

Abstract

The thesis addresses the problem of space-time codebook design for communication in multiple-input multiple-output (MIMO) wireless systems. The realistic and challenging non-coherent setup (channel state information is absent at the receiver) is considered. A generalized likelihood ratio test (GLRT)-like detector is assumed at the receiver and contrary to most existing approaches, an arbitrary correlation structure is allowed for the additive Gaussian observation noise. A theoretical analysis of the probability of error is derived, for both the high and low signal-to-noise ratio (SNR) regimes. This leads to a codebook design criterion which shows that optimal codebooks correspond to optimal packings in a Cartesian product of projective spaces. The actual construction of the codebooks involves solving a high-dimensional, nonlinear, nonsmooth optimization problem which is tackled here in two phases: a convex semi-definite programming (SDP) relaxation furnishes an initial point which is then refined by an iterative subgradient-like geodesic descent algorithm exploiting the Riemannian geometry imposed by the power constraints on the space-time codewords. New codebooks are obtained by this method and their performance is shown to outperform previous state-of-art solutions. In fact, for some particular configurations, these new constellations attain the Rankin bound and are therefore provably optimal. The thesis also contains new theoretical results on the capacity (mutual information) of multiple-antenna wireless links in the low SNR regime. The impact of channel and noise correlation on the mutual information is obtained for the on-off and Gaussian signaling. The main conclusion is that mutual information is maximized when both the transmit and receive antennas are fully correlated.

Keywords: Multiple-input multiple-output (MIMO) systems, non-coherent communications, space-time constellations, Grassmannian packings, equiangular tight frame (ETF), channel capacity.

Resumo

A tese aborda o problema do desenho de códigos espaço-temporais para sistemas de comunicação Multiple-Input Multiple-Output (MIMO) sem fios. Considera-se o contexto realista e desafiante da recepção não-coerente (realização do canal é desconhecida no receptor). O detector conhecido como Generalized Likelihood Ratio Test (GLRT) é implementado no receptor e, ao contrário da maioria das abordagens actuais, permite-se uma estrutura de correlação arbitrária para o ruído Gaussiano de observação. Apresenta-se uma análise teórica para a probabilidade de erro do detector, em ambos os regimes assintóticos de relação sinal-ruído (SNR) alta e baixa. Essa análise conduz a um critério de optimalidade para desenho de códigos e permite uma re-interpretação geométrica como um problema de empacotamento óptimo num producto Cartesiano de espaço projectivos. A construção dos códigos implica a resolução de um problema de optimização não-linear, não-diferenciável e de dimensão elevada que foi atacado aqui em duas fases. A primeira fase explora uma relaxação convexa do problema original para obter uma estimativa inicial. A segunda fase, refina essa estimativa através de um algoritmo iterativo de descida de gradiente ao longo de geodésicas: explora-se assim a geometria Riemmaniana imposta pela restrições de potência sobre os códigos espaço-temporais. Mostra-se que o desempenho dos novos códigos obtidos por este método excede o das soluções previamente conhecidas. De facto, para algumas configurações particulares, estas novas constelações realizam o limiar de Rankin e são por isso garantidamente óptimas. Esta tese também contém novos resultados teóricos sobre a capacidade (informação mútua) de ligações sem-fios com múltiplas antenas no regime de baixa SNR. O impacto de correlação do canal e do ruído sobre a informação mútua é obtido para as sinalizações on-off e Gaussiana. A conclusão principal é que a informação mútua é maximizada quando ambas as antenas do transmissor e receptor estão totalmente correlacionadas.

Palavras-chave: MIMO, comunicação não-coerente, códigos espaço-temporais, variedades Grassmannianas, capacidade do canal.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. João Xavier, without him this thesis would have never been written. He has been the major source of input, constructive criticism and encouragement. His sharp mind, patience, interest and support have been helpful during the work on this thesis. I feel honored of becoming his first graduate student and am very grateful for the past four years.

I would like to express my gratitude to Profs. Victor Barroso and João Sentieiro for the invitation to come to Lisbon and for all friendly support and help they have given me all this time.

I thank many current and former students at Institute for Systems and Robotics (ISR) who helped to create a friendly and creative atmosphere, especially Nuno Silva, Nuno Órfão, Dejan Milutinovic, Paulo Lopes, Tiago Patrão, Tiago Barroso, João Leonardo, João Sousa, Sebastien Bausson, Augusto Santos, Cesaltina Ricardo, Pedro Pedrosa, and Profs. João Pedro Gomes, João Paulo Costeira and Rui Dinis.

I express gratitude to all my friends that helped make all these years very enjoyable. Particularly, I would like to mention Milan Jovicić, Vladimir Borčić, Slobodan Tanacković, Natasa Marjanović, Vesna Prosinecki and Svetislav Momcilović.

I gratefully acknowledge the financial support provided by the ISR, Instituto Superior Técnico, Lisbon, and by the Fundação para a Ciência e Tecnologia (FCT), grant SFRH/BD/12809/2003.

My greatest and heartfelt thanks must go to Biljana Mijatović and Ljuba Damjanović for their help at the beginning of my PhD.

Last, but not least, I would like to thank my family — my parents Vidosava and Mirko, my sister Marija and my brother Djordje — and Raquel for their love, encouragement and patience. Especially, I feel obliged to thank my parents for made me appreciate the importance of education. This thesis is dedicated to them.

Marko Beko

Lisboa, December 2007

Contents

Notation	xvii
Symbols	xvii
Acronyms	xix
1 Introduction	1
1.1 Non-Coherent MIMO Communications	4
1.2 Thesis Outline and Contributions	7
1.2.1 Chapter 2	8
1.2.2 Chapter 3	9
1.2.3 Chapter 4	10
2 Receiver Design and Codebook Construction in the High SNR Regime	11
2.1 Chapter Summary	11
2.2 Problem Formulation	11
2.2.1 A Note on Pre-Whitening	17
2.3 Considerations About the New Codebook Merit Function	22
2.3.1 Optimality of Unitary Codewords for the White Noise Case	22
2.3.2 Codebook Design as a Grassmannian Packing	26
2.3.3 Maximal Diversity Analysis	27
2.4 Codebook Construction	29
2.5 Results	36
2.6 Conclusions	42

3 Capacity and Error Probability Analysis of Non-Coherent MIMO Systems in the Low SNR Regime	47
3.1 Chapter Summary	47
3.2 Random Fading Channel: the Low SNR Mutual Information Analysis . . .	47
3.2.1 Mutual Information: On-Off Signaling	49
3.2.2 Mutual Information: Gaussian Modulation	52
3.3 Deterministic Fading Channel: the Low SNR PEP Analysis	54
3.3.1 Results	57
3.4 Conclusions	69
4 Conclusions and Future Work	73
4.1 Conclusions	73
4.2 Future Work	74
A Pairwise Error Probability for Fast Fading in the High SNR Regime	79
B Optimization Problem	83
C Calculating Gradients	85
D Mutual Information for On-Off Signaling in the Low SNR Regime	89
E Optimization Problem for On-Off Signalling	93
F Mutual Information for Gaussian Signaling in the Low SNR Regime	99
G Optimization Problem for Gaussian Signalling	103
H Pairwise Error Probability for Fast Fading in the Low SNR Regime	109

List of Figures

1.1	MIMO system	3
2.1	Geometrical interpretation of $\mathbf{\Pi}_j^\perp \mathcal{X}_i$	16
2.2	Phase 2: optimizes a non-smooth function on a manifold	35
2.3	Category 1 - spatio-temporally white observation noise: $T = 8, M = 3,$ $N = 1, K = 256, \mathbf{\Upsilon} = \mathbf{I}_{NT}$. Plus-solid curve:our codes; circle-dashed curve:unitary codes.	37
2.4	Category 1 - spatio-temporally white observation noise: $T = 8, M = 2,$ $N = 1, K = 256, \mathbf{\Upsilon} = \mathbf{I}_{NT}$. Plus-solid curve: our codes; circle-dashed curve: unitary codes.	40
2.5	Category 1 - spatio-temporally white observation noise: $T = 8, M = 1,$ $N = 1, K = 256, \mathbf{\Upsilon} = \mathbf{I}_{NT}$. Plus-solid curve: our codes; circle-dashed curve: unitary codes.	41
2.6	Category 1 - spatio-temporally white observation noise: $T = 4, M = 2,$ $N = 2, K = 256, \mathbf{\Upsilon} = \mathbf{I}_{NT}$. Plus-solid curve: our codes; circle-dashed curve: unitary codes found in [32]. The codes use GLRT receiver.	42
2.7	Category 2 - spatially white - temporally colored: $T = 8, M = 2, N =$ $1, K = 67, \boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.1; \text{zeros}(3,1)]$. Solid curves: our codes; dashed curves: unitary codes; dotted curve: codes obtained by the heuristic (2.12); dash-dotted curve: codes obtained by the heuristic (2.13); plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.	44

- 2.8 Category 2 - spatially white - temporally colored: $T = 8$, $M = 2$, $N = 1$, $K = 256$, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$. Solid curves: our codes; dashed curves: unitary codes; dotted curve: codes obtained by the heuristic (2.12); dash-dotted curve: codes obtained by the heuristic (2.13); plus signed curves: GLRT receiver; square signed curves: Bayesian receiver. . . . 44
- 2.9 Category 2 - spatially white - temporally colored: $T = 8$, $M = 2$, $N = 1$, $K = 32$, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$. Solid curves: our codes; dashed curves: unitary codes; dotted curve: codes obtained by the heuristic (2.12); dash-dotted curve: codes obtained by the heuristic (2.13); plus signed curves: GLRT receiver; square signed curves: Bayesian receiver. . . . 45
- 2.10 Category 2 - spatially white - temporally colored: $T = 4$, $M = 2$, $N = 2$, $K = 256$, $\boldsymbol{\rho} = [1; 0.8 ; 0.6; 0.1]$. Solid curves: our codes; dashed curves: unitary codes found in [32]; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver. 45
- 2.11 Category 3: $T = 8$, $M = 2$, $N = 2$, $K = 32$. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver. 46
- 2.12 Category 3: $T = 8$, $M = 2$, $N = 2$, $K = 67$. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver. 46
- 3.1 Category 1 - spatio-temporal white observation noise: Solid signed curve-our codes for $K = 16$, $T = 2$, $M = 1$, dashed signed curve-Borran's codes for $K = 16$, $T = 2$, $M = 1$, solid circled curve-our codes for $K = 8$, $T = 2$, $M = 1$, dashed circled curve-Borran's codes for $K = 8$, $T = 2$, $M = 1$ 60
- 3.2 Category 1 - spatio-temporal white observation noise: $T = 8$, $K = 256$, SNR = 0 dB. Solid curve-our codes for $M = 1$, dashed curve-our codes for $M = 2$, dash-dotted curve-our codes for $M = 3$. All codes use GLRT receiver. 61

- 3.3 Category 1 - spatio-temporal white observation noise: $T = 8$, SNR = -6 dB. Solid curve-our codes for $M = 1$ and $K = 32$, dashed curve-our codes for $M = 1$ and $K = 67$, solid-circled curve-our codes for $M = 2$ and $K = 32$, dashed-circled curve-our codes for $M = 2$ and $K = 67$. All codes use GLRT receiver. 62
- 3.4 Category 1 - spatio-temporal white observation noise: Solid signed curve-our codes for $K = 32$, $T = 4$, $M = 1$, dashed signed curve-Borran's codes for $K = 32$, $T = 4$, $M = 2$, solid circled curve-our codes for $K = 16$, $T = 3$, $M = 1$, dashed circled curve-Borran's codes for $K = 16$, $T = 3$, $M = 2$. . . 63
- 3.5 Category 1 - spatio-temporal white observation noise: $T = 2$, $P = 0.5$, rate = 1 b/s/Hz, $\mathbf{K}_r = \mathbf{I}_N$. Correlated Rayleigh fading: solid curve-our 5 point single beam constellation with unequal priors for $M = 3$ and $\text{rank}(\mathbf{K}_t)=1$, solid circled curve-Srinivasan's 5 point single beam constellation with unequal priors [47] for $M = 3$ and $\text{rank}(\mathbf{K}_t)=1$. Uncorrelated Rayleigh fading: dashed curve-our 5 point single beam constellation with unequal priors for $M = 3$ and $\mathbf{K}_t = \mathbf{I}_M$, dashed circled curve-Srinivasan's 5 point single beam constellation with unequal priors for $M = 3$ and $\mathbf{K}_t = \mathbf{I}_M$. Dash-dotted curve-our 4 point constellation for $M = 1$ with equal priors. Our and Srinivasan's 5 point constellations use *maximum a-posteriori* (MAP) receiver, our 4 point constellation uses GLRT receiver. 64
- 3.6 Category 2 - spatially white - temporally colored: $T = 8$, $K = 256$, SNR = -10 dB, $\boldsymbol{\rho}=[1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$. Solid curve-our codes for $M = 1$ adapted to $\boldsymbol{\rho}=[1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$, solid-circled curve-our codes for $M = 2$ adapted to $\boldsymbol{\rho}=[1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$, dashed curve-our codes for $M = 1$ adapted to $\boldsymbol{\rho}=[1; \text{zeros}(7,1)]$, dashed-circled curve-our codes for $M = 2$ adapted to $\boldsymbol{\rho}=[1; \text{zeros}(7,1)]$. All codes use GLRT receiver. 65

- 3.7 Category 2 - spatially white - temporally colored: $T = 8$, $K = 67$, SNR = -10 dB, $\rho = [1; 0.85; 0.6; 0.35; 0.1; \text{zeros}(3,1)]$. Solid curve-our codes for $M = 1$ adapted to $\rho = [1; 0.85; 0.6; 0.35; 0.1; \text{zeros}(3,1)]$, solid-circled curve-our codes for $M = 2$ adapted to $\rho = [1; 0.85; 0.6; 0.35; 0.1; \text{zeros}(3,1)]$, dashed curve-our codes for $M = 1$ adapted to $\rho = [1; \text{zeros}(7,1)]$, dashed-circled curve-our codes for $M = 2$ adapted to $\rho = [1; \text{zeros}(7,1)]$. All codes use GLRT receiver. 66
- 3.8 Category 2 - spatially white - temporally colored: $T = 8$, $K = 32$, SNR = -10 dB, $\rho = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$. Solid curve-our codes for $M = 1$ adapted to $\rho = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$, solid-circled curve-our codes for $M = 2$ adapted to $\rho = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$, dashed curve-our codes for $M = 1$ adapted to $\rho = [1; \text{zeros}(7,1)]$, dashed-circled curve-our codes for $M = 2$ adapted to $\rho = [1; \text{zeros}(7,1)]$. All codes use GLRT receiver. 67
- 3.9 Category 2 - spatially white - temporally colored: $T = 6$, SNR = -6 dB, $\rho = [1; 0.85; 0.6; 0.35; 0.1; 0]$. Solid, dash-dotted and dashed curve-our 8 point constellations that match the noise statistics for $M = 1$, $M = 2$ and $M = 3$, respectively. Plus-signed dotted curve-our 8 point constellation that is constructed for the spatio-temporal white noise case ($\Upsilon = \mathbf{I}_{TN}$) and $M = 1$. Plus-signed solid curve-our 17 point constellation that match the noise statistics and $M = 1$. Dashed-circled curve-our 17 point constellation that is constructed for $\Upsilon = \mathbf{I}_{TN}$ and $M = 1$. Our 8 point constellations use GLRT receiver, our 17 point constellations use MAP receiver. 68

- 3.10 Category 2 - spatially white - temporally colored: $\mathbf{K}_r = \mathbf{I}_N$. Correlated Rayleigh fading: dotted plus-signed curve-our 17 point single beam constellation with unequal priors for $M = 3$ and $\text{rank}(\mathbf{K}_t)=1$, dash-dotted curve-our 8 point single beam constellation with equal priors for $M = 3$ and $\text{rank}(\mathbf{K}_t)=1$. Uncorrelated Rayleigh fading: dashed curve-our 17 point single beam constellation with unequal priors for $M = 3$ and $\mathbf{K}_t = \mathbf{I}_M$, solid curve-our 8 point single beam constellation with equal priors for $M = 3$ and $\mathbf{K}_t = \mathbf{I}_M$. All codes use MAP receiver. 69
- 3.11 Category 3 - Solid curve-our codes for $M = 2$ adapted to colored noise, dashed curve-our codes for $M = 1$ adapted to colored noise, dash-dotted curve-our codes for $M = 1$ adapted to white noise. All codes use GLRT receiver. 70
- 3.12 Category 3 - Solid curve-our codes for $M = 2$ adapted to colored noise, dashed curve-our codes for $M = 1$ adapted to colored noise, dash-dotted curve-our codes for $M = 1$ adapted to white noise. All codes use GLRT receiver. 71
- 3.13 Category 3 - Solid-circled curve-our 17 point codes with unequal priors [47] adapted to colored noise, plus-signed solid curve-our 8 point codes with equal priors adapted to colored noise, solid curve-our 8 point codes with equal priors adapted to colored noise, dashed-circled curve-our 17 point codes with unequal priors adapted to white noise, plus-signed dashed curve-our 8 point codes with equal priors adapted to white noise, dashed curve-our 8 point codes with equal priors adapted to white noise. Circled, signed, and 8-point code curves use MAP, ML and GLRT receivers, respectively. 71

List of Tables

2.1	Codebook Design Algorithm	30
2.2	GDA Algorithm	34
2.3	PACKING IN COMPLEX PROJECTIVE SPACE: We compare our best configurations (MB) of K points in $\mathbb{P}^{T-1}(\mathbb{C})$ against the Tropp codes (JAT) and Rankin bound [28]. The packing radius of an ensemble is measured as the acute angle between the closest pair of lines. Minus sign symbol (-) means that no packing is available for specific pair (T, K)	38
2.4	PACKING IN COMPLEX PROJECTIVE SPACE: We compare our best configurations (MB) of K points in $\mathbb{P}^{T-1}(\mathbb{C})$ against the Tropp codes (JAT) and Rankin bound [28]. The packing radius of an ensemble is measured as the acute angle between the closest pair of lines.	39
2.5	PACKING IN COMPLEX PROJECTIVE SPACE: We compare our best configurations (MB) of K points in $\mathbb{P}^{T-1}(\mathbb{C})$ against Rankin bound [28]. The packing radius of an ensemble is measured as the acute angle between the closest pair of lines.	43

Notation

For reference purposes, some of the most common symbols and acronyms used throughout this thesis are listed below.

Symbols

$[A]_{i,j}$	the ij -th element of matrix A
A^*	the complex conjugate of A
A^T	the transpose of A
A^H	the conjugate transpose (Hermitian) of A
A^{-1}	the inverse of A
$A^{\frac{1}{2}}$	the Hermitian square root of the positive semidefinite matrix A
$\text{tr}(A)$	the trace of A
$\det(A)$	the determinant of A
$A \otimes B$	the Kronecker product of A and B
$A \odot B$	the Schur-Hadamard product of A and B
$\text{rank}(A)$	the rank of A
I_n	the $n \times n$ identity matrix
$\mathbf{0}_{m \times n}$	the $m \times n$ zero matrix
$\mathbf{1}_n$	the n -dimensional column vector with all entries equal to one
$\lambda_{\max}(A)$	the maximum eigenvalue of Hermitian A
$\lambda_{\min}(A)$	the minimum eigenvalue of Hermitian A
$\text{vec}(A)$	the vector obtained by stacking the columns of A on top of each other, from left to right

$\ \mathbf{A}\ $	the Frobenius norm; $\ \mathbf{A}\ = \sqrt{\text{tr}(\mathbf{A}^H \mathbf{A})}$
$\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	multivariate circularly symmetric, complex Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	real-valued Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$P(A)$	the probability of the event A
$\mathcal{Q}(x)$	the Gaussian \mathcal{Q} -function: if $x \sim \mathcal{N}(0, 1)$, then $P(x > t) = \mathcal{Q}(t)$
$p(\cdot)$	probability density function
$o(x)$	if $f(x) = o(x)$, then $\lim_{x \rightarrow 0} \frac{f(x)}{x} = 0$
$\Re\{\cdot\}$	the real part
$\Im\{\cdot\}$	the imaginary part
$\mathbf{A} \succeq \mathbf{B}$	the matrix $\mathbf{A} - \mathbf{B}$ is positive semidefinite
$\stackrel{d}{=}$	equality in distribution
\sim	distributed according to
$\log_a x$	the base- a logarithm of x ; when a is omitted it denotes the natural logarithm of x
\mathbb{C}	the set of complex numbers
\mathbb{C}^N	the set of N -dimensional complex vectors
$\mathbb{C}^{M \times N}$	the set of $M \times N$ complex matrices
$\text{diag}(\mathbf{A})$	the column vector obtained by extracting the diagonal elements of \mathbf{A}
$\mathbb{E}_p[\cdot]$	the expectation with respect to the probability density function p ; p omitted whenever no confusion can occur
i	the imaginary unit; $i = \sqrt{-1}$
\approx	approximately equal to
δ_{ij}	the Kronecker symbol; equal to one if $i = j$ and zero otherwise
\mathbb{R}	the set of real numbers
\mathbb{R}^N	the set of N -dimensional real vectors
$\mathbb{R}^{M \times N}$	the set of $M \times N$ real matrices

\mathbb{S}^{n-1} the unit sphere in \mathbb{R}^n

Acronyms

4G	fourth-generation
BER	bit error rate
BLAST	Bell labs layered space-time
CSI	channel state information
CDMA	code-division multiple access
DAB	digital audio broadcasting
dB	decibel
DVB	digital video broadcasting
ETF	equiangular tight frame
EVD	eigenvalue decomposition
GDA	geodesic descent algorithm
GLRT	generalized likelihood ratio test
iid	independent and identically distributed
ISI	intersymbol interference
LDC	linear dispersion codes
LMI	linear matrix inequality
MAN	metropolitan area network
MAP	maximum a posteriori
ML	maximum likelihood
MIMO	multiple-input multiple-output
MMSE	minimum mean-square error
OFDM	orthogonal frequency division multiplexing
PEP	pairwise error probability
pdf	probability density function
SDP	semi-definite programming

SER	symbol error rate
SISO	single-input single-output
SNR	signal-to-noise ratio
SIC	successive interference cancelation
SVD	singular value decomposition
STC	space-time codes
STBC	space-time block codes
STTC	space-time trellis codes
TDMA	time-division multiple-access
UB	union bound
WLAN	wireless local area network
w.l.o.g.	without loss of generality
w.r.t.	with respect to

Chapter 1

Introduction

The demand for mobile communications systems with high data rates and improved link quality for a variety of applications has dramatically increased in recent years. Although the benefits of using multi-antenna receivers have been known for a long time, the diversity and rate gains attainable using multiple antennas at both transmit and receive sides have been understood only recently. Winters [1] was among the first to prove that multiple-input multiple-output (MIMO) systems can provide a capacity increase. Paulraj and Kailath [2] demonstrated that the capacity of cellular code-division multiple access (CDMA) systems equipped with multiple antennas at both transmit and receive sides can increase considerably with respect to single-input single output (SISO) systems. Then Telatar [3] proved the fundamental results on the capacity of flat-fading MIMO channels. These results were independently derived and extended with practical considerations by Foschini et al. [4]. The main finding of these information theoretic analyzes was that at high signal-to-noise ratio (SNR) the capacity of multiple antenna channels increases linearly with the smaller of the number of transmit and receive antennas. This has led to a great deal of research on space-time codes (STC) to exploit both spatial and temporal diversity to maximize channel capacity. The key development of the STC was originally revealed in [5] in the form of trellis codes, which required a multidimensional Viterbi algorithm at the receiver for decoding. These codes, called space-time trellis codes (STTC), provide a diversity gain equal to the number of transmit antennas in addition to a coding gain that depends on the complexity of the code (i.e., number of states in the trellis) without any loss in bandwidth efficiency. When the number of antennas is fixed, the decoding

complexity of STTC (measured by the number of trellis states at the decoder) increases exponentially as a function of the diversity level and transmission rate. In addressing the issue of decoding complexity, Alamouti [6] discovered a remarkable space-time block coding scheme for transmission with two antennas. This scheme supports maximum likelihood (ML) detection based on linear processing and scalar detection at the receiver. The very simple structure and linear processing of the Alamouti construction makes it a very attractive scheme that is currently part of wideband CDMA and CDMA-2000 standards. Tarokh et al. [7], by using orthogonal designs to create analogs of the Alamouti codes for more than two transmit antennas, laid down the theory of the space-time block codes (STBC). Their aim was also ML decoding with only linear processing at the receiver, and this is the function of the orthogonal structure. As the number of transmit antennas increases, the data rate available with orthogonal designs becomes unattractive. Hence the recent focus on nonorthogonal linear codes designs such as linear dispersion codes (LDC) [8] and the Golden code [9]. Bell labs layered space-time (BLAST) codes [4] can be regarded as a special class of STBC where streams of independent data are transmitted over different antennas, thus maximizing the average data rate over the MIMO system. There are various layered space-time architectures depending upon whether error control coding is used or not and by the way modulated symbols are assigned to the transmit antennas. Such architectures include the vertical [10], horizontal [11], diagonal [11] and threaded layered space-time architectures [12]. In order to perform symbol detection, the receiver must unmix the channel, in one of several possible ways. The complexity of ML decoding can be high when many antennas or higher order modulations are used. Enhanced variants of this like sphere decoding [13] have been proposed. Another popular decoding strategy proposed along vertical BLAST is known as nulling and canceling which resembles the successive interference cancellation (SIC) proposed for multiuser detection in CDMA receivers [14]. While there are several receiver architectures that can support the full degrees of freedom of the channel, nulling and canceling in combination with minimum mean-square error (MMSE) estimation achieves capacity. See, e.g., [15, Chapter 8] for more details.

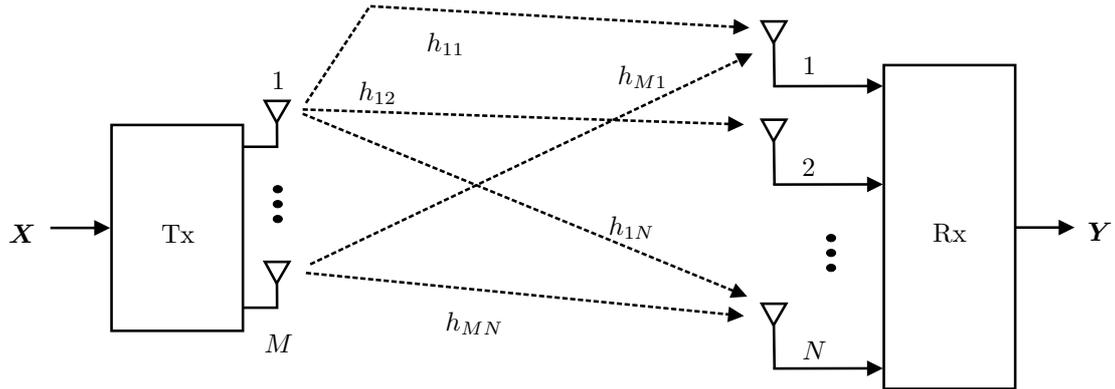


Figure 1.1: MIMO system

Since publication, space-time coding has quickly received recognition by academia, industry and governments. In the last few years, a large number of books and papers on space-time coding have been published worldwide. A detailed overview of some fundamental techniques, along with a survey of core contributions to the field and the generalizations to frequency-selective fading channels, can be found in [16, 17, 18, 19, 20, 21, 22].

The decoding of STC requires knowledge of the channel at the receiver. In slowly fading scenarios, when the fading channel coefficients remain approximately constant for many symbol intervals, channel stability enables the receiver to be trained (by sending training or pilot symbols or sequences) in order to acquire the channel state information (CSI). This is usually referred to as *coherent detection*. However, in fast fading scenarios, fading coefficients change into new, almost independent values before being learned by the receiver through training signals. Using multiple antennas at the transmitter increases the number of parameter to be estimated at the receiver which makes this problem more serious. This makes the *non-coherent* detection mode, where the receiver detects the transmitted symbols without having information about the realization of the channel, an attractive option for these fast fading scenarios.

In this thesis, we deal exclusively with non-coherent communication. In the sequel, a summary of the state-of-the-art in the non-coherent MIMO communications is provided.

1.1 Non-Coherent MIMO Communications

Previous work. The capacity of non-coherent multiple antenna systems was studied in [23, 24]. Under the additive white observation Gaussian noise and Rayleigh channel assumptions, it has been shown that the signal matrix that achieves capacity can be written as $\mathbf{S} = \mathbf{\Phi}\mathbf{V}$, where $\mathbf{\Phi}$ is an $T \times T$ isotropically distributed unitary matrix, and \mathbf{V} is an independent $T \times M$ real, nonnegative, diagonal matrix¹ with M and T denoting the number of transmit antennas and the length of the coherence interval, respectively; also, it has been proven that at high SNR, or when T is much bigger than M , capacity can be achieved by using a constellation of unitary matrices as codebooks. Furthermore, in [25] has been shown that, under the assumption of equal-energy codewords and high SNR, scaled unitary codebooks optimize the union bound (UB) on the error probability. Hence, at high SNR unitary constellations are optimal from both the capacity and symbol error probability viewpoints. Optimal unitary constellations correspond to optimal packings in Grassmann manifolds [26]. In [27, 28], a systematic method for designing unitary space-time constellations was presented. In [29], Sloane's algorithms [30] for producing sphere packings in real Grassmannian space have been extended to complex Grassmannian space. For a small number of transmit antennas, by using *chordal distance* as the design criterion, the corresponding constellations improve on the bit error rate (BER) when compared with the unitary space-time constellations presented in [27]. In [31] the problem of designing signal constellations for the multiple antenna non-coherent Rayleigh fading channel has been examined. The asymptotic UB on the probability of error has been considered, which, consequently, gave rise to a different notion of distance on the Grassmann manifold. By doing this, a method of iteratively designing signals, called successive updates, has been introduced. The signals obtained therein are, in contrast to [27, 29], guaranteed to achieve the full diversity order of the channel. In [32] a family of space-time codes suited for non-coherent MIMO systems was presented. These codes use all the degrees of freedom of the system, and they are constructed as codes on the Grassmann manifold

¹In calling the oblong matrix \mathbf{V} diagonal, it means that only the elements along its main diagonal may be nonzero.

by the exponential map. Recently, in [33, 34] some sub-optimal simplified decodings for the class of unitary space-time codes obtained via the exponential map were presented. In [37], the authors considered non-coherent communication over a frequency-flat MIMO Rayleigh block-fading channel. Using a subspace perturbation analysis, an appropriate metric for the distance between Grassmannian constellation points is determined, and a greedy technique for designing constellations that resemble the isotropic distribution is then proposed. The inherent geometric structure of these constellations is used to develop a novel suboptimum detector. The performance of this detector is comparable to that of the ML detector, but it requires less computational effort. An interested reader is referred to, e.g., [18, Chapter 10], for a summary of the most quoted propositions in the non-coherent MIMO literature.

Low SNR MIMO systems have recently attracted attention of scientific community. One of the reasons stems from the fact that in the third-generation mobile data systems almost 40% of geographical locations experience receiver SNR levels below 0 dB while only less than 10% display levels above 10 dB. High SNR requirement, besides its low power efficiency, cannot always be satisfied due to the power limitations in the mobile device. Also, recent technological advances have led to the emergence of small, low-power, and possibly mobile devices which, when deployed in large numbers, have the ability to form an intelligent (sensor) network which can monitor large areas, detect the presence or absence of targets, etc. This motivates the analysis and construction of communication schemes which can cope with the low SNR regime. See [38, 39, 40] for a more thorough discussion of this topic.

Low SNR MIMO systems when CSI is available at the receiver have been treated in [38]. The interplay of rate, bandwidth, and power is analyzed in the region of energy per bit close to its minimum value. The scenario where no CSI is available at the receiver has been considered in [41]. It has been shown that the optimal signaling at low SNR achieves the same capacity as the known channel case for single transmit antenna systems. Verdu, in [42], has shown that knowledge of the first and second derivatives of capacity at low SNR give us insight on bandwidth and energy efficiency for signal transmission. More precisely,

these quantities tell us how spectral efficiency grows with energy-per-bit. In [43], a formula for the second-order expansion of the input-output mutual information at low SNR is obtained, whereas in [44] the capacity and the reliability function as the peak constraint tends to zero are considered for a discrete-time memoryless channel with peak constrained inputs. Similar results to [43, 44] have been obtained in [45] under weaker assumptions on the input signals. In the same work, Rao and Hassibi have demonstrated that the on-off signaling presented in [41] generalizes to the multi-antenna setting and attains the known channel capacity. The tradeoff between communication rate and average probability of decoding error using a framework of error-exponent theory has been investigated in [46]. It is argued that the advantage of having multiple antennas is best realized when the fading is fully correlated, i.e., a performance gain of MN and a peakiness gain of M^2NT can be achieved where M , N and T represent the number of transmit, receive antennas, and the length of the coherence interval, respectively. The symbol error probability point of view for the analysis of low SNR non-coherent independent and identically distributed (iid) Rayleigh channel is more recent, although, Hochwald, et al. [27] had reported that in the low SNR and Rayleigh fading channel it seems one should employ only one transmit antenna. Borran et al. [39], under the assumption of equally probable codewords, presented a technique that uses Kullback-Liebler divergence between the probability density functions induced at the receiver by distinct transmitted codewords as a design criterion for codebook design. In low SNR condition, their constellation points occupy multiple level (signal points lie in concentric spheres) with a point usually in the origin. The codes thereby constructed were shown to perform better than some existing non-coherent codebook constructions in low SNR, namely [27]. Srinivasan, et al. [47], considered the case of single transmit antenna in the low SNR regime. Using the information theoretic results over the low SNR non-coherent iid Rayleigh fading channel under an average power constraint (c.f. [45, 46]), they allow for codewords with unequal priors in a code and optimize over prior probabilities to achieve better performance. This results in constellations that assume a point in the origin with probability $\frac{1}{2}$, with the probabilities of the points lying in the sphere being equal. In [48], the correlated Rayleigh fading model was studied and it

was shown that at any SNR, any single antenna performs better when used with suitable precoding in a MIMO correlated Rayleigh fading than in a single-input multiple output SIMO channel. Consequently, code designs that exploit the correlations in the transmit antennas in the MIMO case to provide gains over the corresponding SIMO case in the low SNR regime were presented.

In the sequel, the motivation for the research presented in this thesis is discussed.

Motivation. The techniques aforementioned can not be readily extended to the more realistic and challenging scenario, where the Gaussian observation noise has an arbitrary correlation structure. The assumption of spatio-temporal Gaussian observation noise is common, as there are at least two reasons for making it. First, it yields mathematical expressions that are relatively easy to deal with. Second, in some scenarios it can be justified via the central limit theorem. Although customary, the assumption of spatio-temporal *white* Gaussian observation noise is clearly an approximation. In general, in realistic scenarios, the noise term might have very rich correlation structure, e.g, see pp. 554 in [15], pp. 100 in [18], pp. 10, 159, 171 in [19] and [38]. The generalization to arbitrary noise covariance matrices encompasses many scenarios of interest as special cases: spatially colored or not jointly with temporally colored or not observation noise, multiuser environment, etc. Intuitively, unitary space-time constellations are not the optimal ones for this scenario.

1.2 Thesis Outline and Contributions

The thesis is divided into 4 chapters. We summarize the content of each chapter, besides the current one which gives the motivation and outline of this dissertation. For each chapter, we also refer the publications (conference and journal papers) that it has given rise.

In more detail, the outline of this thesis is as follows.

1.2.1 Chapter 2

In Chapter 2, the problem of space-time codebook design for non-coherent communications in multiple-antenna wireless systems and high SNR regime is addressed. In this work, we look for a more practical code design criterion based on error probability, rather than capacity analysis. The calculus of the exact expression for the average error probability for the general non-coherent systems seems not to be tractable. Instead, we consider pairwise error probability (PEP) in high SNR regime, and use it to find a code design criterion (a merit function) for an arbitrarily given noise correlation structure.

Contribution. Our contributions in this area are summarized in the following:

1. The main contribution of this chapter is a new technique that systematically designs space-time codebooks for non-coherent multiple-antenna communication systems. In contrast with other approaches, the channel matrix is modeled as an unknown deterministic parameter at both the receiver and the transmitter, and the Gaussian observation noise is allowed to have an arbitrary correlation structure, known by the transmitter and the receiver. In general correlated noise environments, computer simulations show that the space-time codes obtained with our method significantly outperform those already known which were constructed for spatio-temporally white noise case. We recall that codebook constructions for arbitrary noise correlation structures were not previously available and this demonstrates the interest of the codebook design methodology introduced herein.
2. For the special case of spatio-temporal white observation noise, our codebooks recover the previously known unitary structure, namely the codes in [27] (in fact, our codes are marginally better). Also, for this specific scenario and $M = 1$ we show that the problem of finding good codes coincides with the very well known packing problem in the complex projective space. We compare our best configurations against the codes in [28] and the Rankin bound. We manage to improve the best known results and in some cases actually provide optimal packings in complex projective spaces which attain the Rankin upper bound.

3. Theoretical analysis leading to an upper bound on PEP in the high SNR scenario for the Gaussian observation noise with an arbitrary correlation structure.

Publications. The results of this work have been published in

M. Beko, J. Xavier and V. Barroso, “Codebook design for non-coherent communication in multiple-antenna systems,” in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France, 2006.

as well in the form of a journal paper in

M. Beko, J. Xavier and V. Barroso, “Non-coherent Communication in Multiple-Antenna Systems: Receiver design and Codebook construction,” *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5703 - 5715, Dec. 2007.

1.2.2 Chapter 3

In Chapter 3, we study the non-coherent MIMO channel in the low SNR regime from the capacity and PEP viewpoints. The novel aspect is that we allow the Gaussian observation noise to have an arbitrary correlation structure, albeit known to the transmitter and the receiver.

Contribution. In the following, we summarize our contributions in this area:

1. The spatially correlated non-coherent MIMO block Rayleigh fading channel is analyzed. This extends the approach in [45] as we take into account both channel and noise correlation. The impact of channel and noise correlation on the mutual information is obtained for the on-off and Gaussian signaling. The main conclusion is that mutual information is maximized when both the transmit and receive antennas are fully correlated. This shows that MIMO systems can actually be beneficial in the low SNR regime. We also argue that the on-off signaling is optimal for this multi-antenna setting.
2. Contrary to most approaches for the low SNR regime, the channel matrix is assumed deterministic, i.e., no stochastic model is attached to it. A low SNR analysis

of the PEP for the GLRT receiver is introduced, and a codebook design criterion which takes into account the information about noise correlation is obtained. For the special case of single transmit antenna and spatio-temporal white Gaussian noise, it is shown that the problem of finding good codes corresponds to the very well known packing problem in the complex projective space [28]. New space-time constellations for some particular wireless scenarios are constructed. We argue that one should construct codebooks for just one transmit antenna that match the noise statistics. Computer simulations show that these new codebooks are also of interest for Bayesian receivers which decode constellations with non-uniform priors.

Publications. The material contained in this chapter has been published in

M. Beko, J. Xavier and V. Barroso, “Codebook design for the non-coherent GLRT receiver and low SNR MIMO block fading channel,” in *Proc. of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Cannes, France, 2006.

M. Beko, J. Xavier and V. Barroso, “Capacity and error probability analysis of non-coherent MIMO systems in the low SNR regime,” in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, HI USA, 2007.

In addition, a journal paper extending the previous results was submitted as

M. Beko, J. Xavier and V. Barroso, “Further results on the capacity and error probability analysis of non-coherent MIMO systems in the low SNR regime,” accepted for publication in *IEEE Transactions on Signal Processing*.

1.2.3 Chapter 4

This chapter concludes the thesis summarizing the main obtained results and enumerating the future lines of work.

Chapter 2

Receiver Design and Codebook Construction in the High SNR Regime

2.1 Chapter Summary

The chapter is organized as follows. In section 2.2, we introduce the data model and formulate the problem addressed in this chapter. We describe the structure of our non-coherent receiver and discuss the selection of the codebook design criterion. In section 2.3, before addressing the codebook design problem we draw some conclusions about the design criterion defined in section 2.2. In section 2.4, we propose a new algorithm that systematically designs non-coherent space-time constellations for an arbitrarily given noise covariance matrix and any M , N , K and T , respectively, number of transmitter antennas, number of receiver antennas, size of codebook, and channel coherence interval. In section 2.5, we present codebook constructions for several important special cases and compare their performance with state-of-art solutions. Section 2.6 presents the main conclusions of this chapter.

2.2 Problem Formulation

Data model and assumptions. The communication system comprises M transmit and N receive antennas and we assume a block flat fading channel model with coherence interval T . That is, we assume that the fading coefficients remain constant during blocks of T consecutive symbol intervals, and change into new, independent values at the end

of each block. It is an accurate representation of many time-division multiple-access (TDMA), frequency-hopping, or block-interleaved systems. See, e.g., [27] for more details. In complex base band notation we have the model

$$\mathbf{Y} = \mathbf{X}\mathbf{H}^H + \mathbf{E}, \quad (2.1)$$

where \mathbf{X} is the $T \times M$ matrix of transmitted symbols (the matrix \mathbf{X} is called hereafter a space-time codeword), \mathbf{Y} is the $T \times N$ matrix of received symbols, \mathbf{H}^H is the $M \times N$ matrix of channel coefficients (the operator H is used for the sake of convenience), and \mathbf{E} is the $T \times N$ matrix of zero-mean additive observation noise. In \mathbf{Y} , time indexes the rows and space (receive antennas) indexes the columns. We shall work under the following assumptions:

- A1. (Channel matrix)** The matrix \mathbf{H} is not known at the receiver neither at the transmitter, and no stochastic model is assumed for it;
- A2. (Transmit power constraint)** The codeword \mathbf{X} is chosen from a finite codebook $\mathcal{C} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$ known to the receiver, where K is the size of the codebook. We impose the power constraint $\text{tr}(\mathbf{X}_k^H \mathbf{X}_k) = 1$ for each codeword. Furthermore, we assume that $T \geq M$ and each codeword is of full rank, i.e., $\text{rank}(\mathbf{X}) = M$;
- A3. (Noise distribution)** The observation noise at the receiver is zero mean and obeys circular complex Gaussian statistics, that is, $\text{vec}(\mathbf{E}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Upsilon})$. The noise covariance matrix $\mathbf{\Upsilon} = \text{E}[\text{vec}(\mathbf{E})\text{vec}(\mathbf{E})^H]$ is known at the transmitter and at the receiver ($\text{vec}(\mathbf{E})$ stacks all columns of the matrix \mathbf{E} on the top of each other, from left to right).

Remark that in assumption **A3**, we let the data model depart from the customary assumption of spatio-temporal *white* Gaussian observation noise. Also, note that one cannot perform “pre-whitening” in order to revert the colored case ($\mathbf{\Upsilon} \neq \mathbf{I}_{TN}$) into the spatio-temporal white noise case ($\mathbf{\Upsilon} = \mathbf{I}_{TN}$). To see this, let’s consider two systems where system 1 is described by

$$\mathbf{Y}_1 = \mathbf{X}_1\mathbf{H}^H + \mathbf{E}_1, \quad (2.2)$$

with $\mathbf{e}_1 = \text{vec}(\mathbf{E}_1) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Upsilon})$, and system 2 is given by

$$\mathbf{Y}_2 = \mathbf{X}_2 \mathbf{H}^H + \mathbf{E}_2, \quad (2.3)$$

with $\mathbf{e}_2 = \text{vec}(\mathbf{E}_2) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{TN})$. The systems (2.2) and (2.3) are equivalent to

$$\mathbf{y}_1 = \text{vec}(\mathbf{Y}_1) = (\mathbf{I}_N \otimes \mathbf{X}_1) \text{vec}(\mathbf{H}^H) + \mathbf{e}_1, \quad (2.4)$$

$$\mathbf{y}_2 = \text{vec}(\mathbf{Y}_2) = (\mathbf{I}_N \otimes \mathbf{X}_2) \text{vec}(\mathbf{H}^H) + \mathbf{e}_2 \quad (2.5)$$

respectively (the symbol \otimes denotes the Kronecker product). After pre-whitening, from (2.4) we get

$$\widetilde{\mathbf{y}}_1 = \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{y}_1 = \mathbf{\Upsilon}^{-\frac{1}{2}} (\mathbf{I}_N \otimes \mathbf{X}_1) \text{vec}(\mathbf{H}^H) + \widetilde{\mathbf{e}}_1 \quad (2.6)$$

with $\widetilde{\mathbf{e}}_1 = \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{e}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{TN})$. From (2.5) and (2.6) we deduce that the systems 1 and 2 are not equivalent, i.e., the unitary constellations (which are optimal for spatio-temporally white noise at high SNR) cannot be employed by performing suitable pre-whitening because it breaks down the structure of the constellation. A more detailed discussion on this point can be found in subsection 2.2.1 .

Receiver. According to the system model (2.1) and the assumptions above mentioned, the conditional probability density function (pdf) of the received vector $\mathbf{y} = \text{vec}(\mathbf{Y})$, given the transmitted matrix \mathbf{X} and the unknown realization of the channel $\mathbf{g} = \text{vec}(\mathbf{H}^H)$, is given by

$$p(\mathbf{y}|\mathbf{X}, \mathbf{g}) = \frac{\exp\{-\|\mathbf{y} - (\mathbf{I}_N \otimes \mathbf{X})\mathbf{g}\|_{\mathbf{\Upsilon}^{-1}}^2\}}{\pi^{TN} \det(\mathbf{\Upsilon})},$$

where the notation $\|\mathbf{z}\|_{\mathbf{A}}^2 = \mathbf{z}^H \mathbf{A} \mathbf{z}$ was used.

Since no stochastic model is attached to the channel propagation matrix, the receiver faces a multiple hypothesis testing problem with the channel \mathbf{H} as a deterministic nuisance parameter. We assume a generalized likelihood ratio test (GLRT) receiver which decides the index k of the codeword as the index \widehat{k} such that

$$\begin{aligned} \widehat{k} &= \underset{k=1,2,\dots,K}{\text{argmax}} \quad p(\mathbf{y}|\mathbf{X}_k, \widehat{\mathbf{g}}_k) \\ &= \underset{k=1,2,\dots,K}{\text{argmin}} \quad \left\| \mathbf{y} - \widetilde{\mathbf{X}}_k \widehat{\mathbf{g}}_k \right\|_{\mathbf{\Upsilon}^{-1}}^2 \end{aligned}$$

where

$$\widehat{\mathbf{X}}_k = \mathbf{I}_N \otimes \mathbf{X}_k \quad \text{and} \quad \widehat{\mathbf{g}}_k = (\mathbf{x}_k^H \mathbf{x}_k)^{-1} \mathbf{x}_k^H \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{y} \quad (2.7)$$

with $\mathbf{x}_k = \mathbf{\Upsilon}^{-\frac{1}{2}} \widehat{\mathbf{X}}_k$ denoting the whitened version of $\widehat{\mathbf{X}}_k$. The GLRT [49, 50, 51] is composed of a bank of K parallel processors where the k -th processor assumes the presence of the k -th codeword and computes the likelihood of the observation, after replacing the channel by its ML estimate. The GLRT detector chooses the codeword associated with the processor exhibiting the largest likelihood of the observation. We note that the GLRT performs sub-optimally when compared with the ML receiver, as the latter can exploit the knowledge of channel statistics'. However, since assumption **A1** is in force, the GLRT yields an attractive (implementable) solution in the present setup. Note also that, for the special case of unitary constellations, i.e., $\mathbf{X}_k^H \mathbf{X}_k = \frac{1}{M} \mathbf{I}_M$ for all k , spatio-temporal white Gaussian noise and iid Rayleigh fading, it is readily shown that the two receivers coincide. Due to the respective expression for the ML estimate of the channel, equation (2.7), we note that since each codeword of the codebook has full rank (assumption **A2**), the channel estimate is well defined.

Codebook design criterion. In this chapter, our goal is to design a codebook $\mathcal{C} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$ of size K for the current setup. A codebook \mathcal{C} is a point in the space

$$\mathcal{M} = \{(\mathbf{X}_1, \dots, \mathbf{X}_K) : \text{tr}(\mathbf{X}_k^H \mathbf{X}_k) = 1\}.$$

Note that the space \mathcal{M} can be viewed as multi-dimensional torus, i.e, the Cartesian product of K unit-spheres :

$$\mathcal{M} = \mathbb{S}^{2TM-1} \times \dots \times \mathbb{S}^{2TM-1} \quad (K \text{ times})$$

and each codeword \mathbf{X}_k belongs to $\mathbb{C}^{T \times M}$. The symbol \mathbb{S}^{n-1} denotes the unit sphere in \mathbb{R}^n . First, we must adopt a merit function $f : \mathcal{M} \rightarrow \mathbb{R}$ which gauges the quality of each constellation \mathcal{C} . The average error probability for a specific \mathcal{C} would be the natural choice, but the theoretical analysis seems to be intractable. Instead, as usual [24]- [27], we rely on a PEP study to construct our merit function. For the special case of unitary codebooks ($\mathbf{X}_k^H \mathbf{X}_k = \frac{1}{M} \mathbf{I}_M$), spatio-temporal white Gaussian noise ($\mathbf{\Upsilon} = \mathbf{I}_{TN}$) and iid Rayleigh

fading, the exact expression and Chernoff upper bound for the PEP have been derived in [24]. However, the calculus of these expressions for the general case, i.e., arbitrary matrix constellations \mathcal{C} and noise correlation matrix $\mathbf{\Upsilon}$, seems to be burdensome. As in [24]- [27], in this chapter, we focus on the high SNR regime. Namely, we resort to the asymptotic expression of the PEP in the high SNR regime, for arbitrary \mathcal{C} and $\mathbf{\Upsilon}$. To start the PEP analysis, we consider a codebook with only two codewords, i.e., $\mathcal{C} = \{\mathbf{X}_1, \mathbf{X}_2\}$. Let $P_{\mathbf{X}_i \rightarrow \mathbf{X}_j}$ be the probability of the GLRT receiver deciding \mathbf{X}_j when \mathbf{X}_i is sent. It can be shown (see Appendix A) that at sufficiently high SNR we have the approximation

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \approx \mathcal{Q}\left(\frac{1}{\sqrt{2}} \sqrt{\mathbf{g}^H \mathbf{L}_{ij} \mathbf{g}}\right), \quad (2.8)$$

with

$$\mathbf{L}_{ij} = \mathbf{x}_i^H \mathbf{\Pi}_j^\perp \mathbf{x}_i \quad \text{and} \quad \mathbf{\Pi}_j^\perp = \mathbf{I}_{TN} - \mathbf{x}_j (\mathbf{x}_j^H \mathbf{x}_j)^{-1} \mathbf{x}_j^H$$

where $\mathcal{Q}(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ is the \mathcal{Q} -function and $\mathbf{\Pi}_j^\perp$ is the orthogonal projector onto the orthogonal complement of the column space of \mathbf{x}_j .

Equation (2.8) shows that the probability of misdetecting \mathbf{X}_i for \mathbf{X}_j , depends on the channel realization $\mathbf{g} = \text{vec}(\mathbf{H}^H)$ and on the relative geometry of the codewords \mathbf{x}_i and \mathbf{x}_j . We can decouple the action of \mathbf{g} and \mathbf{L}_{ij} as follows: using the inequality

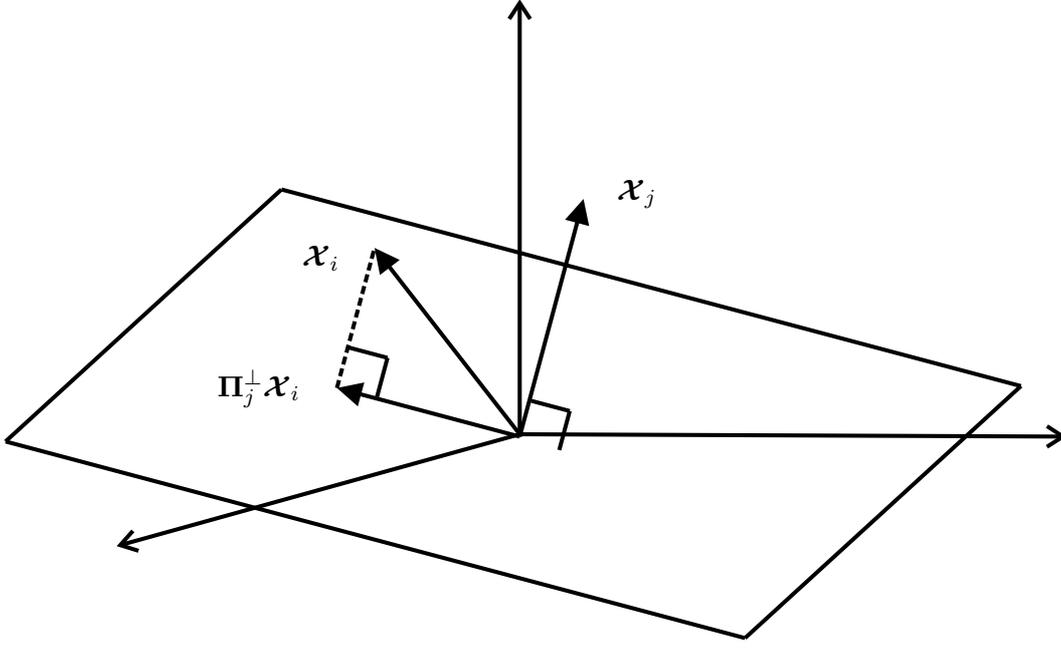
$$\mathbf{g}^H \mathbf{L}_{ij} \mathbf{g} \geq \lambda_{\min}(\mathbf{L}_{ij}) \|\mathbf{g}\|^2,$$

which is an equality when $M = 1$ and $\mathbf{\Upsilon} = \mathbf{I}_{NT}$, and the fact that $\mathcal{Q}(\cdot)$ is monotonically non-increasing, we have the upper bound on the PEP for high SNR

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \leq \mathcal{Q}\left(\frac{1}{\sqrt{2}} \|\mathbf{g}\| \sqrt{\lambda_{\min}(\mathbf{L}_{ij})}\right). \quad (2.9)$$

We cannot control the power of the channel $\mathbf{g} = \text{vec}(\mathbf{H}^H)$, but we can design codebooks aiming at maximizing $\lambda_{\min}(\mathbf{L}_{ij})$. We see that, if $M = 1$ and $\mathbf{\Upsilon} = \mathbf{I}_{NT}$, the bound in (2.9) is attained for arbitrary \mathcal{C} .

Geometrical interpretation. This latter objective has a clear geometric interpretation. Define $\mathbf{V} = \mathbf{\Pi}_j^\perp \mathbf{x}_i$. Then $\mathbf{\Pi}_j^\perp \mathbf{x}_i$ is the orthogonal projection of \mathbf{x}_i onto the orthogonal complement of $\text{span}\{\mathbf{x}_j\}$ (the span of $\mathbf{x}_j : TN \times MN$ is the linear subspace in \mathcal{C}^{TN} spanned by its MN columns, or, equivalently, the range of the matrix \mathbf{x}_j), see figure 2.1.

Figure 2.1: Geometrical interpretation of $\Pi_j^\perp \mathbf{x}_i$.

Now, note that

$$\mathbf{L}_{ij} = \mathbf{V}^H \mathbf{V} = (\Pi_j^\perp \mathbf{x}_i)^H (\Pi_j^\perp \mathbf{x}_i)$$

is the corresponding Gram matrix and

$$\sqrt{\det(\mathbf{V}^H \mathbf{V})} = \sqrt{\lambda_{\min}(\mathbf{V}^H \mathbf{V}) \cdot \dots \cdot \lambda_{\max}(\mathbf{V}^H \mathbf{V})} \geq \lambda_{\min}(\mathbf{V}^H \mathbf{V})^{\frac{MN}{2}}.$$

Hence, by maximizing $\lambda_{\min}(\mathbf{V}^H \mathbf{V})$, we are increasing a lower bound on $\sqrt{\det(\mathbf{V}^H \mathbf{V})}$ which is proportional to the volume of the parallelepiped spanned by the columns of the $\Pi_j^\perp \mathbf{x}_i$. That is, we are trying to place \mathbf{x}_i in the orthogonal complement of $\text{span}\{\mathbf{x}_j\}$.

Problem formulation. Following a worst-case approach, we are led from (2.9) to define the codebook merit function

$$f : \mathcal{M} \rightarrow \mathbb{R} \quad \text{and} \quad \mathcal{C} = \{\mathbf{X}_1, \dots, \mathbf{X}_K\} \mapsto f(\mathcal{C})$$

as

$$f(\mathcal{C}) = \min\{f_{ij}(\mathcal{C}) : 1 \leq i \neq j \leq K\} \quad (2.10)$$

where $f_{ij}(\mathcal{C}) = \lambda_{\min}(\mathbf{L}_{ij}(\mathcal{C}))$. Constructing an optimal codebook $\mathcal{C} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$ amounts to solving the optimization problem

$$\mathcal{C}^* = \arg \max_{\mathcal{C} \in \mathcal{M}} f(\mathcal{C}). \quad (2.11)$$

The problem defined in (2.11) is a high-dimensional, non-linear and non-smooth optimization problem. As an example, for a codebook of size $K = 256$ the number of f_{ij} functions is $K(K - 1) = 65280$. Also, for $T = 8$ and $M = 2$, there are $2KTM = 8192$ real variables to optimize.

The problem in (2.11) is a non-smooth optimization problem because the objective function f , as the pointwise minimum of several f_{ij} 's, is in general not smooth at points where the minimum is attained by several f_{ij} 's. In our case, each f_{ij} is not even smooth, due to the λ_{\min} operator. For an illustrative example, consider $\phi : \mathbb{R} \rightarrow \mathbb{R}$,

$$\phi(t) = \lambda_{\min} \left(\begin{bmatrix} t & 0 \\ 0 & -t \end{bmatrix} \right).$$

Although the matrix involved is a smooth function of its entries, $\phi(t) = -|t|$ is not smooth everywhere. Moreover, note that we have

$$f(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K) = f(\mathbf{X}_1 e^{i\theta_1}, \mathbf{X}_2 e^{i\theta_2}, \dots, \mathbf{X}_K e^{i\theta_K})$$

for any $\theta_k \in \mathbb{R}$ and $k = 1, \dots, K$. This means that f depends on each \mathbf{X}_k ($\|\mathbf{X}_k\| = 1$) only through the line spanned by it (i.e., $\{\lambda \mathbf{X}_k : \lambda \in \mathbb{C}\}$).

2.2.1 A Note on Pre-Whitening

It may not be immediately obvious why the ‘‘pre-whitening’’ device cannot be employed here. After all, this is a common trick in signal processing for generalizing solutions formulated for white noise to the colored noise setup. However, since it cannot be done in our situation, in the sequel, we try to provide more detailed explanations.

(i) As pointed out, ‘‘pre-whitening’’ cannot be performed as it changes the constellation’s structure. Essentially, this summarizes our argument in equations (2.2)-(2.6). We will now furnish another viewpoint on this matter.

When one asks if “pre-whitening” can be performed, one is asking if the solution to the white noise case can be “transported” to the colored case. That is, one is asking if the codebook construction problem for the colored case has an equivalent reformulation in the white noise setup. Of course, if this were true, then it would be sufficient to have a tool to construct codebooks for the white noise case, say, our tool solving the optimization problem stated in equation (2.11) for the white noise case. We argue that, in our situation, we cannot find such an equivalent reformulation.

To demonstrate our claim, consider the special case $M = 1$ (single transmit antenna), $N = 1$ (single receive antenna) and $T \geq 2$. As mentioned above, suppose that a tool is available to solve problem (2.11) for the white noise case $\mathbf{\Upsilon} = \mathbf{I}_T$. Denoting codebooks by

$$\mathbf{C} = [\mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_K] : T \times K$$

(each \mathbf{x}_k is a codeword) this means precisely that we have a tool to solve the optimization problem (for any chosen T and K)

$$P_1 : \quad \max_{\text{diag}(\mathbf{C}^H \mathbf{C}) = \mathbf{1}_T} f_{\mathbf{I}_T}(\mathbf{C})$$

where $\mathbf{1}_T = (1, 1, \dots, 1)^T$ (T times), $\text{diag}(\mathbf{A})$ extracts the diagonal of matrix \mathbf{A} and, for any positive-definite $T \times T$ matrix $\mathbf{\Sigma}$, we use the notation

$$f_{\mathbf{\Sigma}}(\mathbf{C}) = \min \left\{ \mathbf{x}_i^H \mathbf{\Sigma}^{-1} \mathbf{x}_i - \mathbf{x}_i^H \mathbf{\Sigma}^{-1} \mathbf{x}_j (\mathbf{x}_j^H \mathbf{\Sigma}^{-1} \mathbf{x}_j)^{-1} \mathbf{x}_j^H \mathbf{\Sigma}^{-1} \mathbf{x}_i : i \neq j \right\}.$$

Now, solving the codebook construction problem (2.11) for a general noise correlation matrix $\mathbf{\Upsilon}$ corresponds to solving the optimization problem

$$P_2 : \quad \max_{\text{diag}(\mathbf{C}^H \mathbf{C}) = \mathbf{1}_T} f_{\mathbf{\Upsilon}}(\mathbf{C}).$$

We now try reformulate problem P_2 into the format P_1 . We start by noticing that we have the identity

$$f_{\mathbf{\Upsilon}}(\mathbf{C}) = f_{\mathbf{I}_T}(\mathbf{\Upsilon}^{-1/2} \mathbf{C}).$$

Thus, problem P_2 is equivalent, through the change-of-variables $\mathbf{D} = \mathbf{\Upsilon}^{-1/2}\mathbf{C}$ (“pre-whitening”), to the problem

$$P_3 : \quad \max_{\text{diag}(\mathbf{D}^H \mathbf{\Upsilon} \mathbf{C}) = \mathbf{1}_T} f_{I_T}(\mathbf{D}).$$

The solution of problem P_3 cannot be extracted from the one in problem P_1 . Note that both problems have the same objective function but, whereas problem P_1 searches the codewords over a unit-sphere, problem P_3 searches them over an ellipsoid. This is a different problem.

Just as a side remark, note that a similar phenomenon (problems became inequivalent as a unit-sphere constraint is changed to an ellipsoidal one) occurs when one tries to project a point $x_0 \in \mathbb{R}^n$ onto a sphere or onto an ellipsoid, i.e.,

$$P_a : \quad \min_{\mathbf{x}^T \mathbf{x} = 1} \frac{1}{2} \|\mathbf{x} - \mathbf{x}_0\|^2 \quad P_b : \quad \min_{\mathbf{x}^T \mathbf{A} \mathbf{x} = 1} \frac{1}{2} \|\mathbf{x} - \mathbf{x}_0\|^2$$

where $\mathbf{A} : n \times n$ is positive-definite. It is well known that problem P_b cannot be reformulated as problem P_a : it is known that P_b does not admit a closed-form solution, whereas P_a does (radial projection).

The above argument used our tool but this is not restrictive: the key-point here is that “pre-whitening” the data model changes the power constraints (which constitute an important part of the problem formulation) and therefore changes the structure of the optimal constellation.

(ii) To address further this question we reproduce here equations (2.4) and (2.6)

$$\begin{aligned} \mathbf{y}_1 &= \text{vec}(\mathbf{Y}_1) = (\mathbf{I}_N \otimes \mathbf{X}_1) \text{vec}(\mathbf{H}^H) + \mathbf{e}_1 \\ \widetilde{\mathbf{y}}_1 &= \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{y}_1 = \mathbf{\Upsilon}^{-\frac{1}{2}} (\mathbf{I}_N \otimes \mathbf{X}_1) \text{vec}(\mathbf{H}^H) + \widetilde{\mathbf{e}}_1 \end{aligned}$$

where \mathbf{y}_1 corresponds to a colored noise system ($\mathbf{e}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Upsilon})$) and $\widetilde{\mathbf{y}}_1$ represents the “pre-whitened” system ($\widetilde{\mathbf{e}}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{NT})$). We recall that a white noise system corresponds to equation (2.5), i.e.,

$$\mathbf{y}_2 = \text{vec}(\mathbf{Y}_2) = (\mathbf{I}_N \otimes \mathbf{X}_2) \text{vec}(\mathbf{H}^H) + \mathbf{e}_2$$

where $\mathbf{e}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{NT})$. We have a solution for model \mathbf{y}_2 (reference [27]). We want to use it in the model $\tilde{\mathbf{y}}_1$. However, the structure of the model $\tilde{\mathbf{y}}_1$ matches the structure of the model \mathbf{y}_2 if and only if (iff) the signal $\mathbf{\Upsilon}^{-1/2}(\mathbf{I}_N \otimes \mathbf{X}_1)$ can be put in the block-diagonal format $\mathbf{I}_N \otimes \mathbf{X}_2$. This is possible iff $\mathbf{\Upsilon}$ has the special structure

$$\mathbf{\Upsilon} = \mathbf{I}_N \otimes \mathbf{\Sigma}$$

(from a physical viewpoint, corresponds to spatially uncorrelated receive antennas). Thus, for general $\mathbf{\Upsilon}$, we cannot “transport” a solution obtained for the data model \mathbf{y}_2 to the data model $\tilde{\mathbf{y}}_1$. But, let’s address the special case $\mathbf{\Upsilon} = \mathbf{I}_N \otimes \mathbf{\Sigma}$. We would have

$$\tilde{\mathbf{y}}_1 = \left(\mathbf{I}_N \otimes \left(\mathbf{\Sigma}^{-\frac{1}{2}} \mathbf{X}_1 \right) \right) \text{vec}(\mathbf{H}^H) + \tilde{\mathbf{e}}_1.$$

Now, the codebook construction solution provided in [27] for the data model \mathbf{y}_2 consists in selecting \mathbf{X}_2 from unitary constellations. Employing this solution in the data model $\tilde{\mathbf{y}}_1$ corresponds to making $\mathbf{\Sigma}^{-1/2} \mathbf{X}_1$ unitary. That is, suppose $\mathcal{C} = \{\mathbf{U}_1, \dots, \mathbf{U}_K\}$ ($\mathbf{U}_k^H \mathbf{U}_k = \frac{1}{M} \mathbf{I}_M$) is an optimal unitary codebook for \mathbf{y}_2 , i.e., the codeword \mathbf{X}_2 is selected within \mathcal{C} . Then, with respect to the data model $\tilde{\mathbf{y}}_1$ one should take $\mathbf{\Sigma}^{-1/2} \mathbf{X}_1 \in \mathcal{C}$, or, equivalently, the codeword \mathbf{X}_1 should be taken from

$$\tilde{\mathcal{C}} = \left\{ \mathbf{\Sigma}^{1/2} \mathbf{U}_1, \dots, \mathbf{\Sigma}^{1/2} \mathbf{U}_K \right\}.$$

The main problem is here is that the codebook $\tilde{\mathcal{C}}$ does not verify the power constraint, i.e., in general, we will not have

$$\text{tr}(\mathbf{U}_k^H \mathbf{\Sigma} \mathbf{U}_k) = 1 \quad \text{for all } k = 1, 2, \dots, K$$

for generic unitary matrices \mathbf{U}_k . One (sub-optimal) way around this is to enforce the power constraint and pass to the codebook

$$\hat{\mathcal{C}} = \left\{ \frac{\mathbf{\Sigma}^{1/2} \mathbf{U}_1}{\|\mathbf{\Sigma}^{1/2} \mathbf{U}_1\|}, \dots, \frac{\mathbf{\Sigma}^{1/2} \mathbf{U}_K}{\|\mathbf{\Sigma}^{1/2} \mathbf{U}_K\|} \right\}. \quad (2.12)$$

Another way around is to define the following codebook

$$\bar{\mathcal{C}} = \left\{ \mathbf{\Sigma}^{1/2} \mathbf{V}_1, \mathbf{\Theta} \mathbf{\Sigma}^{1/2} \mathbf{V}_1, \dots, \mathbf{\Theta}^{K-1} \mathbf{\Sigma}^{1/2} \mathbf{V}_1 \right\} \quad (2.13)$$

where the $T \times M$ matrix \mathbf{V}_1 is such that $\mathbf{V}_1^H \mathbf{V}_1 = \frac{1}{M} \mathbf{I}_M$ and $\text{tr}(\mathbf{V}_1^H \boldsymbol{\Sigma} \mathbf{V}_1) = 1$. The matrix $\boldsymbol{\Theta}$ is a $T \times T$ diagonal matrix whose diagonal elements are $e^{i2\pi u_1/K}, \dots, e^{i2\pi u_T/K}$ with the coefficients u_1, \dots, u_T presented in [27]. Note that $\boldsymbol{\Theta}$ is a unitary matrix and that $\boldsymbol{\Theta}^K = \mathbf{I}_T$. Clearly, the codebook $\bar{\mathcal{C}}$ satisfies the power constraint in the assumption **A2**.

We emphasize that $\hat{\mathcal{C}}$ and $\bar{\mathcal{C}}$ were heuristically obtained (do not satisfy an optimality criterion). Their performance will be assessed by a computer simulation, please see figures 2.7-2.9 in section 2.5.

(iii) Under the assumption **A1**, no statistical model is attached to the channel matrix. Anyhow, we present an analysis for the case when $\text{vec}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{MN})$. Due to the assumption **A3**, $\text{vec}(\mathbf{E}) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Upsilon})$ for some $\boldsymbol{\Upsilon} \succ \mathbf{0}$. We start by remarking that the model in (2.1) can be rewritten as

$$\mathbf{Y}^H = \mathbf{H} \mathbf{X}^H + \mathbf{E}^H,$$

and hence as

$$\check{\mathbf{y}} = \mathcal{G} \check{\mathbf{x}} + \check{\mathbf{e}}, \quad (2.14)$$

where $\check{\mathbf{y}} = \text{vec}(\mathbf{Y}^H)$, $\check{\mathbf{x}} = \text{vec}(\mathbf{X}^H)$, $\check{\mathbf{e}} = \text{vec}(\mathbf{E}^H)$ and $\mathcal{G} = \mathbf{I}_T \otimes \mathbf{H}$. From the perspective of most communications objectives, the system described in (2.14) is equivalent to the system

$$\bar{\mathbf{y}} = \bar{\mathcal{G}} \check{\mathbf{x}} + \bar{\mathbf{e}}, \quad (2.15)$$

where $\bar{\mathbf{e}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{TN})$, and

$$\bar{\mathcal{G}} = \mathbf{L}^{-1} \mathcal{G}, \quad (2.16)$$

where \mathbf{L} is a Cholesky factor of $\boldsymbol{\Upsilon}$; i.e., $\boldsymbol{\Upsilon} = \mathbf{L} \mathbf{L}^H$. This represents the channel model in which the first step of the receiver processing is to pre-whiten the noise. It is clear from (2.16) that the statistics of the elements of $\bar{\mathcal{G}}$ are different from those of \mathcal{G} and, hence, unitary constellations cannot be employed.

2.3 Considerations About the New Codebook Merit Function

Before addressing the codebook design problem (2.10) we draw in this section some conclusions about the codebook merit function f in (2.10). In subsection 2.3.1, we show that, for the special case of spatio-temporally white noise and $K = 2$, the unitary constellations are the optimal ones with respect to f . In subsection 2.3.2, we show that, when restricting attention to unitary codebooks, our codebook design criterion corresponds to a packing problem in the Grassmannian space with respect to *spectral distance*, for the white noise case. In subsection 2.3.3, we argue that the design method proposed in (2.11) guarantees that the constructed constellation has full diversity when $T \geq 2M$.

2.3.1 Optimality of Unitary Codewords for the White Noise Case

Consider a codebook with two codewords $\mathcal{C} = \{\mathbf{X}_1, \mathbf{X}_2\}$. We want to maximize $f(\mathcal{C}) = \min\{\lambda_{\min}(\mathbf{L}_{ij}(\mathcal{C})) : i \neq j\}$ subject to $\text{tr}(\mathbf{X}_k^H \mathbf{X}_k) = 1$. We rewrite $\mathbf{L}_{ij}(\mathcal{C})$ as

$$\mathbf{L}_{ij}(\mathcal{C}) = (\mathbf{x}_i^H \mathbf{x}_i)^{\frac{1}{2}} (\mathbf{I}_{MN} - \mathbf{U}_i^H \mathbf{U}_j \mathbf{U}_j^H \mathbf{U}_i) (\mathbf{x}_i^H \mathbf{x}_i)^{\frac{1}{2}} \quad (2.17)$$

where $\mathbf{U}_i = \mathbf{x}_i (\mathbf{x}_i^H \mathbf{x}_i)^{-\frac{1}{2}}$, $\mathbf{U}_j = \mathbf{x}_j (\mathbf{x}_j^H \mathbf{x}_j)^{-\frac{1}{2}}$. That is, \mathbf{U}_i contains an orthonormal basis for the subspace spanned by the columns of \mathbf{x}_i . Notice that $\mathbf{U}_j^H \mathbf{U}_j = \mathbf{U}_i^H \mathbf{U}_i = \mathbf{I}_{MN}$. To proceed with the analysis we use an useful fact from the cosine sine (CS) decomposition, see [52] pp. 199: if $\mathbf{U}_i, \mathbf{U}_j$ are $TN \times MN$ matrices with orthonormal columns ($T \geq M$), then there exist $MN \times MN$ unitary matrices \mathbf{W}_1 and \mathbf{W}_2 , and a $TN \times TN$ unitary matrix \mathbf{Q} with the following properties:

(i) If $2MN \leq TN$ ($2M \leq T$), then

$$\mathbf{Q}\mathbf{U}_i\mathbf{W}_1 = \begin{bmatrix} \mathbf{I}_{MN} \\ \mathbf{0}_{MN} \\ \mathbf{0}_{(TN-2MN) \times MN} \end{bmatrix}, \quad \mathbf{Q}\mathbf{U}_j\mathbf{W}_2 = \begin{bmatrix} \mathbf{C}_{ij} \\ \mathbf{S}_{ij} \\ \mathbf{0}_{(TN-2MN) \times MN} \end{bmatrix} \quad (2.18)$$

where \mathbf{C}_{ij} is a diagonal $MN \times MN$ matrix with diagonal entries $\cos \alpha_1, \dots, \cos \alpha_{MN}$, $0 \leq \alpha_1 \leq \dots \leq \alpha_{MN} \leq \frac{\pi}{2}$, and $\mathbf{S}_{ij}^2 + \mathbf{C}_{ij}^2 = \mathbf{I}_{MN}$. Now, using (2.18) we can write

$$\mathbf{W}_2^H \mathbf{U}_j^H \mathbf{Q}^H \mathbf{Q} \mathbf{U}_i \mathbf{W}_1 = \mathbf{W}_2^H \mathbf{U}_j^H \mathbf{U}_i \mathbf{W}_1 = \mathbf{C}_{ij} \Rightarrow \mathbf{U}_j^H \mathbf{U}_i = \mathbf{W}_2 \mathbf{C}_{ij} \mathbf{W}_1^H, \quad (2.19)$$

so α_i for $i = 1, \dots, MN$ are the *principal angles* between the subspaces spanned by \mathbf{U}_i and \mathbf{U}_j . Due to Ostrowski's theorem pp. 224, 225 in [53], and equations (2.17) and (2.19), it is not difficult to see that the following inequality holds

$$\lambda_{\min}(\mathbf{L}_{ij}) \geq \lambda_{\min}(\mathbf{x}_i^H \mathbf{x}_i) \lambda_{\min}(\mathbf{S}_{ij}^2). \quad (2.20)$$

Clearly, from (2.20), we deduce that in order to minimize an upper bound on PEP in high SNR regime, one should simultaneously increase $\lambda_{\min}(\mathbf{x}_i^H \mathbf{x}_i)$ and $\lambda_{\min}(\mathbf{S}_{ij}^2)$. Unfortunately, the right-hand side of the inequality (2.20) does not offer much insight into the form of the optimal codebook for the case of arbitrary noise covariance matrix $\mathbf{\Upsilon}$ (even for the case $K = 2$). One of the reasons originates from the fact that pairwise error probabilities are not symmetric for this general case. Hence, in the following, we treat the specific case of spatio-temporal white Gaussian observation noise to find out what conclusions can we draw about the form of the optimal codebook.

Special case (spatio-temporal white noise): $\mathbf{\Upsilon} = \mathbf{I}_{TN}$, $2M \leq T$. Remark that using (2.7), (2.9), and for $\mathbf{\Upsilon} = \mathbf{I}_{TN}$, we have

$$\mathbf{L}_{ij} = \mathbf{I}_N \otimes \left(\mathbf{X}_i^H \mathbf{X}_i - \mathbf{X}_i^H \mathbf{X}_j (\mathbf{X}_j^H \mathbf{X}_j)^{-1} \mathbf{X}_j^H \mathbf{X}_i \right).$$

Hence,

$$\lambda_{\min}(\mathbf{L}_{ij}) = \lambda_{\min} \left(\mathbf{X}_i^H \mathbf{X}_i - \mathbf{X}_i^H \mathbf{X}_j (\mathbf{X}_j^H \mathbf{X}_j)^{-1} \mathbf{X}_j^H \mathbf{X}_i \right). \quad (2.21)$$

From (2.21), an immediate conclusion is that the code design criterion in (2.11) does not depend on the number of receive antennas N . Because $T \geq 2M$ (in particular $T \geq M$), using a thin singular value decomposition (SVD), we can write $\mathbf{X}_i = \mathbf{V}_i \mathbf{D}_i \mathbf{W}_i$ and $\mathbf{X}_j = \mathbf{V}_j \mathbf{D}_j \mathbf{W}_j$ where \mathbf{V}_i and \mathbf{V}_j are $T \times M$ unitary (orthonormal) matrices, \mathbf{W}_i and \mathbf{W}_j are $M \times M$ unitary matrices, and \mathbf{D}_i , \mathbf{D}_j are $M \times M$ real nonnegative diagonal matrices. It is not difficult to see that

$$\lambda_{\min}(\mathbf{L}_{ij}) = \lambda_{\min} \left(\mathbf{D}_i^2 - \mathbf{D}_i \mathbf{V}_i^H \mathbf{V}_j \mathbf{V}_j^H \mathbf{V}_i \mathbf{D}_i \right). \quad (2.22)$$

As we can see from (2.22), the matrices \mathbf{W}_i and \mathbf{W}_j do not appear in the expression. This implies that any optimal constellation can be described in the form $\mathbf{X}_i = \mathbf{V}_i \mathbf{D}_i$.

We now show that for two symbol constellations ($K = 2$) the unitary constellations are optimal in the sense of maximizing the codebook merit function defined in (2.10). Toward this end, note that for $\mathbf{X}_1 = \mathbf{V}_1 \mathbf{D}_1$ and $\mathbf{X}_2 = \mathbf{V}_2 \mathbf{D}_2$

$$f(\mathbf{X}_1, \mathbf{X}_2) = \min\{f_{12}(\mathbf{X}_1, \mathbf{X}_2), f_{21}(\mathbf{X}_1, \mathbf{X}_2)\} \leq f_{12}(\mathbf{X}_1, \mathbf{X}_2) = \lambda_{\min}(\mathbf{L}_{12}), \quad (2.23)$$

where \mathbf{V}_1 and \mathbf{V}_2 are $T \times M$ unitary (orthonormal) matrices, and $\mathbf{D}_1, \mathbf{D}_2$ are $M \times M$ real nonnegative diagonal matrices. Since $\mathbf{V}_1, \mathbf{V}_2$ are $T \times M$ matrices with orthonormal columns and $2M \leq T$, as before, we know that there exist $M \times M$ unitary matrices \mathbf{W}_1 and \mathbf{W}_2 , and a $T \times T$ unitary matrix \mathbf{Q} with the following properties [52]:

$$\mathbf{Q} \mathbf{V}_1 \mathbf{W}_1 = \begin{bmatrix} \mathbf{I}_M \\ \mathbf{0}_M \\ \mathbf{0}_{(T-2M) \times M} \end{bmatrix}, \quad \mathbf{Q} \mathbf{V}_2 \mathbf{W}_2 = \begin{bmatrix} \mathbf{C}_{12} \\ \mathbf{S}_{12} \\ \mathbf{0}_{(T-2M) \times M} \end{bmatrix} \quad (2.24)$$

where \mathbf{C}_{12} is a diagonal $M \times M$ matrix with diagonal entries $\cos \beta_1, \cos \beta_2, \dots, \cos \beta_M$, $0 \leq \beta_1 \leq \dots \leq \beta_M \leq \frac{\pi}{2}$, and $\mathbf{S}_{12}^2 + \mathbf{C}_{12}^2 = \mathbf{I}_M$. Substituting (2.24) in (2.23) yields

$$\begin{aligned} \lambda_{\min}(\mathbf{L}_{12}) &= \lambda_{\min}(\mathbf{D}_1 \mathbf{W}_1 \mathbf{S}_{12}^2 \mathbf{W}_1^H \mathbf{D}_1) \\ &= \lambda_{\min}(\mathbf{S}_{12} \mathbf{W}_1^H \mathbf{D}_1^2 \mathbf{W}_1 \mathbf{S}_{12}) \leq \lambda_{\min}(\mathbf{D}_1^2) \lambda_{\max}(\mathbf{S}_{12}^2) \end{aligned} \quad (2.25)$$

where (2.25) is valid due to Ostrowski's theorem. Since $\lambda_{\min}(\mathbf{D}_1^2) \leq \frac{1}{M} \text{tr}(\mathbf{X}_1^H \mathbf{X}_1) = \frac{1}{M}$, and also using (2.23) and (2.25) we have the upper bound on the codebook merit function for $K = 2$:

$$f(\mathbf{X}_1, \mathbf{X}_2) \leq \frac{1}{M}. \quad (2.26)$$

Since we want to maximize the codebook merit function, from (2.23) and (2.26) we can list some of the conditions for it to happen:

1. The constellation of unitary matrices is optimal, i.e., $\mathbf{D}_1 = \mathbf{D}_2 = \frac{1}{\sqrt{M}} \mathbf{I}_M$ and $\mathbf{X}_1^H \mathbf{X}_1 = \mathbf{X}_2^H \mathbf{X}_2 = \frac{1}{M} \mathbf{I}_M$.
2. We want \mathbf{V}_1 and \mathbf{V}_2 to be separated as much as possible. The optimal scenario is when $\beta_1 = \frac{\pi}{2}$, the case when codewords \mathbf{X}_1 and \mathbf{X}_2 are mutually orthogonal, i.e., $\mathbf{X}_2^H \mathbf{X}_1 = 0$.

In this case, the inequality sign in (2.26) can be replaced with an equality sign. Thus, we showed that for the special case of spatio-temporally white noise and $K = 2$ the unitary constellations are the optimal ones with respect to our codebook design criterion f . We recall that the unitary structure was also shown to be optimal in [23, 24, 25] from both the capacity and asymptotic UB on the probability of error minimization viewpoints.

(ii) For $M \leq T < 2M$, then

$$QU_iW_1 = \begin{bmatrix} \mathbf{I}_{TN-MN} & \mathbf{0}_{(TN-MN) \times (2MN-TN)} \\ \mathbf{0}_{(2MN-TN) \times (TN-MN)} & \mathbf{I}_{2MN-TN} \\ \mathbf{0}_{TN-MN} & \mathbf{0}_{(TN-MN) \times (2MN-TN)} \end{bmatrix}$$

$$QU_jV_1 = \begin{bmatrix} \mathbf{C}_{ij} & \mathbf{0}_{(TN-MN) \times (2MN-TN)} \\ \mathbf{0}_{(2MN-TN) \times (TN-MN)} & \mathbf{I}_{2MN-TN} \\ \mathbf{S}_{ij} & \mathbf{0}_{(TN-MN) \times (2MN-TN)} \end{bmatrix}$$

where \mathbf{C}_{ij} is a diagonal $(TN - MN) \times (TN - MN)$ matrix with diagonal entries $\cos \alpha_1, \cos \alpha_2, \dots, \cos \alpha_{TN-MN}$, $0 \leq \alpha_1 \leq \dots \leq \alpha_{TN-MN} \leq \frac{\pi}{2}$, and $\mathbf{S}_{ij}^2 + \mathbf{C}_{ij}^2 = \mathbf{I}_{TN-MN}$.

Repeating the analysis which has performed previously (for the case $2M \leq T$) leads to

$$\mathbf{L}_{ij} = (\mathbf{x}_i^H \mathbf{x}_i)^{\frac{1}{2}} \mathbf{W}_1 \begin{bmatrix} \mathbf{S}_{ij}^2 & \mathbf{0}_{(TN-MN) \times (2MN-TN)} \\ \mathbf{0}_{(2MN-TN) \times (TN-MN)} & \mathbf{0}_{2MN-TN} \end{bmatrix} \mathbf{W}_1^H (\mathbf{x}_i^H \mathbf{x}_i)^{\frac{1}{2}}. \quad (2.27)$$

Given that $T < 2M$, we see that the lower right block of zeros in the middle matrix in the right-hand side of (2.27) is non-void. Thus, $\lambda_{\min}(\mathbf{L}_{ij}) = 0$ and plugging this in (2.9) yields the upper-bound $P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \leq \mathcal{Q}(0) = 0.5$ which holds irrespective of the choice of codewords. Thus, we cannot extract a guideline for codebook construction in this case. This motivates the following assumption.

A4. (Length of channel coherence) In this work, the length of the coherence interval T is at least as twice as large as the number of transmit antennas M : $T \geq 2M$.

The preceding assumption is not surprising since, for the special case $\mathbf{Y} = \mathbf{I}_{TN}$, Rayleigh fading and in high SNR scenario, it is known that the length of the coherence interval has to be necessarily at least as twice as large as the number of transmit antennas ($2M \leq T$) to achieve full order of diversity MN [25], but also, from the capacity viewpoint

it is found that there is no point in using more than $\frac{T}{2}$ transmit antenna when one wants to maximize the number of degrees of freedom [26].

2.3.2 Codebook Design as a Grassmannian Packing

It is instructive to compare our codebook construction criterion with the one proposed in [24, 27] defined as

$$\mathcal{C}^* = \arg \min_{\mathcal{C} \in \mathcal{N}} \max_{1 \leq i \neq j \leq K} \text{tr}(\mathbf{X}_i^H \mathbf{X}_j \mathbf{X}_j^H \mathbf{X}_i) \quad (2.28)$$

where the constraint space is the set of unitary codebooks

$$\mathcal{N} = \left\{ (\mathbf{X}_1, \dots, \mathbf{X}_K) : \mathbf{X}_k^H \mathbf{X}_k = \frac{1}{M} \mathbf{I}_M \right\}.$$

It is readily seen that (2.28) is equivalent to

$$\mathcal{C}^* = \arg \min_{\mathcal{C} \in \mathcal{N}} \max_{1 \leq i \neq j \leq K} \cos^2 \theta_{ij,1} + \dots + \cos^2 \theta_{ij,M} \quad (2.29)$$

where $0 \leq \theta_{ij,1} \leq \dots \leq \theta_{ij,M} \leq \pi/2$ denote the principal angles between \mathbf{X}_i and \mathbf{X}_j . In order to compare our approach with the one proposed in [24, 27], we must temporarily adopt the signal model assumptions in [24, 27], i.e., we consider white noise ($\mathbf{\Upsilon} = \mathbf{I}_{NT}$) and also unitary codebooks. In this setup, our codebook construction criterion in (2.11), simplifies to

$$\mathcal{C}^* = \arg \min_{\mathcal{C} \in \mathcal{N}} \max_{1 \leq i \neq j \leq K} \cos^2 \theta_{ij,1}. \quad (2.30)$$

It is clear that both criteria in (2.29) and (2.30) aim at building codebooks by reducing the pairwise “spatial crosstalk” between distinct codewords. The distinction lies in how this crosstalk is measured: the strategy in (2.29) looks at the average of the principal angles and corresponds to the Grassmannian chordal distance [27], whereas our criterion in (2.30) considers the worst-case and leads to the Grassmannian spectral distance [28]. We recall that, as defined in [28], the squared spectral distance of two linear subspaces of the same dimension, say $\mathcal{L}_i, \mathcal{L}_j \subset \mathbb{C}^n$, is given by $\sin^2 \theta_{ij,1}$ where $\theta_{ij,1}$ is the minimal principal angle between \mathcal{L}_i and \mathcal{L}_j . It can be computed as follows: if the matrices $\mathbf{U}_i, \mathbf{U}_j$ contain in their columns an orthonormal basis for $\mathcal{L}_i, \mathcal{L}_j$, respectively, then $\sin^2 \theta_{ij,1} = 1 - \sigma_{ij}^2$ where

σ_{ij} is the maximal singular value of $\mathbf{U}_i^H \mathbf{U}_j$. Given this definition, it follows that $\sin^2 \theta_{ij,1}$ corresponds to the squared spectral distance between the codewords \mathbf{X}_i and \mathbf{X}_j (more precisely, between their respective range spaces). Please refer to [28] and [54] for more details on packing problems in Grassmannian space. The reader is referred to [55] for a more in depth discussion on the geometry of complex Grassmann manifolds regarding distance, cut and conjugate locus, etc.

We note that, for this particular scenario, the criterion presented in (2.28) is easier to deal with mathematically. Also, from (2.30) we see that, for $M = 1$, the problem of finding good codes coincides with the very well known packing problem in the complex projective space [28].

2.3.3 Maximal Diversity Analysis

The method proposed in (2.10) and (2.11) is a numerical method as, for example, the ones of [27] and [29]. A very interesting theoretical point is whether the method guarantees the maximal diversity. Remark that the aforementioned references were not able to guarantee the maximal diversity to their codes. In the following, we provide an analysis of this topic.

We will assume uncorrelated Rayleigh fading channel, i.e., $h_{ij} \stackrel{iid}{\sim} \mathcal{CN}(0, \sigma^2)$. In this case,

$$\text{SNR} = \frac{\text{E} \left[\|\mathbf{X}_k \mathbf{H}^H\|^2 \right]}{\text{E} \left[\|\mathbf{E}\|^2 \right]} = \frac{N\sigma^2}{\text{tr}(\mathbf{Y})}.$$

For a two point constellation ($K = 2$), the probability of error is given by

$$\begin{aligned} P_e &= 0.5 \text{E}[P_{\mathbf{X}_1 \rightarrow \mathbf{X}_2} + P_{\mathbf{X}_2 \rightarrow \mathbf{X}_1}] \\ &= 0.5 \text{E} \left[\mathcal{Q} \left(\frac{1}{\sqrt{2}} \sqrt{\mathbf{g}^H \mathbf{L}_{12} \mathbf{g}} \right) + \mathcal{Q} \left(\frac{1}{\sqrt{2}} \sqrt{\mathbf{g}^H \mathbf{L}_{21} \mathbf{g}} \right) \right] \end{aligned} \quad (2.31)$$

where the expectations are over the channel $\mathbf{g} = \text{vec}(\mathbf{H}^H)$. Note that $\mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{MN})$.

We proceed as

$$P_e \leq 0.5 \mathbb{E} \left[\exp\left\{-\frac{1}{4} \mathbf{g}^H \mathbf{L}_{12} \mathbf{g}\right\} \right] + 0.5 \mathbb{E} \left[\exp\left\{-\frac{1}{4} \mathbf{g}^H \mathbf{L}_{21} \mathbf{g}\right\} \right] \quad (2.32)$$

$$= 0.5 \left(\frac{1}{\det(\mathbf{I}_{MN} + \frac{\sigma^2}{4} \mathbf{L}_{12})} + \frac{1}{\det(\mathbf{I}_{MN} + \frac{\sigma^2}{4} \mathbf{L}_{21})} \right) \quad (2.33)$$

$$\leq 0.5 \left(\frac{1}{\left(1 + \frac{\sigma^2}{4} \lambda_{\min}(\mathbf{L}_{12})\right)^{MN}} + \frac{1}{\left(1 + \frac{\sigma^2}{4} \lambda_{\min}(\mathbf{L}_{21})\right)^{MN}} \right) \quad (2.34)$$

$$\leq \text{SNR}^{-MN} \left(\frac{4N}{\text{tr}(\mathbf{Y}) f_{12}} \right)^{MN} \quad (2.35)$$

where $f_{12} = \min\{\lambda_{\min}(\mathbf{L}_{12}), \lambda_{\min}(\mathbf{L}_{21})\}$. The inequality (2.32) is the Chernoff upper bound

$$\mathcal{Q}(x) \leq e^{-x^2/2}$$

applied to (2.31), see section 2.1.5 in [56]. The equation (2.33) is due to the following formula: if $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{K})$ and $\mathbf{L} = \mathbf{L}^H \succeq \mathbf{0}$, then

$$\mathbb{E}[\exp\{-\mathbf{z}^H \mathbf{L} \mathbf{z}\}] = \frac{\exp\{-\boldsymbol{\mu}^H \mathbf{L} (\mathbf{I} + \mathbf{K} \mathbf{L})^{-1} \boldsymbol{\mu}\}}{\det(\mathbf{I} + \mathbf{K} \mathbf{L})},$$

see equation (5.59), page 194 in [21]. Inequality (2.34) is obtained by using the fact that, for an $n \times n$ Hermitian matrix $\mathbf{L} = \mathbf{L}^H \succeq \mathbf{0}$, we have

$$\det(\mathbf{I}_n + \mathbf{L}) \geq (1 + \lambda_{\min}(\mathbf{L}))^n.$$

Now, if the range space of the codewords \mathbf{X}_1 and \mathbf{X}_2 do not share a line then both \mathbf{L}_{12} and \mathbf{L}_{21} are positive-definite, i.e.,

$$\lambda_{\min}(\mathbf{L}_{12}) > 0 \quad \lambda_{\min}(\mathbf{L}_{21}) > 0.$$

Thus, at high SNR (channel power $\sigma^2 \rightarrow +\infty$), we have

$$1 + \frac{\sigma^2}{4} \lambda_{\min}(\mathbf{L}_{ij}) \approx \frac{\sigma^2}{4} \lambda_{\min}(\mathbf{L}_{ij}).$$

Note the importance of having $\lambda_{\min}(\mathbf{L}_{ij}) > 0$ in the above approximation. From this, it is straightforward to attain (2.35).

This line of reasoning is readily extended to the UB:

$$\begin{aligned}
P_e &\leq \text{UB} = \frac{1}{K} \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \\
&\leq \frac{1}{K} \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K \text{SNR}^{-MN} \left(\frac{4N}{\text{tr}(\mathbf{\Upsilon}) \lambda_{\min}(\mathbf{L}_{ij})} \right)^{MN} \\
&\leq (K-1) \text{SNR}^{-MN} \left(\frac{4N}{\text{tr}(\mathbf{\Upsilon}) f} \right)^{MN}
\end{aligned}$$

where

$$f = \min \{ \lambda_{\min}(\mathbf{L}_{ij} : i \neq j) \}.$$

What all this analysis shows is that, as long as we have a codebook $\mathcal{C} = \{\mathbf{X}_1, \dots, \mathbf{X}_K\}$ which makes the matrices \mathbf{L}_{ij} ($i \neq j$) definite-positive (for which $T \geq 2M$ is necessary, see subsection 2.3.1), then full-diversity is secured. But, what our design method tries to achieve is exactly the maximization of

$$\min \{ \lambda_{\min}(\mathbf{L}_{ij} : i \neq j) \},$$

see equations (2.10) and (2.11). Although we are not able to furnish a theoretical proof that our algorithm actually guarantees full diversity for any combination of T , K , M , N and $\mathbf{\Upsilon}$, in all simulations presented afterwards the condition $f > 0$ is fulfilled.

2.4 Codebook Construction

We propose a two-phase methodology to tackle the optimization problem in (2.11). In phase one, we start by solving a convex semi-definite programming (SDP) relaxation to obtain a rough estimate of the optimal codebook. Phase two refines it through a geodesic descent optimization algorithm (GDA) which efficiently exploits the Riemannian geometry of the constraints. Suppose a codebook of size K is desired. In table 2.1, page 30 we give the strategy that has shown to be effective.

The algorithm presented in table 2.1 is of the greedy type. We now explain Steps (3)

input:	M, N, T, K, Υ
step 1)	Choose the first codeword (randomly generated, etc);
step 2)	Set $k = 2$;
step 3)	Perform SDP relaxation to obtain k -th codeword;
step 4)	Set $k = k + 1$;
step 5)	if $k \leq K$, return to Step 3);
step 6)	Run the geodesic descent algorithm (GDA) to obtain the final codebook;
output:	The matrix $\mathcal{X} = [\text{vec}(\mathbf{X}_1) \ \dots \ \text{vec}(\mathbf{X}_K)]$

Table 2.1: Codebook Design Algorithm

and (6), respectively, in more detail.

Phase 1: SDP relaxation. This phase constructs a sub-optimal codebook $\mathcal{C}^* = \{\mathbf{X}_1^*, \dots, \mathbf{X}_K^*\}$. The codebook is constructed incrementally. We start assuming that we are in a possession of a suboptimal codebook of size $k - 1$, while we are interested in a suboptimal codebook of size k , where $k = 2, 3, \dots, K$. We obtain a suboptimal codebook of size k by retaining the first $k - 1$ codewords. Hence, we solve the optimization problem in the sequel consecutively $K - 1$ times. There are several strategies for choosing the first codeword \mathbf{X}_1^* , e.g, randomly generated, filling columns of the matrix with eigenvectors associated to the smallest eigenvalues of the noise covariance matrix, etc. Addition of a new codeword consists in solving a SDP. Let $\mathcal{C}_{k-1}^* = \{\mathbf{X}_1^*, \dots, \mathbf{X}_{k-1}^*\}$ be the codebook at the $k - 1^{\text{th}}$ stage. The new codeword is found by solving

$$\begin{aligned}
\mathbf{X}_k^* &= \arg \max_{\text{tr}(\mathbf{X}_k^H \mathbf{X}_k) = 1} f(\mathbf{X}_1^*, \dots, \mathbf{X}_{k-1}^*, \mathbf{X}_k) \\
&= \arg \max_{\text{tr}(\mathbf{X}_k^H \mathbf{X}_k) = 1} \min_{1 \leq m \leq k-1} \{\lambda_{\min}(\mathbf{L}_{mk}), \lambda_{\min}(\mathbf{L}_{km})\}.
\end{aligned} \tag{2.36}$$

We can show that the optimization problem defined in (2.36) is equivalent to

$$(\mathfrak{X}_k^*, \text{vec}(\mathbf{X}_k^*), t^*) = \arg \max t \tag{2.37}$$

with the following constraints

$$\begin{aligned} \text{LMI}_{A_m}(\mathfrak{X}_k, \text{vec}(\mathbf{X}_k), t) &\succeq \mathbf{0}, \quad m = 1, \dots, k-1 \\ \text{LMI}_{B_m}(\mathfrak{X}_k, \text{vec}(\mathbf{X}_k), t) &\succeq \mathbf{0}, \quad m = 1, \dots, k-1 \\ \text{tr}(\mathfrak{X}_k) &= 1, \quad \mathfrak{X}_k = \text{vec}(\mathbf{X}_k)\text{vec}^H(\mathbf{X}_k) \end{aligned} \quad (2.38)$$

where the abbreviations $\text{LMI}_{A_m}(\mathfrak{X}_k, \text{vec}(\mathbf{X}_k), t)$ and $\text{LMI}_{B_m}(\mathfrak{X}_k, \text{vec}(\mathbf{X}_k), t)$ denote linear (actually, affine) matrix inequalities in the variables \mathfrak{X}_k , $\text{vec}(\mathbf{X}_k)$, and t of type A and B , respectively, for $m = 1, \dots, k-1$. The proof and the meaning of the LMI's of type A and B are given in Appendix B.

Due to the rank condition in (2.38) (note that the equations $\mathfrak{X}_k = \text{vec}(\mathbf{X}_k)\text{vec}^H(\mathbf{X}_k)$ and $\text{tr}(\mathfrak{X}_k) = 1$ imply that $\text{rank}(\mathfrak{X}_k) = 1$), the design of the codewords, once again, translates into a difficult nonlinear optimization problem. However, relaxing this restriction as

$$\mathfrak{X}_k \succeq \text{vec}(\mathbf{X}_k)\text{vec}^H(\mathbf{X}_k) \quad (2.39)$$

and rewriting (2.39) as

$$\begin{bmatrix} \mathfrak{X}_k & \text{vec}(\mathbf{X}_k) \\ \text{vec}^H(\mathbf{X}_k) & 1 \end{bmatrix} \succeq \mathbf{0}$$

the optimization problem in (2.37) becomes

$$(\mathfrak{X}_k^*, \text{vec}(\mathbf{X}_k^*), t^*) = \arg \max t \quad (2.40)$$

with the constraints

$$\begin{aligned} \text{LMI}_{A_m}(\mathfrak{X}_k, \text{vec}(\mathbf{X}_k), t) &\succeq \mathbf{0}, \quad m = 1, \dots, k-1 \\ \text{LMI}_{B_m}(\mathfrak{X}_k, \text{vec}(\mathbf{X}_k), t) &\succeq \mathbf{0}, \quad m = 1, \dots, k-1 \\ \text{tr}(\mathfrak{X}_k) &= 1, \quad \begin{bmatrix} \mathfrak{X}_k & \text{vec}(\mathbf{X}_k) \\ \text{vec}^H(\mathbf{X}_k) & 1 \end{bmatrix} \succeq \mathbf{0}. \end{aligned}$$

The rank 1 relaxation is usually known as the Shor relaxation [57]. The optimization problem in (2.40) is a convex one in the variables \mathfrak{X}_k , $\text{vec}(\mathbf{X}_k)$ and t . Remark that for $K = 256$, $M = 2$, $N = 2$, $T = 8$ and in the last passage through the loop, i.e., for $k = K$,

the output variable \mathbf{x}_k is of dimension 16×16 (does not depend on N and K) and the number of linear matrix inequality constraints that needs to be defined is of order K . To solve the optimization problem in (2.40) we used the *Self-Dual-Minimization* package SeDuMi 1.1 [58]. Once the problem defined in (2.40) is solved we need to extract the k^{th} codeword from the output variable \mathbf{x}_k . Toward this end, we adopt a technique similar to [62]. The technique consists in generating independent realizations of random vectors that follow a Gaussian distribution with zero mean and covariance matrix \mathbf{x}_k , i.e., $\mathbf{z}_l \stackrel{iid}{\sim} \mathcal{CN}(\mathbf{0}, \mathbf{x}_k)$, for $l = 1, 2, \dots, L$, where L is a parameter to be chosen (in all simulations herein presented we assumed $L = 1000$). After forcing norm 1, i.e., $\mathbf{v}_l = \mathbf{z}_l / \|\mathbf{z}_l\|$ for $l = 1, 2, \dots, L$, we choose the k -th codeword, $\mathbf{X}_k^* = \text{ivec}(\mathbf{v}_{l^*})$ where

$$l^* = \underset{l = 1, 2, \dots, L}{\arg \max} f(\mathbf{X}_1^*, \mathbf{X}_2^*, \dots, \mathbf{X}_{k-1}^*, \text{ivec}(\mathbf{v}_l)). \quad (2.41)$$

The operation “ivec” operates as an inverse of “vec” (reshapes the TM -dimensional vector into a $T \times M$ matrix). Note that \mathbf{X}_k^* is a valid codeword because $\text{tr}(\mathbf{X}_k^{*H} \mathbf{X}_k^*) = 1$. We are clearly dealing with a suboptimal solution for a codebook.

Phase 2: Geodesic Descent Algorithm. Problem (2.11) requires the optimization of a non-smooth function over the smooth manifold \mathcal{M} (Cartesian product of K spheres). After phase 1, i.e., having solved the optimization problem (2.40) consecutively $K - 1$ times, for $k = 2, 3, \dots, K$, we are now in possession of a suboptimal codebook of size K . To refine it we resort to an iterative algorithm, which we call GDA (geodesic descent algorithm). In table 2.2 we explain the GDA in more detail.

Let \mathcal{C}_k be the k^{th} iterate (the initialization \mathcal{C}_0 is furnished by phase 1). Note that the power constraint $\text{tr}(\mathbf{X}_i^H \mathbf{X}_i) = 1$, for $i = 1, 2, \dots, K$, can be equivalently written as

$$\mathbf{x}_i^T \mathbf{x}_i = 1,$$

where

$$\mathbf{x}_i = \begin{bmatrix} \Re \{ \text{vec}(\mathbf{X}_i) \} \\ \Im \{ \text{vec}(\mathbf{X}_i) \} \end{bmatrix} \in \mathbb{R}^{2TM},$$

and $\Re \{ \cdot \}$ and $\Im \{ \cdot \}$ denote the real and imaginary part of a complex quantity, respectively.

In step 3 each \mathbf{x}_i , $i = 1, \dots, K$ is used to construct the vector \mathbf{x} . In step 4 we identify the

input: The matrix $\mathcal{X} = [\text{vec}(\mathbf{X}_1) \ \dots \ \text{vec}(\mathbf{X}_K)]$

step 1) Determine the value of the merit function, cost = $f(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K)$;

step 2) Initialize $\epsilon = 10^{-5}$;

step 3) Construct the vector $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_K \end{bmatrix} = \begin{bmatrix} \Re\{\text{vec}(\mathbf{X}_1)\} \\ \Im\{\text{vec}(\mathbf{X}_1)\} \\ \vdots \\ \Re\{\text{vec}(\mathbf{X}_K)\} \\ \Im\{\text{vec}(\mathbf{X}_K)\} \end{bmatrix}$;

step 4) Determine z , the number of combinations $(\mathbf{X}_i, \mathbf{X}_j)$, $1 \leq i \neq j \leq K$, such that $f_{ij}(\mathcal{C}) = \lambda_{\min}(\mathbf{L}_{ij}(\mathcal{C}))$ falls into the interval $[\text{cost}, \text{cost} + \epsilon]$, i.e. f_{ij} attains the minimum. These combinations $(\mathbf{X}_i, \mathbf{X}_j)$ are called the active ones;

step 5) Determine the gradient, $\nabla f_{i_a j_a}(\mathbf{x})$, for every active combination $(\mathbf{X}_{i_a}, \mathbf{X}_{j_a})$, $1 \leq i_a \neq j_a \leq K, a = 1, 2, \dots, z$;

step 6) Construct the gradient matrix

$$\mathbf{G} = \begin{bmatrix} \nabla^T f_{i_1 j_1}(\mathbf{x}) \\ \vdots \\ \nabla^T f_{i_z j_z}(\mathbf{x}) \end{bmatrix}_{z \times 2KTM}$$

step 7) Construct the matrix

$$\mathbb{H} = \begin{bmatrix} \mathbf{x}_1^T & 0 & \cdots & 0 \\ 0 & \mathbf{x}_2^T & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \mathbf{x}_K^T \end{bmatrix}_{K \times 2KTM}$$

step 8) Solve the linear program

$$(\mathbf{d}^*, s^*) = \begin{array}{ll} \arg \max & s; \\ \mathbf{G}\mathbf{d} \geq s\mathbf{1}_{z \times 1} & \\ \mathbb{H}\mathbf{d} = \mathbf{0}_{K \times 1} & \\ -\mathbf{1}_{2KTM \times 1} \leq \mathbf{d} \leq +\mathbf{1}_{2KTM \times 1} & \end{array}$$

step 9) If $s \leq 0$, Go to Step (16);

step 10) Initialize $\beta = 0.9$, $c = 0$, $c_{max} = 400$ and $t = 1$;

<p>step 11) Construct the geodesic</p> $\boldsymbol{\gamma}(t) = \begin{bmatrix} \mathbf{x}_1(t) \\ \vdots \\ \mathbf{x}_K(t) \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \cos(\ \mathbf{d}_1\ t) + \frac{\mathbf{d}_1}{\ \mathbf{d}_1\ } \sin(\ \mathbf{d}_1\ t) \\ \vdots \\ \mathbf{x}_K \cos(\ \mathbf{d}_K\ t) + \frac{\mathbf{d}_K}{\ \mathbf{d}_K\ } \sin(\ \mathbf{d}_K\ t) \end{bmatrix};$ <p>step 12) Determine temporary value of the merit function, $\text{tempcost} = f(\mathbf{X}_1(t), \mathbf{X}_2(t), \dots, \mathbf{X}_K(t))$, where $\mathbf{X}_i(t) = \text{ivec}(\mathbf{x}_i(t)(1:TM) + j\mathbf{x}_i(t)(TM+1:2TM))$ for $i = 1, 2, \dots, K$;</p> <p>step 13) If $\text{tempcost} > \text{cost}$, then $\text{cost} = \text{tempcost}$, $\mathbf{x}_i = \mathbf{x}_i(t)$ for $i=1, 2, \dots, K$. Return to Step (3);</p> <p>step 14) Increment c, update $t = \beta^c$;</p> <p>step 15) If $c \leq c_{max}$, Return to Step (12);</p> <p>step 16) Return the matrix $\mathcal{X} = [\text{vec}(\mathbf{X}_1) \ \dots \ \text{vec}(\mathbf{X}_K)]$, where $\mathbf{X}_i = \text{ivec}(\mathbf{x}_i(1:TM) + j\mathbf{x}_i(TM+1:2TM))$ for $i = 1, 2, \dots, K$;</p> <p>output: The matrix $\mathcal{X} = [\text{vec}(\mathbf{X}_1) \ \dots \ \text{vec}(\mathbf{X}_K)]$</p>

Table 2.2: GDA Algorithm

index set \mathcal{A} of “active” constraint pairs (i, j) , i.e., $\mathcal{A} = \{(i, j) : f_{ij}(\mathcal{C}_k) \leq f(\mathcal{C}_k) + \epsilon\}$ where ϵ is arbitrary small (in all simulations herein presented we have chosen $\epsilon = 10^{-5}$). In step 8 we check if there is an ascent direction \mathbf{d} simultaneously for all functions f_{ij} with $(i, j) \in \mathcal{A}$. We know that if it exists \mathbf{d} such that $\nabla^T f_{i_a j_a}(\mathbf{x}) \mathbf{d} > 0$, for $1 \leq i_a \neq j_a \leq K, a = 1, 2, \dots, z$, we can try to improve our cost function locally. In order to solve the optimization problem in step 8 we need to determine the gradient $\nabla f_{i_a j_a}$. In Appendix C, we give its respective expression. This ascent direction \mathbf{d} is searched within $T_{\mathcal{C}_k} \mathcal{M}$, the tangent space to \mathcal{M} at \mathcal{C}_k , and consists in solving a linear program. To ensure that \mathbf{d} belongs to $T_{\mathcal{C}_k} \mathcal{M}$, the constraint $\mathbb{H}\mathbf{d} = \mathbf{0}_{K \times 1}$ (equivalently, $\mathbf{x}_i^T \mathbf{d}_i = 0$ for $i=1, 2, \dots, K$) in step 8 is introduced. The constraint $-\mathbf{1}_{2KTM \times 1} \leq \mathbf{d} \leq +\mathbf{1}_{2KTM \times 1}$ bounds the solution of the linear program in step 8. If there is no such ascent direction, the algorithm stops. Otherwise, we perform an Armijo search for $f(\mathcal{C})$ along the geodesic which emanates from \mathcal{C}_k in the direction \mathbf{d} , see figure 2.2. This Armijo search determines \mathcal{C}_{k+1} and we repeat the loop. From the expression for the geodesic in step 11, it is easy to see that we travel along the surface of

the sphere S^{2TM-1} , i.e., $\mathbf{x}_i(t)^T \mathbf{x}_i(t) = 1$ for every $i=1,2,\dots,K$.

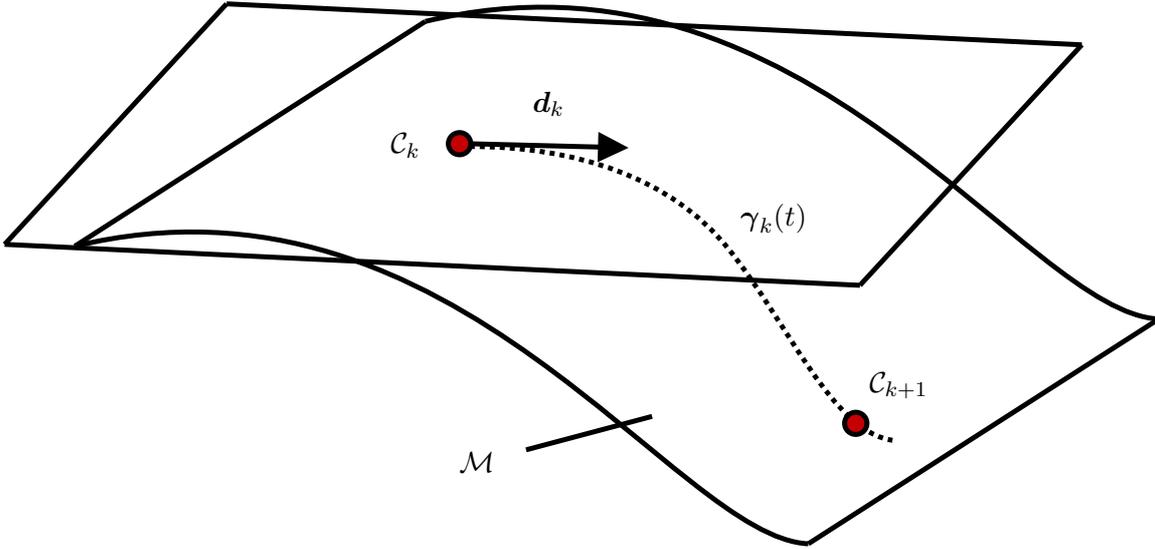


Figure 2.2: Phase 2: optimizes a non-smooth function on a manifold

A geodesic is nothing but the generalization of a straight line in Euclidean space to a curved surface [63]. In loose terms, GDA resembles a sub-gradient method and consequently, the algorithm usually converges slowly near local minimizers (sublinearly). Note however that this is not a serious drawback since codebooks can be generated off-line. It is important to notice that other non-geodesic based approaches are also of interest in this setup. See, e.g., [64] for more details.

The parameter ϵ in step 2 controls the complexity of the optimization problem in step 8. A too small ϵ implies slow convergence of the algorithm, whereas a big ϵ increases the complexity of the linear program (by increasing z , the number of active functions f_{ij}). For a codebook of size $K = 256$, and $T = 8$, $M = 2$, the gradient matrix \mathbf{G} can be of size 10000×8000 (remark that $z_{\max} = K(K-1) = 65280$). Although the matrix G is a sparse matrix, it is preferable to impose it to be of moderate size too. The choice of ϵ made in step 2 controls that.

Remark: The utility of the step 3 (SDP) in table 2.1 for large K is an open issue. Based on numerical experiments, we have found it quite useful for small and moderate sized codebooks. For example, for the real case, $M = 1$ and $T = 2$, the step 3 provides us the optimal codebook for $K = 2^p$ where $p = 1, 2, \dots$. In this case there is no need to use step 6

of the algorithm. In all simulations herein presented the procedure presented in table 2.1 has been implemented.

2.5 Results

We have constructed codes for three special categories of noise covariance matrices Υ . In all simulations we assumed a Rayleigh fading model for the channel matrix, i.e., $h_{ij} \stackrel{iid}{\sim} \mathcal{CN}(0, \sigma^2)$.

First category: spatio-temporal white observation noise

In the first category the spatio-temporal white observation noise case is considered, i.e., $\Upsilon = \mathbf{E}[\text{vec}(\mathbf{E}) \text{vec}(\mathbf{E})^H] = \mathbf{I}_{NT}$. First, we compare our codes with the codes presented in [27]. We considered scenarios with coherence interval $T = 8$, $M = 1, 2$ and 3 transmit antennas, $N = 1$ receive antennas and a codebook with $K = 256$ codewords. Let

$$\text{dist} = \frac{1}{K} \sum_{k=1}^K \sqrt{\text{tr} \left(\left(\mathbf{X}_k^H \mathbf{X}_k - \frac{1}{M} \mathbf{I}_M \right)^2 \right)}$$

denote the average distance of our codebooks from the constellation of unitary matrices. For $M = 2$, $T = 8$ and $K = 256$, the average distance obtained was $\text{dist} = 1.6 \cdot 10^{-3}$, while for $M = 3$, $T = 8$ and $K = 256$, the average distance was $\text{dist} = 1.3 \cdot 10^{-2}$. As expected (see subsection 2.3.1), the algorithm converged to constellations of almost unitary matrices. In figures 2.3–2.5, we show the symbol error rate (SER) versus

$$\text{SNR} = \mathbf{E} \left[\|\mathbf{X}_k \mathbf{H}^H\|^2 \right] / \mathbf{E} \left[\|\mathbf{E}\|^2 \right] = N\sigma^2 / \text{tr}(\Upsilon).$$

The solid-plus and dashed-circle curves represent performances of codes constructed by our method, and unitary codes respectively. As we can see, our codebook constructions replicate the performance of [27] for these particular cases, with just marginal improvements. Note that, for unitary constellations, iid Rayleigh fading and white spatio-temporal observation noise, the GLRT and the Bayesian receiver in [27] coincide (the Bayesian receiver takes into account the statistics of the channel). This, in conjunction with the fact that our codebook is almost unitary, explains the comparable performance of the two approaches. For $M = 1$, in tables 2.3-2.5 we compare our results with [28] for $T = 2, 3, \dots, 6$.

We manage to improve the best known results and in some cases actually provide optimal packings, i.e., equiangular tight frames (ETFs), which attain the Rankin upper bound (equivalently, the Welch lower bound [65]). For the sake of completeness a comparison with unitary codes found in [32] is also performed. Figure 2.6 shows the result of the experiment for $T = 4$, $M = 2$, $N = 2$ and $K = 256$. We observe that our codes show almost no improvement over the constructions presented in [32].

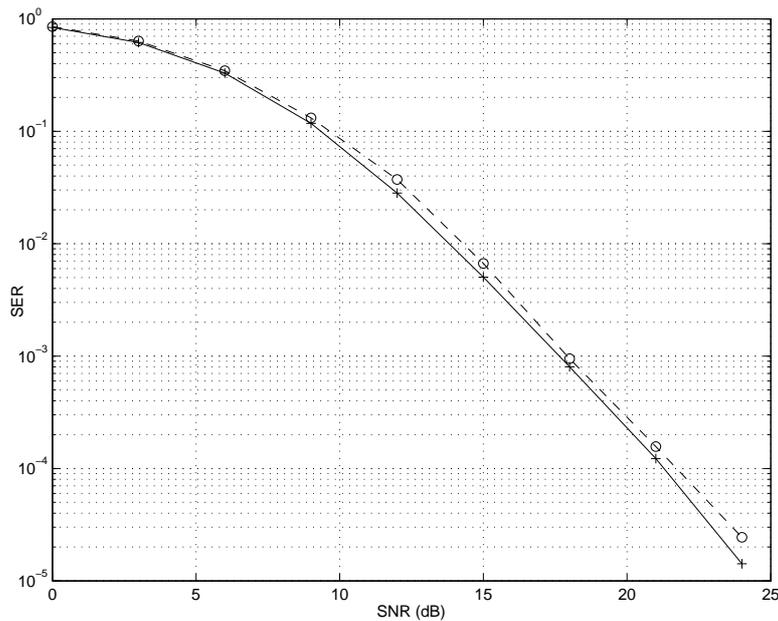


Figure 2.3: Category 1 - spatio-temporally white observation noise: $T = 8$, $M = 3$, $N = 1$, $K = 256$, $\mathbf{\Upsilon} = \mathbf{I}_{NT}$. Plus-solid curve:our codes; circle-dashed curve:unitary codes.

Second category: spatially white-temporally colored observation noise

The second category corresponds to spatially white-temporally colored observation noise, i.e., $\mathbf{\Upsilon} = \mathbf{I}_N \otimes \Sigma(\boldsymbol{\rho})$ where the vector $\boldsymbol{\rho} : T \times 1$ is the first column of an Hermitian Toeplitz matrix $\Sigma(\boldsymbol{\rho})$. To the best of our knowledge, we are not aware of any work that treats the problem of codebook constructions in the presence of spatially white-temporally colored observation noise. Hence, we compare our codes designed (adapted) to this specific scenario with unitary codes [27]. The goal here is to demonstrate the increase of performance obtained by matching the codebook construction to the noise statistics. In figures 2.7–2.9 the solid curves represent the performance of codes constructed by our method, while the dashed curves represent the performance of unitary codes. In either

		PACKING RADII (DEGREES)		
T	K	MB	JAT	Rankin
2	3	60	60	60
2	4	54.74	54.74	54.74
2	5	45.00	45.00	52.24
2	6	45.00	45.00	50.77
2	7	38.93	38.93	49.80
2	8	37.43	37.41	49.11
2	9	35.26	–	48.59
2	10	33.07	–	48.19
2	11	31.72	–	47.87
2	12	31.72	–	47.61
2	13	28.24	–	47.39
2	14	27.83	–	47.21
2	15	26.67	–	47.05
2	16	25.97	–	46.91
3	4	70.53	70.53	70.53
3	5	64.26	64.00	65.91
3	6	63.43	63.43	63.43
3	7	61.87	61.87	61.87
3	8	60.00	60.00	60.79
3	9	60.00	60.00	60.00
3	10	54.74	54.73	59.39
3	11	54.74	54.73	58.91
3	12	54.74	54.73	58.52
3	13	51.38	51.32	58.19
3	14	50.36	50.13	57.92
3	15	49.80	49.53	57.69
3	16	49.61	49.53	57.49
3	17	49.13	49.10	57.31
3	18	48.12	48.07	57.16

Table 2.3: PACKING IN COMPLEX PROJECTIVE SPACE: We compare our best configurations (MB) of K points in $\mathbb{P}^{T-1}(\mathbb{C})$ against the Tropp codes (JAT) and Rankin bound [28]. The packing radius of an ensemble is measured as the acute angle between the closest pair of lines. Minus sign symbol (-) means that no packing is available for specific pair (T, K) .

		PACKING RADII (DEGREES)		
T	K	MB	JAT	Rankin
4	5	75.52	75.52	75.52
4	6	70.89	70.88	71.57
4	7	69.29	69.29	69.30
4	8	67.79	67.78	67.79
4	9	66.31	66.21	66.72
4	10	65.74	65.71	65.91
4	11	64.79	64.64	65.27
4	12	64.68	64.24	64.76
4	13	64.34	64.34	64.34
4	14	63.43	63.43	63.99
4	15	63.43	63.43	63.69
4	16	63.43	63.43	63.43
5	6	78.46	78.46	78.46
5	7	74.55	74.52	75.04
5	8	72.83	72.81	72.98
5	9	71.33	71.24	71.57
5	10	70.53	70.51	70.53
5	11	69.73	69.71	69.73
5	12	69.04	68.89	69.10
5	13	68.38	68.19	68.58
5	14	67.92	67.66	68.15
5	15	67.48	67.37	67.79
5	16	67.08	66.68	67.48
5	17	66.82	66.53	67.21
5	18	66.57	65.87	66.98
5	19	66.57	65.75	66.77
5	20	66.42	65.77	66.59
5	21	66.42	65.83	66.42
5	22	65.91	65.87	66.27
5	23	65.91	65.90	66.14
5	24	65.91	65.91	66.02
5	25	65.91	65.91	65.91

Table 2.4: PACKING IN COMPLEX PROJECTIVE SPACE: We compare our best configurations (MB) of K points in $\mathbb{P}^{T-1}(\mathbb{C})$ against the Tropp codes (JAT) and Rankin bound [28]. The packing radius of an ensemble is measured as the acute angle between the closest pair of lines.

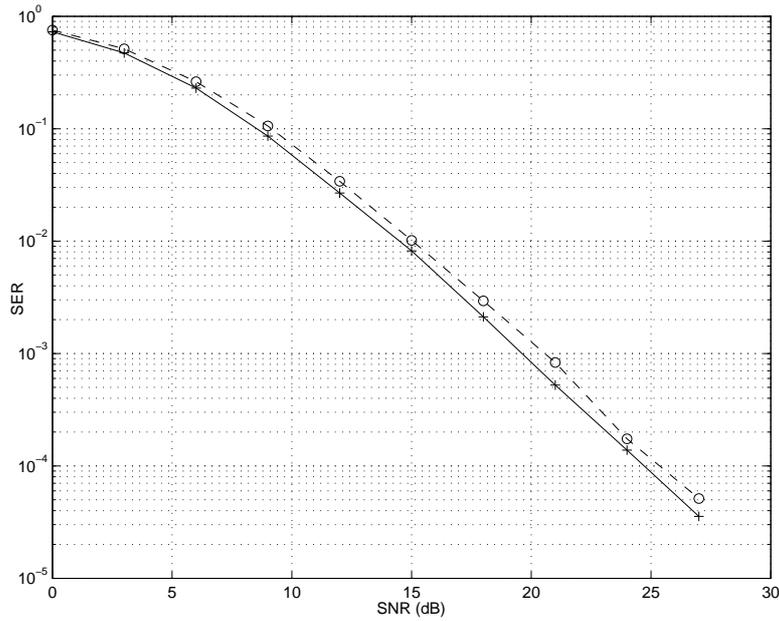


Figure 2.4: Category 1 - spatio-temporally white observation noise: $T = 8$, $M = 2$, $N = 1$, $K = 256$, $\Upsilon = \mathbf{I}_{NT}$. Plus-solid curve: our codes; circle-dashed curve: unitary codes.

case, the plus sign indicates that the GLRT receiver is implemented. The square sign indicates that the Bayesian receiver is implemented. Figure 2.7 plots the result of the experiment for $T = 8$, $M = 2$, $N = 1$, $K = 67$ and $\boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.1; \text{zeros}(3,1)]$. It can be seen that for $\text{SER} = 10^{-3}$, our codes demonstrate a gain of 3 dB when compared with the unitary codes. Figure 2.8 plots the result of the experiment for $T = 8$, $M = 2$, $N = 1$, $K = 256$ and $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$. For $\text{SER} = 10^{-3}$ our codes demonstrate gain of 2 dB when compared with unitary codes. Figure 2.9 plots the result of the experiment for $T = 8$, $M = 2$, $N = 1$, $K = 32$ and $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$. For $\text{SER} = 10^{-3}$, our codes demonstrate gain of 3 dB when compared with the unitary codes. In figures 2.7- 2.9, the dotted curve represents the performance of the codes obtained by the heuristic codebook $\hat{\mathcal{C}}$ defined in (2.12), and dash-dotted curve represents the performance of the codes obtained by the heuristic codebook $\bar{\mathcal{C}}$ defined in (2.13). As can be seen, this leads to a degradation of performance, even with respect to plain unitary codebooks (dashed curves).

We have also performed a comparison with unitary codes found in [32]. Figure 2.10 shows the result of the experiment for $T = 4$, $M = 2$, $N = 2$, $K = 256$ and $\boldsymbol{\rho} = [1;$

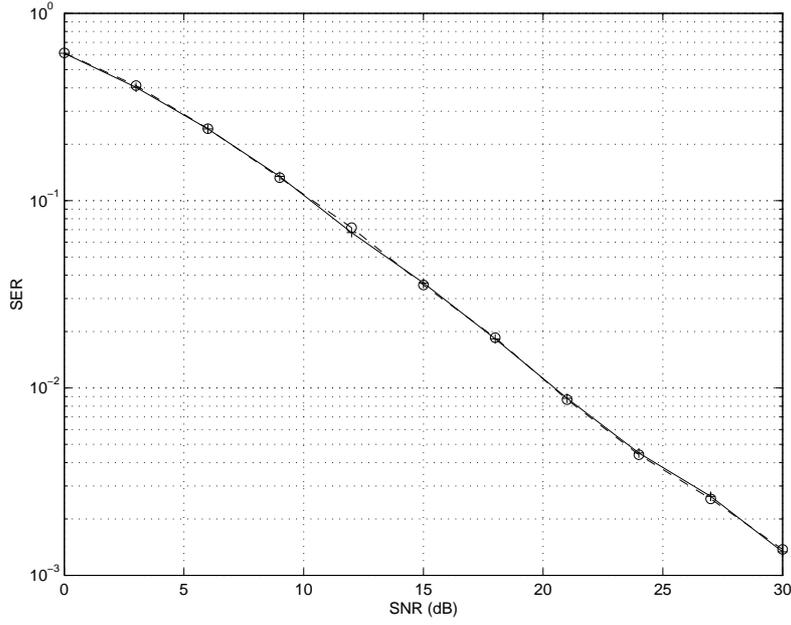


Figure 2.5: Category 1 - spatio-temporally white observation noise: $T = 8$, $M = 1$, $N = 1$, $K = 256$, $\mathbf{\Upsilon} = \mathbf{I}_{NT}$. Plus-solid curve: our codes; circle-dashed curve: unitary codes.

0.8 ; 0.6; 0.1]. We see that for $\text{SER} = 10^{-3}$ our codes demonstrate a gain of 2 dB when compared with the unitary codes.

Third category: $\mathbf{E} = \mathbf{s} \boldsymbol{\alpha}^T + \mathbf{E}_{\text{temp}}$

In the third category, we considered the case where the noise matrix is of the form $\mathbf{E} = \mathbf{s} \boldsymbol{\alpha}^T + \mathbf{E}_{\text{temp}}$. This models an interfering source \mathbf{s} (with known covariance matrix $\mathbf{\Upsilon}_s$) where the complex vector $\boldsymbol{\alpha}$ is the known channel attenuation between each receive antenna and the interfering source. The matrix \mathbf{E}_{temp} has a noise covariance matrix belonging to the second category. Thus, the noise covariance matrix is given by $\mathbf{\Upsilon} = \boldsymbol{\alpha} \boldsymbol{\alpha}^H \otimes \mathbf{\Upsilon}_s + \mathbf{I}_N \otimes \Sigma(\boldsymbol{\rho})$. As for the second category, we compare our codes adapted to this particular scenario with unitary codes. In figures 2.11–2.12 the solid curves represent performance of codes constructed by our method, while the dashed curves represent performance of unitary codes [27]. Figure 2.11 plots the result of the experiment for $T = 8$, $M = 2$, $N = 2$, $K = 32$, $\mathbf{s} = [1; 0.7; 0.4; 0.15; \text{zeros}(4,1)]$, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$ and $\boldsymbol{\alpha} = [-1.146 + 1.189i; 1.191 - 0.038i]$. For $\text{SER} = 10^{-3}$, once again our codes demonstrate a gain of more than 2 dB gain when compared with the unitary codes. Figure 2.12 plots the result of the experiment for $T = 8$, $M = 2$, $N = 2$, $K = 67$, $\boldsymbol{\rho} = [1; 0.7; 0.4; 0.15;$

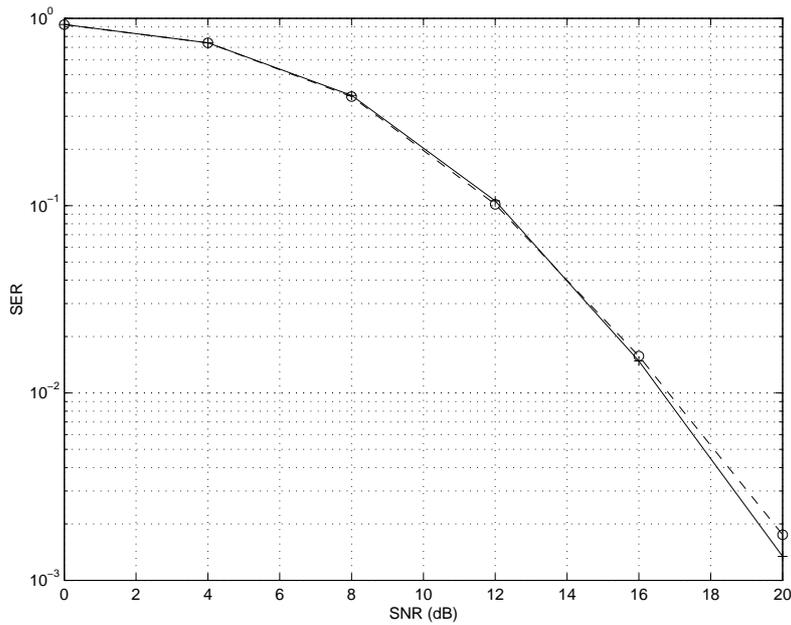


Figure 2.6: Category 1 - spatio-temporally white observation noise: $T = 4$, $M = 2$, $N = 2$, $K = 256$, $\mathbf{Y} = \mathbf{I}_{NT}$. Plus-solid curve: our codes; circle-dashed curve: unitary codes found in [32]. The codes use GLRT receiver.

zeros(4,1)], $\mathbf{s} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$ and $\boldsymbol{\alpha} = [-0.4534 + 0.0072i; 0.4869 + 1.9728i]$. For $\text{SER} = 10^{-3}$, our codes demonstrate a gain of more than 1.5 dB gain when compared with the unitary codes.

As a final remark, in the sequel, we offer a possible, intuitive explanation for the obtained results.

It is a known fact that correlation in noise is beneficial to overall system performance. Hence, even if one could employ unitary constellations, we strongly believe that it should not be done since, by doing it, we would “revert” to the white noise case and possible gain that originates from the fact that the noise is correlated would be lost. We believe that the solution is in somehow finding the way which, when constructing codebooks, explicitly takes into account the information about the noise correlation.

2.6 Conclusions

We addressed the problem of codebook construction for non-coherent communication in multiple-antenna wireless systems. In contrast with other related approaches, the Gaussian observation noise may have an arbitrary correlation structure. The non-coherent receiver

		PACKING RADII (DEGREES)	
T	K	MB	Rankin
6	7	80.41	80.41
6	8	77.06	77.40
6	9	75.52	75.52
6	10	74.20	74.21
6	11	73.22	73.22
6	12	72.45	72.45
6	13	71.82	71.83
6	14	71.31	71.32
6	15	70.87	70.89
6	16	70.53	70.53
6	17	70.10	70.21
6	18	69.73	69.94
6	19	69.40	69.70

Table 2.5: PACKING IN COMPLEX PROJECTIVE SPACE: We compare our best configurations (MB) of K points in $\mathbb{P}^{T-1}(\mathbb{C})$ against Rankin bound [28]. The packing radius of an ensemble is measured as the acute angle between the closest pair of lines.

operates according to the GLRT principle. A methodology for designing space-time codebooks for this non-coherent setup, taking the probability of error of the detector in the high SNR regime as the code design criterion, is proposed. We have presented a two-phase greedy approach to solve the resulting high-dimensional, nonlinear and non-smooth optimization problem. The first phase solves a convex SDP relaxation to obtain a suboptimal codebook. The second phase refines it through a geodesic descent optimization algorithm which efficiently exploits the Riemannian geometry of the constraints. Computer simulations show that our codebooks are marginally better than state-of-art known solutions for the special case of spatio-temporal white Gaussian observation noise but significantly outperform them in the correlated noise environments. This shows the relevance of the codebook construction tool proposed herein.

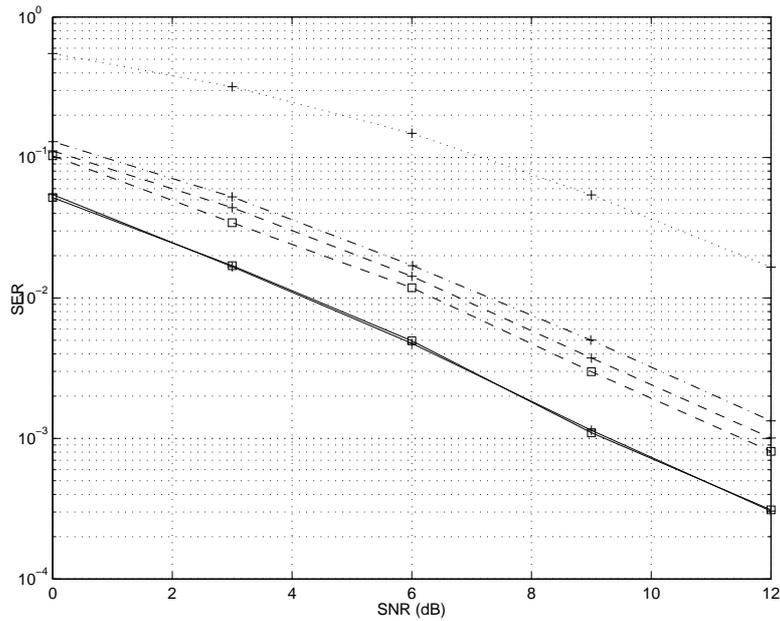


Figure 2.7: Category 2 - spatially white - temporally colored: $T = 8$, $M = 2$, $N = 1$, $K = 67$, $\boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.1; \text{zeros}(3,1)]$. Solid curves: our codes; dashed curves: unitary codes; dotted curve: codes obtained by the heuristic (2.12); dash-dotted curve: codes obtained by the heuristic (2.13); plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

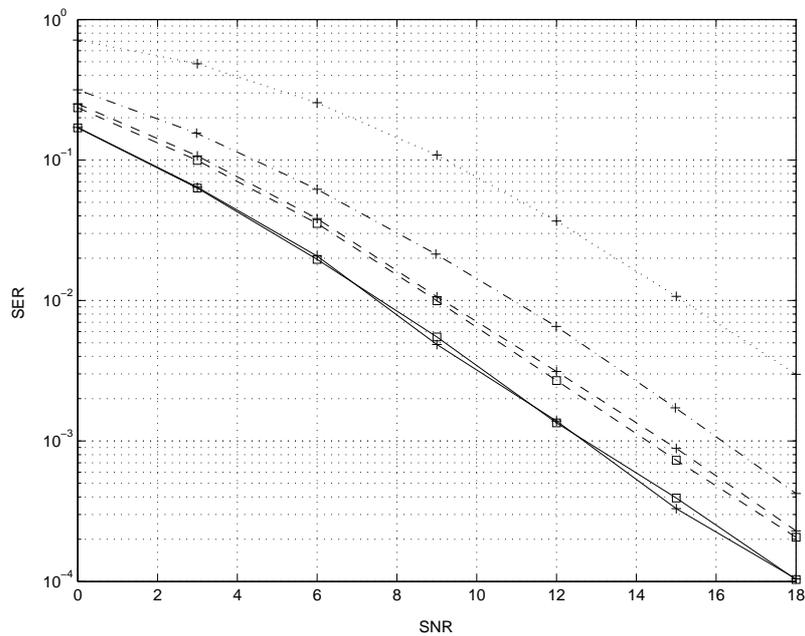


Figure 2.8: Category 2 - spatially white - temporally colored: $T = 8$, $M = 2$, $N = 1$, $K = 256$, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$. Solid curves: our codes; dashed curves: unitary codes; dotted curve: codes obtained by the heuristic (2.12); dash-dotted curve: codes obtained by the heuristic (2.13); plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

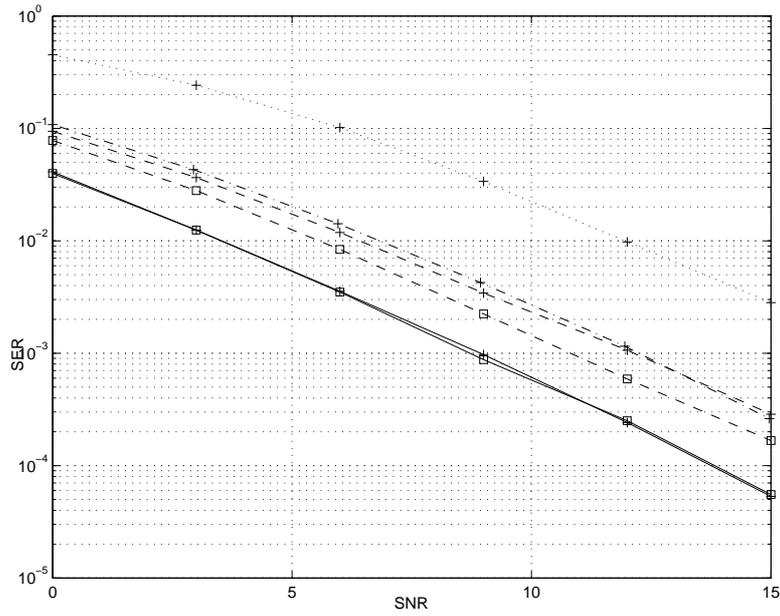


Figure 2.9: Category 2 - spatially white - temporally colored: $T = 8$, $M = 2$, $N = 1$, $K = 32$, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$. Solid curves: our codes; dashed curves: unitary codes; dotted curve: codes obtained by the heuristic (2.12); dash-dotted curve: codes obtained by the heuristic (2.13); plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

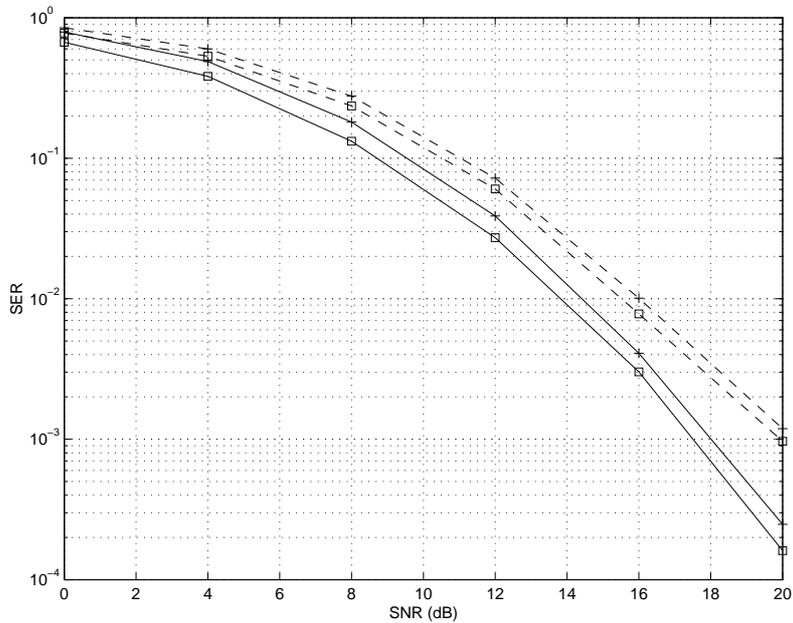


Figure 2.10: Category 2 - spatially white - temporally colored: $T = 4$, $M = 2$, $N = 2$, $K = 256$, $\boldsymbol{\rho} = [1; 0.8; 0.6; 0.1]$. Solid curves: our codes; dashed curves: unitary codes found in [32]; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

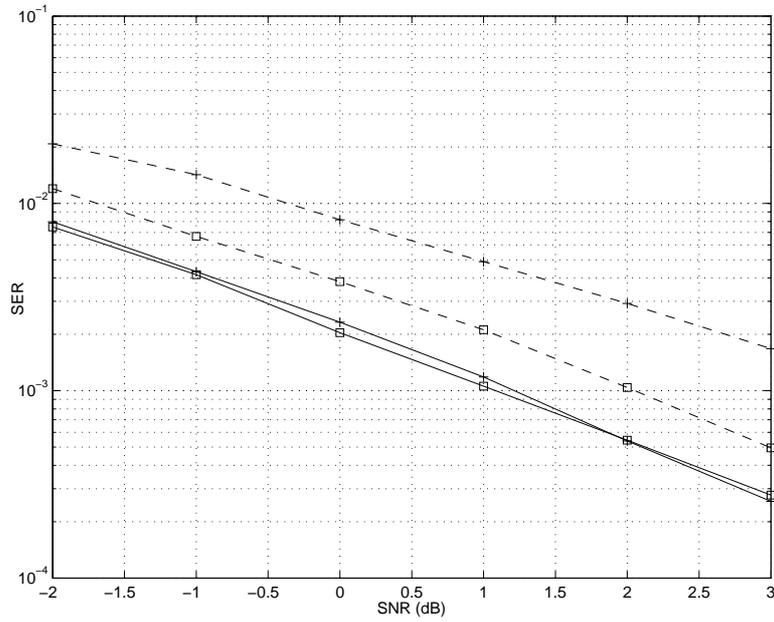


Figure 2.11: Category 3: $T = 8$, $M = 2$, $N = 2$, $K = 32$. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

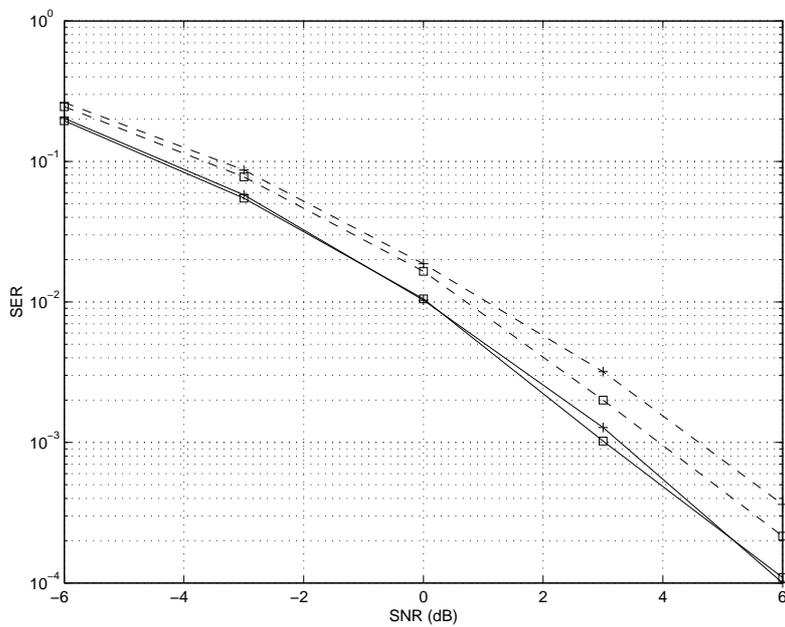


Figure 2.12: Category 3: $T = 8$, $M = 2$, $N = 2$, $K = 67$. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

Chapter 3

Capacity and Error Probability Analysis of Non-Coherent MIMO Systems in the Low SNR Regime

3.1 Chapter Summary

The chapter is organized as follows. In section 3.2, the spatially correlated non-coherent MIMO block Rayleigh fading channel is analyzed. The impact of channel and noise correlation on the mutual information is obtained for the on-off and Gaussian signaling. In section 3.3, contrary to most approaches for the low SNR regime, a low SNR analysis of the PEP for the GLRT receiver is introduced, and a codebook design criterion which takes into account the information about noise correlation is obtained. New space-time constellations for some particular wireless scenarios are constructed. Computer simulations show that these new codebooks are also of interest for Bayesian receivers which decode constellations with non-uniform priors. Section 3.4 contains the main conclusions of this chapter.

3.2 Random Fading Channel: the Low SNR Mutual Information Analysis

Data model and assumptions. We focus on a communication system comprising M transmit and N receive antennas over a narrowband flat Rayleigh fading channel. We assume a block fading channel model which is widely used in the MIMO literature [23, 45, 46, 66], with coherence interval T . In complex base band notation we have the system

model

$$\mathbf{Y} = \mathbf{X}\mathbf{H}^H + \mathbf{E}, \quad (3.1)$$

where \mathbf{X} is the $T \times M$ matrix of transmitted symbols, \mathbf{Y} is the $T \times N$ matrix of received symbols, \mathbf{H}^H is the $M \times N$ matrix of channel coefficients, and \mathbf{E} is the $T \times N$ matrix of zero-mean additive observation noise. For the sake of convenience, let $\mathcal{H} = \mathbf{H}^H$. We work under the following assumptions:

B1. (Channel matrix) The popular *separable (Kronecker) spatial correlation* model [38, 66, 67] is used, i.e., $\mathcal{H} = \sqrt{\frac{\rho}{M}} \mathbf{K}_t^{\frac{1}{2}} \mathbf{H}_w (\mathbf{K}_r^T)^{\frac{1}{2}}$ where \mathbf{H}_w is a $M \times N$ matrix comprised of statistically independent $\mathcal{CN}(0, 1)$ entries and the coefficient ρ is the model parameter proportional to the SNR. We assume that the matrix \mathbf{H}_w remains fixed for the coherence time T after which it changes to a new independent value. The correlation coefficients between the M (N) transmit (receive) antennas are assembled into an $M \times M$ ($N \times N$) positive semidefinite Hermitian correlation matrix \mathbf{K}_t (\mathbf{K}_r^T ; the operator T is used for the sake of convenience). The matrix \mathbf{H}_w is not known at the receiver nor at the transmitter, but its distribution is, in addition to \mathbf{K}_t and \mathbf{K}_r . This model is appropriate for the scenarios where only the objects surrounding the transmitter and the receiver cause the correlation of the local antenna elements, while they have no impact on the correlation at the other end of the link. The model has been found to be satisfactory in certain scenarios [68, 69]. We would also like to point out that there exist other spatial correlation models that take into account coupling between transmit and receive sides, see [70] and references therein. Although these models may characterize realistic channels more accurately for some scenarios (in these cases, the Kronecker model leads to capacity underestimations), in this work, we adopt the Kronecker model since we believe that it represents a good compromise between analytical tractability and validity of the channel. For a fair comparison of different correlation cases, we assume that $\text{tr}(\mathbf{K}_t) = M$ and $\text{tr}(\mathbf{K}_r) = N$;

B2. (Transmit power constraint) We impose the power constraint $\text{E}[\text{tr}(\mathbf{X}^H \mathbf{X})] \leq$

TM ;

B3. (Noise distribution) The noise covariance matrix $\mathbf{\Upsilon} = \mathbb{E}[\text{vec}(\mathbf{E}) \text{vec}(\mathbf{E})^H]$ is known at the transmitter and at the receiver. Also, without loss of generality (w.l.o.g.), we assume $\text{tr}(\mathbf{\Upsilon}) = NT$.

Note that in **B3**, we again let the data model depart from the customary assumption of spatio-temporal white Gaussian observation noise.

3.2.1 Mutual Information: On-Off Signaling

In [45], it has been demonstrated that the on-off signaling presented in [41], where the single transmit antenna systems were considered, generalizes to the multi-antenna setting and attains the ergodic channel capacity for the coherent case. Here, we show that this is also the case for the correlated Rayleigh fading channel model with arbitrary noise covariance matrix. Furthermore, we maximize the mutual information with respect to (w.r.t.) the input signal \mathbf{X}_{on} , \mathbf{K}_t and \mathbf{K}_r . Hence, we view both \mathbf{K}_t and \mathbf{K}_r as system parameters which we can introduce and track. The on-off signaling is defined as: for any $\epsilon > 1$ and assuming $\rho < 1$,

$$\mathbf{X} = \begin{cases} \mathbf{X}_{on} \rho^{-\frac{\epsilon}{2}} & \text{with probability (w.p.) } \rho^\epsilon \\ \mathbf{0}_{T \times M} & \text{w.p. } 1 - \rho^\epsilon \end{cases}$$

With an analysis similar to [45], presented in Appendix D, it can be shown that at sufficiently low SNR the mutual information between \mathbf{Y} and \mathbf{X} up to first order in ρ is given by

$$I(\mathbf{Y}; \mathbf{X}) = \frac{\rho}{M} \text{tr}(\mathbf{\Upsilon}^{-1} (\mathbf{K}_r \otimes \mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H)) + o(\rho). \quad (3.2)$$

Note that for the special case of spatio-temporal white observation noise and uncorrelated Rayleigh fading channel case, i.e., $\mathbf{\Upsilon} = \mathbf{I}_{TN}$, $\mathbf{K}_t = \mathbf{I}_M$ and $\mathbf{K}_r = \mathbf{I}_N$ our result in (3.2) recovers the finding in [45]. In that case, the maximal mutual information (per channel use) is equal to

$$\frac{1}{T} I(\mathbf{Y}; \mathbf{X}) = \rho N + o(\rho).$$

Now, we address the maximization of the mutual information w.r.t. \mathbf{X}_{on} , \mathbf{K}_t and \mathbf{K}_r , i.e.,

$$\begin{aligned} & \max \quad \text{tr}(\mathbf{\Upsilon}^{-1}(\mathbf{K}_r \otimes \mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H)) \\ & \text{tr}(\mathbf{X}_{on} \mathbf{X}_{on}^H) \leq TM, \\ & \mathbf{K}_t \in \mathcal{P}_M, \mathbf{K}_r \in \mathcal{P}_N \end{aligned} \quad (3.3)$$

where

$$\mathcal{P}_n = \{\mathbf{Q} : n \times n \text{ matrix such that } \mathbf{Q} \succeq \mathbf{0} \text{ and } \text{tr}(\mathbf{Q}) = n\}. \quad (3.4)$$

In Appendix E we show that the maximum in (3.3) is attained by

$$\widehat{\mathbf{X}}_{on} = \sqrt{TM} [\hat{\mathbf{x}} \quad \mathbf{0}_{T \times (M-1)}], \widehat{\mathbf{K}}_r = N \hat{\mathbf{u}} \hat{\mathbf{u}}^H, [\widehat{\mathbf{K}}_t]_{i,i} = M \delta_{i1} \quad (3.5)$$

where

$$\begin{aligned} (\hat{\mathbf{u}}, \hat{\mathbf{x}}) = & \arg \max \quad (\mathbf{u} \otimes \mathbf{x})^H \mathbf{\Upsilon}^{-1} (\mathbf{u} \otimes \mathbf{x}) \\ & \mathbf{u} \in \mathbb{C}^N, \|\mathbf{u}\| = 1, \\ & \mathbf{x} \in \mathbb{C}^T, \|\mathbf{x}\| = 1 \end{aligned} \quad (3.6)$$

with $\delta_{ij} = 1$ for $i = j$ and zero otherwise. The notation $[\mathbf{M}]_{i,i}$ represents the entry of the matrix \mathbf{M} on the position (i, i) . Note that $\widehat{\mathbf{K}}_t$ is a diagonal matrix. The optimization problem in (3.6) always admits a solution (maximization of a continuous function over a compact set) but, in general, a closed form solution is not available. The exception is the case when the noise covariance matrix $\mathbf{\Upsilon}$ has a Kronecker structure, say $\mathbf{\Upsilon} = \mathbf{\Upsilon}_1 \otimes \mathbf{\Upsilon}_2$ for some $N \times N$ matrix $\mathbf{\Upsilon}_1$ and $T \times T$ matrix $\mathbf{\Upsilon}_2$. In that situation, the optimal $\hat{\mathbf{u}}$ (resp. $\hat{\mathbf{x}}$) can be taken as any unit-norm eigenvector associated with the minimal eigenvalue of $\mathbf{\Upsilon}_1$ (resp. $\mathbf{\Upsilon}_2$). For example, in the case of spatially white-temporally colored observation noise, i.e., $\mathbf{\Upsilon} = \mathbf{I}_N \otimes \mathbf{\Sigma}$ for some $T \times T$ positive definite matrix $\mathbf{\Sigma}$, the vector $\hat{\mathbf{x}}$ is the unit-norm eigenvector associated with the minimal eigenvalue of $\mathbf{\Sigma}$ (in other words, we transmit the codeword in the direction that is least affected by the noise). For the choice in (3.5), the maximal mutual information (per channel use) is equal to

$$\frac{1}{T} I(\mathbf{Y}; \mathbf{X}) = \rho N M \hat{\lambda} + o(\rho). \quad (3.7)$$

where $\hat{\lambda} = (\hat{\mathbf{u}} \otimes \hat{\mathbf{x}})^H \mathbf{\Upsilon}^{-1} (\hat{\mathbf{u}} \otimes \hat{\mathbf{x}})$.

Remarks. From (3.5) it is clear that both the transmit and receive antennas should be made as correlated as possible, as both the optimal \mathbf{K}_t and \mathbf{K}_r have rank one. Note

that in (3.7) the mutual information is proportional to the number of transmit antennas M . This is in sharp contrast with the case of uncorrelated Rayleigh fading channel model for which it has been shown that the maximal mutual information is independent of the number of transmit antennas [45]. Also, since

$$\text{tr}(\mathbf{\Upsilon}^{-1}) = \sum_{i=1}^{NT} 1/\lambda_i \geq \sum_{i=1}^{NT} (2 - \lambda_i) = NT,$$

where λ_i 's are the eigenvalues of $\mathbf{\Upsilon}$, we can w.l.o.g. assume that, e.g., $[\mathbf{\Upsilon}^{-1}]_{1,1} \geq 1$ (otherwise $\text{tr}(\mathbf{\Upsilon}^{-1}) < NT$). Then, by choosing $\mathbf{u}_1 = [1 \ \mathbf{0}_{1 \times (N-1)}]^T$ and $\mathbf{x}_1 = [1 \ \mathbf{0}_{1 \times (T-1)}]^T$ we have

$$\hat{\lambda} \geq (\mathbf{u}_1 \otimes \mathbf{x}_1)^H \mathbf{\Upsilon}^{-1} (\mathbf{u}_1 \otimes \mathbf{x}_1) = [\mathbf{\Upsilon}^{-1}]_{1,1} \geq 1.$$

This result confirms the general principle that correlated noise is beneficial from the capacity point of view. See, e.g., pp. 100 in [18] for more details. In practice, by changing the antenna separation one can control the eigenvalues of \mathbf{K}_t and \mathbf{K}_r , but not their eigenvectors. See [46, 66, 71, 72], [18, section 6.3.3]. As a consequence, the result presented herein has to be interpreted as the upper bound on the channel capacity. For guidelines for optimizing antenna spacing in the case when there is little to no local scattering around the base station and the mobile is assumed to be in a rich scattering environment, see [72]. For the physical conditions on the antenna spacing under which the Kronecker model is appropriate, the reader is referred to pp. 98-100 in [71] for more details. Nevertheless, the previous conclusion holds for the case when we can not manipulate the eigenvectors of \mathbf{K}_r (which are still available to the transmitter through a feedback link): in this case, from (E.11) we obtain

$$\frac{1}{T} I(\mathbf{Y}; \mathbf{X}) = \rho N M \lambda_{\max} \left(\mathbf{F}_{i^*} \hat{\mathbf{\Upsilon}}^{-1} \mathbf{F}_{i^*}^H \right) + o(\rho)$$

where $\hat{\mathbf{\Upsilon}}$, \mathbf{F}_i and i^* are defined in (E.5), (E.5) and (E.8), respectively. It is clear that $\lambda_{\max} \left(\mathbf{F}_{i^*} \hat{\mathbf{\Upsilon}}^{-1} \mathbf{F}_{i^*}^H \right) \geq 1$ (otherwise $\text{tr}(\mathbf{\Upsilon}^{-1}) < NT$). Thus, correlated noise is beneficial from the capacity point of view in this case too. A short exercise would show that the first order term in (3.7) corresponds to that of the capacity when the channel is known to the receiver and the noise covariance matrix $\mathbf{\Upsilon}$ is arbitrary (when $\mathbf{\Upsilon} = \mathbf{I}_{NT}$, then

$\hat{\lambda} = 1$, from (3.7) we retrieve the result on pp. 94 in [18] where the coherent correlated Rayleigh fading channel has been treated). With coherent reception, we know that the mutual information is maximized if the input $\text{vec}(\mathbf{X})$ is circularly symmetric, complex Gaussian distribution, i.e., $\text{vec}(\mathbf{X}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{P}_{coh})$ for some covariance matrix \mathbf{P}_{coh} with $\text{tr}(\mathbf{P}_{coh}) \leq TM$ such that the power constraint in the assumption **B2** is satisfied (note that we may assume w.l.o.g. that $\text{vec}(\mathbf{X})$ is zero-mean). In that case, the channel capacity is given by

$$\mathbb{E}_{\mathbf{h}_w} \left[\log_2 \det \left(\mathbf{I}_{TN} + \frac{\rho}{M} \mathbf{\Upsilon}^{-\frac{1}{2}} (\mathcal{L} \otimes \mathbf{I}_T) \mathbf{P}_{coh} (\mathcal{L}^H \otimes \mathbf{I}_T) \mathbf{\Upsilon}^{-\frac{1}{2}} \right) \right] \quad (3.8)$$

where $\mathcal{L} = \mathbf{K}_r^{\frac{1}{2}} \mathbf{H}_w^T (\mathbf{K}_t^T)^{\frac{1}{2}}$ and $\mathbf{h}_w = \text{vec}(\mathbf{H}_w)$. It can be readily shown (by maximizing the first-order expansion of (3.8) w.r.t. \mathbf{P}_{coh} , \mathbf{K}_t and \mathbf{K}_r) that the optimal covariance matrix $\hat{\mathbf{P}}_{coh}$ is equal to

$$\hat{\mathbf{P}}_{coh} = \text{vec}(\hat{\mathbf{X}}_{on}) \text{vec}^H(\hat{\mathbf{X}}_{on}), \quad (3.9)$$

with $\hat{\mathbf{X}}_{on}$, $\hat{\mathbf{K}}_t$ and $\hat{\mathbf{K}}_r$ defined as in (3.5).

3.2.2 Mutual Information: Gaussian Modulation

In this subsection, we compute the low SNR mutual information for the more realistic and practical case of Gaussian modulation. Let $\mathbf{x} = \text{vec}(\mathbf{X})$ be a zero-mean random variable with covariance matrix \mathbf{P} that follows a circularly symmetric, complex Gaussian distribution, i.e., $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{P})$. Clearly, in order to meet the power constraint in the assumption **B2**, $\text{tr}(\mathbf{P}) \leq TM$. Then, at sufficiently low SNR, the mutual information between \mathbf{Y} and \mathbf{X} up to second order in ρ is given by

$$I(\mathbf{Y}; \mathbf{X}) = \frac{\rho^2}{2M^2} \text{tr} \left(\mathbb{E}[\mathbf{Z}^2] - (\mathbb{E}[\mathbf{Z}])^2 \right) + o(\rho^2) \quad (3.10)$$

$$= \frac{\rho^2}{2M^2} \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^N \sum_{z=1}^N \lambda_k \lambda_z \text{tr}(\hat{\mathbf{\Upsilon}}_{kz} \tilde{\mathbf{P}}_{ij}) \text{tr}(\hat{\mathbf{\Upsilon}}_{zk} \tilde{\mathbf{P}}_{ji}) + o(\rho^2), \quad (3.11)$$

where

$$\mathbf{Z} = \mathbf{\Upsilon}^{-\frac{1}{2}} (\mathbf{K}_r \otimes \mathbf{X} \mathbf{K}_t \mathbf{X}^H) \mathbf{\Upsilon}^{-\frac{1}{2}}, \quad \tilde{\mathbf{P}} = \left((\mathbf{K}_t^T)^{\frac{1}{2}} \otimes \mathbf{I}_T \right) \mathbf{P} \left((\mathbf{K}_t^T)^{\frac{1}{2}} \otimes \mathbf{I}_T \right),$$

$$\tilde{\mathbf{P}}_{ij} = \mathbf{E}_i \tilde{\mathbf{P}} \mathbf{E}_j^H, \quad \hat{\mathbf{\Upsilon}}_{kz} = \mathbf{F}_k \hat{\mathbf{\Upsilon}}^{-1} \mathbf{F}_z^H$$

and λ_k 's, for $k = 1, \dots, N$, are the eigenvalues of \mathbf{K}_r . The matrices $\widehat{\mathbf{\Upsilon}}$ and \mathbf{F}_k , for $k = 1, \dots, N$, are defined in (E.5), whereas the $T \times TM$ matrix \mathbf{E}_i , for $i = 1, \dots, M$, is given by

$$\mathbf{E}_i = \mathbf{e}_i^T \otimes \mathbf{I}_T, \quad (3.12)$$

where \mathbf{e}_i represents the i -th column of the identity matrix \mathbf{I}_M . The proof is given in Appendix F. We now address the optimization problem

$$\begin{aligned} \max_{\substack{\mathbf{P} \succeq \mathbf{0}, \text{tr}(\mathbf{P}) \leq TM, \\ \mathbf{K}_t \in \mathcal{P}_M, \mathbf{K}_r \in \mathcal{P}_N}} & \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^N \sum_{z=1}^N \lambda_k \lambda_z \text{tr}(\widehat{\mathbf{\Upsilon}}_{kz} \widetilde{\mathbf{P}}_{ij}) \text{tr}(\widehat{\mathbf{\Upsilon}}_{zk} \widetilde{\mathbf{P}}_{ji}). \end{aligned} \quad (3.13)$$

It can be shown that the maximum of (3.13) is attained by the following signaling scheme: the optimal correlation matrices $\widehat{\mathbf{K}}_r$ and $\widehat{\mathbf{K}}_t$ are defined as in (3.5), and the optimal covariance matrix $\widehat{\mathbf{P}}$ is given by

$$\widehat{\mathbf{P}} = TM \mathbf{K}_P \otimes \hat{\mathbf{x}} \hat{\mathbf{x}}^H \quad (3.14)$$

where the vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{x}}$ are, as before, solutions of the optimization problem (3.6). The $M \times M$ constant matrix \mathbf{K}_P has all the entries equal to zero except the entry (1,1) which is one. The proof is given in Appendix G. In this case, the mutual information (per channel use) is given by

$$\frac{1}{T} I(\mathbf{Y}; \mathbf{X}) = \frac{\rho^2}{2} N^2 T M^2 \hat{\lambda}^2 + o(\rho^2). \quad (3.15)$$

where $\hat{\lambda} = (\hat{\mathbf{u}} \otimes \hat{\mathbf{x}})^H \mathbf{\Upsilon}^{-1} (\hat{\mathbf{u}} \otimes \hat{\mathbf{x}})$.

Remarks. In [45] it has been proved that for the uncorrelated Rayleigh fading channel only one transmit antenna should be employed. Here, we see from (3.15) that having more transmit (M) and receive (N) antennas can actually enhance the channel performance in terms of capacity significantly in the correlated setup. We see that the mutual information is proportional to $M^2 N^2$, whereas in [45] the increase is only linear in the number of the receive antennas. Hence, by making the antennas as correlated as possible the total gain is $M^2 N$. Remark that although the mathematics for calculating $\widehat{\mathbf{P}}$, $\widehat{\mathbf{K}}_t$ and $\widehat{\mathbf{K}}_r$ are quite involved, our findings are not surprising since we see that the optimal values

correspond to those of the coherent correlated Rayleigh fading channel case (it is easy to check that $\widehat{\mathbf{P}} = \widehat{\mathbf{P}}_{coh}$ defined in (3.9)). The conclusions herein presented are in concordance with [46, 66] and with the results of the previous subsection where it has been shown that channel correlation and correlated noise can actually improve the channel performance.

3.3 Deterministic Fading Channel: the Low SNR PEP Analysis

Data model and assumptions. We retain the data model (3.1), but the presumptions under which we work are the following:

- C1. (Channel matrix)** The matrix \mathcal{H} is not known at the receiver neither at the transmitter, and no stochastic model is assumed for it;
- C2. (Transmit power constraint)** The codeword \mathbf{X} is chosen from a finite codebook $\mathcal{C} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$ known to the receiver, where K is the size of the codebook. We impose the power constraint $\text{tr}(\mathbf{X}_k^H \mathbf{X}_k) = 1$ for each codeword. We further assume that each codeword is of full rank;
- C3. (Noise distribution)** As in the assumption **B3**, the noise covariance matrix $\mathbf{\Upsilon} = \text{E}[\text{vec}(\mathbf{E}) \text{vec}(\mathbf{E})^H]$ is known at the transmitter and at the receiver.

Receiver. Under the above conditions, the conditional pdf of the received vector $\mathbf{y} = \text{vec}(\mathbf{Y})$, given the transmitted matrix \mathbf{X} and the unknown realization of the channel $\mathbf{g} = \text{vec}(\mathcal{H})$, is given by

$$p(\mathbf{y}|\mathbf{X}, \mathbf{g}) = k \exp\{-\|\mathbf{y} - (\mathbf{I}_N \otimes \mathbf{X})\mathbf{g}\|_{\mathbf{\Upsilon}^{-1}}^2\},$$

where $k = 1/(\pi^{TN} \det(\mathbf{\Upsilon}))$ and the notation $\|\mathbf{z}\|_{\mathbf{A}}^2 = \mathbf{z}^H \mathbf{A} \mathbf{z}$ is used. Since no stochastic model is attached to the channel propagation matrix, the receiver faces a multiple hypothesis testing problem with the channel \mathcal{H} as a deterministic nuisance parameter. Hence, we shall assume a GLRT receiver. The GLRT [51] is composed of a bank of K parallel processors where the k -th processor assumes the presence of the k -th codeword and computes the likelihood of the observation, after replacing the channel by its ML estimate.

The GLRT detector chooses the codeword associated with the processor exhibiting the largest likelihood of the observation, i.e.,

$$\hat{k} = \operatorname{argmax}\{p(\mathbf{y}|\mathbf{X}_k, \hat{\mathbf{g}}_k) : k = 1, 2, \dots, K\}$$

where

$$\hat{\mathbf{g}}_k = (\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k)^{-1} \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\Upsilon}^{-\frac{1}{2}} \mathbf{y}$$

with $\boldsymbol{\mathcal{X}}_k = \boldsymbol{\Upsilon}^{-\frac{1}{2}} (\mathbf{I}_N \otimes \mathbf{X}_k)$ represents the ML estimate of the channel. Due to the respective expression for $\hat{\mathbf{g}}_k$, we note that since each codeword of the codebook has full rank (presumption **C2**), the ML channel estimate is well defined.

Low SNR analysis. In the sequel, a low SNR analysis of the PEP is introduced. Let $P_{\mathbf{X}_i \rightarrow \mathbf{X}_j}$ be the probability of the GLRT receiver deciding \mathbf{X}_j when \mathbf{X}_i is sent. It can be shown that for $T \geq 2M$

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \approx P(Y > \mathbf{g}^H \mathbf{L}_{ij} \mathbf{g}), \quad (3.16)$$

with

$$\mathbf{L}_{ij} = \boldsymbol{\mathcal{X}}_i^H \boldsymbol{\Pi}_j^\perp \boldsymbol{\mathcal{X}}_i, \boldsymbol{\Pi}_j^\perp = \mathbf{I}_{TN} - \boldsymbol{\mathcal{X}}_j (\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j)^{-1} \boldsymbol{\mathcal{X}}_j^H$$

and $Y = \sum_{m=1}^{MN} \sin \alpha_m (|a_m|^2 - |b_m|^2)$ where a_m, b_m are iid circular complex Gaussian random variables with zero mean and unit variance, i.e., $a_m, b_m \stackrel{iid}{\sim} \mathcal{CN}(0, 1)$ for $m = 1, \dots, MN$. The angles α_m are the *principal angles* between the subspaces spanned by $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{X}}_j$. The proof is given in Appendix H. For the case of spatio-temporal white observation noise, i.e., $\boldsymbol{\Upsilon} = \mathbf{I}_{NT}$ and $M = 1$, from (3.16) we have

$$P_{\mathbf{x}_i \rightarrow \mathbf{x}_j} \approx P\left(\sum_{i=1}^N (|a_i|^2 - |b_i|^2) > \|\mathbf{g}\|^2 \sin \alpha_1\right), \quad (3.17)$$

where we assume $\sin \alpha_1 \neq 0$ (remark that for $\boldsymbol{\Upsilon} = \mathbf{I}_{NT}$ there are maximum M different principal angles where each of them is of multiplicity N). In Chapter 2 we derive the expression for the PEP in the high SNR regime and $T \geq 2M$. For $M = 1$ and $\boldsymbol{\Upsilon} = \mathbf{I}_{NT}$, it is given by

$$P_{\mathbf{x}_i \rightarrow \mathbf{x}_j} = \mathcal{Q}\left(\frac{1}{\sqrt{2}} \|\mathbf{g}\| \sin \alpha_{ij}\right). \quad (3.18)$$

Equations (3.17) and (3.18) show that the probability of misdetecting \mathbf{x}_i for \mathbf{x}_j depends on the channel \mathbf{g} , but more important, on the relative geometry of the codewords \mathbf{x}_i

and \mathbf{x}_j . Since $P_{\mathbf{x}_i \rightarrow \mathbf{x}_j} = P_{\mathbf{x}_j \rightarrow \mathbf{x}_i}$ (a feature of the scenario $M = 1$ and $\mathbf{\Upsilon} = \mathbf{I}_{NT}$), the PEPs are symmetric which gives rise to a intuitive distance measure. Hence, by analyzing the PEP in both extreme cases (low and high SNR) it is clear that one wishes to make the codewords \mathbf{x}_i and \mathbf{x}_j as separate as possible, i.e., the problem of finding good codes corresponds to the very well known packing problem in the complex projective space [28, 74]. Unfortunately, from (3.16) it seems difficult to propose a codebook design criteria for $M > 1$ and $\mathbf{\Upsilon} \neq \mathbf{I}_{TN}$. One of the reasons originates from the fact that PEPs are not symmetric for this general case. Hence, as usually, we resort to an upper bound on the PEP. From (3.16), an upper bound on the PEP is readily derived

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \leq P(Z > \|\mathbf{g}\|^2 \lambda_{\min}(\mathbf{L}_{ij})), \quad (3.19)$$

where $Z = \sum_{m=1}^{MN} |a_m|^2$. The bound in (3.19) is admittedly loose, but allows us to come up with a workable codebook design criterion. The simulation results below will assess its effectiveness. By invoking the second part of the theorem on pp. 200 in [52], the case when $M \leq T < 2M$, and then by repeating the analysis of the case $T \geq 2M$ presented in Appendix H, it is straightforward to see that the matrix \mathbf{L}_{ij} is rank deficient. This can seriously effect the error performance of the system since, by interpreting (3.16), one wants to maximize $\mathbf{g}^H \mathbf{L}_{ij} \mathbf{g}$. Thus, as in the high SNR regime and GLRT receiver, when designing constellation for arbitrary $\mathbf{\Upsilon}$ and the low SNR regime, we take $T \geq 2M$. Also, remark that for $M \leq T < 2M$ the bound in (3.19) is not applicable since $\lambda_{\min}(\mathbf{L}_{ij}) = 0$.

Codebook construction methodology. Denoting a codebook by $\mathcal{C} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$ we are led to the following optimization problem

$$\mathcal{C}^* = \underset{\mathcal{C} \in \mathcal{M}}{\operatorname{arg\,max}} f(\mathcal{C}) \quad (3.20)$$

where

$$f : \mathcal{M} \rightarrow \mathbb{R}, \mathcal{C} = \{\mathbf{X}_1, \dots, \mathbf{X}_K\} \mapsto f(\mathcal{C})$$

and

$$f(\mathcal{C}) = \min\{f_{ij}(\mathcal{C}) : 1 \leq i \neq j \leq K\}$$

with $f_{ij}(\mathcal{C}) = \lambda_{\min}(\mathbf{L}_{ij})$. A codebook \mathcal{C} is a point in the space

$$\mathcal{M} = \{(\mathbf{X}_1, \dots, \mathbf{X}_K) : \text{tr}(\mathbf{X}_k^H \mathbf{X}_k) = 1\}.$$

Remark that the codebook design criterion in (3.20) is equivalent to the one for the high SNR regime that has been proposed in Chapter 2, see (2.11). Consequently, the algorithm presented in table 2.1 will be employed to construct codebooks for some particular wireless scenarios.

3.3.1 Results

Using the codebook construction criterion methodology in (2.11), we have constructed codes for three special categories of noise covariance matrices $\mathbf{\Upsilon}$. If not stated otherwise, in all simulations we assume uncorrelated Rayleigh fading model for the channel matrix, i.e., $h_{ij} \stackrel{iid}{\sim} \mathcal{CN}(0, \sigma^2)$.

First category: spatio-temporal white observation noise

We are not aware of any work concerning the low SNR non-coherent MIMO scenario employing a GLRT receiver. Hence, we shall compare the performance of our codes and our GLRT receiver with the codes assuming a Rayleigh fading channel with equally probable codewords [39] and ML receiver. We also show that our codes are of great interest for the constellations with unequal priors [47, 48].

- Constellations with equal priors for $M = 1$. In figure 3.1 we compared our codes and our GLRT receiver against the codes found in [39] with the ML receiver proposed therein. We considered the cases where the coherence interval $T = 2$, SNR=7 dB and codebooks with $K = 8$ and $K = 16$ codewords. The solid and dashed curves represent our codes, and Borran codes respectively. As we can see for $K = 8$, although the Borran's codes assume the knowledge of actual $\text{SNR} = \mathbb{E}[\|\mathbf{x}_k \mathbf{h}^H\|^2] / \mathbb{E}[\|\mathbf{E}\|^2] = 7$ dB, our codebook constructions can save up to 3 receive antennas at symbol error rate (SER) of $2 \cdot 10^{-3}$. The same figure plots the results of a similar experiment for $K = 16$. It can be seen that for $\text{SER} = 2 \cdot 10^{-2}$, our codes demonstrate a saving of 6 receive antennas when compared with Borran's codes.

- Constellations with equal priors and $M > 1$. We present some results to study the impact of employing $M > 1$ transmit antennas in the low SNR regime. First, we compare our codebook constructions obtained by the method presented in [73, 74] for $M = 1$, $M = 2$ and $M = 3$. Figure 3.2 shows the result of the performance comparisons of our 256-point constellations for $T = 8$ and SNR = 0 dB. It can be seen that for SER = $2 \cdot 10^{-3}$, our codes for $M = 1$ can spare 1 receive antennas when comparing with our codes constructed for $M = 2$, and nearly 4 receive antennas compared with our codes constructed for $M = 3$. Figure 3.3 plots the result of the experiment for $T = 8$, SNR = -6 dB, $K = 32$ and $M = 1, 2$. We see that at SER = $2 \cdot 10^{-2}$, our codes for $M = 1$ demonstrate a saving of 8 receive antennas when compared with our codes constructed for $M = 2$. The same plot presents the result of the experiment for $T = 8$, SNR = -6 dB, $K = 67$ and $M = 1, 2$. For SER = $4 \cdot 10^{-2}$, our codes for $M = 1$ can spare 9 receive antennas compared with our codes constructed for $M = 2$. Then, we compare our codebook constructions for $M = 1$ against Borran's codes with $M = 2$. Figure 3.4 plots the result of the experiment. The solid signed and the solid circled curve show the performance of our codes for $K = 32$, $T = 4$, $M = 1$, and $K = 16$, $T = 3$, $M = 1$, respectively. The dashed signed and the dashed circled curve represent the performances of the Borran's codes for $K = 32$, $T = 4$, $M = 2$ and $K = 16$, $T = 3$, $M = 2$, respectively. For 32-point constellations, we see that our codes can save 7 receive antennas at SER = $4 \cdot 10^{-2}$. For 16-point constellations, we witness the gain of more than 10 receive antennas at SER = 10^{-1} . We think that the results presented in the figures 3.2- 3.4 further strengthen the motivation of using a single transmit antenna codebooks in the low SNR regime when GLRT receiver is employed.

- Constellations with unequal priors. Now, we depart from our GLRT receiver and show that our codebook designs for $M = 1$ are nevertheless of interest for schemes that allow for non-uniform priors, e.g., the Bayesian receiver in [47, 48]. In figure 3.5 we show the results of the simulations. We considered the case where the coherence interval $T = 2$, $P = 0.5$ and rate = 1 bps/Hz, with $P = \text{E}[\text{tr}(\mathbf{X}_k^H \mathbf{X}_k)]$. For simplicity, we assume $\mathbf{K}_r = \mathbf{I}_N$. We consider codes with codewords of the form $\mathbf{X}_k = [\mathbf{x}_k \quad \mathbf{0}_{T \times (M-1)}]$ since this form of the code resembles the capacity achieving distribution at sufficiently

low SNR presented in section 3.2. We call them single beam constellations. First, we treat the case of uncorrelated Rayleigh fading channel, i.e., $\mathbf{K}_t = \mathbf{I}_M$ where $M = 3$. Note that in this case only one transmit antenna is effectively used. The dashed and dashed-circled curves represent our codes, and Srinivasan's 5 point constellations with unequal priors [47] (the constellations assume a point in the origin with probability $\frac{1}{2}$, with the probabilities of the points lying in the sphere being equal). Next, we assume the correlated Rayleigh fading case with $\mathbf{K}_t = \widehat{\mathbf{K}}_t$ with $\widehat{\mathbf{K}}_t$ defined in section 3.2 (when referring to $\mathbf{K}_t = \widehat{\mathbf{K}}_t$ case, we simply write $\text{rank}(\mathbf{K}_t)=1$). The solid and solid-circled curves represent our codes, and Srinivasan's 5 point constellations with unequal priors. As expected, high improvements are possible when codes are used in correlated MIMO scenarios which is in concordance with the information-theoretic result presented herein. The gain of our 5 point constellations with unequal priors compared with Srinivasan's codes is due to the fact that we use optimal packings in complex projective space (in the outer sphere), whereas Srinivasan uses optimal packings in the real projective space (one can expect larger gains as K increases, where K represents the number of the codewords on the sphere). The improvement obtained can be explained by the optimality of our designed packings. Rankin bound is an upper bound on the packing radius of K subspaces in the Grassmanian space $G(M, \mathcal{C}^T)$. When $M = 1$, the bound applies to packings in the complex projective space, and in this case it holds

$$\min \{\sin^2 \alpha_{ij} : 1 \leq i \neq j \leq K\} \leq \frac{T-1}{T} \frac{K}{K-1}$$

where α_{ij} is the acute angle between codewords \mathbf{x}_i and \mathbf{x}_j . Please refer to [28] for more details. One can easily check that our designed codebook indeed meets the Rankin bound which is $\frac{2}{3}$ for $T = 2$ and $K = 4$. Our codebook is represented in the following matrix

$$\begin{bmatrix} 0.4946 - 0.6268i & -0.2375 + 0.5533i \\ -0.8183 - 0.4446i & -0.3392 + 0.1328i \\ 0.4908 - 0.4101i & 0.7326 + 0.2329i \\ -0.0955 - 0.2776i & -0.8817 + 0.3693i \end{bmatrix}.$$

The dash-dotted curve represents our 4 point constellation with equal priors and $M = 1$, and is plotted only to confirm that if the receiver knows the channel statistics, then constellations with non-uniform priors are the best option.

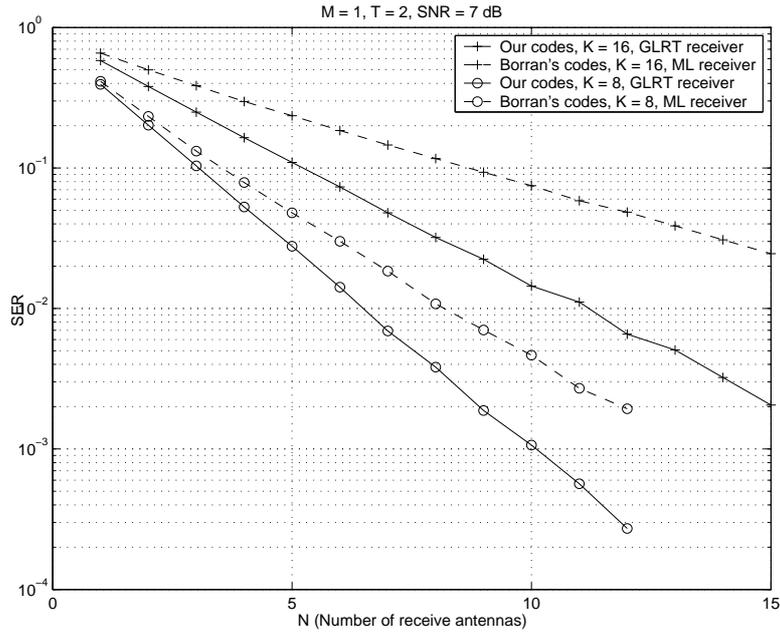


Figure 3.1: Category 1 - spatio-temporal white observation noise: Solid signed curve-our codes for $K = 16$, $T = 2$, $M = 1$, dashed signed curve-Borran's codes for $K = 16$, $T = 2$, $M = 1$, solid circled curve-our codes for $K = 8$, $T = 2$, $M = 1$, dashed circled curve-Borran's codes for $K = 8$, $T = 2$, $M = 1$.

Second category: spatially white-temporally colored observation noise

The second category corresponds to spatially white-temporally colored observation noise, i.e., $\mathbf{Y} = \mathbf{I}_N \otimes \Sigma(\boldsymbol{\rho})$ where the vector $\boldsymbol{\rho} : T \times 1$ is the first column of the Toeplitz matrix $\Sigma(\boldsymbol{\rho})$. To the best of our knowledge, we are not aware of any work that treats the problem of codebook constructions in the presence of spatially white-temporally colored observation noise for the low SNR regime. Hence, we compare our codes designed (adapted) to this specific scenario with codes designed when the presence of spatio-temporal white observation noise is assumed. The goal here is to demonstrate the increase of performance obtained by matching the codebook construction to the noise statistics.

- Constellations with equal priors. Figure 3.6 shows the result of the experiment for $T = 8$, $K = 256$, $\text{SNR} = -10$ dB and $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; 0; 0; 0; 0]$. The solid and solid-circled curve represent our codes adapted to the noise statistics for $M = 1$ and $M = 2$, respectively. The dashed and dashed-circled curve represent the performance of our codes adapted to $\boldsymbol{\rho} = [1; \text{zeros}(7,1)]$ for $M = 1$ and $M = 2$, respectively. We see that for SER

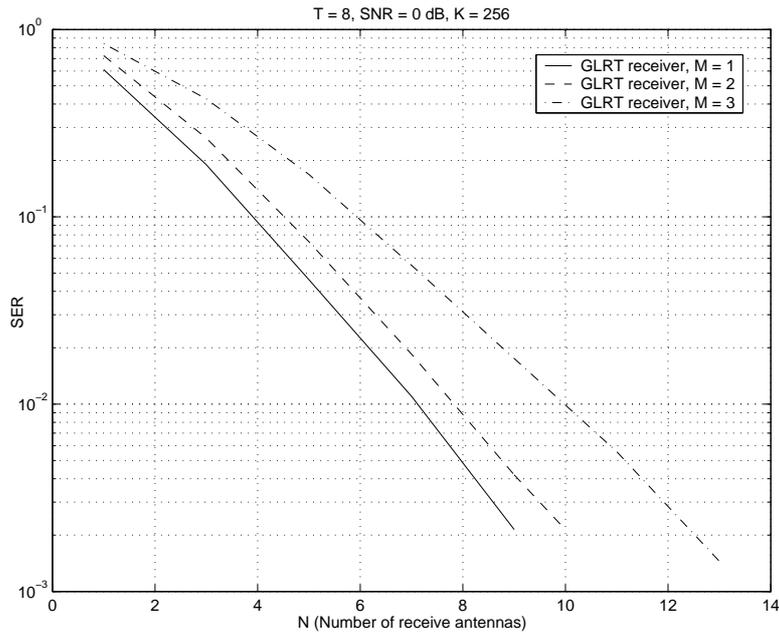


Figure 3.2: Category 1 - spatio-temporal white observation noise: $T = 8$, $K = 256$, $\text{SNR} = 0$ dB. Solid curve-our codes for $M = 1$, dashed curve-our codes for $M = 2$, dash-dotted curve-our codes for $M = 3$. All codes use GLRT receiver.

$= 5 \cdot 10^{-3}$, our $M = 1$ codes adapted to the noise statistics demonstrate the gain of 5 receive antennas over our $M = 2$ adapted codes, and 7 receive antennas over $M = 1$ noise mismatched codes. Figure 3.7 plots the result of the experiment for $T = 8$, $\text{SNR} = -10$ dB, $K = 67$ and $\boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.10; 0; 0; 0]$. The solid and solid-circled curve represent the performance of our codes that match the noise statistics for $M = 1$ and $M = 2$, respectively. The dashed and dashed-circled curve show the performance of our codes adapted to $\boldsymbol{\rho} = [1; \text{zeros}(7,1)]$. For $\text{SER} = 2 \cdot 10^{-3}$, our $M = 1$ codes adapted to the noise statistics demonstrate the gain of 3 receive antennas over our $M = 2$ adapted codes, and more than 6 receive antennas over $M = 1$ noise mismatched codes. Figure 3.8 shows the result of the experiment for $T = 8$, $K = 32$, $\text{SNR} = -10$ dB and $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; 0; 0; 0; 0]$. The solid and solid-circled curve represent our codes adapted to the noise statistics for $M = 1$ and $M = 2$, respectively. The dashed and dashed-circled curve represent the performance of our codes adapted to $\boldsymbol{\rho} = [1; \text{zeros}(7,1)]$ for $M = 1$ and $M = 2$, respectively. We witness that for $\text{SER} = 3 \cdot 10^{-3}$, our $M = 1$ codes adapted to the noise statistics demonstrate the gain of 4 receive antennas over our $M = 2$ adapted codes, and more than 8 receive antennas over $M = 1$ noise mismatched codes. We conclude

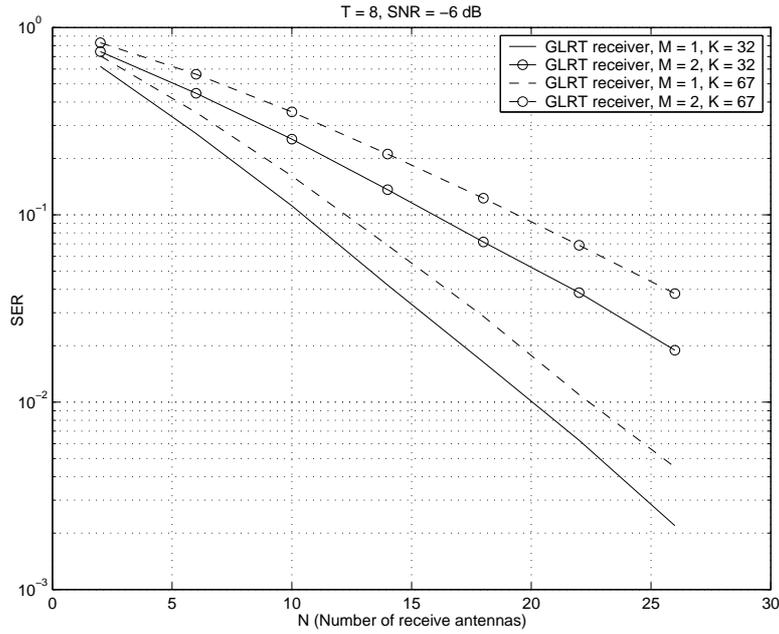


Figure 3.3: Category 1 - spatio-temporal white observation noise: $T = 8$, $\text{SNR} = -6$ dB. Solid curve-our codes for $M = 1$ and $K = 32$, dashed curve-our codes for $M = 1$ and $K = 67$, solid-circled curve-our codes for $M = 2$ and $K = 32$, dashed-circled curve-our codes for $M = 2$ and $K = 67$. All codes use GLRT receiver.

that for sufficiently low SNR one should construct codebook constellations with just one transmit antenna that match the noise statistics.

- Constellations with unequal priors. Although our primal goal in this work is to address the deterministic channel case, figure 3.9 further shows that our codebook designs for $M = 1$ are also of interest for MAP receivers that assume knowledge of the channel statistics. Figure 3.9 plots the result of the experiment for $T = 6$, $\text{SNR} = -6$ dB and $\rho = [1; 0.85; 0.6; 0.35; 0.1; 0]$. The solid, dash-dotted and dashed line represent the performances of our eight point constellations that match the noise statistics, when the GLRT receiver is implemented for $M = 1$, $M = 2$ and $M = 3$, respectively. The plus-signed dotted line represents the performance of our eight point constellation that is constructed for the spatio-temporal white noise case ($\Upsilon = \mathbf{I}_{TN}$), when GLRT receiver is implemented and $M = 1$. The plus-signed solid curve represent our 17 point constellation that match the noise statistics and $M = 1$. The dashed-circled curve shows the performance of our 17 point constellation that is constructed for $\Upsilon = \mathbf{I}_{TN}$ and $M = 1$. Both 17 point constellations are with unequal priors [47] (there is a point in the origin with probability $\frac{1}{2}$, with

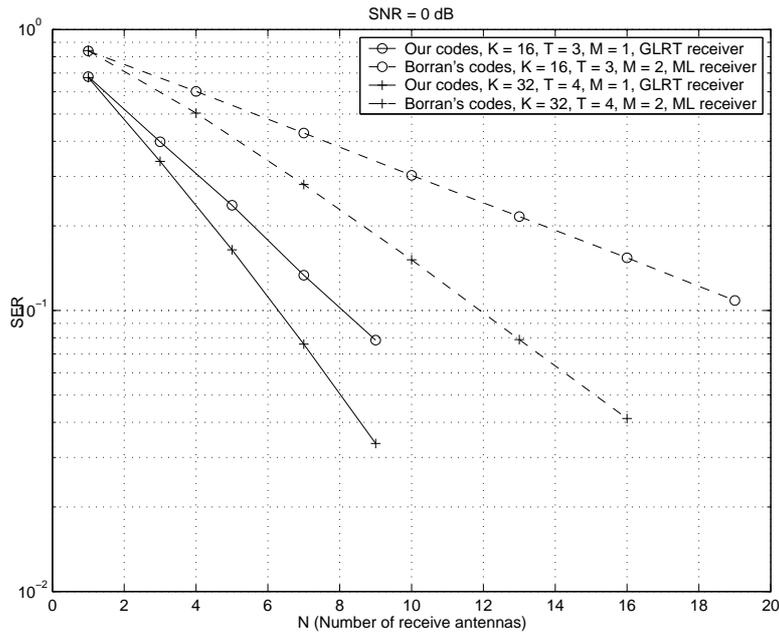


Figure 3.4: Category 1 - spatio-temporal white observation noise: Solid signed curve-our codes for $K = 32$, $T = 4$, $M = 1$, dashed signed curve-Borran's codes for $K = 32$, $T = 4$, $M = 2$, solid circled curve-our codes for $K = 16$, $T = 3$, $M = 1$, dashed circled curve-Borran's codes for $K = 16$, $T = 3$, $M = 2$.

the probabilities of the points lying in the sphere being equal), and they use MAP receiver. The gain that 17 point constellation with unequal priors demonstrate over 8 point constellation with equal priors can be explained by the fact that the signaling scheme proposed in [47] only resembles optimal, the capacity achieving distribution. The information theoretic results presented here, over the low SNR non-coherent Rayleigh fading channel with arbitrary noise correlation structure under an average power constraint, suggest that the capacity achieving distribution becomes peaky. We see that for SER of $2 \cdot 10^{-4}$, we can save two receive antennas when we compare our 17 point constellation matched to the noise statistics with the mismatched constellation constructed for $\mathbf{\Upsilon} = \mathbf{I}_{TN}$. Also, as expected, for SER of $2 \cdot 10^{-4}$, and $M = 1$, two receive antennas can be spared when we compare our eight point constellation matched to the noise statistics with the mismatched constellation constructed for $\mathbf{\Upsilon} = \mathbf{I}_{TN}$.

Next, we consider correlated Rayleigh fading channel case. We treat single beam constellations, i.e., codes with codewords of the form $\mathbf{X}_k = [\mathbf{x}_k \quad \mathbf{0}_{T \times (M-1)}]$, and we show that significant improvements are possible over uncorrelated Rayleigh fading. Figure 3.10

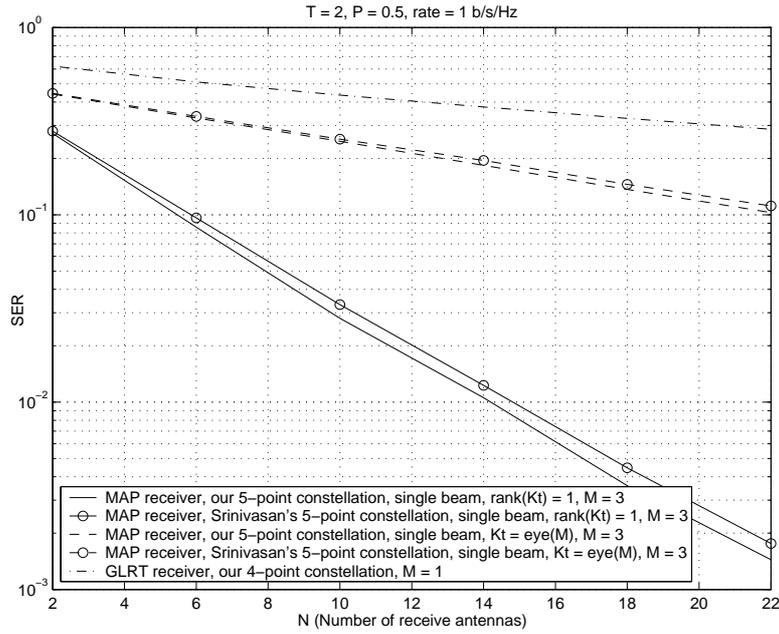


Figure 3.5: Category 1 - spatio-temporal white observation noise: $T = 2$, $P = 0.5$, rate = 1 b/s/Hz, $\mathbf{K}_r = \mathbf{I}_N$. Correlated Rayleigh fading: solid curve-our 5 point single beam constellation with unequal priors for $M = 3$ and $\text{rank}(\mathbf{K}_t)=1$, solid circled curve-Srinivasan's 5 point single beam constellation with unequal priors [47] for $M = 3$ and $\text{rank}(\mathbf{K}_t)=1$. Uncorrelated Rayleigh fading: dashed curve-our 5 point single beam constellation with unequal priors for $M = 3$ and $\mathbf{K}_t = \mathbf{I}_M$, dashed circled curve-Srinivasan's 5 point single beam constellation with unequal priors for $M = 3$ and $\mathbf{K}_t = \mathbf{I}_M$. Dash-dotted curve-our 4 point constellation for $M = 1$ with equal priors. Our and Srinivasan's 5 point constellations use *maximum a-posteriori* (MAP) receiver, our 4 point constellation uses GLRT receiver.

plots the result of the experiment. We considered the case where the coherence interval $T = 6$, $P = 0.1$, $\mathbf{K}_r = \mathbf{I}_N$, $\boldsymbol{\rho}=[1; 0.85; 0.6; 0.35; 0.1; 0]$ and rate = 0.5 bps/Hz, with $P = \text{E}[\text{tr}(\mathbf{X}_k^H \mathbf{X}_k)]$. The correlated Rayleigh fading case where $\mathbf{K}_t = \widehat{\mathbf{K}}_t$ with $\widehat{\mathbf{K}}_t$ defined in section 3.2 is assumed (again, the notation $\text{rank}(\mathbf{K}_t)=1$ implies that the case when $\mathbf{K}_t = \widehat{\mathbf{K}}_t$ is treated). The dotted plus-signed and dash-dotted curve represent our 17 point single beam constellation with unequal priors [47], and our 8 point single beam constellation with equal priors, respectively. Then, we investigate the case of uncorrelated Rayleigh fading channel, i.e., $\mathbf{K}_t = \mathbf{I}_M$ where $M = 3$. The dashed curve represents our 17 point single beam constellation with unequal priors and the solid curve represents our 8 point single beam constellation with equal priors. For $\text{SER} = 10^{-2}$, our 8 and 17 point codes in the correlated regime perform substantially better than in uncorrelated regime

(more than 10 receive antennas can be spared if one decides to make the transmit antennas fully correlated). This is in concordance with the results presented in section 3.2 where it has been proved that channel correlation can actually improve the performance of the channel in arbitrary noise environment. We witness the effect of our 17 point codes losing their superiority over 8 point codes as the number of receive antennas increases. This can be justified by the fact that the signaling scheme proposed in [47] only resembles the optimal, capacity achieving distribution.

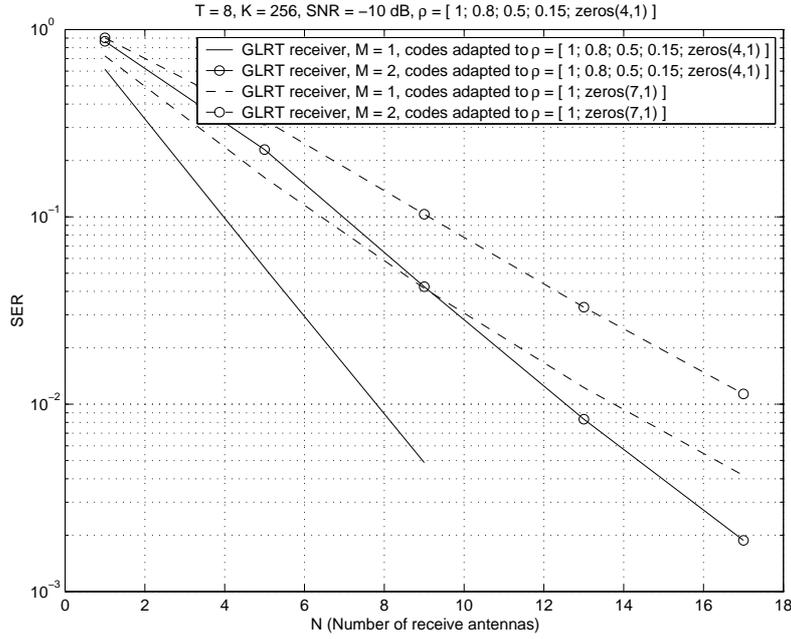


Figure 3.6: Category 2 - spatially white - temporally colored: $T = 8$, $K = 256$, $\text{SNR} = -10$ dB, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$. Solid curve-our codes for $M = 1$ adapted to $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$, solid-circled curve-our codes for $M = 2$ adapted to $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$, dashed curve-our codes for $M = 1$ adapted to $\boldsymbol{\rho} = [1; \text{zeros}(7,1)]$, dashed-circled curve-our codes for $M = 2$ adapted to $\boldsymbol{\rho} = [1; \text{zeros}(7,1)]$. All codes use GLRT receiver.

Third category: $\mathbf{E} = \mathbf{s} \boldsymbol{\alpha}^T + \mathbf{E}_{\text{temp}}$

We considered the case where the noise matrix is of the form $\mathbf{E} = \mathbf{s} \boldsymbol{\alpha}^T + \mathbf{E}_{\text{temp}}$. This models an interfering source \mathbf{s} (with known covariance matrix $\boldsymbol{\Upsilon}_s$) where the complex vector $\boldsymbol{\alpha}$ is the known channel attenuation between each receive antenna and the interfering source. The matrix \mathbf{E}_{temp} has a noise covariance matrix belonging to the second category. Thus, the noise covariance matrix is given by $\boldsymbol{\Upsilon} = \boldsymbol{\alpha} \boldsymbol{\alpha}^H \otimes \boldsymbol{\Upsilon}_s + \mathbf{I}_N \otimes \boldsymbol{\Sigma}(\boldsymbol{\rho})$. As in the

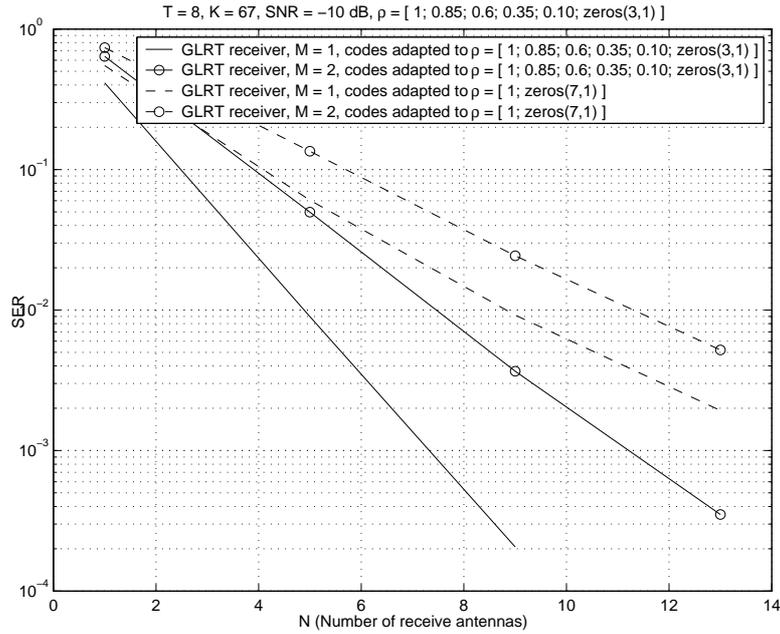


Figure 3.7: Category 2 - spatially white - temporally colored: $T = 8$, $K = 67$, $\text{SNR} = -10$ dB, $\boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.1; \text{zeros}(3,1)]$. Solid curve—our codes for $M = 1$ adapted to $\boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.1; \text{zeros}(3,1)]$, solid-circled curve—our codes for $M = 2$ adapted to $\boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.1; \text{zeros}(3,1)]$, dashed curve—our codes for $M = 1$ adapted to $\boldsymbol{\rho} = [1; \text{zeros}(7,1)]$, dashed-circled curve—our codes for $M = 2$ adapted to $\boldsymbol{\rho} = [1; \text{zeros}(7,1)]$. All codes use GLRT receiver.

second category, we shall compare our codes adapted to this specific scenario with codes designed for spatio-temporal white observation noise. We demonstrate the increase of performance obtained by matching the codebook construction to the noise statistics.

- Constellations with equal priors. Figure 3.11 plots the result of the experiment for $T = 8$, $N = 2$, $K = 32$, $\mathbf{s} = [1; 0.7; 0.4; 0.15; 0; 0; 0; 0]$, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; 0; 0; 0; 0]$ and $\boldsymbol{\alpha} = [-1.146 + 1.189i; 1.191 - 0.038i]$. For $\text{SER} = 10^{-2}$ we experience the gain of 3 dB when we compare the one transmit antenna constellation, constructed taking into account the noise statistics, with the one transmit constellation constructed for $\boldsymbol{\Upsilon} = \mathbf{I}_{TN}$. The conclusion we draw here, as before, is that for sufficiently low SNR one should construct codebook constellations with just one transmit antenna that match the noise statistics. Also, as expected, the $M = 2$ codebook construction, adapted to noise statistics, outperforms the one antenna constellation as SNR increases. Figure 3.12 shows the result of the experiment for $T = 4$, $N = 2$, $K = 16$, $\mathbf{s} = [1; 0.7; 0.4; 0]$, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0]$ and $\boldsymbol{\alpha} = [-0.433 + 0.125i; -1.665 + 0.288i]$. The solid (dashed) curve represent our codes for $M = 2$ ($M = 1$) adapted

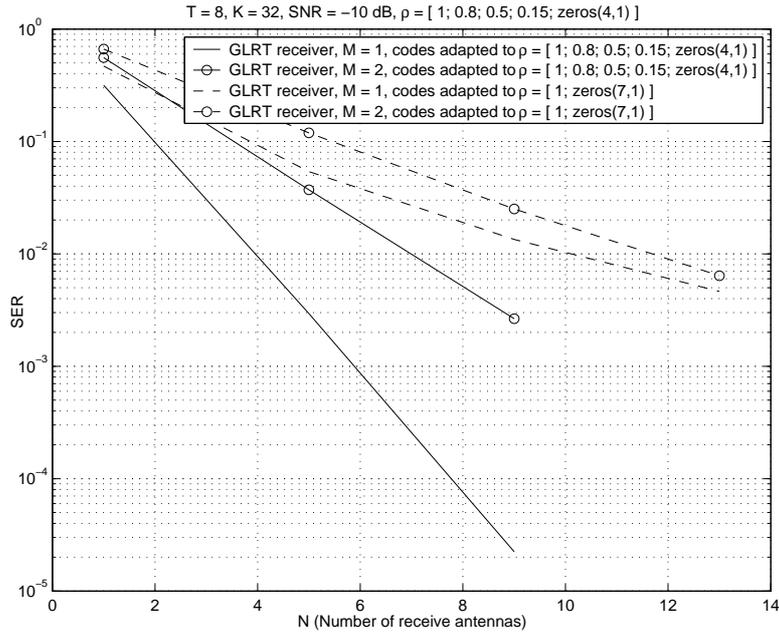


Figure 3.8: Category 2 - spatially white - temporally colored: $T = 8$, $K = 32$, $\text{SNR} = -10$ dB, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$. Solid curve-our codes for $M = 1$ adapted to $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$, solid-circled curve-our codes for $M = 2$ adapted to $\boldsymbol{\rho} = [1; 0.8; 0.5; 0.15; \text{zeros}(4,1)]$, dashed curve-our codes for $M = 1$ adapted to $\boldsymbol{\rho} = [1; \text{zeros}(7,1)]$, dashed-circled curve-our codes for $M = 2$ adapted to $\boldsymbol{\rho} = [1; \text{zeros}(7,1)]$. All codes use GLRT receiver.

to colored noise, respectively, and the dash-dotted curve represents our codes for $M = 1$ adapted to spatio-temporal white observation noise. For $\text{SER} = 10^{-2}$ we witness the gain of 1.5 dB when we compare the one transmit antenna constellation constructed taking into account the noise statistics, with the one transmit constellation constructed for $\mathbf{Y} = \mathbf{I}_{TN}$. Again, as SNR increases, the $M = 2$ codebook construction, adapted to the noise statistics, outperforms the one antenna constellation.

The foregoing results for the cases when the noise matrix is of the form of the second and third category give rise to the following conclusion:

For a GLRT receiver, at sufficiently low SNR, one should construct codebook constellations with just one transmit antenna, but which are adapted to the noise statistics.

- Constellations with unequal priors. As for the case when the noise covariance matrix belongs to the second category, we demonstrate that our codebook designs for $M = 1$ are of interest for MAP receivers that assume knowledge of the channel statistics. Figure 3.13 plots the result of the experiment for $T = 4$, $M = 1$, $N = 2$, $\mathbf{s} = [1; 0.7; 0.4; 0]$, $\boldsymbol{\rho} =$

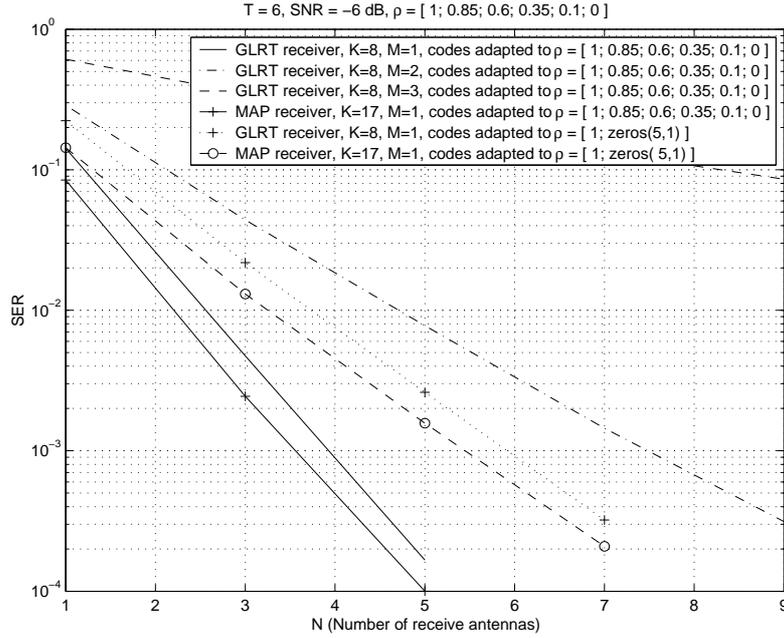


Figure 3.9: Category 2 - spatially white - temporally colored: $T = 6$, $\text{SNR} = -6$ dB, $\boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.1; 0]$. Solid, dash-dotted and dashed curve-our 8 point constellations that match the noise statistics for $M = 1$, $M = 2$ and $M = 3$, respectively. Plus-signed dotted curve-our 8 point constellation that is constructed for the spatio-temporal white noise case ($\boldsymbol{\Upsilon} = \mathbf{I}_{TN}$) and $M = 1$. Plus-signed solid curve-our 17 point constellation that match the noise statistics and $M = 1$. Dashed-circled curve-our 17 point constellation that is constructed for $\boldsymbol{\Upsilon} = \mathbf{I}_{TN}$ and $M = 1$. Our 8 point constellations use GLRT receiver, our 17 point constellations use MAP receiver.

$[1; 0.8; 0.5; 0]$, $\boldsymbol{\alpha} = [-0.4326 + 0.1253i; -1.6656 + 0.2877i]$. Solid-circled curve represents our 17 point codes with unequal priors [47] adapted to colored noise, and use MAP receiver. The plus-signed solid curve represents our 8 point codes with equal priors adapted to colored noise, and use ML receiver. The solid curve represents our 8 point codes with equal priors adapted to colored noise, and use GLRT receiver. The dashed-circled curve represents our 17 point codes with unequal priors adapted to white noise, and use MAP receiver. The plus-signed dashed curve represents our 8 point codes with equal priors adapted to white noise, and use ML receiver. The dashed curve represents our 8 point codes with equal priors adapted to white noise, and use GLRT receiver. For SNR from -5 dB to 5 dB, our one transmit antenna constellations adapted to the noise statistics demonstrate the gain of 2 dB when compared to the codes designed for the white noise case. For $\text{SER} = 10^{-2}$, our 17 point codes demonstrate the gain of 1 dB when compared to

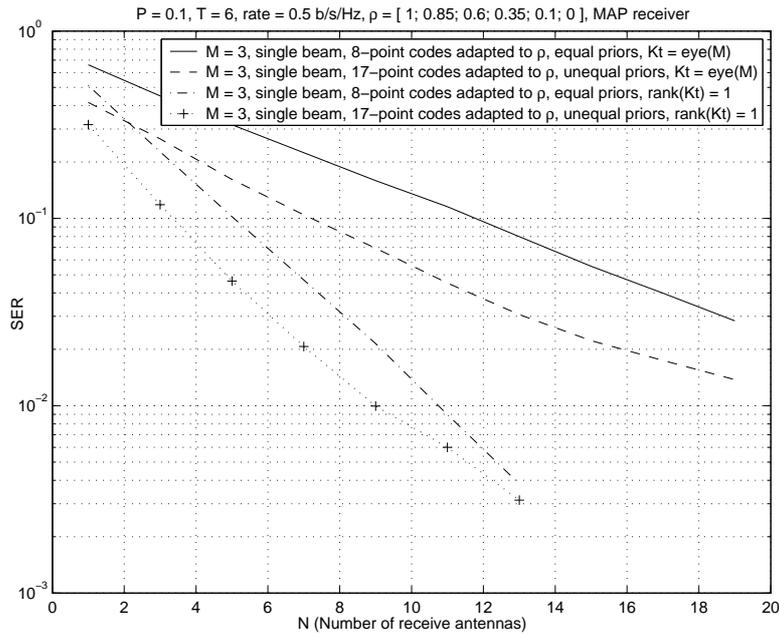


Figure 3.10: Category 2 - spatially white - temporally colored: $\mathbf{K}_r = \mathbf{I}_N$. Correlated Rayleigh fading: dotted plus-signed curve-our 17 point single beam constellation with unequal priors for $M = 3$ and $\text{rank}(\mathbf{K}_t)=1$, dash-dotted curve-our 8 point single beam constellation with equal priors for $M = 3$ and $\text{rank}(\mathbf{K}_t)=1$. Uncorrelated Rayleigh fading: dashed curve-our 17 point single beam constellation with unequal priors for $M = 3$ and $\mathbf{K}_t = \mathbf{I}_M$, solid curve-our 8 point single beam constellation with equal priors for $M = 3$ and $\mathbf{K}_t = \mathbf{I}_M$. All codes use MAP receiver.

the 8 point codes, which is in concordance with the theoretic-information results presented in this chapter. Also, we see that our GLRT receiver performs sub-optimally w.r.t. the ML receiver.

3.4 Conclusions

We have studied the MIMO channel in the low SNR regime from two perspectives: capacity and PEP analysis. The novel aspect is that we allow the Gaussian observation noise to have an arbitrary correlation structure. From the capacity analysis perspective for correlated Rayleigh fading channel, we have shown that, by maximizing the mutual information for the on-off and Gaussian signalings over the system's parameters (antenna correlation), the transmit (receive) antennas should be made as correlated as possible. Further, we have presented the PEP analysis for the low SNR deterministic channel setup and have shown how the noise statistics could be taken into account when constructing codebook

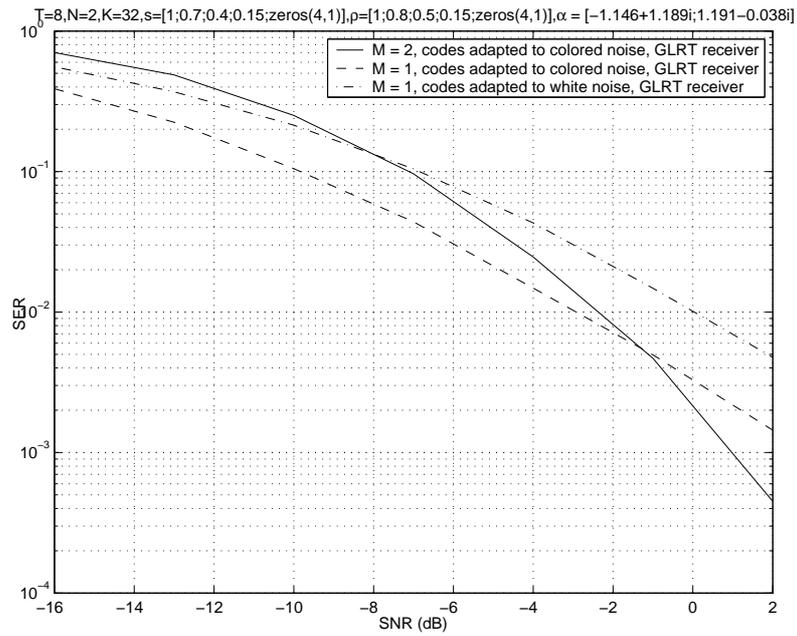


Figure 3.11: Category 3 - Solid curve-our codes for $M = 2$ adapted to colored noise, dashed curve-our codes for $M = 1$ adapted to colored noise, dash-dotted curve-our codes for $M = 1$ adapted to white noise. All codes use GLRT receiver.

constellations. We argued that one should construct codebooks for just one transmit antenna that match the noise statistics.

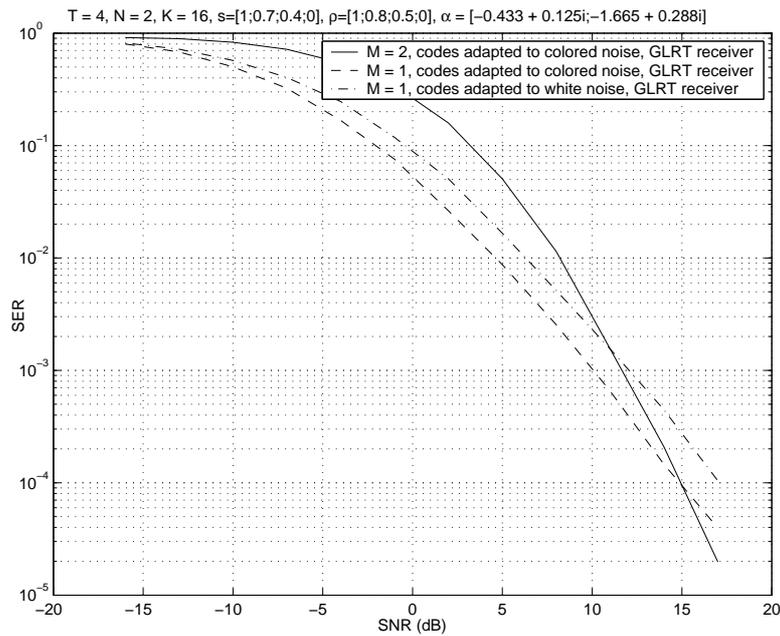


Figure 3.12: Category 3 - Solid curve-our codes for $M = 2$ adapted to colored noise, dashed curve-our codes for $M = 1$ adapted to colored noise, dash-dotted curve-our codes for $M = 1$ adapted to white noise. All codes use GLRT receiver.

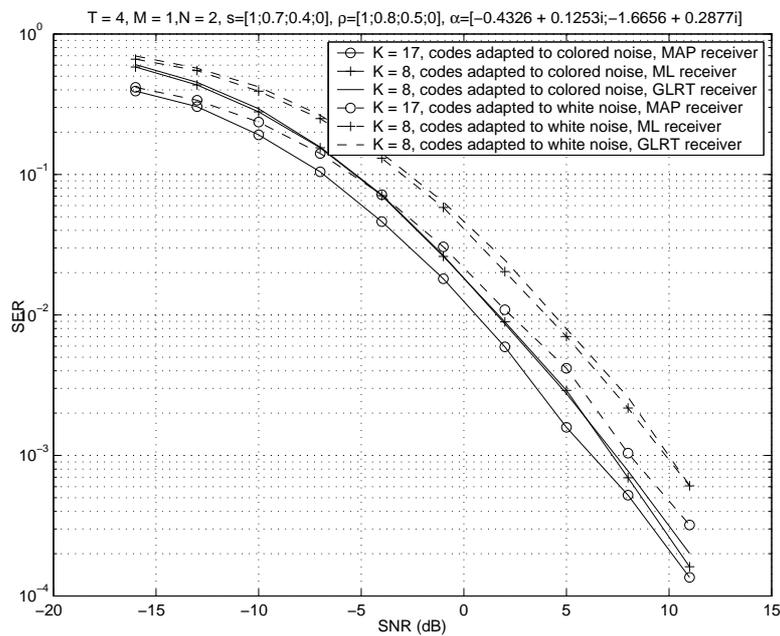


Figure 3.13: Category 3 - Solid-circled curve-our 17 point codes with unequal priors [47] adapted to colored noise, plus-signed solid curve-our 8 point codes with equal priors adapted to colored noise, solid curve-our 8 point codes with equal priors adapted to colored noise, dashed-circled curve-our 17 point codes with unequal priors adapted to white noise, plus-signed dashed curve-our 8 point codes with equal priors adapted to white noise, dashed curve-our 8 point codes with equal priors adapted to white noise. Circled, signed, and 8-point code curves use MAP, ML and GLRT receivers, respectively.

Chapter 4

Conclusions and Future Work

4.1 Conclusions

The research conducted in this thesis can be divided in two parts: high and low SNR regime.

In Chapter 2, we addressed the problem of space-time codebook design for non-coherent communications in multiple-antenna wireless systems and high SNR regime. In contrast with other approaches, the channel matrix was modeled as an unknown deterministic parameter at both the receiver and the transmitter, and the Gaussian observation noise was allowed to have an arbitrary correlation structure, known by the transmitter and the receiver. In order to handle the unknown deterministic space-time channel, a GLRT receiver was considered. A new methodology for space-time codebook design under this non-coherent setup was proposed. This optimizes the probability of error of the GLRT receiver's detector in the high SNR regime, thus solving a high-dimensional nonlinear non-smooth optimization problem in a two-step approach: **(i)** firstly, a convex SDP relaxation of the codebook design problem yields a rough estimate of the optimal codebook; **(ii)** this is then refined through a geodesic descent optimization algorithm that exploits the Riemannian geometry imposed by the power constraints on the space-time codewords. The results obtained through computer simulations illustrate the advantages of our method. For the specific case of spatio-temporal white observation noise, our codebook constructions replicate the performance of state-of-art known solutions. The main point here is that our methodology permits to extend the codebook construction to any given correlated noise environment. The simulation results show the good performance of these new

designed codes in colored noise setups.

In Chapter 3, the non-coherent single-user MIMO channel in the low SNR regime was investigated from two viewpoints: capacity and probability of error analysis. The novelty in both viewpoints is that an arbitrary correlation structure is allowed for the Gaussian observation noise. First, we looked at the capacity of the spatially correlated Rayleigh fading channel. We investigated the impact of channel and noise correlation on the mutual information for the on-off and Gaussian signaling schemes. Our results establish that, in the low SNR regime, mutual information is maximized when the transmit antennas are fully correlated (the same holds for the receive array). Then, we considered the deterministic channel setup and perform a PEP analysis for the GLRT receiver. This leads to a codebook design criterion on which we base the construction of new space-time constellations. Their performance were assessed by computer simulations and we argued that one should construct codebooks for just one transmit antenna that match the noise statistics. As a byproduct, we showed that our codebooks are also of interest for Bayesian receivers which decode constellations with non-uniform priors.

4.2 Future Work

There are number of possibilities for future research. In the following, some research directions are mentioned:

1. An important outcome of this thesis is the conclusion that is worthwhile to treat the codebook design problem in the case when the Gaussian observation noise is allowed to have an arbitrary correlation structure, since we have seen that if the perfect knowledge of the noise correlation is available, significant improvements over unitary constellations in general correlated noise environments are possible. Potential next questions to consider are the following. The designed codebooks do not possess any structure that could alleviate a reduction in the size of the memory needed to store them. This is especially important in the high SNR regime where, due to large capacity, one should construct high-rate constellations. (In the low-medium SNR regime, constellations with relatively small cardinalities suffice since the channel ca-

capacities are also relatively small). In practice, the storage of a high-rate unstructured codebook becomes infeasible because the size of the codebook increases exponentially with the data rate. Hence, the design of structured constellations, which will consequently allow tractable encoding, may be a topic for future research. Likewise, there is a concern w.r.t. the decoding complexity. The decoding complexity of the codebook constructions is, unfortunately, exponential in the data rate. Since, in this work, the decoding complexity did not represent a constraint, the design of non-coherent STC with reduced complexity algorithms remains an important open problem that deserves further investigation.

2. Throughout the thesis, the noise covariance matrix is assumed to be perfectly known at the site where the codebook design is performed. An assessment of the sensitivity of the codebook design to mismatch between the noise covariance assumed in the design and the noise covariance encountered in practice could be a good topic for future work.
3. The research presented in the thesis has been conducted on channels that are assumed to be flat fading. This can be considered true for low symbol rate or narrow bandwidth communications systems. For higher symbol rate transmissions, fading is frequency selective and so poses a greater challenge given limited link budget and severity of wireless environment. We plan to devote our attention to orthogonal frequency division multiplexing (OFDM) which has emerged as a very promising technique to provide high data-rate transmission over broadband MIMO channels. Its relevance is testified by being chosen as the standard interface for digital audio broadcasting (DAB), terrestrial digital video broadcasting (DVB), wireless local area networks (WLANs), and wireless metropolitan area networks (MANs). One of the key advantages of OFDM over traditional single-carrier modulation and CDMA is its low computational complexity in practical implementations. MIMO-OFDM is therefore a particularly promising candidate for future fourth-generation (4G) wireless networks. An interesting research direction is the design of optimum modulation

and coding schemes for MIMO-OFDM in both high and low SNR regime. We are particularly interested in the non-coherent case. Our goal is to find modulation and coding schemes for MIMO-OFDM that exploit the available spatial and frequency diversity, being robust for a wide range of realistic channels.

4. In Chapter 2, new packings in the complex projective space have been presented. In some cases we were able to construct optimal packings that meet the Rankin bound (so-called ETFs). Equivalently, an ETF can be seen as a sequence of K unit vectors in a T -dimensional Euclidian space whose pairwise absolute inner products meet the well-known Welch lower bound [65]. ETFs have applications in communications, coding theory and sparse approximation [78], [79]. As an example, in [80] it has been shown that an ETF provides an error correcting code that is maximally robust against two erasures. ETFs also play an important role in multiuser communication systems. It is known that ETFs achieve the capacity of a Gaussian channel and satisfy an interference invariance property [81]. Numerical evidence indicates that complex ETFs do not exist for most pairs (T, K) . An interesting open question is to find conditions on (T, K) that rule out the existence of general complex ETFs. In the other direction, it seems that a *maximal complex ETF* exists for each natural number T (in this specific case $K = T^2$). Resolving this conjecture represents a challenging topic for future work.

5. In Chapter 3, section 3.2, the capacity analysis of MIMO fading channels in the low SNR regime for the case when the channel correlation matrix is modeled by the Kronecker product of the transmit and receive correlation matrix was presented. The generalization of the analysis to the case when the channel covariance matrix has an arbitrary correlation structure represents a possible topic for future work. Moreover, an extension of the analysis presented in Chapter 3 to the cases when peak and fourth-order moment signal constraints are imposed in order to determine what signaling should be applied to the input, how many transmit antennas should be employed and what is the optimal value of the channel covariance matrix is an

interesting topic for future research.

Appendix A

Pairwise Error Probability for Fast Fading in the High SNR Regime

In this appendix, we derive the expression for the asymptotic (high SNR regime) pairwise error probability for fast fading presented in (2.8).

If \mathbf{X}_i is transmitted, then the probability that the receiver decides in favor of \mathbf{X}_j is:

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} = P(\mathbf{z}_i^H \mathbf{\Upsilon}^{-1} \mathbf{z}_i > \mathbf{z}_j^H \mathbf{\Upsilon}^{-1} \mathbf{z}_j) \quad (\text{A.1})$$

where for $k \in \{i, j\}$

$$\mathbf{z}_k = \mathbf{y} - \widetilde{\mathbf{X}}_k \widehat{\mathbf{g}}_k,$$

$$\widetilde{\mathbf{X}}_k = \mathbf{I}_N \otimes \mathbf{X}_k, \quad \mathbf{x}_k = \mathbf{\Upsilon}^{-\frac{1}{2}} \widetilde{\mathbf{X}}_k$$

$$\mathbf{y} = \text{vec}(\mathbf{Y}) = \widetilde{\mathbf{X}}_i \mathbf{g} + \mathbf{e}, \quad \mathbf{e} = \text{vec}(\mathbf{E}),$$

and

$$\widehat{\mathbf{g}}_k = (\mathbf{x}_k^H \mathbf{x}_k)^{-1} \mathbf{x}_k^H \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{y}$$

is the ML estimate of the channel when \mathbf{X}_k is transmitted. The unknown realization of the channel is denoted by $\mathbf{g} = \text{vec}(\mathbf{H}^H)$.

Let $\mathbf{S}_i = (\mathbf{x}_i^H \mathbf{x}_i)^{-1} \mathbf{x}_i^H \mathbf{\Upsilon}^{-\frac{1}{2}}$. Thus,

$$\begin{aligned} \mathbf{z}_i &= \mathbf{y} - \widetilde{\mathbf{X}}_i \widehat{\mathbf{g}}_i \\ &= \underbrace{(\mathbf{I}_{TN} - \widetilde{\mathbf{X}}_i \mathbf{S}_i)}_{\mathbf{P}_i} (\widetilde{\mathbf{X}}_i \mathbf{g} + \mathbf{e}) \\ &= \mathbf{P}_i \mathbf{e}. \end{aligned} \quad (\text{A.2})$$

Similarly, it can be shown that

$$\mathbf{z}_j = \mathbf{\Delta} \mathbf{g} + \mathbf{P}_j \mathbf{e} \quad (\text{A.3})$$

where $\mathbf{P}_j = \mathbf{I}_{TN} - \widetilde{\mathbf{X}}_j \mathbf{S}_j$ and $\mathbf{\Delta} = \mathbf{P}_j \widetilde{\mathbf{X}}_i$. Hence, substituting (A.2) and (A.3) in (A.1) we have

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} = P(\underbrace{e^H (\mathbf{P}_i^H \mathbf{\Upsilon}^{-1} \mathbf{P}_i - \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j)}_{\mathbf{T}} e - e^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta} \mathbf{g} - \mathbf{g}^H \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j e > \underbrace{\mathbf{g}^H \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta} \mathbf{g}}_{\lambda}). \quad (\text{A.4})$$

Unfortunately, the expression (A.4) cannot be simplified analytically. We shall analyze (A.4) in the high SNR regime where the linear term of \mathbf{e} is dominant, i.e., the quadratic term of \mathbf{e} is negligible. In the following, we explain better the preceding claim.

First note that SNR and $\mathbf{\Upsilon}$ are linked by the formula

$$\text{SNR} = \frac{\mathbb{E} [\|\mathbf{X}_k \mathbf{H}^H\|^2]}{\mathbb{E} [\|\mathbf{E}\|^2]} = \frac{\text{tr}(\mathbf{H}^H \mathbf{H} \mathbb{E}[\mathbf{X}_k^H \mathbf{X}_k])}{\text{tr}(\mathbf{\Upsilon})}.$$

In the above equation, the channel was assumed deterministic (we used assumption **A1**). Moreover, once the codebook is fixed, we have $\mathbb{E}[\mathbf{X}_k^H \mathbf{X}_k] = \mathbf{R}$ for a certain positive-definite matrix \mathbf{R} . Thus,

$$\text{SNR} = \frac{\kappa}{\text{tr}(\mathbf{\Upsilon})}, \quad \kappa := \text{tr}(\mathbf{H}^H \mathbf{H} \mathbf{R}),$$

and saying that $\text{SNR} \rightarrow +\infty$ corresponds to saying that $\mathbf{\Upsilon} \rightarrow \mathbf{0}$. Our argument is that when $\text{SNR} \rightarrow +\infty$ the linear term

$$\mathbf{z} = -e^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta} \mathbf{g} - \mathbf{g}^H \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j e$$

dominates (in the mean square sense) the quadratic term $e^T \mathbf{T} e$ in (A.4).

As the random vector $\mathbf{e} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Upsilon})$ and other quantities in (A.4), e.g., \mathbf{P}_j , $\mathbf{\Delta}$, depend on $\mathbf{\Upsilon}$ we admit that our claim is not readily acceptable. However, if we work out the dependencies, it can be validated. We propose the following exercise. Let $\mathbf{\Upsilon} = \mathbf{\Upsilon}(t) = t \mathbf{\Upsilon}_0$ for some (fixed) Hermitian positive definite matrix $\mathbf{\Upsilon}_0$ and $t > 0$. Taking $t \downarrow 0$ corresponds to making $\text{SNR} \rightarrow +\infty$. Remark that

$$\mathbf{e} \stackrel{d}{=} \mathbf{\Upsilon}^{1/2} \mathbf{z}$$

where $\stackrel{d}{=}$ means equality in distribution and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{TN})$. Now, if we unwind the definitions of \mathbf{P}_j , etc, and go through tedious but straightforward computations, it can be seen that (A.4) boils down to

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} = P\left(\mathbf{z}^H \mathbf{C} \mathbf{z} - \frac{1}{\sqrt{t}}(\mathbf{z}^H \mathbf{c} + \mathbf{c}^H \mathbf{z}) > \frac{1}{t} \eta\right) \quad (\text{A.5})$$

where \mathbf{C} , \mathbf{c} and η denote constants with respect to t (they do not depend on t) which are not explicitly given here. Thus, the random variables

$$\mathbf{z}^H \mathbf{C} \mathbf{z} \quad \text{and} \quad \mathbf{z}^H \mathbf{c} + \mathbf{c}^H \mathbf{z}$$

have fixed power (not depending on t , i.e., SNR). Since $t > 0$ we can rewrite (A.5) as

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} = P\left(t \mathbf{z}^H \mathbf{C} \mathbf{z} - \sqrt{t}(\mathbf{z}^H \mathbf{c} + \mathbf{c}^H \mathbf{z}) > \eta\right). \quad (\text{A.6})$$

When $t \downarrow 0$ (SNR $\rightarrow +\infty$) we have $t \ll \sqrt{t}$ and that's how we justify the approximation

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \approx P\left(-\sqrt{t}(\mathbf{z}^H \mathbf{c} + \mathbf{c}^H \mathbf{z}) > \eta\right) \quad (\text{A.7})$$

which corresponds to (A.6) above. As a side remark, note that (A.7) confirms that $P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \rightarrow 0$ as $t \downarrow 0$, thanks to the fact that $\eta > 0$.

The above argument is not new. We were inspired by similar manipulations from the book [19], see pp. 56.

Therefore,

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \approx P\left(\underbrace{-e^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \Delta \mathbf{g} - \mathbf{g}^H \Delta^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j e}_z > \lambda\right). \quad (\text{A.8})$$

We see that z is a real Gaussian variable with zero mean (because $\mathbb{E}[e] = 0$) and unknown variance σ^2 , i.e., $z \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$, which will be calculated in sequel:

$$\begin{aligned} \sigma^2 &= \mathbb{E}[z^2] = \mathbb{E}[(e^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \Delta \mathbf{g} + \mathbf{g}^H \Delta^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j e)^2] \\ &= \mathbb{E}[\mathbf{g}^H \Delta^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j e e^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \Delta \mathbf{g} + e^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \Delta \mathbf{g} \mathbf{g}^H \Delta^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j e]. \end{aligned}$$

Continuing with analysis,

$$\begin{aligned} \mathbb{E}[e^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \Delta \mathbf{g} \mathbf{g}^H \Delta^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j e] &= \text{tr}(\mathbb{E}[e e^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \Delta \mathbf{g} \mathbf{g}^H \Delta^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j]) \\ &= \mathbf{g}^H \Delta^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j \mathbf{\Upsilon} \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \Delta \mathbf{g} \end{aligned}$$

which implies

$$\begin{aligned}\sigma^2 &= 2\mathbf{g}^H \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j \mathbf{\Upsilon} \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta} \mathbf{g} \\ &= 2\mathbf{g}^H \widetilde{\mathbf{X}}_i^H \underbrace{\mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j \mathbf{\Upsilon} \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j}_{\mathbf{R}} \widetilde{\mathbf{X}}_i \mathbf{g}.\end{aligned}\quad (\text{A.9})$$

It is known that, if $z \sim \mathcal{N}(0, \sigma^2)$, then

$$P(z > \lambda) = \mathcal{Q}\left(\frac{\lambda}{\sigma}\right) \quad (\text{A.10})$$

where $\mathcal{Q}(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$. The matrix \mathbf{R} in (A.9) can be simplified and it can be easily shown that

$$\mathbf{R} = \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j = \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{\Pi}_j^\perp \mathbf{\Upsilon}^{-\frac{1}{2}}. \quad (\text{A.11})$$

where $\mathbf{\Pi}_j^\perp = \mathbf{I}_{TN} - \mathbf{X}_j (\mathbf{X}_j^H \mathbf{X}_j)^{-1} \mathbf{X}_j^H$ is the orthogonal projector onto the orthogonal complement of the column space of \mathbf{X}_j . Using (A.11) and substituting it in (A.9) we have

$$\sigma^2 = 2\mathbf{g}^H \widetilde{\mathbf{X}}_i^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j \widetilde{\mathbf{X}}_i \mathbf{g} = 2\mathbf{g}^H \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta} \mathbf{g}$$

which implies

$$\sigma = \sqrt{2\mathbf{g}^H \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta} \mathbf{g}}. \quad (\text{A.12})$$

Equations (A.4), (A.8), (A.10) and (A.12) result in

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \approx \mathcal{Q}\left(\frac{\lambda}{\sigma}\right) = \mathcal{Q}\left(\frac{\mathbf{g}^H \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta} \mathbf{g}}{\sqrt{2\mathbf{g}^H \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta} \mathbf{g}}}\right) = \mathcal{Q}\left(\frac{1}{\sqrt{2}} \sqrt{\mathbf{g}^H \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta} \mathbf{g}}\right). \quad (\text{A.13})$$

Let $\mathbf{L}_{ij} = \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta}$. Thus,

$$\mathbf{L}_{ij} = \mathbf{\Delta}^H \mathbf{\Upsilon}^{-1} \mathbf{\Delta} = \widetilde{\mathbf{X}}_i^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j \widetilde{\mathbf{X}}_i. \quad (\text{A.14})$$

Hence, due to (A.11) and (A.14) it holds

$$\mathbf{L}_{ij} = \widetilde{\mathbf{X}}_i^H \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{\Pi}_j^\perp \mathbf{\Upsilon}^{-\frac{1}{2}} \widetilde{\mathbf{X}}_i = \mathbf{X}_i^H \mathbf{\Pi}_j^\perp \mathbf{X}_i. \quad (\text{A.15})$$

Equations (A.13), (A.14) and (A.15) result in (2.8). This completes the proof.

Appendix B

Optimization Problem

In this section, we prove that the equivalent formulation of the optimization problem (2.36) is given by (2.37).

The optimization problem in (2.36) can be rewritten in the following way

$$\begin{aligned}
 (\mathbf{X}_k^*, t^*) = & \arg \max_t & (B.1) \\
 & \lambda_{\min}(\mathbf{L}_{mk}) \geq t, m = 1, \dots, k-1 & (A) \\
 & \lambda_{\min}(\mathbf{L}_{km}) \geq t, m = 1, \dots, k-1 & (B) \\
 & \text{tr}(\mathbf{X}_k^H \mathbf{X}_k) = 1,
 \end{aligned}$$

where

$$\begin{aligned}
 \mathbf{L}_{ij} &= \mathbf{x}_i^H \mathbf{\Pi}_j^\perp \mathbf{x}_i, \mathbf{\Pi}_j^\perp = \mathbf{I}_{TN} - \mathbf{x}_j(\mathbf{x}_j^H \mathbf{x}_j)^{-1} \mathbf{x}_j^H, \\
 \mathbf{x}_i &= \mathbf{\Upsilon}^{-\frac{1}{2}} \widetilde{\mathbf{X}}_i \text{ and } \widetilde{\mathbf{X}}_i = \mathbf{I}_N \otimes \mathbf{X}_i.
 \end{aligned}$$

Approach: Define $\mathfrak{x}_k = \text{vec}(\mathbf{X}_k) \text{vec}^H(\mathbf{X}_k)$. We are going to show that both (A) and (B) can be written as LMI's with respect to \mathfrak{x}_k , $\text{vec}(\mathbf{X}_k)$ and t .

(A) Note that

$$\lambda_{\min}(\mathbf{L}_{mk}) \geq t \Leftrightarrow \mathbf{L}_{mk} - t\mathbf{I}_{MN} \succeq \mathbf{0}.$$

Since the matrix $\mathbf{L}_{mk} - t\mathbf{I}_{MN} = \mathbf{x}_m^H \mathbf{x}_m - \mathbf{x}_m^H \mathbf{x}_k (\mathbf{x}_k^H \mathbf{x}_k)^{-1} \mathbf{x}_k^H \mathbf{x}_m - t\mathbf{I}_{MN}$ is the Schur complement [84] of $\mathbf{x}_k^H \mathbf{x}_k$ in

$$\begin{bmatrix} \mathbf{x}_k^H \mathbf{x}_k & \mathbf{x}_k^H \mathbf{x}_m \\ \mathbf{x}_m^H \mathbf{x}_k & \mathbf{x}_m^H \mathbf{x}_m - t\mathbf{I}_{MN} \end{bmatrix}$$

we have the following equivalence (we assumed that \mathbf{X}_k is of full column rank):

$$\lambda_{\min}(\mathbf{L}_{mk}) \geq t \Leftrightarrow \begin{bmatrix} \mathbf{x}_k^H \mathbf{x}_k & \mathbf{x}_k^H \mathbf{x}_m \\ \mathbf{x}_m^H \mathbf{x}_k & \mathbf{x}_m^H \mathbf{x}_m - t\mathbf{I}_{MN} \end{bmatrix} \succeq \mathbf{0}. \quad (B.2)$$

• Let $[M]_{i,j}$ denote the ij -th element of the matrix M and e_i represent the i -th column of the identity matrix I_{MN} . Then,

$$[\mathcal{X}_k^H \mathcal{X}_k]_{i,j} = e_i^H \mathcal{X}_k^H \mathcal{X}_k e_j = e_i^H \widetilde{\mathbf{X}}_k^H \mathbf{\Upsilon}^{-1} \widetilde{\mathbf{X}}_k e_j = \text{tr} \left(\mathbf{\Upsilon}^{-1} \widetilde{\mathbf{X}}_k e_j \left(\widetilde{\mathbf{X}}_k e_i \right)^H \right) \quad (\text{B.3})$$

As $\widetilde{\mathbf{X}}_k = I_N \otimes \mathbf{X}_k$, there exists matrix \mathbf{K} of size $TMN^2 \times TM$ such that $\text{vec}(\widetilde{\mathbf{X}}_k) = \mathbf{K} \text{vec}(\mathbf{X}_k)$, see [90]. Hence,

$$\widetilde{\mathbf{X}}_k e_j = \text{vec} \left(\widetilde{\mathbf{X}}_k e_j \right) = (e_j^T \otimes I_{TN}) \text{vec}(\widetilde{\mathbf{X}}_k) = (e_j^T \otimes I_{TN}) \mathbf{K} \text{vec}(\mathbf{X}_k). \quad (\text{B.4})$$

Substituting (B.4) in (B.3) we have

$$\begin{aligned} [\mathcal{X}_k^H \mathcal{X}_k]_{i,j} &= \text{tr} \left(\mathbf{\Upsilon}^{-1} (e_j^T \otimes I_{TN}) \mathbf{K} \text{vec}(\mathbf{X}_k) \left((e_i^T \otimes I_{TN}) \mathbf{K} \text{vec}(\mathbf{X}_k) \right)^H \right) \\ &= \text{tr} (\mathbf{B}_{ij} (I_{TN}) \mathfrak{X}_k), \end{aligned} \quad (\text{B.5})$$

where we define $\mathbf{B}_{ij}(\Phi) = \mathbf{K}^H (e_i \otimes I_{TN}) \mathbf{\Upsilon}^{-\frac{1}{2}} \Phi \mathbf{\Upsilon}^{-\frac{1}{2}} (e_j^T \otimes I_{TN}) \mathbf{K}$.

• Similarly,

$$[\mathcal{X}_m^H \mathcal{X}_k]_{i,j} = e_i^H \mathcal{X}_m^H \mathcal{X}_k e_j = e_i^H \mathcal{X}_m^H \mathbf{\Upsilon}^{-\frac{1}{2}} \widetilde{\mathbf{X}}_k e_j = e_i^H \mathcal{X}_m^H \mathbf{\Upsilon}^{-\frac{1}{2}} (e_j^T \otimes I_{TN}) \mathbf{K} \text{vec}(\mathbf{X}_k). \quad (\text{B.6})$$

(B) By repeating the analysis for the case (A) we have:

$$\begin{aligned} [\mathbf{L}_{km}]_{i,j} = e_i^H \mathcal{X}_k^H \mathbf{\Pi}_m^\perp \mathcal{X}_k e_j &= e_i^H \widetilde{\mathbf{X}}_k^H \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{\Pi}_m^\perp \mathbf{\Upsilon}^{-\frac{1}{2}} \widetilde{\mathbf{X}}_k e_j \\ &= \text{tr} \left(\mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{\Pi}_m^\perp \mathbf{\Upsilon}^{-\frac{1}{2}} \widetilde{\mathbf{X}}_k e_j \left(\widetilde{\mathbf{X}}_k e_i \right)^H \right). \end{aligned}$$

Using (B.4) we obtain

$$[\mathbf{L}_{km}]_{i,j} = \text{tr} \left(\mathbf{K}^H (e_i \otimes I_{TN}) \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{\Pi}_m^\perp \mathbf{\Upsilon}^{-\frac{1}{2}} (e_j^T \otimes I_{TN}) \mathbf{K} \text{vec}(\mathbf{X}_k) \text{vec}^H(\mathbf{X}_k) \right).$$

Hence,

$$[\mathbf{L}_{km}]_{i,j} = \text{tr} \left(\mathbf{B}_{ij}(\mathbf{\Pi}_m^\perp) \mathfrak{X}_k \right). \quad (\text{B.7})$$

Combining (B.2), (B.3), (B.5), (B.6) and (B.7) we conclude that both (A) and (B) can be written as LMI's with respect to the variables \mathfrak{X}_k , $\text{vec}(\mathbf{X}_k)$ and t . Consequently, the optimization problems (2.36) and (2.37) are equivalent. This concludes the proof.

Appendix C

Calculating Gradients

In this section, we calculate gradient to be used in (2.42). Although the function f_{ij} assumes complex valued entries, that is

$$f_{ij} : \underbrace{\mathbb{C}^{T \times M} \times \dots \times \mathbb{C}^{T \times M}}_K \rightarrow \mathbb{R} \quad f_{ij}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K) = \lambda_{\min}(\mathbf{L}_{ij})$$

where

$$\mathbf{L}_{ij} = \boldsymbol{\chi}_i^H \boldsymbol{\Pi}_j^\perp \boldsymbol{\chi}_i, \boldsymbol{\Pi}_j^\perp = \mathbf{I}_{TN} - \boldsymbol{\chi}_j (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} \boldsymbol{\chi}_j^H, \boldsymbol{\chi}_i = \boldsymbol{\Upsilon}^{-\frac{1}{2}} \widetilde{\mathbf{X}}_i$$

and $\widetilde{\mathbf{X}}_i = \mathbf{I}_N \otimes \mathbf{X}_i$, we shall treat f_{ij} as a function of the real and imaginary components of $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$, i.e.

$$f_{ij} : \underbrace{\mathbb{R}^{T \times M} \times \dots \times \mathbb{R}^{T \times M}}_{2K} \rightarrow \mathbb{R},$$

$$f_{ij}(\Re\{\mathbf{X}_1\}, \Im\{\mathbf{X}_1\}, \Re\{\mathbf{X}_2\}, \Im\{\mathbf{X}_2\}, \dots, \Re\{\mathbf{X}_K\}, \Im\{\mathbf{X}_K\}) = \lambda_{\min}(\mathbf{L}_{ij}).$$

Let λ_{\min} be a simple eigenvalue of the Hermitian matrix $\mathbf{L}_{ij}(\mathcal{C}_0)$, and let \mathbf{u}_0 be an associated unit-norm eigenvector, so that $\mathbf{L}_{ij}(\mathcal{C}_0)\mathbf{u}_0 = \lambda_{\min}(\mathbf{L}_{ij}(\mathcal{C}_0))\mathbf{u}_0$. The differential df_{ij} , computed at the point \mathcal{C}_0 , is given by, pp. 162 in [90]

$$df_{ij} = \mathbf{u}_0^H d\mathbf{L}_{ij} \mathbf{u}_0.$$

where $d\mathbf{L}_{ij}$ denotes the differential of the map $\mathcal{C} \mapsto \mathbf{L}_{ij}(\mathcal{C})$ computed at the point \mathcal{C}_0 .

Define $\mathbf{K}_j = \boldsymbol{\Upsilon}^{-\frac{1}{2}} \boldsymbol{\Pi}_j^\perp \boldsymbol{\Upsilon}^{-\frac{1}{2}}$. Hence,

$$d\mathbf{L}_{ij} = (d\widetilde{\mathbf{X}}_i)^H \mathbf{K}_j \widetilde{\mathbf{X}}_i + \widetilde{\mathbf{X}}_i^H \mathbf{K}_j d\widetilde{\mathbf{X}}_i + \widetilde{\mathbf{X}}_i^H d\mathbf{K}_j \widetilde{\mathbf{X}}_i$$

and

$$\begin{aligned} df_{ij} = \mathbf{u}_0^H d\mathbf{L}_{ij} \mathbf{u}_0 &= \mathbf{u}_0^H \left((d\widetilde{\mathbf{X}}_i)^H \mathbf{K}_j \widetilde{\mathbf{X}}_i + \widetilde{\mathbf{X}}_i^H \mathbf{K}_j d\widetilde{\mathbf{X}}_i + \widetilde{\mathbf{X}}_i^H d\mathbf{K}_j \widetilde{\mathbf{X}}_i \right) \mathbf{u}_0 \\ &= \Re \left\{ \text{tr} \left((d\widetilde{\mathbf{X}}_i)^H \underbrace{2\mathbf{K}_j \widetilde{\mathbf{X}}_i \mathbf{u}_0 \mathbf{u}_0^H}_{\mathbf{C}_i} \right) \right\} + \mathbf{u}_0^H \widetilde{\mathbf{X}}_i^H d\mathbf{K}_j \widetilde{\mathbf{X}}_i \mathbf{u}_0. \end{aligned} \quad (\text{C.1})$$

Continuing with analysis,

$$\begin{aligned} d\mathbf{K}_j &= \underbrace{-\mathbf{\Upsilon}^{-\frac{1}{2}} (d\boldsymbol{\chi}_j (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} \boldsymbol{\chi}_j^H + \boldsymbol{\chi}_j (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} (d\boldsymbol{\chi}_j)^H) \mathbf{\Upsilon}^{-\frac{1}{2}}}_{\mathbf{K}_{j1}} - \\ &\quad \underbrace{-\mathbf{\Upsilon}^{-\frac{1}{2}} \boldsymbol{\chi}_j d \left((\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} \right) \boldsymbol{\chi}_j^H \mathbf{\Upsilon}^{-\frac{1}{2}}}_{\mathbf{K}_{j2}}. \end{aligned} \quad (\text{C.2})$$

Using the equality $d(\mathbf{A}^{-1}) = -\mathbf{A}^{-1} d\mathbf{A} \mathbf{A}^{-1}$ [90], we can write

$$d \left((\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} \right) = -(\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} \left((d\boldsymbol{\chi}_j)^H \boldsymbol{\chi}_j + \boldsymbol{\chi}_j^H d\boldsymbol{\chi}_j \right) (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1}. \quad (\text{C.3})$$

Substituting (C.3) and (C.2) in (C.1) we get

$$\mathbf{u}_0^H \widetilde{\mathbf{X}}_i^H d\mathbf{K}_j \widetilde{\mathbf{X}}_i \mathbf{u}_0 = \mathbf{u}_0^H \widetilde{\mathbf{X}}_i^H \mathbf{K}_{j1} \widetilde{\mathbf{X}}_i \mathbf{u}_0 + \mathbf{u}_0^H \widetilde{\mathbf{X}}_i^H \mathbf{K}_{j2} \widetilde{\mathbf{X}}_i \mathbf{u}_0$$

with

$$\begin{aligned} \mathbf{u}_0^H \widetilde{\mathbf{X}}_i^H \mathbf{K}_{j1} \widetilde{\mathbf{X}}_i \mathbf{u}_0 &= -2\Re \left\{ \mathbf{u}_0^H \boldsymbol{\chi}_i^H \boldsymbol{\chi}_j (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} (d\boldsymbol{\chi}_j)^H \boldsymbol{\chi}_i \mathbf{u}_0 \right\} \\ &= \Re \left\{ \text{tr} \left((d\widetilde{\mathbf{X}}_j)^H \underbrace{-2\mathbf{\Upsilon}^{-\frac{1}{2}} \boldsymbol{\chi}_i \mathbf{u}_0 \mathbf{u}_0^H \boldsymbol{\chi}_i^H \boldsymbol{\chi}_j (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1}}_{\mathbf{C}_{j1}} \right) \right\} \end{aligned}$$

and

$$\begin{aligned} \mathbf{u}_0^H \widetilde{\mathbf{X}}_i^H \mathbf{K}_{j2} \widetilde{\mathbf{X}}_i \mathbf{u}_0 &= 2\Re \left\{ \mathbf{u}_0^H \boldsymbol{\chi}_i^H \boldsymbol{\chi}_j (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} (d\boldsymbol{\chi}_j)^H \boldsymbol{\chi}_j (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} \boldsymbol{\chi}_j^H \boldsymbol{\chi}_i \mathbf{u}_0 \right\} \\ &= \Re \left\{ \text{tr} \left((d\widetilde{\mathbf{X}}_j)^H \mathbf{C}_{j2} \right) \right\} \end{aligned}$$

where $\mathbf{C}_{j2} = 2\mathbf{\Upsilon}^{-\frac{1}{2}} \boldsymbol{\chi}_j (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} \boldsymbol{\chi}_j^H \boldsymbol{\chi}_i \mathbf{u}_0 \mathbf{u}_0^H \boldsymbol{\chi}_i^H \boldsymbol{\chi}_j (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1}$. Define $\mathbf{C}_j = \mathbf{C}_{j1} + \mathbf{C}_{j2}$. Thus,

$$df_{ij} = \Re \left\{ \text{tr} \left((d\widetilde{\mathbf{X}}_i)^H \mathbf{C}_i \right) \right\} + \Re \left\{ \text{tr} \left((d\widetilde{\mathbf{X}}_j)^H \mathbf{C}_j \right) \right\}.$$

Note that $d\widetilde{\mathbf{X}}_i = \mathbf{I}_N \otimes d\mathbf{X}_i$, then

$$df_{ij} = \Re \{ \text{tr} ((d\mathbf{X}_i)^H \overline{\mathbf{C}}_i) \} + \Re \{ \text{tr} ((d\mathbf{X}_j)^H \overline{\mathbf{C}}_j) \}$$

where $\overline{\mathbf{C}}_i = \sum_{k=1}^N \mathbf{C}_{ik}$ and \mathbf{C}_{ik} is a diagonal block of the matrix \mathbf{C}_i of size $T \times M$, i.e.,

$$\mathbf{C}_i = \begin{bmatrix} \mathbf{C}_{i1} & * & \dots & * \\ * & \mathbf{C}_{i2} & \dots & * \\ * & * & \ddots & * \\ * & * & * & \mathbf{C}_{iN} \end{bmatrix}.$$

Remark that the matrix \mathbf{C}_i is of size $TN \times MN$. Now, it is straightforward to identify the gradient. Hence, the gradient is given by

$$\nabla f_{ij}(\mathbf{x}) = \begin{bmatrix} \mathbf{0}_{(i-1)c \times 1} \\ \Re \{ \text{vec}(\overline{\mathbf{C}}_i) \} \\ \Im \{ \text{vec}(\overline{\mathbf{C}}_i) \} \\ \mathbf{0}_{(j-i-1)c \times 1} \\ \Re \{ \text{vec}(\overline{\mathbf{C}}_j) \} \\ \Im \{ \text{vec}(\overline{\mathbf{C}}_j) \} \\ \mathbf{0}_{(K-j)c \times 1} \end{bmatrix}_{2KTM \times 1}$$

$$\text{for } 1 \leq i \neq j \leq K \text{ and } c = 2TM, \text{ where } \mathbf{x} = \begin{bmatrix} \Re \{ \text{vec}(\mathbf{X}_1) \} \\ \Im \{ \text{vec}(\mathbf{X}_1) \} \\ \vdots \\ \Re \{ \text{vec}(\mathbf{X}_K) \} \\ \Im \{ \text{vec}(\mathbf{X}_K) \} \end{bmatrix}.$$

Appendix D

Mutual Information for On-Off Signaling in the Low SNR Regime

In this section, we show that the expression for the mutual information between \mathbf{Y} and \mathbf{X} up to first order in ρ , for on-off signaling and sufficiently low SNR, is given by (3.2). Since

$$I(\mathbf{Y}; \mathbf{X}) = h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}),$$

we shall find the mutual information by computing the conditional entropy of \mathbf{Y} given \mathbf{X} , i.e., $h(\mathbf{Y}|\mathbf{X})$, and in sequel, the entropy of \mathbf{Y} , i.e., $h(\mathbf{Y})$.

• We start by computing $h(\mathbf{Y}|\mathbf{X})$. Given \mathbf{X} , $\mathbf{y} = \text{vec}(\mathbf{Y})$ is a zero-mean complex Gaussian with covariance $\mathbb{E}[\mathbf{y}\mathbf{y}^H|\mathbf{X}] = \mathbf{K}_r \otimes \frac{\rho}{M} \mathbf{X} \mathbf{K}_t \mathbf{X}^H + \mathbf{\Upsilon}$ and the conditional pdf of \mathbf{Y} given \mathbf{X} is described by

$$p(\mathbf{Y}|\mathbf{X}) = \frac{\exp\{-\mathbf{y}^H (\mathbf{K}_r \otimes \frac{\rho}{M} \mathbf{X} \mathbf{K}_t \mathbf{X}^H + \mathbf{\Upsilon})^{-1} \mathbf{y}\}}{\pi^{TN} \det (\mathbf{K}_r \otimes \frac{\rho}{M} \mathbf{X} \mathbf{K}_t \mathbf{X}^H + \mathbf{\Upsilon})}. \quad (\text{D.1})$$

From (D.1) it is possible to compute the conditional entropy $h(\mathbf{Y}|\mathbf{X})$. By definition (we use logarithms to base e):

$$\begin{aligned}
h(\mathbf{Y}|\mathbf{X}) &= -\mathbb{E}_{p(\mathbf{Y},\mathbf{X})}[\log p(\mathbf{Y}|\mathbf{X})] \\
&= NT \log \pi + \mathbb{E}_{p(\mathbf{Y},\mathbf{X})}[\mathbf{y}^H \left(\mathbf{K}_r \otimes \frac{\rho}{M} \mathbf{X} \mathbf{K}_t \mathbf{X}^H + \mathbf{\Upsilon} \right)^{-1} \mathbf{y}] + \log \det(\mathbf{\Upsilon}) + \\
&\quad + \mathbb{E}_{p(\mathbf{Y},\mathbf{X})}[\log \det \left(\mathbf{I}_{TN} + \frac{\rho}{M} \mathbf{Z} \right)] \\
&\quad \left(\text{where } \mathbf{Z} = \mathbf{\Upsilon}^{-\frac{1}{2}} \left(\mathbf{K}_r \otimes \mathbf{X} \mathbf{K}_t \mathbf{X}^H \right) \mathbf{\Upsilon}^{-\frac{1}{2}} \right) \tag{D.2} \\
&= NT \log \pi + \log \det(\mathbf{\Upsilon}) + \rho^\epsilon \log \det \left(\mathbf{I}_{TN} + \frac{\rho^{1-\epsilon}}{M} \mathbf{Z}_{on} \right) + \\
&\quad + \mathbb{E}_{p(\mathbf{X})}[\mathbb{E}_{p(\mathbf{Y}|\mathbf{X})}[\text{tr} \left(\mathbf{y} \mathbf{y}^H \left(\mathbf{K}_r \otimes \frac{\rho}{M} \mathbf{X} \mathbf{K}_t \mathbf{X}^H + \mathbf{\Upsilon} \right)^{-1} \right)]] \\
&\quad \left(\text{where } \mathbf{Z}_{on} = \mathbf{\Upsilon}^{-\frac{1}{2}} \left(\mathbf{K}_r \otimes \mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H \right) \mathbf{\Upsilon}^{-\frac{1}{2}} \right) \\
&\approx NT \log \pi + \log \det(\mathbf{\Upsilon}) + \mathbb{E}_{p(\mathbf{X})}[\text{tr}(\mathbf{I}_{NT})] + \rho^\epsilon \log \prod_{i=1}^r \frac{\rho^{1-\epsilon}}{M} \lambda_i \\
&\quad \text{as } \rho^{1-\epsilon} \text{ is large for } \epsilon > 1 \text{ and } \rho \text{ small,} \\
&\quad \lambda_i \text{'s are the positive eigenvalues of } \mathbf{Z}_{on} \text{ and } r \text{ is its rank} \\
&= NT \log \pi e + \log \det(\mathbf{\Upsilon}) + o(\rho), \quad \text{since } \lim_{\rho \rightarrow 0^+} \rho^\epsilon \log \rho = 0 \text{ for } \epsilon > 1. \tag{D.3}
\end{aligned}$$

• Next, we compute $h(\mathbf{Y})$. We have

$$p(\mathbf{Y}) = \rho^\epsilon \frac{\exp\{-\mathbf{y}^H \left(\mathbf{K}_r \otimes \frac{\rho^{1-\epsilon}}{M} \mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H + \mathbf{\Upsilon} \right)^{-1} \mathbf{y}\}}{\pi^{TN} \det \left(\mathbf{K}_r \otimes \frac{\rho^{1-\epsilon}}{M} \mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H + \mathbf{\Upsilon} \right)} + (1 - \rho^\epsilon) \frac{\exp\{-\mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y}\}}{\pi^{TN} \det(\mathbf{\Upsilon})}. \tag{D.4}$$

It is not difficult to see that the second term of (D.4) is much larger than the first (since ρ is small and $\epsilon > 1$, ρ^ϵ is small, the determinant in the denominator of the first term is large, whereas the numerator is bounded above by one), therefore

$$\begin{aligned}
h(\mathbf{Y}) &= -\mathbb{E}_{p(\mathbf{Y})}[\log p(\mathbf{Y})] \\
&\approx -\log(1 - \rho^\epsilon) + NT \log \pi + \log \det(\mathbf{\Upsilon}) + \mathbb{E}_{p(\mathbf{Y})}[\mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y}] \\
&\approx \rho^\epsilon + NT \log \pi + \log \det(\mathbf{\Upsilon}) + \mathbb{E}_{p(\mathbf{Y})}[\text{tr}(\mathbf{y} \mathbf{y}^H \mathbf{\Upsilon}^{-1})] \tag{D.5}
\end{aligned}$$

$$= NT \log \pi e + \log \det(\mathbf{\Upsilon}) + \frac{\rho}{M} \text{tr}(\mathbf{\Upsilon}^{-1} (\mathbf{K}_r \otimes \mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H)) + o(\rho). \tag{D.6}$$

The step (D.6) is valid because

$$\mathbb{E}_{p(\mathbf{Y})}[\text{tr}(\mathbf{y} \mathbf{y}^H \mathbf{\Upsilon}^{-1})] = \mathbb{E}_{p(\mathbf{X}, \mathbf{h}_w, \mathbf{e})}[\text{tr} \left(f(\mathbf{X}, \mathbf{h}_w, \mathbf{e}) f(\mathbf{X}, \mathbf{h}_w, \mathbf{e})^H \mathbf{\Upsilon}^{-1} \right)] \tag{D.7}$$

where

$$f(\mathbf{X}, \mathbf{h}_w, \mathbf{e}) = \left(\mathbf{K}_r^{\frac{1}{2}} \otimes \sqrt{\frac{\rho}{M}} \mathbf{X} \mathbf{K}_t^{\frac{1}{2}} \right) \mathbf{h}_w + \mathbf{e}, \quad \mathbf{h}_w = \text{vec}(\mathbf{H}_w)$$

and $\mathbf{e} = \text{vec}(\mathbf{E})$. Since \mathbf{X} , \mathbf{h}_w and \mathbf{e} are independent random variables, from (D.7) we easily get

$$\mathbb{E}_{p(\mathbf{Y})}[\text{tr}(\mathbf{y}\mathbf{y}^H \mathbf{\Upsilon}^{-1})] = NT + \frac{\rho}{M} \text{tr}(\mathbf{\Upsilon}^{-1} (\mathbf{K}_r \otimes \mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H)). \quad (\text{D.8})$$

Substituting (D.8) in (D.5) results in (D.6). Then, using (D.3) and (D.6) we have

$$I(\mathbf{Y}; \mathbf{X}) = h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}) = \frac{\rho}{M} \text{tr}(\mathbf{\Upsilon}^{-1} (\mathbf{K}_r \otimes \mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H)) + o(\rho),$$

as required.

Appendix E

Optimization Problem for On-Off Signalling

In this Appendix, we prove that the maximum of the optimization problem defined in (3.3) is attained by the signaling scheme presented in (3.5) and (3.6).

Approach: It is known that if the constraints are independent (in the sense that each constraint function depends only on one of the variables), we can always maximize (minimize) a function by first maximizing (minimizing) over some of the variables, and then maximizing (minimizing) over the remaining ones. See, e.g., pp. 133 in [84] for more details. This simple and general procedure will help us to transform the problem in (3.3) into equivalent forms.

Let p^* represent the maximum of (3.3), i.e.,

$$p^* = \max_{\substack{\text{tr}(\mathbf{X}_{on}\mathbf{X}_{on}^H) \leq TM \\ \mathbf{\Lambda}_t \in \mathcal{D}_M, \mathbf{\Lambda}_r \in \mathcal{D}_N, \mathbf{U}_t \in \mathcal{U}_M, \mathbf{U}_r \in \mathcal{U}_N \\ \mathbf{K}_t = \mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^H, \mathbf{K}_r = \mathbf{U}_r\mathbf{\Lambda}_r\mathbf{U}_r^H}} \text{tr}(\mathbf{\Upsilon}^{-1}(\mathbf{K}_r \otimes \mathbf{X}_{on}\mathbf{K}_t\mathbf{X}_{on}^H)) \quad (\text{E.1})$$

where

$$\mathcal{D}_n = \{\mathbf{E} : n \times n \text{ diagonal matrix such that } \mathbf{E} \succeq \mathbf{0} \text{ and } \text{tr}(\mathbf{E}) = n\}, \quad (\text{E.2})$$

$$\mathcal{U}_n = \{\mathbf{F} : n \times n \text{ unitary matrix, i.e., } \mathbf{F}^H\mathbf{F} = \mathbf{F}\mathbf{F}^H = \mathbf{I}_n\}, \quad (\text{E.3})$$

\mathbf{U}_t (\mathbf{U}_r) is the $M \times M$ ($N \times N$) unitary matrix of the eigenvectors of \mathbf{K}_t (\mathbf{K}_r) and $\mathbf{\Lambda}_t$ ($\mathbf{\Lambda}_r$) is the diagonal matrix of the corresponding eigenvalues (for which we assume that they are arranged in the decreasing order). That is, $\mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^H$ ($\mathbf{U}_r\mathbf{\Lambda}_r\mathbf{U}_r^H$) represents the eigenvalue decomposition (EVD) of the matrix \mathbf{K}_t (\mathbf{K}_r). Let $\widehat{\mathbf{X}}_{on}$, $\widehat{\mathbf{U}}_r$, $\widehat{\mathbf{\Lambda}}_r$, $\widehat{\mathbf{U}}_t$ and $\widehat{\mathbf{\Lambda}}_t$

denote optimal values for the matrices \mathbf{X}_{on} , \mathbf{U}_r , $\mathbf{\Lambda}_r$, \mathbf{U}_t and $\mathbf{\Lambda}_t$, respectively. There are two main steps in the proof. In the first step, $\widehat{\mathbf{\Lambda}}_r$, $\widehat{\mathbf{U}}_t$ and $\widehat{\mathbf{\Lambda}}_t$ will be determined for given \mathbf{X}_{on} , \mathbf{U}_r , whereas the second step furnishes $\widehat{\mathbf{X}}_{on}$ and $\widehat{\mathbf{U}}_r$.

Step 1. We start by rewriting the problem in (E.1) in the equivalent form,

$$p^* = \max_{\substack{\text{tr}(\mathbf{X}_{on}\mathbf{X}_{on}^H) \leq TM \\ \mathbf{\Lambda}_t \in \mathcal{D}_M, \mathbf{\Lambda}_r \in \mathcal{D}_N, \mathbf{U}_t \in \mathcal{U}_M, \mathbf{U}_r \in \mathcal{U}_N}} \sum_{i=1}^N \lambda_i \text{tr} \left(\mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H \mathbf{F}_i \widehat{\mathbf{\Upsilon}}^{-1} \mathbf{F}_i^H \right), \quad (\text{E.4})$$

where λ_i is the diagonal element of $\mathbf{\Lambda}_r$ on the position (i, i) ,

$$\widehat{\mathbf{\Upsilon}} = (\mathbf{U}_r^H \otimes \mathbf{I}_T) \mathbf{\Upsilon} (\mathbf{U}_r \otimes \mathbf{I}_T), \quad \mathbf{F}_i = \mathbf{f}_i^T \otimes \mathbf{I}_T, \quad (\text{E.5})$$

and \mathbf{f}_i represents the i -th column of \mathbf{I}_N .

Idea: The idea that we shall use in order to determine the optimal $\mathbf{\Lambda}_r$, \mathbf{U}_t and $\mathbf{\Lambda}_t$ is to relax the problem in (E.4) (hence, by doing this, the optimal value of the new, relaxed problem will be equal or larger than of the original one in (E.4)), and then to show that the maximum of the relaxed problem is also attainable for the original problem. The relaxed problem is defined as

$$q^* = \max_{\substack{\mathbf{A} = \mathbf{A}^H \succeq \mathbf{0}, \text{tr}(\mathbf{A}) \leq TM^2 \\ \mathbf{\Lambda}_r \in \mathcal{D}_N, \mathbf{U}_r \in \mathcal{U}_N}} \sum_{i=1}^N \lambda_i \text{tr} \left(\mathbf{A} \mathbf{F}_i \widehat{\mathbf{\Upsilon}}^{-1} \mathbf{F}_i^H \right), \quad (\text{E.6})$$

The motivation behind the definition of the new variable \mathbf{A} , a $T \times T$ Hermitian positive semidefinite matrix that obeys the constraint $\text{tr}(\mathbf{A}) \leq TM^2$, is the fact that $\mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H$ is also a Hermitian positive semidefinite matrix with the same upper bound on the trace constraint, i.e., $\text{tr}(\mathbf{X}_{on} \mathbf{K}_t \mathbf{X}_{on}^H) \leq TM^2$ (which is easily verified since $\text{tr}(\mathbf{X}_{on} \mathbf{X}_{on}^H) \leq TM$). Clearly, $p^* \leq q^*$. The problem in (E.6) can be readily maximized by maximizing it first w.r.t. $\mathbf{\Lambda}_r$, and then w.r.t. \mathbf{A} . We shall maximize (E.6) w.r.t. $\mathbf{\Lambda}_r$ by invoking the following fact: if a convex function f is defined on the bounded, closed convex set Ω , then, if f has a maximum over Ω it is achieved at an extreme point of Ω (see, e.g., pp. 119 in [92]). In our case, the function f is a linear combination of λ_i 's and the set $\Omega = \{(\lambda_1, \dots, \lambda_N) : N \times 1 \text{ vectors such that } \sum_{i=1}^N \lambda_i = N \text{ and } \lambda_i \geq 0\}$. It is readily seen that the extreme points of Ω are the vectors whose all entries are zero except one

which is equal to N . Regarding maximization of (E.6) w.r.t. \mathbf{A} , the Fan's theorem is used, see pp. 17 in [95]. That is, beamforming in the direction of $\lambda_{\max}(\mathbf{F}_i \hat{\mathbf{Y}}^{-1} \mathbf{F}_i^H)$ is performed such that an upper bound on $\text{tr}(\mathbf{A} \mathbf{F}_i \hat{\mathbf{Y}}^{-1} \mathbf{F}_i^H)$ is attained ($\text{tr}(\mathbf{A} \mathbf{F}_i \hat{\mathbf{Y}}^{-1} \mathbf{F}_i^H) \leq \text{tr}(\mathbf{A}) \lambda_{\max}(\mathbf{F}_i \hat{\mathbf{Y}}^{-1} \mathbf{F}_i^H)$). Hence,

$$\left[\hat{\mathbf{\Lambda}}_r \right]_{i,i} = N \delta_{ii^*}, \quad \hat{\mathbf{A}} = TM^2 \mathbf{U}_{i^*} \mathbf{K}_A \mathbf{U}_{i^*}^H, \quad (\text{E.7})$$

where the matrix \mathbf{U}_{i^*} is the $T \times T$ unitary matrix of the eigenvectors of $\mathbf{F}_{i^*} \hat{\mathbf{Y}}^{-1} \mathbf{F}_{i^*}^H$ (that is, we assume that $\mathbf{U}_{i^*} \mathbf{D}_{i^*} \mathbf{U}_{i^*}^H$ represents the EVD of the matrix $\mathbf{F}_{i^*} \hat{\mathbf{Y}}^{-1} \mathbf{F}_{i^*}^H$ where the matrix \mathbf{D}_{i^*} is the diagonal matrix of the corresponding eigenvalues sorted in the decreasing order), the $T \times T$ constant matrix \mathbf{K}_A has all the entries equal to zero except the (1,1) entry which is one, $\delta_{ij} = 1$ for $i = j$ and zero otherwise, the matrix \mathbf{F}_{i^*} is defined as in (E.5), and

$$i^* = \arg \max_{i=1, \dots, N} \lambda_{\max}(\mathbf{F}_i \hat{\mathbf{Y}}^{-1} \mathbf{F}_i^H). \quad (\text{E.8})$$

For the choice in (E.7) and from (E.6), we get

$$q^* = NTM^2 \max_{\mathbf{U}_r \in \mathcal{U}_N} \lambda_{\max}(\mathbf{F}_{i^*} \hat{\mathbf{Y}}^{-1} \mathbf{F}_{i^*}^H). \quad (\text{E.9})$$

Next, we prove that the optimal value of the relaxed problem in (E.9) is attainable for the original problem in (E.4). To see this, we define $\hat{\mathbf{X}}_{on}$, $\hat{\mathbf{U}}_t$ and $\hat{\mathbf{\Lambda}}_t$ in the following way,

$$\hat{\mathbf{X}}_{on} = \sqrt{TM} \mathbf{U}_{i^*} \mathbf{K}_X, \quad \hat{\mathbf{U}}_t = \mathbf{I}_M, \quad \left[\hat{\mathbf{\Lambda}}_t \right]_{j,j} = M \delta_{j1}, \quad \text{for } j = 1, \dots, M \quad (\text{E.10})$$

together with $\hat{\mathbf{\Lambda}}_r$ as in (E.7). The $T \times M$ constant matrix \mathbf{K}_X has all the entries equal to zero except the (1,1) entry which is one, and i^* is defined in (E.8). In that case,

$$p^* = q^* = NTM^2 \max_{\mathbf{U}_r \in \mathcal{U}_N} \lambda_{\max}(\mathbf{F}_{i^*} \hat{\mathbf{Y}}^{-1} \mathbf{F}_{i^*}^H). \quad (\text{E.11})$$

Remark that for the choice in (E.10) the power constraint on the transmitted codeword is satisfied with equality, $\hat{\mathbf{K}}_t$ is a diagonal matrix, and both $\hat{\mathbf{K}}_t$ and $\hat{\mathbf{K}}_r$ are rank one matrices.

Step 2. In the first step, we have determined $\hat{\mathbf{K}}_t$ and $\hat{\mathbf{\Lambda}}_r$. It remains now to compute $\hat{\mathbf{X}}_{on}$ and $\hat{\mathbf{U}}_r$. First, note that from (E.10) we can write

$$\hat{\mathbf{X}}_{on} = \sqrt{TM} [\mathbf{x} \quad \mathbf{0}_{T \times (M-1)}], \quad (\text{E.12})$$

where the unit norm vector $\mathbf{x} \in \mathbb{C}^T$ is the eigenvector associated to the maximal eigenvalue of the matrix $\mathbf{F}_{i^*} \widehat{\mathbf{\Upsilon}}^{-1} \mathbf{F}_{i^*}^H$, i.e., it is the first column of the unitary matrix \mathbf{U}_{i^*} . Hence, we need to find such a unit norm vector \mathbf{x} and a unitary matrix \mathbf{U}_r that maximize p^* in (E.11). Let

$$\widehat{\lambda} = \max_{\mathbf{U}_r \in \mathcal{U}_N} \lambda_{\max} \left(\mathbf{F}_{i^*} \widehat{\mathbf{\Upsilon}}^{-1} \mathbf{F}_{i^*}^H \right).$$

Then the following equalities hold:

$$\widehat{\lambda} = \max_{\mathbf{U}_r \in \mathcal{U}_N} \max_{i=1, \dots, N} \max_{\mathbf{x} \in \mathbb{C}^T, \|\mathbf{x}\|=1} \mathbf{x}^H \left(\mathbf{F}_i \widehat{\mathbf{\Upsilon}}^{-1} \mathbf{F}_i^H \right) \mathbf{x} \quad (\text{E.13})$$

$$= \max_{\mathbf{U}_r \in \mathcal{U}_N} \max_{i=1, \dots, N} \max_{\mathbf{x} \in \mathbb{C}^T, \|\mathbf{x}\|=1} \mathbf{x}^H (\mathbf{f}_i^T \otimes \mathbf{I}_T) (\mathbf{U}_r^H \otimes \mathbf{I}_T) \mathbf{\Upsilon}^{-1} (\mathbf{U}_r \otimes \mathbf{I}_T) (\mathbf{f}_i \otimes \mathbf{I}_T) \mathbf{x}$$

(from (E.5), $\widehat{\mathbf{\Upsilon}} = (\mathbf{U}_r^H \otimes \mathbf{I}_T) \mathbf{\Upsilon} (\mathbf{U}_r \otimes \mathbf{I}_T)$ and $\mathbf{F}_i = \mathbf{f}_i^T \otimes \mathbf{I}_T$)

$$= \max_{\mathbf{U}_r = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_N] \in \mathcal{U}_N} \max_{i=1, \dots, N} \max_{\mathbf{x} \in \mathbb{C}^T, \|\mathbf{x}\|=1} (\mathbf{u}_i^H \otimes \mathbf{x}^H) \mathbf{\Upsilon}^{-1} (\mathbf{u}_i \otimes \mathbf{x}) \quad (\text{E.14})$$

$$= \max_{\substack{\mathbf{u} \in \mathbb{C}^N, \|\mathbf{u}\|=1, \\ \mathbf{x} \in \mathbb{C}^T, \|\mathbf{x}\|=1}} (\mathbf{u} \otimes \mathbf{x})^H \mathbf{\Upsilon}^{-1} (\mathbf{u} \otimes \mathbf{x}). \quad (\text{E.15})$$

The passage from (E.14) to (E.15) is valid due to the following arguments. Let

$$p_1 = \max_{\mathbf{U}_r = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_N] \in \mathcal{U}_N} \max_{i=1, \dots, N} \max_{\mathbf{x} \in \mathbb{C}^T, \|\mathbf{x}\|=1} f_{1i}(\mathbf{u}_i, \mathbf{x}) \quad (\text{E.16})$$

and

$$p_2 = \max_{\substack{\mathbf{u} \in \mathbb{C}^N, \|\mathbf{u}\|=1, \\ \mathbf{x} \in \mathbb{C}^T, \|\mathbf{x}\|=1}} f_2(\mathbf{u}, \mathbf{x}) \quad (\text{E.17})$$

where the functions $f_{1i}(\mathbf{u}_i, \mathbf{x}) = (\mathbf{u}_i^H \otimes \mathbf{x}^H) \mathbf{\Upsilon}^{-1} (\mathbf{u}_i \otimes \mathbf{x})$, for $i = 1, \dots, N$, and $f_2(\mathbf{u}, \mathbf{x}) = (\mathbf{u}^H \otimes \mathbf{x}^H) \mathbf{\Upsilon}^{-1} (\mathbf{u} \otimes \mathbf{x})$. Remark that $f_{1i}(\mathbf{u}_i, \mathbf{x})$, for $i = 1, \dots, N$, and $f_2(\mathbf{u}, \mathbf{x})$ are continuous functions (as a consequence, $f_1(\mathbf{u}_1, \dots, \mathbf{u}_N, \mathbf{x}) = \max(f_{11}(\mathbf{u}_1, \mathbf{x}), \dots, f_{1N}(\mathbf{u}_N, \mathbf{x}))$ is also continuous because pointwise maximum of continuous functions is a continuous function). Moreover, $f_1(\mathbf{u}_1, \dots, \mathbf{u}_N, \mathbf{x})$ and $f_2(\mathbf{u}, \mathbf{x})$ are defined over the product spaces

$\mathcal{CS}_1 = \{(\mathbf{J}, \mathbf{t}) : \text{where } \mathbf{J} \in \mathcal{U}_N \text{ and } \mathbf{t} \in \mathbb{S}^{2T-1}\}$ and $\mathcal{CS}_2 = \{(\mathbf{w}, \mathbf{v}) : \text{where } \mathbf{w} \in \mathbb{S}^{2N-1} \text{ and } \mathbf{v} \in \mathbb{S}^{2T-1}\}$, respectively (the symbol \mathbb{S}^{n-1} denotes the unit sphere in \mathbb{R}^n), that are compact spaces (both \mathcal{U}_m and \mathbb{S}^n , for some $m, n \in \mathbb{N}$, are subclasses of the Stiefel manifold which is a compact space, and every product of compact spaces is a compact space by the Tychonoff's theorem, see [96]). Due to the Bolzano-Weierstrass' fundamental existence theorem (see, e.g., pp. 654 in [91]) we know that both $f_1(\mathbf{u}_1, \dots, \mathbf{u}_N, \mathbf{x})$ and $f_2(\mathbf{u}, \mathbf{x})$ achieve the minimum and the maximum on \mathcal{CS}_1 and \mathcal{CS}_2 , respectively. Now, we need to show that $p_1 = p_2$. To prove this, we show that $p_1 \leq p_2$, but also $p_2 \leq p_1$. Let's assume w.l.o.g. that the maximum in (E.16) is achieved for some $\hat{\mathbf{u}}_1$ (the unit vector that is the first column of the unitary matrix \mathbf{U}_r) and $\hat{\mathbf{x}}$. Then, by taking $\mathbf{u} = \hat{\mathbf{u}}_1$ and $\mathbf{x} = \hat{\mathbf{x}}$ in (E.17), we have $p_1 = f_2(\hat{\mathbf{u}}_1, \hat{\mathbf{x}}) \leq p_2$. Similarly, let's assume that the maximum in (E.17) is achieved for some $\hat{\mathbf{u}}$ and $\hat{\mathbf{x}}$. We know that we can construct an unitary matrix \mathbf{U}_r with $\hat{\mathbf{u}}$ as its first column. Hence, $p_2 = f_1(\hat{\mathbf{u}}, \mathbf{u}_1, \dots, \mathbf{u}_{N-1}, \hat{\mathbf{x}}) \leq p_1$ where \mathbf{u}_i , for $i = 1, \dots, N-1$, is the $(i+1)$ -th column of the matrix \mathbf{U}_r . Thus, $p_1 = p_2$.

Let $(\hat{\mathbf{u}}, \hat{\mathbf{x}})$ be the solution pair of (E.15), i.e.,

$$(\hat{\mathbf{u}}, \hat{\mathbf{x}}) = \underset{\substack{\mathbf{u} \in \mathbb{C}^N, \|\mathbf{u}\| = 1, \\ \mathbf{x} \in \mathbb{C}^T, \|\mathbf{x}\| = 1}}{\arg \max} (\mathbf{u} \otimes \mathbf{x})^H \mathbf{\Upsilon}^{-1} (\mathbf{u} \otimes \mathbf{x}). \quad (\text{E.18})$$

Then, using (E.12) and the fact that $\hat{\mathbf{u}}$ is the i^* -th column of $\widehat{\mathbf{U}}_r$ (where i^* is defined in (E.8) and due to the passage from (E.14) to (E.15) we can w.l.o.g. assume that $i^* = 1$, although any other choice of i^* would just change $\widehat{\mathbf{U}}_r$ but not $\widehat{\mathbf{K}}_r$), we have

$$\widehat{\mathbf{X}}_{on} = \sqrt{TM} [\hat{\mathbf{x}} \quad \mathbf{0}_{T \times (M-1)}], \quad \widehat{\mathbf{K}}_r = N \hat{\mathbf{u}} \hat{\mathbf{u}}^H. \quad (\text{E.19})$$

Equations (E.7), (E.10), (E.18) and (E.19) complete the proof.

Appendix F

Mutual Information for Gaussian Signaling in the Low SNR Regime

In this section, we show that the expression for the mutual information between \mathbf{Y} and \mathbf{X} up to second order in ρ , for Gaussian signaling and sufficiently low SNR, is given by (3.11). But, before that, we prove the validity of (3.10).

(A) Here, we shall show that (3.10) holds. As in Appendix D, we find the mutual information by computing the conditional entropy $h(\mathbf{Y}|\mathbf{X})$ and the entropy of \mathbf{Y} , i.e., $h(\mathbf{Y})$.

- We start by computing $h(\mathbf{Y}|\mathbf{X})$. By repeating the analysis for the on-off signaling in Appendix D, we readily find

$$\begin{aligned} h(\mathbf{Y}|\mathbf{X}) &= NT \log \pi e + \log \det(\mathbf{\Upsilon}) + \mathbb{E}[\log \det \left(\mathbf{I}_{TN} + \frac{\rho}{M} \mathbf{Z} \right)] \\ &\approx NT \log \pi e + \log \det(\mathbf{\Upsilon}) + \frac{\rho}{M} \text{tr}(\mathbb{E}[\mathbf{Z}]) - \frac{\rho^2}{2M^2} \text{tr}(\mathbb{E}[\mathbf{Z}^2]), \end{aligned} \quad (\text{F.1})$$

where the matrix \mathbf{Z} is defined in (D.2).

- Next, we compute $h(\mathbf{Y})$. Since $\text{vec}(\mathbf{X}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{P})$ for some matrix \mathbf{P} , and $p(\mathbf{Y}|\mathbf{X})$ in (D.1) is a function whose derivatives w.r.t. ρ at $\rho = 0$ can be calculated, we can expand $p(\mathbf{Y})$ and $\log p(\mathbf{Y})$ to second order as

$$p(\mathbf{Y}) = p(\mathbf{Y}, 0) + \rho p'(\mathbf{Y}, 0) + \frac{\rho^2}{2} p''(\mathbf{Y}, 0) + o(\rho^2), \quad (\text{F.2})$$

$$\log p(\mathbf{Y}) \approx \log p(\mathbf{Y}, 0) + \rho \frac{p'(\mathbf{Y}, 0)}{p(\mathbf{Y}, 0)} + \frac{\rho^2}{2} \left(\frac{p''(\mathbf{Y}, 0)}{p(\mathbf{Y}, 0)} - \left(\frac{p'(\mathbf{Y}, 0)}{p(\mathbf{Y}, 0)} \right)^2 \right) \quad (\text{F.3})$$

where $p'(\mathbf{Y}, 0)$ and $p''(\mathbf{Y}, 0)$ denote the first and second partial derivative of $p(\mathbf{Y})$ w.r.t. ρ evaluated at $\rho = 0$, respectively. Due to (F.2) and (F.3) we have the following approxi-

mation of $h(\mathbf{Y})$:

$$h(\mathbf{Y}) = -\mathbb{E}_{p(\mathbf{Y})}[\log p(\mathbf{Y})] \approx a + \rho b + \frac{\rho^2}{2}c \quad (\text{F.4})$$

where

$$a = -\int p(\mathbf{Y}, \mathbf{0}) \log p(\mathbf{Y}, \mathbf{0}) d\mathbf{Y}, \quad b = -\int (p'(\mathbf{Y}, \mathbf{0}) \log p(\mathbf{Y}, \mathbf{0}) + p'(\mathbf{Y}, \mathbf{0})) d\mathbf{Y}, \quad (\text{F.5})$$

$$c = -\int \left(p''(\mathbf{Y}, \mathbf{0}) + p''(\mathbf{Y}, \mathbf{0}) \log p(\mathbf{Y}, \mathbf{0}) + \frac{(p'(\mathbf{Y}, \mathbf{0}))^2}{p(\mathbf{Y}, \mathbf{0})} \right) d\mathbf{Y}. \quad (\text{F.6})$$

We next show that the integrals in (F.6) comprising $p''(\mathbf{Y}, \mathbf{0})$ are equal to zero. To this end, remark that, using (F.2)

$$\int p(\mathbf{Y}) d\mathbf{Y} = \int \left(p(\mathbf{Y}, 0) + \rho p'(\mathbf{Y}, 0) + \frac{\rho^2}{2} p''(\mathbf{Y}, 0) + o(\rho^2) \right) d\mathbf{Y} = 1$$

that implies

$$\int p^{(n)}(\mathbf{Y}, 0) d\mathbf{Y} = 0, \quad n = 1, 2. \quad (\text{F.7})$$

since $p(\mathbf{Y}, 0) = p(\mathbf{Y}, 0|\mathbf{X})$ is pdf of a zero-mean complex Gaussian with covariance $\mathbf{\Upsilon}$.

Then, from (F.2) and by repeating the analysis of (D.7) and (D.8), it holds

$$\begin{aligned} \mathbb{E}_{p(\mathbf{Y})}[\mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y}] &= \int \mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y} \left(p(\mathbf{Y}, 0) + \rho p'(\mathbf{Y}, 0) + \frac{\rho^2}{2} p''(\mathbf{Y}, 0) + o(\rho^2) \right) d\mathbf{Y} \\ &= NT + \frac{\rho}{M} \text{tr}(\mathbb{E}[\mathbf{Z}]). \end{aligned} \quad (\text{F.8})$$

From (F.8), we draw the following conclusions:

$$\int \mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y} p(\mathbf{Y}, 0) d\mathbf{Y} = NT, \quad \int \mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y} p'(\mathbf{Y}, 0) d\mathbf{Y} = \frac{1}{M} \text{tr}(\mathbb{E}[\mathbf{Z}]), \quad (\text{F.9})$$

$$\int \mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y} p''(\mathbf{Y}, 0) d\mathbf{Y} = 0. \quad (\text{F.10})$$

Thus, equations (F.9) and (F.10) combined with (F.4), (F.5), (F.6) and (F.7), together with the fact that $\log p(\mathbf{Y}, 0) = -TN \log \pi - \log \det(\mathbf{\Upsilon}) - \mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y}$, lead to

$$h(\mathbf{Y}) \approx NT \log \pi e + \log \det(\mathbf{\Upsilon}) + \frac{\rho}{M} \text{tr}(\mathbb{E}[\mathbf{Z}]) - \frac{\rho^2}{2} \int \frac{(p'(\mathbf{Y}, \mathbf{0}))^2}{p(\mathbf{Y}, \mathbf{0})} d\mathbf{Y}. \quad (\text{F.11})$$

We now approximate $p(\mathbf{Y}|\mathbf{X})$ in (D.1) to first order in ρ . To this end, we use the following approximations:

$$\left(\mathbf{I}_{TN} + \frac{\rho}{M} \mathbf{Z} \right)^{-1} \approx \mathbf{I}_{TN} - \frac{\rho}{M} \mathbf{Z}, \quad (\text{F.12})$$

$$\left(\det\left(\mathbf{I}_{TN} + \frac{\rho}{M}\mathbf{Z}\right)\right)^{-1} \approx \left(1 + \frac{\rho}{M}\text{tr}(\mathbf{Z})\right)^{-1} \approx 1 - \frac{\rho}{M}\text{tr}(\mathbf{Z}), \quad (\text{F.13})$$

$$e^{\rho t} \approx 1 + \rho t \quad (\text{for some } t) \quad (\text{F.14})$$

which are valid for sufficiently small ρ . Hence, using (D.1),(F.12), (F.13) and (F.14), the approximated $p(\mathbf{Y}|\mathbf{X})$ is given by

$$p(\mathbf{Y}|\mathbf{X}) \approx \frac{1}{\pi^{TN} \det \mathbf{\Upsilon}} e^{-\mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y}} \left(1 + \frac{\rho}{M} \mathbf{y}^H \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{Z} \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{y}\right) \left(1 - \frac{\rho}{M} \text{tr}(\mathbf{Z})\right), \quad (\text{F.15})$$

where the matrix \mathbf{Z} is defined in (D.2). From (F.15), $p'(\mathbf{Y}|\mathbf{X}, \rho = 0)$ (henceforth, we write $p'(\mathbf{Y}|\mathbf{X}, 0)$) is readily calculated,

$$p'(\mathbf{Y}|\mathbf{X}, 0) = \frac{1}{M \pi^{TN} \det \mathbf{\Upsilon}} e^{-\mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y}} \left(\mathbf{y}^H \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{Z} \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{y} - \text{tr}(\mathbf{Z})\right). \quad (\text{F.16})$$

In order to find $p'(\mathbf{Y}, \rho = 0)$ we take the expectation of (F.16) over \mathbf{X} leading to

$$p'(\mathbf{Y}, 0) = \frac{1}{M \pi^{TN} \det \mathbf{\Upsilon}} e^{-\mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y}} \left(\mathbf{y}^H \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{K} \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{y} - \text{tr}(\mathbf{E}[\mathbf{Z}])\right) \quad (\text{F.17})$$

where $\mathbf{K} = \mathbf{E}[\mathbf{Z}]$. This implies

$$\begin{aligned} \int \frac{(p'(\mathbf{Y}, 0))^2}{p(\mathbf{Y}, 0)} d\mathbf{Y} &= \frac{1}{M^2} \int \frac{e^{-\mathbf{y}^H \mathbf{\Upsilon}^{-1} \mathbf{y}}}{\pi^{TN} \det \mathbf{\Upsilon}} \left(\mathbf{y}^H \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{K} \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{y} - \text{tr}(\mathbf{E}[\mathbf{Z}])\right)^2 d\mathbf{Y} \\ &\quad (\text{change of variables: } \mathbf{v} = \mathbf{\Upsilon}^{-\frac{1}{2}} \mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{TN}) \text{ for } \rho = 0) \\ &= \frac{1}{M^2} \int \frac{e^{-\mathbf{v}^H \mathbf{v}}}{\pi^{TN}} (\mathbf{v}^H \mathbf{K} \mathbf{v} - \text{tr}(\mathbf{E}[\mathbf{Z}]))^2 d\mathbf{v} \\ &= \frac{1}{M^2} \left(\mathbf{E}_{p(\mathbf{v})}[(\mathbf{v}^H \mathbf{K} \mathbf{v})^2] - 2 \cdot \text{tr}(\mathbf{E}[\mathbf{Z}]) \mathbf{E}_{p(\mathbf{v})}[\mathbf{v}^H \mathbf{K} \mathbf{v}] + \text{tr}^2(\mathbf{E}[\mathbf{Z}])\right) \\ &= \frac{1}{M^2} \text{tr} \left((\mathbf{E}[\mathbf{Z}])^2 \right). \end{aligned} \quad (\text{F.18})$$

The step (F.18) is valid due to the following result (see, e.g., pp. 564 in [93]): if $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, then for \mathbf{A}, \mathbf{B} Hermitian matrices

$$\mathbf{E}[\mathbf{y}^H \mathbf{A} \mathbf{y} \mathbf{y}^H \mathbf{B} \mathbf{y}] = \text{tr}(\mathbf{A}\mathbf{C})\text{tr}(\mathbf{B}\mathbf{C}) + \text{tr}(\mathbf{A}\mathbf{C}\mathbf{B}\mathbf{C}). \quad (\text{F.19})$$

Equations (F.18) and (F.11) combined with (F.1) result in (3.10) as required.

(B) We now prove that (3.11) holds.

First, we calculate $\text{tr}(\mathbf{E}[\mathbf{Z}^2])$, where $\mathbf{Z} = \mathbf{\Upsilon}^{-\frac{1}{2}} (\mathbf{K}_r \otimes \mathbf{X} \mathbf{K}_t \mathbf{X}^H) \mathbf{\Upsilon}^{-\frac{1}{2}}$. Remark that $\text{tr}(\mathbf{E}[\mathbf{Z}^2]) = \text{tr}(\mathbf{E}[\mathbf{F}])$ where $\mathbf{F} = \left(\widehat{\mathbf{\Upsilon}}^{-1} (\mathbf{\Lambda}_r \otimes \widetilde{\mathbf{X}} \widetilde{\mathbf{X}}^H)\right)^2$, the matrix $\widehat{\mathbf{\Upsilon}}$ is defined in (E.5),

$\mathbf{K}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$ (as in Appendix E) and $\widetilde{\mathbf{X}} = \mathbf{X} \mathbf{K}_t^{\frac{1}{2}}$. If the vector $\widetilde{\mathbf{x}}_i$, for $i = 1, \dots, M$, represents the i -th column of the matrix $\widetilde{\mathbf{X}}$, then it holds

$$\begin{aligned} \text{tr}(\mathbb{E}[\mathbf{Z}^2]) &= \sum_{i=1}^M \sum_{j=1}^M \mathbb{E}[\text{tr}(\widehat{\mathbf{\Upsilon}}^{-1}(\mathbf{\Lambda}_r \otimes \widetilde{\mathbf{x}}_i \widetilde{\mathbf{x}}_i^H) \widehat{\mathbf{\Upsilon}}^{-1}(\mathbf{\Lambda}_r \otimes \widetilde{\mathbf{x}}_j \widetilde{\mathbf{x}}_j^H))] \\ &= \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^N \sum_{z=1}^N \lambda_k \lambda_z \mathbb{E}[\text{tr}(\widehat{\mathbf{\Upsilon}}_{kz} \widetilde{\mathbf{x}}_i \widetilde{\mathbf{x}}_i^H \widehat{\mathbf{\Upsilon}}_{zk} \widetilde{\mathbf{x}}_j \widetilde{\mathbf{x}}_j^H)], \end{aligned}$$

where $\widehat{\mathbf{\Upsilon}}_{kz} = \mathbf{F}_k \widehat{\mathbf{\Upsilon}}^{-1} \mathbf{F}_z^H$, for $k, z = 1, \dots, N$, the matrix \mathbf{F}_k is defined in (E.5), and λ_k 's are the eigenvalues of the matrix $\mathbf{\Lambda}_r$. Let $\Phi(\mathbf{A}, \mathbf{B}) = \mathbb{E}[\mathbf{y}^H \mathbf{A} \mathbf{y} \mathbf{y}^H \mathbf{B} \mathbf{y}]$ where $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$. Then, by using (F.19), it is easy to check that

$$\begin{aligned} \Phi(\mathbf{A}, \mathbf{A}^H) &= \frac{1}{4} (\Phi(\mathbf{A} + \mathbf{A}^H, \mathbf{A} + \mathbf{A}^H) + \Phi(j\mathbf{A} - j\mathbf{A}^H, j\mathbf{A} - j\mathbf{A}^H)) \\ &= \text{tr}(\mathbf{A}\mathbf{C})\text{tr}(\mathbf{A}^H\mathbf{C}) + \text{tr}(\mathbf{A}\mathbf{C}\mathbf{A}^H\mathbf{C}) \end{aligned}$$

for any matrix \mathbf{A} . This, and using the facts that $\widehat{\mathbf{\Upsilon}}_{kz} = \widehat{\mathbf{\Upsilon}}_{zk}^H$ and $\widetilde{\mathbf{x}}_i = \mathbf{E}_i \widetilde{\mathbf{x}}$ (where $\widetilde{\mathbf{x}} = \text{vec}(\widetilde{\mathbf{X}})$ and \mathbf{E}_i is defined in (3.12)), implies

$$\text{tr}(\mathbb{E}[\mathbf{Z}^2]) = \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^N \sum_{z=1}^N \lambda_k \lambda_z \left(\text{tr}(\widehat{\mathbf{\Upsilon}}_{kz} \widetilde{\mathbf{P}}_{ij}) \text{tr}(\widehat{\mathbf{\Upsilon}}_{zk} \widetilde{\mathbf{P}}_{ji}) + \text{tr}(\widehat{\mathbf{\Upsilon}}_{kz} \widetilde{\mathbf{P}}_{ii} \widehat{\mathbf{\Upsilon}}_{zk} \widetilde{\mathbf{P}}_{jj}) \right), \quad (\text{F.20})$$

where $\widetilde{\mathbf{P}}_{ij} = \mathbf{E}_i \widetilde{\mathbf{P}} \mathbf{E}_j^H$ and $\widetilde{\mathbf{P}} = \left((\mathbf{K}_t^T)^{\frac{1}{2}} \otimes \mathbf{I}_T \right) \mathbf{P} \left((\mathbf{K}_t^T)^{\frac{1}{2}} \otimes \mathbf{I}_T \right)$.

It can readily be shown that

$$\begin{aligned} \text{tr}(\mathbb{E}[\mathbf{Z}^2]) &= \sum_{i=1}^M \sum_{j=1}^M \text{tr}(\mathbf{\Upsilon}^{-1}(\mathbf{K}_r \otimes \mathbf{E}_i \widetilde{\mathbf{P}} \mathbf{E}_i^H) \mathbf{\Upsilon}^{-1}(\mathbf{K}_r \otimes \mathbf{E}_j \widetilde{\mathbf{P}} \mathbf{E}_j^H)) \\ &= \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^N \sum_{z=1}^N \lambda_k \lambda_z \text{tr}(\widehat{\mathbf{\Upsilon}}_{kz} \widetilde{\mathbf{P}}_{ii} \widehat{\mathbf{\Upsilon}}_{zk} \widetilde{\mathbf{P}}_{jj}). \end{aligned} \quad (\text{F.21})$$

Combining (F.20) with (F.21) results in (3.11). This concludes the proof.

Appendix G

Optimization Problem for Gaussian Signalling

In this Appendix, we prove that the maximum of the optimization problem defined in (3.13) is attained by $\widehat{\mathbf{P}}$, defined in (3.14), and by $\widehat{\mathbf{K}}_t$ and $\widehat{\mathbf{K}}_r$ presented in (3.5). As in Appendix E, let $\mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^H$ ($\mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$) represents the EVD of the matrix \mathbf{K}_t (\mathbf{K}_r), and let $\widehat{\mathbf{P}}$, $\widehat{\mathbf{U}}_r$, $\widehat{\mathbf{\Lambda}}_r$, $\widehat{\mathbf{U}}_t$ and $\widehat{\mathbf{\Lambda}}_t$ denote the optimal values of the matrices \mathbf{P} , \mathbf{U}_r , $\mathbf{\Lambda}_r$, \mathbf{U}_t and $\mathbf{\Lambda}_t$, respectively. We repeat the analysis presented in Appendix E. In the first step, $\widehat{\mathbf{\Lambda}}_r$ is calculated, whereas the second step determines $\widehat{\mathbf{P}}$, $\widehat{\mathbf{U}}_t$, $\widehat{\mathbf{\Lambda}}_t$ and $\widehat{\mathbf{U}}_r$.

Step 1. Note that

$$\sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^N \sum_{z=1}^N \lambda_k \lambda_z \text{tr}(\widehat{\mathbf{Y}}_{kz} \widetilde{\mathbf{P}}_{ij}) \text{tr}(\widehat{\mathbf{Y}}_{zk} \widetilde{\mathbf{P}}_{ji}) = \lambda^T \mathbf{A} \lambda \quad (\text{G.1})$$

where the vector $\lambda = [\lambda_1 \ \lambda_2 \ \dots \ \lambda_N]^T$. The entry (k, z) , for $k, z = 1, \dots, N$, of the $N \times N$ Hermitian matrix \mathbf{A} is given by $[\mathbf{A}]_{k,z} = \sum_{i=1}^M \sum_{j=1}^M \text{tr}(\widehat{\mathbf{Y}}_{kz} \widetilde{\mathbf{P}}_{ij}) \text{tr}(\widehat{\mathbf{Y}}_{zk} \widetilde{\mathbf{P}}_{ji})$.

Approach: We shall show that the matrix \mathbf{A} is positive semidefinite. In that case, the function $\lambda^T \mathbf{A} \lambda$ is convex on \mathcal{D}_N defined in (E.2), and the maximum is achieved at an extreme point of \mathcal{D}_N (see pp. 119 in [92]). To this end, let's introduce the result that will be used in the proof: if a $NT \times NT$ matrix \mathbf{Q} is positive semidefinite, then $N \times N$ matrix \mathbf{R} , with

$$[\mathbf{R}]_{k,z} = \sum_{i=1}^T \sum_{j=1}^T \mathbf{Q}((k-1)T + i, (z-1)T + j)$$

for $k, z = 1, \dots, N$, is also positive semidefinite. In words, $[\mathbf{R}]_{k,z}$ is equal to the sum of all elements of the (k, z) -th block of the matrix \mathbf{Q} (the matrix \mathbf{Q} consists in N^2 disjoint

blocks where each of the blocks is of size $T \times T$). Now, remark that

$$\mathbf{R} = (\mathbf{I}_N \otimes \mathbf{1}_T)^T \mathbf{Q} (\mathbf{I}_N \otimes \mathbf{1}_T)$$

where the $T \times 1$ vector $\mathbf{1}_T$ is the vector of all ones (from now on, we write $\mathbf{R} = \text{sum}(N, T, \mathbf{Q})$).

Hence, the matrix \mathbf{R} is positive semidefinite. The subsequent series of results of matrix analysis proves that \mathbf{A} is positive semidefinite:

1. Let \mathbf{A} represent the $NTM \times NTM$ matrix obtained from $\hat{\mathbf{Y}}^{-1}$ defined in (E.5) as explained in the following. Seen as a block matrix with N^2 disjoint blocks, where each of them is of size $MT \times MT$, any (k, z) -th block of \mathbf{A} , for $k, z = 1, \dots, N$, consists of M^2 identical disjoint sub-blocks of dimension $T \times T$. For the (k, z) -th block of \mathbf{A} , the corresponding sub-block is $\hat{\mathbf{Y}}_{kz}^*$, i.e., the complex conjugate of $\hat{\mathbf{Y}}_{kz} = \mathbf{F}_k \hat{\mathbf{Y}}^{-1} \mathbf{F}_z^H$, where \mathbf{F}_k is defined in (E.5). Note that \mathbf{A} is positive semidefinite. For any nonzero complex vector $\mathbf{v} = [\mathbf{v}_1^T \ \dots \ \mathbf{v}_{MN}^T]^T$, where $\mathbf{v}_i \in \mathbb{C}^T$ for $i = 1, \dots, MN$, it holds: $\mathbf{v}^H \mathbf{A} \mathbf{v} = \mathbf{u}^H (\hat{\mathbf{Y}}^{-1})^* \mathbf{u} \geq 0$, where $\mathbf{u} = [\mathbf{u}_1^T \ \dots \ \mathbf{u}_N^T]^T$ with $\mathbf{u}_j = \sum_{i=1}^M \mathbf{v}_{(j-1)M+i}$ for $j = 1, \dots, N$. Since $\hat{\mathbf{Y}}^{-1} \succ \mathbf{0}$, then $(\hat{\mathbf{Y}}^{-1})^* \succ \mathbf{0}$ and \mathbf{A} is positive semidefinite;
2. Define $\mathbf{B} = \mathbf{1}_{N \times N} \otimes \tilde{\mathbf{P}}$ where $\mathbf{1}_{N \times N}$ is the $N \times N$ matrix of all ones. Then, \mathbf{B} is positive semidefinite since the Kronecker product of two positive (semi)definite matrices is positive (semi)definite (see pp. 245 in [94]). Also, the matrix \mathbf{C} defined as $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$, the symbol \odot represents the Schur (Hadamard) product, is positive semidefinite since both \mathbf{A} and \mathbf{B} are positive semidefinite, and the Schur product of two positive (semi)definite matrices is also positive (semi)definite (see pp. 458 in [53]);
3. Next, we define the matrix \mathbf{D} as $\mathbf{D} = \text{sum}(MN, T, \mathbf{C})$ which is positive semidefinite. Remark that $\text{sum}(1, T, \hat{\mathbf{Y}}_{kz}^* \odot \tilde{\mathbf{P}}_{ij}) = \text{tr}(\hat{\mathbf{Y}}_{kz}^H \tilde{\mathbf{P}}_{ij})$. Now, it is easy to see that the matrix \mathbf{A} can be presented as follows: $\mathbf{A} = \text{sum}(N, M, \mathbf{D} \odot \mathbf{D}^*)$ since $\sum_{i=1}^M \sum_{j=1}^M \text{tr}(\hat{\mathbf{Y}}_{zk} \tilde{\mathbf{P}}_{ij}) \text{tr}(\hat{\mathbf{Y}}_{kz} \tilde{\mathbf{P}}_{ji}) = \sum_{i=1}^M \sum_{j=1}^M \text{tr}(\hat{\mathbf{Y}}_{kz} \tilde{\mathbf{P}}_{ij}) \text{tr}(\hat{\mathbf{Y}}_{zk} \tilde{\mathbf{P}}_{ji})$, for $k, z = 1, \dots, N$ and Hermitian $\hat{\mathbf{Y}}^{-1}$ and $\tilde{\mathbf{P}}$.

Thus, as stated, the matrix \mathbf{A} is positive semidefinite and the maximum is achieved at an

extreme point of \mathcal{D}_N . So, the optimal value of Λ_r is given by

$$\left[\widehat{\Lambda}_r\right]_{k,k} = N\delta_{kk^*} \quad (\text{G.2})$$

for $k = 1, \dots, N$, where

$$\begin{aligned} (k^*, \widehat{\mathbf{P}}) = & \arg \max_{k=1, \dots, N} \sum_{i=1}^M \sum_{j=1}^M \text{tr}(\widehat{\mathbf{Y}}_{kk} \widetilde{\mathbf{P}}_{ij}) \text{tr}(\widehat{\mathbf{Y}}_{kk} \widetilde{\mathbf{P}}_{ji}). \\ & \mathbf{P} = \mathbf{P}^H \succeq \mathbf{0}, \text{tr}(\mathbf{P}) \leq TM \end{aligned} \quad (\text{G.3})$$

The optimization problem in (3.14) now becomes

$$\begin{aligned} \max_{\substack{U_r \in \mathcal{U}_N, \mathbf{K}_t \in \mathcal{P}_M, \\ \mathbf{P} = \mathbf{P}^H \succeq \mathbf{0}, \text{tr}(\mathbf{P}) \leq TM}} & N^2 \sum_{i=1}^M \sum_{j=1}^M \text{tr}(\widehat{\mathbf{Y}}_{k^*k^*} \widetilde{\mathbf{P}}_{ij}) \text{tr}(\widehat{\mathbf{Y}}_{k^*k^*} \widetilde{\mathbf{P}}_{ji}). \end{aligned} \quad (\text{G.4})$$

Step 2. In the second step, we shall determine the optimal values of \mathbf{K}_t , \mathbf{P} and U_r .

Approach: Similarly to the analysis of Appendix E in order to calculate $\widehat{\mathbf{K}}_t$, $\widehat{\mathbf{P}}$ and \widehat{U}_r we relax the problem in (G.4), find an upper bound on the relaxed problem and show that this bound is attainable for the original problem. Let p^* be the maximum of the problem in (G.4), and let

$$\overline{\mathbf{P}} = \left(\mathbf{I}_M \otimes \widehat{\mathbf{Y}}_{k^*k^*}^{\frac{1}{2}} \right) \widetilde{\mathbf{P}} \left(\mathbf{I}_M \otimes \widehat{\mathbf{Y}}_{k^*k^*}^{\frac{1}{2}} \right)$$

with $\text{tr}(\overline{\mathbf{P}}) \leq K = TM^2 \lambda_{\max}(\widehat{\mathbf{Y}}_{k^*k^*})$. Then, the relaxed problem is defined as

$$q^* = \max_{U_r \in \mathcal{U}_N, \overline{\mathbf{P}} = \overline{\mathbf{P}}^H \succeq \mathbf{0}, \text{tr}(\overline{\mathbf{P}}) \leq K} N^2 \sum_{i=1}^M \sum_{j=1}^M \text{tr}(\overline{\mathbf{P}}_{ij}) \text{tr}(\overline{\mathbf{P}}_{ji}). \quad (\text{G.5})$$

where

$$\begin{aligned} \overline{\mathbf{P}}_{ij} &= \mathbf{E}_i \overline{\mathbf{P}} \mathbf{E}_j^H = \widehat{\mathbf{Y}}_{k^*k^*}^{\frac{1}{2}} \widetilde{\mathbf{P}}_{ij} \widehat{\mathbf{Y}}_{k^*k^*}^{\frac{1}{2}}, \\ \widetilde{\mathbf{P}} &= \left((\mathbf{K}_t^T)^{\frac{1}{2}} \otimes \mathbf{I}_T \right) \mathbf{P} \left((\mathbf{K}_t^T)^{\frac{1}{2}} \otimes \mathbf{I}_T \right), \end{aligned}$$

and the matrix \mathbf{E}_i , for $i = 1, \dots, M$, is defined in (3.12). The matrix $\widehat{\mathbf{Y}}_{k^*k^*}^{\frac{1}{2}}$ is the square root of $\widehat{\mathbf{Y}}_{k^*k^*}$, i.e., it is the $T \times T$ matrix such that $\left(\widehat{\mathbf{Y}}_{k^*k^*}^{\frac{1}{2}} \right)^2 = \widehat{\mathbf{Y}}_{k^*k^*}$. Clearly, $p^* \leq q^*$.

Now, note that, as $\text{tr}(\overline{\mathbf{P}}) \leq K$ it holds

$$\sum_{i=1}^M \text{tr}^2(\overline{\mathbf{P}}_{ii}) + 2 \sum_{1 \leq k < j \leq M} \text{tr}(\overline{\mathbf{P}}_{kk}) \text{tr}(\overline{\mathbf{P}}_{jj}) \leq K^2.$$

Also, remark that

$$\sum_{i=1}^M \sum_{j=1}^M \text{tr}(\bar{\mathbf{P}}_{ij})\text{tr}(\bar{\mathbf{P}}_{ji}) = \sum_{i=1}^M \text{tr}^2(\bar{\mathbf{P}}_{ii}) + 2 \sum_{1 \leq k < j \leq M} \text{tr}(\bar{\mathbf{P}}_{kj})\text{tr}(\bar{\mathbf{P}}_{jk}).$$

Then, if we define

$$f_{kj} = \sum_{1 \leq k < j \leq M} \text{tr}(\bar{\mathbf{P}}_{kk})\text{tr}(\bar{\mathbf{P}}_{jj}) - \sum_{1 \leq k < j \leq M} \text{tr}(\bar{\mathbf{P}}_{kj})\text{tr}(\bar{\mathbf{P}}_{jk}),$$

it holds

$$q^* \leq \max_{\substack{\mathbf{U}_r \in \mathcal{U}_N \\ \bar{\mathbf{P}} = \bar{\mathbf{P}}^H \succeq \mathbf{0}, \text{tr}(\bar{\mathbf{P}}) \leq K}} N^2 K^2 - 2N^2 f_{kj} \leq N^2 K^2$$

due to the fact proved in the sequel that

$$\text{tr}(\bar{\mathbf{P}}_{kk})\text{tr}(\bar{\mathbf{P}}_{jj}) \geq \text{tr}(\bar{\mathbf{P}}_{kj})\text{tr}(\bar{\mathbf{P}}_{jk})$$

for every pair (k, j) where $k, j = 1, \dots, M$.

Proof: We start by noting that

$$\left([\mathbf{e}_k \ \mathbf{e}_j]^T \otimes \mathbf{I}_T \right) \bar{\mathbf{P}} \left([\mathbf{e}_k \ \mathbf{e}_j] \otimes \mathbf{I}_T \right) = \begin{bmatrix} \bar{\mathbf{P}}_{kk} & \bar{\mathbf{P}}_{kj} \\ \bar{\mathbf{P}}_{jk} & \bar{\mathbf{P}}_{jj} \end{bmatrix} \succeq \mathbf{0}$$

where \mathbf{e}_k represents the k -th column of \mathbf{I}_M . Now, let a_i, b_i, b_i^* and c_i , for $i = 1, \dots, T$, represent the diagonal entries of the matrices $\bar{\mathbf{P}}_{kk}, \bar{\mathbf{P}}_{kj}, \bar{\mathbf{P}}_{jk}$ and $\bar{\mathbf{P}}_{jj}$, respectively (i.e., $[\bar{\mathbf{P}}_{kk}]_{i,i} = a_i, [\bar{\mathbf{P}}_{kj}]_{i,i} = b_i, [\bar{\mathbf{P}}_{jk}]_{i,i} = b_i^*$ and $[\bar{\mathbf{P}}_{jj}]_{i,i} = c_i$). It is not difficult to see that

$$(\mathbf{I}_2 \otimes \mathbf{k}_i^T) \begin{bmatrix} \bar{\mathbf{P}}_{kk} & \bar{\mathbf{P}}_{kj} \\ \bar{\mathbf{P}}_{jk} & \bar{\mathbf{P}}_{jj} \end{bmatrix} (\mathbf{I}_2 \otimes \mathbf{k}_i) = \begin{bmatrix} a_i & b_i \\ b_i^* & c_i \end{bmatrix} \succeq \mathbf{0}$$

for $i = 1, \dots, T$, where \mathbf{k}_i represents the i -th column of \mathbf{I}_T . Let $\mathbf{S}_i = \begin{bmatrix} a_i & b_i \\ b_i^* & c_i \end{bmatrix}$. Hence,

$$\sum_{i=1}^T \mathbf{S}_i = \begin{bmatrix} \text{tr}(\bar{\mathbf{P}}_{kk}) & \text{tr}(\bar{\mathbf{P}}_{kj}) \\ \text{tr}(\bar{\mathbf{P}}_{jk}) & \text{tr}(\bar{\mathbf{P}}_{jj}) \end{bmatrix} \succeq \mathbf{0}$$

from which the desired inequality is readily obtained. Remark that we have used the fact that the sum of positive semidefinite matrices is a positive semidefinite matrix. Thus, we have $p^* \leq q^* \leq N^2 K^2$. We now prove that an upper bound on the relaxed problem in (G.5) is attainable for the original problem in (G.4). To see this, we define $\hat{\mathbf{P}}$ and $\hat{\mathbf{K}}_t$ on the following way,

$$\hat{\mathbf{P}} = TM\mathbf{K}_P \otimes \mathbf{U}_{k^*} \mathbf{K}_{P_1} \mathbf{U}_{k^*}^H, \quad \left[\hat{\mathbf{K}}_t \right]_{j,j} = M\delta_{j1}, \quad (\text{G.6})$$

where we assume that $\mathbf{U}_{k^*} \mathbf{D}_{k^*} \mathbf{U}_{k^*}^H$ represents the EVD of $\widehat{\mathbf{\Upsilon}}_{k^*k^*}$ (the matrix \mathbf{D}_{k^*} is the diagonal matrix of the corresponding eigenvalues sorted in the decreasing order), the $M \times M$ ($T \times T$) constant matrix $\mathbf{K}_{\mathcal{P}}$ ($\mathbf{K}_{\mathcal{P}_1}$) has all the entries equal to zero except the $(1, 1)$ entry which is one, and k^* is defined in (G.3). Note that $\widehat{\mathbf{K}}_t$ is a diagonal matrix. In that case,

$$p^* = q^* = N^2 T^2 M^4 \max_{\mathbf{U}_r \in \mathcal{U}_N} \lambda_{\max}^2(\widehat{\mathbf{\Upsilon}}_{k^*k^*}) \quad (\text{G.7})$$

where, from (G.3) and (G.6),

$$k^* = \arg \max_{k=1, \dots, N} \lambda_{\max}(\widehat{\mathbf{\Upsilon}}_{k^*k^*}).$$

By repeating the analysis of the on-off signaling in Appendix E (equations (E.13), (E.14) and (E.15)), we easily find

$$\widehat{\mathbf{P}} = TM\mathbf{K}_{\mathcal{P}} \otimes \hat{\mathbf{x}}\hat{\mathbf{x}}^H, \quad \widehat{\mathbf{K}}_r = N\hat{\mathbf{u}}\hat{\mathbf{u}}^H \quad (\text{G.8})$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{u}}$ are the solutions of the optimization problem in (3.6). Equations (G.6) and (G.8) conclude the proof.

Appendix H

Pairwise Error Probability for Fast Fading in the Low SNR Regime

In this appendix, we derive the expression for the low SNR regime pairwise error probability presented in (3.16). In Appendix A, it has been shown that if \mathbf{X}_i is transmitted, then the probability that the receiver decides in favor of \mathbf{X}_j is:

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} = P(e^H(\mathbf{P}_i^H \mathbf{\Upsilon}^{-1} \mathbf{P}_i - \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j) \mathbf{e} - 2 \Re\{e^H \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \Delta \mathbf{g}\} > \mathbf{g}^H \mathbf{L}_{ij} \mathbf{g}) \quad (\text{H.1})$$

where, for $k \in \{i, j\}$,

$$\begin{aligned} \mathbf{P}_k &= \mathbf{I}_{TN} - \widetilde{\mathbf{X}}_k \mathbf{Z}_k, \quad \widetilde{\mathbf{X}}_k = \mathbf{I}_N \otimes \mathbf{X}_k, \\ \mathbf{Z}_k &= (\boldsymbol{\chi}_k^H \boldsymbol{\chi}_k)^{-1} \boldsymbol{\chi}_k^H \mathbf{\Upsilon}^{-\frac{1}{2}}, \quad \boldsymbol{\chi}_k = \mathbf{\Upsilon}^{-\frac{1}{2}} \widetilde{\mathbf{X}}_k, \\ \Delta &= \mathbf{P}_j \widetilde{\mathbf{X}}_i, \quad \mathbf{e} = \text{vec}(\mathbf{E}), \quad \mathbf{L}_{ij} = \boldsymbol{\chi}_i^H \mathbf{\Pi}_j^\perp \boldsymbol{\chi}_i \end{aligned}$$

and $\mathbf{\Pi}_j^\perp = \mathbf{I}_{TN} - \boldsymbol{\chi}_j (\boldsymbol{\chi}_j^H \boldsymbol{\chi}_j)^{-1} \boldsymbol{\chi}_j^H$. The unknown realization of the channel is denoted by $\mathbf{g} = \text{vec}(\mathcal{H})$, whereas $\Re\{a\}$ represents the real part of the complex number a . Unfortunately, the expression (H.1) cannot be simplified analytically. Hence, we shall analyze (H.1) in the low SNR regime where the linear term of \mathbf{e} is negligible (see Appendix A for the analysis of (H.1) in the high SNR regime). Therefore, at sufficiently low SNR,

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \approx P(e^H(\mathbf{P}_i^H \mathbf{\Upsilon}^{-1} \mathbf{P}_i - \mathbf{P}_j^H \mathbf{\Upsilon}^{-1} \mathbf{P}_j) \mathbf{e} > \mathbf{g}^H \mathbf{L}_{ij} \mathbf{g}). \quad (\text{H.2})$$

It is easy to see that, for $k \in \{i, j\}$,

$$e^H \mathbf{P}_k^H \mathbf{\Upsilon}^{-1} \mathbf{P}_k \mathbf{e} \stackrel{d}{=} \mathbf{z}^H \mathbf{\Pi}_k^\perp \mathbf{z}$$

where $\mathbf{z} = \Upsilon^{-\frac{1}{2}} \mathbf{e} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{TN})$ and $\stackrel{d}{=}$ means equal in distribution. Then, from (H.2) it holds

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \approx P(\mathbf{z}^H (\mathbf{U}_j \mathbf{U}_j^H - \mathbf{U}_i \mathbf{U}_i^H) \mathbf{z} > \mathbf{g}^H \mathbf{L}_{ij} \mathbf{g}). \quad (\text{H.3})$$

where, for $k \in \{i, j\}$, $\mathbf{U}_k = \boldsymbol{\chi}_k (\boldsymbol{\chi}_k^H \boldsymbol{\chi}_k)^{-\frac{1}{2}}$. That is, \mathbf{U}_k contains an orthonormal basis for the subspace spanned by the columns of $\boldsymbol{\chi}_k$. Notice that $\mathbf{U}_k^H \mathbf{U}_k = \mathbf{I}_{MN}$. To proceed with the analysis we use the known fact from [52] pp. 199: if $\mathbf{U}_i, \mathbf{U}_j$ are $TN \times MN$ matrices with orthonormal columns ($T \geq M$), then there exist $MN \times MN$ unitary matrices \mathbf{W}_1 and \mathbf{W}_2 , and a $TN \times TN$ unitary matrix \mathbf{Q} with the following properties:

If $2MN \leq TN$ ($2M \leq T$), then

$$\mathbf{Q} \mathbf{U}_i \mathbf{W}_1 = [\mathbf{I}_{MN} \quad \mathbf{0}_{MN} \quad \mathbf{0}_{MN \times (TN-2MN)}]^T \quad (\text{H.4})$$

and

$$\mathbf{Q} \mathbf{U}_j \mathbf{W}_2 = [\mathbf{C}_{ij} \quad \mathbf{S}_{ij} \quad \mathbf{0}_{MN \times (TN-2MN)}]^T \quad (\text{H.5})$$

where \mathbf{C}_{ij} is a diagonal $MN \times MN$ matrix with diagonal entries $\cos \alpha_i$, $i = 1, \dots, MN$, $0 \leq \alpha_1 \leq \dots \leq \alpha_{MN} \leq \frac{\pi}{2}$, and $\mathbf{S}_{ij}^2 + \mathbf{C}_{ij}^2 = \mathbf{I}_{MN}$. Now, using (H.4) and (H.5) we have

$$\mathbf{W}_2^H \mathbf{U}_j^H \mathbf{Q}^H \mathbf{Q} \mathbf{U}_i \mathbf{W}_1 = \mathbf{C}_{ij} \Rightarrow \mathbf{U}_j^H \mathbf{U}_i = \mathbf{W}_2 \mathbf{C}_{ij} \mathbf{W}_1^H,$$

so α_i , for $i = 1, \dots, MN$, are the *principal angles* between the subspaces spanned by \mathbf{U}_i and \mathbf{U}_j . It can be readily shown that

$$\mathbf{z}^H (\mathbf{U}_j \mathbf{U}_j^H - \mathbf{U}_i \mathbf{U}_i^H) \mathbf{z} \stackrel{d}{=} \mathbf{c}^H \mathbf{D} \mathbf{c},$$

where

$$\mathbf{D} = \begin{bmatrix} -\mathbf{S}_{ij}^2 & \mathbf{C}_{ij} \mathbf{S}_{ij} \\ \mathbf{C}_{ij} \mathbf{S}_{ij} & \mathbf{S}_{ij}^2 \end{bmatrix}$$

and $\mathbf{c} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{2MN})$. Let $\mathbf{c} = [a_1 \quad b_1 \quad \dots \quad a_{MN} \quad b_{MN}]^T$ where $a_m, b_m \stackrel{iid}{\sim} \mathcal{CN}(0, 1)$ for $m = 1, \dots, MN$. Hence, using the fact that $\pm \sin \alpha_i$, for $i = 1, \dots, MN$, are the eigenvalues of the matrix \mathbf{D} , from (H.3) we easily get

$$P_{\mathbf{X}_i \rightarrow \mathbf{X}_j} \approx P\left(\sum_{m=1}^{MN} \sin \alpha_m (|a_m|^2 - |b_m|^2) > \mathbf{g}^H \mathbf{L}_{ij} \mathbf{g}\right)$$

as required.

Bibliography

- [1] J. H. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, pp. 871-878, June. 1987.
- [2] A. Paulraj and T. Kailath "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception," *US Patent*, 5 345 599, 1994.
- [3] I. Telatar, "Capacity of multi-antenna Gaussian channels," *Technical Memorandum, AT&T Bell Laboratories*, 1995.
- [4] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41-59, 1996.
- [5] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744-765, Mar. 1998.
- [6] S. A. Alamouti, "A simple transmit diversity technique for wireless communication," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.
- [7] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 145-1467, July 1999.
- [8] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1804-1824, July 2002.

- [9] J. -C. Belfiore, G. Rekaya, and E. Viterbo, "The Golden code: A 2×2 full rate space-time code with non-vanishing determinant," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1432-1436, Apr. 2005.
- [10] G. D. Golden, C. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14-16, Jan. 1999.
- [11] G.J. Foschini, D. Chizhik, M. J. Gans, C. Papadias, R. A. Valenzuela, " Analysis and performance of some basic space-time architectures," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 303-320, Apr. 2003.
- [12] J. -C. Belfiore, G. Rekaya, and E. Viterbo, "A new approach to layered space-time coding and signal processing," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2321 - 2334, Sep. 2001.
- [13] M. O. Damen, A. Chkeif, and J. -C. Belfiore, "Lattice codes decoder for space-time codes," *IEEE Communication Letters*, vol. 4, no. 5, pp. 161-163, May 2000.
- [14] S. Verdú, *Multiuser Detection.*, Cambridge University Press, 1998.
- [15] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.*, Cambridge University Press, 2005.
- [16] A. Hottinen, O. Tirkkonen, and R. Wichman, *Multi-antenna transceiver techniques for 3G and beyond.*, John Wiley & Sons, 2003.
- [17] D. Gesbert, M. Shafi, Da-Shan Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 281-302, April 2003.
- [18] H. Bolcskei, D. Gesbert, C. B. Papadias, and A. -J. van der Veen (eds.), *Space-Time Wireless Systems: From Array Processing to MIMO Communications.*, Cambridge University Press, 2006.

- [19] E. G. Larsson and P. Stoica, *Space-Time Block Codes for Wireless Communications.*, Cambridge University Press, 2003.
- [20] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications.*, Cambridge University Press, 2003.
- [21] S. Barbarosa, *Multiantenna Wireless Communication Systems.*, Artech House, 2005.
- [22] B. Vucetic and J. Yuan, *Space-Time Coding.*, Wiley, 2003.
- [23] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139-157, Jan. 1999.
- [24] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat-fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543-564, Mar. 2000.
- [25] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2383-2399, Sept. 2001.
- [26] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359-383, Feb. 2002.
- [27] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 1962-1973, Sep. 2000.
- [28] J. A. Tropp, "Topics in sparse approximation", *Ph.D. dissertation: Univ. Texas at Austin*, 2004.

- [29] D. Agrawal, T. J. Richardson, and R. L. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2618-2626, Sept. 2001.
- [30] N. J. A. Sloane, "Packing planes in four dimensions and other mysteries," in *Proc. of Algebraic Combinatorics and Related Topics*, Yamagata, Japan, 1997.
- [31] M. L. McCloud, M. Brehler, and M. K. Varanasi, "Signal design and convolutional coding for noncoherent space-time communication on the block-Rayleigh-fading channel," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1186-1194, May 2002.
- [32] I. Kammoun and J. -C. Belfiore, "A new family of Grassmann space-time codes for noncoherent MIMO systems," *IEEE Communications Letters*, vol. 7, no. 11, pp. 528-530, Nov. 2003.
- [33] A. M. Cipriano, I. Kammoun, and J. -C. Belfiore, "Simplified decoding for some noncoherent codes over the Grassmannian," in *Proc. of IEEE International Conference on Communications (ICC)*, Seoul, Korea, 2005.
- [34] A. M. Cipriano and I. Kammoun, "Linear approximation of the exponential map with application to simplified detection of noncoherent Systems," in *Proc. of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Cannes, France, 2006.
- [35] R. H. Gohary and T. N. Davidson, "Noncoherent MIMO communication: Grassmannian constellations and efficient detection," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Chicago, IL USA, 2004.
- [36] R. H. Gohary and T. N. Davidson, "On efficient non-coherent detection of Grassmannian constellations," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, 2005.

- [37] R. H. Gohary and T. N. Davidson, "Non-coherent MIMO communication: Grassmannian constellations and efficient detection," submitted to *IEEE Transactions on Information Theory*.
- [38] A. Lozano, A. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2527-2544, Oct. 2003.
- [39] M. J. Borran, A. Sabharwal, and B. Aazhang, "On design criteria and construction of non-coherent space-time constellations," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2332-2351, Oct. 2003.
- [40] S. A. Aldosari and J. M. F. Moura, "Detection in Sensor Networks: The Saddlepoint Approximation," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 327-340, Jan. 2007.
- [41] I. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh fading channels," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1290-1301, May 2001.
- [42] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319-1343, June 2002.
- [43] V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1567-1580, Aug. 2004.
- [44] B. Hajek and V. Subramaniam, "Capacity and reliability function for small peak signal constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 4, pp. 828-839, Apr. 2002.
- [45] C. Rao and B. Hassibi, "Analysis of multiple-antenna wireless links at low SNR," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 2123-2130, Sep. 2004.

- [46] X. Wu and R. Srikant, "MIMO channels in the low SNR regime: communication rate, error exponent and signal peakiness," in *Proc. of IEEE Information Theory Workshop (ITW)*, San Antonio, TX USA, 2004.
- [47] S. G. Srinivasan and M. K. Varanasi, "Constellation Design with Unequal Priors and New Distance Criteria for the Low SNR Noncoherent Rayleigh Fading Channel," in *Proc. of Conference on Information Sciences and Systems*, The Johns Hopkins University, Baltimore, MD USA, 2005.
- [48] S. G. Srinivasan and M. K. Varanasi, "Code design for the low SNR noncoherent MIMO block Rayleigh fading channel," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, 2005.
- [49] H. L. Van Trees, *Detection, Estimation, and Modulation Theory.*, Part I New York: Wiley, 1968.
- [50] S. M. Kay, *Fundamentals of Statistical Signal Processing.*, Detection Theory, Prentice Hall, 1998.
- [51] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis.*, New York: Addison-Wesley Publishing Co., 1990.
- [52] R. Bhatia, *Matrix Analysis.*, Springer-Verlag, 1997.
- [53] R. A. Horn and C. R. Johnson, *Matrix Analysis.*, Cambridge University Press, 1985.
- [54] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experimental Mathematics*, vol. 5, pp. 139-159, Sept. 1996.
- [55] S. Berceanu, "On the geometry of complex Grassmann manifold, its noncompact dual and coherent states," *Bulletin of the Belgian Mathematical Society*, vol. 4, no. 2, pp. 205-243, 1997.
- [56] J. G. Proakis, *Digital Communications.*, McGraww Hill, 2001.
- [57] N. Shor, "Dual quadratic estimates in polynomial and Boolean programming," *Annals of Operation Research*, vol. 25, pp. 163-168, 1990.

- [58] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, 1999. (Updated for Version 1.05, <http://sedumi.mcmaster.ca>)
- [59] I. Polik, "Addendum to the Sedumi user guide version 1.1," <http://sedumi.mcmaster.ca>
- [60] D. Peaucelle, "Users's guide for SeDuMi interface 1.04," <http://www.laas.fr/peaucell/SeDuMiInt.html>
- [61] Y. Labit, D. Peaucelle and D. Henrion, "SeDuMi interface 1.02 : a tool for solving LMI problems with SeDuMi," in *Proc. of IEEE International Symposium on Computer-Aided Control System Design (CACSD)*, Glasgow, Scotland, 2002.
- [62] M. X. Goemans, "Semidefinite programming in combinatorial optimization," *Mathematical Programming*, vol. 79, pp. 143-161, 1997.
- [63] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303-353, 1998.
- [64] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 635-650, Mar. 2002.
- [65] L. R. Welch, "Lower bounds on the maximum cross-correlation of signals," *IEEE Transactions on Information Theory*, vol. IT-20, no. 3, pp. 397-399, May 1974.
- [66] S. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 990-997, May 2005.
- [67] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 637-650, Mar. 2002.

- [68] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, and K. Huber, "Multiple input multiple output measurements and modeling in Manhattan," in *Proc. of IEEE Vehicular Technology Conference (VTC)*, Vancouver, BC Canada, 2002.
- [69] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211-1226, Aug. 2002.
- [70] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 90-100, Jan. 2006.
- [71] X. Wu, "Wireless communications in the energy-limited regime", *Ph.D. dissertation: Univ. Illinois at Urbana-Champaign*, 2004.
- [72] G. Barriac and U. Madhow, "Space-time communication for OFDM with implicit channel feedback," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3111 - 3129, Dec. 2004.
- [73] M. Boko, J. Xavier, and V. Barroso, "Codebook design for non-coherent communication in multiple-antenna systems," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France, 2006.
- [74] M. Boko, J. Xavier, and V. Barroso, "Non-coherent communication in multiple-antenna systems: Receiver design and codebook construction," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5703 - 5715, Dec. 2007.
- [75] M. Boko, J. Xavier, and V. Barroso, "Codebook design for the non-coherent GLRT receiver and low SNR MIMO block fading channel," in *Proc. of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Cannes, France, 2006.
- [76] M. Boko, J. Xavier, and V. Barroso, "Capacity and error probability analysis of non-coherent MIMO systems in the low SNR regime," in *Proc. of IEEE International*

- Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, HI USA, 2007.
- [77] M. Beko, J. Xavier, and V. Barroso, "Further results on the capacity and error probability analysis of non-coherent MIMO systems in the low SNR regime," accepted for publication in *IEEE Transactions on Signal Processing*.
- [78] T. Strohmer and R. W. Heath, "Grassmannian frames with applications to coding and communication," *Applied and Computational Harmonic Analysis*, vol. 14, no. 3, pp. 257-275, May 2003.
- [79] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231-2242, Oct. 2004.
- [80] R. B. Holmes and V. I. Paulsen, "Optimal frames for erasures," *Linear Algebra and its Applications*, vol. 377, pp. 31-51, Jan. 2004.
- [81] R. W. Heath, T. Strohmer, and A. J. Paulraj, "Grassmannian signatures for CDMA systems," in *Proc. of IEEE Global Telecommunications Conference (Globecom)*, San Francisco, CA USA, 2003.
- [82] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 1926 - 1934, Dec. 2002.
- [83] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. of 37th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL USA, 1999.
- [84] S. Boyd and L. Vandenberghe, *Convex Optimization.*, Cambridge University Press, 2004.
- [85] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49-95, 1996.

- [86] L. Vandenberghe and S. Boyd, "SP: Software for semidefinite programming. User's guide, beta version," K.U. Leuven and Stanford University, October 1994.
- [87] K. C. Toh, M. J. Todd, and R. H. Tutuncu, "SDPT3 - A Matlab software package for semidefinite programming, version 2.1," *Optimisation Methods and Software*, vol. 11, pp. 545-581, 1999.
- [88] W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry.*, New York: Academic Press, 2003.
- [89] M. P. Carmo, *Geometria Riemanniana.* 2nd ed., Instituto de Matemática Pura e Aplicada, 1988.
- [90] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics.*, Revised Edition, John Wiley & Sons, 1999.
- [91] D. P. Bertsekas, *Nonlinear Programming.*, Athena Scientific, 1999.
- [92] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming.*, Addison-Wesley Publishing Company, 1972.
- [93] S. M. Kay, *Fundamentals of Statistical Signal Processing.*, Estimation Theory, Prentice Hall, 1993.
- [94] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis.*, Cambridge University Press, 1991.
- [95] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples.*, Springer-Verlag, 2000.
- [96] J. Munkres, *Topology.*, Prentice Hall, 2000.
- [97] J. Salo, M. Herdin, H. M. El-Sallabi, and P. Vainikainen, "Statistical analysis of the multiple scattering radio channel," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11, pp. 3114 - 3124, Nov. 2006.

-
- [98] J. Salo, M. Herdin, H. M. El-Sallabi, and P. Vainikainen, "Impact of double-Rayleigh fading on system performance," in *Proc. of 1st International Symposium on Wireless Pervasive Computing*, Phuket, Thailand, 2006.
- [99] T. M. Cover and J. A. Thomas, *Elements of information theory.*, Wiley series in telecommunications. John Wiley & Sons, 1991.
- [100] CommWeb. Wireless industry statistics, 2001.
- [101] T. S. Rappaport, A. Annamalai, R. M. Buehrer, and W. H. Tranter, "Wireless communications: Past events and a future perspective," *IEEE Communications Magazine*, vol. 40, no. 5, pp. 148-161, May 2002.