulated as the problem in (115). Instead of applying IPM solvers, an augmented Lagrangian method can be applied directly to the mentioned decoding problem. However, despite the work in [18], the MLSDP problem have an “explicit rank constraint". In other words, the size of the matrix $\mathbf{R}$ is precisely determined by the rank constraint.

The other difference is that there are some inequality constraints in the optimization problem (107). The augmented Lagrangian method in [18] should be generalized to handle linear inequality constraints. The same technique as in [74] can be approached by solving directly for updates of the Lagrange multipliers for both equality and inequality constraints.

The updates of λ_i can be solved by treating a constraint $\operatorname{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R}) \geq b_i$ as the constraint $\operatorname{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R}) = b_i + s_i$ and $s_i \geq 0$. For inequality constraint, this results in

$$\lambda_i = \max(\lambda_i - \sigma(\operatorname{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R}) - b_i), 0). \tag{118}$$

This update satisfies $\lambda_i \geq 0$ for Lagrange multipliers for inequality constraints. Computing the Lagrangian and the gradient of the Lagrangian are also straightforward for inequality constraints [74]. The second and third terms of the Lagrangian change: for each constraint i, the Lagrangian term $-\lambda_i(\text{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R}) - b_i) + \frac{\sigma}{2}(\text{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R}) - b_i)^2$ is unchanged if $\sigma(\text{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R}) - b_i) \leq \lambda_i$. Otherwise, this term becomes $-\lambda_i^2/2\sigma$. Similarly, in the computation of the gradient of Lagrangian, only the term $2(\lambda_i - \sigma(\text{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R}) - b_i))\mathbf{A}_i\mathbf{R}$ is contributed if $\sigma(\text{trace}(\mathbf{R}^T\mathbf{A}_i\mathbf{R}) - b_i) \leq \lambda_i$. Otherwise, nothing is added to the gradient.

6.5 Integer Solution - Matrix Nearness Problem

Solving the relaxed decoding problems results in the solution $\tilde{\mathbf{U}}$. In general, this matrix is not in $\mathcal{E}_{N\times K}$. The condition $\mathbf{U}\mathbf{e}_K = \mathbf{e}_N$ is satisfied. However, the elements are between 0 and 1. This matrix has to be converted to a 0-1 matrix by finding a matrix in $\mathcal{E}_{N\times K}$ which is satisfies a condition of nearness to this matrix. Matrix approximation problems typically measure the distance between matrices with a norm. The Frobenius and spectral norms are common choices as they are analytically tractable.

Therefore, in order to find the nearest solution in $\mathcal{E}_{N\times K}$ to the solution of the relaxed problem $\tilde{\mathbf{U}}$, it is proposed to solve

$$\min_{\mathbf{U}\in\,\mathcal{E}_{N\times K}} ||\mathbf{U} - \tilde{\mathbf{U}}||_{\mathbb{F}}^{\mathbf{2}}, \tag{119}$$

where $||\mathbf{A}||_{\mathbb{F}}^2$ is the Frobenius norm of the matrix $\mathbf{A}$ which is defined as $||\mathbf{A}||_{\mathbb{F}}^2 = \text{trace}(\mathbf{A}\mathbf{A}^T)$, and

$$\begin{aligned} ||\mathbf{U} - \tilde{\mathbf{U}}||_{\mathbb{F}}^{\mathbf{2}} &= \text{trace}\left((\mathbf{U} - \tilde{\mathbf{U}})(\mathbf{U} - \tilde{\mathbf{U}})^{\mathbf{T}}\right) \\ &= N - 2\text{trace}(\tilde{\mathbf{U}}\mathbf{U}^{\mathbf{T}}) + \text{trace}(\tilde{\mathbf{U}}\tilde{\mathbf{U}}^{\mathbf{T}}). \end{aligned} \tag{120}$$

The last equality is due to the fact that for any $\mathbf{U} \in \mathcal{E}_{N\times K}$, $\text{diag}(\mathbf{U}\mathbf{U}^T) = \mathbf{e}_N$, see (107). Therefore, after removing the constants, finding the integer solution is the solution of

$$\max_{\mathbf{U}\in\,\mathcal{E}_{N\times K}} \text{trace}(\tilde{\mathbf{U}}\mathbf{U}^{\mathbf{T}}) \tag{121}$$

Consider the maximization problem

$$\begin{aligned} \max \quad & \text{trace}(\tilde{\mathbf{U}}\mathbf{U}^{\mathbf{T}}) \\ s.t. \quad & \mathbf{U}\mathbf{e}_K = \mathbf{e}_N \\ & 0 \leq \mathbf{U} \leq 1, \end{aligned} \tag{122}$$

where $\leq$ in the last constraint is element-wise. This problem is a linear programming problem with linear constraints. Therefore, the optimum solution is a corner point meaning that the constraints are satisfied with equality at the optimum point. In other words, at the optimum point, $\mathbf{U} \in \mathcal{E}_{N\times K}$. Therefore, to find the solution for (121), simply, the linear problem (122) can be solved, which is strongly polynomial time. To improve this result, the randomization algorithms, introduced in Chapter 5, can be further applied.

6.6 Simulation Results

The proposed MLSDP relaxation model (113) in a 4×4 MIMO system employing 16-QAM is simulated. Figure 23 shows the performance of the proposed method vs. the performance of the VLSDP Model III in Chapter 5 and the previous known methods in WES [143], SiLu [121], and YZZX [150]. As it can be seen, the proposed method outperforms all other convex sub-optimal methods.

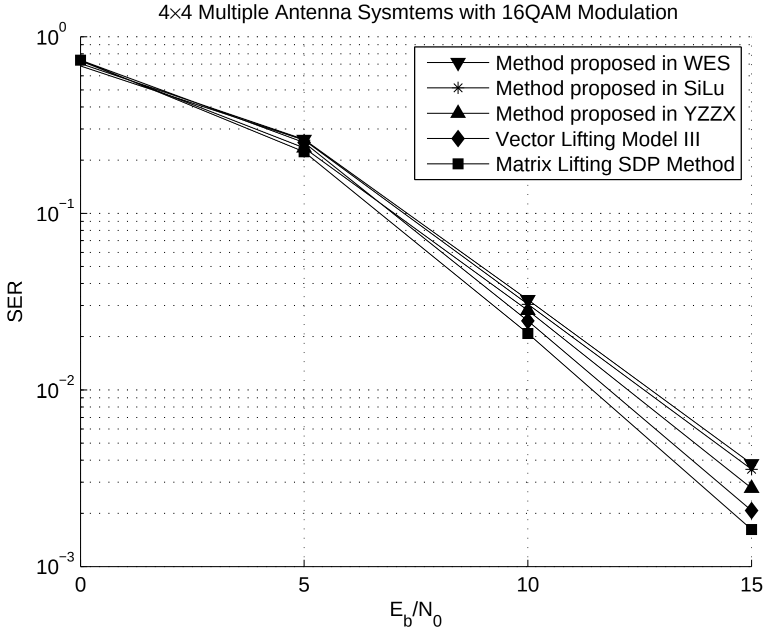

Figure 23: Performance of the proposed MLSDP method in a 4×4 MIMO system employing 16-QAM compared to the previous known methods in WES [143], SiLu [121], and YZZX [150]

The worst case complexity of the proposed method solved by IPMs is a polynomial function of the number of antennas (similar to the analysis in Chapter 5). In the optimization problem of (107), where $N \leq K$, the dimension of the matrix variable $\mathbf{W}$ is $m = O(K)$ and the number of constraints is $p = O(K^2)$. Similar to Chapter 5, it can be easily seen that a solution to (113) can be found in at most $O(K^{5.5})$ arithmetic operations (utilizing the sparsity of the rank-one constraint matrices), where the computational complexity of the VLSDP model in Chapter 5, [143], [150], [121] are $O(N^{5.5}K^{5.5})$, $O(N^{6.5}K^{6.5})$, $O(K^2N^{4.5} + K^3N^{3.5})$, and $O(N^{3.5})$ respectively.

Note that for the equivalent optimization problem (109), where $K \leq N$, the computational complexity is at most $\mathcal{O}(N^{5.5})$. It must be emphasized that depending on values of N and K, the optimization problem (107) or (109) which results in less computational complexity can be implemented.

Note that many of the constraints have very simple structures. This property can be used to develop an interior-point optimization algorithm fully exploiting the constraint structures of the problem, thereby getting complexity order better than that of using a general purpose solver such as SeDuMi or SDPA. Moreover, the complexity of the proposed method can be further reduced by implementing the augmented Lagrangian method for large problem sizes.

6.7 Conclusion

A sub-optimum method for decoding in MIMO systems based on MLSDP is introduced. The advantage of this structure lies in its efficient implementations, whereas there is fewer redundant constraints to add under this relaxation. Simulation results show that the proposed algorithm outperforms all known convex quasi-ML decoding methods in the literature.

CHAPTER 7

EPILOGUE

7.1 Conclusion

It is known that lattice codes can achieve the capacity in an AWGN channel [82, 83]. Some general lattice code were designed based on fixed dimensional classical lattices [28] or based on algebraic error correcting codes [46]. However, the lattice labeling and lattice decoding problems associated with these lattice codes make them almost impractical to be implemented for common communication applications. In this thesis, some lattice codes for specific applications are designed such that some desirable geometrical properties are provided.

In OFDM systems, a lattice code with low PAPR is designed. The shaping region for the lattice code is a cube whose boundary is along the basis defined by a Hadamard matrix. The lattice labeling and decoding algorithms are based on SNF decomposition of an integer matrix. Due to the recursive structure of the Hadamard matrix these algorithms can be implemented very efficiently. In multiple antenna broadcast systems, a lattice code with low average transmit energy is introduced. This design is more complex compared to that in the OFDM system, since the labeling algorithm should be such that the users can decode their data independent of each other. By defining some redundancy, an SLM technique is introduced which results in a lattice code with low average transmit energy. The decoding algorithm is a simple rounding algorithm.

In MIMO systems usually simple lattice codes are used. However, due to the channel and noise effects, the decoding problem in these systems is an NP-hard problem. In the second part of this thesis, several sub-optimal decoding algorithms are introduced. They introduce a wealth of tradeoff between complexity and performance. They offer a better performance compared to other sub-optimal method reported in the literature with the same order of computational complexity.

In summary, in this thesis, lattice codes are designed in different communication applications such that they provide different geometrical properties and they also increase the data rate of wireless communication networks. The focus of these designs is in fast and simple implementations for the corresponding lattice labeling and decoding algorithms.

7.2 Future Work

Over the last two decades, there have been significant advances in producing powerful optimization models and efficient algorithmic tools as well as software. Recently these advances have begun to significantly impact various applied science and engineering fields, where efficient optimization is essential. The goal of this proposal is to apply SDP tools to solve several core problems in signal processing and communications. The previous work-horse optimization algorithms in communication systems suffer from slow convergence, difficulty in finding global optima, and sensitivity to the algorithm initialization and stepsize selection, especially when applied to many naturally ill-conditioned or non-convex communication problems. Examples of such problems include decoding in multiple antenna systems, decoding of binary block codes, and designing codes in delay-constrained networks. One powerful way to avoid these problems is to derive an SDP relaxation of the original non-convex formulation. Then, we can be guaranteed of finding the globally optimal design "efficiently" without the usual headaches of stepsize selection, algorithm initialization and local minima. Therefore, the challenge is to find SDP relaxations which efficiently model different problems, and then, to find additional constraints that improve the system performance.

7.2.1 Multiple Antenna Decoding

In this thesis, several sub-optimum SDP relaxation models are introduced. However, the bounds achieved by the relaxation models for high dimension problems are always loose. To achieve tight bounds for high dimensional problems, the branch and bound methods and SDP relaxation method can be combined to develop better SDP relaxations. Compared to the global optimization SDP methods, the proposed method achieves a better bounds. Moreover, the SDP branch and bound decoding algorithm can be implemented more efficiently compared to previous decoding methods based on conventional branch and bound methods.

7.2.2 Decoding Binary Block Codes by SDP

Binary Block codes have been widely used in different communication applications. These codes are sequences of zero and one which follow a certain algebraic pattern. Sub-optimum decoding of these codes can be implemented by linear programming. However, the corresponding decoding problem can be reformulated as a quadratic binary optimization problem. Then, SDP algorithms can be applied to find a better solution for the decoding problem. The advantage of the proposed formulation is that potentially the symmetry properties of the SDP problem can be exploited. Therefore, matrix algebra can be used to reduce the size of this SDP problem. Using this approach simplifies the SDP decoding while it achieves a better performance.

7.2.3 Finding Optimal Precoding Scheme for Broadcast Systems

In order to provide a simple and independent decoding for users in a MIMO broadcast system, a precoding has been applied at the transmitter side. Recently, some researchers have focused on finding the optimal precoder for different setups. Since, the general optimization problem is hard, several simplifications have been considered in defining the problem. These simplifications will significantly limit a broadcast system. In continue, the design of an optimal "non-linear" precoder can be considered. By redefining the constraints to limit the signal to interference ratio, an SDP relaxation problem will result, which can be solved efficiently.

7.2.4 Code Design for Delay Limited Networks based on SDP

In transmitting data packets in a communication network, some packets may be lost. In order to deal with these dropped packets, the receiver send an acknowledgement when it receives a packet. If the transmitter does not receive any thing back, it will send the packet again. Consider a network that has a hard upper bound on acceptable delays for transmitting packets. The problem of encoding transmitting packets can be considered such that increasingly more accurate approximation of the original packet is obtained. Having a set of labels, the problem of deciding which codeword to assign to each label can be relaxed to an SDP optimization problem. By efficient solution of this problem, different codes with desired properties can be obtained.

APPENDIX A

PROOFS

This appendix is devoted to the detailed proofs for the lemmas and theorems in the thesis.

A.1 Part I - PAPR Reduction Proofs

A.1.1 Theorem 2

Based on (15), $\det(\mathbf{H}_{2^n}) = \det(\mathbf{D}_{2^n})$, because the matrices $\mathbf{U}_{2^n}$ and $\mathbf{V}_{2^n}$ are unimodular and their determinants are one. To prove this theorem, the induction can be used. For a 2×2 Hadamard matrix,

$$\det(\mathbf{D}_2) = \det\left(\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right) = 2 = 2^{1\times 2^{1-1}}. \tag{123}$$

It is assumed that the claim is valid for a $2^k \times 2^k$ Hadamard matrix. Based on (15), for a $2^{k+1} \times 2^{k+1}$ Hadamard matrix,

$$\mathbf{D}_{2^{k+1}} = \begin{bmatrix} \mathbf{D}_{2^k} & 0 \\ 0 & 2\mathbf{D}_{2^k} \end{bmatrix}$$

$$\begin{aligned} \Rightarrow \det(\mathbf{D}_{2^{k+1}}) &= \det(\mathbf{D}_{2^k}) \times 2^{2^k} \times \det(\mathbf{D}_{2^k}) \\ &= 2^{2^k} \times (\det(\mathbf{D}_{2^k}))^2 \\ &= 2^{2^k} \times \left(2^{k2^{k-1}}\right)^2 = 2^{(k+1)2^{(k+1)-1}}. \end{aligned} \tag{124}$$

A.2 Part II - Precoding Proofs

A.2.1 Theorem 3

The matrix $\mathbf{Q}$ is positive semi-definite, i.e., $\boldsymbol{\mu}^T \mathbf{Q} \boldsymbol{\mu} \geq \mathbf{0}$ for any vector $\boldsymbol{\mu}$. Therefore, in order to minimize $E\{\gamma\}$ in (35), $E\{\mathbf{u}\} = \boldsymbol{\mu} = \mathbf{0}$.

In designing input constellations for communication systems, the objective is to reduce the average transmit energy, while keeping a fixed entropy for the input constellation. On the other hand, this is equivalent to keeping a fixed average energy while the entropy is maximized. Therefore, instead of minimizing $E\{\gamma\} = tr(\mathbf{Q}\boldsymbol{\Sigma})$ given $\mathcal{H}(\mathbf{u}) =$ fixed, the equivalent problem of maximization of $\mathcal{H}(\mathbf{u})$ is considered, given a fixed average transmit energy, $E\{\gamma\} = tr(\mathbf{Q}\boldsymbol{\Sigma}) = K$, where K is a known constant.

In [39], it has been shown that for a random vector $\mathbf{u}$ with zero mean and covariance matrix $\boldsymbol{\Sigma}$,

$$\mathcal{H}(\mathbf{u}) \leq \frac{1}{2}\log(2\pi e)^{M}|\boldsymbol{\Sigma}|$$

with equality iff $\mathbf{u}$ is a multivariate Gaussian vector. Therefore, a Gaussian random vector with zero mean and covariance $\boldsymbol{\Sigma}$ is desired such that

$$\begin{aligned} \max \quad & \log|\boldsymbol{\Sigma}| \\ s.t. \quad & tr\,(\mathbf{Q}\boldsymbol{\Sigma}) = K \\ & \boldsymbol{\Sigma} \succ 0 \end{aligned} \tag{125}$$

By considering $\mathbf{Q} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathbf{T}}$ and $\boldsymbol{\Sigma}' = \mathbf{U}^{T}\boldsymbol{\Sigma}\mathbf{U}$, it is easy to show that the optimization problem (125) is equivalent to

$$\begin{aligned} \max \quad & \log|\boldsymbol{\Sigma}'| \\ s.t. \quad & tr(\boldsymbol{\Lambda}\boldsymbol{\Sigma}') = K \\ & \boldsymbol{\Sigma}' \succ 0. \end{aligned} \tag{126}$$

The Hadamard inequality states that, for a Hermitian positive definite matrix $\boldsymbol{\Sigma}' = [\sigma'_{ij}]$, $|\boldsymbol{\Sigma}'| \leq \Pi_i \sigma'_{ii}$, with equality iff $\boldsymbol{\Sigma}'$ is a diagonal matrix. Therefore, in order to maximize the objective in (126), it is assumed that the covariance matrix $\boldsymbol{\Sigma}'$ is a diagonal matrix with diagonal elements σ_i^2. Hence, the optimization problem (126) can be written as

$$\begin{aligned} \max \quad & \sum_{i=1}^{M} \log\sigma_i^2 \\ s.t. \quad & \sum_{i=1}^{M}(\lambda_i\sigma_i^2) = K \end{aligned} \tag{127}$$

The optimum solution for (127) is $\sigma_i^2 = \frac{K}{M\lambda_i}$. Therefore, for independent Gaussian random variables with variance σ_i^2 in (127),

$$\begin{aligned} \mathcal{H}(\mathbf{u}) &= \frac{1}{2}\log(2\pi e)^{M}|\boldsymbol{\Sigma}| \\ &= \frac{1}{2}\sum_{i=1}^{M}\log\left(2\pi e\frac{K}{M\lambda_i}\right) \\ &= \log(\mathbb{V}) = M\mathcal{H}, \end{aligned} \tag{128}$$

or equivalently,

$$\log 2\pi e\frac{K}{M\sqrt[M]{\Pi\lambda_i}} = 2\mathcal{H}. \tag{129}$$

Considering (177) and (129),

$$K = M \sqrt[M]{\Pi\lambda_i}\sigma^2,$$

which results in

$$\sigma_i^2 = \frac{\sqrt[M]{\Pi\lambda_i}}{\lambda_i}\sigma^2 \quad i = 1, \cdots, M. \tag{130}$$

Therefore, the optimum solution of (125) would be

$$\begin{aligned} \boldsymbol{\Sigma} &= \sqrt[M]{\Pi\lambda_i}\sigma^2\mathbf{U}\begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \vdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_M} \end{bmatrix}\mathbf{U}^T \\ &= \sqrt[M]{\Pi\lambda_i}\sigma^2\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T = \sqrt[M]{\Pi\lambda_i}\sigma^2\mathbf{H}\mathbf{H}^T \end{aligned} \tag{131}$$

A.2.2 Theorem 4

The following lemmas will help in proving the main results.

Lemma 12

$$E\left\{N^{\frac{r}{D}}\gamma_{r,N}^F\right\} = \int_{\mathbb{R}_+}\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N dv \tag{132}$$

Proof:

$$E\left\{N^{\frac{r}{D}}\gamma_{r,N}^F\right\} = E\left\{N^{\frac{r}{D}} \min_{1\le i\le N} \|\mathbf{s}^{(i)}\|^r\right\} \tag{133}$$

But, if V is a nonnegative random variable, then

$$E\{V\} = \int_{\mathbb{R}_+} \mathbb{P}\{V \ge v\}dv, \tag{134}$$

since

$$\begin{aligned} \int_{\mathbb{R}_+} \mathbb{P}\{V \ge v\}dv &= \int_{\mathbb{R}_+} E\left\{\mathbf{1}_{\{V\ge v\}}\right\}dv \\ &= E\left\{\int_{\mathbb{R}_+} \mathbf{1}_{\{V\ge v\}}dv\right\} \\ &= E\{V\} \end{aligned}$$

Combining (134) and (133) yields

$$\begin{aligned} E\left\{N^{\frac{r}{D}}\gamma_{r,N}^F\right\} &= \int_{\mathbb{R}_+} \mathbb{P}\left\{N^{\frac{r}{D}} \min_{1\le i\le N} \|\mathbf{s}^{(i)}\|^r \ge v\right\}dv \\ &= \int_{\mathbb{R}_+} \mathbb{P}\left\{\bigcap_{i=1}^{N}\{N^{\frac{r}{D}}\|\mathbf{s}^{(i)}\|^r \ge v\}\right\}dv \end{aligned} \tag{135}$$

Since $\mathbf{s}^{(i)}$'s are i.i.d.,

$$\begin{aligned} E\left\{N^{\frac{r}{D}}\gamma_{r,N}^{F}\right\} &= \int_{\mathbb{R}_+}\left(\mathbb{P}\left\{N^{\frac{r}{D}}\|\mathbf{s}^{(1)}\|^r \geq v\right\}\right)^N dv \\ &= \int_{\mathbb{R}_+}\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N dv, \end{aligned} \tag{136}$$

which completes the proof. ■

Lemma 13 *For any $\rho > 0$, define $\mathcal{A}_\rho = \{v \in \mathbb{R}_+; v^{1/r}/N^{1/D} \leq \rho\}$ and set*

$$g_\rho := \inf_{\delta\in(0,\rho]} \frac{F\left(\mathcal{B}_D(0,\delta)\right)}{Vol\left(\mathcal{B}_D(0,\delta)\right)}.$$

Then,

$$\begin{aligned} &\int_{\mathcal{A}_\rho}\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N dv \leq \\ &\int_{\mathcal{A}_\rho}\exp\left(-B_D g_\rho v^{D/r}\right) dv. \end{aligned} \tag{137}$$

Proof: One has

$$\begin{aligned} &\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N \\ = \ &\exp\left(N\ln\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)\right) \\ \leq \ &\exp\left(-NF\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right) \\ \leq \ &\exp\left(-B_D\frac{F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)}{\mathrm{Vol}\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)}v^{D/r}\right), \end{aligned} \tag{138}$$

where $B_1 = 2$ and $B_D = \mathrm{Vol}\left(\mathcal{B}_D(0,1)\right) = \pi^{D/2}/\Gamma(1 + D/2)$ for $D = 2, \cdots$. By the definition of the set $\mathcal{A}$,

$$\begin{aligned} &\int_{\mathcal{A}_\rho}\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N dv \\ \leq \ &\int_{\mathcal{A}_\rho}\exp\left(-B_D\inf_{\delta\in(0,\rho]}\frac{F\left(\mathcal{B}_D(0,\delta)\right)}{\mathrm{Vol}\left(\mathcal{B}_D(0,\delta)\right)}v^{D/r}\right) dv \\ = \ &\int_{\mathcal{A}_\rho}\exp\left(-B_D g_\rho v^{D/r}\right) dv. \end{aligned} \tag{139}$$

■

Proof for Theorem 4: According to Lemma 12,

$$E\left\{N^{\frac{r}{D}}\gamma^F_{r,N}\right\} = \int_{\mathbb{R}_+}\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N dv \tag{140}$$

For any $\rho > 0$ define $\bar{\mathcal{A}}_\rho$ as the complement region of $\mathcal{A}$. Therefore,

$$\begin{aligned} E\left\{N^{\frac{r}{D}}\gamma^F_{r,N}\right\} =& \int_{\mathcal{A}_\rho}\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N dv \\ &+ \int_{\bar{\mathcal{A}}_\rho}\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N dv \end{aligned} \tag{141}$$

Based on Lemma 13,

$$\begin{aligned} &\int_{\mathcal{A}_\rho}\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N dv \\ \leq & \int_{\mathcal{A}_\rho} \exp\left(-B_D g_\rho v^{D/r}\right) dv. \end{aligned} \tag{142}$$

Note that the integral in (142) is limited (The proof is easy and it is similar to the approach in [25]). As $N \longrightarrow \infty$, $\mathcal{A}_\rho \longrightarrow \mathbb{R}_+$. Therefore,

$$\begin{aligned} &\int_{\mathcal{A}_\rho}\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N dv \longrightarrow \\ &\int_{\mathbb{R}_+} \exp\left(-B_D g_\rho v^{D/r}\right) dv, \\ \text{and} \\ &\int_{\bar{\mathcal{A}}_\rho}\left(1 - F\left(\mathcal{B}_D(0, \frac{v^{\frac{1}{r}}}{N^{\frac{1}{D}}})\right)\right)^N dv \longrightarrow \\ &0. \end{aligned} \tag{143}$$

In [64], it is shown that

$$\int_{\mathbb{R}_+} \exp\left(-B_D g_\rho v^{D/r}\right) dv = B_D^{-\frac{r}{D}}\Gamma(1 + \frac{r}{D})g_\rho^{-\frac{r}{D}} \tag{144}$$

Therefore,

$$\lim_{N\to\infty} E\left\{N^{\frac{r}{D}}\gamma^F_{r,N}\right\} = B_D^{-\frac{r}{D}}\Gamma(1 + \frac{r}{D})g_\rho^{-\frac{r}{D}} \tag{145}$$

A.2.3 Theorem 6

According to (49), when $N \longrightarrow \infty$,

$$E_{Broadcast} = \frac{1}{L} E\left\{\gamma_{2,N}^{U(\bar{\mathcal{R}}')}\right\} = \frac{1}{L} B_D^{-\frac{2}{D}} \Gamma(1 + \frac{2}{D}) N^{-\frac{2}{D}} \mathrm{Vol}(\bar{\mathcal{R}}')^{\frac{2}{D}}, \qquad (146)$$

where $\mathrm{Vol}(\bar{\mathcal{R}}') = (\sqrt{\Pi\lambda_j})^L \mathrm{Vol}(\bar{\mathcal{R}})$, $\bar{\mathcal{R}} = \bar{\mathcal{R}}_1 \times \bar{\mathcal{R}}_2 \times \cdots \times \bar{\mathcal{R}}_K$, and $\mathrm{Vol}(\bar{\mathcal{R}}_i) = N_i \mathrm{Vol}(\mathcal{B}_{D_i}(0, \sqrt{D_i}\sigma_i))$. Therefore,

$$\mathrm{Vol}(\bar{\mathcal{R}}) = N_1 B_{D_1} D_1^{\frac{D_1}{2}} \sigma_1^{D_1} \cdots N_K B_{D_K} D_K^{\frac{D_K}{2}} \sigma_K^{D_K}, \qquad (147)$$

where $B_{D_i} = \pi^{D_i/2}/\Gamma(1 + D_i/2)$ is the volume of a D_i-dimensional sphere of radius one ($D = LM = \sum D_i$ and $N = \prod N_i$).

$$\begin{aligned} \mathrm{Vol}(\bar{\mathcal{R}}) &= NB_{D_1} \cdots B_{D_K} \prod_{i=1}^{K} D_i^{\frac{D_i}{2}} \prod_{i=1}^{K} \sigma_i^{D_i} \\ &= NB_D \left(\frac{B_{D_1} \cdots B_{D_K}}{B_D}\right) \prod_{i=1}^{K} (2Ln_i)^{\frac{D_i}{2}} \prod_{i=1}^{K} \sigma_i^{D_i} \\ &= N2^{\frac{D}{2}} L^{\frac{D}{2}} B_D \left(\frac{B_{D_1} \cdots B_{D_K}}{B_D} \prod_{i=1}^{K} n_i^{\frac{D_i}{2}}\right) \prod_{i=1}^{K} \sigma_i^{D_i} \qquad (148) \end{aligned}$$

The term in parenthesis can be simplified

$$\begin{aligned} \frac{B_{D_1} \cdots B_{D_K}}{B_D} \prod_{i=1}^{K} n_i^{\frac{D_i}{2}} &= \frac{\Gamma(1 + D/2)}{\prod \Gamma(1 + D_i/2)} \prod_{i=1}^{K} n_i^{\frac{D_i}{2}} \\ &= \frac{(LM/2)!}{\prod (Ln_i)!} \prod_{i=1}^{K} n_i^{\frac{D_i}{2}} \\ &= \frac{\sqrt{\pi LM} \frac{(LM/2)^{LM/2}}{e^{LM/2}}}{\prod_{i=1}^{K} \sqrt{2\pi Ln_i} \frac{(Ln_i)^{Ln_i}}{e^{Ln_i}}} \prod_{i=1}^{K} n_i^{Ln_i} \\ &= \frac{\sqrt{\pi LM}(M/2)^{LM/2}}{\prod_{i=1}^{K} \sqrt{2\pi Ln_i}}, \qquad (149) \end{aligned}$$

where the last equality is based on applying stirling's formula (since $L \longrightarrow \infty$). On the other hand[1],

$$\begin{aligned} & \lim_{L\longrightarrow\infty}\left(\frac{B_{D_1}\cdots B_{D_K}}{B_D}\prod_{i=1}^{K} n_i^{\frac{D_i}{2}}\right)^{\frac{2}{LM}} \\ = & \lim_{L\longrightarrow\infty}\left(\frac{\sqrt{\pi LM}}{\prod_{i=1}^{K}\sqrt{2\pi L n_i}}\right)^{\frac{2}{LM}}\frac{M}{2} \\ = & \frac{M}{2}. \end{aligned} \tag{150}$$

Therefore, for large enough L, the volume of region $\bar{\mathcal{R}}$ can be written as

$$\begin{aligned} \mathrm{Vol}(\bar{\mathcal{R}}) &= NB_D D^{\frac{D}{2}}\prod_{i=1}^{K}\sigma_i^{D_i} \\ &= N\mathrm{Vol}\left(\mathcal{B}_D(0,\ \sqrt{D}\sqrt[D]{\prod_{i=1}^{K}\sigma_i^{D_i})}\right) \end{aligned} \tag{151}$$

It is easy to show that, in order to provide entropy $\mathcal{H} = \frac{1}{M}\log(\mathbb{V})$ per each dimension,

$$\sigma^M = \sigma_1^{2n_i}\cdots\sigma_K^{2n_K},$$

where σ^2 is the variance of a Gaussian random variable with entropy $\mathcal{H}$ and

$$\mathrm{Vol}(\bar{\mathcal{R}}) = N\mathrm{Vol}\left(\mathcal{B}_D(0,\ \sqrt{D}\sigma)\right).$$

Therefore, the average transmit energy in (146) can be written as

$$\begin{aligned} E_{Broadcast} &= \frac{1}{L}B_D^{-\frac{2}{D}}\Gamma(1+\frac{2}{D})N^{-\frac{2}{D}}\mathrm{Vol}(\bar{\mathcal{R}}')^{\frac{2}{D}} \\ &= M\sqrt[M]{\Pi\lambda_j}\sigma^2, \end{aligned} \tag{152}$$

where $\sigma^M = \sigma_1^{2n_i}\cdots\sigma_K^{2n_K}$.

A.3 Part III - Lattice Code Decoding Proofs

A.3.1 Lemma 7

Let $\mathbf{U} \in \mathcal{M}_{N\times K}$ and $\mathbf{U}\mathbf{e}_K = \mathbf{e}_N$. Since $\mathbf{G}$ is a $(K-1)\times K$ matrix containing a basis of the orthogonal complement of the vector of all ones, i.e., $\mathbf{G}\mathbf{e}_K = \mathbf{0}_{K-1}$, and

$$\mathbf{F}\mathbf{e}_K = \mathbf{e}_N, \tag{153}$$

[1] $\lim_{L\longrightarrow\infty}\frac{1}{LM}\left[\ln\pi LM - \sum_{i=1}^{K}\ln 2\pi L n_i\right] = 0$

therefore,

$$\mathbf{U} = \mathbf{F} + \hat{\mathbf{U}}\mathbf{G}, \tag{154}$$

where $\hat{\mathbf{U}} \in \mathcal{M}_{N\times(K-1)}$. From

$$\begin{aligned}\mathbf{F} &= \frac{1}{K}(\mathbf{E}_{N\times K} - \mathbf{E}_{N\times(K-1)}\mathbf{G}) \\ &= \left[\, \mathbf{0}_{N\times(K-1)} \middle| \mathbf{e}_N \,\right],\end{aligned} \tag{155}$$

and

$$\hat{\mathbf{U}}\mathbf{G} = \left[\ \hat{\mathbf{U}} \ \middle| \ -\hat{\mathbf{U}}\mathbf{e}_{K-1}\ \right] \tag{156}$$

it follows that $\hat{\mathbf{U}} = \mathbf{U}(1 : N, 1 : (K - 1))$.

A.3.2 Theorem 8

Let $\mathbf{Y} \in \mathcal{F}$ be an extreme point of $\mathcal{F}$, i.e.

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{u}} = \left[\begin{array}{c|c} 1 & \mathbf{x}^T \\ \hline \mathbf{x} & \mathbf{x}\mathbf{x}^T \end{array}\right], \tag{157}$$

for some $\mathbf{x} = \text{vec}(\mathbf{X}^T)$, $\mathbf{X} \in \mathcal{E}_{N\times K}$. From Lemma 7, it follows that every matrix $\mathbf{X} \in \mathcal{E}_{N\times K}$ is of the form $\mathbf{X} = \mathbf{F} + \tilde{\mathbf{X}}\mathbf{G}$ where $\tilde{\mathbf{X}} = \mathbf{X}(1 : K - 1, 1 : N)$. From the properties of the Kronecker product (see [55]), it is known that

$$\text{vec}(\mathbf{ACB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{C}).$$

Therefore, it follows that

$$\begin{aligned}\mathbf{x} = \text{vec}(\mathbf{X}^T) &= \frac{1}{K}(\mathbf{e}_{KN} - (\mathbf{I}_N \otimes \mathbf{G}^T)\mathbf{e}_{(K-1)N}) \\ &\quad + (\mathbf{I}_N \otimes \mathbf{G}^T)\tilde{\mathbf{x}},\end{aligned} \tag{158}$$

where $\tilde{\mathbf{x}} = \text{vec}(\tilde{\mathbf{X}}^T)$. Let $\mathbf{p}^T := \left[\ 1 \quad \tilde{\mathbf{x}}^T\ \right]$,

$$\mathbf{b}_{\mathbf{V}} = \frac{1}{K}(\mathbf{e}_{KN} - (\mathbf{I}_N \otimes \mathbf{G}^T)\mathbf{e}_{(K-1)N}),$$

and

$$\mathbf{W} := \left[\ \mathbf{b}_{\mathbf{V}} \ \middle| \ \mathbf{I}_N \otimes \mathbf{G}^T\ \right]. \tag{159}$$

Therefore, $\mathbf{x} = \mathbf{W}\mathbf{p}$, and

$$\mathbf{Y} = \left[\begin{array}{c|c} 1 & \mathbf{p}^T\mathbf{W}^T \\ \hline \mathbf{W}\mathbf{p} & \mathbf{W}\mathbf{p}\mathbf{p}^T\mathbf{W}^T \end{array}\right] = \hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T, \tag{160}$$

where $\mathbf{R} := \mathbf{p}\mathbf{p}^T$, i.e.

$$\mathbf{R} = \left[\begin{array}{c|c} 1 & \tilde{\mathbf{x}}^T \\ \hline \tilde{\mathbf{x}} & \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \end{array}\right] \succeq \mathbf{0}. \tag{161}$$

Since $\tilde{\mathbf{x}}$ is a binary vector, it follows that $r_{ij} \in \{0, 1\},\ \forall i, j \in \{0, \ldots, N(K-1)\}$, and $\text{diag}(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T) = \tilde{\mathbf{x}}$. The proof follows analogously for any convex combination of the extreme points from $\mathcal{F}$.

A.3.3 Theorem 9

Fix $\mathbf{u} \in \mathcal{U}$ and let

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{u}} = \begin{bmatrix} 1 \\ \mathbf{u} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{u}^T \end{bmatrix}.$$

Considering the constraint on this vector in (67), the vector $\mathbf{u}$ is divided into N *sub-vectors* of length K. In each sub-vector all the elements are zero except one of the elements which is one. Therefore, there are K^N different binary vectors in the set $\mathcal{U}$. Consider the entries of the 0-th row of $\mathbf{Y}$. Note that $y_{0,j} = 1$ means that the j-th element of $\mathbf{u}$ is 1. In addition, there is only one element equal to 1 in each sub-vector. Therefore, there are K^{N-1} such vectors, and the components of the 0-th row of $\hat{\mathbf{Y}}$ are given by

$$\hat{y}_{0,j} = \frac{1}{K^N} K^{N-1} = \frac{1}{K}.$$

Now consider the entries of $\mathbf{Y}$ in the other rows, $y_{i,j}$.

1. If $i = j$, then, $y_{i,j} = 1$ means that the i-th element of the vector $\mathbf{u}$ is 1 and there are K^{N-1} such vectors; therefore, the diagonal elements are

$$\hat{y}_{i,i} = \frac{1}{K^N} K^{N-1} = \frac{1}{K}.$$

2. If $i = K(p-1) + q,\ \ j = K(p-1) + r, q \neq r, q, r \in \{1, \cdots, K\}, p \in \{1, \cdots, N\}$, i.e. the element is an off-diagonal element in a diagonal block, then, $y_{i,j} = 1$ means that the i-th and the j-th elements of $\mathbf{u}$ in a sub-vector should be 1 and this is not possible. Therefore, this element is always zero.

3. Otherwise, the elements of the off-diagonal blocks of $\mathbf{Y}$ are considered. Then, $y_{i,j} = 1$ means that the i-th and the j-th elements of $\mathbf{u}$ in two different sub-vectors are 1 and there are K^{N-2} such vectors; therefore, the elements of the off-diagonal blocks are

$$\hat{y}_{i,i} = \frac{1}{K^N} K^{N-2} = \frac{1}{K^2}.$$

This proves the representation of $\hat{\mathbf{Y}}$ in (i) in Theorem 9.

It can be easily shown that

$$\left[\begin{array}{c|c} 1 & \mathbf{0}_n^T \\ \hline -\frac{1}{K}\mathbf{e}_n & \mathbf{I}_n \end{array}\right] \hat{\mathbf{Y}} \left[\begin{array}{c|c} 1 & -\frac{1}{K}\mathbf{e}_n^T \\ \hline \mathbf{0}_n & \mathbf{I}_n \end{array}\right] = \left[\begin{array}{c|c} 1 & \mathbf{0}_n^T \\ \hline \mathbf{0}_n & \hat{\mathbf{W}} \end{array}\right], \tag{162}$$

where $\hat{\mathbf{W}} = \dfrac{1}{K^2}\mathbf{I}_N \otimes (K\mathbf{I}_K - \mathbf{E}_K)$. Note that $\operatorname{rank}(\hat{\mathbf{Y}}) = 1 + \operatorname{rank}(\hat{\mathbf{W}})$.
The eigenvalues of $\mathbf{I}_N$ are 1, with multiplicity N and the eigenvalues of $K\mathbf{I}_K - \mathbf{E}_K$ are K, with multiplicity $K - 1$, and 0. Note that the eigenvalues of a Kronecker product are given by the Kronecker product of the eigenvalues [55]. Therefore, the eigenvalues of $\hat{\mathbf{W}}$ are $\frac{1}{K}$, with multiplicity $N(K - 1)$, and 0, with multiplicity N. Therefore,

$$\operatorname{rank}(\hat{\mathbf{Y}}) = 1 + \operatorname{rank}(\hat{\mathbf{W}}) = N(K - 1) + 1. \tag{163}$$

This proves (ii) in Theorem 9.

By (162) and (163), it can be easily seen that the eigenvalues of $\hat{\mathbf{Y}}$ are $\dfrac{1}{K}$, with multiplicity $N(K - 1)$, $\dfrac{K + N}{K}$, and 0, with multiplicity N. This proves (iii) in Theorem 9.

The only constraint that defines the minimal face is $\mathbf{U}\mathbf{e}_K = \mathbf{e}_N$, or equivalently $(\mathbf{I}_N \otimes \mathbf{e}_K)\mathbf{u} = \mathbf{e}_N$. By multiplying of both sides by $\mathbf{u}^T$ and using the fact that $\mathbf{u}$ is a binary vector,

$$(\mathbf{I}_N \otimes \mathbf{e}_K)\mathbf{u}\mathbf{u}^T = \mathbf{e}_N \left(\operatorname{diag}(\mathbf{u}\mathbf{u}^T)\right)^T. \tag{164}$$

This condition is equivalent to

$$\mathbf{T}\hat{\mathbf{Y}} = \mathbf{0}. \tag{165}$$

Note that $\operatorname{rank}(\mathbf{T}) = N$. Therefore,

$$\mathcal{N}(\hat{\mathbf{Y}}) = \{\mathbf{u} : \mathbf{u} \in \mathcal{R}(\mathbf{T}^T)\}.$$

This proves (iv) in Theorem 9.

Since $\operatorname{rank}(\hat{\mathbf{V}}) = N(K-1)+1$ and using Theorem 8, the columns of $\hat{\mathbf{V}}$ span the range space of $\hat{\mathbf{Y}}$. This proves (v) in Theorem 9.

APPENDIX B

REVIEW ON SOME PRECODING METHODS

In this appendix, the average transmit energy for different methods based on the probabilistic view point of (35) is calculated. Each case corresponds to different regions $\mathcal{R}$ (or $\mathcal{R}'$) and marginal probability distributions. In all these cases, it is assumed that the vector $\mathbf{u}$ is selected uniformly over a hypercube centered at the origin with side length of $2A$, $\mathcal{R} = C_M(0, 2A)$. The energy of the vector $\mathbf{v} = \sqrt{\mathbf{\Lambda}}\mathbf{U}^T\mathbf{u}$ is equal to the transmit energy of vector $\mathbf{s}$, $\gamma = \mathbf{s}^T\mathbf{s} = \mathbf{u}^T\mathbf{Q}\mathbf{u} = \mathbf{u}^T\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\mathbf{u} = \mathbf{v}^T\mathbf{v}$. In the following, in some cases, the region $\mathcal{R}'$ is defined based on the vector $\mathbf{v}$.

B.1 Case (I) - Independent Uniform Marginal Probability Distribution

Assume that $\mathbf{u}$ is a random vector in $\mathbb{R}^M$ such that its elements are i.i.d. random variables with a uniform distribution between $-A$ and A. In order to provide the entropy of $\mathcal{H}$, there should be

$$\mathcal{H} = \log 2A \quad \text{or equivalently} \quad A = 2^{\mathcal{H}-1}. \tag{166}$$

According to (35), the average anergy of the transmit signal corresponding to $\mathbf{u}$ is

$$E\{\gamma\} = tr(\mathbf{Q\Sigma}) = tr(\mathbf{U\Lambda U}^T\mathbf{\Sigma}) = tr(\mathbf{\Lambda\Sigma}') \tag{167}$$

where $\mathbf{\Sigma}' = \mathbf{U}^T\mathbf{\Sigma}\mathbf{U}$. Note that $\mathbf{\Sigma} = \frac{A^2}{3}\mathbf{I}$ and since $\mathbf{U}$ is a unitary matrix $\mathbf{\Sigma}' = \frac{A^2}{3}\mathbf{I}$. Therefore, the average energy of transmit vector corresponding to $\mathbf{u}$ is

$$E_{cube} = E\{\gamma\} = \frac{A^2}{3}\sum_{i=1}^{M}\lambda_i = \frac{2^{2\mathcal{H}}}{12}\sum_{i=1}^{M}\lambda_i. \tag{168}$$

This case corresponds to a conventional block constellation, i.e., there is a uniform distribution over region $\mathcal{R}$ which is an M-dimensional cube centered at the origin with side length of $2A$, $\mathcal{R} = C_M(0, 2A)$. Therefore, the region $\mathcal{R}'$ for the auxiliary vector $\mathbf{v}$ is an orthotope centered at the origin and along the eigenvectors of $\mathbf{Q}$ (see Figure 24). This orthotope is the rotated version of the hypercube $\mathcal{R}$ where each side is multiplied by $\sqrt{\lambda_i}$. It is clear that $\text{Vol}(\mathcal{R}') = \mathbb{V}\sqrt{\Pi\lambda_i} = (2A)^M\sqrt{\Pi\lambda_i}$. When the channel condition is poor, this ortotope is stretched, resulting in a large average transmit energy.

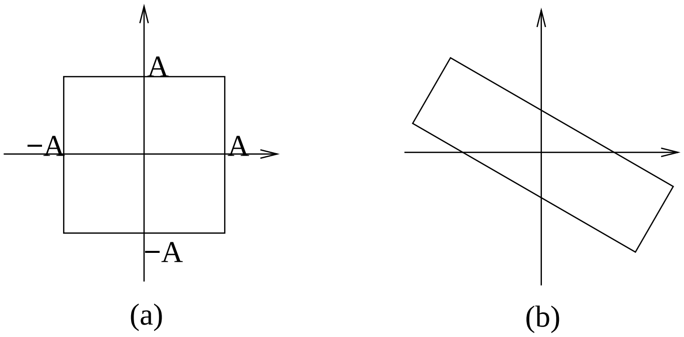

Figure 24: Schematic region $\mathcal{R}$ for the uniform distribution and its corresponding auxiliary region $\mathcal{R}'$.

B.2 Case (II) - Regularization

In [114], a method is introduced to reduce the average transmit energy. In this technique, a multiple of the identity matrix is added to the channel matrix before inverting in (33), i.e.,

$$\mathbf{s} = \mathbf{H^T}(\mathbf{HH^T} + \alpha\mathbf{I_M})^{-\mathbf{1}}\mathbf{u}, \tag{169}$$

where it is shown that $\alpha_{opt} = M/\rho$. Since $\mathbf{HH}^T = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T$, the auxiliary vector $\mathbf{v}$ can be written as[1]

$$\mathbf{v} = \frac{\sqrt{\mathbf{\Lambda}}}{\alpha\mathbf{\Lambda} + \mathbf{I}_M}\mathbf{U}\mathbf{u}.$$

Here, the region $\mathcal{R}'$ is an orthotope whose sides compared to the one in Case (I) are smaller (the sides are divided by $1 + \alpha\lambda > 1$). This orthotope is called a *reduced-orthotope* compared to the one in Case (I). By assuming the same uniform distribution for $\mathbf{u}$, the average transmit energy of this signal is

$$E_{reg} = E\{\gamma\} = \frac{A^2}{3}\sum_{i=1}^{M}\frac{\lambda_i}{(\alpha\lambda_i + 1)^2}, \tag{170}$$

where for the same reason is less than E_{cube}.

In the channel inversion technique, a linear transformation $\mathbf{T}$ is desired such that it maps the region $\mathcal{R}$ to some other region. This region should be such that when its points are multiplied by $\mathbf{H}^{-1}$ the average transmit energy is minimized, while the minimum mean square error in the receivers are minimized. In other words, a matrix $\mathbf{T}$ is desired such that $\mathbf{s} = \mathbf{H}^{-1}\mathbf{T}\mathbf{u}$ with minimum error in decoding regarding the interference caused by $\mathbf{T}$ and with minimum average transmit energy for the new region $\mathbf{T}\mathcal{R}$.

[1] The operations on the diagonal matrices are based on the operations on the diagonal elements.

In [65], it is shown that in order to satisfy these conditions, the matrix $\mathbf{H}^{-1}\mathbf{T}$ should be a Wiener filter, i.e.,

$$\mathbf{H}^{-1}\mathbf{T} = \mathbf{H}^{T}(\mathbf{H}\mathbf{H}^{\mathbf{T}} + \alpha\mathbf{I}_{\mathbf{M}})^{\mathbf{-1}}.$$

By using matrix inversion lemma [53],

$$\mathbf{H}^{T}(\mathbf{H}\mathbf{H}^{\mathbf{T}} + \alpha\mathbf{I}_{\mathbf{M}})^{\mathbf{-1}} = \mathbf{H}^{\mathbf{-1}}\left(\mathbf{I}_{\mathbf{M}} - \alpha(\mathbf{H}\mathbf{H}^{\mathbf{T}} + \alpha\mathbf{I}_{\mathbf{M}})^{\mathbf{-1}}\right).$$

Therefore,

$$\mathbf{T} = \left(\mathbf{I}_{M} - \alpha(\mathbf{H}\mathbf{H}^{\mathbf{T}} + \alpha\mathbf{I}_{\mathbf{M}})^{\mathbf{-1}}\right).$$

This simple argument emphasis that the method in [114] proposes a lattice code defined by the region $\mathbf{T}\mathcal{R}$. This lattice code is the best lattice code that can be achieved by linear transformation which has the minimum transmit energy while minimizes the mean square error at the decoding process in the receiver side.

B.3 Case (III) - Perturbation Technique

In [115], a perturbation method is introduced in order to reduce the average energy of the transmit signal. The data is transmitted by judiciously adding an integer vector offset. Instead of sending the transmit signal in (33), the perturbed version of this signal is transmitted as

$$\mathbf{s} = \mathbf{H}^{-1}(\mathbf{u} + \tau\mathbf{l}), \tag{171}$$

where τ is a positive real number and it is chosen large enough so that the receivers may apply the modulo function. The vector $\mathbf{l}$ is chosen such that the average energy γ is minimized.

Let $\mathbf{u}$ be a point selected uniformly over $C_M(0, 2A)$ and $\tau = 2A$. Therefore, vector $\mathbf{s}$, based on (33), is selected uniformly over a polytop whose sides are along the columns of the matrix $\mathbf{H}^{-1}$. This polytope is actually a fundamental region of the lattice $\tau\mathbf{H}^{-1}$ centered at zero (for definition of fundamental region refer to [27]).

By using (171), the hypercubic region of $\mathcal{R}$ is expanded and the vectors $\mathbf{l}$ are found in the expanded region such that the vectors $\mathbf{s}$ with the minimum energy are selected. In other words, by using perturbation in (171), vector $\mathbf{s}$ is uniformly selected over a region which is known as the Voronoi region of the lattice $\tau\mathbf{H}^{-1}$. The Voronoi region of a lattice is the set of points in $\mathbb{R}^M$ which are closer to the origin than any other points in the lattice [27]. Denote by $\mathcal{V}(\tau\mathbf{H}^{-1})$ the Voronoi region of $\tau\mathbf{H}^{-1}$. Therefore, the average transmit energy of perturbation technique [115] would be

$$E_{perturb} = E(\gamma) = \int_{\mathcal{V}(\tau\mathbf{H}^{-1})} ||x||^2 dF(x), \tag{172}$$

where $F(x)$ is the uniform distribution over $\mathcal{V}(\tau\mathbf{H}^{-1})$.

Since the probability distribution over the Voronoi region is uniform, the average energy in (172) can be formulated by the second moment of $\tau\mathbf{H}^{-1}$ [27]. The dimensionless second moment of the Voronoi region $\mathcal{V}(\tau\mathbf{H}^{-1})$ is defined as

$$\begin{aligned} G(\mathcal{V}(\tau\mathbf{H}^{-1})) &= \frac{1}{M}\frac{\int_{\mathcal{V}(\tau\mathbf{H}^{-1})}\|x\|^2dx}{\text{Vol}(\mathcal{V}(\tau\mathbf{H}^{-1}))^{1+\frac{2}{M}}} \\ &= \frac{1}{M}\frac{\int_{\mathcal{V}(\tau\mathbf{H}^{-1})}\|x\|^2dF(x)}{\text{Vol}(\mathcal{V}(\tau\mathbf{H}^{-1}))^{\frac{2}{M}}} \qquad (173) \end{aligned}$$

Therefore,

$$E_{perturb} = MG(\mathcal{V}(\tau\mathbf{H}^{-1}))\left(\text{Vol}(\mathcal{V}(\tau\mathbf{H}^{-1}))\right)^{\frac{2}{M}}, \qquad (174)$$

where

$$\text{Vol}(\mathcal{V}(\tau\mathbf{H}^{-1})) = \sqrt{\det(\mathbf{Q})}(2A)^M. \qquad (175)$$

In [27], it is shown that $G(\mathcal{V}(\tau\mathbf{H}^{-1})) \geq G_M$ where

$$\begin{aligned} &\frac{1}{(M+2)\pi}\Gamma\left(\tfrac{M}{2}+1\right)^{\frac{2}{M}} \leq G_M \leq \\ &\frac{1}{M\pi}\Gamma\left(\tfrac{M}{2}+1\right)^{\frac{2}{M}}\Gamma\left(1+\tfrac{2}{M}\right) \qquad (176) \end{aligned}$$

Generally, the method in [115] changes the region of the transmit signal **s** form a fundamental region to the Voronoi region of $\tau\mathbf{H}^{-1}$. The closer the Voronoi region is to a ball, the smaller the value of the average transmit energy would be. However, the Voronoi region of $\tau\mathbf{H}^{-1}$ is fixed and it can be far away from a ball. The methods in [129] and [144] change the region of the transmit signal **s** form a fundamental region to another fundamental region which is more similar to a ball.

B.4 Case (IV) - Independent Gaussian Marginal Probability Distribution

Now, consider the theoretical case that **u** is a random vector in $\mathbb{R}^M$ such that its elements are i.i.d. random variables with a Gaussian distribution with zero mean and variance σ^2, i.e. $N(0, \sigma^2\mathbf{I}_M)$. Satisfying the entropy $\mathcal{H}$ per dimension requires that

$$\mathcal{H} = \frac{1}{2}\log 2\pi e\sigma^2. \qquad (177)$$

The average transmit energy of **u** is $E(\gamma) = tr(\mathbf{\Lambda\Sigma}')$. The covariance matrix is a diagonal matrix since the elements are i.i.d., $\mathbf{\Sigma} = \sigma^2\mathbf{I}$. By multiplication of a unitary matrix, there are i.i.d. Gaussian elements with $\mathbf{\Sigma}' = \sigma^2\mathbf{I}$. Therefore, the average energy is

$$E_{sphere} = E(\gamma) = \sigma^2\sum_{i=1}^{M}\lambda_i = \frac{2^{2\mathcal{H}}}{2\pi e}\sum_{i=1}^{M}\lambda_i. \qquad (178)$$

According to (168) and (178), $\mathcal{G}_{sphere} = \frac{E_{cube}}{E_{sphere}} = \frac{\pi e}{6}$, which is known as the shaping gain (the advantage of using Gaussian distribution instead of uniform distribution). The idea of shaping has been frequently used in different applications to generate signals with desired properties.

APPENDIX C

LAGRANGIAN DUALITY

In this appendix, it is shown that Lagrangian duality can be used to derive the SDP relaxation problem (77). First, the dual for the constraints of (65) is found, and then, an SDP relaxation from the dual of the homogenized Lagrangian dual is derived. Finally, the obtained relaxation is projected onto the minimal face. The resulting relaxation is equivalent to the relaxation (77).

It is easy to show that the minimization problem in (65) is equivalent to

$$\begin{aligned} \min \quad & \mathbf{u}^T(\mathbf{Q}\otimes\mathbf{S})\mathbf{u} + 2\mathrm{vec}(\mathbf{C})^T\mathbf{u} \\ s.t. \quad & (\mathbf{I}_N\otimes\mathbf{e}_K^T)\mathbf{u} = \mathbf{e}_N \\ & u_i^2 = u_i \quad \forall i = 1,\cdots,n. \end{aligned} \tag{179}$$

According to [146], for an accurate semi-definite solution, zero-one constraints should be formulated as quadratic constraints. Therefore,

$$\begin{aligned} \min \quad & \mathbf{u}^T(\mathbf{Q}\otimes\mathbf{S})\mathbf{u} + 2\mathrm{vec}(\mathbf{C})^T\mathbf{u} \\ s.t. \quad & ||(\mathbf{I}_N\otimes\mathbf{e}_K^T)\mathbf{u} - \mathbf{e}_N||^2 = 0 \\ & u_i^2 = u_i \quad \forall i = 1,\cdots,n. \end{aligned} \tag{180}$$

First, the constraints are added to the objective function using lagrange multipliers λ and $\tilde{\mathbf{w}} = [\tilde{w}_1,\cdots,\tilde{w}_n]^T$:

$$\begin{aligned} \mu_O = \min_{\mathbf{u}}\max_{\lambda,\tilde{\mathbf{w}}} \Big\{ & \mathbf{u}^T(\mathbf{Q}\otimes\mathbf{S})\mathbf{u} + 2\mathrm{vec}(\mathbf{C})^T\mathbf{u} \\ & + \lambda\left(\mathbf{u}^T(\mathbf{I}_N\otimes\mathbf{E}_K)\mathbf{u} - 2\mathbf{e}_{NK}^T\mathbf{u} + N\right) \\ & + \sum_{i=1}^{n}\tilde{w}_i\left(u_i^2 - u_i\right)\Big\}. \end{aligned} \tag{181}$$

Interchanging min and max yields

$$\begin{aligned} \mu_O \geq \mu_{\mathcal{L}} = \max_{\lambda,\tilde{\mathbf{w}}}\min_{\mathbf{u}} \Big\{ & \mathbf{u}^T(\mathbf{Q}\otimes\mathbf{S})\mathbf{u} + 2\mathrm{vec}(\mathbf{C})^T\mathbf{u} \\ & + \lambda\left(\mathbf{u}^T(\mathbf{I}_N\otimes\mathbf{E}_K)\mathbf{u} - 2\mathbf{e}_{NK}^T\mathbf{u} + N\right) \\ & + \sum_{i=1}^{n}\tilde{w}_i\left(u_i^2 - u_i\right)\Big\}. \end{aligned} \tag{182}$$

Next, the objective function is homogenized by multiplying it with a constrained scalar u_0 and then increasing the dimension of the problem by 1. Homogenization simplifies the transition to a semi-definite programming problem. Therefore,

$$\begin{aligned}\mu_O \geq \mu_{\mathcal{L}} = \max_{\lambda,\tilde{\mathbf{w}}} \min_{\mathbf{u},u_0^2=1} \Big\{ & \mathbf{u}^T \left[\mathbf{Q} \otimes \mathbf{S} + \lambda \mathbf{I}_N \otimes \mathbf{E}_K + \mathrm{Diag}(\tilde{\mathbf{w}})\right] \mathbf{u} \\ & - \left(2\lambda \mathbf{e}_{NK}^T - 2\mathrm{vec}(\mathbf{C})^T + \tilde{\mathbf{w}}^T\right) u_0 \mathbf{u} \\ & + \lambda N \Big\}, \end{aligned} \tag{183}$$

where $\mathrm{Diag}(\tilde{\mathbf{w}})$ is a diagonal matrix with $\tilde{\mathbf{w}}$ as its diagonal elements. By introducing a Lagrange multiplier w_0 for the constraint on u_0, the lower bound $\mu_{\mathcal{R}}$ is obtained

$$\begin{aligned}\mu_O \geq \mu_{\mathcal{L}} \geq \mu_{\mathcal{R}} = \max_{\lambda,\tilde{\mathbf{w}},w_0} \min_{u_0,\mathbf{u}} \Big\{ & \mathbf{u}^T \left[\mathbf{Q} \otimes \mathbf{S} + \lambda \mathbf{I}_N \otimes \mathbf{E}_K + \mathrm{Diag}(\tilde{\mathbf{w}})\right] \mathbf{u} \\ & - \left(2\lambda \mathbf{e}_{NK}^T - 2\mathrm{vec}(\mathbf{C})^T + \tilde{\mathbf{w}}^T\right) u_0 \mathbf{u} \\ & + \lambda N u_0^2 + w_0 \left(u_0^2 - 1\right) \Big\}. \end{aligned} \tag{184}$$

Note that both inequalities can be strict, i.e. there can be duality gaps in each of the Lagrangian relaxations. Also, the multiplication of $\lambda \mathbf{E}_N$ by u_0^2 is a multiplication by 1. Now, by grouping the quadratic, linear, and constant terms together and defining $\tilde{\mathbf{u}}^T = \left[u_0, \mathbf{u}^T\right]^T$ and $\mathbf{w}^T = \left[w_0, \tilde{\mathbf{w}}^T\right]^T$, the following relaxation is obtained:

$$\mu_{\mathcal{R}} = \max_{\lambda,\mathbf{w}} \min_{\tilde{\mathbf{u}}} \left\{ \tilde{\mathbf{u}}^T [\mathcal{L}_{\mathbf{Q}} + \mathrm{Arrow}(\mathbf{w}) + \lambda \mathcal{L}_\lambda] \tilde{\mathbf{u}} - w_0 \right\}, \tag{185}$$

where

$$\begin{aligned}\mathcal{L}_\lambda &= \begin{bmatrix} N & -\mathbf{e}_{NK}^T \\ -\mathbf{e}_{NK} & \mathbf{I}_N \otimes (\mathbf{E}_K) \end{bmatrix}, \\ \mathrm{Arrow}(\mathbf{w}) &= \begin{bmatrix} w_0 & -\frac{1}{2}\mathbf{w}_{1:n}^T \\ -\frac{1}{2}\mathbf{w}_{1:n} & \mathrm{Diag}(\mathbf{w}_{1:n}) \end{bmatrix}, \\ \text{and } \mathcal{L}_{\mathbf{Q}} &= \begin{bmatrix} 0 & \mathrm{vec}(\mathbf{C})^T \\ \mathrm{vec}(\mathbf{C}) & \mathbf{Q} \otimes \mathbf{S} \end{bmatrix}. \end{aligned} \tag{186}$$

Note that the additional row and column generated by the homogenization of the problem is referred as the 0-th row and column. There is a hidden semi-definite constraint in (185), i.e. the inner minimization problem is bounded below only if the Hessian of the quadratic form is positive semi-definite. In this case, the quadratic form has minimum value 0. This yields the following SDP problem:

$$\begin{aligned} \max \quad & - w_0 \\ s.t. \quad & \mathcal{L}_{\mathbf{Q}} + \mathrm{Arrow}(\mathbf{w}) + \lambda \mathcal{L}_\lambda \succeq 0. \end{aligned} \tag{187}$$

The desired SDP relaxation of (180) is the Lagrangian dual of (187). By introducing the $(n+1)\times(n+1)$ dual matrix variable $\mathbf{Y}\succeq 0$, the dual program to the SDP (187) would be

$$\begin{aligned} \min\ & \operatorname{trace} \mathcal{L}_{\mathbf{Q}}\mathbf{Y} \\ s.t.\ & \operatorname{diag}(\mathbf{Y}) = (1, \mathbf{Y}_{0,1:n})^T \\ & \operatorname{trace}\mathcal{L}_{\lambda}\mathbf{Y} = 0 \\ & \mathbf{Y}\succeq 0, \end{aligned} \tag{188}$$

where the first constraint represents the zero-one constraints in (180) by guaranteeing that the diagonal and 0-th column (row) are identical (matrix $\mathbf{Y}$ is indexed from 0); and the constraint $(\mathbf{I}_N\otimes\mathbf{e}_K^T)\mathbf{u}=\mathbf{e}_N$ is represented by the constraint $\operatorname{trace}\mathcal{L}_{\lambda}\mathbf{Y}=0$. Note that if the matrix $\mathbf{Y}$ is restricted to be rank-one in (188), i.e.

$$\mathbf{Y} = \begin{bmatrix} 1 \\ \mathbf{u} \end{bmatrix}\begin{bmatrix} 1 & \mathbf{u}^T \end{bmatrix},$$

for some $\mathbf{u}\in\mathbb{R}^n$, then the optimal solution of (188) provides the optimal solution, $\mathbf{u}$, for (180).

Since the matrix $\mathcal{L}_{\lambda}\neq 0$ is a positive semi-definite matrix; therefore, to satisfy the constraint in (188), $\mathbf{Y}$ has to be singular. This means the feasible set of the primal problem in (188) has no interior [156] and an IPM may never converge. However, a simple structured matrix can be found in the relative interior of the feasible set in order to project (and regularize) the problem into a smaller dimension.

As mentioned before, the rank-one matrices are the extreme points of the feasible set of the problem in (188) and the minimal face of the feasible set that contains all these points shall be found [156].

From Theorems 8 and 9, it can be concluded that $\mathbf{Y}\succeq 0$ is in the minimal face if and only if $\mathbf{Y}=\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T$, for some $\mathbf{R}\succeq 0$. By substituting $\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T$ for $\mathbf{Y}$ in the SDP relaxation (188), the following projected SDP relaxation is obtained which is the same as the SDP relaxation in (77):

$$\begin{aligned} \mu_{R1} = \min\ & \operatorname{trace}\,(\hat{\mathbf{V}}^T\mathcal{L}_{\mathbf{Q}}\hat{\mathbf{V}})\mathbf{R} \\ \text{s.t.}\ & \operatorname{diag}(\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T) = (1, (\hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{V}}^T)_{0,1:n})^T \\ & \mathbf{R}\succeq 0. \end{aligned} \tag{189}$$

Note that the constraint $\operatorname{trace}(\hat{\mathbf{V}}^T\mathcal{L}_{\lambda}\hat{\mathbf{V}})\mathbf{R}=0$ is dropped since it is always satisfied, i.e. $\mathcal{L}_{\lambda}\hat{\mathbf{V}}=0$.

REFERENCES

[1] Agrell, E., Eriksson, T., Vardy, A., and Zeger, K., "Closest point search in lattices," *IEEE Transaction on Information Theory*, vol. 48, pp. 2201–2214, Aug. 2002.

[2] Agrell, E., Lassing, J., Ström, E. G., and Ottosson, T., "On the optimality of the binary reflected gray code," *IEEE Trans. on Inform. Theory*, vol. 50, pp. 3170–3182, Dec 2004.

[3] Alizadeh, F., Haeberly, J.-P. A., and Overton, M., "Primal-dual interior-point methods for semidefinite programming: Stability, convergence, and numerical results," *SIAM Journal on Optimization*, vol. 8, no. 3, pp. 746–768, 1998.

[4] Babai, L., "On Lovasz' lattice reduction and the nearest lattice point problem," *Combinatorica 6*, pp. 1–13, 1986.

[5] Balas, E., Ceria, S., and Cornuejols, G., "A lift-and-project cutting plane algorithm for mixed 0–1 programs," *Mathematical Programming*, vol. 58, pp. 295–324, 1993.

[6] Banihashemi, A. H. and Khandani, A. K., "Lattice decoding using the Korkin-Zolotarev reduced basis," Tech. Rep. UW-E&CE 95-12, Elec. & Comp. Eng. Dept., Univ. of Waterloo, Waterloo, Ont., Canada, 1995.

[7] Banihashemi, A. H. and Khandani, A. K., "On the Complexity of Decoding Lattices Using the Korkin-Zolotarev Reduced Basis," *IEEE Transactions in Information Theory*, vol. IT-44, pp. 162–171, January 1998.

[8] Barvinok, A., "Problems of distance geometry and convex properties of quadratic maps," *Discrete Computational Geometry*, vol. 13, pp. 189–202, 1995.

[9] Bauml, R. W., Fischer, R. F. H., and Huber, J. B., "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *Electronic Letters*, vol. 32, pp. 2056–2057, 1996.

[10] Bayer-Fluckiger, E., "Lattices and number fields," *Contemporary Mathematics*, vol. 241, pp. 69–84, 1999.

[11] Beck, A., "Quadratic matrix programming," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 1224–1238, 2007.

[12] Belfiore, J.-C. and Rekaya, G., "Quaternionic lattices for space-time coding," in *ITW2003*, (Paris), April 2003.

[13] Belfiore, J.-C., Rekaya, G., and Viterbo, E., "The Golden code: A 2×2 full rate space-time code with non-vanishing determinants," in *IEEE International Symposium on Information Theory*, 2004.

[14] Benson, S., Ye, Y., and Zhang, X., "Mixed linear and semidefinite programming for combinatorial and quadratic optimization," *Optimization Methods and Software*, vol. 11-12, pp. 515–544, 1999.

[15] Benson, S. J. and Ye, Y., "DSDP5 User Guide Ű The Dual-Scaling Algorithm for Semidefinite Programming," Tech. Rep. ANL/MCS-TM-255, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, 2004. Available via the WWW site at http://www.mcs.anl.gov/~benson.

[16] Bernardini, R. and Manduchi, R., "On the Reduction of Multidimensional DFT to Separabile DFT by Smith Normal form Theorem," *Signal Processing Letter*, vol. 5, May-June 1994.

[17] Breiling, M., Muller-Weinfurtner, S. H., and Hubber, J. B., "SLM Peak-Power Reduction Without Explicit Side Information," *IEEE Commun. Lett.*, vol. 5, pp. 239–241, June 2001.

[18] Burer, S. and Monteiro, R. D. C., "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Mathematical Programming (Series B)*, vol. 95, pp. 329–357, 2003.

[19] C. P. Schnorr, M. E., "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Mathematical Programming*, no. 66, pp. 181–191, 1994.

[20] Callard, A., Khandani, A. K., and Saleh, A., "Vector precoding with mmse for the fast fading and quasi-static multi-user broadcast channel," in *Conference on Information Sciences and Systems*, 2006.

[21] Carson, N. and Gulliver, T. A., "Peak-to-Average Power Ratio Reduction of OFDM Using Repeat-Accumulate Codes and Selective Mapping," *2002 IEEE International Symposium on Information Theory (ISIT 2002)*, p. 244, June 30-July 5 2002.

[22] Chan, A. M. and Lee, I., "A new reduced-complexity sphere decoder for multiple antenna systems," in *IEEE International Conference on Communications*, pp. 460–464, 2002.

[23] Cochet, P. Y. and Serpollet, R., "Digital transformfor a selective channel estimation," in *the IEEE International Conference on Communication*, IEEE, IEEE, 1998.

[24] Cohen, H., *Graduate Texts in Mathematics: A Course in Computational Algebraic Number Theory*, vol. 138. Spring-Verlag, 1993.

[25] Cohort, P., "Limit theorems for random normalized distortion," *Annals of Applied Probability*, vol. 14, p. 118Ű143, Mar. 2004.

[26] Collings, I. and Clarkson, I., "A low-complexity lattice-based low-PAR transmission scheme for DSL channels," *IEEE Trans. on Commun.*, vol. 52, pp. 755 – 764, May 2004.

[27] Conway, J. H. and Sloane, N. J. A., “Voronoi, regions of lattices, second moments of polytopes, and quantization,” *IEEE Trans. on Info. Theory*, vol. 28, pp. 211–226, Mar. 1982.

[28] Conway, J. H. and Sloane, N. J. A., *Sphere packings, lattices, and groups*. New York: Springer-Verlag, 2nd ed., 1993.

[29] Costa, M., “Writing on dirty paper,” *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 439–441, May 1983.

[30] Damen, M. O., Abed-Meriam, K., and Belfiore, J.-C., “A generalized lattice decoder for asymmetrical space-time communication architecture,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing 2000 (ICASSP 2000)*, pp. 2581–2584, 2000.

[31] Damen, M. O., Chkeif, A., and Belfiore, J.-C., “Lattice Code Decoder for Space-Time Codes,” *IEEE Communications Letters*, vol. 4, pp. 161–163, May 2000.

[32] Damen, M. O., Chkeif, A., and Belfiore, J.-C., “Sphere Decoding of Space-Time Codes,” in *ISIT 2000*, (Sorrento, Italy), p. 362, June 25-30 2000.

[33] Damen, M. O., El-Gamal, H., and Cairo, G., “On maximum-likelihood detection and the search for the closest lattice point,” *IEEE Transactions on Information Theory*, vol. 49, pp. 2389–2402, Oct. 2003.

[34] Damen, M. O., Gamal, H. E., and Caire, G., “MMSE-GDFE Lattice Decoding for Under-determined Linear Channels,” in *Conference on Information Sciences and Systems (CISS04)*, March 2004.

[35] David, H. A. and Nagaraja, H. N., *Order Statistics*. Wiley Series in Probability and Statistics, Hoboken, N.J., Wiley-Interscience, 3 ed., 2003.

[36] Davis, J. A. and Jedwab, J., “Peak-to-mean power control in OFDM, Golay complementary sequences, and Reed-Muller codes,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 2397–2417, Nov. 1999.

[37] de Klerk, E., *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*, vol. 65 of *Applied Optimization*. Kluwer Academic Publishers, March 2002.

[38] Debbah, M., Muquet, B., de Courville, M., Muck, M., Simoens, S., and Loubaton, P., “A MMSE successive interference cancellation scheme for a new adjustable hybrid spread OFDM system,” in *IEEE VTC*, (Tokyo (Japan)), pp. 745–749, Spring 2000.

[39] Dembo, A., Cover, T. M., and Thomas, J. A., “Information theoretic inequalities,” *IEEE Trans. on Info. Theory*, vol. 37, pp. 1501–1518, Nov. 1991.

[40] Ding, Y. and Wolkowicz, H., “A Matrix-lifting Semidefinite Relaxation for the Quadratic Assignment Problem,” Tech. Rep. CORR 06-22, Department of Combinatorics & Optimization, University of Waterloo, 2006.

[41] Duel-Hallen, A., Holtzman, J., and Zvonar, Z., "Multiuser detection for CDMA systems," *IEEE Personal Communications Magazine*, pp. 46–58, April 1995.

[42] Eatvelt, J. V., Wade, G., and Tomlinson, M., "Peak to average power reduction for OFDM schemes by selective scrambling," *Electronic Letters*, vol. 32, pp. 1963–1964, 1996.

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

[44] Fischer, R., Windpassinger, C., Lampe, A., and Huber, J., "Tomlinson-harashima precoding in space-time transmission for low-rate backward channel," in *IEEE International Zurich Seminar on Broadband Communications, Access, Transmission, Networking*, (Zurich, Switzerland), pp. 7–1–7–6, 2002.

[45] Foschini, G. J., "Layered Space-Time Architecture for Wireless Communication in a Fading Environment when using Multi-Element Antennas," *Bell Labs Technical Journal*, pp. 41–59, Autumn 1996.

[46] G. D. Forney ,Jr., "Coset Codes - Part I: Introduction and Geometrical Classification," *IEEE Tran. Inform. Theory*, vol. 34, pp. 1123–1151, September 1988.

[47] G. D. Forney, Jr., "Multidimensional constellations - Part II: Voronoi constellation," *IEEE J. Select. Areas Commun*, vol. SAC-7, pp. 941–958, August 1989.

[48] G. D. Forney, Jr. and Trott, M. D., "The dynamics of group codes: state spaces, trellis diagrams, and canonical encoders," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 1491–1513, Sept. 1993.

[49] G. D. Forney, Jr. and Wei, F., "Multidimensional constellations - Part I: Introduction, Figures of merit, and Generalized Cross Constellations," *IEEE J. Select. Areas Commun.*, vol. SAC-7, pp. 877–892, August 1989.

[50] Ginis, G. and Cioffi, J. M., "Vectored transmission for digital subscriber line systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, p. 1085Ű1104, June 2002.

[51] Goemans, M. and Williamson, D., "Improved approximation algorithms for maximum cut and satisfability problem using semidefinite programming," *Journal of ACM*, vol. 42, pp. 1115–1145, 1995.

[52] Golden, G. D., Foschini, G. J., Valenzuela, R. A., and Wolniansky, P. W., "Detection Algorithm and Initial Laboratory Results using V-BLAST Space-Time Communication Architecture," *Electronics Letters*, vol. 35, pp. 14–16, January 1999.

[53] Golub, G. and Loan, C. V., *Matrix Computations*. North Oxford Academic, 1986.

[54] Graf, S. and Luschgy, H., *Foundation of Quantization for Probability Distributions*, vol. 1730 of *Lecture Notes in Mathematics*. Springer, 2000.

[55] Graham, A., *Kronecker Products and Matrix Calculus with Applications, Mathematics and its Applications*. Ellis Horwood Limited, Chichester, 1981.

[56] Hadley, S., Rendl, F., and Wolkowicz, H., "A new lower bound via projection for the quadratic assignment problem," *Math. Oper. Res.*, vol. 17, no. 3, pp. 727–739, 1992.

[57] Hassibi, B. and Hochwald, B., "High-rate codes that are linear in space and time," *IEEE Trans. Info. Theory*, vol. 48, pp. 1804–1824, July 2002.

[58] Hassibi, B. and Vikalo, H., "On Sphere Decoding Algorithm. Part I. Expected Complexity." Submitted to IEEE Transaction on Signal Processing, 2003.

[59] Hassibi, B. and Vikalo, H., "On Sphere Decoding Algorithm. Part II. generalization, second order statistics and applications to communications." Submitted to IEEE Transaction on Signal Processing, 2003.

[60] Haustein, T., von Helmolt, C., Jorswieck, E., Jungnickel, V., and Pohl, V., "Performance of mimo systems with channel inversion," in *55th IEEE Veh. Technol. Conf.*, vol. 1, (Birmingham, AL), p. 35Ű39, May 2002.

[61] Helfrich, B., "Algorithms to construct Minkowski reduced and Hermit reduced lattice bases," *Theoretical Computer Sci. 41*, pp. 125–139, 1985.

[62] Henkel, W. and B.Wagner, "Another application for trellis shaping: PAR reduction for DMT (OFDM)," *IEEE Trans. on Comm.*, vol. 48, p. 1471Ű1476, Sept. 2000.

[63] Hochwald, B. M. and ten Brink, S., "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, pp. 389 –399, March 2003.

[64] Hoffmann-Jørgensen, J., *Probability witha view toward statistics*, vol. 1. Chapman and Hall, 1994.

[65] Honig, M., Madhow, U., and Verdu, S., "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 994–960, July 1995.

[66] Jaldén, J. and Ottersten, B., "On the complexity of sphere decoding in digital communications," *IEEE Trans. on Signal Proc.*, vol. 53, no. 4, pp. 1474–1484, 2005.

[67] Kannan, R., "Improved algorithms on integer programming and related lattice problems," in *Proc. 15th Annu. ACM Symp. on Theory of Computing*, pp. 193–206, 1983.

[68] Kannan, R., "Minkowski s convex body theorem and integer programming," *Math. of Operations Res.*, vol. 12, p. 415 440, Aug. 1987.

[69] Kohno, R., Imai, H., and Hatori, M., "Cancellation techniques of co-channel interference in asynchronous spread spectrum multiple access systems," *Trans. Elect. and Comm. in Japan*, vol. 66, pp. 416–423, May 1983.

[70] Kojima, M., Fujisawa, K., Nakata, K., and Yamashita, M., "SDPA(SemiDefinite Programming Algorithm) UserŠs Mannual Ů Version 6.00," tech. rep., Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo, 152-0033, Japan, 2002. Available via the WWW site at http://sdpa.is.titech.ac.jp/SDPA/.

[71] Korkine, A. and Zolotareff, G., "Sur les formes quadratiques," *Mathematische Annalen*, vol. 6, pp. 366–389, 1873 (in French).

[72] Krongold, B. S. and Jones, D. L., "PAR Reduction in OFDM via Active Constellation Extension," *IEEE Trans. Broadcasting*, vol. 49, pp. 258–268, Sept. 2003.

[73] Krongold, B. S. and Jones, D. L., "An Active-Set Approach for OFDM PAR Reduction via Tone Reservation," *IEEE Trans. on Signal Processing*, vol. 52, pp. 495–509, Feb. 2004.

[74] Kulis, B., Surendran, A. C., and Platt, J. C., "Fast low-rank semidefinite programming for embedding and clustering," in *Eleventh International Conference on Artifical Intelligence and Statistics, AISTATS 2007*, March 2007.

[75] Kwok, H. K., *Shape Up: Peak-Power Reduction via Constellation Shaping*. PhD thesis, University of Illinois, 2001.

[76] Kwok, H. K. and Jones, D. L., "PAR Reduction for Hadamard Transform-Based OFDM," in *34th Conference on Signal, Systems, and Computers*, (Princeton, NJ), March 15-17 2000.

[77] Kwok, H. K. and Jones, D. L., "PAR Reduction via Constellation Shaping," in *2000 International Symposium on Information Theory*, (Sorrento, Italy), June 25-30 2000.

[78] Lancaster, P. and Tismenetsky, M., *The theory of matrices*. Academic Press, 1985. Second Edition with Applications.

[79] Lee, Y.-L., You, Y.-H., Jeon, W.-G., Paik, J.-H., and Song, H.-K., "Peak-to-Average Power Ratio in MIMO-OFDM Systems Using Selective Mapping," *IEEE Comm. Lett.*, vol. 7, pp. 575–577, Dec. 2003.

[80] Lenstra, A. K., Lenstra, H. W., and Lovász, L., "Factoring polynomials with rational coefficients," *Mathematische Annalen*, vol. 261, pp. 515–534, 1982.

[81] Li, X. and Cimini, L. J., "Effects of clipping and filtering on the performance of OFDM," *IEEE Comm. Lett.*, vol. 2, pp. 131–133, May 1998.

[82] Linder, T., Schlegel, C., and Zeger, K., "Corrected proof of de BudaŚs theorem," *IEEE Trans. Inform. Theory*, pp. 1735–1737, Sept. 1993.

[83] Loeliger, H. A., "Averaging bounds for lattices and linear codes," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1767–1773, Nov. 1997.

[84] Lovasz, L., "On the shannon capacity of a graph," *IEEE Trans. on Info. Theory*, vol. 25, pp. 1–7, Jan. 1979.

[85] Luo, Z. Q., Luo, X., and Kisialiou, M., "An Efficient Quasi-Maximum Likelihood Decoder for PSK Signals," in *ICASSP '03*, 2003.

[86] Ma, W.-K., Ching, P. C., and Ding, Z., "Semidefinite relaxation based multiuser detection for m-ary psk multiuser systems," *IEEE Trans. on Signal Processing*, 2004.

[87] Maddah-Ali, M. A., Mobasher, A., and Khandani, A. K., "On the Fairest Corner Point of the MIMO-BC Capacity Region," in *the 43rd Annual Allerton Conference on Communication, Control, and Computing*, (Monticello, IL, USA), Sept. 2005.

[88] Maddah-Ali, M. A., Mobasher, A., and Khandani, A. K., "Using Polymatroids to provide Fairness in Multi-User Systems," in *IEEE International Symposium on Information Theory, ISIT'06*, (Seattle, WA, US), July 2006.

[89] Maddah-Ali, M. A., Mobasher, A., and Khandani, A. K., "Fairness in Multiuser Systems with Polymatroid Capacity Region," *Submitted to IEEE Trans. on Info. Theory*, Jul. 2007. Revised, Expected publication.

[90] Matthai, A. M. and Provost, S. B., *Quadratic Forms in Random Variables: Theory and Applications*, vol. 126 of *Statistics: textbooks and monographs*. New York: Marcel Dekker, Inc., 1992.

[91] Mobasher, A., Taherzadeh, M., Sotirov, R., and Khandani, A. K., "A Near Maximum Likelihood Decoding Algorithm for MIMO Systems Based on Graph Partitioning," in *IEEE International Symposium on Information Theory*, Sept. 4-9 2005.

[92] Mobasher, A., Taherzadeh, M., Sotirov, R., and Khandani, A. K., "A Randomization Method for Quasi-Maximum Likelihood Decoding," in *Canadian Workshop on Information Theory*, Jun. 5-8 2005.

[93] Mobasher, A., Taherzadeh, M., Sotirov, R., and Khandani, A. K., "An Efficient Quasi-Maximum Likelihood Decoding for Finite Constellations," in *Conference on Information Sciences and Systems (CISS) 2005*, March 16-18 2005.

[94] Mobasher, A., Taherzadeh, M., Sotirov, R., and Khandani, A. K., "An Efficient Quasi-Maximum Likelihood Decoding for Finite Constellations," Tech. Rep. UW-E&CE 2005-01, Department of E&CE, University of Waterloo, 2005. Available via the WWW site at http://www.cst.uwaterloo.ca/~amin.

[95] Mobasher, A., Taherzadeh, M., Sotirov, R., and Khandani, A. K., "A near maximum likelihood decoding algorithm for mimo systems based on semi-definite programming," *to appear in IEEE Trans. on Info. Theory*, Nov. 2007.

[96] Mobasher, A. and Khandani, A. K., "PAPR Reduction in OFDM Systems Using Constellation Shaping," in 22^{nd} *Biennial Symposium on Comm.*, (Kingstone, ON, Canada), June 2004.

[97] Mobasher, A. and Khandani, A. K., "PAPR Reduction Using Integer Structures in OFDM Systems," in *IEEE Vehicular Technology Conference*, (Los Angeles, CA, USA), Sept. 2004.

[98] Mobasher, A. and Khandani, A. K., "Integer-Based Constellation Shaping Method for PAPR Reduction in OFDM Systems," *IEEE Trans. on Comm.*, vol. 54, pp. 119–127, Jan. 2006.

[99] Mobasher, A. and Khandani, A. K., "Matrix-Lifting Semi-Definite Programming for Decoding in Multiple Antenna Systems," in *the* 10^{th} *Canadian Workshop on Information Theory, CWIT'07*, (Edmonton, Alberta, Canada), June 2007.

[100] Mobasher, A. and Khandani, A. K., "Matrix-Lifting Semi-Definite Programming for Decoding in Multiple Antenna Systems," *Submitted to IEEE Trans. on Info. Theory*, 2007.

[101] Mobasher, A. and Khandani, A. K., "Precoding in Multiple-Antenna Broadcast Systems with a Probabilistic Viewpoint," in *the* 10^{th} *Canadian Workshop on Information Theory (CWIT'07)*, (Edmonton, Alberta, Canada), June 2007.

[102] Mobasher, A. and Khandani, A. K., "Probabilistic Behavior of Average Transmit Energy in Broadcast Systems with Precoding," *Submitted to IEEE Trans. on Info. Theory*, 2007.

[103] Mow, W. H., "Universal Lattice Decoding: Principle and Recent Advances," *Wireless Communications and Mobile Computing, Special Issue on Coding and Its Applications in Wireless CDMA Systems*, vol. 3, pp. 553–569, Aug. 2003.

[104] Muller, S. H. and Hubber, J. B., "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *Electron. Lett.*, vol. 33, pp. 368–369, Feb. 1997.

[105] Newman, M., "The Smith Normal Form," *Linear ALgebra Appl.*, vol. 254, pp. 367–381, 1997.

[106] Nikopour, H., Khandani, A. K., and Jamali, S. H., "Turbo Coded OFDM Transmission over Nonlinear Channel," tech. rep., University of Waterloo, 2004.

[107] Ochiai, H., "A Novel Trellis Shaping Design with Both Peak and Average Power Reduction for OFDM Systems," *Submitted to IEEE Trans. on Comm.*, 2003.

[108] Ochiai, H. and Imai, H., "On the Distribution of the Peak-to-Average Power Ratio in OFDM Signals," *IEEE Trans. on Commun.*, vol. 49, pp. 282–289, Feb. 2001.

[109] Ochiai, H. and Imai, H., "Performance Analysis of Deliberately Clipped OFDM Signals," *IEEE Trans. on Commun.*, vol. 50, pp. 89–101, Jan. 2002.

[110] Oggier, F. and Viterbo, E., *Algebraic number theory and code design for Rayleigh fading channels*, vol. 1 of *Foundations and Trends in Communications and Information Theory*. Now publishers, Dec. 2004.

[111] Pataki, G., “Algorithms for cone-optimization problems and semi-definite programming,” tech. rep., Graduate School of Industrial Administration, Carnegie Mellon University, 1994.

[112] Pataki, G., “On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalue,” *Mathematics of Operations Research*, vol. 23, pp. 339–358, 1998.

[113] Patterson, K., “Generalized Reed-Muller codes and power control in OFDM modulation,” *IEEE Trans. Inform. Theory*, vol. 46, pp. 104–120, Jan. 2000.

[114] Peel, C. B., Hochwald, B. M., and Swindlehurst, A. L., “A vector-perturbation technique for near-capacity multiple-antenna multi-user communications-Part I: Channel Inversion and Regularization,” *IEEE Trans. on Comm.*, vol. 53, Jan. 2005.

[115] Peel, C. B., Hochwald, B. M., and Swindlehurst, A. L., “A vector-perturbation technique for near-capacity multiple-antenna multi-user communications-Part II: Perturbation,” *IEEE Trans. on Comm.*, vol. 53, Mar. 2005.

[116] Saeedi, H., Sharif, M., and Marvasti, F., “Clipping noise cancellation in OFDM systems using oversampled signal reconstruction,” *IEEE Comm. Lett.*, vol. 6, pp. 73–75, Feb. 2002.

[117] Sato, H., “An outer bound on the capacity region of broadcast channels,” *IEEE Trans. Info Theory*, vol. 24, p. 374Ű377, May 1978.

[118] Schneider, K. S., “Optimum detection of code division multiplexed signals,” *IEEE Trans. Aerospace Elect. Sys.*, vol. AES-15, pp. 181–185, January 1979.

[119] Schnorr, C. P., “A hierarchy of polynomial time lattice basis reduction algorithms,” *Theoretical Computer Sci. 53*, pp. 201–224, 1987.

[120] Sharif, M., Gharavi-Alkhansari, M., and Khalaj, B. H., “On the peak-to-average power of OFDM signals based on oversampling,” *IEEE Trans. on Comm.*, vol. 51, pp. 72–78, Jan. 2003.

[121] Sidiropoulos, N. D. and Luo, Z.-Q., “A semidefinite relaxation approach to mimo detection for high-order qam constellations,” *IEEE SIGNAL PROCESSING LETTERS*, vol. 13, Sept. 2006.

[122] Smith, H. J. S., “On systems of linear indeterminate equations and congruences,” *Phil. Trans. Roy. Soc. London*, vol. 151, pp. 293–326, 1861.

[123] Sotirov, R. and Rendl, F., “Bounds for the Quadratic Assignment Problem Using the Bundle Method,” tech. rep., Department of Mathematics, University of Klagenfurt, Austria, 2003. Available at http://www.ms.unimelb.edu.au/~rsotirov.

[124] Steingrimsson, B., Luo, T., and Wong, K. M., “Soft quasi-maximum-likelihood detection for multiple-antenna wireless channels,” *IEEE Transactions on Signal Processing*, vol. 51, pp. 2710– 2719, Nov. 2003.

[125] Steingrimsson, B., Luo, Z. Q., and Wong;, K. M., "Quasi-ML detectors with soft output and low complexity for PSK modulated MIMO channels," in *4th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2003)*, pp. 427 – 431, 15-18 June 2003.

[126] Storjohann, A. and Labahn, G., "A Fast Las Vegas Algorithm for Computing the Smith Normal Form of a Polynomial Matrix," *Technical Report CS-94-43, University of Waterloo*, Nov. 1994.

[127] Sturm, J., "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, 1999.

[128] Sturm, J., "Implementation of interior point methods for mixed semidefinite and second order cone optimization problems," *Optimization Methods and Software*, vol. 17, no. 6, pp. 1105–1154, 2002.

[129] Taherzadeh, M., Mobasher, A., and Khandani, A. K., "Communication over MIMO broadcast channels using lattice-basis reduction," *to appear in IEEE Trans. on Info. Theory*, Dec. 2007.

[130] Taherzadeh, M., Mobasher, A., and Khandani, A. K., "LLL reduction achieves the receive diversity in mimo decoding," *to appear in IEEE Trans. on Info. Theory*, Dec. 2007.

[131] Taherzadeh, M., Mobasher, A., and Khandani, A. K., "Communication Over MIMO Broadcast Channels Using Lattice-Basis Reduction," in *the 42nd Annual Allerton Conference on Communication, Control, and Computing*, (Monticello, IL, USA), Oct. 2004.

[132] Taherzadeh, M., Mobasher, A., and Khandani, A. K., "Lattice-Basis Reduction Achieves the Precoding Diversity in MIMO Broadcast Systems," in *the* 39*th* *Conference on Information Sciences and Systems, CISS'05*, (Baltimore, MD, USA), Mar. 2005.

[133] Taherzadeh, M., Mobasher, A., and Khandani, A. K., "LLL Lattice-Basis Reduction Achieves Maximum Diversity In MIMO Systems," in *IEEE International Symposium on Information Theory, ISIT'05*, (Adelaide, Australia), Sept. 2005.

[134] Telatar, E., "Capacity of multi-antenna gaussian channels," *European Trans. on Telecomm. ETT*, vol. 10, pp. 585–596, November 1999.

[135] Tellado, J., *Peak to average power reduction for multicarrier modulation*. PhD thesis, Stanford University, Stanford CA, 2000.

[136] Tellambura, C., "Computation of the Continuous-Time PAR of an OFDM Signal with BPSK Subcarriers," *IEEE Communications Letters*, vol. 5, pp. 185–187, May 2001.

[137] Todd, M., "Semidefinite Optimization," *Acta Numerica*, vol. 10, pp. 515–560, 2001.

[138] Trees, H. V., *Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 1968.

[139] VAN NEE, R., VAN ZELST, A., and AWATER, G., "Maximum Likelihood Decoding in a Space Division Multiplexing System," *IEEE VTC2000*, vol. 45, pp. 6–10., July 1999.

[140] VISWANATH, P. and TSE, D., "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Info Theory*, pp. 1912–1921, August 2003.

[141] VITERBO, E. and BOUTROS, J., "A Universal Lattice Code Decoder for Fading Channels," *IEEE Transactions on Information Theory*, vol. 45, pp. 1639–1642, July 1999.

[142] WANG, R. and GIANNAKIS, G. B., "Approaching MIMO Capacity with Reduced Complexity Soft Sphere-Decoding," in *Wireless Comm. and Networking Conf*, (Atlanta, GA), March 21-25, 2004.

[143] WIESEL, A., ELDAR, Y. C., and SHAMAI, S., "Semidefinite relaxation for detection of 16-QAM signaling in mimo channels," *IEEE Signal Processing Letters*, vol. 12, Sept. 2005.

[144] WINDPASSINGER, C., FISCHER, R. F. H., and HUBER, J. B., "Lattice-reduction-aided broadcast precoding," in *5th International ITG Conference on Source and Channel Coding (SCC)*, (Erlangen, Germany), pp. 403–408, January 2004.

[145] WINDPASSINGER, C. and FISCHER, R. F. H., "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *ITW2003*, (Paris, France), March 31-April 4 2003.

[146] WOLKOWICZ, H., SAIGAL, R., and VANDENBERGHE, L., *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. Kluwer, 2000.

[147] WOLKOWICZ, H. and ZHAO, Q., "Semidefinite programming relaxations for the graph partitioning problem," *Discrete Applied Mathematics 96-97*, pp. 461–479, 1999.

[148] XIE, Z., SHORT, R. T., and RUSHFORTH, C. K., "A family of suboptimum detectors for coherent multi-user communications," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 683–690, May 1990.

[149] YANG, K. and CHANG, S., "Peak-to-Average Power Control in OFDM Using Standard Arrays of Linear Block Codes," *IEEE Commun. Lett.*, vol. 7, pp. 174–176, Apr. 2003.

[150] YANG, Y., ZHAO, C., ZHOU, P., and XU, W., "Mimo detection of 16-qam signaling based on semidefinite relaxation," *to appear in IEEE SIGNAL PROCESSING LETTERS*, 2007.

[151] YAO, H. and WORNELL, G. W., "Lattice-reduction-aided detectors for MIMO communication systems," in *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE*, vol. 1, pp. 424–428, 17-21 Nov 2002.

[152] YE, Y., *Interior Point Algorithms: Theory and Analysis*. Wiley Interscience series in discrete Mathematics and Optimization, John Wiley & Sons, 1997.

[153] YU, H. and WEI, G., "Computation of the continuous-time PAR of an OFDM signal," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP'03)*, (Hong Kong, China), pp. IV–529–31, IEEE, IEEE, 6-10 April 2003.

[154] Yu, W. and Cioffi, J., “Trellis precoding for the broadcast channel,” in *IEEE Global Telecommunications Conference (Globecom)*, vol. 2, pp. 1344–1348, 2001.

[155] Zamir, R., Shamai, S., and Erez, U., “Nested linear/lattice codes for structured multiterminal binning,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 1250–1276, June 2002.

[156] Zhao, Q., Karisch, S., Rendl, F., and Wolkowicz, H., “Semidefinite programming relaxation for the quadratic assignment problem,” *J. Combinatorial Optimization*, vol. 2, pp. 71–109, 1998.

INDEX

VITA

Amin Mobasher received his B.Sc. and M.Sc. degrees in electrical engineering from Sharif University of Technology (SUT), Tehran, Iran, in 2000 and 2002, respectively. He has been awarded his Ph.D. degree in electrical and computer engineering at the University of Waterloo, Waterloo, ON, Canada in Dec. 2007. His research interests are detection and decoding for multiple antenna systems, optimization in communication applications, and signal processing.

He is the recipient of several awards including University of Waterloo Doctoral Thesis Completion Award, NSERC Industrial R&D Fellowship, Ontario Graduate Scholarship (OGS), and President's Graduate Scholarship. He has been served as the webmaster and committee member in the student committee of IEEE Information Theory Society. He is also an active member in IEEE Kitchener-Waterloo section.