PhD–FSTM–2021–011
The Faculty of Sciences, Technology and Medicine

DISSERTATION

Defence held on 04/02/2021 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN INFORMATIQUE

by

Alireza Haqiqatnejad

Born on 28 July 1990 in Esfahan (Iran)

ENHANCED SIGNAL SPACE DESIGN FOR MULTIUSER MIMO INTERFERENCE CHANNELS

Dissertation defense committee

Dr Björn Ottersten, dissertation supervisor
Professor, Université du Luxembourg

Dr A. Lee Swindlehurst
Professor, University of California, Irvine

Dr Gabriele Lenzini, Chairman
Associate professor, Université du Luxembourg

Dr Pei Xiao
Professor, University of Surrey

Dr Farbod Kayhan, Vice Chairman
Lecturer, University of Surrey

“If you read all the time what other people have done, you will think the way they thought. If you want to think new thoughts that are different, then do what a lot of creative people do - get the problem reasonably clear and then refuse to look at any answers until you have thought the problem through carefully how you would do it, how you could slightly change the problem to be the correct one.”

Richard Hamming
Mathematician

To my family, my friends, and my love

Abstract

Multiuser precoding techniques are critical to handle the co-channel interference, also known as multiuser interference (MUI), in the downlink of multiuser multi-antenna wireless systems. The convention in designing multiuser precoding schemes has been to treat the MUI as an undesired received signal component. Consequently, the design attempts to suppress the MUI by exploiting the channel state information (CSI), regardless of the instantaneous users' data symbols. In contrast, it has been shown that the MUI may not always be undesired or destructive as it is possible to exploit the constructive part of the interference or even converting the interfering components into constructive interference (CI) by instantaneously exploiting the users' intended data symbols. As a result, the MUI can be transformed into a useful source of power that constructively contributes to the users' received signals. This observation has turned the viewpoint on multiuser precoding from conventional approaches towards more sophisticated designs that further exploit the data information (DI) in addition to the CSI, referred to as symbol-level precoding (SLP). The SLP schemes can improve the multiuser system's overall performance in terms of various metrics, such as power efficiency, symbol error rate, and received signal power. However, such improvement comes with several practical challenges, for example, the need for setting the modulation scheme in advance, increased computational complexity at the transmitter, and sensitivity to CSI and other system uncertainties. The main goal of this thesis is to address these challenges in the design of an SLP scheme.

The existing design formulations for the CI-based SLP problem consider a specific signal constellation; therefore, the design needs to set the modulation scheme in advance. In this thesis, we first elaborate on optimal and relaxed approaches to exploit the CI in a novel systematic way. This study enables us to develop a generic framework for the SLP design problem, which can be used for modulation schemes with constellations of any given shape and order. Depending on the design criterion, the proposed framework can offer significant gains in the power consumption at the transmitter side or the received signal power and the symbol error rate at the receiver side without increasing the complexity, compared to the state-of-the-art schemes. Next, to address the high computational complexity issue, we simplify the design process and propose approximate

yet computationally-efficient solutions performing relatively close to the optimal design. We further propose an optimized accelerated FPGA design that allows the real-time implementation of our SLP technique in high-throughput communications systems. Remarkably, the accelerated design enjoys the same per-symbol complexity order as that of the zero-forcing (ZF) precoding scheme. Next, we address the problem of robust SLP design under system uncertainties. In particular, we focus on two sources of uncertainty, namely, the channel and the design process. The related problems are tackled by adopting worst-case and stochastic design approaches and appropriately redefining the precoding optimization problem. The resulting robust schemes can effectively deal with system uncertainties while preserving reliability and power efficiency in the multiuser communications system, at the cost of a slightly increased complexity. Finally, we broaden our scope to new technologies such as millimeter wave (mmWave) communications and massive multiple-input multiple-output (MIMO) systems and revisit the SLP problem for low-cost energy-efficient transmitter architectures. The precoding design problem is more challenging particularly in such scenarios as the related hardware restrictions impose additional (often intractable) constraints on the problem. The restrictions are typically due to the use of finite-resolution analog-to-digital converters (DAC) or analog components such as switches and/or phase shifters. Two well-known design strategies are considered in this thesis, namely, quantized (finite-alphabet) precoding and hybrid analog-digital precoding. We tackle the related problems through adopting efficient design mechanisms and optimization algorithms, which are novel for the SLP schemes. The proposed techniques are shown to improve the system's energy efficiency compared to the state-of-the-art.

Acknowledgements

The days of my Ph.D. at the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, have been a wonderful experience armed me with new skills both in my professional and personal life. I would like to express my sincere gratitude to many people, without whom I would not have completed this Ph.D. Thesis.

First of all, I would like to thank my supervisors Prof. Björn Ottersten and Dr. Farbod Kayhan for accepting me as a Ph.D. student, and further, for their consistent support, encouragement and guidance throughout the past four years. I also had the great pleasure of working with Prof. Shahram ShahbazPanahi (firstly as a visiting Professor at SnT and later as a professor with the University of Ontario Institute of Technology, Canada), whose ingenious advice and invaluable contributions facilitated the completion of this thesis. I am also grateful to my CET committee member Dr. Bhavani Shankar Mysore, for his valuable suggestions. The constructive discussions during the ESSTIMS meetings and insightful suggestions by Dr. Nader Alagha, from the European Space Agency (ESA), and Dr. Joel Grotz, from the Société Européenne des Satellites (SES), are also gratefully acknowledged. Besides, I very much appreciate Prof. Lee Swindlehurst, from the University of California, Irvine, and Prof. Pei Xiao, from the University of Surrey, for kindly accepting to participate in my dissertation committee, and Prof. Gabriele Lenzini for kindly accepting to serve as the Chairman of my dissertation defense.

I would like to extend my special gratitude to my former colleagues at SnT, Dr. Maha Alodeh, Dr. Ahmad Gharanjik, Dr. Ashkan Kalantari, Dr. Nicolò Mazzali, and Dr. Stefano Andrenacci for thoughtfully sharing their knowledge and facilitating the beginning steps of this thesis. I would also like to thank my collaborators Dr. Jevgenij Krivochiza, Dr. Sumit Gautam, Dr. Juan Merlano Duncan, and Prof. Symeon Chatzinotas. I also wish to thank all my colleagues at SnT for providing such an amazing work environment. Many thanks to Dr. Alberto Mengali, Dr. Vahid Joroughi, and Dr. Saeid Sedighi for helpful time-to-time discussions.

I am deeply grateful and indebted to my family, who set me off on the road to this Ph.D. a long time ago and never wavered in their support and understanding. And special thanks to Saharnaz, whose support and patience never let me down.

Finally, I would like to gratefully acknowledge the generous funding of my Ph.D. by the Fonds National de la Recherche (FNR, Luxembourg National Research Fund) via the University of Luxembourg.

List of Abbreviations

ACM	adaptive coding and modulation
ADC	analog-to-digital converter
ADMM	Alternating direction method of multipliers
AM-AM	amplitude-to-amplitude
AM-PM	amplitude-to-phase
APGD	accelerated projected gradient descent
AWGN	additive white Gaussian noise
BCD	block coordinate descent
BER	bit error rate
BLP	block-level precoding
BS	base station
CE	constant-envelope
CSCG	circularly symmetric complex Gaussian
CDL	clock-driven logic
CF-SLP	closed-form symbol-level precoding
CI	constructive interference
CIR	constructive interference region
CRN	cognitive radio network
CSI	channel state information

DAC	digital-to-analog converter
DAS	distributed antenna system
DI	data information
DPCIR	distance-preserving constructive interference region
EPM	exact penalty method
FLOP	floating-point operation
FPGA	field-programmable gate array
FSK	frequency-shift keying
HDL	hardware description language
HLS	high-level synthesis
ICF-SLP	improved closed-form symbol-level precoding
IP	intellectual property
KKT	Karush-Kuhn-Tucker
LCQP	linearly-constrained quadratic programming
LLR	log-likelihood ratio
LMI	linear matrix inequality
LP	linear programming
MDPCIR	minimum distance-preserving constructive interference region
MF	matched filter
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
ML	maximum likelihood
MMSE	minimum mean square error
mmWave	Millimeter wave
MODCOD	modulation and coding
MRT	maximum ratio transmission
MUI	multiuser interference

MU-MIMO	multiuser multiple-input multiple-output
NNLS	non-negative least squares
NR	noise-robust
NSPC	non-strict phase constraints
PA	power amplifier
PAM	pulse-amplitude modulation
PAPR	peak-to-average power ratio
PHY	physical-layer
PMF	probability mass function
PSK	phase-shift keying
PSS	phase shifter selection
QAM	quadrature-amplitude modulation
QoS	quality-of-service
QP	quadratic programming
QPSK	quadrature phase-shift keying
RF	radio frequency
RTL	register-transfer level
RZF	regularized zero-forcing
SDP	semi-definite programming
SEP	symbol error probability
SER	symbol error rate
SINR	signal-to-interference-plus-noise ratio
SLP	symbol-level precoding
SNR	signal-to-noise ratio
SOC	second-order cone
SOCP	second-order cone programming
SPC	strict phase constraints

SS	symbol-scaling
SWIPT	simultaneous wireless information and power transfer
TDD	time-division duplex
UBCIR	union bound constructive interference region
UE	user equipment
WF	Wiener filter
ZF	zero-forcing

Notations

j	Imaginary unit $\mathrm{j} = \sqrt{-1}$
π	Pi number equals ≈ 3.1416
$\ln(x)$	Natural logarithm of x
e^x	Exponential function of x
$\mathbb{E}\{\cdot\}$	Statistical expectation
$\Pr\{\cdot\}$	Probability function
$\mathrm{Re}(\cdot)$	Real part of a complex input
$\mathrm{Im}(\cdot)$	Imaginary part of a complex input
$a, \mathbf{a}, \mathbf{A}, \mathcal{A}$	A scalar, a vector, a matrix, a set
$\lvert\mathcal{A}\rvert$	Cardinality of set $\mathcal{A}$
$\mathcal{A}\backslash\mathcal{B}$	The set of all the elements in $\mathcal{A}$ excluding those in common with $\mathcal{B}$
$\lvert a\rvert$	Modulus of scalar a
a^*	Conjugate of complex scalar a
$\mathbf{a}^{\mathrm{T}}, \mathbf{A}^{\mathrm{T}}$	Transpose of vector $\mathbf{a}$, transpose of matrix $\mathbf{A}$
$\mathbf{a}^{\mathrm{H}}, \mathbf{A}^{\mathrm{H}}$	Conjugate transpose of vector $\mathbf{a}$, conjugate transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of square matrix $\mathbf{A}$
$\mathbf{A}^{\dagger}$	Moore-Penrose Inverse of matrix $\mathbf{A}$
$\mathbf{A} \succeq 0$	Matrix $\mathbf{A}$ is positive semidefinite
$\mathbf{A} \succeq \mathbf{B}$	Matrix $\mathbf{A} - \mathbf{B}$ is positive semidefinite

$\mathbf{A} \circ \mathbf{B}$	Hadamard (element-wise) product of matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathrm{Tr}(\mathbf{A})$	Trace of matrix $\mathbf{A}$
$\lvert\mathbf{A}\rvert$	Determinant of square matrix $\mathbf{A}$
$\mathrm{rank}(\mathbf{A})$	Rank of matrix $\mathbf{A}$
$\mathcal{R}(\mathbf{A})$	Range space of matrix $\mathbf{A}$
$\mathrm{vec}(\mathbf{A})$	The column vector obtained by stacking the columns of matrix $\mathbf{A}$
$\mathbf{a} \succeq \mathbf{b}$	Element-wise inequality between vectors $\mathbf{a}$ and $\mathbf{b}$
$\mathrm{diag}(\mathbf{a})$	Diagonal matrix with the elements of vector $\mathbf{a}$ on the main diagonal
$\lVert\mathbf{a}\rVert$	ℓ_2-norm (Euclidean norm) of vector $\mathbf{a}$
$\lVert\mathbf{a}\rVert_1$	ℓ_1-norm of vector $\mathbf{a}$
$\lVert\mathbf{a}\rVert_\infty$	ℓ_∞-norm of vector $\mathbf{a}$
$\lVert\mathbf{A}\rVert_{\mathrm{F}}$	Frobenius norm of matrix $\mathbf{A}$
$\lVert\mathbf{A}\rVert_2$	Spectral norm of matrix $\mathbf{A}$
$\mathbb{R}$	The set of real numbers
$\mathbb{R}_+$	The set of non-negative real numbers
$\mathbb{C}$	The set of complex numbers
$\mathbb{R}^{n \times m}$	The set of $n \times m$ matrices with real-valued entries
$\mathbb{C}^{n \times m}$	The set of $n \times m$ matrices with complex-valued entries
$\mathbf{I}_n$	Identity matrix of dimension n
$\mathbf{1}_{n \times m}$	All-ones matrix of dimension $n \times m$
$\mathbf{0}_{n \times m}$	All-zeros matrix of dimension $n \times m$
$\mathbf{I}, \mathbf{1}, \mathbf{0}$	Identity, all-ones or all-zeros matrix of appropriate dimension
max	Maximize
min	Minimize
argmax	Maximizing argument
argmin	Minimizing argument
s.t.	Subject to

List of Tables

1.1 Summary of the CI-based SLP techniques in the literature. 48

3.1 Normal vectors corresponding to QPSK symbols. 67
3.2 Different design formulations for the DPCIR-based SLP power minimization problem. 79
3.3 Number of constraints and solution time for different SLP schemes. 87
3.4 Complexities of the proposed methods for the SLP max-min SINR design. 92

4.1 Actual complexity in FLOPs for different SLP designs. 103
4.2 Dominating complexity order for different SLP designs. 103
4.3 Execution time of different precoding schemes. 108

5.1 Actual arithmetic complexity of Algorithm 3. 115
5.2 Interface specifications of the IP core. 116
5.3 Structure of the RTL data ports. 117
5.4 Performance estimates of the non-optimized HDL design. 119
5.5 Performance estimates of the optimized HDL design. 120
5.6 Resource utilization of the non-optimized HDL design on the Xilinx Kintex-7 xc7k410tffv900-2 FPGA. 120
5.7 Resource utilization of the optimized HDL design on the Xilinx Kintex-7 xc7k410tffv900-2 FPGA. 121

6.1 Proposed worst-case/stochastic robust CI constraints. 144
6.2 Complexity comparison of the non-robust and the proposed robust SLP designs. 146

8.1 Average percentage of inactive binary constraints in the RQSLP solution for different values of N_t with $N_\mathrm{u} = 4$ and $b = 1$. 189
8.2 Average number of iterations with computations of dimension $2N_\mathrm{t}$ till convergence of the QSLP algorithm. 196

9.1 Complexity comparison of different hybrid SLP schemes. 212

List of Figures

2.1 Downlink MU-MIMO system with multiuser precoding. 50
2.2 Typical CIRs for the QPSK constellation. 56

3.1 A simplified block diagram for the downlink MU-MIMO channel with SLP. 63
3.2 An illustration of CIRs for an AWGN-optimized 8-ary constellation. . . . 68
3.3 Geometry of DPCIRs for a boundary constellation point. 72
3.4 Feasibility probability of the SLP design as a function of the power budget for different target SINRs with $(N_\mathrm{t}, N_\mathrm{u}) = (4, 4)$. 86
3.5 Transmit power versus target SINR with $(N_\mathrm{t}, N_\mathrm{u}) = (8, 8)$ for (a) 8-PSK constellation; (b) the optimized 8-ary constellation. 87
3.6 Performance comparison for a system with $(N_\mathrm{t}, N_\mathrm{u}) = (8, 8)$: (a) Average per-user achievable throughput as a function of target rate; (b) Average symbol error probability versus target SINR. 88
3.7 Scatter plot of the noise-free received signals taken form 8-PSK constellation in a system with $(N_\mathrm{t}, N_\mathrm{u}) = (8, 8)$ and $p = 15$ dBW. 89
3.8 The worst-user received SINR among $N_\mathrm{u} = 4$ users as a function of the power budget for (a) 8-PSK constellation; (b) the optimized 8-ary constellation; (c) 16-QAM constellation. 90
3.9 Worst-user received SINR as a function of system dimension and power budget. 91
3.10 Number of iterations until convergence of the BCD algorithm as a function of (a) system dimension; (b) power budget for three constellations with different orders. 91

4.1 Accuracy of the CF-SLP solution for different system dimensions. 104
4.2 Comparison of the total transmit power resulted from different precoding schemes versus target SINR with (a) $N_\mathrm{t} = N_\mathrm{u} = 4$; (b) $N_\mathrm{t} = N_\mathrm{u} = 8$; (c) $N_\mathrm{t} = N_\mathrm{u} = 16$. 105
4.3 Transmit power versus target SINR with $N_\mathrm{t} = N_\mathrm{u} = 8$. 106
4.4 Transmit power and time complexity versus number of users with $\beta = 6/5$.107

5.1 Experimental probability mass function of L. 113
5.2 IP block design of the low-complexity SLP technique. 117
5.3 Data flow of the IP core sampled in the LabVIEW environment. 118
5.4 Intended symbols and noise-free received signals obtained by simulating the HDL design of Algorithm 3. 122
5.5 Block diagram of the simulated communication system. 123
5.6 Scatter plot of the users' received signals at SINR = 0 dB. 124
5.7 Performance comparison of different precoding designs as a function of target SINR for a system with QPSK modulation and $N_\mathrm{t} = N_\mathrm{u} = 4$: (a) Average per-user symbol error rate; (b) Average total transmit power. . . 125
5.8 Power efficiency as a function of target SINR with QPSK modulation and $N_\mathrm{t} = N_\mathrm{u} = 4$. 126

6.1 Plot of $\rho(\upsilon)$, $\alpha(\upsilon)$ and $\psi(\upsilon)$ as a function of the violation probability. . . . 146
6.2 Average transmission power of the non-robust and the worst-case robust SLP schemes versus SINR target for a system with $N_\mathrm{t} = 6$. 149
6.3 Scatter plot of the noise-free received signals at $\gamma = 10$ dB with a fixed channel, $\xi^2 = 0.005$ and $\upsilon = 0.05$. 150
6.4 Average per-user received SINR versus target SINR with $\xi^2 = 0.005$ and $\upsilon = 0.05$. 151
6.5 Average received SINR at $\gamma = 15$ dB with $\xi^2 = 0.005$ and $\upsilon = 0.05$. 152
6.6 Probability of outage versus target SINR under two different settings with $\xi^2 = 0.005$, $\upsilon = 0.05$ and $\xi^2 = 0.01$, $\upsilon = 0.2$. 152
6.7 Average transmission power versus target SINR with $\xi^2 = 0.005$ and $\upsilon = 0.05$. 153
6.8 Energy efficiency comparison of different SLP schemes versus target SINR with $\xi^2 = 0.005$ and $\upsilon = 0.05$. 154
6.9 Average symbol error rate per user versus target SINR for two different scenarios with $\xi^2 = 0.005$, $\upsilon = 0.05$ and $\xi^2 = 0.01$, $\upsilon = 0.2$. 155
6.10 Average transmission power as a function of uncertainty variance with $\gamma = 10$ dB and $\upsilon = 0.05$. 155
6.11 Per-user energy efficiency as a function of uncertainty variance with $\gamma = 10$ dB and $\upsilon = 0.05$. 156
6.12 Feasibility rate as a function of violation probability at $\gamma = 10$ dB with $\xi^2 = 0.01$. 157
6.13 Scatter plot of the noise-free received signals at $\gamma = 5$ dB with a fixed channel, $\upsilon = 0.1$ and $\xi^2 = 0.001$. 158
6.14 Energy efficiency comparison of BLP and SLP schemes versus target SINR with $\upsilon = 0.1$ and $\xi^2 = 0.001$. 159
6.15 Per-user energy efficiency as a function of violation probability with $\gamma = 5$ dB and $\xi^2 = 0.001$. 160

6.16 Feasibility rate of different robust precoding schemes as a function of uncertainty variance under two settings with $\gamma = 5$ dB, $\upsilon = 0.1$ and $\gamma = 10$ dB, $\upsilon = 0.05$. 160

7.1 The considered system model where the output of the symbol-level precoder is subject to linear distortion before being transmitted to the users. 165
7.2 Energy efficiency comparison of different SLP schemes under linear distortions. 170

8.1 The considered transmission scheme with symbol-level precoding where each I/Q channel undergoes quantization via finite-resolution DACs. . . . 178
8.2 The DPCIRs are depicted in green color for the optimized 8-ary constellation. 180
8.3 The loss in optimality due to the penalized formulation with soft CI constraints for $N_\mathrm{u} = 4$: (a) Deviation form the CI constraints versus the number of transmit antennas N_t; (b) Comparison of γ^* obtained from the original and the reformulated problem versus N_t. 189
8.4 The users' noise-free received signal with $(N_\mathrm{t}, N_\mathrm{u}) = (16, 4)$ at a transmit SNR of 0 dB, i.e., $p/\sigma^2 = 1$. 190
8.5 Average per-user BER versus transmit SNR for an MU-MIMO system with (a) $(N_\mathrm{t}, N_\mathrm{u}) = (16, 4)$; (b) $(N_\mathrm{t}, N_\mathrm{u}) = (64, 8)$. 192
8.6 Average users' BER versus the number of transmit antennas N_t for a fixed number of users $N_\mathrm{u} = 8$ at a transmit SNR of 5 dB. 192
8.7 Power efficiency as a function of N_t for a fixed number of users $N_\mathrm{u} = 8$ at SNR = 5 dB. 194
8.8 Average number of outer iterations of the QSLP method to reach a squared error of 10^{-4} as a function of transmit SNR with $(N_\mathrm{t}, N_\mathrm{u}) = (16, 4)$. . . . 195
8.9 Average number of outer and inner iterations for the QSLP method to reach a squared error of 10^{-4} as a function of transmit SNR in linear scale.195

9.1 Schematic diagram of the considered hybrid transmitter architecture with fully-connected switching and phase-shifting networks. 203
9.2 Convergence behavior of Algorithm 6 versus iteration number for $\mu = 10^{-4}$ and $\vartheta = 1.1$. 210
9.3 Power consumption of different hybrid SLP schemes as a function of N_l at SNR $= -5$ dB with $N_\mathrm{t} = 64$ and $N_\mathrm{u} = 4$. 216
9.4 Average percentage of deactivated phase shifters as a function of N_l at SNR $= -5$ dB with $N_\mathrm{t} = 64$ and $N_\mathrm{u} = 4$. 217
9.5 Average per-user symbol error rate versus transmit SNR with $(N_\mathrm{t}, N_\mathrm{l}, N_\mathrm{u}) = (64, 8, 8)$. 218
9.6 Per-user spectral efficiency versus transmit SNR with $(N_\mathrm{t}, N_\mathrm{l}, N_\mathrm{u}) = (64, 8, 8)$.218
9.7 Energy efficiency versus transmit SNR with $(N_\mathrm{t}, N_\mathrm{l}, N_\mathrm{u}) = (64, 8, 8)$. . . . 220
9.8 Energy efficiency at SNR $= -5$ dB as a function of N_l with $N_\mathrm{t} = 64$ and $N_\mathrm{u} = 4$. 220

9.9 Average number of iterations till convergence of the proposed hybrid SLP algorithm as a function of N_l with $N_\mathrm{t} = 64$. 222

Contents

Acknowledgements **6**

Acknowledgements **8**

List of Abbreviations **11**

Notations **15**

List of Tables **17**

List of Figures **19**

Preface **27**

Support of the thesis 27

Publications 27

Journal papers 27

Letters 28

Conference Papers 28

Publications not Included in the Thesis 29

Patents 29

1 Introduction **31**

1.1 Background 31

1.2 Motivation, Problem Definition and Methodology 35

1.3 Literature on Symbol-Level Precoding 37

1.4 Thesis Outline and Contributions 41

2 Multiuser Precoding in Unicast MU-MIMO Downlink Systems **49**

2.1 Preliminaries on MU-MIMO Interference Channels 49

2.2 Linear Block-Level Precoding 51

2.2.1 Spatial Matched Filter (MF) 52

2.2.2 Zero-Forcing (ZF) 52
2.2.3 Regularized Zero-Forcing (RZF) 53
2.2.4 Wiener Filter (WF) 53
2.3 General Families of Objective-Oriented Precoding 54
2.3.1 SINR-Constrained Power Minimization Problem 54
2.3.2 Power-Constrained Precoding Design 55
2.4 Symbol-Level Precoding 56

3 A Generic Design Framework for Constructive Interference Based Symbol-Level Precoding 59
3.1 Introduction 60
3.2 System Model and Problem Definition 62
3.2.1 Interpretation of Symbol-Level SINR Constraints 63
3.2.2 Definition of the SLP Design Problem with CI Constraints 64
3.3 Constructive Interference Regions 65
3.3.1 Distance Preserving Constructive Interference Regions 67
3.3.2 Union Bound Constructive Interference Regions 68
3.4 Characterization of DPCIRs 70
3.5 Design Formulations for the SLP Problem 74
3.5.1 DPCIR-based SLP Power Minimization 75
3.5.2 UBCIR-based SLP Power Minimization 78
3.5.3 DPCIR-based SLP SINR Balancing 79
3.6 Simulation Results 84
3.7 Conclusions 92

4 Computationally-Efficient Symbol-Level Precoding–Part I: Derivation 93
4.1 Introduction 94
4.2 System Model 95
4.3 Optimal Solution Structure of the SINR-Constrained Power Minimization SLP Design 96
4.3.1 Optimality Conditions for the SLP Power Minimization Problem . 97
4.4 Low-Complexity SLP Design 99
4.4.1 Closed-Form Approximate Solution 99
4.4.2 Improved Closed-Form Approximate Solution 100
4.5 Computational Complexity Analysis 102
4.6 Simulation Results 103
4.7 Conclusions 107

5 Computationally-Efficient Symbol-Level Precoding–Part II: Implementation 109
5.1 Introduction 111
5.2 Overview of the CF-SLP design 112
5.3 Low-Complexity Implementation of CF-SLP 113
5.4 FPGA Design 115

5.4.1 RTL I/O Ports Description 117
5.4.2 Resource Utilization and Timing Estimates 118
5.4.3 Design Validation 121
5.5 Numerical and Simulation Results 123
5.6 Conclusions 126

6 Robust Symbol-Level Precoding under System Uncertainties – Part I: Channel Uncertainty 129
6.1 Introduction 130
6.2 System and Uncertainty Model 132
6.2.1 Bounded Uncertainty Region 133
6.2.2 Stochastic Uncertainty Region 134
6.3 Robust CI Formulation with Imperfect CSI 135
6.3.1 Worst-Case Robust Formulation 135
6.3.2 Stochastic Robust Formulation 137
6.3.3 Relative Tightness Comparison 144
6.4 Robust SLP Optimization Problem 146
6.4.1 Computational Complexity Analysis 147
6.5 Simulation Results 148
6.6 Conclusions 161

7 Robust Symbol-Level Precoding under System Uncertainties – Part II: Design Uncertainty 163
7.1 Introduction 164
7.2 System Model and Problem Definition 164
7.3 Worst-Case Design Formulation 166
7.4 Simulation Results 170
7.5 Conclusions 171

8 Quantized Symbol-Level Precoding for Massive MU-MIMO Systems 173
8.1 Introduction 174
8.2 System Model 176
8.2.1 Signal Model 177
8.2.2 Quantization Model 178
8.3 Problem Formulation 179
8.4 Quantized Symbol-Level Precoding Design 181
8.4.1 Computational Complexity Analysis 185
8.5 Special Case: One-Bit Quantized Precoding 187
8.6 Simulation Results 188
8.7 Conclusions 194

9 Hybrid Symbol-Level Precoding for mmWave MU-MIMO Systems 197
9.1 Introduction . . . 198
9.2 System, Signal and Channel Model . . . 201
9.2.1 System and Signal Model . . . 201
9.2.2 Multiuser mmWave Channel Model . . . 203
9.3 Hybrid Symbol-Level Precoding Design . . . 204
9.3.1 Hybrid Precoder with Phase Shifter Selection . . . 205
9.3.2 Convergence Analysis . . . 208
9.3.3 Analysis of Computational Complexity . . . 210
9.4 Energy Efficiency Analysis . . . 211
9.5 Simulation Results . . . 214
9.6 Conclusions . . . 221

10 Concluding Remarks and Future Work 225
10.1 Main Conclusions . . . 225
10.2 Future Work . . . 228

Appendix A Appendices for Chapter 3 231
A.1 Proof of Lemma 2 . . . 231
A.2 Proof of Lemma 3 . . . 231
A.3 Proof of Theorem 4 . . . 232
A.4 Proof of Lemma 6 . . . 233
A.5 Proof of Lemma 5 . . . 234

Appendix B Appendices for Chapter 4 235
B.1 Proof of Lemma 7 . . . 235

Appendix C Appendices for Chapter 6 237
C.1 Proof of equality (**b**) in (6.26) . . . 237
C.2 Derivation of an equivalent SOC formulation for safe approximation II . . 238

Appendix D Appendices for Chapter 7 241
D.1 Proof of Lemma 14 . . . 241
D.2 Proof of Theorem 15 . . . 242
D.3 Proof of Lemma 16 . . . 244

Appendix E Appendices for Chapter 8 247
E.1 Proof of Lemma 17 . . . 247
E.2 Proof of Lemma 19 . . . 247
E.3 Derivation of an approximate upper bound on L . . . 249

Appendix F Appendices for Chapter 9 251
F.1 Proof of Lemma 20 . . . 251

Bibliography 253

Preface

This Ph.D. Thesis has been carried out from July, 2017 to January, 2021, at the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg, under the supervision of Prof. Björn Ottersten and Dr. Farbod Kayhan. The time-to-time evaluations of the Ph.D. Thesis were duly performed by the CET members Prof. Björn Ottersten, Dr. Farbod Kayhan, and Dr. Bhavani Shankar Mysore.

Support of the Thesis

This Ph.D. Thesis has been fully supported by the Luxembourg National Research Fund under CORE Junior project: C16/IS/11332341 Enhanced Signal Space opTImization for satellite comMunication Systems (ESSTIMS). The partial support from SIGCOM is also gratefully acknowledged.

Publications

The original publications that have been produced during the period of Ph.D. candidacy is listed below. These publications are referred to in the text by J ≡ Journal, L ≡ Letter, C ≡ Conference, and P ≡ Patent.

Journal Papers

[J1] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Symbol-Level Precoding Design Based on Distance Preserving Constructive Interference Regions," IEEE Transactions on Signal Processing, vol. 66, no. 22,pp. 5817-5832, Nov. 2018.

[J2] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Robust SINR-Constrained Symbol-Level Multiuser Precoding with Imperfect Channel Knowledge," IEEE Transactions on Signal Processing, vol. 68, no.1, pp. 1837-1852, Mar. 2020.

[J3] Alireza Haqiqatnejad, Farbod Kayhan, Shahram ShahbazPanahi, and Björn Ottersten, "Finite-Alphabet Symbol-Level Multiuser Precoding for Massive MU-MIMO Downlink," Submitted to IEEE Transactions on Signal Processing in August 2020.

[J4] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Energy-Efficient Hybrid Symbol-Level Precoding for Large-Scale mmWave Multiuser MIMO Systems," IEEE Transactions on Communications, doi: 10.1109/TCOMM.2021.3058967.

[J5] Alireza Haqiqatnejad, Jevgenij Krivochiza, Juan Merlano Duncan, Symeon Chatzinotas, and Björn Ottersten, "Design Optimization for Low-Complexity FPGA Implementation of Symbol-Level Multiuser Precoding," Accepted for Publication in IEEE ACCESS, Feb. 2021.

Letters

[L1] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Constructive Interference for Generic Constellations," IEEE Signal Processing Letters, vol. 25, no. 4, pp. 586-590, Apr. 2018.

[L2] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Power Minimizer Symbol-Level Precoding: A Closed-Form Sub-Optimal Solution," IEEE Signal Processing Letters, vol. 25, no. 11, pp. 1730-1734, Sep. Nov. 2018.

Conference Papers

[C1] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Symbol-Level Precoding Design for Max-Min SINR in Multiuser MISO Broadcast Channels," in Proc. 19th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, Greece, Jun. 2018.

[C2] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Robust Design of Power Minimizing Symbol-Level Precoder under Channel Uncertainty," in Proc. IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, Dec. 2018.

[C3] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "An Approximate Solution for Symbol-Level Multiuser Precoding Using Support Recovery," in Proc. 20th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, Jul. 2019.

[C4] Alireza Haqiqatnejad, Shahram ShahbazPanahi, and Björn Ottersten, "A Worst-Case Performance Optimization Based Design Approach to Robust Symbol-Level Precoding for Downlink MU-MIMO," in Proc. 7th IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, Canada, Nov. 2019.

[C5] Alireza Haqiqatnejad, Farbod Kayhan, Shahram ShahbazPanahi, and Björn Ottersten, “One-Bit Quantized Constructive Interference Based Precoding for Massive Multiuser MIMO Downlink,” in Proc. IEEE International Conference on Communications (ICC), Virtual Conference, Jun. 2020.

[C6] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, “Energy-Efficient Hybrid Symbol-Level Precoding via Phase Shifter Selection in mmWave MU-MIMO Systems,” in Proc. IEEE Global Communications Conference (GLOBECOM), Taipei, Taiwan, Dec. 2020.

Publications not Included in the Thesis

[C7] Sumit Gautam, Jevgenij Krivochiza, Alireza Haqiqatnejad, Symeon Chatzinotas, and Björn Ottersten, “Boosting SWIPT via Symbol-Level Precoding,” in Proc. 21st IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Virtual Conference, May 2020.

[C8] Farbod Kayhan, Alireza Haqiqatnejad, Joel Grotz, Nader Alagha, “Symbol vs Block Level Precoding in Multibeam Satellite Systems,” in Proc. 36th International Communications Satellite Systems Conference (ICSSC), Niagara Falls, Canada, 2018.

[C9] Alireza Haqiqatnejad and Farbod Kayhan, “Unified Satellite and Terrestrial ACM Design,” in Proc. 35th International Communications Satellite Systems Conference (ICSSC), Trieste, Italy, 2017.

Patents

[P1] Farbod Kayhan, Alireza Haqiqatnejad, Bhavani Shankar, and Björn Ottersten, ”Method and Device for Adaptive Coding and Modulation”, Filed on Oct. 2018 in Luxembourg, Publication Number: WO/2019/073029

Chapter 1

Introduction

This chapter introduces the problem of interest in this thesis. The motivation, contribution and organization of the thesis are presented in the subsequent sections. A brief review of the relevant literature in the last section closes this chapter.

1.1 Background

Multiuser interference (MUI) is one of the major performance-limiting factors in simultaneously serving multiple users in the same time-frequency resource block over a wireless multiple-input multiple-output (MIMO) downlink channel. The MUI, which is also known as co-channel interference, may adversely affect the downlink transmission's achievable rate. One approach to mitigate the MUI at the transmit side is to pre-compensate for its undesired effect on the received signal, which is commonly known as multiuser transmit beamforming, or precoding [1]. Multiuser precoding techniques address this issue by handling the co-channel interference via processing and spatially multiplexing the users' intended data streams prior to transmission. In principle, the interference mitigation capabilities of multiuser precoding schemes are enabled by employing multiple antennas at the transmitter, providing sufficient degrees of freedom to manage the MUI.

When perfect transmit-side channel state information (CSI) is non-causally available, it is well known that dirty paper coding (DPC) can achieve the sum-rate capacity of the MU-MIMO broadcast channel at an impractically high computational complexity. In addition to simple linear precoding schemes such as matched filter (MF), also known as maximum ratio transmission (MRT) [2], zero-forcing (ZF) or regularized zero-forcing (RZF) [3,4], and Wiener filter (WF), also known as minimum mean square error (MMSE) [5], extensive research focusing on practical yet efficient multiuser precoding techniques have been reported in the literature; see, e.g., [1, 6–10] and the references therein. Most of the proposed techniques in this line of research fall within the category of objective-oriented precoding approaches where closed-form solutions do not exist.

In general, an objective-oriented multiuser precoding design can be expressed as a constrained optimization problem [6, 7]. The design problem aims to balance some system-centric and user-centric objectives/requirements, depending on the network's operator strategy. Power and sum-rate are often regarded as system-centric criteria [9]. Transmit power is considered, for example, to control the inter-cell interference in multicell wireless networks, and sum-rate is a measure of the overall system performance. On the other hand, a frequently-used user-centric criterion is signal-to-interference-plus-noise ratio (SINR), which is an effective measure of quality-of-service (QoS) in multiuser interference channels [11]. In particular, both bit error rate (BER) and capacity, which are two relevant criteria from a practical point of view, are closely related to maximizing SINR [10]. Considering different types of optimization criteria, some well-known formulations for the multiuser precoding problem are QoS-constrained power minimization [12,13], SINR balancing [8,10,14], and (weighted) sum-rate maximization [9,15,16]. In the sequel, we primarily focus on the power minimization problem with SINR constraints and the power-constrained SINR balancing problem based on a max-min fairness criterion.

The existing multiuser precoding schemes can be broadly classified into two groups, namely, block-level and symbol-level techniques. In designing a block-level precoder, the convention is to exploit the transmit-side CSI in order to suppress the MUI, regardless of the users' data symbols. A crucial assumption is therefore the availability of instantaneous or stochastic CSI at the transmitter [17], based on which a block-level precoder has to be recalculated depending on the channel coherence time. On the contrary, it has been shown that the MUI might not always be harmful or destructive as it is possible to exploit the constructive part of the interference [18], or even converting the interfering components into constructive interference (CI) by instantaneously exploiting the users' data symbols [19]. By doing do, the MUI can be turned into an additional source of power that constructively contributes to the users' received signals and improves the overall performance of a multiuser system. As a result, the conventional viewpoint on multiuser precoding can turn from block-level approaches towards a more sophisticated design that further exploits the data information (DI), which is readily available at the transmitter, in addition to the CSI. Such a design approach is commonly referred to as symbol-level precoding (SLP) [20–22] and has been shown to be a promising alternative for multiuser precoding schemes.

In principle, the CI is identified based on the philosophy that a noise-free received signal can be decoded correctly not necessarily when it is close enough to the intended symbol, rather, as long as it lies within the correct decision region even far away from the target symbol. This has been the underlying rule of thumb in defining various CI regions (CIR); see, e.g., [20, 21, 23]. The CIRs are typically defined with the aim of enhancing or guaranteeing a certain level of detection accuracy for the users.

To benefit from the potential advantages of the CI, one needs to process the transmit signal specifically for every set of users' intended symbols, i.e., on a symbol-by-symbol basis. The idea of designing the multiuser precoder on a symbol-level basis and converting the MUI into CI was first introduced in [24], and then elaborated in [20] and [21], where the definition of CI was formalized. Earlier research in this direction decomposes the MUI into constructive and destructive parts and attempts to exploit the CI by canceling the destructive interference via linear precoding techniques such as ZF and RZF [18]. This design strategy was further improved in [19], where the destructive interference is not canceled but converted to the CI such that it is aligned with the desired symbols through a symbol-specific rotation matrix. This alignment led to one of the first definitions of CIRs, referred to as CIRs with strict phase constraints. The strict phase constraints were later adopted in [21] for an elaborated design of the SLP scheme, and then improved in [25] by defining relaxed CIRs where the phase of the received signal is allowed to have a limited angular deviation as long as it stays within the correct detection region.

The SLP design problem is usually expressed as a constrained optimization problem. Accordingly, the symbol-level design of a multiuser precoder involves solving an optimization problem for instantaneous realizations of the users' data symbols. The optimization constraints are defined so that the noise-free received signal of each user is pushed into the predefined CIR. Therefore, formulation of the optimization problem, and particularly the constraints, depend on the adopted modulation scheme (i.e., the signal constellation). The objective function, on the other hand, depends on the design criterion.

Compared to conventional block-level precoding techniques, it has been shown that an SLP scheme can achieve significant gains at the cost of higher transmitter complexity [19, 20, 26], but without re-designing the receiver. While the precoder's linear structure can be preserved under an SLP scheme, one may also adopt a non-linear structure by forming a virtual multicast formulation to directly design the precoded transmit signal [21], instead of calculating the precoding matrix. Some other advantages offered by an SLP technique are described below:

- The symbol-level design of a multiuser precoder can significantly improve the system performance in terms of power efficiency, symbol error rate (SER), and users' received SINR, depending on the design criterion. These potential improvements originate from two sources. Firstly, due to the CIRs, the precoding optimization problem enjoys a larger search space (feasible region) to find the optimal precoded signal compared to the conventional block-level precoding schemes. Secondly, the symbol-by-symbol precoding design approach provides additional degrees of freedom to handle the MUI particularly for each set of symbols.

- A linear block-level precoding scheme can only satisfy the design constraints when averaged over a block of symbols; however, the constraints might be violated instantaneously. This issue becomes problematic, e.g., in systems with strict hardware-related restrictions such as peak per-antenna power constraints or unit-modulus signal constraints. On the contrary, the SLP techniques can instantaneously guarantee the design constraints at a symbol-level scale as the precoder is specifically designed for each symbol period.

- Unlike traditional small-scale transmitter architectures that employ highly linear and power-inefficient amplifiers, the MIMO system deployments with large-scale antenna arrays require power-efficient amplifiers due to practical considerations regarding the associated cost and power consumption. However, power-efficient amplifiers typically show poor linearity characteristics, and therefore, require the input signals to have a low peak-to-average power ratio (PAPR). As a result, the per-antenna transmit power needs to have adequate dynamic properties in order to limit the associated non-linearity effects. The SLP schemes are capable of compensating for such non-linearity/imperfections on a symbol-level basis. For example, constant-envelope (CE) SLP is an effective design approach to achieve a favorable unit PAPR. Furthermore, the SLP design may target optimizing the power dynamics, such as the dynamic range and PAPR, by minimizing the instantaneous peak transmit power, both in temporal (i.e., per-symbol) and spatial (i.e., per-antenna) dimensions. By doing so, one can limit the performance degradation caused by a non-linear amplifier due to the amplitude-to-amplitude (AM-AM) and the amplitude-to-phase (AM-PM) distortion.

- The structure of each user's receiver is independent of the SLP design strategy. For example, in the case of downlink transmission with equiprobable signaling in a multiuser system, each user can detect its intended symbol by applying an optimal single-user detection rule, e.g., the maximum-likelihood (ML) detector. Moreover, an SLP scheme can reduce the receiver's complexity as no further post-processing or compensation is needed on the users' received signals, which is a considerable advantage in the likely case where the users have limited computational capabilities. This is mainly due to the fact that the noise-free signal received by each user is accommodated onto the intended CIR.

- The number of transmit antennas fundamentally limits the number of simultaneously served users (i.e., multiplexed data streams) by a linear block-level precoding scheme. In contrast, the SLP techniques support multiplexing more data streams than the number of transmit antennas [27, 28], while the system performance can be preserved or even improved.

To have a complete introduction to the SLP design paradigm, we close this section by pointing out some of the disadvantages and practical challenges in implementing a symbol-level precoder. These challenges include the need for setting the modulation scheme in advance, a substantially increased computational burden at the transmitter, sensitivity to system uncertainties (e.g., CSI errors) and sub-optimality of SINR pilots and log-likelihood ratio (LLR) calculation algorithms; see [29, 30]. In the next section, we elaborate more on the first three challenges and define the problems of interest in this thesis.

1.2 Motivation, Problem Definition and Methodology

The main motivations behind the work carried out in this thesis are explained in four parts, and the related problems are defined accordingly. As mentioned earlier, the potential performance improvements offered by an SLP technique comes with some practical challenges that need to be properly addressed. Among the subsequent motivations, the first three ones have been raised to address the existing design challenges, while the last one aims to propose power-efficient design approaches according to new precoding architectures.

Firstly, none of the existing research has attempted to define the CI constraints in the optimization problem independently of the signal constellation from which the users' intended symbols are taken. As a consequence, an essential concern in designing the SLP techniques is that the modulation scheme needs to be set in advance as it recasts the design problem. This particularly becomes an issue in systems with an adaptive coding and modulation (ACM) technique where the modulation scheme is not fixed. Motivated by such an issue, in this thesis, we aim to obtain a universal design formulation that applies to modulation schemes with any given constellation shape and order. To do so, we primarily assume a generic geometry for the constellation set and attempt to describe the CIRs systematically such that the resulting description is invariant in form to the constellation type. Meanwhile, an important design consideration is to describe the CIRs in an easy-to-handle convex form. We further aim to utilize these systematic descriptions to formulate the CI restrictions as convex constraints, and ultimately, provide a convex design formulation for the SLP problem.

Secondly, one of the main challenges in optimally designing a symbol-level precoder is its high computation cost. The SLP design, in its original form, is accomplished by solving an optimization problem for every instantaneous set of users' symbols. As a result, the downlink transmission in each symbol period requires a specific precoding design that imposes a relatively high computational complexity on the transmitter. To address this issue, in this thesis, we aim to find more computationally-efficient solutions for the SLP problem based on our generic design framework. For this purpose, we attempt to simplify the design problem and exploit the optimal solution structure. The crucial consideration in deriving such solutions is the required computational complexity, which may be achieved by sacrificing the design optimality and lead us to an approximate solution.

Thirdly, it is known that the multiuser precoding schemes are quite sensitive to the system uncertainties. In the particular case with SLP techniques, achieving CI at the receiver side relies highly on the accuracy of the system parameters, e.g., the transmit-side CSI knowledge. In this regard, the problem of robust SLP design under imperfect CSI becomes of importance in order to guarantee the reliable performance of the downlink system. However, this problem has not been well investigated in the literature. More specifically, while some worst-case design approaches have been proposed under bounded CSI errors, there were no published results (at the time of performing this work) studying the robust design of SLP by adopting a stochastic model for the CSI uncertainty. Considering that a stochastic uncertainty model adequately captures an imperfect channel estimation process, a stochastic robust SLP design is of high practical interest; however, the literature lacks such a design approach. This thesis aims to address this gap by redefining the CI restrictions as probabilistic constraints, meaning that the CI is guaranteed with a certain probability. The probabilistic form of CI restrictions may lead us to computationally intractable constraints that have to be handled via stochastic optimization approaches. Moreover, we also address the problem in scenarios where the design process is subject to uncertainty, e.g., due to finite precision of the underlying design and implementation technology. The results of this part can point the research community to address new practical challenges in robust design when the design parameters are subject to uncertainty.

Finally, to respond to the consistently growing traffic and data rate demand of wireless users, the communications systems have shifted towards new technologies such as employing large-scale antenna arrays, known as massive MIMO, and operating at millimeter (mmWave) frequencies. However, deployment of these new technologies requires a large number of analog, digital, or mixed signal components, possibly operating in the 30-300 GHz frequency band. Consequently, power efficiency has become a serious practical concern in implementing such systems, which has pointed the research community to focus on low-cost and power-efficient transceiver architectures. In the context of (multiuser) precoding, quantized and hybrid analog-digital precoding schemes have been introduced to improve the system's cost and power efficiency by, respectively, employing low-resolution digital-to-analog converters (DAC) and analog components such as phase shifters or switches. The literature on SLP schemes also includes some published work to date addressing the design problem of quantized or hybrid precoding on a symbol-level basis. Nonetheless, there is still room to develop more power-efficient SLP techniques by investigating different precoding architectures or more efficient algorithms. Motivated by this, we aim to address both quantized and hybrid CI-based SLP design problems in this thesis. The quantized SLP problem comes with several difficulties, such as discrete-domain optimization variables due to finite-resolution DACs. To tackle this difficulty, we attempt to find an equivalent binary problem and solve it via efficient binary optimization algorithms. On the other hand, to design the hybrid scheme, we investigate the architecture with a combination of switches and phase shifters in the analog precoder. Such an architecture has not been investigated in the literature for a symbol-level hybrid precoder. We aim to design the digital precoder and the switching network on a symbol-level basis while adopting a block-level design for the phase-shifting network. Similarly, the hybrid design problem of interest can be formulated and solved via efficient binary optimization tools. The use of the switching network along with the CI-based design allows us to further improve the system's power efficiency compared to the state-of-the-art.

1.3 Literature on Symbol-Level Precoding

The literature on CI-based SLP has become very rich within the recent past years and contains both fundamental research, e.g., on the design problem and CI definition, and application-based studies in different wireless communications scenarios. This section covers both these research areas and provides an extensive review of the existing work.

Earlier research on the CI-based precoding decomposes the MUI into constructive and destructive parts and attempts to exploit the CI by canceling the destructive interference via simple linear precoding techniques such as ZF and RZF [18, 31]. In [31] and [18], the authors propose a selective precoding scheme that preserves the CI but eliminates the destructive component via ZF. This design strategy was further improved in [19], where the destructive interference is not canceled but converted to the CI such that it is aligned with the desired symbols through a symbol-specific rotation matrix, called correlation-rotation precoding. This alignment led to one of the first definitions of CIRs, referred to as CIRs with strict phase constraints (SPC), which has later been shown to be sub-optimal. In contrast to the selective precoding in [31] and [18] that exploits the MUI only when it is constructive, the correlation-rotation precoding controls the MUI so that the entire interference becomes constructive for each user. The SPCs has later been adopted in [21] for an elaborated design of the SLP scheme, and then improved in [25] by defining relaxed CIRs where the phase of the received signal is allowed to have a limited angular deviation as long as it stays within the correct detection region.

The CI-enhanced design approach has further been applied to the non-linear Tomlinson-Harashima precoding (THP) in [32, 33] and vector-perturbation (VP) precoding in [34]. The work in [32] introduces a complex scaling factor for the first user in a way that the interfering signals are more properly aligned with the intended symbols, where the scaling factor can be optimized to reduce the transmit signal power. This scheme has further been improved in [33] where complex scaling factors are considered for several users, and not only the first user. Moreover, in [34], the authors apply the CI concept into VP precoding by substituting a linear symbol-scaling (SS) precoder for the perturbation vector search.

Beyond the above schemes that are mostly based on simple linear precoding methods, a broader group of CI-based SLP techniques fall into the category of objective-oriented precoding. These SLP techniques aim to further improve the multiuser downlink system performance via optimization tools [20, 21, 25, 35–38]. The CI-based SLP schemes in [21] and [36] are essentially based on the idea of correlation-rotation and adopt CIRs with SPCs, but aim to improve the performance by avoiding the underlying ZF operation. In [21], two design formulations, namely, power minimization and SINR balancing, are studied and solved using iterative optimization algorithms. The SPCs are evolved in [20] and [37], where CIRs with not necessarily strict phase alignments are defined and characterized for PSK constellations. Due to introducing a larger search space (i.e., feasible region) for the precoding optimization problem, the CIRs with non-strict phase constraints (NSPC) lead to higher performance gains compared to the previous schemes. This type of CIRs has been widely adopted in designing various precoding schemes; see, e.g., [27, 28, 39–42]. In addition to the CIRs with NSPC, a sort of sub-optimal relaxed CIRs is introduced in [25, 38], where the so-called phase margin allows for an angular deviation from the intended symbol's phase.

Most of the above research considers phase-shift keying (PSK) constellations in formulating the SLP design problem and utilizes the CIRs with strict or non-strict phase constraints. However, the same CIR types cannot be applied to quadrature amplitude modulations (QAM) due to the different geometry of the associated constellation set. In particular, QAM constellations may have some inner symbols, i.e., symbols with bounded decision regions, which limits the applicability of CI. Furthermore, they may not be fully compliant with the phase-constrained definition of CIRs due to their non-circular shape. Extended definitions of CIRs, being applicable to multi-level modulations such as QAM, have been studied in [26–28, 41–44]. Based on the philosophy behind the definition of CIRs, only the outer constellation symbols can benefit from exploiting CI. Accordingly, one might not expect the CI-based SLP schemes to perform promisingly for high-order QAM modulations. In contrast, the work in [28] shows that substantial gains can still be achieved, even for a 64-QAM constellation. This is due to the fact that the CI-based precoding further alleviates the noise enhancement issue, which is known to be more prominent for high-order QAM modulations.

It is also worth mentioning that while most of the existing CI-based SLP designs focus on the noise-free received signals, there has been some work taking the effect of noise into account in designing the precoder, leading to noise-robust (NR) definitions for CIRs [39, 45, 46]. Among the other previous work addressing the SLP problem, we refer to peak per-antenna power minimization SLP [47], MMSE approach to SLP [48], and CI-based multi-group multicast beamforming [49].

As previously mentioned, CIRs are constellation-dependent regions. In light of this, all the above SLP schemes consider a specific signal constellation, e.g., PSK or QAM, to formulate the CI-based precoding problem; therefore, the design needs to set the modulation scheme in advance. This issue becomes of particular importance in systems employing adaptive coding and modulation (ACM) techniques, where different combinations of modulation and coding (MODCOD) schemes are used to achieve different target spectral efficiencies with a certain granularity. However, the literature lacks a generic CIR definition, and accordingly, an SLP design framework that supports any given modulation scheme. Moreover, most of the above objective-oriented SLP techniques are based on solving an optimization problem via an iterative algorithm. An important concern entangled with these SLP designs is the required high computational complexity. This becomes even more challenging when one takes into account that such an optimization problem has to be solved specifically for every set of users' data symbols. With the intention of addressing this issue, there has been a particular research focus on deriving more computationally-efficient SLP designs, e.g., [28, 40, 50]. Although these solutions drive the SLP techniques one step further towards being implemented in practical scenarios, each of them has been tailored for a specific constellation, and therefore, they still have the problem of non-seamless operation in ACM systems.

The concept of CI exploitation via SLP techniques has also been introduced into various wireless communication scenarios and applications, e.g., cognitive radio networks (CRN) [51–55], cooperative multi-hop MIMO relaying [56], simultaneous wireless information and power transfer (SWIPT) [57–60], directional modulation [46,61,62], physical-layer (PHY) security [58,61–64], full-duplex communications [65–67], distributed antenna systems (DAS) [68], faster-than-Nyquist signaling based on spatio-temporal CI [69–71], CE precoding [72–75], antenna selection schemes [76–79], hybrid analog-digital precoding [75, 80–83], quantized precoding with low-resolution DACs [84–90], and non-linear systems [91–93].

The literature mentioned above on the CI-based SLP design problem is categorized and summarized in Table 1.1. This table also includes the addressed problems in the subsequent chapters of this thesis and the references to the corresponding publications (shown in blue), indicating the relevance and positioning of our work within the relevant literature.

Related Work

As mentioned earlier, the problem of SLP design for MU-MIMO downlink systems has been widely studied in several contexts within the literature. Below, we classify and refer to the existing research publications that are directly related to the work carried out in this thesis.

Symbol-Level Precoding for MU-MIMO Downlink Systems

Most of the existing research in the literature formulates and solves the SLP problem specifically for a given modulation scheme with either a single-level or multi-level constellation set. In this regard, the majority of work studies the SLP design problem for the phase-shift keying (PSK) modulations, e.g., in [20, 21, 36–38, 61, 94, 108–110]. The problem has also been extended to multi-level modulations, e.g., quadrature amplitude modulations (QAM), in [26–28, 41–43, 105], and amplitude and phase-shift keying (APSK), in [70]. The design formulation in the above techniques depends on the constellation shape and order of the adopted modulation scheme.

From a different perspective, the objective-oriented precoding problems can be classified based on their design objective and constraints. Among different types of design criteria, we refer to two well-known formulations, namely, the power minimization problem with SINR constraints and the power-constrained SINR balancing problem via max-min fair criterion. The SLP problem minimizing the total transmit power has been studied in [20, 21, 36–38, 61], and the minimization of peak per-antenna transmit power is addressed in [94]. On the other hand, the SINR balancing problem for SLP schemes has been addressed in [20, 21, 28, 41]. For example, in [21], the non-convex SLP max-min SINR is solved using its relation to the power minimization via a bisection search. This problem is also studied in [20] and a second-order cone programming (SOCP) formulation is proposed.

Low-Complexity Symbol-Level Precoding Design and Implementation

In this line of research, some efforts have been made towards deriving low-complexity solutions to the SLP design problem, e.g., [28,34,40,50,111,112], and accordingly, some other studies have addressed efficient hardware demonstrations of these low-complexity SLP techniques, e.g., [113,114]. In [40], the authors propose an iterative algorithm with a closed-form update equation for the SLP problem with a max-min fair design criterion, where the algorithm is shown to converge to the optimal solution in a few iterations. The authors in [111] show that, given a perturbation of the target users' symbols, the SLP power minimization design is equivalent to the ZF precoding. In another work [50], the power minimization SLP problem is addressed with strict phase constraints on the received signals, and a computationally-efficient approximate solution is suggested for this particular case with a phase-shift keying (PSK) modulation scheme. The authors demonstrate an FPGA-accelerated design of this solution in [115], indicating that it is capable of providing a high symbol throughput in a real-time operation mode.

Robust Symbol-Level Precoding under System Uncertainties

Robust SLP design under channel uncertainty (i.e., imperfect CSI knowledge) has been addressed in some recent work. Worst-case robust SLP approaches are proposed in [20] and [58,63] for unsecured and secured wireless systems, respectively, aiming to design the symbol-level precoder under norm-bounded CSI errors based on the power minimization and max-min fair criteria. In [64], the authors develop an SLP design to enhance both PHY security against eavesdropping and the quality of legitimate transmissions in MU-MIMO wiretap systems, where the design is studied under different assumptions on the availability of CSI at the legitimate and eavesdropping channels, including a bounded CSI uncertainty model. However, it is important to note that as far as the SLP power minimization problem is concerned, the bounded uncertainty model may not yield an efficient solution. This modeling ultimately leads to a worst-case conservatism, which inherently increases the transmission power, though enhancing the users' received SINR and symbol error probability. To address stochastic channel uncertainties in the SLP design, in [116], a sphere bounding scheme is proposed for robust SLP power minimization with probabilistic CI constraints, where the probabilistic constraints are transformed into a tractable second-order cone (SOC) form and are tightened to achieve a lower SER but at the cost of a higher transmitted power. In another work published in [117], the problem of robust SLP design is addressed by considering quantized transmit-side CSI. The problem is solved by decomposing the inter-user interference into predictable and unpredictable (due to the quantization error) parts, where an upper bound is derived for the latter part. Targeting CI at the receiver side, the design aligns the predictable interference to achieve much higher received power over the derived upper bound, and ultimately, lower symbol error rates (SER) for the users. It is also worth mentioning that a precoding optimization problem with outage probability constraints based on a symbol-level approach is presented in [45], therein the goal is to achieve robustness to the receiver noise, but not to channel uncertainties.

On the other hand, to the best of our knowledge, the robust SLP optimization problem under design uncertainty (or more specifically, under linear distortion of the precoded signal) has not been addressed in the literature.

Symbol-Level Precoding for Low-Cost Transmitter Architectures

The problem of quantized SLP design for massive MU-MIMO downlink systems equipped with low-resolution DACs has been addressed in some recent work. In particular, the case with one-bit DACs has become an attractive research direction due to its simplicity and efficiency in terms of power consumption and hardware cost. In this line of work, in [84] and [85, 86], the authors propose SLP schemes under unit-modulus constraints dictated by the one-bit DACs, for PSK and QAM signaling. Another one-bit quantized SLP scheme is proposed in [87], where the design objective is to minimize the users' SER by minimizing the maximum (among the users) distance between a received signal and its corresponding target symbol up to a scaling factor. This scaling factor can be classified as a special case of CI regions with strict phase constraints. Moreover, the SLP problem with low-resolution DACs is addressed in [88], where the design objective is defined based on a mean square error (MSE) criterion rather than the CI constraints.

On the other hand, symbol-level design approaches to hybrid precoding are not well studied for large-scale mmWave MU-MIMO systems. Hybrid SLP design under mmWave hardware limitations has been addressed in some recent work [80, 81, 83]. In [80], the authors adopt a disjoint sub-optimal approach to optimize the digital and analog precoders with a focus on the analog precoder design, where different techniques are studied and compared. Power-efficient transmitter architectures, including antenna selection and analog-only, are studied for symbol-level precoding in [75], where it has been shown that the analog-only design can outperform the other schemes especially when the transmit array size is much larger than the number of UEs. An even more cost-effective hybrid structure is considered in [81] where the baseband digitally precoded signal is subject to one-bit quantization due to the use of low-cost one-bit DACs for each RF chain. The joint optimization of digital and analog symbol-level precoders is addressed in [83], where the authors exploit the symbol-based design of the phase-shifting network to achieve the performance of the fully-digital precoder. In practice, the design needs to switch between the phase states of the variable phase shifters at the symbol rate.

1.4 Thesis Outline and Contributions

The contributions of this thesis can be categorized into four main parts, which are organized into seven chapters. Briefly speaking, in this thesis, we address four different challenges in designing the multiuser precoder on a symbol-level basis, namely, design generality, computational complexity, system uncertainties, and hardware limitations. First, in Chapter 3, we elaborate a general framework for the SLP design problem based on different CI region types. This framework has been used frequently within the subsequent chapters in formulating the precoding design problems. Chapter 4 and Chapter

5 address the computationally-efficient design of SLP, respectively, from theoretical and implementation points of view. We study robust design of SLP under system uncertainties, including channel and design imperfections, respectively in Chapter 6 and Chapter 7. Finally, in Chapter 8 and Chapter 9, we revisit the SLP problem in low-cost transmitter architectures, where practical limitations on the number of RF chains or the resolution of DACs are imposed on the design problem.

Chapter 2: Multiuser Precoding in Unicast MU-MIMO Downlink Systems

This chapter describes the problem of multiuser precoding in multi-antenna downlink systems and provides an overview of different design approaches to this problem.

Chapter 3: A Generic Design Framework for Constructive Interference Based Symbol-Level Precoding

This chapter provides a general framework to formulate and solve the SLP problem over an MU-MIMO downlink channel. First, we consider generic modulation schemes with constellation sets of any shape and size and elaborate on optimal and relaxed CIRs. We characterize two types of CIRs, namely, distance-preserving CIR (DPCIR) and union bound CIR (UBCIR) and provide a systematic way to describe these regions as convex sets. We then confine ourselves to DPCIRs and perform a comprehensive study which allows us to derive several properties for these regions. Using these properties, we first show that any signal in a given DPCIR has a norm greater than or equal to the norm of the corresponding constellation point if and only if the convex hull of the constellation contains the origin. It is followed by proving that the power of the noise-free received signal in a DPCIR is a monotonic strictly increasing function of two parameters relating to the infinite Voronoi edges. Using the convex representations of DPCIRs and UBCIRs, we formulate two design problems, namely, the SLP power minimization with SINR constraints, and the SLP SINR balancing problem under max-min fairness criterion. We show that the SLP power minimization problem, minimizing either sum or peak (per-antenna) transmit power, can always be formulated as a convex QP. We further derive a simplified reformulation of this problem which is more computationally-efficient. Our simulation results indicate that the DPCIRs and UBCIRs allow further reduction of the transmit power compared to the state-of-the-art without increasing the computational complexity at the transmitter or receiver. The SLP max-min SINR problem, on the other hand, is non-convex in its original form, and hence is difficult to tackle. We propose alternative optimization approaches, including SDP formulation and BCD optimization. We finally discuss and evaluate the loss due to the proposed alternative methods through extensive simulation results. The material presented in this chapter has been partially published by the author in the following references:

[23] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, “Constructive Interference for Generic Constellations,” IEEE Signal Processing Letters, vol. 25, no. 4, pp. 586-590, Apr. 2018.

[95] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, “Symbol-Level Precoding Design Based on Distance Preserving Constructive Interference Regions,” IEEE Transactions on Signal Processing, vol. 66, no. 22,pp. 5817-5832, Nov. 2018.

[96] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, “Symbol-Level Precoding Design for Max-Min SINR in Multiuser MISO Broadcast Channels,” in Proc. 19th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, Greece, Jun. 2018.

Chapter 4: Computationally-Efficient Symbol-Level Precoding – Part I: Derivation

This chapter proposes two approximate yet computationally-efficient solutions for the SLP design problem. First, we study the optimal solution to the multiuser SLP design for minimization of the total transmit power under given SINR requirements. We adopt the DPCIRs and derive a simplified reformulation of the problem in the form of a standard NNLS design. Then, we analyze the optimal solution structure using the KKT optimality conditions. This leads us to obtain a computationally-efficient closed-form approximate SLP solution (CF-SLP). Meanwhile, we obtain a necessary and sufficient condition under which the power minimizer SLP is equivalent to the conventional ZF precoding. Our simulation results show that the CF-SLP technique provides significant gains over the ZF scheme and performs quite close to the optimal SLP in scenarios with a relatively small number of users; however, it shows poor performance for large numbers of transmit antennas and users. To address this drawback, we build on the CF-SLP technique to derive an improved approximate closed-form solution, named ICF-SLP, using the conditions for nearly perfect recovery of the optimal solution support. Through simulation results, we show that in comparison with the CF-SLP technique, the ICF-SLP method significantly enhances the system's performance with a slight increase in complexity. In particular, the ICF-SLP method successfully resolves the drawback of the CF-SLP technique by performing relatively close to the optimal SLP in systems with large numbers of transmit antennas and users. We also compare our computationally-efficient solutions with a fast-converging iterative NNLS algorithm, where the ICF-SLP method shows competitive performance in terms of both accuracy and complexity of the design compared to the iterative algorithm's solution. Analytical and numerical discussions on the complexities of different SLP schemes verify the computational efficiency of the proposed solutions. We show that the CF-SLP and ICF-SLP techniques enjoy a substantial reduction in the computation time compared to the optimal solution. The material presented in this chapter has been partially published in the following references:

[97] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, “Power Minimizer Symbol-Level Precoding: A Closed-Form Sub-Optimal Solution,” IEEE Signal Processing Letters, vol. 25, no. 11, pp. 1730-1734, Sep. Nov. 2018.

[98] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, “An Approximate Solution for Symbol-Level Multiuser Precoding Using Support Recovery,” in Proc. 20th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, Jul. 2019.

Chapter 5: Computationally-Efficient Symbol-Level Precoding – Part II: Implementation

To address the high computation cost of solving the SLP problem, this chapter develops and validates a low-complexity FPGA design of SLP, targeted for real-time implementations in high-throughput communication systems. The work of this chapter builds on the CF-SLP method presented in Chapter 4, and thus, the resulting design is constellation-independent which makes it appropriate for seamless handling of ACM schemes. We enable the FPGA design by expressing the proposed approximate solution in a closed-form algorithmic way and translating it to hardware description language (HDL). We then optimize the HDL code for accelerated performance and generate the HDL core to be deployed on FPGA. We provide the synthesis report for the generated HDL core, including performance, resource utilization, and interface descriptions. In order to validate our design, we simulate an uncoded transmission scheme over a downlink multiuser channel using the LabVIEW software, where the SLP HDL core is implemented as a clock-driven logic (CDL) unit. Our simulation results show that a throughput of 100 Mega symbols per second per user can be achieved for a 4×4 system with QPSK signaling via the HDL design of the proposed approximate SLP solution. We further use the MATLAB software to produce numerical results for the conventional ZF and the optimal SLP techniques as benchmarks for comparison; thereby, it is shown that the proposed low-complexity FPGA implementation offers an improvement of up to 50 percent in power efficiency compared to the ZF precoding, while it enjoys the same per-symbol complexity order as that of the ZF technique. We also evaluate the loss of the HDL implementation due to the approximation-induced and arithmetic inaccuracies with respect to the optimal SLP solution. The material presented in this chapter has been submitted for review in the following reference:

[99] Alireza Haqiqatnejad, Jevgenij Krivochiza, Juan Merlano Duncan, Symeon Chatzinotas, and Björn Ottersten, “Design Optimization for Low-Complexity FPGA Implementation of Symbol-Level Multiuser Precoding,” Accepted for Publication in IEEE ACCESS, Feb. 2021.

Chapter 6: Robust Symbol-Level Precoding under System Uncertainties – Part I: Channel Uncertainty

This chapter addresses robust design of SLP for the MU-MIMO downlink wireless channels when imperfect CSI is available at the transmitter. We consider two well-known models for the CSI imperfection, namely, bounded and stochastic uncertainty. Our design objective is to minimize the total (per-symbol) transmission power subject to CI constraints as well as the users' QoS requirements in terms of SINR. Assuming bounded channel uncertainties, we obtain a convex CI constraint based on the worst-case robust analysis, whereas in the case of stochastic uncertainties, we define probabilistic CI constraints in order to achieve robustness to statistically-known CSI errors. Since these probabilistic constraints are difficult to handle, we resort to their convex approximations given in the form of tractable deterministic robust constraints. Three convex approximations are derived based on different conservatism levels, among which one is introduced as a benchmark for comparison. We show that each of our proposed approximations is tighter than the other under specific robustness settings, while both always outperform the benchmark. Using the proposed CI constraints, we formulate a robust SLP design problem as an SOCP. Extensive simulation results are provided to validate our analytical results and to make comparisons with conventional block-level robust precoding schemes. We show that the robust design of symbol-level precoder leads to improved performance in terms of energy efficiency at the cost of increasing the computational complexity by an order equal to the number of users in the large system limit, compared to the non-robust design. The material presented in this chapter has been partially published in the following references:

[100] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Robust SINR-Constrained Symbol-Level Multiuser Precoding with Imperfect Channel Knowledge," IEEE Transactions on Signal Processing, vol. 68, no.1, pp. 1837-1852, Mar. 2020.

[101] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Robust Design of Power Minimizing Symbol-Level Precoder under Channel Uncertainty," in Proc. IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, Dec. 2018.

Chapter 7: Robust Symbol-Level Precoding under System Uncertainties – Part II: Design Uncertainty

This chapter addresses the optimization problem of SLP in MU-MIMO downlink wireless systems where the precoder's output is subject to partially-known distortions, e.g., due to finite precision of the underlying design and implementation technology. We assume a linear distortion model with bounded additive noise. The original SINR-constrained SLP problem minimizing the total transmit power is first reformulated as a penalized unconstrained problem, referred to as the relaxed robust formulation. We then adopt a worst-case design approach to protect the users' intended symbols and the targeted constructive interference with a desired level of confidence. Due to the non-convexity of the relaxed robust formulation, we propose an iterative algorithm based on the block coordinate ascent-descent method. We show through simulation results that the proposed robust design is flexible in the sense that the CI constraints can be relaxed to keep a desirable balance between achievable rate and power consumption. Remarkably, the robust formulation yields more energy-efficient solutions for appropriate choices of the penalty parameter, compared to the original SLP problem. The material presented in this chapter has been partially published in the following reference:

[102] Alireza Haqiqatnejad, Shahram ShahbazPanahi, and Björn Ottersten, "A Worst-Case Performance Optimization Based Design Approach to Robust Symbol-Level Precoding for Downlink MU-MIMO," in Proc. 7th IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, Canada, Nov. 2019.

Chapter 8: Quantized Symbol-Level Precoding for Massive MU-MIMO Systems

This chapter proposes a finite-alphabet SLP technique for massive MU-MIMO downlink systems equipped with finite-resolution DACs of any precision. We adopt a CI-based max-min fair design criterion which aims to maximize the minimum instantaneous received SINR among the users while ensuring a CI constraint for each user under the restriction that the output of the precoder is a vector with finite-alphabet discrete elements. Due to the latter restriction, the design problem is an NP-hard QP with discrete variables, and hence, is difficult to solve. We tackle this difficulty by reformulating the problem in several steps into an equivalent continuous-domain biconvex form, including equivalent representations for discrete and binary constraints. Our final biconvex reformulation is obtained via an exact penalty approach and can efficiently be solved using a standard cyclic BCD algorithm. We evaluate the performance of the proposed finite-alphabet precoding for DACs with different resolutions and show that employing low-resolution DACs can lead to higher power efficiencies. In particular, we focus on a setup with one-bit DACs and show through simulation results that compared to the existing schemes, the proposed design can achieve significant SINR gains. We further provide analytical and numerical analyses of complexity and show that our proposed algorithm is computationally-efficient as it typically needs only a few tens of iterations to converge. The material presented in this chapter has been partially published or submitted for review in the following references:

[103] Alireza Haqiqatnejad, Farbod Kayhan, Shahram ShahbazPanahi, and Björn Ottersten, "Finite-Alphabet Symbol-Level Multiuser Precoding for Massive MU-MIMO Downlink," Submitted to IEEE Transactions on Signal Processing in August 2020.

[104] Alireza Haqiqatnejad, Farbod Kayhan, Shahram ShahbazPanahi, and Björn Ottersten, "One-Bit Quantized Constructive Interference Based Precoding for Massive Multiuser MIMO Downlink," in Proc. IEEE International Conference on Communications (ICC), Virtual Conference, Jun. 2020.

Chapter 9: Hybrid Symbol-Level Precoding for mmWave MU-MIMO Systems

This chapter addresses the SLP design problem for a mmWave downlink MU-MIMO wireless system where the transmitter is equipped with a large-scale antenna array.

The high cost and power consumption associated with the massive use of RF chains prohibit fully-digital implementation of the precoder. Therefore, we consider a hybrid analog-digital architecture where a small-sized baseband precoder is followed by two successive networks of analog on-off switches and variable phase shifters according to a fully-connected structure. The use of the switching network allows us to implement a phase shifter selection mechanism. We jointly optimize the baseband precoder and the states of the switching network on a symbol-level basis, i.e., by exploiting both the CSI and the instantaneous data symbols. In contrast, the phase-shifting network is designed only based on the CSI due to practical considerations. Our approach to this joint optimization is to minimize the Euclidean distance between the optimal fully-digital and the hybrid SLP schemes. The phase shifter selection mechanism allows for significant power-savings in the analog precoder by switching some of the phase shifters off according to the switches' instantaneously optimized states. Our numerical results indicate that up to half of the phase shifters can be switched off, on average, in systems where the number of transmit antennas is much larger that the number of RF chains and users. We provide an analysis of energy efficiency by adopting appropriate power consumption models for the analog precoder and show that the energy efficiency of precoding can substantially be improved thanks to the phase shifter selection approach, compared to the fully-digital and state-of-the-art hybrid symbol-level schemes. The material presented in this chapter has been partially published or submitted for review in the following references:

[106] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Energy-Efficient Hybrid Symbol-Level Precoding for Large-Scale mmWave Multiuser MIMO Systems," IEEE Transactions on Communications, doi: 10.1109/TCOMM.2021.3058967.

[107] Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten, "Energy-Efficient Hybrid Symbol-Level Precoding via Phase Shifter Selection in mmWave MU-MIMO Systems," in Proc. IEEE Global Communications Conference (GLOBECOM), Taipei, Taiwan, Dec. 2020.

Chapter 10: Concluding Remarks and Future Work

This chapter concludes the thesis and suggests some possible extensions to the current work.

Table 1.1: Summary of the CI-based SLP techniques in the literature.

Reference	Scenario	Closed-form	CIR	Modulation	CSI	Design
[18, 19]	MU-MIMO	No	SPC	PSK	Perfect	Perfect
[34]	"	"	SS	"	(Im)perfect	"
[20, 37]	"	"	NSPC	"	"	"
[21, 36, 94]	"	"	SPC	"	Perfect	"
[40, 41]	"	Yes	NSPC	"	"	"
[48]	"	No	"	"	"	"
[39, 45]	"	"	NSPC+NR	"	"	"
[26, 43]	"	"	SS	QAM	"	"
[27, 28, 41]	"	Yes	NSPC+SS	"	"	"
[46]	"	No	NSPC+NR	"	"	"
[91]	"	"	SPC	APSK	"	"
[35]	"	"	NSPC	PSK+QAM	"	"
Ch. 3: [23, 95, 96]	"	"	DP+UB	Any	"	"
Ch. 4,5: [97–99]	"	Yes	DP	Any	"	"
Ch. 6: [100, 101]	"	No	DP	PSK	Imperfect	"
Ch. 7: [102]	"	"	DP	Any	Perfect	Imperfect
[51–53]	CRN	No	SPC	PSK	Perfect	Perfect
[54, 55]	"	"	NSPC	"	"	"
[57, 59, 60]	SWIPT	No	NSPC	PSK	Perfect	Perfect
[58]	SWIPT + PHY Security	No	NSPC+SS	PSK+QAM	(Im)perfect	Perfect
[61, 62]	PHY Security	No	SPC+NSPC	PSK	Perfect	Perfect
[63, 64]	"	"	NSPC	"	(Im)perfect	"
[65, 67]	Full-Duplex	No	NSPC	PSK	Perfect	Perfect
[65, 67]	"	"	"	"	Imperfect	"
[66]	"	"	NSPC+SS	PSK+QAM	(Im)perfect	"
[68]	DAS + PHY Security	No	NSPC	PSK	Imperfect	Perfect
[70]	Spatio-Temporal CI	No	NSPC	PSK	Perfect	Perfect
[69, 71]	"	"	SS	QAM	"	"
[72–74]	CE	No	NSPC	PSK	Perfect	Perfect
[91–93]	Non-linear	No	SPC	APSK	Perfect	Perfect
[76–78]	Antenna Selection	No	SPC	PSK	Perfect	Perfect
[79]	"	"	NSPC	"	(Im)perfect	"
[86]	Quantized	No	NSPC	PSK	Perfect	Perfect
[84, 89]	"	"	NSPC+SS	"	"	"
[90]	"	"	SS	PSK+QAM	"	"
[85]	"	"	NSPC+SS	"	"	"
Ch. 8: [103, 104]	"	"	DP	Any	"	"
[80, 82, 83]	Hybrid	No	NSPC	PSK	Perfect	Perfect
[105]	"	"	NSPC+SS	PSK+QAM	"	"
Ch. 9: [106, 107]	"	"	DP	Any	"	"
[81]	Quantized + Hybrid	No	SS	PSK	Perfect	Perfect

Chapter 2

Multiuser Precoding in Unicast MU-MIMO Downlink Systems

In this chapter, we explain the problem of multiuser precoding in multi-antenna downlink systems and provide an overview of different design approaches to this problem.

2.1 Preliminaries on MU-MIMO Interference Channels

Consider the downlink of a unicast MU-MIMO system, where independent data streams are intended for multiple users and have to be simultaneously transmitted in the same time-frequency resource block. A well-known technique to perform the downlink transmission in such a system is to exploit the transmitter's multi-antenna structure and spatially multiplex the users' data stream, known as multiuser precoding. Let us confine ourselves to a setup where the transmitter is equipped with an array of N_{t} antennas and communicates with N_{u} single-antenna users, each supporting single-stream transmission. The main functionality of a multiuser precoder is to map N_{u} data symbols onto N_{t} transmit antennas; however, it comes with some other considerations and objectives in order to improve the downlink system's performance, as we will see later.

Let $s_i[n]$ denote the discrete-time data symbol intended for the ith user at symbol period n, where $i \in \{1, 2, ..., N_{\mathrm{u}}\}$ and $\mathbb{E}\{s_i[n]s_i[n]^*\} = 1$. A multiuser precoder may have either a linear or non-linear structure, i.e., the precoder may be a linear function of the users' symbols $\{s_i[n]\}_{i=1}^{N_{\mathrm{u}}}$ or not. Assuming a linear structure, we can express the precoder as an $N_{\mathrm{t}} \times N_{\mathrm{u}}$ matrix, denoted by $\mathbf{W}$, mapping a linear combination of the symbols $\{s_i[n]\}_{i=1}^{N_{\mathrm{u}}}$ onto each transmit antenna. Let the precoding matrix be constructed as $\mathbf{W} \triangleq [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{N_{\mathrm{u}}}]^{\mathrm{T}}$, with $\mathbf{w}_i \in \mathbb{C}^{N_{\mathrm{t}} \times 1}$ denoting the precoding vector for the ith user. In fact, the precoding vector $\mathbf{w}_i$ contains the complex weights of the symbol $\mathbf{s}_i[n]$ on each of N_{t} antennas, and therefore, can be viewed as the intended precoded signal for the ith user. Note that each complex weight modifies both the amplitude and the phase of $\mathbf{s}_i[n]$. The overall precoded vector, describing the complex baseband symbol-sampled

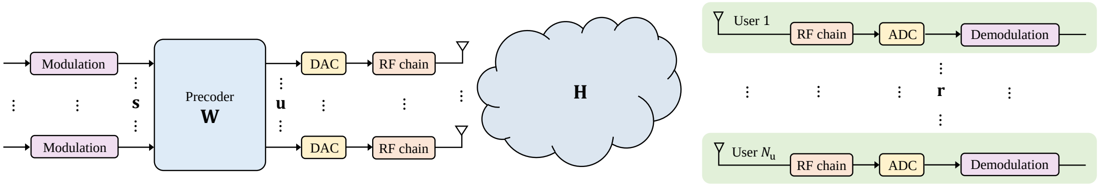


Figure 2.1: Downlink MU-MIMO system with multiuser precoding.

signal for each transmit antenna at symbol period n, is denoted by $\mathbf{u}[n]$ and can be expressed as

$$\mathbf{u}[n] = \mathbf{W}\mathbf{s}[n] = \sum_{i=1}^{N_\mathrm{u}} \mathbf{w}_i s_i[n], \tag{2.1}$$

where $\mathbf{s}[n] \triangleq [s_1[n], s_2[n], ..., s_{N_\mathrm{u}}[n]]^\mathrm{T}$ collects the data symbols for all N_u users. Assume that the precoded signal $\mathbf{u}$ is passed through frequency-flat fading channels towards the users. Let the $1 \times N_\mathrm{t}$ vector $\mathbf{h}_i$ denote the complex coefficients describing the gains and the phases of the propagation channels between N_t transmit antennas and the ith user. Accordingly, the received signal by each user i is given by

$$r_i[n] = \mathbf{h}_i\mathbf{W}\mathbf{s}[n] + z_i[n] = \mathbf{h}_i \sum_{i=1}^{N_\mathrm{u}} \mathbf{w}_i s_i[n] + z_i[n], \quad i = 1, 2, ..., N_\mathrm{u}, \tag{2.2}$$

where $z_i[n] \sim \mathcal{CN}(0, \sigma_i^2)$ represents the additive circularly symmetric complex Gaussian noise at the ith user's receiver. The received signal model (2.2) can be written in a compact form as

$$\mathbf{r}[n] = \mathbf{H}\mathbf{W}\mathbf{s}[n] + \mathbf{z}[n], \tag{2.3}$$

where $\mathbf{z}[n] \triangleq [z_1[n], z_2[n], ..., z_{N_\mathrm{u}}[n]]^\mathrm{T}$, $\mathbf{r}[n] \triangleq [r_1[n], r_2[n], ..., r_{N_\mathrm{u}}[n]]^\mathrm{T}$, and $\mathbf{H} \triangleq [\mathbf{h}_1^\mathrm{T}, \mathbf{h}_2^\mathrm{T}, ..., \mathbf{h}_{N_\mathrm{u}}^\mathrm{T}]^\mathrm{T}$ denotes the channel matrix. The considered multiuser system is depicted in Fig. 2.1. Recall that the transmit signal intended for the ith user is $\mathbf{w}_i s_i[n]$. After down-conversion, matched filtering, and symbol sampling, by decomposing the received signal r_i into different components, we can write

$$r_i[n] = \underbrace{\mathbf{h}_i\mathbf{w}_i s_i[n]}_{\text{desired}} + \underbrace{\mathbf{h}_i \sum_{j\neq i} \mathbf{w}_j s_j[n]}_{\text{MUI}} + z_i[n], \quad i = 1, 2, ..., N_\mathrm{u}, \tag{2.4}$$

in which the MUI term refers to the multiuser interference caused by simultaneous transmission to all the other users than i. From (2.4), it follows that the jth user contributes to the received signal of the user i via the term $\mathbf{h}_i\mathbf{w}_j s_j[n]$, which is in general an undesired contribution. Therefore, the average SINR of the ith user, denoted by SINR_i, can be defined as the ratio between the desired and undesired signal power received by the user i averaged over the symbol time, i.e.,

$$\mathrm{SINR}_i \triangleq \frac{P_{\mathrm{S},i}}{P_{\mathrm{I},i} + P_{\mathrm{N},i}}, \quad i = 1, 2, ..., N_\mathrm{u}, \tag{2.5}$$

where $P_{\mathrm{N},i} = \sigma_i^2$ denotes the noise power, and $P_{\mathrm{S},i}$ and $P_{\mathrm{I},i}$ respectively denote the average desired received signal power and the interference power for the ith user, which are obtained as

$$P_{\mathrm{S},i} \triangleq \mathbb{E}\left\{\mathbf{h}_i\mathbf{w}_i s_i[n] s_i^*[n]\mathbf{w}_i^\mathrm{H}\mathbf{h}_i^\mathrm{H}\right\} = \mathbf{h}_i\mathbf{w}_i \,\mathbb{E}\left\{s_i[n]s_i^*[n]\right\}\, \mathbf{w}_i^\mathrm{H}\mathbf{h}_i^\mathrm{H} = \mathbf{h}_i\mathbf{w}_i\mathbf{w}_i^\mathrm{H}\mathbf{h}_i^\mathrm{H} \tag{2.6}$$

and

$$P_{\mathrm{I},i} \triangleq \mathbb{E}\left\{\left\|\mathbf{h}_i\Big(\sum\nolimits_{j\neq i}\mathbf{w}_j s_j[n]\Big)\right\|^2\right\} = \mathbf{h}_i\Big(\sum\nolimits_{j\neq i}\mathbf{w}_j\mathbb{E}\left\{s_j[n]s_j^*[n]\right\}\mathbf{w}_j^\mathrm{H}\Big)\mathbf{h}_i^\mathrm{H} = \sum_{j\neq i}\mathbf{h}_i\mathbf{w}_j\mathbf{w}_j^\mathrm{H}\mathbf{h}_i^\mathrm{H}, \tag{2.7}$$

where in deriving (2.7), it is assumed that the symbols $\{s_i[n]\}_{i=1}^{N_\mathrm{u}}$ are mutually uncorrelated, i.e., $\mathbb{E}\{s_i[n]s_j^*[n]\} = 0$ for all $i, j \in \{1, 2, ..., N_\mathrm{u}\}, i \neq j$. Accordingly, we have

$$\mathrm{SINR}_i = \frac{\mathbf{h}_i\mathbf{w}_i\mathbf{w}_i^\mathrm{H}\mathbf{h}_i^\mathrm{H}}{\sum_{j\neq i}\mathbf{h}_i\mathbf{w}_j\mathbf{w}_j^\mathrm{H}\mathbf{h}_i^\mathrm{H} + \sigma_i^2}, \quad i = 1, 2, ..., N_\mathrm{u}, \tag{2.8}$$

It can be inferred from (2.8) that at low transmit SNRs, the undesired signal power is dominated by the receiver noise, whereas the MUI becomes dominant in the high SNR regime. In any case, the MUI may degrade the multiuser system's performance if not managed properly. Therefore, another critical functionality of multiuser precoding techniques is interference management. One may attempt to manage the MUI in different ways, e.g., by partially mitigating, eliminating, or converting it to a useful signal component. Some of these approaches are reviewed in the sequel.

2.2 Linear Block-Level Precoding

We begin by reviewing some simple linear block-level precoding schemes. Note that by block-level precoding, we mean the cases in which the design problem considers some statistical performance metrics averaged over a block of symbols, leading to precoding solutions that make use of the channel matrix $\mathbf{H}$ but not the users' data symbols. Therefore, a common assumption in derivation of the following multiuser precoders is that the transmitter perfectly knows the instantaneous channel matrix $\mathbf{H}$. For notational simplicity, we drop the symbol time index n from the formulations presented in the rest of this chapter.

2.2.1 Spatial Matched Filter (MF)

The matched filter (MF) precoding [5], also known as maximum ratio transmission (MRT) [2], aims to maximize the received SNR while being subject to an average transmit power constraint. The corresponding optimization problem can be expressed as

$$\max_{\mathbf{W}} \frac{\left|\mathbb{E}\left\{\mathbf{s}^{\mathrm{H}}\mathbf{r}\right\}\right|^2}{\mathbb{E}\left\{\|\mathbf{z}\|^2\right\}} \quad \text{s.t.} \quad \mathbb{E}\left\{\|\mathbf{W}\mathbf{s}\|^2\right\} = p, \tag{2.9}$$

where p is a fixed average transmit power. The solution to this optimization problem can simply be obtained as

$$\mathbf{W}_{\mathrm{MF}} = \eta_{\mathrm{MF}}\, \mathbf{H}^{\mathrm{H}}, \tag{2.10}$$

where the power normalization factor η_{MF} is given by

$$\eta_{\mathrm{MF}} = \sqrt{\frac{p}{\mathrm{Tr}(\mathbf{H}\mathbf{H}^{\mathrm{H}})}}. \tag{2.11}$$

Note that the noise variance does not appear in (2.11), and therefore, the MF precoder does not take the noise characteristics into account.

2.2.2 Zero-Forcing (ZF)

As the name might suggest, the zero-forcing (ZF) precoder aims to completely cancel the MUI such that $\mathbf{H}\mathbf{W} = \mathbf{I}$ [4]. Accordingly, the precoding matrix is derived by solving the following optimization problem:

$$\min_{\mathbf{W}} \quad \mathbb{E}\left\{\|\mathbf{W}\mathbf{s}\|^2\right\} \quad \text{s.t.} \quad \mathbf{H}\mathbf{W} = \mathbf{I}. \tag{2.12}$$

The solution to (2.12) is given by

$$\mathbf{W} = \mathbf{H}^{\mathrm{H}} \left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{-1}. \tag{2.13}$$

Given $\mathbf{W}$ in (2.13), the resulting transmit power depends only on the channel matrix $\mathbf{H}$, and hence, cannot be controlled. To eliminate such dependence and enforce a fixed transmit power, one usually considers the solution (2.13) up to a scaling factor, i.e.,

$$\mathbf{W}_{\mathrm{ZF}} = \eta_{\mathrm{ZF}}\, \mathbf{H}^{\mathrm{H}} \left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{-1}, \tag{2.14}$$

where

$$\eta_{\mathrm{ZF}} = \sqrt{\frac{p}{\mathrm{Tr}\left((\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\right)}}. \tag{2.15}$$

Using the ZF precoder in (2.14), the received signal vector is equal to

$$\mathbf{r} = \eta_{\mathrm{ZF}}\, \mathbf{H}\mathbf{W}_{\mathrm{ZF}}\mathbf{s} + \mathbf{z} = \eta_{\mathrm{ZF}}\, \mathbf{s} + \mathbf{z}, \tag{2.16}$$

from which we can see that the ZF precoding scheme leads to an interference-free received signal at the receiver of each user. In case the channel matrix $\mathbf{H}$ is ill-conditioned, i.e., the ratio between its maximum and minimum singular value is rather large, the matrix inversion in (2.14) results in a relatively small value for the scaling factor η_{ZF} compared to the noise power; a disadvantage which is referred to as noise enhancement. In general, the ZF precoding scheme is known to be power inefficient. In addition, it is outperformed by the MF precoder at low transmit SNRs.

2.2.3 Regularized Zero-Forcing (RZF)

To resolve the power efficiency issue with the ZF precoder, the regularized inversion method can be applied in deriving the precoding matrix [118]. The resulting precoding scheme is called regularized zero-forcing (RZF) and the corresponding precoding matrix is given by

$$\mathbf{W}_{\mathrm{RZF}} = \eta_{\mathrm{RZF}}\, \mathbf{H}^{\mathrm{H}} \left(\mathbf{H}\mathbf{H}^{\mathrm{H}} + \alpha \mathbf{I} \right)^{-1}, \tag{2.17}$$

where α is referred to as the regularizing factor, and the scaling factor η_{RZF} can be obtained to enforce a fixed average transmit power of p as

$$\eta_{\mathrm{RZF}} = \sqrt{\frac{p}{\mathrm{Tr}\left((\mathbf{H}\mathbf{H}^{\mathrm{H}} + \alpha \mathbf{I})^{-1} \right)}}. \tag{2.18}$$

Finding the optimal value of α depends on the design objective and, in general, is not straightforward. One may choose α to maximize an approximation of the received SINR in the limiting case where $N_{\mathrm{u}} \to \infty$, as in [3], leading to $\alpha = \sigma^2 N_{\mathrm{u}}/p$ (assuming $\sigma_1 = \sigma_2 = ... = \sigma_{N_{\mathrm{u}}} \triangleq \sigma$). Note that using $\mathbf{W}_{\mathrm{RZF}}$ to precode the users' data symbols, the signal received by the ith user is no longer a scaled version of s_i as with the ZF precoding; rather, it may include some non-zero MUI components from the other users.

2.2.4 Wiener Filter (WF)

It is known that the MF and ZF precoding schemes are two extreme designs with one outperforming the other at very low or high transmit SNRs. On the contrary, the Wiener filter (WF) precoding technique [5], also known as the minimum mean square error (MMSE), balances the system performance by taking the noise statistics into account. The WF design problem aims to minimize the variance of the difference between the intended and received symbols of the users, while the transmit power is limited by p, i.e.,

$$\min_{\mathbf{W},\eta} \quad \mathbb{E}\left\{ \|\mathbf{s} - \eta^{-1}\mathbf{r}\|^2 \right\} \quad \text{s.t.} \quad \mathbb{E}\left\{ \|\mathbf{W}\mathbf{s}\|^2 \right\} = p, \tag{2.19}$$

where η is a weighting design variable. The optimization problem (2.19) admits a closed-form solution given by

$$\mathbf{W}_{\mathrm{WF}} = \eta_{\mathrm{WF}} \mathbf{G}^{-1} \mathbf{H}^{\mathrm{H}}, \tag{2.20}$$

where $\mathbf{G} \triangleq \mathbf{H}^{\mathrm{H}}\mathbf{H} + \left(\sigma^2 N_{\mathrm{u}}/p\right)\mathbf{I}$, and the scaling factor η_{WF} is obtained as

$$\eta_{\mathrm{WF}} = \sqrt{\frac{p}{\mathrm{Tr}\left(\mathbf{H}\mathbf{G}^{-2}\mathbf{H}^{\mathrm{H}}\right)}}, \tag{2.21}$$

It follows from (2.20) that at extremely low and high transmit SNRs, i.e., in the limiting cases where $p/\sigma^2 \to 0$ and $p/\sigma^2 \to \infty$, the WF precoder respectively converges to the MF and the ZF precoding schemes. These observations can be verified by applying the matrix inversion lemma; see [119]. Note, however, that in order to design the WF precoder, the transmitter must be aware of the noise properties at the receiver of each user such as the variance σ^2. It is also worth noting that the WF precoder in (2.20) appears to be very similar in structure to the RZF scheme in (2.17). In fact, one may consider the former as a special case of the latter scheme. Recall that the optimal α for the RZF scheme is obtained only for large N_{u}. In this particular case, the WF and RZF schemes become identical; however, it is not the case in general.

2.3 General Families of Objective-Oriented Precoding

When some specific system objective or constraint are given in a multiuser wireless system, more sophisticated objective-oriented precoding schemes become of interest. In such scenarios, the precoder can be designed to optimize the given performance objective, while being constrained to some other system-specific or user-centric restrictions. Starting from the block-level schemes, in this section, we overview some general formulations for such precoding problems.

2.3.1 SINR-Constrained Power Minimization Problem

If the multiuser system requirements can be met via the available resources, one may further attempt to reduce the transmitted power through solving the power minimization problem. This problem is typically constrained by some SINR requirements that are intended to be achieved for the users. Consequently, the problem aims to find a precoding solution that meets the given SINR constraints by consuming as little transmit power as possible. The power minimization design is known to be a relatively straightforward problem with a simple optimal solution structure [7, 120]. The corresponding optimization problem is given by

$$\min_{\mathbf{w}_1,\ldots,\mathbf{w}_{N_{\mathrm{u}}}} \quad \sum_{i=1}^{N_{\mathrm{u}}} \|\mathbf{w}_i\|^2 \quad \text{s.t.} \quad \mathrm{SINR}_i \geq \gamma_i,\ i = 1, 2, \ldots, N_{\mathrm{u}}, \tag{2.22}$$

where γ_i denotes the target SINR for the ith user. The SINR constraints in (2.22) are not convex in the presented form; however, they can be recast as convex constraints as follows. Using (2.8), we can express the constraint $\mathrm{SINR}_i \geq \gamma_i$ as

$$\frac{\mathbf{h}_i \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} \mathbf{h}_i^{\mathrm{H}}}{\sum_{j \neq i} \mathbf{h}_i \mathbf{w}_j \mathbf{w}_j^{\mathrm{H}} \mathbf{h}_i^{\mathrm{H}} + \sigma_i^2} \geq \gamma_i. \tag{2.23}$$

After some straightforward algebraic steps, one can rewrite (2.23) as

$$\mathbf{h}_i \left(\sum_{j \neq i} \mathbf{w}_j \mathbf{w}_j^{\mathrm{H}} - \frac{1}{\gamma_i} \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} \right) \mathbf{h}_i^{\mathrm{H}} + \sigma_i^2 \leq 0, \tag{2.24}$$

which is a convex second-order cone (SOC) constraint. Having a quadratic objective function, the optimization problem (2.22) can then be expressed in a convex form and solved using off-the-shelf convex optimization algorithms [121].

2.3.2 Power-Constrained Precoding Design

The power-constrained precoding design problem becomes relevant in the case where the transmit power is a strict restriction in the system. The design problem aims to maximize some performance metric that is, typically, a function of the users' target SINRs, while the total transmit power is constrained by p. In general, the corresponding optimization problem can be expressed as

$$\max_{\mathbf{w}_1, \ldots, \mathbf{w}_{N_\mathrm{u}}} \quad f\Big(\mathrm{SINR}_1, \mathrm{SINR}_2, ..., \mathrm{SINR}_{N_\mathrm{u}}\Big) \quad \text{s.t.} \quad \sum_{i=1}^{N_\mathrm{u}} \|\mathbf{w}_i\|^2 \leq p, \tag{2.25}$$

where $f(\cdot)$ is strictly increasing in SINR_i for any $i \in \{1, 2, ..., N_\mathrm{u}\}$. Unlike the power minimization problem, the power-constrained design problems in the form of (2.25) are known to be difficult to solve, or even NP-hard for some common choices of the objective function $f(\cdot)$, e.g., the sum-rate function given as

$$f\Big(\mathrm{SINR}_1, \mathrm{SINR}_2, ..., \mathrm{SINR}_{N_\mathrm{u}}\Big) = \sum_{i=1}^{N_\mathrm{u}} \log_2(1 + \mathrm{SINR}_i).$$

Another common objective function is based on the max-min fair criterion and is expressed as

$$f\Big(\mathrm{SINR}_1, \mathrm{SINR}_2, ..., \mathrm{SINR}_{N_\mathrm{u}}\Big) = \min_i \; \{\mathrm{SINR}_i\}_{i=1}^{N_\mathrm{u}},$$

which aims to maximize the minimum achievable SINR among all the users. The resulting design formulation is often called the SINR balancing problem. It is worth noting that, for a given fixed transmit power, at low transmit SNRs where the noise dominates the system, the solution to the SINR balancing problem approaches the MF precoding scheme to maximize the desired received signal power at the users. On the other hand, in the high transmit SNR regime where the system is interference-limited, the solution tends to that of the ZF precoding to cancel the MUI. At an arbitrary SNR value, the

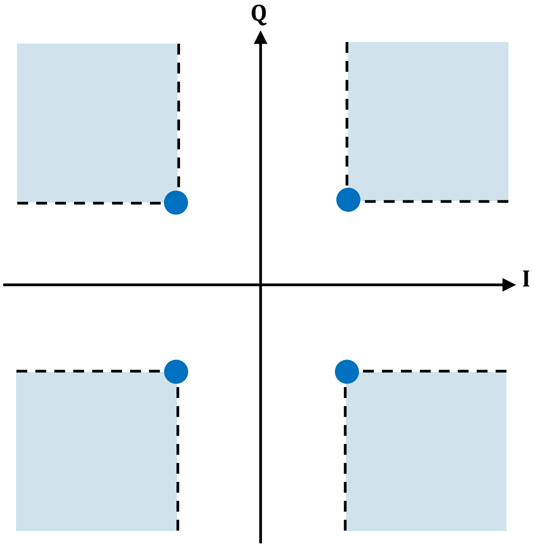


Figure 2.2: Typical CIRs for the QPSK constellation.

optimal max-min SINR precoding finds a balance between these two extreme solutions.

2.4 Symbol-Level Precoding

The block-level precoding schemes treat the MUI as an undesired received signal component and aim to suppress it in an efficient way. In contrast, in designing the precoder, one may attempt to manipulate the MUI such that it constructively contributes to the desired signal of each user, i.e., by exploiting the constructive interference (CI). Such a design approach falls within another category of precoding design strategies, known as symbol-level precoding (SLP), which is the main focus of this thesis. The term "symbol-level" refers to the fact that in order to exploit the CI, one needs to design the precoder particularly for every set of the users' symbols, i.e., each realization of the symbol vector $\mathbf{s}$. The CI is defined based on the philosophy that a noise-free received signal can be decoded correctly not necessarily when it is close enough to the intended symbol, rather, as long as it lies within the correct decision region even far away from the target symbol. Accordingly, the CI regions (CIR) are typically defined as those regions that satisfy this philosophy. An example illustration of CIRs is shown in Fig. 2.2.

The block-level schemes use statistical objectives and constraints in the design problem of the precoder. In practical systems, these statistical measures can be realized over sufficiently many symbol periods, e.g., a large enough block of symbols, which is usually the case in practice. For instance, the assumption $\mathbb{E}\{s_i s_i^*\} = 1$ is often used in simplifying the optimization problems, e.g., in defining the total transmit power as $\sum_{i=1}^{N_\mathrm{u}} \|\mathbf{w}_i\|^2$. As a result, the precoding matrix turns out to be only a function of the channel matrix $\mathbf{H}$ and not the symbol vector $\mathbf{s}$. In contrast, the objectives and constraints in the optimization problem of an SLP scheme are of an instantaneous per-symbol type. For example, one may consider the instantaneous transmit power or received SINR in defining the design problem. Moreover, the statistical assumptions on the users' symbols are no loner useful, e.g., $\mathbb{E}\{s_i s_i^*\} = 1$ or the assumption of having uncorrelated symbols. As a result, one may expect the symbol vector $\mathbf{s}$ to appear in the precoding solution. Therefore, due to the dependence of the precoder's output on the instantaneous users' symbols, a symbol-level precoder is a function of both $\mathbf{H}$ and $\mathbf{s}$.

As a consequence of exploiting the CI at the receiver of each user, the users' received signals can be decomposed as

$$r_i = \overbrace{\mathbf{h}_i\mathbf{w}_i s_i + \underbrace{\mathbf{h}_i \sum_{j\neq i} \mathbf{w}_j s_j}_{\text{MUI}}}^{\text{desired}} + z_i, \quad i = 1, 2, ..., N_{\text{u}}. \tag{2.26}$$

The decomposition in (2.26) holds if an essential constraint is included in the SLP design problem, namely, the CI constraint. Accordingly, the precoding vectors $\{\mathbf{w}_i\}_{i=1}^{N_{\text{u}}}$ has to be designed so that the desired signal component in (2.26) lies within the CIR that corresponds to the symbol s_i, for all $i = 1, 2, ..., N_{\text{u}}$, e.g., the blue regions in Fig. 2.2 for the QPSK symbols. In this case, the instantaneous received SINRs of the users can be redefined as

$$\text{SINR}_i = \frac{\mathbf{h}_i\mathbf{W}\mathbf{s}\mathbf{s}^{\text{H}}\mathbf{W}^{\text{H}}\mathbf{h}_i^{\text{H}}}{\sigma_i^2}, \quad i = 1, 2, ..., N_{\text{u}}, \tag{2.27}$$

In symbol-level precoded downlink transmission, none of the users will experience destructive interference from the other users. As a result, the term SINR translates to signal-plus-interference-to-noise ratio, and thus, may be considered equivalent to the conventional SNR in an interference-free system.

Broadly speaking, the criteria used to design a block-level precoding scheme can also be used to formulate an SLP design problem. Accordingly, the optimization problem for an SINR-constrained power minimization SLP design can be written as

$$\min_{\mathbf{W}} \quad \|\mathbf{W}\mathbf{s}\|^2 \quad \text{s.t.} \quad \text{CI constraints}, \tag{2.28}$$

where the CI constraints are typically given as $\eta\, \mathbf{h}_i\mathbf{W}\mathbf{s} \in \mathcal{D}_i$ with $\mathcal{D}_i$ denoting a particular CIR, and the scaling factor η is in general a function of the noise variance σ_i^2 and the target SINR γ_i. Moreover, the power-constrained SLP optimization problem can be expressed as

$$\min_{\mathbf{W}} \quad f\Big(\text{SINR}_1, \text{SINR}_2, ..., \text{SINR}_{N_{\text{u}}}\Big) \quad \text{s.t.} \quad \|\mathbf{W}\mathbf{s}\|^2 \leq p, \quad \text{CI constraints}. \tag{2.29}$$

One may also cast the SLP problem by forming a virtual multicast formulation to directly design the precoded transmit signal $\mathbf{u}$ instead of calculating the precoding matrix $\mathbf{W}$, leading to a non-linear structure for the precoder. In this case, for example, the power minimization SLP problem can be rewritten as

$$\min_{\mathbf{u}} \quad \|\mathbf{u}\|^2 \quad \text{s.t.} \quad \text{CI constraints}, \tag{2.30}$$

The precoded signal **u** obtained by solving the problem (2.30) can not be uniquely decomposed as a linear combination of the precoding vectors. However, using the relation $\mathbf{u} = \mathbf{W}\mathbf{s}$, one can obtain a rank-one (not necessarily unique) precoding matrix as

$$\mathbf{W} = \left(\mathbf{s}^{\mathrm{T}} \otimes \mathbf{I}_{N_{\mathrm{t}}}\right) \mathbf{u}, \tag{2.31}$$

In general, it can be shown that the two design formulations in (2.28) and (2.30) lead to identical solutions for the symbol-level precoder [20, 21]. In this thesis, we mainly use the virtual multicast formulation to design the precoder, i.e., we directly optimize the precoded vector **u** as a function of the channel matrix **H**, the users' symbols **s**, and the other system parameters.

Chapter 3

A Generic Design Framework for Constructive Interference Based Symbol-Level Precoding

In this chapter, we study the problem of SLP design in the downlink of an MU-MIMO channel. We first consider generic modulation schemes with constellation sets of any shape and size and elaborate on optimal and relaxed CIRs. We define two types of CIRs, namely, distance-preserving CIR (DPCIR) and union bound CIR (UBCIR) and provide a systematic way to describe these regions as convex sets. We then confine ourselves to DPCIRs and perform a comprehensive study which allows us to derive several properties for these regions. Using these properties, we first show that any signal in a given DPCIR has a norm greater than or equal to the norm of the corresponding constellation point if and only if the convex hull of the constellation contains the origin. It is followed by proving that the power of the noise-free received signal in a DPCIR is a monotonic strictly increasing function of two parameters relating to the infinite Voronoi edges. Using the convex representations of DPCIRs and UBCIRs, we formulate two design problems, namely, the SLP power minimization with SINR constraints, and the SLP SINR balancing problem under max-min fairness criterion. We show that the SLP power minimization problem, minimizing either sum or peak (per-antenna) transmit power, can always be formulated as a convex quadratic programming (QP). We further derive a simplified reformulation of this problem which is more computationally efficient. Our simulation results indicate that the DPCIRs and UBCIRs allow further reduction of the transmit power compared to the state-of-the-art without increasing the computational complexity at the transmitter or receiver. The SLP max-min SINR problem, on the other hand, is non-convex in its original form, and hence is difficult to tackle. We propose alternative optimization approaches, including semidefinite programming (SDP) formulation and block coordinate descent (BCD) optimization. We finally discuss and evaluate the loss due to the proposed alternative methods through extensive simulation results.

3.1 Introduction

The symbol-level design of a multiuser precoder generally involves an optimization problem for each possible combination of the users' data symbols. The optimization constraints are so defined to push each user's (noise-free) received signal into the corresponding predefined CIR, enhancing (or guaranteeing a certain level of) the users' symbol detection accuracy. Therefore, formulation of the optimization problem, and particularly the constraints, depend on the adopted modulation scheme (i.e. signal constellation). The objective function, on the other hand, depends on the design criterion.

In general, objective-oriented multiuser precoding design aims at keeping a balance between some system-centric and user-centric objectives/requirements, depending on the network's operator strategy [6, 7]. Power and sum-rate are often regarded as system-centric quantities [9]. Transmit power is considered, for example, to control the inter-cell interference in multi-cell wireless networks, and sum-rate is a measure of the overall system performance. On the other hand, as a user-centric criterion, SINR is an effective QoS measure in multiuser interference channels [11]. In particular, both BER and capacity, which are two relevant criteria from a practical point of view, are closely related with maximizing SINR [10]. Considering different types of optimization criteria, some well-known formulations for the multiuser precoding problem are QoS-constrained power minimization [12, 13], SINR balancing [8, 10, 14], and (weighted) sum-rate maximization [9, 15, 16]. In this work, we primarily focus on the power minimization problem with SINR constraints and the SINR balancing problem using max-min fair criterion.

The SLP problem minimizing the total transmit power has been studied for various constellations, including PSK [20, 21, 36–38, 61, 94, 108–110], QAM [26–28, 41–43, 105], and APSK [70]. For PSK constellations, the minimization of peak per-antenna transmit power is addressed in [94].

SINR balancing in MU-MIMO systems is generally more challenging and has been widely investigated for conventional precoding schemes; see, e.g., [8, 10, 11, 122, 123]. This problem has been addressed in both multicast (single data stream) and unicast (multiple independent data streams) downlink scenarios. The problem is not convex in general and is known to be NP-hard [11]. To address this difficulty, several alternative optimization approaches have been proposed in the literature. In particular, for downlink unicast channels, it is shown in [10] that the power minimization and the max-min SINR are inverse problems. We kindly refer the readers to [96] for a short review on SINR balancing in conventional multiuser precoding.

The SINR balancing problem for SLP schemes has not been addressed extensively in the literature. In [21], the non-convex SLP max-min SINR is solved using its relation to the power minimization via a bisection search. The method is only applicable to PSK constellations (more precisely, to constant envelope modulations) and suffers from high computational complexity. This problem is also addressed in [20] and a second-order cone programming (SOCP) formulation is proposed for PSK constellations. Nevertheless, there is no general solution method or convex formulation for the SLP max-min SINR problem being valid for all generic (two-dimensional) constellation sets.

In this work, we first study optimal and relaxed CIRs for a generic constellation which leads us to introducing several types of CIRs, such as distance-preserving CIR (DPCIR) and union bound CIR (UBCIR). We specifically focus on DPCIRs and fully characterize their geometry based on the Voronoi regions of the constellation set. We also prove several properties for these regions which will subsequently be used to address the SLP design problems. The main property states that the norm of any signal in a given unbounded DPCIR is a monotonically increasing function of two parameters related to the corresponding infinite Voronoi edges, under the necessary and sufficient condition that the convex hull of the constellation contains the origin. We address both the power minimization and the SINR balancing problems in SLP. We show that the power minimization SLP problem can always be formulated as a convex optimization problem defined on the DPCIRs and UBCIRs. We provide explicit convex formulations for the SLP power minimization problem and compare their performances and computational complexities with the state-of-the-art techniques, where it will be shown that the proposed formulations can provide performance gains in terms of transmit power compared to the existing schemes, with relatively reduced complexities. We then study the SLP design criterion from a system-level point of view and discuss the feasibility of QoS provisioning in a resource-constrained multiuser downlink channel through deriving a sufficient feasibility condition. Moreover, using the properties of DPCIRs, we show that by fixing a subset of variables in the optimization problem, the SLP max-min SINR can be treated as a convex problem. Accordingly, we propose more tractable alternative optimization approaches, which result in competitive sub-optimal solutions for the original problem. Two methods are proposed and evaluated, namely, semidefinite programming (SDP) formulation and block coordinate descent (BCD) optimization. In summary, the main contributions of this chapter are as follows:

1. Considering generic modulation schemes with constellation sets of any shape and size, we define DPCIRs and UBCIRs as, respectively, optimal and relaxed CIRs.

2. We provide a systematic way to describe the DPCIRs and UBCIRs as convex sets and show that the SLP power minimization problem, minimizing either sum or peak (per-antenna) transmit power, can always be formulated as a quadratic programming (QP) defined on these regions.

3. We derive several properties for DPCIRs in order to further improve the SLP techniques and simplify the involved optimization problems.

4. We obtain a simple feasibility condition for the SLP power minimization problem, which is of practical importance in a realistic scenario as it determines whether the power minimization problem is applicable or not.

5. By rearranging the DPCIR-based SLP power minimization, we obtain an equivalent formulation with a reduced problem size.

6. We propose alternative sub-optimal solutions for the SLP max-min SINR problem enhancing the performance of the multiuser system in terms of the worst-user SINR

compared to the existing schemes, while being less computationally complex.

7. All the definitions and optimization problems are provided in a general form for constellation sets which are indifferent to both the shape and the order of constellation.

The rest of this chapter is organized as follows. In Section 3.2, we describe the system model and define the design problems of interest. In Section 3.3, optimal and relaxed CIRs are introduced and characterized for a generic constellation set. This is followed by proving various properties for these regions in Section 3.4. We address the SLP design problems in Section 3.5. Discussions on the power minimization problem and the proposed alternatives for the SINR balancing design are also included in this section. In Section 3.6, we present the simulation results. Finally, we conclude the chapter in Section 3.7.

3.2 System Model and Problem Definition

We consider the downlink of an MU-MIMO unicast channel, where a single base station (BS) sends independent data streams to N_{u} users in the same time-frequency resource block. The BS is equipped with N_{t} transmit antennas while each user has a single receive antenna. The number of simultaneously served users N_{u} is limited by the number of BS's antennas, i.e., $N_{\mathrm{u}} \leq N_{\mathrm{t}}$. A frequency-flat block-fading channel is assumed between the BS's transmit antennas and the ith user, where the complex channel vector is denoted by $\mathbf{h}_i \in \mathbb{C}^{1\times N_{\mathrm{t}}}$. It is further assumed that perfect channel knowledge is available at the BS, and that $\mathbb{E}\{\mathbf{h}_i^{\mathrm{H}}\mathbf{h}_j\} = \mathbf{0}$ for all $i, j = 1, ..., N_{\mathrm{u}}$ and $i \neq j$. At a given symbol period, independent data symbols $\{s_i\}_{i=1}^{N_{\mathrm{u}}}$ are intended to be transmitted to N_{u} users (throughout this chapter, we drop the symbol's period index to simplify the notations), where s_i denotes the discrete-time target symbol for the ith user. Each symbol s_i is drawn from a finite equiprobable two-dimensional constellation set. Without loss of generality, for all the users, we assume an M-ary constellation set $\mathcal{X} = \{x_m | x_m \in \mathbb{C}\}_{m=1}^{M}$ with unit average power, i.e., $(1/M)\sum_{m=1}^{M} |x_m|^2 = 1$. The user's symbol vector $\mathbf{s}$ is mapped to N_{t} transmit antennas. This is done by a symbol-level precoder, yielding the transmit vector $\mathbf{u} = [u_1, \ldots, u_{N_{\mathrm{t}}}]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{t}}\times 1}$, which implicitly contains the data symbols $\{s_i\}_{i=1}^{N_{\mathrm{u}}}$, as depicted in Fig. 3.1. Considering a complex baseband symbol-sampled model, under the above assumptions, the ith user's received signal is given by

$$r_i = \mathbf{h}_i\mathbf{u} + z_i, \quad i = 1, 2, ..., N_{\mathrm{u}}, \tag{3.1}$$

where $z_i \sim \mathcal{CN}(0, \sigma_i^2)$ denotes the complex additive white Gaussian noise (AWGN) at the ith receiver. From the received scalar r_i, the user i may detect its own symbol s_i by applying the single-user maximum-likelihood (ML) decision rule. Notice that the structure of the users' receivers is not affected by employing the symbol-level precoder at the transmitter.

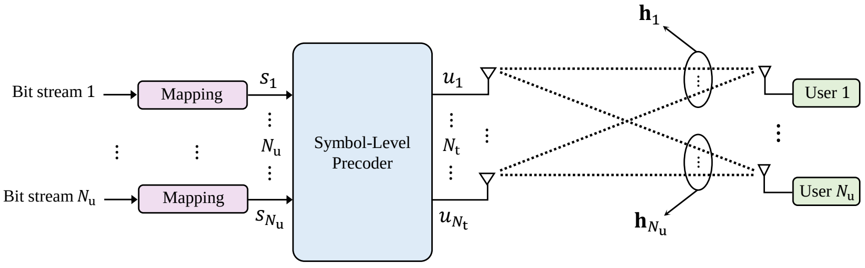


FIGURE 3.1: A simplified block diagram for the downlink MU-MIMO channel with SLP.

3.2.1 Interpretation of Symbol-Level SINR Constraints

The functionality of symbol-level precoder is to instantaneously design the transmit signal for each symbol period based on a CI-constrained optimization problem. The solution of this problem, i.e., the precoded vector $\mathbf{u}$, is in general a function of instantaneous data information (DI) and channel state information (CSI) as well as a set of given system constraints or user-specific requirements.

In a downlink MU-MIMO system, the convention is to define the SINR of each user as the ratio between the desired received signal power and the power of interfering components (due to multiplexing the users' data streams) plus noise power. On the other hand, the SLP design generally aims at forcing all the received signal components to constructively interfere at the receiver of each user. This can be interpreted as having no destructive interference at none of the receivers, i.e., SINR turns into signal-to-noise ratio (SNR) with CI contributing to the desired signal power. Therefore, in the context of SLP, SINR translates to signal-plus-interference-to-noise ratio, and hence, is equivalent to the conventional SNR. Nevertheless, in the rest of this chapter, we continue to use "SINR", as it has been commonly used in this context. In a formal way, it follows from (3.1) that the instantaneous received SINR of the ith user at a given symbol period is equal to

$$\mathrm{SINR}_i = \frac{\mathbf{u}^\mathrm{H}\mathbf{h}_i^\mathrm{H}\mathbf{h}_i\mathbf{u}}{\sigma_i^2}. \tag{3.2}$$

The user-specific requirements in a multiuser system are individual target SINRs that guarantee the reliable communication for all the users. It should, however, be noted that the given target SINRs typically refer to long-term (e.g., block-level) SINRs, i.e., the average received SINR over a block of symbols. Therefore, based on the instantaneous SINRs in (3.2), the following average SINR constraints has to be imposed on the design problem:

$$\mathbb{E}\{\mathrm{SINR}_i\} \geq \gamma_i,\ i = 1, 2, ..., N_\mathrm{u}, \tag{3.3}$$

where γ_i is the required SINR for the ith user, and the expectation is taken with respect to the symbol time over the entire block. Note that while the time index is dropped for simplicity of notation, the precoded vector $\mathbf{u}$ is a function of the symbol time. By substituting (3.2) for the instantaneous SINRs, the inequality (3.3) is equivalent to

$$\mathbb{E}\{\mathbf{u}^{\mathrm{H}}\mathbf{h}_i^{\mathrm{H}}\mathbf{h}_i\mathbf{u}\} \geq \sigma_i^2\gamma_i, \quad i = 1, 2, ..., N_{\mathrm{u}}. \tag{3.4}$$

For sufficiently large blocks (which is often the case in practice), we have $\mathbb{E}\{s_i s_i^*\} \rightarrow 1$ for all $i = 1, ..., N_{\mathrm{u}}$. Hence, it is sufficient for the block-level SINR constraints in (3.4) to be met if

$$\mathbf{u}^{\mathrm{H}}\mathbf{h}_i^{\mathrm{H}}\mathbf{h}_i\mathbf{u} \geq \sigma_i^2\gamma_i\, s_i s_i^*, \quad i = 1, 2, ..., N_{\mathrm{u}}, \tag{3.5}$$

which are referred to as symbol-level SINR constraints. One may think of these symbol-level constraints in (3.5) as a conservative way to meet the block-level SINR requirements in (3.3).

3.2.2 Definition of the SLP Design Problem with CI Constraints

In SLP design, the DI is exploited by optimizing the precoded transmit vector such that the noise-free received signal of each user is located in a predefined CIR that corresponds to the user's intended symbol. The CIRs are typically defined so that they preserve or even enhance the users' symbol detection accuracy compared to the original constellation set; see, e.g., [22] and [23].

For each user i, the noise-free received signal, i.e., $\mathbf{h}_i\mathbf{u}$, is pushed by the precoder into the corresponding CIR up to a scaling factor that depends on the given SINR requirement. Accordingly, for a generic constellation, the CI-constrained SLP power minimization problem with individual user-specific SINR constraints can be formulated as

$$\begin{aligned} \min_{\mathbf{u}} \quad & f(\mathbf{u}) \\ \text{s.t.} \quad & \mathbf{h}_i\mathbf{u} \in \sigma_i\sqrt{\gamma_i}\,\mathcal{D}_i,\ i = 1, 2, ..., N_{\mathrm{u}}, \end{aligned} \tag{3.6}$$

where $\mathcal{D}_i$ denotes the CIR associated with symbol s_i which is typically defined in a way that it pushes $\mathbf{h}_i\mathbf{u}$ away from the corresponding decision boundaries (i.e., pushes $\mathbf{h}_i\mathbf{u}$ deeper into the decision region of s_i). An explicit definition for $\mathcal{D}_i$, in general, depends on the type of CIR and will be provided in the next section. The objective function $f(\mathbf{u})$ in (3.6) can be either $\mathbf{u}^{\mathrm{H}}\mathbf{u}$ or $\|\mathbf{u}\|_\infty^2$, depending on whether the total or the peak (per-antenna) transmit power is minimized. It is important to note that a sufficient (but not necessary) condition under which the optimal solution of (3.6) satisfies the SINR constraints in (3.5) is that the amplitude of any point in $\mathcal{D}_i$ is at least equal to $|s_i| = \sqrt{s_i s_i^*}$, for all $i = 1, ..., N_{\mathrm{u}}$, i.e.,

$$\sigma_i^2\gamma_i\, x x^* \geq \sigma_i^2\gamma_i\, s_i s_i^*, \quad \forall x \in \mathcal{D}_i. \tag{3.7}$$

The SLP SINR balancing problem, on the other hand, aims to serve all the users in a fair manner under a given system-centric restriction, which is usually the total transmit power. In particular, with the max-min fair criterion, the goal is to maximize the worst SINR among all the users subject to a total power constraint. This leads to the following formulation:

$$\begin{aligned} \max_{\mathbf{u}} \quad & \min_{i} \left\{ \frac{\mathbf{u}^{\mathrm{H}}\mathbf{h}_i^{\mathrm{H}}\mathbf{h}_i\mathbf{u}}{\sigma_i^2} \right\}_{i=1}^{N_{\mathrm{u}}} \\ \text{s.t.} \quad & \mathbf{h}_i\mathbf{u} \in \sigma_i \mathcal{D}_i, \; i = 1, 2, ..., N_{\mathrm{u}}, \\ & \mathbf{u}^{\mathrm{H}}\mathbf{u} \leq p, \end{aligned} \tag{3.8}$$

where p denotes the power budget. It should be noted that, in practice, the value of p may be given as the average (over a block of symbols) available power for the downlink transmission, while the power constraint in (3.8) controls the instantaneous total transmit power in each symbol period. This is a sufficient constraint to meet the average power budget, but clearly it is not necessary and has been considered in order to simplify the problem.

We will reformulate and discuss both problems (3.6) and (3.8) in Section 3.5, using explicit mathematical representations for the CI constraints. To this end, in the next section, we present a detailed study of the CIRs to obtain such mathematical representations, and further, to exploit their properties in order to properly form the constraints of the SLP problems.

3.3 Constructive Interference Regions

In this section, we define several types of CIRs and describe them in a systematic way based on the ML decision regions of the constellation $\mathcal{X}$. Hereafter, we denote each complex-valued constellation point by its equivalent real-valued vector form, and thus the set of symbols in $\mathcal{X}$ is denoted by $\{\mathbf{x}_m | \mathbf{x}_m \in \mathbb{R}^2\}_{m=1}^{M}$ where $\mathbf{x}_m = [\mathrm{Re}(x_m), \mathrm{Im}(x_m)]^{\mathrm{T}}$ for all $m = 1, 2, ..., M$.

For the assumed equiprobable constellation set $\mathcal{X}$, the ML decision rule for the constellation set $\mathcal{X}$ has a geometric interpretation; it corresponds to the Voronoi regions of $\mathcal{X}$ which are bounded by hyperplanes. Assuming a given constellation point $\mathbf{x}_m$ and one of its neighboring points $\mathbf{x}_k$ (the neighboring points are referred to those points that share an ML decision boundary with $\mathbf{x}_m$), the hyperplane separating the Voronoi region of $\mathbf{x}_m$ from that of $\mathbf{x}_k$ is given by $\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{a}_{m,k}^{\mathrm{T}}\mathbf{x} = b_{m,k}^{(\mathrm{ML})}\}$, where $\mathbf{a}_{m,k} \in \mathbb{R}^2$, $\mathbf{a}_{m,k} \neq \mathbf{0}$, and $b_{m,k}^{(\mathrm{ML})} \in \mathbb{R}$. This hyperplane represents the ML decision boundary (Voronoi edge) between $\mathbf{x}_m$ and $\mathbf{x}_k$, which splits $\mathbb{R}^2$ into two halfspaces (note that hyperplanes are infinite lines in $\mathbb{R}^2$). The halfspace that extends towards $\mathbf{x}_m$, and thus, contains the decision region of $\mathbf{x}_m$ is represented as

$$\mathcal{H}_{m,k}^{(\mathrm{ML})} = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{a}_{m,k}^{\mathrm{T}}\mathbf{x} \geq b_{m,k}^{(\mathrm{ML})}\}, \tag{3.9}$$

where $\mathbf{a}_{m,k}$ is the inward normal and $b_{m,k}^{(\mathrm{ML})}$ determines the offset from the origin. The ML decision region (Voronoi region) of $\mathbf{x}_m$ is then given by intersecting the all halfspaces in the form of (3.9) over the neighboring points of $\mathbf{x}_m$, i.e.,

$$\mathcal{D}_m^{(\mathrm{ML})} = \bigcap_{k \in \mathcal{J}_m} \mathcal{H}_{m,k}^{(\mathrm{ML})} = \left\{ \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{a}_{m,k}^{\mathrm{T}} \mathbf{x} \geq b_{m,k}^{(\mathrm{ML})}, i \in \mathcal{J}_m \right\}, \tag{3.10}$$

where $\mathcal{J}_m = \{k | \mathbf{x}_k \in \mathcal{S}_m\}$ and $\mathcal{S}_m$ denotes the set of neighboring points of $\mathbf{x}_m$, i.e., the set of points having a common decision boundary with $\mathbf{x}_m$, with $|\mathcal{S}_m| = |\mathcal{J}_m| = M_m$. Depending on the relative geometry of $\mathbf{x}_m$ in $\mathcal{X}$, the Voronoi region (3.10) can be either an unbounded polyhedron, if $\mathbf{x}_m$ is an outer constellation point, or a bounded polyhedron, if $\mathbf{x}_m$ is an inner point. We will elaborate on this aspect in more detail in the next sections. In any case, it can be easily shown that a Voronoi region is always a convex set [121]. The Voronoi region (3.10) can be expressed in a more compact form as

$$\mathcal{D}_m^{(\mathrm{ML})} = \left\{ \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{A}_m \mathbf{x} \succeq \mathbf{b}_m^{(\mathrm{ML})} \right\}, \tag{3.11}$$

where $\succeq$ denotes elementwise inequality, and $\mathbf{A}_m$ and $\mathbf{b}_m$ respectively contain $\mathbf{a}_{m,k}^{\mathrm{T}}$ and $b_{m,k}^{(\mathrm{ML})}$ for all $k \in \mathcal{J}_m$, i.e.,

$$\mathbf{A}_m = \begin{bmatrix} \mathbf{a}_{m,1}^{\mathrm{T}} \\ \vdots \\ \mathbf{a}_{m,M_m}^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{M_m \times 2}, \quad \mathbf{b}_m^{(\mathrm{ML})} = \begin{bmatrix} b_{m,1}^{(\mathrm{ML})} \\ \vdots \\ b_{m,M_m}^{(\mathrm{ML})} \end{bmatrix} \in \mathbb{R}^{M_m}. \tag{3.12}$$

Each normal vector $\mathbf{a}_{m,k}$ in (3.12) is orthogonal to the Voronoi edge shared by $\mathbf{x}_m$ and $\mathbf{x}_k$, and thus, it can be obtained as $\mathbf{a}_{m,k} = \mathbf{x}_m - \mathbf{x}_k$ (or any non-zero scalar multiplication of $\mathbf{x}_m - \mathbf{x}_k$). Furthermore, this Voronoi edge passes through the point $(\mathbf{x}_m + \mathbf{x}_k)/2$, and therefore according to [121, p. 27], the corresponding offset $b_{m,k}^{(\mathrm{ML})}$ in (3.10) can be obtained by simple vector algebra as

$$b_{m,k}^{(\mathrm{ML})} = \frac{1}{2} \mathbf{a}_{m,k}^{\mathrm{T}} (\mathbf{x}_m + \mathbf{x}_k). \tag{3.13}$$

Note that $b_{m,k}^{(\mathrm{ML})}$ is found such that the orthogonal distance between $\mathbf{x}_m$ and the corresponding Voronoi edge is equal to half of the distance between $\mathbf{x}_m$ and $\mathbf{x}_k$. By changing $b_{m,k}$ to $b_{m,k}^{(\mathrm{ML})} + \delta$, where $\delta \geq 0$, we get a new hyperplane displaced by

$$\Delta = \frac{\delta}{\|\mathbf{a}_{m,k}\|}, \tag{3.14}$$

in the direction of $\mathbf{a}_{m,k}$ such that it is parallel to the original hyperplane. As an example, in Table 3.1, we show the normal vector corresponding to a symbol $\mathbf{x}_m$ taken from a QPSK constellation. Note that the normal vectors given in Table 3.1 are normalized such that they have a unit Euclidean norm.

According to the definition of CI [20,21], the CIR of $\mathbf{x}_m$ should be a subset of $\mathcal{D}_{m,\mathrm{ML}}$. In this work, we propose a construction method such that each CIR is obtained by

TABLE 3.1: Normal vectors corresponding to QPSK symbols.

$\mathbf{x}_m$	$\mathbf{a}_{m,1}^{\mathrm{T}}$	$\mathbf{a}_{m,2}^{\mathrm{T}}$
$0.7071 + \mathrm{j}0.7071$	$[+1, 0]$	$[0, +1]$
$-0.7071 + \mathrm{j}0.7071$	$[-1, 0]$	$[0, +1]$
$-0.7071 - \mathrm{j}0.7071$	$[-1, 0]$	$[0, -1]$
$0.7071 - \mathrm{j}0.7071$	$[+1, 0]$	$[0, -1]$

displacement of the hyperplanes contributing to $\mathcal{D}_{m,\mathrm{ML}}$. The displacement value δ must be chosen carefully as it determines the margins from the Voronoi decision boundaries and thus affect the symbol error probability (SEP). It is clear that for a fixed signal-to-noise ratio (SNR), reducing the margins would result in a higher SEP. On the other hand, from (3.6) it is inferred that for a given target SNR, having narrower margins provides a larger feasible region for $\mathbf{u}$ and possibly results in a lower transmit power.

3.3.1 Distance Preserving Constructive Interference Regions

We call a CIR distance-preserving (DPCIR) if it does not decrease the original distances between the constellation points. As a consequence, the achievable SEP will be always lower than that of the original constellation. Let $d_{m,k} = \|\mathbf{x}_m - \mathbf{x}_k\|$ denote the Euclidean distance between the points $\mathbf{x}_m$ and $\mathbf{x}_k$ from which the distance-preserving margin is equal to $d_{m,k}/2$. Then, the value of δ can simply be obtained from (3.14) by substituting $\Delta = d_{m,k}/2$. Accordingly, the distance-preserving halfspaces corresponding to $\mathbf{x}_m$ can be expressed as

$$\mathcal{H}_{m,k}^{(\mathrm{DP})} = \left\{ \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{a}_{m,k}^{\mathrm{T}} \mathbf{x} \geq b_{m,k}^{(\mathrm{ML})} + b_{m,k}^{(\mathrm{DP})} \right\}, \quad k \in \mathcal{J}_m, \tag{3.15}$$

where $b_{m,k}^{(\mathrm{DP})} = d_{m,k} \|\mathbf{a}_{m,k}\|/2$. Intersecting (3.15) over all the neighboring points of $\mathbf{x}_m$ yields

$$\mathcal{D}_m^{(\mathrm{DP})} = \bigcap_{k \in \mathcal{J}_m} \mathcal{H}_{m,k}^{(\mathrm{DP})} = \left\{ \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{a}_{m,k}^{\mathrm{T}} \mathbf{x} \geq b_{m,k}^{(\mathrm{ML})} + b_{m,k}^{(\mathrm{DP})}, i \in \mathcal{J}_m \right\}. \tag{3.16}$$

Therefore, the compact representation for the DPCIR associated with $\mathbf{x}_m$ is obtained as

$$\mathcal{D}_m^{(\mathrm{DP})} \triangleq \left\{ \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{A}_m \mathbf{x} \succeq \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})} \right\}, \tag{3.17}$$

where $\mathbf{b}_m^{(\mathrm{DP})} \in \mathbb{R}^{M_m}$ is the vector containing $d_{m,k}\|\mathbf{a}_{m,k}\|/2$ for all $k \in \mathcal{J}_m$. From (3.17), it follows that if $\mathbf{x}_m$ is an inner constellation point, $\mathcal{D}_m^{(\mathrm{DP})}$ is only composed of $\mathbf{x}_m$ itself, and thus, the noise-free received signal should exactly locate at $\mathbf{x}_m$. In case where $\mathbf{x}_m$ is an outer point, $\mathcal{D}_m^{(\mathrm{DP})}$ will be a convex cone with a vertex at $\mathbf{x}_m$, and the noise-free received signal can lie anywhere on this cone. Figure 3.2 illustrates the DPCIRs (regions

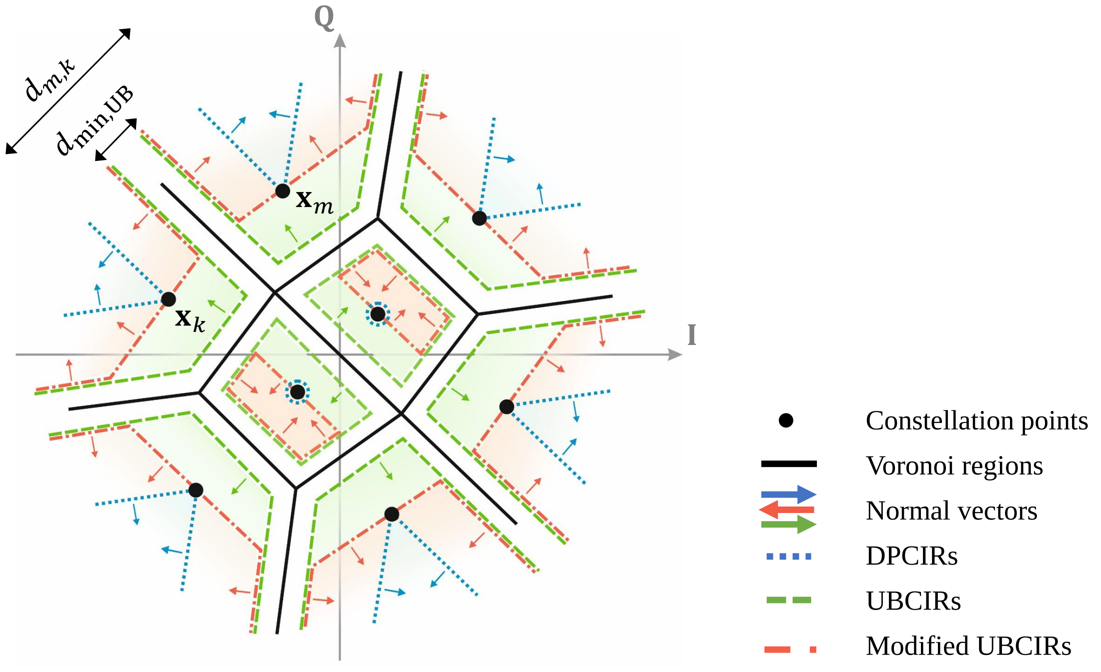


FIGURE 3.2: An illustration of CIRs for an AWGN-optimized 8-ary constellation.

in blue) for an 8-ary constellation set which is optimized over an AWGN channel using the method presented in [124]. Note that if SEP is not allowed to increase, then DPCIRs are optimal and correspond to the CIRs introduced for PSK and QAM constellations in [20] and [105], respectively.

3.3.2 Union Bound Constructive Interference Regions

In practice, the users may have some flexibility in terms of SEP. In such cases, one can relax the DPCIRs as long as a given target SEP is guaranteed. By doing so, we may have larger solution spaces for the SLP problem (3.6), and possibly lower transmit powers are achievable. The relaxation can be done by bringing the CIR hyperplanes closer to the Voronoi decision boundaries.

In what follows, we use the union bound on SEP to determine how close the CIR hyperplanes can get to the Voronoi boundaries. A tractable form of the union bound, known as the nearest neighbor union bound (NNUB), is given in [125] by

$$P_{\mathrm{e}} \leq \left(\frac{1}{M}\sum_{m} M_m\right) Q\left(\frac{d_{\min}}{2\sigma}\right), \tag{3.18}$$

where $Q(v) \triangleq (1/\sqrt{2\pi})\int_v^{\infty} e^{-y^2/2}dy$ is the standard Q-function, σ and P_{e} respectively denote the noise standard deviation and SEP, and $d_{\min}$ is the minimum distance of the constellation defined as

$$d_{\min} \triangleq \min \{ d_{m,k} | \mathbf{x}_m, \mathbf{x}_k \in \mathcal{X}, m, k = 1, 2, ..., M, m \neq k \} . \tag{3.19}$$

The NNUB provides a tight theoretical bound on SEP which is quite close to the exact SEP at high SNRs. Note that in our model, the received signal r_i can be treated as the output of an AWGN channel, and therefore, the NNUB (3.18) is applicable. Using (3.18), for a given P_e, we define the distance threshold $d_{\min,\mathrm{UB}}$ as

$$d_{\min}^{(\mathrm{UB})} = 2\sigma Q^{-1} \left(\frac{M \, P_\mathrm{e}}{\sum_m M_m} \right) . \tag{3.20}$$

where $Q^{-1}(\cdot)$ is the inverse Q-function. The value of $d_{\min}^{(\mathrm{UB})}$ determines how far the noise-free received signal is allowed to be distanced from the desired symbol without violating the target SEP. In other words, $d_{\min}^{(\mathrm{UB})}$ as defined in (3.20) is the smallest minimum distance by which the worst SEP performance is guaranteed to be P_e. This further provides us with the intervals $[d_{\min}^{(\mathrm{UB})}, d_{m,k}]$ from which we can choose the relaxed distances; note, however, that the most power-efficient choice is $d_{\min}^{(\mathrm{UB})}$. We refer to these regions as union bound CIRs (UBCIR). It is worth also noting that in the general case of having unequal user-specific target SEPs, UBCIRs can be defined separately for each user.

In the case of UBCIRs, the displacement value δ can be obtained from (3.14) by substituting $\delta = d_{\min}^{(\mathrm{UB})}/2$. Accordingly, the union bound halfspaces corresponding to $\mathbf{x}_m$ are given by

$$\mathcal{H}_{m,k}^{(\mathrm{UB})} = \left\{ \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{a}_{m,k}^{\mathrm{T}} \mathbf{x} \geq b_{m,k}^{(\mathrm{ML})} + b_{m,k}^{(\mathrm{UB})} \right\} , \quad k \in \mathcal{J}_m, \tag{3.21}$$

where $b_{m,k}^{(\mathrm{UB})} = d_{\min}^{(\mathrm{UB})} \|\mathbf{a}_{m,k}\|/2$. We then intersect (3.21) over all the neighboring points of $\mathbf{x}_m$ to obtain

$$\mathcal{D}_m^{(\mathrm{UB})} = \bigcap_{k \in \mathcal{J}_m} \mathcal{H}_{m,k}^{(\mathrm{UB})} = \left\{ \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{a}_{m,k}^{\mathrm{T}} \mathbf{x} \geq b_{m,k}^{(\mathrm{ML})} + b_{m,k}^{(\mathrm{UB})}, i \in \mathcal{J}_m \right\} . \tag{3.22}$$

Equivalently, the UBCIR associated with $\mathbf{x}_m$ can be represented in a compact form as

$$\mathcal{D}_m^{(\mathrm{UB})} \triangleq \left\{ \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{A}_m \mathbf{x} \succeq \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{UB})} \right\} , \tag{3.23}$$

with $\mathbf{b}_m^{(\mathrm{UB})} \in \mathbb{R}^{M_m}$ containing $d_{\min}^{(\mathrm{UB})} \|\mathbf{a}_{m,k}\|/2$ for all $k \in \mathcal{J}_m$. Note that, in general, the shapes of UBCIRs for a given constellation depend on the Voronoi regions, as illustrated in Fig. 3.2 (regions in green).

The region $\mathcal{D}_m^{(\mathrm{UB})}$ as defined in (3.23) may not fulfill the amplitude condition implied by (3.7). Therefore, we should consider an additional constraint for each constellation point $\mathbf{x}_m$. In fact, this amplitude condition is satisfied if the relaxed CIR corresponding to $\mathbf{x}_m$ is a subset of the complementary region of the disc centered at the origin passing through $\mathbf{x}_m$. However, intersecting such a region with $\mathcal{D}_m^{(\mathrm{UB})}$ yields a non-convex set. An approximate alternative is to consider the outward halfspace generated by the hyperplane tangent to the disc at $\mathbf{x}_m$. This halfspace can be identified by a normal vector $\mathbf{a}_{m,0}$ parallel to $\mathbf{x}_m$ and the offset constant $b_{m,0}^{(\mathrm{ML})} = \mathbf{a}_{m,0}^{\mathrm{T}}\mathbf{x}_m$. Subsequently, $\mathbf{A}_m$, $\mathbf{b}_m^{(\mathrm{ML})}$ and $\mathbf{b}_m^{(\mathrm{UB})}$ in (3.23) are replaced with $\tilde{\mathbf{A}}_m = [\mathbf{A}_m; \mathbf{a}_{m,0}]$, $\tilde{\mathbf{b}}_m^{(\mathrm{ML})} = [\mathbf{b}_m^{(\mathrm{ML})}; b_{m,0}^{(\mathrm{ML})}]$, $\tilde{\mathbf{b}}_m^{(\mathrm{UB})} = [\mathbf{b}_m^{(\mathrm{UB})}; 0]$, respectively, where $[\,\cdot\,;\,\cdot\,]$ denotes concatenation by rows. Loosely speaking, we also refer to these modified regions as UBCIRs in the rest of this chapter, which are shown in Fig. 3.2 in red color.

The definitions of DPCIR and UBCIR are valid for all generic constellations as they depend only on the Voronoi regions. We further point out that one may relax the DPCIRs such that the distance between each CIR boundary and the corresponding Voronoi edge is $d_{\min}$. In this case, the upper bound on SEP provided by the NNUB (3.18) remains unchanged with respect to the original constellation as the constellation's minimum distance is preserved. Such relaxed regions can be referred to as minimum distance-preserving CIRs (MDPCIR). For a constellation point $\mathbf{x}_m$, if there exists at least one neighboring point $\mathbf{x}_k$ with $d_{m,k} > d_{\min}$, the corresponding MDPCIR will be larger than $\mathcal{D}_m^{(\mathrm{DP})}$, but not larger than $\mathcal{D}_m^{(\mathrm{UB})}$. Therefore, in the rest of this chapter, we only focus on DPCIRs and UBCIRs.

3.4 Characterization of DPCIRs

In this section, we provide a comprehensive study of DPCIRs and fully characterize these regions by deriving some of their properties. The main results of this section are stated in Lemma 2, Lemma 3 and Theorem 4.

As mentioned earlier, DPCIRs are defined so that they preserve the Euclidean distances between the constellation points, i.e., they do not increase the SEPs of the users. By definition, any point belonging to the DPCIR of a particular constellation point has an increased distance to all the other constellation points in $\mathcal{X}$. In the following, a systematic representation of DPCIRs based on the ML decision regions of the constellation set $\mathcal{X}$ is provided, which will help us to further study their characteristics.

We start by expanding the compact representation of DPCIRs in (3.17). The vector $\mathbf{b}_m^{(\mathrm{DP})}$, which uniquely describes $\mathcal{D}_m^{(\mathrm{DP})}$, is constructed as

$$\mathbf{b}_m^{(\mathrm{DP})} = \frac{1}{2}\begin{bmatrix} d_{m,1}\|\mathbf{a}_{m,1}\| \\ \vdots \\ d_{m,M_m}\|\mathbf{a}_{m,M_m}\| \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \|\mathbf{x}_m - \mathbf{x}_{m,1}\|^2 \\ \vdots \\ \|\mathbf{x}_m - \mathbf{x}_{m,M_m}\|^2 \end{bmatrix}, \tag{3.24}$$

where $\mathbf{x}_{m,k}$, for $k \in \mathcal{J}_m$, denotes a neighboring constellation point of $\mathbf{x}_m$ distanced by $d_{m,k}$. Furthermore, matrix $\mathbf{A}_m$ collecting the normal vectors of the ML decision boundaries and vector $\mathbf{b}_m^{(\mathrm{ML})}$ containing the offsets from the origin are given by

$$\mathbf{A}_m = \begin{bmatrix} \mathbf{a}_{m,1}^{\mathrm{T}} \\ \vdots \\ \mathbf{a}_{m,M_m}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} (\mathbf{x}_m - \mathbf{x}_{m,1})^{\mathrm{T}} \\ \vdots \\ (\mathbf{x}_m - \mathbf{x}_{m,M_m})^{\mathrm{T}} \end{bmatrix}, \quad \mathbf{b}_m^{(\mathrm{ML})} = \frac{1}{2}\begin{bmatrix} \mathbf{a}_{m,1}^{\mathrm{T}}(\mathbf{x}_m + \mathbf{x}_{m,1}) \\ \vdots \\ \mathbf{a}_{m,M_m}^{\mathrm{T}}(\mathbf{x}_m + \mathbf{x}_{m,M_m}) \end{bmatrix}, \tag{3.25}$$

After some straightforward algebraic steps on (3.25) and (3.24), we obtain

$$\mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})} = \begin{bmatrix} (\mathbf{x}_m - \mathbf{x}_{m,1})^{\mathrm{T}}\mathbf{x}_m \\ \vdots \\ (\mathbf{x}_m - \mathbf{x}_{m,M_m})^{\mathrm{T}}\mathbf{x}_m \end{bmatrix}. \tag{3.26}$$

Using (3.25) and (3.26), we can simplify the representation in (3.17) and describe the DPCIR of $\mathbf{x}_m$ as

$$\mathcal{D}_m^{(\mathrm{DP})} = \left\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{A}_m(\mathbf{x} - \mathbf{x}_m) \succeq \mathbf{0}\right\}, \tag{3.27}$$

The simplified representation in (3.27) describes $\mathcal{D}_m^{(\mathrm{DP})}$ as a vector space originated at $\mathbf{x}_m$ and (non-negatively) spanned by the row vectors of $\mathbf{A}_m$. It is straightforward to show that the following properties hold for DPCIRs:

Property 1. *For all $\mathbf{x}_m \in \mathcal{X}$ and any $\mathbf{x} \in \mathcal{D}_m^{(\mathrm{DP})}$, we have*

i. $\mathcal{D}_m^{(\mathrm{DP})} \subseteq \mathcal{D}_m^{(\mathrm{ML})}$.

ii. $\|\mathbf{x} - \mathbf{y}\|_2 \geq \|\mathbf{x}_m - \mathbf{x}_k\|_2 = d_{m,k}$, $\forall \mathbf{x}_k \in \mathcal{X}, \forall \mathbf{y} \in \mathcal{D}_k^{(\mathrm{DP})}$.

As a special case of Property 1-ii for $\mathbf{y} = \mathbf{x}_k$, we have

$$\|\mathbf{x} - \mathbf{x}_k\|_2 \geq \|\mathbf{x}_m - \mathbf{x}_k\|_2, \quad \forall \mathbf{x}_k \in \mathcal{X}, \tag{3.28}$$

where (3.28) holds with equality only when $\mathbf{x} = \mathbf{x}_m$.

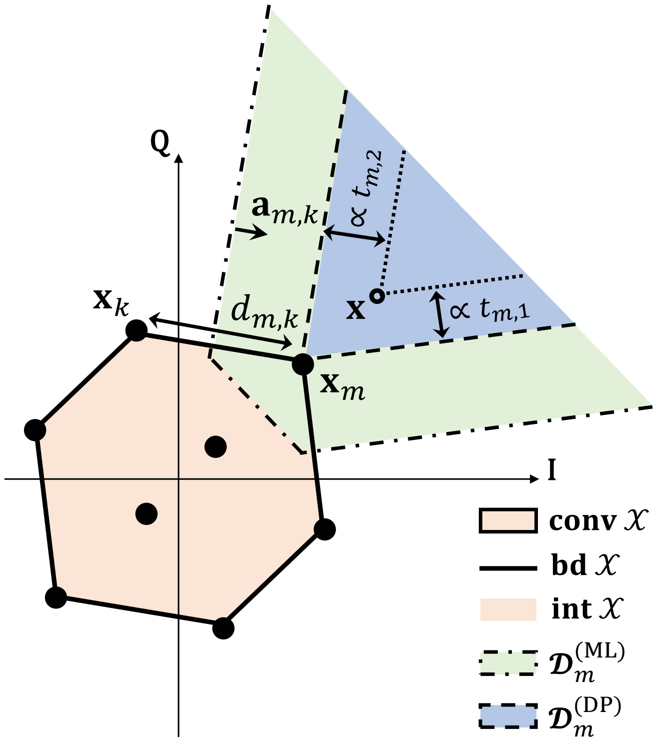
The convex hull of $\mathcal{X}$, denoted by $\mathbf{conv}\mathcal{X}$, refers to the smallest convex set containing $\mathcal{X}$ and can be simply derived from the constellation set $\mathcal{X}$. The set of points belonging to the boundary of $\mathbf{conv}\mathcal{X}$ is denoted by $\mathbf{bd}\mathcal{X}$, and the set of interior points of $\mathbf{conv}\mathcal{X}$, i.e., $\mathbf{conv}\mathcal{X} \backslash \mathbf{bd}\mathcal{X}$, is denoted by $\mathbf{int}\mathcal{X}$. An illustrative example of the these sets for the optimized 8-ary constellation in [124] is shown in Fig. 3.3. It follows from (3.16) that if $\mathcal{D}_m^{(\mathrm{ML})}$ is bounded, then $\mathcal{D}_m^{(\mathrm{DP})} = \mathbf{x}_m$, which means that all the inequalities in (3.16) are satisfied with equality. On the other hand, for an unbounded $\mathcal{D}_m^{(\mathrm{ML})}$, the associated $\mathcal{D}_m^{(\mathrm{DP})}$ is an unbounded polyhedron, or more specifically, a polyhedral angle as depicted in Fig. 3.3, which can be explicitly characterized using the two following lemmas.

Lemma 1. *A point $\mathbf{x}_m \in \mathcal{X}$ lies on the boundary of (or is a vertex of) $\mathbf{conv}\mathcal{X}$ if and only if its Voronoi region $\mathcal{D}_m^{(\mathrm{ML})}$ is unbounded [126, Lemma 2.2].*

Lemma 2. *For every $\mathbf{x}_m \in \mathcal{X}$ with unbounded $\mathcal{D}_m^{(\mathrm{ML})}$, $\mathcal{D}_m^{(\mathrm{DP})}$ is a polyhedral angle with a vertex at $\mathbf{x}_m$ and two infinite edges starting from $\mathbf{x}_m$, where each of its edges is perpendicular to one of the two line segments connecting $\mathbf{x}_m$ to its two neighboring points on $\mathbf{bd}\mathcal{X}$.*

Proof. See Appendix A.1. □

For any $\mathbf{x}_m \in \mathbf{bd}\mathcal{X}$, Lemma 2 implicitly states that $\mathcal{D}_m^{(\mathrm{DP})}$ is not affected by changing the geometry of any point $\mathbf{x}_k \in \mathbf{int}\mathcal{X}$, as well as by adding a new constellation point



(a)

FIGURE 3.3: Geometry of DPCIRs for a boundary constellation point.

on either $\mathbf{bd}\mathcal{X}$ or $\mathbf{int}\mathcal{X}$. This is because the direction of $\mathbf{a}_{m,k}$ remains unchanged for all $\mathbf{x}_k \in \{\mathcal{S}_m \cap \mathbf{bd}\mathcal{X}\}$ under the above operations. Next, we prove that the norm of any point in a DPCIR is always greater than or equal to the norm of the corresponding vertex if and only if the convex hull of the constellation includes the origin. It should be noted that this is a rather light condition, as all well-known constellations in the literature with $M \geq 4$ have at least one point in each quadrant and therefore their convex hull contains the origin.

Lemma 3. *For any constellation point $\mathbf{x}_m \in \mathcal{X}$, we have $\|\mathbf{x}\| \geq \|\mathbf{x}_m\|, \forall \mathbf{x} \in \mathcal{D}_m^{(\mathrm{DP})}$ if and only if* $\mathbf{conv}\mathcal{X}$ *contains the origin. Equality is achieved only when $\mathbf{x} = \mathbf{x}_m$.*

Proof. See Appendix A.2. □

To proceed, it is more convenient to rewrite the linear inequalities in (3.17) as an equivalent set of linear equations. To do so, we introduce a non-negative vector $\mathbf{t}_m$ and describe the region $\mathcal{D}_m^{(\mathrm{DP})}$ as

$$\mathcal{D}_m^{(\mathrm{DP})} = \left\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{A}_m\mathbf{x} = \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})} + \mathbf{t}_m, \mathbf{t}_m \in \mathbb{R}_+^{M_m}\right\}, \tag{3.29}$$

The linear equations in (3.29) indicate that any $\mathbf{x} \in \mathcal{D}_m^{(\mathrm{DP})}$ can be represented as the intersection point of M_m displaced hyperplanes, each of which being parallel to one of M_m boundaries of $\mathcal{D}_m^{(\mathrm{DP})}$ but has a different offset due to the vector $\mathbf{t}_m$. Accordingly, for an inner constellation point $\mathbf{x}_m \in \mathbf{int}\mathcal{X}$, we always have $\mathcal{D}_m^{(\mathrm{DP})} = \mathbf{x}_m$, which is the unique solution to $\mathbf{A}_m\mathbf{x} = \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})}$, i.e.,

$$\mathcal{D}_m^{(\mathrm{DP})} = \left\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{A}_m\mathbf{x} = \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})}\right\}. \tag{3.30}$$

It then follows from (3.29) that

$$\mathbf{t}_m = \mathbf{0}, \quad \forall \mathbf{x}_m \in \mathbf{int}\mathcal{X}. \tag{3.31}$$

It can be easily verified that for any $\mathbf{x}_m \in \mathbf{int}\mathcal{X}$, the region $\mathcal{D}_m^{(\mathrm{DP})}$ is bounded by $M_m \geq 3$ hyperplanes of which at least two are not parallel. This allows us to represent $\mathcal{D}_m^{(\mathrm{DP})}$ as the intersection point of these two non-parallel hyperplanes by considering $\mathbf{t}_m = \mathbf{0}$. Consequently, $\mathbf{A}_m$ can be written as a 2×2 matrix with two linearly independent rows, and thus, is invertible.

On the other hand, it is shown that a hyperplane in a set of hyperplanes describing the boundaries of a polyhedron is redundant if the corresponding polyhedron remains unchanged by removing that hyperplane [127, p. 9]. Therefore, in the rest, we consider the minimal set of hyperplanes that are sufficient to describe $\mathcal{D}_m^{(\mathrm{DP})}$ by removing from (3.29) the equalities that come from a redundant hyperplane. As a result, and based on Lemma 2, for any $\mathbf{x}_m \in \mathbf{bd}\mathcal{X}$, the associated region $\mathcal{D}_m^{(\mathrm{DP})}$ is spanned by two normal vectors corresponding to the (infinite) boundaries of $\mathcal{D}_m^{(\mathrm{DP})}$, i.e.,

$$\mathcal{D}_m^{(\mathrm{DP})} = \left\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{A}_m\mathbf{x} = \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})} + \mathbf{t}_m, \mathbf{t}_m \in \mathbb{R}_+^2\right\}, \tag{3.32}$$

from which any point $\mathbf{x} \in \mathcal{D}_m^{(\mathrm{DP})}$ can be specified by two non-negative coefficients as

$$\mathbf{t}_m = [t_{m,1}, t_{m,2}]^{\mathrm{T}} \in \mathbb{R}_+^2, \quad \forall \mathbf{x}_m \in \mathbf{bd}\mathcal{X}, \tag{3.33}$$

In case the two infinite hyperplanes corresponding to $\mathcal{D}_m^{(\mathrm{DP})}$ are not parallel, matrix $\mathbf{A}_m$ can simply be formed as a 2×2 invertible matrix with two linearly independent rows. It should be further noted that this representation also covers the special case where the two infinite boundary hyperplanes are parallel to each other, e.g., in quadrature amplitude modulation (QAM) constellations. In such a case, both $t_{m,1}$ and $t_{m,2}$ are constrained to be always zero. However, the region $\mathcal{D}_m^{(\mathrm{DP})}$, which is a half-line starting from the constellation point $\mathbf{x}_m$, can be spanned by a non-negative scalar indicating the offset of a virtual hyperplane orthogonal to the two existing infinite boundaries (which preserves the non-singularity of $\mathbf{A}_m$). Thereby, any point $\mathbf{x} \in \mathcal{D}_m^{(\mathrm{DP})}$ can be specified as $\mathbf{t}_m = [t_{m,1}, 0]^{\mathrm{T}}, t_{m,1} \in \mathbb{R}_+$. It is also important to note that while our derivations have been presented for two-dimensional constellation sets, the same concepts can be generalized for both one-dimensional, e.g., pulse amplitude modulation (PAM), and multi-dimensional, e.g., frequency shift keying (FSK), modulation schemes. In general, one may define

$$\mathcal{D}_m^{(\mathrm{DP})} = \left\{ \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \mathbf{A}_m \mathbf{x} = \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})} + \mathbf{t}_m, \mathbf{t}_m \in \mathbb{R}_+^n \right\}, \tag{3.34}$$

with n denoting the dimensionality of the constellation set. In this general case, a number of n normal vectors (each corresponding to a bounding hyperplane) is sufficient to span the entire region, i.e., any $\mathbf{x} \in \mathcal{D}_m^{(\mathrm{DP})}$ can be specified by an n-dimensional vector $\mathbf{t}_m$ as

$$\mathbf{t}_m = [t_{m,1}, t_{m,2}, ..., t_{m,n}]^{\mathrm{T}} \in \mathbb{R}_+^n. \tag{3.35}$$

Accordingly, $\mathbf{A}_m$, $\mathbf{b}_m^{(\mathrm{ML})}$ and $\mathbf{b}_m^{(\mathrm{DP})}$ in (3.34) are constructed with appropriate dimensions. In the special case of PAM constellation with one-dimensional DPCIRs, we have $\mathbf{t}_m = t_m \in \mathbb{R}_+$. Finally, we state the following theorem which will be of essential use in formulating the SLP design problems in the next section.

Theorem 4. *For any constellation point* $\mathbf{x}_m \in \mathbf{bd}\mathcal{X}$ *with* $\mathcal{D}_m^{(\mathrm{DP})}$ *as represented in* (3.29), *function* $f(\mathbf{x}) = \|\mathbf{x}\|$ *over its domain* $\mathcal{D}_m^{(\mathrm{DP})}$ *is a monotonic strictly increasing function of each element of* $\mathbf{t}_m$ *if and only if* $\mathbf{conv}\mathcal{X}$ *contains the origin.*

Proof. See Appendix A.3 □

It is worth noting that Theorem 4 can also be generalized for more generic CIRs, namely, UBCIRs and MDPCIRs as defined in Section 3.3. In both cases, the Euclidean norm of any point belonging to these regions is strictly increasing in exactly two coefficients related to the two infinite bounding hyperplanes.

3.5 Design Formulations for the SLP Problem

In this section, using the convex descriptions and properties of DPCIRs and UBCIRs provided in Section 3.3 and Section 3.4, we formulate the multiuser precoding optimization problems on a symbol-level basis. In particular, we are interested in formulating two well-known design problems, namely, power minimization and SINR balancing. As discussed in Section 3.3, the DPCIRs can be explicitly obtained for all generic constellations as they depend only on the Voronoi regions. This enables us to arrange the design problems in a general form which is indifferent to the constellation's shape and order.

We start by noting that for any user $i = 1, ..., N_{\mathrm{u}}$, the intended data symbol s_i corresponds to one of the points $\{\mathbf{x}_m\}_{m=1}^M$ in $\mathcal{X}$, and thus, we denote $\mathbf{s}_i = [\mathrm{Re}(s_i), \mathrm{Im}(s_i)]^{\mathrm{T}}$. For the brevity of notations, we denote by i the index of matrix $\mathbf{A}$ and vectors $\mathbf{b}^{(\mathrm{ML})}$ and $\mathbf{b}^{(\mathrm{DP})}$ that correspond to $\mathbf{s}_i$. Furthermore, we define the index set $\mathcal{I} = \{i \,|\, \mathbf{s}_i \in \mathbf{bd}\mathcal{X}\}$ referring to those users with a symbol on the boundary of $\mathcal{X}$. In the rest of this chapter, we use the following equivalent real-valued notations:

$$\bar{\mathbf{u}} = \begin{bmatrix} \mathrm{Re}(\mathbf{u}) \\ \mathrm{Im}(\mathbf{u}) \end{bmatrix} \in \mathbb{R}^{2N_{\mathrm{t}} \times 1}, \quad \mathbf{H}_i = \begin{bmatrix} \mathrm{Re}(\mathbf{h}_i) & -\mathrm{Im}(\mathbf{h}_i) \\ \mathrm{Im}(\mathbf{h}_i) & \mathrm{Re}(\mathbf{h}_i) \end{bmatrix} \in \mathbb{R}^{2 \times 2N_{\mathrm{t}}}, \quad i = 1, ..., N_{\mathrm{u}}, \tag{3.36}$$

Note that with the equivalent notations in (3.36), vector $\mathbf{H}_i\bar{\mathbf{u}}$ denotes the real-valued noise-free received signal at the ith user's receiver. Moreover, it is easy to check that $\mathbf{u}^{\mathrm{H}}\mathbf{u} = \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}}$. We also denote by

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}_1\mathbf{H}_1 \\ \vdots \\ \mathbf{A}_{N_\mathrm{u}}\mathbf{H}_{N_\mathrm{u}} \end{bmatrix} \in \mathbb{R}^{2N_\mathrm{u}\times 2N_\mathrm{t}},\ \mathbf{b}^{(\mathrm{ML})} = [\mathbf{b}_1^{(\mathrm{ML})}, ..., \mathbf{b}_{N_\mathrm{u}}^{(\mathrm{ML})}]^{\mathrm{T}} \in \mathbb{R}^{2N_\mathrm{u}},$$

$$\mathbf{b}^{(\mathrm{DP})} = [\mathbf{b}_1^{(\mathrm{DP})}, ..., \mathbf{b}_{N_\mathrm{u}}^{(\mathrm{DP})}]^{\mathrm{T}} \in \mathbb{R}^{2N_\mathrm{u}},\ \mathbf{t} = [\mathbf{t}_1, ..., \mathbf{t}_{N_\mathrm{u}}]^{\mathrm{T}} \in \mathbb{R}^{2N_\mathrm{u}},$$

the vectors and matrices collecting the channel and CI parameters for all N_u users. It should be noted that, in general, the number of rows in matrix $\mathbf{G}$, as well as the number of entries in vectors $\mathbf{b}^{(\mathrm{ML})}$ and $\mathbf{b}^{(\mathrm{DP})}$, are equal to the summation of the number of neighboring constellation points of $\mathbf{s}_i$ over all $i = 1, ..., N_\mathrm{u}$, i.e., $\sum_{i=1}^{N_\mathrm{u}} M_i$. However, as a consequence of (3.30)-(3.33), only two hyperplanes (linear equations) are sufficient to entirely span $\mathcal{D}_i^{(\mathrm{DP})}$ for all $i = 1, ..., N_\mathrm{u}$. This allows us to reduce the problem's dimensionality to $2N_\mathrm{u}$, including the constraints $\mathbf{t}_i = \mathbf{0}$ for any $i \notin \mathcal{I}$.

3.5.1 DPCIR-based SLP Power Minimization

In a realistic multiuser scenario, the power minimization problem might be relevant if the required QoS (e.g. SINR) of all the users can be guaranteed through the available transmission resources in the system. For a detailed discussion on the rationale behind the power minimization problem, we kindly refer the readers to [8]. Accordingly, in this section we first study the relevance of the SLP power minimization problem. For this purpose, we consider a power-restricted scenario in which the downlink transmission is supposed to provide each user with a given target SINRs, while the BS is subject to a total power constraint. This can be interpreted as a feasibility problem based on the given power constraint and the users' SINR requirements. Through this problem, one may examine whether the given SINR requirements are achievable or not, i.e., whether the spatial multiplexing to serve multiple users is meaningful. Otherwise, the system operator decides to relax the other constraints, e.g., decrease the number of users or increase the total power budget. In the sequel, we first express a feasibility problem for the considered scenario and then formulate the power minimization problem.

Let us first focus on DPCIRs. By substituting $\mathbf{H}_i\bar{\mathbf{u}}$ for $\mathbf{x}$ in (3.29) and scaling the ML and distance-preserving offsets $\mathbf{b}_i^{(\mathrm{ML})}$ and $\mathbf{b}_i^{(\mathrm{DP})}$ to satisfy the SINR requirements in (3.5), we can write the CI constraint for the ith user as

$$\mathbf{A}_i\mathbf{H}_i\bar{\mathbf{u}} = \sigma_i\sqrt{\gamma_i}\left(\mathbf{b}_i^{(\mathrm{ML})} + \mathbf{b}_i^{(\mathrm{DP})}\right) + \mathbf{t}_i, \quad \begin{cases} \mathbf{t}_i \succeq \mathbf{0} & i \in \mathcal{I}, \\ \mathbf{t}_i = \mathbf{0} & i \notin \mathcal{I}. \end{cases} \tag{3.37}$$

Taking all N_u users into account, the compact CI constraint (3.37) imposes a total number of $2N_\mathrm{u}$ constraints on the problem. Therefore, the corresponding feasibility problem can be expressed as

$$\begin{aligned}
&\text{find} \quad \bar{\mathbf{u}} \\
&\text{s.t.} \quad \mathbf{A}_i\mathbf{H}_i\bar{\mathbf{u}} = \sigma_i\sqrt{\gamma_i}\ \left(\mathbf{b}_i^{(\mathrm{ML})} + \mathbf{b}_i^{(\mathrm{DP})}\right) + \mathbf{t}_i,\ i = 1, ..., N_\mathrm{u}, \\
&\qquad\ \mathbf{t}_i = \mathbf{0},\ i \notin \mathcal{I}, \\
&\qquad\ \mathbf{t}_i \succeq \mathbf{0},\ i \in \mathcal{I}, \\
&\qquad\ \bar{\mathbf{u}}^\mathrm{T}\bar{\mathbf{u}} \leq p,
\end{aligned} \tag{3.38}$$

where the forth constraint imposes the power restriction on the transmit signal. By defining the $2N_\mathrm{u}\times 2N_\mathrm{u}$ diagonal matrices $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, ..., \sigma_{N_\mathrm{u}})\otimes\mathbf{I}_2$ and $\mathbf{\Gamma} = \mathrm{diag}(\sqrt{\gamma_1}, ..., \sqrt{\gamma_{N_\mathrm{u}}})\otimes\mathbf{I}_2$, we can rewrite problem (3.38) in a more compact form as

$$\begin{aligned}
&\text{find} \quad \bar{\mathbf{u}} \\
&\text{s.t.} \quad \mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma}\mathbf{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) + \mathbf{t}, \\
&\qquad\ \mathbf{t}_i = \mathbf{0},\ i \notin \mathcal{I}, \\
&\qquad\ \mathbf{t}_i \succeq \mathbf{0},\ i \in \mathcal{I}, \\
&\qquad\ \bar{\mathbf{u}}^\mathrm{T}\bar{\mathbf{u}} \leq p.
\end{aligned} \tag{3.39}$$

A sufficient condition under which there exists at least a feasible point for (3.39) can be obtained according to the following proposition.

Proposition 5. *The feasibility problem* (3.39) *has at least one solution for* $N_\mathrm{u} \leq N_\mathrm{t}$ *if*

$$\left\|\mathbf{G}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right)\right\|^2 \leq p, \tag{3.40}$$

where $\mathbf{G}^\dagger = \mathbf{G}^\mathrm{T}(\mathbf{G}\mathbf{G}^\mathrm{T})^{-1}$ *is the Moore-Penrose inverse of* $\mathbf{G}$.

Proof. See Appendix A.5. □

If a solution to (3.39) exists, then the relevant problem is to further reduce the transmit power, which is known as power minimization. The precoder is designed to minimize either the total or the peak (per-antenna) transmit power. The latter objective is more realistic as, in practice, many systems are subject to individual per-antenna power constraints [4, 94]. Accordingly, the DPCIR-based SLP problem minimizing the total transmit power can be formulated as a linearly-constrained quadratic programming (LCQP), i.e.,

$$\begin{aligned}
&\min_{\bar{\mathbf{u}},\mathbf{t}} \quad \bar{\mathbf{u}}^\mathrm{T}\bar{\mathbf{u}} \\
&\text{s.t.} \quad \mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma}\mathbf{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) + \mathbf{t}, \\
&\qquad\ \mathbf{t}_i = \mathbf{0},\ i \notin \mathcal{I}, \\
&\qquad\ \mathbf{t}_i \succeq \mathbf{0},\ i \in \mathcal{I},
\end{aligned} \tag{3.41}$$

which has $2N_\mathrm{t}+2N_\mathrm{u}$ real-valued variables stacked in vectors $\bar{\mathbf{u}}$ and $\mathbf{t}$ and $4N_\mathrm{u}$ constraints. Many algorithms are known to efficiently solve an LCQP, e.g., interior-point, active-set, and gradient methods [121, 128]. Denoting the optimal solution of (3.41) by $\bar{\mathbf{u}}^*$, the feasibility problem (3.39) guarantees that $\bar{\mathbf{u}}^{*\mathrm{T}}\bar{\mathbf{u}}^* \leq p$. On the other hand, by replacing $\bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}}$ with $\|\bar{\mathbf{u}}\|^2_{\infty,\mathbb{C}}$, the design objective aims to minimize the peak per-antenna transmit power, where by $\|\cdot\|_{\infty,\mathbb{C}}$ we mean the infinity norm over equivalent complex-valued elements. This variant of the SLP power optimization problem has also convex objective function and constraints. Hence, it is a convex problem and can be efficiently solved using off-the-shelf algorithms [121]. The feasibility problem (3.39) can further be extended to the case with peak per-antenna power constraints if one substitutes $\|\bar{\mathbf{u}}\|^2_{\infty,\mathbb{C}}$ for $\bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}}$, and p/N_{t} for p. In such a case, the feasibility condition is given by

$$\|\mathbf{G}^{\dagger}\mathbf{\Sigma}\mathbf{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})}+\mathbf{b}^{(\mathrm{DP})}\right)\|^2_{\infty,\mathbb{C}} \leq p/N_{\mathrm{t}}. \tag{3.42}$$

It is worth noting that if the condition in (3.42) holds true, then the feasibility condition in Proposition 5 is also met given the norm inequality

$$\|\mathbf{G}^{\dagger}\mathbf{\Sigma}\mathbf{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})}+\mathbf{b}^{(\mathrm{DP})}\right)\| \leq \sqrt{N_{\mathrm{t}}}\,\|\mathbf{G}^{\dagger}\mathbf{\Sigma}\mathbf{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})}+\mathbf{b}^{(\mathrm{DP})}\right)\|_{\infty,\mathbb{C}}.$$

It is possible to further simplify the power minimization problem (in terms of problem size) by reducing the number of optimization variables and constraints as follows.

Lemma 6. *The LCQP in* (3.41) *can be reduced to*

$$\min_{\mathbf{t}\succeq\mathbf{0}} \quad \left\|\mathbf{G}^{\dagger}\Big(\mathbf{\Sigma}\mathbf{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})}+\mathbf{b}^{(\mathrm{DP})}\right)+\mathbf{W}\mathbf{t}\Big)\right\|^2, \tag{3.43}$$

for $N_{\mathrm{u}} \leq N_{\mathrm{t}}$, where $\mathbf{W}$ is a $2N_{\mathrm{u}} \times 2N_{\mathrm{u}}$ diagonal matrix with a diagonal element being one if it corresponds to a symbol in $\mathcal{I}$, and zero otherwise, i.e.,

$$\mathbf{W} \triangleq \mathrm{diag}(w_1, ..., w_{N_{\mathrm{u}}}) \otimes \mathbf{I}_2, \quad w_i = \begin{cases} 1, & \mathbf{s}_i \in \mathbf{bd}\mathcal{X}, \\ 0, & \mathbf{s}_i \in \mathbf{int}\mathcal{X}. \end{cases}, \quad i = 1, 2, ..., N_{\mathrm{u}}. \tag{3.44}$$

The optimal precoded vector $\bar{\mathbf{u}}^*$ is then obtained as

$$\bar{\mathbf{u}}^* = \mathbf{G}^{\dagger}\Big(\mathbf{\Sigma}\mathbf{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})}+\mathbf{b}^{(\mathrm{DP})}\right)+\mathbf{W}\mathbf{t}^*\Big), \tag{3.45}$$

where $\mathbf{t}^*$ is the solution to (3.43)*.*

Proof. See Appendix A.4. □

The reduced LCQP in (3.43) involves $2N_{\mathrm{u}}$ variables and $2N_{\mathrm{u}}$ constraints, and therefore, its solution is more computationally efficient than that of the original LCQP in (3.41). Moreover, problem (3.43) can be classified as a non-negative least squares (NNLS) optimization, which can be solved using several efficient methods, e.g., fast NNLS algorithm [129]. We will elaborate more on this formulation in Chapter 4 and derive two low-complexity approximate solutions for the SLP power minimization problem.

Before proceeding to the next section, we provide some equivalent design formulations for the SLP power minimization problem. These equivalent formulations will be of frequent use in later chapters. Using the convex representation of DPCIRs in (3.27), by substituting $\mathbf{H}_i\bar{\mathbf{u}}$ for $\mathbf{x}$ and replacing $\mathbf{x}_m$ with the scaled symbol $\sigma_i\sqrt{\gamma_i}\,\mathbf{s}_i$, we can write the CI constraint for the ith user as

$$\text{C1}: \quad \mathbf{A}_i\left(\mathbf{H}_i\bar{\mathbf{u}} - \sigma_i\sqrt{\gamma_i}\,\mathbf{s}_i\right) \succeq \mathbf{0}, \tag{3.46}$$

or equivalently,

$$\mathbf{A}_i\left(\mathbf{H}_i\bar{\mathbf{u}} - \sigma_i\sqrt{\gamma_i}\,\mathbf{s}_i\right) = \mathbf{t}_i, \quad \mathbf{t}_i \succeq \mathbf{0}, \tag{3.47}$$

where $\mathbf{t}_i = \mathbf{0}$ is imposed for inner constellation symbols, i.e., $\mathbf{s}_i \in \mathbf{int}\mathcal{X}$. Collecting the CI constraints (3.47) for all $i \in \{1, 2, ..., N_\mathrm{u}\}$ into a matrix form, we obtain

$$\text{C2}: \quad \mathbf{A}(\mathbf{H}\bar{\mathbf{u}} - \mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s}) = \mathbf{W}\mathbf{t}, \quad \mathbf{t} \succeq \mathbf{0}, \tag{3.48}$$

where we have used the following notations:

$$\mathbf{A} \triangleq \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{N_\mathrm{u}} \end{bmatrix} \in \mathbb{R}^{2N_\mathrm{u}\times 2N_\mathrm{u}}, \quad \mathbf{H} \triangleq \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_{N_\mathrm{u}} \end{bmatrix} \in \mathbb{R}^{2N_\mathrm{u}\times 2N_\mathrm{t}}, \quad \mathbf{t} \triangleq \begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \\ \vdots \\ \mathbf{t}_{N_\mathrm{u}} \end{bmatrix} \in \mathbb{R}^{2N_\mathrm{u}\times 1},$$

and $\mathbf{\Sigma} \triangleq \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_{N_\mathrm{u}})\otimes\mathbf{I}_2$, $\mathbf{\Gamma} \triangleq \mathrm{diag}(\sqrt{\gamma_1}, \sqrt{\gamma_2}, ..., \sqrt{\gamma_{N_\mathrm{u}}})\otimes\mathbf{I}_2$, and $\mathbf{s} \triangleq [\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_{N_\mathrm{u}}]^\mathrm{T}$. As we have shown earlier in Section 3.4, a sub-matrix $\mathbf{A}_i$ can always be formed as an invertible matrix, so does $\mathbf{A}$. Hence, the compact CI constraint (3.48) can be rewritten as

$$\text{C3}: \quad \mathbf{H}\bar{\mathbf{u}} = \mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s} + \mathbf{A}^{-1}\mathbf{W}\mathbf{t}, \quad \mathbf{t} \succeq \mathbf{0}. \tag{3.49}$$

where

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{N_\mathrm{u}}^{-1} \end{bmatrix}.$$

The CI constraints C1 to C3 are all equivalent and can be used interchangeably in formulating the SLP design problem. For the sake of convenience, we have summarized the corresponding design formulations in Table 3.2.

3.5.2 UBCIR-based SLP Power Minimization

In an analogous way, we can utilize the UBCIRs to formulate the SLP design problem. Since the derivation steps are similar to those taken in the previous section, in the following, we only present the final design formulation for the SLP problem.

Table 3.2: Different design formulations for the DPCIR-based SLP power minimization problem.

Name	Formulation
P1	$\min_{\bar{\mathbf{u}}} \quad \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \quad \text{s.t.} \quad \mathbf{A}_i\left(\mathbf{H}_i\bar{\mathbf{u}} - \sigma_i\sqrt{\gamma_i}\,\mathbf{s}_i\right) \succeq \mathbf{0}, \quad i = 1, 2, ..., N_{\mathrm{u}}$
P2	$\min_{\bar{\mathbf{u}},\mathbf{t}} \quad \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \quad \text{s.t.} \quad \mathbf{A}(\mathbf{H}\bar{\mathbf{u}} - \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{s}) = \mathbf{W}\mathbf{t}, \quad \mathbf{t} \succeq \mathbf{0}$
P3	$\min_{\bar{\mathbf{u}},\mathbf{t}} \quad \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \quad \text{s.t.} \quad \mathbf{H}\bar{\mathbf{u}} = \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{s} + \mathbf{A}^{-1}\mathbf{W}\mathbf{t}, \quad \mathbf{t} \succeq \mathbf{0}$

Using the modified convex representation for UBCIRs (3.23) described in Section 3.3.2, we can cast the UBCIR-based power minimization SLP problem as

$$\begin{aligned} \min_{\bar{\mathbf{u}}} \quad & \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{H}\bar{\mathbf{u}} \succeq \boldsymbol{\Sigma}\boldsymbol{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{UB})}\right), \end{aligned} \tag{3.50}$$

Due to the convex quadratic objective function and linear inequality constraints, the optimization problem (3.50) is a convex LCQP, and therefore, can efficiently be solved via standard algorithms [121]. In general, the number of constraints in (3.50) varies from $3N_{\mathrm{u}}$ to $N_{\mathrm{u}} + \sum_i M_i$, depending on the adopted modulation scheme.

3.5.3 DPCIR-based SLP SINR Balancing

In a downlink scenario where power is a strict transmit limitation, fairness might be a relevant design criterion [10]. In this section, we are interested in a max-min fair criterion under which the SLP design problem aims at maximizing the worst SINR among the users constrained by a total transmit power p. Assuming the CIRs to be distance-preserving, the problem is not convex in its original form. Therefore, we first provide an overview and discuss the methods presented in the literature to solve the SLP max-min SINR problem. Then, we derive several alternate convex formulations for this problem. All the proposed methods are simulated in Section 3.6 with a detailed discussion on the complexity and performance.

One may tackle the SLP max-min SINR by exploiting its connection to the power minimization problem, as proposed in [21]. By considering the DPCIR-based design as a generalization of [21], this method iteratively solves the following problem:

$$\begin{aligned} \bar{\mathbf{u}}_{\mathrm{PM}}(\boldsymbol{\Gamma}^*) = \operatorname*{argmin}_{\bar{\mathbf{u}},\mathbf{t}} \quad & \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \\ \text{s.t.} \quad & \mathbf{G}\bar{\mathbf{u}} = \boldsymbol{\Sigma}\boldsymbol{\Gamma}^*\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) + \mathbf{t}, \\ & \mathbf{t}_i = \mathbf{0},\ i \notin \mathcal{I}, \\ & \mathbf{t}_i \succeq \mathbf{0},\ i \in \mathcal{I}, \end{aligned} \tag{3.51}$$

where $\mathbf{\Gamma}^* = \operatorname{diag}\left(\sqrt{\gamma_1^*}, ..., \sqrt{\gamma_{N_\mathrm{u}}^*}\right) \otimes \mathbf{I}_2$ is the input vector of target SINRs given by the optimal solution to the problem:

$$\begin{aligned} \bar{\mathbf{u}}_{\mathrm{SB}}(p) = \underset{\bar{\mathbf{u}},\mathbf{\Gamma},\mathbf{t}}{\operatorname{argmax}} \ & \min_i \, \{\gamma_i\}_{i=1}^{N_\mathrm{u}} \\ \text{s.t.} \ & \mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma}\mathbf{\Gamma}\ \left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) + \mathbf{t}, \\ & \mathbf{t}_i = \mathbf{0},\ i \notin \mathcal{I}, \\ & \mathbf{t}_i \succeq \mathbf{0},\ i \in \mathcal{I}, \\ & \bar{\mathbf{u}}^\mathrm{T}\bar{\mathbf{u}} \le p, \end{aligned} \tag{3.52}$$

until the solution to (3.51) converges to p. It can be inferred that the power optimization problem (3.51) and the max-min SINR (3.52) are related as

$$\bar{\mathbf{u}}_{\mathrm{PM}}(\mathbf{\Gamma}^*) = \bar{\mathbf{u}}_{\mathrm{SB}}\Big(\bar{\mathbf{u}}_{\mathrm{PM}}(\mathbf{\Gamma}^*)^\mathrm{T}\bar{\mathbf{u}}_{\mathrm{PM}}(\mathbf{\Gamma}^*)\Big). \tag{3.53}$$

In fact, to guarantee the SINR requirement $\bar{\mathbf{u}}^\mathrm{T}\mathbf{H}_i^\mathrm{T}\mathbf{H}_i\bar{\mathbf{u}} \ge \sigma_i^2\gamma_i$ through the first constraint of (3.52), the variables γ_i in (3.52) manipulate the instantaneous average power of the constellations from which $\mathcal{D}_i^{(\mathrm{DP})}$ are constructed for all $i = 1, 2, ..., N_\mathrm{u}$. This is a conservative way to guarantee that the instantaneous achieved SINRs satisfy $\mathbb{E}\{\bar{\mathbf{u}}^\mathrm{T}\mathbf{H}_i^\mathrm{T}\mathbf{H}_i\bar{\mathbf{u}}\}/\sigma_i^2 \ge \gamma_i$ for all $i = 1, 2, ..., N_\mathrm{u}$, which is typically desired in conventional multiuser precoding [9]. The optimal solution γ_i^*, however, pushes $\mathbf{H}_i\bar{\mathbf{u}}$ into $\sqrt{\gamma_i^*}\,\mathcal{D}_i^{(\mathrm{DP})}$ rather than $\mathcal{D}_i^{(\mathrm{DP})}$. Since γ_i^* is a function of the user's symbol $\mathbf{s}_i$, it varies over symbol time, which limits the applicability of this method to constant envelope modulations. For modulation schemes with generic constellations, possibly having inner points with bounded decision regions, the ith receiver needs to be aware of the value of γ_i^* in each symbol period in order to correctly detect $\mathbf{s}_i$, which might be quite impractical as this value has to be updated at the symbol rate. It is important to note that we are not allowed to reformulate (3.52) by excluding the constraints related to the symbols $i \notin \mathcal{I}$, as the power optimization problem (3.51) needs to take all the users' symbols into account in order to guarantee the given SINR requirements for all the users.

Assuming identical noise variances across the receivers, i.e., $\sigma_i^2 = \sigma^2$ for all $i = 1, 2, ..., N_\mathrm{u}$, the symbol-level SINR for the ith user is proportional to the instantaneous received power at the ith receiver within each symbol period. As a result, the DPCIR-based SLP max-min SINR problem can be expressed as

$$\begin{aligned} \max_{\bar{\mathbf{u}},\mathbf{t}} \ & \min_i \left\{\bar{\mathbf{u}}^\mathrm{T}\mathbf{H}_i^\mathrm{T}\mathbf{H}_i\bar{\mathbf{u}}\right\}_{i\in\mathcal{I}} \\ \text{s.t.} \ & \mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma}\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) + \mathbf{t}, \\ & \mathbf{t}_i = \mathbf{0},\ i \notin \mathcal{I}, \\ & \mathbf{t}_i \succeq \mathbf{0},\ i \in \mathcal{I}, \\ & \bar{\mathbf{u}}^\mathrm{T}\bar{\mathbf{u}} \le p. \end{aligned} \tag{3.54}$$

By introducing a slack variable λ, we can rewrite (3.54) as

$$\begin{aligned} \max_{\bar{\mathbf{u}},\mathbf{t},\lambda\geq 0} \quad & \lambda \\ \text{s.t.} \quad & \mathbf{G}\bar{\mathbf{u}} = \boldsymbol{\Sigma}\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) + \mathbf{t}, \\ & \bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}} \geq \lambda, \ i \in \mathcal{I}, \\ & \mathbf{t}_i = \mathbf{0}, \ i \notin \mathcal{I}, \\ & \mathbf{t}_i \succeq \mathbf{0}, \ i \in \mathcal{I}, \\ & \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \leq p, \end{aligned} \tag{3.55}$$

which is not convex due to the second set of constraints. To tackle this problem, we use the properties of DPCIRs derived in Section 3.3. Accordingly, any point in $\mathcal{D}_i^{(\mathrm{DP})}$ can be uniquely specified by $\mathbf{t}_i = [t_{i,1}, t_{i,2}]^{\mathrm{T}} \in \mathbb{R}_+^2$ for all $\mathbf{s}_i \in \mathbf{bd}\mathcal{X}$. It further follows from Theorem 4 that $\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}} = \|\mathbf{H}_i\bar{\mathbf{u}}\|^2$ is strictly increasing in each element of $\mathbf{t}_i$ for all $i \in \mathcal{I}$, i.e., letting either $t_{i,1}$ or $t_{i,2}$ be fixed, $\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}}$ is a monotonically increasing function of the other variable. Therefore, given the optimal value of one of the elements, e.g., $t_{i,1}$, for all $i \in \mathcal{I}$, maximizing $\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}}$ is equivalent to maximizing $t_{i,2}$. As a result, by fixing one of the variables $t_{i,1}$ or $t_{i,2}$ for all $i \in \mathcal{I}$, the optimization problem (3.55) can be expressed in a convex form. Let assume $t_{i,1}$ are fixed for all $i \in \mathcal{I}$. Thus, the convex reformulation of (3.55) can be written as

$$\begin{aligned} \max_{\bar{\mathbf{u}},\mathbf{t}\setminus\mathbf{t}_{\mathcal{I},1},\lambda\geq 0} \quad & \lambda \\ \text{s.t.} \quad & \mathbf{G}\bar{\mathbf{u}} = \boldsymbol{\Sigma}\ \left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) + \mathbf{t}, \\ & \mathbf{t}_i = \mathbf{0}, \ i \notin \mathcal{I}, \\ & t_{i,2} \geq \lambda, \ i \in \mathcal{I}, \\ & \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \leq p, \end{aligned} \tag{3.56}$$

where $t_{i,2}$ is substituted for $\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}}$ in (3.55), and $\mathbf{t}_{\mathcal{I},1} \in \mathbb{R}_+^{|\mathcal{I}|}$ denotes the vector collecting $t_{i,1}$ for all $i \in \mathcal{I}$. In theory, achieving the optimum of (3.55) through (3.56) requires an exhaustive search over all possible non-negative values of $t_{i,1}$ for $i \in \mathcal{I}$ and picking the value that maximizes the objective function of (3.56). Alternatively, due to the power restriction induced by p, one may bound and discretize the search interval to do a grid search. This reduces the solution to choose $t_{i,1}$ only from a finite set, but of course, leads us to a sub-optimal solution. Considering an identical search interval for all the users' symbols, let L be the number of discrete values of $t_{i,1}$ for all $i \in \mathcal{I}$, which results in a total number of $L^{|\mathcal{I}|}$ combinations over all $|\mathcal{I}|$ symbols. This means that the number of convex problems to be solved in every symbol period is of order $L^{|\mathcal{I}|}$. In general, the gap to the optimal solution depends on L as well as the accuracy of bounding (i.e., whether the search interval includes the optimal value or not). The output of this grid search approaches the optimum of (3.55) as $L \to \infty$; however, the computational complexity grows exponentially with L. Motivated by the very high and impractical complexity of the grid search method, in the following, we propose two more computationally efficient approaches to solve the SLP max-min SINR problem. The proposed alternative solutions are not equivalent to solving the original problem (3.55), but extensively reduce the computational complexity of the solution compared to the grid search. In Section 3.6, the loss of the proposed approaches with respect to the optimal solution will be evaluated through simulation results.

Semidefinite Programming Formulation

Inspired by the strictly increasing behavior of $\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}}$ with respect to the elements of $\mathbf{t}_i$ for all $i \in \mathcal{I}$, we propose an alternative way to convert (3.55) into a convex problem by replacing the non-convex quadratic constraints on $\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}}$ with affine constraints on $\mathbf{t}_i$, i.e.,

$$\begin{aligned}\max_{\bar{\mathbf{u}},\mathbf{t},\lambda\geq 0} \quad & \lambda \\ \text{s.t.} \quad & \mathbf{G}\bar{\mathbf{u}} = \boldsymbol{\Sigma}\ \left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) + \mathbf{t}, \\ & \mathbf{t}_i = \mathbf{0},\ i \notin \mathcal{I}, \\ & \mathbf{t}_i \succeq \lambda\,\mathbf{1}, i \in \mathcal{I}, \\ & \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \leq p,\end{aligned} \tag{3.57}$$

which can be viewed as jointly maximizing $t_{i,1}$ and $t_{i,2}$ over all $i \in \mathcal{I}$. By Schur complement, problem (3.57) can be written as

$$\begin{aligned}\max_{\bar{\mathbf{u}},\mathbf{t},\lambda\geq 0} \quad & \lambda \\ \text{s.t.} \quad & \mathbf{G}\bar{\mathbf{u}} = \boldsymbol{\Sigma}\ \left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) + \mathbf{t}, \\ & \mathbf{t}_i = \mathbf{0},\ i \notin \mathcal{I}, \\ & \begin{bmatrix} \mathrm{diag}(\mathbf{t}_{\mathcal{I}}) & \mathbf{I}_{2|\mathcal{I}|} \\ \mathbf{I}_{2|\mathcal{I}|} & \lambda\,\mathbf{I}_{2|\mathcal{I}|} \end{bmatrix} \succeq 0, \\ & \begin{bmatrix} 1 & \bar{\mathbf{u}}^{\mathrm{T}} \\ \bar{\mathbf{u}} & p\mathbf{I}_{2N_{\mathrm{t}}} \end{bmatrix} \succeq 0,\end{aligned} \tag{3.58}$$

where $\mathbf{t}_{\mathcal{I}} \in \mathbb{R}_+^{2|\mathcal{I}|}$ is the vector collecting $\mathbf{t}_i$ for all $i \in \mathcal{I}$, and $\succeq 0$ denotes positive semidefinite. Problem (3.58) is a semidefinite programming (SDP) and can be solved using standard algorithms [121]. This convex formulation, however, is not expected to achieve the same solution as compared to the original problem (3.55) since it has a reduced degrees of freedom to maximize the minimum SINR. More precisely, the SDP in (3.58) optimizes $\min\{t_{i,1}, t_{i,2}\}$ instead of optimizing both $t_{i,1}$ and $t_{i,2}$. Nonetheless, the optimal solution of problem (3.58) can be considered as a lower bound on the optimum of the SLP max-min SINR. It is also important to note that the SDP (3.58) is equivalent to the SOCP formulation of SLP SINR balancing proposed for PSK constellations in [20]. However, the SOCP formulation in [20] is not equivalent to the original SLP max-min SINR problem.

Block Coordinate Descent Optimization

To improve the solution of the SDP formulation (3.58), we propose an iterative method based on the block coordinate descent (BCD) algorithm [130]. The BCD algorithm belongs to the family of successive lower-bound maximization methods in which certain approximate version of the objective function is optimized with respect to one block variable at a time, while fixing the rest of the block variables. We denote by $\mathbf{t}_{\mathcal{I},1} \in \mathbb{R}_+^{|\mathcal{I}|}$ and $\mathbf{t}_{\mathcal{I},2} \in \mathbb{R}_+^{|\mathcal{I}|}$ the vectors (blocks) collecting $t_{i,1}$ and $t_{i,2}$ for all $i \in \mathcal{I}$, respectively. Then, the idea behind the BCD algorithm is to successively maximize the worst-user SINR along coordinates $\mathbf{t}_{\mathcal{I},1}$ and $\mathbf{t}_{\mathcal{I},2}$ until convergence of the solution. In more detail, by defining the elementwise monotonically increasing function $f_i : \mathbb{R}_+^2 \mapsto \mathbb{R}$ as

$$f_i(t_{i,1}, t_{i,2}) = \bar{\mathbf{u}}^{\mathrm{T}} \mathbf{H}_i^{\mathrm{T}} \mathbf{H}_i \bar{\mathbf{u}}, \quad i \in \mathcal{I}, \tag{3.59}$$

the objective function of the SLP max-min SINR can be expressed as

$$g(\mathbf{t}_{\mathcal{I},1}, \mathbf{t}_{\mathcal{I},2}) = \min_i \Big\{ f_i(t_{i,1}, t_{i,2}) \Big\}_{i \in \mathcal{I}}. \tag{3.60}$$

In the nth iteration, each block of variables is updated using the following objective functions (the constraints are as before):

$$\mathbf{t}^*_{\mathcal{I},1|n} = \underset{\mathbf{t}_{\mathcal{I},1}}{\operatorname{argmax}} \quad g(\mathbf{t}_{\mathcal{I},1}, \mathbf{t}^*_{\mathcal{I},2|n-1}), \tag{3.61}$$

$$\mathbf{t}^*_{\mathcal{I},2|n} = \underset{\mathbf{t}_{\mathcal{I},2}}{\operatorname{argmax}} \quad g(\mathbf{t}^*_{\mathcal{I},1|n-1}, \mathbf{t}_{\mathcal{I},2}), \tag{3.62}$$

where $\mathbf{t}^*_{\mathcal{I},1|n}$ and $\mathbf{t}^*_{\mathcal{I},2|n}$ respectively denote the optimal solutions of (3.61) and (3.62) obtained from the nth iteration, and $g(\mathbf{t}_{\mathcal{I},1}, \mathbf{t}^*_{\mathcal{I},2|n-1})$ and $g(\mathbf{t}^*_{\mathcal{I},1|n-1}, \mathbf{t}_{\mathcal{I},2})$ are approximate lower bounds on $g(\mathbf{t}_{\mathcal{I},1}, \mathbf{t}_{\mathcal{I},2})$. We adopt a cyclic update rule, i.e., the BCD algorithm cyclically solves the following two SDPs:

$$\begin{aligned} \max_{\bar{\mathbf{u}}, \mathbf{t}_{\mathcal{I},1}, \lambda \geq 0} \quad & \lambda \\ \text{s.t.} \quad & \mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma} \, \big(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\big) + \mathbf{t}, \\ & \mathbf{t}_i = \mathbf{0}, \; i \notin \mathcal{I}, \\ & \begin{bmatrix} \operatorname{diag}(\mathbf{t}_{\mathcal{I},1}) & \mathbf{I} \\ \mathbf{I} & \lambda \mathbf{I} \end{bmatrix} \succeq 0, \\ & \begin{bmatrix} 1 & \bar{\mathbf{u}}^{\mathrm{T}} \\ \bar{\mathbf{u}} & p\mathbf{I} \end{bmatrix} \succeq 0, \end{aligned} \tag{3.63}$$

and

$$\begin{aligned}
\max_{\bar{\mathbf{u}},\mathbf{t}_{\mathcal{I},2},\lambda\geq 0} \quad & \lambda \\
\text{s.t.} \quad & \mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma}\ \left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) + \mathbf{t}, \\
& \mathbf{t}_i = \mathbf{0},\ i \notin \mathcal{I}, \\
& \begin{bmatrix} \mathrm{diag}(\mathbf{t}_{\mathcal{I},2}) & \mathbf{I} \\ \mathbf{I} & \lambda\mathbf{I} \end{bmatrix} \succeq 0, \\
& \begin{bmatrix} 1 & \bar{\mathbf{u}}^{\mathrm{T}} \\ \bar{\mathbf{u}} & p\mathbf{I} \end{bmatrix} \succeq 0,
\end{aligned} \tag{3.64}$$

where the dimensions of identity matrices in (3.63) and (3.64) are the same as in (3.58). Each SDP is solved with respect to one of the blocks $\mathbf{t}_{\mathcal{I},1}$ or $\mathbf{t}_{\mathcal{I},2}$ while the other block is fixed and is given by the solution from the previous iteration. The pseudocode of the proposed method is presented in Algorithm 1, where we have arbitrarily initialized $\mathbf{t}^*_{\mathcal{I},2}$. For all the iterations $n = 1, 2, 3, ...$, we have

$$\mathbf{t}^*_{\mathcal{I},1|n-1} \preceq \mathbf{t}^*_{\mathcal{I},1|n},\ \ \mathbf{t}^*_{\mathcal{I},2|n-1} \preceq \mathbf{t}^*_{\mathcal{I},2|n}, \tag{3.65}$$

and hence

$$\lambda^*_{|n-1} \leq \lambda^*_{|n}, \tag{3.66}$$

where by $\lambda^*_{|n}$ we denote the optimal solution from the nth iteration. The sequence $\{\lambda^*_{|n}\}_{n=1,2,...}$ is therefore guaranteed to converge to a stationary point (i.e., at least a local extremum) of the SLP max-min SINR. As we will see in Section 3.6, the BCD algorithm usually converges within a few iterations.

3.6 Simulation Results

Finally, in this section, we provide some simulation results to validate the analytical discussions in earlier sections and evaluate the performance of the proposed SLP design approaches. We also compare our results with those obtained from the state-of-the-art schemes. In our simulations, we consider a downlink multiuser unicast system with an equal number of transmit and receive antennas, i.e., $N_{\mathrm{t}} = N_{\mathrm{u}}$. The intended symbols for all the users are taken from an identical constellation set. We evaluate the results for three constellations, namely, 8-PSK, optimized 8-ary, and 16-QAM; however, we are particularly interested in the optimized 8-ary constellation since it has a generic shape with unequal distances as well as both bounded and unbounded Voronoi regions. We assume the additive noise component's variance at the receiver of each user to be $\sigma_i^2 = \sigma^2 = 1$ for all $i = 1, ..., N_{\mathrm{u}}$. Furthermore, we assume equal target SINRs $\gamma_i = \gamma$ for all $i = 1, ..., N_{\mathrm{u}}$, when the power minimization is of interest. A quasi-static Rayleigh fading channel is assumed where the complex channel vectors $\mathbf{h}_i$ for $i = 1, ..., N_{\mathrm{u}}$ are generated following an i.i.d. complex Gaussian distribution with zero mean and unit variance, with assumption $\mathbb{E}\{\mathbf{h}_i\mathbf{h}_j^{\mathrm{H}}\} = 0, \forall j = 1, ..., N_{\mathrm{u}}, j \neq i$. As for the BCD algorithm,

Algorithm 1 Block Coordinate Descent Algorithm to solve the SLP max-min SINR

```
input: $\{\mathbf{s}_i\}_{i=1}^{N_\mathrm{u}}, \{\mathbf{h}_i\}_{i=1}^{N_\mathrm{u}}, \mathbf{\Sigma}, p, \epsilon$
initialize: $n \leftarrow 0, \mathbf{t}^*_{\mathcal{I},2|0} \leftarrow \mathbf{0}_{|\mathcal{I}|}$
repeat
    $n \leftarrow n + 1$
    if $n$ is odd then
        $\mathbf{t}^*_{\mathcal{I},2|n} \leftarrow \mathbf{t}^*_{\mathcal{I},2|n-1}$
        solve (3.63)
        return $\lambda^*_{|n}, \mathbf{t}^*_{\mathcal{I},1|n}$
    else
        $\mathbf{t}^*_{\mathcal{I},1|n} \leftarrow \mathbf{t}^*_{\mathcal{I},1|n-1}$
        solve (3.64)
        return $\lambda^*_{|n}, \mathbf{t}^*_{\mathcal{I},2|n}$
    end if
until $|\lambda^*_{|n} - \lambda^*_{|n-1}| \leq \epsilon$
output: $\bar{\mathbf{u}}$
```

we set the terminating condition as $\epsilon = 10^{-3}$ with a maximum number of iterations of 100.

For a power-limited downlink scenario with $N_\mathrm{t} = N_\mathrm{u} = 4$, the feasibility probability of the DPCIR-based SLP scheme is obtained (based on Proposition 5) and shown in Fig. 3.4. The adopted constellation is the optimized 8-ary, and the probabilities are calculated by averaging over all 8^4 possible combinations of the users' symbol vector $\mathbf{s}$, and further averaging over 10^3 randomly generated channel realizations. It can be noticed that for smaller values of γ, the probability of feasibility grows faster as a function of the total transmit power budget. A case-specific example could be wireless systems with adaptive coding and modulation (ACM) capability, such as DVB-S2X broadcasting standard [131]. In DVB-S2X, the target range of SNR for an 8-ary constellation is typically around 5-7 dB over a linear channel (recall that in SLP, SINR translates to the received SNR). In such a system with a total power budget of at least 130 dBW, one can infer from Fig. 3.4 that providing all the users with an SINR (SNR) level of $\gamma = 5$ dB is guaranteed by 90%, and further reduction of transmit power might be possible via the SLP power minimization.

In Fig. 3.5, we plot the average per-antenna transmit power (total power divided by N_t) and the peak per-antenna transmit power obtained from various SLP power minimization techniques for 8-PSK and the optimized 8-ary constellations, respectively. The results are compared to the CI zero-forcing (CIZF), the CI total power minimization (CIPM) [21], and the CI peak power minimization (CIPPM) [94] schemes. Note that the DPCIR-based and UBCIR-based SLP designs minimizing the total transmit power are respectively referred to as "DPCIR-SLP" and "UBCIR-SLP", whereas the SLP designs minimizing the peak per-antenna transmit power are referred to as "DPCIR-SLP-PP"

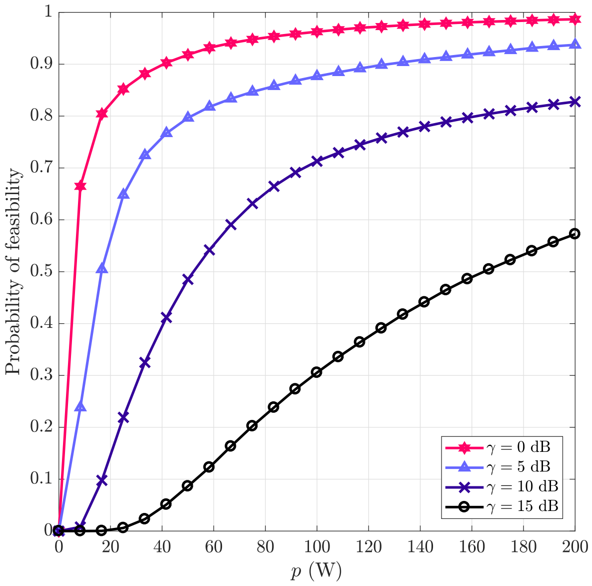
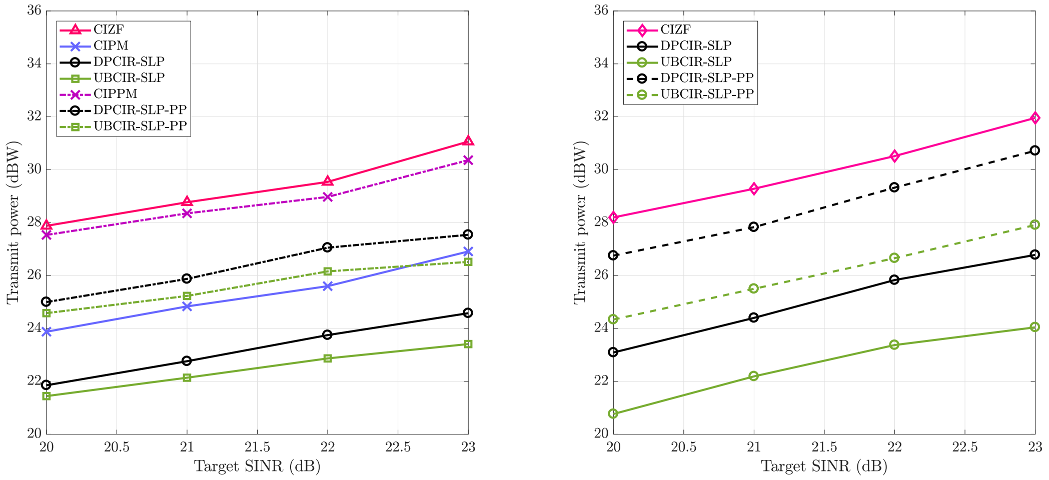


FIGURE 3.4: Feasibility probability of the SLP design as a function of the power budget for different target SINRs with $(N_\mathrm{t}, N_\mathrm{u}) = (4, 4)$.

and “UBCIR-SLP-PP”, respectively. In UBCIR-SLP and UBCIR-SLP-PP, a fixed target SEP of $P_\mathrm{e} = 10^{-3}$ is assumed at all target SINRs.

For the 8-PSK constellation in Fig. 3.5 (a), the transmit powers with DPCIR-SLP and DPCIR-SLP-PP are around 2 dBW less than those obtained by the CIPM and CIPPM schemes, respectively. It should be also noted that the DPCIR-SLP design and the SLP power minimization problem proposed for PSK constellations in [20] are equivalent due to having the same CIRs (as mentioned in Section 3.4). As expected, the UBCIR-SLP and UBCIR-SLP-PP designs with relaxed CIRs are the most power-efficient SLP schemes, both with 1 dBW less transmit power at $\gamma = 23$ dB, compared to the DPCIR-SLP and DPCIR-SLP-PP schemes. This power reduction is achieved in exchange for possibly higher, but upper bounded by 10^{-3} SEPs. To have an estimate of the variance of minimum transmit power due to the random channel matrix $\mathbf{H}$, we have also simulated 700 frames of 100 symbols for the 8-PSK constellation at $\gamma = 21$ dB. The results show a maximum of 4% relative variance compared to those shown in Fig. 3.5 (a). Similar results can be observed in Fig. 3.5 (b) for the optimal transmit powers obtained from different SLP problems with the optimized 8-ary constellation. Note that the CIPM and CIPPM schemes, as formulated in [21] and [94], are not applicable to this constellation.

We also compare the complexities of the SLP power minimization schemes of interest in terms of the average solution time computed by the CVX disciplined convex programming tool (SDPT3 solver) [132]. The relative solution times (normalized by the smallest



(a)

(b)

FIGURE 3.5: Transmit power versus target SINR with $(N_\mathrm{t}, N_\mathrm{u}) = (8, 8)$ for (a) 8-PSK constellation; (b) the optimized 8-ary constellation.

value) and the number of constraints are reported in Table 3.3. As it can be seen, the CIZF scheme offers the lowest time complexity, but it is the least power-efficient SLP design. The DPCIR-SLP and UBCIR-SLP designs, on the other hand, both require a lower solution time than that of the CIPM method, and indeed, they are more power-efficient.

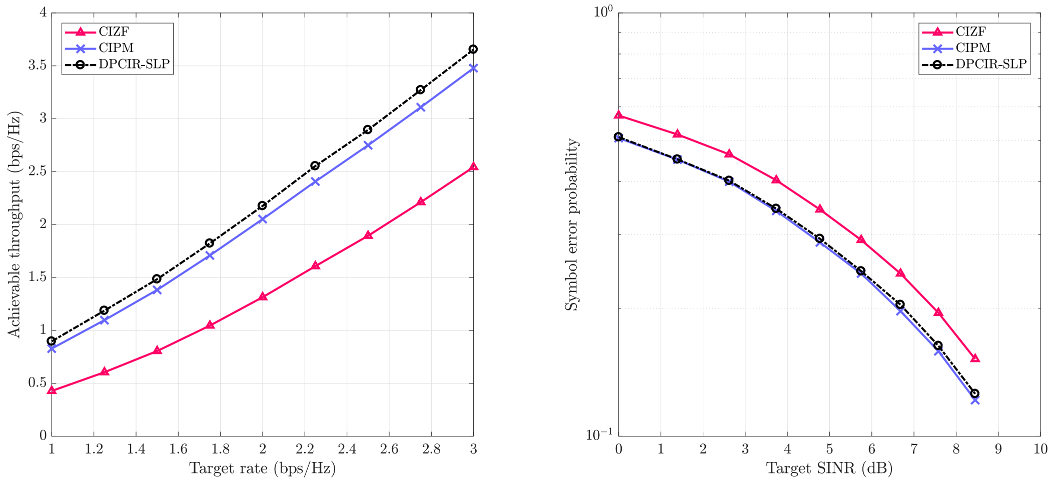
TABLE 3.3: Number of constraints and solution time for different SLP schemes.

	CIZF	CIPM	DPCIR-SLP	UBCIR-SLP
Number of constraints	$2N_\mathrm{u}$	$3N_\mathrm{u}$	$2N_\mathrm{u}$	$3N_\mathrm{u}$
Solution time	1.000	1.330	1.192	1.318

In Fig. 3.6 (a), we plot the average achievable throughput of $N_\mathrm{u} = 8$ users under the SLP power minimization scheme as a function of a given target rate R, where the target rate is related to the SINR requirement as $R = \log_2(1+\gamma)$. The number of BS's transmit antennas is $N_\mathrm{t} = 8$ and an 8-PSK modulation scheme is employed. We define the average achievable throughput for the ith user as

$$(1 - \mathrm{SEP}_i) \log_2 \left(1 + \mathbb{E}\left\{\|\mathbf{h}_i \mathbf{u}\|_2^2\right\}\right), \tag{3.67}$$

where SEP_i is the symbol error probability of the ith user and the expectation has to be taken over an entire block of symbols. In addition to the DPCIR-based SLP design, the results are obtained for two other SLP approaches, namely, constructive interference zero-forcing (CIZF) and constructive interference power minimization (CIPM) [21]. The



(a)

(b)

FIGURE 3.6: Performance comparison for a system with $(N_\mathrm{t}, N_\mathrm{u}) = (8, 8)$: (a) Average per-user achievable throughput as a function of target rate; (b) Average symbol error probability versus target SINR.

proposed DPCIR-based scheme outperforms both CIZF and CIPM. It can also be observed that both the DPCIR-based and the CIPM symbol-level precoders provide higher achievable throughputs than the given target rate. Moreover, under the same scenario, the average symbol error probability over all N_u users is depicted versus the target SINR in Fig. 3.6 (b). As it can be seen, assuming the CI constraints of the SLP power optimization to be distance-preserving causes a very slight difference in the average SEP compared to the CIPM approach (in which the phase of the noise-free received signal is constrained to be aligned with that of the original constellation point). Overall, considering Fig. 3.6 (a), the DPCIR-based SLP shows a better performance than the CIPM in terms of the achievable throughput given by (3.67), where both the shape of the CIRs and the resulting SEP are taken into account.

Figure 3.7 shows the scatter plot of $N_\mathrm{u} \times 10^3$ noise-free received signals in a scenario with $N_\mathrm{t} = N_\mathrm{u} = 8$ and $\gamma = 15$ dBW, where all the transmitted symbols are drawn from 8-PSK constellation and mapped to N_t transmit antennas via a DPCIR-based SLP max-min SINR precoder. In this figure, the black points and the dashed lines represent the constellation points and their corresponding Voronoi regions, respectively. This figure supports the discussion in Section 3.5 regarding the relative geometry of the noise-free received signal in a DPCIR. It can be seen from Fig. 3.7 that the density of signals resulted from the BCD algorithm is higher in areas closer to the boundaries of DPCIRs, while those signals from the SDP formulation being distributed around the bisector (with the majority being located exactly on the bisector). This is a consequence of maximizing the minimum of $t_{i,1}$ and $t_{i,2}$ in (3.58) which, loosely speaking, disregards half of the design degrees of freedoms. On the other hand, as it can be seen from Fig. 3.7,

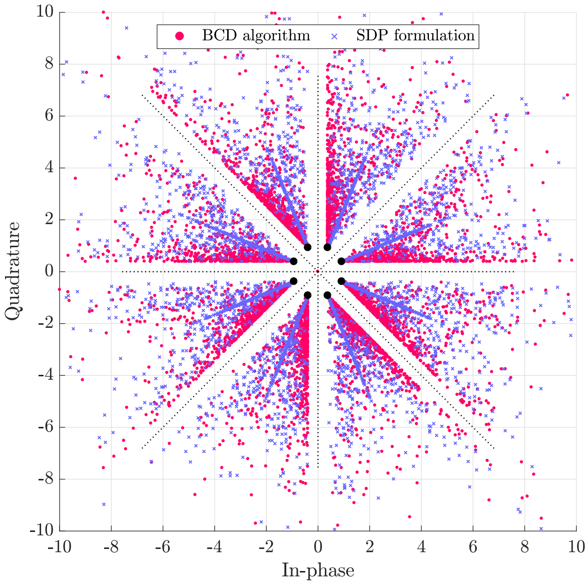
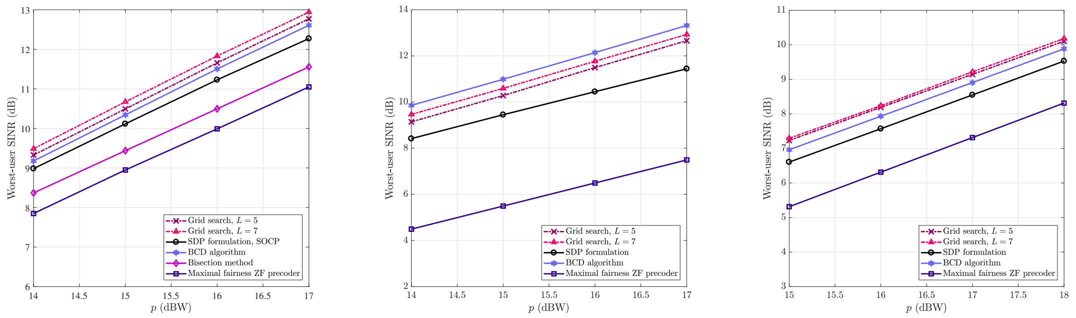


FIGURE 3.7: Scatter plot of the noise-free received signals taken form 8-PSK constellation in a system with $(N_\mathrm{t}, N_\mathrm{u}) = (8, 8)$ and $p = 15$ dBW.

the results obtained from the BCD algorithm are biased towards one of the boundaries in each DPCIR, depending on the initialization step (i.e., whether to initialize $\mathbf{t}_{\mathcal{I},1}$ or $\mathbf{t}_{\mathcal{I},2}$). The exact same plot as in Fig. 3.7 is obtained for the output of the SOCP formulation of SLP max-min SINR in [20].

Figure 3.8 shows the optimized worst-user SINR obtained via different SLP SINR balancing approaches for three constellations 8-PSK, optimized 8-ary and 16-QAM. We further compare the results with those of the maximal fairness zero-forcing precoder in [4], and the bisection algorithm in [21]. The method based on gird search described in Section 3.5 has been used here as a benchmark for comparison. We choose $L = 5$ and $L = 7$ points to search over the interval $[0, 2.5]$. The SDP formulation, while being always superior to the maximal fairness ZF precoding by at least 1 dB, is a lower bound on the optimal solution to the SLP max-min SINR. The BCD algorithm, on the other hand, provides gains of up to 2 dB with respect to the SDP formulation using the optimized 8-ary constellation. The results in Fig. 3.8 (b) further indicate that this iterative method is able to achieve even better solutions compared to the grid search with $L = 7$, but with an extremely lower computational complexity.

We plot, in Fig. 3.9, the worst-user received SINR for a fully-loaded system ($N_\mathrm{t} = N_\mathrm{u}$) as a function of system dimension, where the users' symbols are taken from the optimized 8-ary constellation. As expected, a lower worst SINR is achieved with increasing the system dimension; however, the received SINR drops more slowly with respect to the system dimension for larger power budgets.



(a) (b) (c)

FIGURE 3.8: The worst-user received SINR among $N_{\mathrm{u}} = 4$ users as a function of the power budget for (a) 8-PSK constellation; (b) the optimized 8-ary constellation; (c) 16-QAM constellation.

In Fig. 3.10 (a), we compare the convergence behavior of the BCD algorithm versus system dimension for different power budgets with 8-PSK and the optimized 8-ary constellation. Here, the convergence behavior is shown in terms of the average number of iterations until convergence, i.e., until the terminating condition is met. It can be seen that the BCD algorithm solving the SLP max-min SINR for $N_{\mathrm{t}} = N_{\mathrm{u}} = 4$ converges within a few iterations with an average of up to 6 iterations for $p = 30$ dBW, where each iteration consists of a single SDP. Figure 3.10 (a) also demonstrates a slightly slower convergence behavior for higher values of p which is due to a larger feasible region. Furthermore, in order to evaluate the dependence of the convergence behavior on the constellation size, in Fig. 3.10 (b), we plot the number of iterations for three modulation schemes with different orders. It can be seen from Fig. 3.10 (b) that for a constellation set with narrower unbounded DPCIRs, the BCD algorithm needs fewer iterations to converge. This observation can be justified as a smaller angle between the two distance-preserving boundaries means more alignment between the two block coordinates $\mathbf{t}_{\mathcal{I},1}$ and $\mathbf{t}_{\mathcal{I},2}$. As a result, the BCD algorithm performs fewer recursions among the coordinates. Note that the DPCIR angles for QPSK and 8-PSK constellations are equal to $\pi/2$ and $\pi/4$, respectively.

Complexity comparison between SDP and BCD

In the SDP formulation, a single convex problem has to be solved per symbol period. On the other hand, according to Fig. 3.10 (a), the BCD algorithm converges after 4 iterations (optimized 8-ary) and 6-8 iterations (8-PSK), on average, where each iteration involves solving one SDP. The BCD algorithm, despite having a higher complexity than the SDP formulation, offers gains of 1.5-2.0 dB (optimized 8-ary) and 0.2-0.4 dB (8-PSK) in the worst-user received SINR (see Fig. 3.8). Therefore, the BCD algorithm provides a reasonable complexity-performance tradeoff compared to the SDP formulation. In order to summarize and compare the complexities of the two methods, in Table 3.4, we present

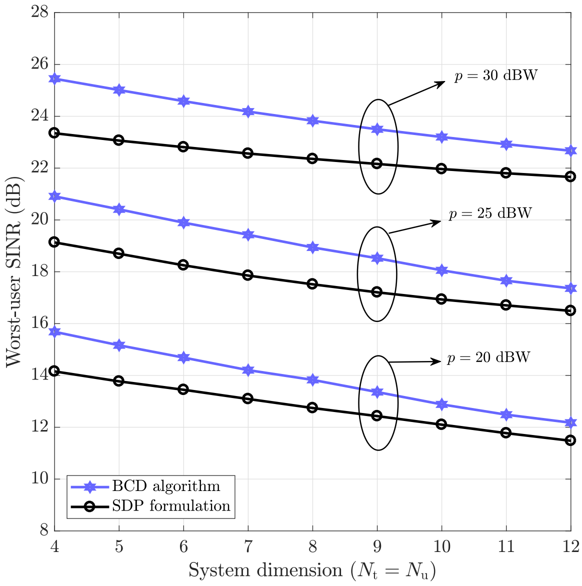
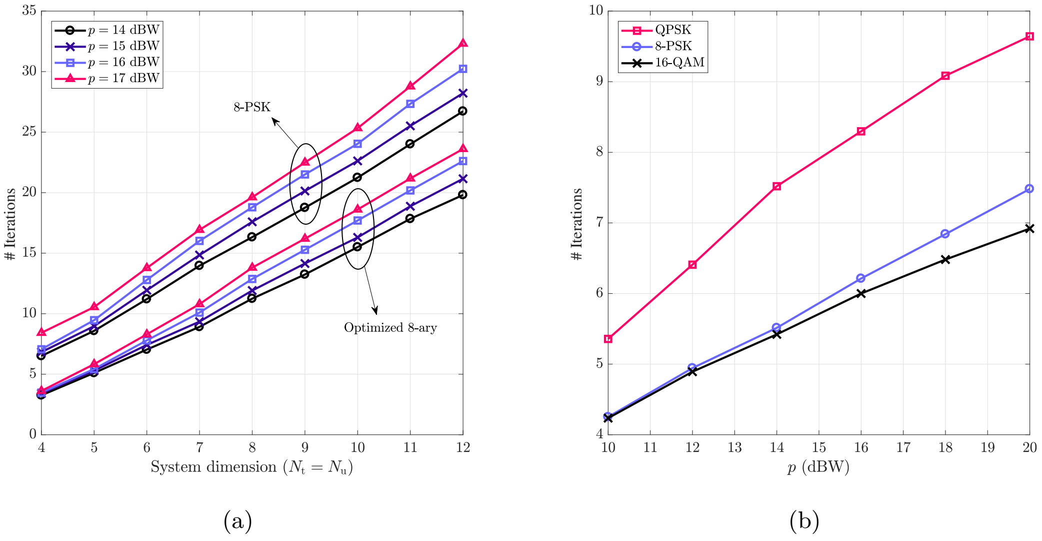


FIGURE 3.9: Worst-user received SINR as a function of system dimension and power budget.



(a)

(b)

FIGURE 3.10: Number of iterations until convergence of the BCD algorithm as a function of (a) system dimension; (b) power budget for three constellations with different orders.

the problem size (in terms of the number of optimization variables) and the number of iterations per symbol period for each method.

Table 3.4: Complexities of the proposed methods for the SLP max-min SINR design.

Method	**Problem size**	**Iteration/symbol period**
SDP formulation	$2N_\mathrm{t} + 2N_\mathrm{u} + 1$	1
BCD algorithm	$2N_\mathrm{t} + 2N_\mathrm{u} - \lvert\mathcal{I}\rvert + 1$	Fig. 3.10

3.7 Conclusions

CIRs are the key to formulate the SLP design problem as they define the constraints to achieve CI at each user's receiver. In this chapter, we first defined the DPCIRs and showed that these regions are optimal when the target SEP is not allowed to increase. In a more flexible setting, we considered relaxed CIRs and guaranteed the target SEP using the union bound, which led us to introduce the UBCIRs. We mainly focused on the DPCIRs and fully characterized these regions for a generic constellation and derived some of their properties. We then addressed two well-known precoding design problems in a downlink multiuser unicast channel, namely, power optimization and SINR balancing, with a symbol-level design approach. Using a systematic description for DPCIRs and UBCIRs, we formulated and discussed the SLP optimization problems. The SINR-constrained SLP power minimization was formulated as a convex problem and studied in a realistic scenario, where a simple feasibility condition was derived. Furthermore, we expressed this optimization in an equivalent form with reduced problem size. Our simulation results indicated that the DPCIR-based and UBCIR-based SLP design formulations can reduce the transmit power consumption without imposing additional complexity on the transmitter compared to the state-of-the-art schemes. For the more challenging and generally non-convex problem of SLP SINR balancing with a max-min fair criterion, the properties of DPCIRs helped us to reformulate the problem in a convex form, which can be solved for a sub-optimal solution. To tackle this problem, we proposed two different approaches, namely, SDP formulation and BCD optimization. We provided a detailed comparison of performance and complexity for the proposed methods.

Chapter 4

Computationally-Efficient Symbol-Level Precoding–Part I: Derivation

While the SLP schemes offer favorable performance gains, e.g., in power-efficiency, they impose a rather high computational complexity on the transmitter. This high complexity comes from the fact that a symbol-level precoder calculates the precoded vector specifically for every set of users' symbols, where this calculation requires solving an optimization problem. In this chapter, we first study the optimal solution to the multiuser SLP design for minimization of the total transmit power under given SINR requirements. We adopt the DPCIRs, as introduced in the previous chapter, and derive a simplified reformulation of the problem in the form of a standard non-negative least squares (NNLS) design. Then, we analyze the structure of the optimal solution using the Karush-Kuhn-Tucker (KKT) optimality conditions. This leads us to obtain a computationally-efficient approximate closed-form SLP solution (CF-SLP). Meanwhile, we obtain a necessary and sufficient condition under which the power minimizer SLP is equivalent to the conventional ZF precoding. Our simulation results show that the CF-SLP technique provides significant gains over the ZF scheme and performs quite close to the optimal SLP in scenarios with a relatively small number of users; however, it shows poor performance for large numbers of transmit antennas and users. To address this drawback, we build on the CF-SLP technique to derive an improved approximate closed-form solution, named ICF-SLP, using the conditions for nearly perfect recovery of the optimal solution support. Through simulation results, we show that in comparison with the CF-SLP technique, the ICF-SLP method significantly enhances the system's performance with a slight increase in complexity. In particular, the ICF-SLP method successfully resolves the drawback of the CF-SLP technique by performing relatively close to the optimal SLP in systems with large numbers of transmit antennas and users. We also compare our computationally-efficient solutions with a fast-converging iterative NNLS algorithm, where the ICF-SLP method shows competitive performance in terms of both accuracy and complexity of the design compared to the iterative algorithm's solution. Analytical and numerical discussions on the complexities of different SLP schemes verify the computational efficiency of the proposed solutions. We show that the CF-SLP and ICF-SLP techniques enjoy a reduction of order 10^3 in the computation time compared to the optimal solution.

4.1 Introduction

The symbol-level design of a multiuser precoder can considerably improve the system's power efficiency. However, it comes with some practical challenges that need to be addressed properly, e.g., a substantially increased computational burden at the transmitter, the need for setting the modulation scheme in advance, and sub-optimality of SINR pilots and log-likelihood ratio (LLR) calculation algorithms; see [29,30]. Among the challenges mentioned above, the increased complexity at the transmitter is one of the main factors that may prohibit the use of SLP schemes in practice; see [19] and [21] for analytical discussions on the computational complexity of the SLP design and [29] for a possible implementation of SLP and the resulting complexity. The high computation cost of SLP is primarily due to the fact that the design needs to be optimized specifically for every set of users' symbols. In high-throughput wireless communication systems, online computation of precoding may suffer from the high complexity of the symbol-level design. On the contrary, an offline (codebook design) computation may lead to an unfavorable computation cost for high-order modulation schemes, even with a moderate number of users [26, 95]. In either case, a relatively large number of optimization problems have to be solved for every realization of the users' symbols. Nonetheless, the considerable performance gain offered by a symbol-level precoder has been motivating to find a more computationally-efficient solution.

In this line of research, some efforts have been made towards deriving low-complexity solutions to the SLP design problem, e.g., [28, 34, 40, 50, 111, 112]. In [40], the authors propose an iterative algorithm with a closed-form update equation for the SLP problem with a max-min fair design criterion, where the algorithm is shown to converge to the optimal solution in a few iterations. The authors in [111] show that, given a perturbation of the target users' symbols, the SLP power minimization design is equivalent to the ZF precoding. In another work [50], the power minimization SLP is addressed with strict phase constraints on the received signals, and a computationally-efficient approximate solution is suggested for this particular case with the PSK modulation schemes. However, the major drawback of the existing methods is the poor performance of the approximate solution for large system dimensions, i.e., large numbers of transmit antennas and users.

In this work, we address the high computational complexity of the SLP problem. We are particularly interested in a power minimization design with SINR constraints, where in formulating the problem, we use the DPCIRs introduced in Chapter 3. The contributions of this work are presented in two chapters. This chapter mainly focuses on the theoretical aspects of the computationally-efficient SLP design problem and proposes two low-complexity algorithms. In Chapter 5, we elaborate on the FPGA design for real-time implementation of the proposed low-complexity SLP algorithms. Accordingly, the main contributions of this chapter are as follows:

1. We transform the SLP problem into an equivalent non-negative least squares (NNLS) design and discuss the optimal solution structure via the Karush-Kuhn-Tucker (KKT) conditions. This leads us to obtain a necessary and sufficient condition under which the SLP design is equivalent to the conventional zero-forcing (ZF) precoding.
2. The KKT conditions for the NNLS design help us to derive a computationally-efficient approximate solution, referred to as CF-SLP, which is given in a closed-form. Through simulation results, we show that the CF-SLP solution performs well close to the optimal SLP scheme for a relatively small number of users, but with a significantly reduced time complexity of order 10^3. However, the main drawback of this approximate solution is its poor performance for large system dimensions.
3. To resolve the performance disadvantage of the CF-SLP method, we improve this solution by applying an additional validation step before calculating the final solution, at the cost of slightly increasing the computational complexity. The new method, named ICF-SLP, significantly improves the system's performance in terms of transmit power. In particular, the gap to the optimal SLP solution remains relatively small, even with increasing the system dimension.
4. We analyze the computational complexities of the CF-SLP and ICF-SLP techniques and compare them with a fast-converging NNLS algorithm solving the SLP problem. Our analyses indicate that the proposed SLP techniques can offer competitive performance compared to the NNLS algorithm while enjoying a lower complexity order.
5. The proposed low-complexity SLP designs are more computationally demanding than simple block-level precoding schemes such as ZF. However, we show that our proposed designs provide substantial gains over the ZF precoding and outperform the classic optimal block-level precoding at high target SINRs. Our results may indeed encourage the use of the proposed SLP designs in practical applications.

The remainder of this chapter is organized as follows. We overview the considered system model in Section 4.2. In Section 4.3, we formulate the SLP optimization problem and discuss the optimal solution structure, which is followed by deriving the optimality conditions. Using these analyses, in Section 4.4, we propose two low-complexity SLP designs and evaluate their computational complexity in Section 4.5. The numerical and simulation results are presented in Section 4.6. Finally, we conclude the chapter in Section 4.7.

4.2 System Model

We mainly consider the same system model as in Chapter 3. As a brief overview, let consider an MU-MIMO downlink system where a BS, equipped with an array of N_t antennas, sends independent data streams to N_u single-antenna users in the same time-frequency resource block, where $N_\mathrm{u} \leq N_\mathrm{t}$. The BS employs an SLP scheme to map independent data symbols $\{s_i\}_{i=1}^{N_\mathrm{u}}$ onto N_t transmit antennas, with s_i denoting the intended symbol for the ith user drawn from a finite equiprobable constellation set. The precoded vector is denoted by $\mathbf{u} = [u_1, \ldots, u_{N_\mathrm{t}}]^\mathrm{T} \in \mathbb{C}^{N_\mathrm{t}\times 1}$. We assume a frequency-flat block-fading channel and denote by row vectors $\mathbf{h}_i \in \mathbb{C}^{1\times N_\mathrm{t}}$ the instantaneous channel coefficients between the BS's antennas and the ith user, for all $i = 1, ..., N_\mathrm{u}$. Accordingly, the received signal at the ith user's receiver can be expressed as

$$r_i = \mathbf{h}_i \mathbf{u} + z_i, \quad i = 1, ..., N_\mathrm{u}, \tag{4.1}$$

where $z_i \sim \mathcal{CN}(0, \sigma_i^2)$ represents the additive complex Gaussian noise at the ith receiver. We define the equivalent real-valued notations as follows:

$$\bar{\mathbf{u}} = \begin{bmatrix} \mathrm{Re}(\mathbf{u}) \\ \mathrm{Im}(\mathbf{u}) \end{bmatrix}, \ \mathbf{H}_i = \begin{bmatrix} \mathrm{Re}(\mathbf{h}_i) & -\mathrm{Im}(\mathbf{h}_i) \\ \mathrm{Im}(\mathbf{h}_i) & \mathrm{Re}(\mathbf{h}_i) \end{bmatrix}, \ \mathbf{s}_i = \begin{bmatrix} \mathrm{Re}(s_i) \\ \mathrm{Im}(s_i) \end{bmatrix}, \quad i = 1, 2, ..., N_\mathrm{u},$$

Hence, the ith user's noise-free received signal is represented by $\mathbf{H}_i \bar{\mathbf{u}} = [\mathrm{Re}(\mathbf{h}_i\mathbf{u}), \mathrm{Im}(\mathbf{h}_i\mathbf{u})]^\mathrm{T}$ for all $i = 1, ..., N_\mathrm{u}$.

4.3 Optimal Solution Structure of the SINR-Constrained Power Minimization SLP Design

We are interested in the SLP power minimization problem constrained by CIRs and the users' SINR requirements. By assuming DPCIRs, we focus on the design formulation P3, which is provided in Section 3.5 as

$$\min_{\bar{\mathbf{u}}, \mathbf{t}\succeq \mathbf{0}} \quad \bar{\mathbf{u}}^\mathrm{T}\bar{\mathbf{u}} \quad \text{s.t.} \quad \mathbf{H}\bar{\mathbf{u}} = \mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s} + \mathbf{A}^{-1}\mathbf{W}\mathbf{t}. \tag{4.2}$$

We further assume that $\mathbf{H}$ is a full row rank matrix with high probability. This results in a bijection between $\bar{\mathbf{u}}$ and $\mathbf{t}$ in (4.2), i.e., for any given $\mathbf{t}$, the least-norm vector $\bar{\mathbf{u}}$ is given by

$$\operatorname*{argmin}_{\bar{\mathbf{u}}} \ \bar{\mathbf{u}}^\mathrm{T}\bar{\mathbf{u}} = \underbrace{\mathbf{H}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s}}_{\bar{\mathbf{u}}_\mathrm{ZF}} + \underbrace{\mathbf{H}^\dagger\mathbf{A}^{-1}\mathbf{W}\mathbf{t}}_{\bar{\mathbf{u}}_\mathrm{SLP}}, \tag{4.3}$$

where $\mathbf{H}^\dagger = \mathbf{H}^\mathrm{T}(\mathbf{H}\mathbf{H}^\mathrm{T})^{-1}$ denotes the Moore-Penrose inverse of $\mathbf{H}$. Equation (4.3) reveals the structure of the minimal-power precoded vector, i.e., the optimal solution to (4.2). Intuitively, it consists of two parts: $\bar{\mathbf{u}}_\mathrm{ZF}$, which is the ZF solution, and $\bar{\mathbf{u}}_\mathrm{SLP}$, which is the CI-dependent part and accounts for the potential gain of SLP compared with the ZF scheme. We can also look at (4.3) from a different point of view by rewriting it as

$$\operatorname*{argmin}_{\bar{\mathbf{u}}} \ \bar{\mathbf{u}}^\mathrm{T}\bar{\mathbf{u}} = \mathbf{H}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\left(\mathbf{s} + \mathbf{\Sigma}^{-1}\mathbf{\Gamma}^{-1}\mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right) \triangleq \mathbf{s}_\mathrm{p}(\mathbf{t}). \tag{4.4}$$

The structure of the least-norm vector $\bar{\mathbf{u}}$ in (4.4) has an interesting interpretation: the SLP design (4.2) is equivalent to implementing the ZF precoding for the perturbed target symbols $\mathbf{s}_\mathrm{p}(\mathbf{t})$, where the perturbation depends on $\mathbf{t}$ and has to be optimized on a symbol-by-symbol basis. The perturbed symbols $\mathbf{s}_\mathrm{p}(\mathbf{t})$ are designed such that they locate within the DPCIRs corresponding to the original target symbols in $\mathbf{s}$. This is accomplished through solving the SLP design in (4.2) which, according to Lemma 6, is equivalent to solving the NNLS problem

$$\mathbf{t}^* = \underset{\mathbf{t}\succeq\mathbf{0}}{\operatorname{argmin}} \quad \left\|\mathbf{H}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s} + \mathbf{H}^\dagger\mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right\|^2, \tag{4.5}$$

and then plugging $\mathbf{t}^*$ into the following closed-form expression to obtain the optimal precoded vector:

$$\bar{\mathbf{u}}^* = \mathbf{H}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s} + \mathbf{H}^\dagger\mathbf{A}^{-1}\mathbf{W}\mathbf{t}^*, \tag{4.6}$$

The NNLS problem, unlike its unconstrained counterpart, is not amenable to a closed-form solution in general due to the non-negativity constraints. Many efficient algorithms solving an NNLS problem can be found in the literature, such as the well-known active set based method proposed by Lawson and Hanson [133], the fast NNLS algorithm (FNNLS) [134], and those based on projected/proximal gradient method [135–137]. However, an NNLS algorithm, in the best known case, requires tens of iterations to converge. For instance, the accelerated gradient method offers a superlinear convergence rate of $\mathcal{O}(1/n^2)$, where n is the number of iterations. With a convex objective function, this translates to a worst-case complexity bound of $\mathcal{O}(1/\sqrt{\epsilon})$ to reach an ϵ-optimal solution. As an illustrative example, using an accelerated projected gradient descent algorithm, it takes nearly 100 iterations to have a residual error of 10^{-3} with respect to the optimum. In a practical SLP application, this process has to be done either for every symbol period or every possible symbol set corresponding to N_u users. This motivates us to derive more computationally-efficient, though possibly approximate, solutions for the SLP design problem. In the next section, we take a closer look at the optimization problem (4.2) in order to obtain the optimality conditions for the SLP power minimization design. The results of the next section will help us in deriving a low-complexity approximate solution in Section 4.4.1.

4.3.1 Optimality Conditions for the SLP Power Minimization Problem

Let us denote $\mathbf{B} \triangleq -\mathbf{H}^\dagger\mathbf{A}^{-1}\mathbf{W}$ and $\mathbf{y} \triangleq \mathbf{H}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s}$. Therefore, the NNLS problem (4.5) can be written in the standard form as

$$\min_{\mathbf{t}\succeq\mathbf{0}} \quad \|\mathbf{y} - \mathbf{B}\mathbf{t}\|^2, \tag{4.7}$$

To obtain the optimality conditions for problem (4.2), we use the method of Lagrange multipliers. Accordingly, the Lagrangian of (4.7) is given by

$$\mathcal{L}(\mathbf{t},\boldsymbol{\lambda}) = \mathbf{y}^\mathrm{T}\mathbf{y} + 2\mathbf{y}^\mathrm{T}\mathbf{B}\mathbf{t} + \mathbf{t}^\mathrm{T}\mathbf{B}^\mathrm{T}\mathbf{B}\mathbf{t} + \boldsymbol{\lambda}^\mathrm{T}\mathbf{t}, \tag{4.8}$$

where $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_{2N_\mathrm{u}}]^\mathrm{T}$ is the vector of the Lagrange multipliers. From (4.8), the Lagrange dual problem can be written as

$$\max_{\boldsymbol{\lambda} \preceq \mathbf{0}} \quad \inf_{\mathbf{t} \succeq \mathbf{0}} \mathcal{L}(\mathbf{t}, \boldsymbol{\lambda}), \tag{4.9}$$

Denoting the primal and dual optimal by $\mathbf{t}^*$ and $\boldsymbol{\lambda}^*$, respectively, the Karush-Kuhn-Tucker (KKT) optimality conditions are given as

$$\begin{aligned} \nabla_{\mathbf{t}} \mathcal{L}(\mathbf{t}^*, \boldsymbol{\lambda}^*) = \mathbf{0}, &\quad \text{(stationarity)} && (4.10a)\\ \mathbf{t}^* \succeq \mathbf{0}, &\quad \text{(primal feasibility)} && (4.10b)\\ \boldsymbol{\lambda}^* \preceq \mathbf{0}, &\quad \text{(dual feasibility)} && (4.10c)\\ \boldsymbol{\lambda}^{*\mathrm{T}} \mathbf{t}^* = 0, &\quad \text{(complementary slackness)} && (4.10d) \end{aligned}$$

Note that since the primal problem (4.7) is convex, strong duality holds and the KKT conditions (4.10a)-(4.10d) are necessary and sufficient [121]. As a consequence, a candidate solution satisfying all the KKT conditions is globally optimal. Let $\mathbf{Q} = \mathbf{Q}^\mathrm{T} \triangleq \mathbf{B}^\mathrm{T}\mathbf{B} = [\mathbf{q}_1, ..., \mathbf{q}_{2N_\mathrm{u}}]^\mathrm{T}$ and $\mathbf{p} \triangleq \mathbf{B}^\mathrm{T}\mathbf{y} = [p_1, ..., p_{2N_\mathrm{u}}]^\mathrm{T}$. Using these new notations, the stationarity condition (4.10a) can be written as $2\mathbf{Q}\mathbf{t}^* + 2\mathbf{p} + \boldsymbol{\lambda}^* = \mathbf{0}$, and therefore,

$$\boldsymbol{\lambda}^* = -2(\mathbf{Q}\mathbf{t}^* + \mathbf{p}). \tag{4.11}$$

It then follows from (4.10c) and (4.11) that

$$\mathbf{Q}\mathbf{t}^* + \mathbf{p} \succeq \mathbf{0}. \tag{4.12}$$

Furthermore, plugging $\boldsymbol{\lambda}^*$ from (4.11) into (4.10d) yields

$$(\mathbf{Q}\mathbf{t}^* + \mathbf{p})^\mathrm{T}\mathbf{t}^* = 0, \tag{4.13}$$

from which by denoting $\mathbf{v} \triangleq \mathbf{Q}\mathbf{t}^* + \mathbf{p} = [v_1, ..., v_{2N_\mathrm{u}}]^\mathrm{T}$ and $\mathbf{t}^* = [t_1^*, ..., t_{2N_\mathrm{u}}^*]^\mathrm{T}$, it follows that

$$\sum_{l=1}^{2N_\mathrm{u}} v_l \, t_l^* = 0. \tag{4.14}$$

Considering (4.10b) and (4.12), we have $v_l \geq 0$ for all $l = 1, ..., 2N_\mathrm{u}$. As a consequence, the optimality condition (4.14) is satisfied if and only if

$$v_l \, t_l^* = 0, \quad \forall l \in \{1, ..., 2N_\mathrm{u}\}. \tag{4.15}$$

In other words, v_l and t_l^* cannot be both non-zero for any specific $l \in \{1, ..., 2N_\mathrm{u}\}$. Based on this observation, the following lemma relates the SLP solution to that of the ZF precoder.

Lemma 7. *The optimal solution to the SLP power minimization* (4.2) *is equal to the solution of ZF if and only if* $\mathbf{p} \succeq \mathbf{0}$.

Proof. See Appendix B.1 □

Lemma 7 provides a necessary and sufficient condition under which the DPCIR-based SLP design has the same solution as that of the ZF scheme. This condition depends on the instantaneous realization of the users' symbols as $\mathbf{p} \succeq \mathbf{0}$ is equivalently met by $\mathbf{B}^{\mathrm{T}}\mathbf{y} \succeq \mathbf{0}$. It can be further inferred from (4.15) and Lemma 7 that as the number of non-zero (i.e., positive) elements of $\mathbf{v}$ decreases, the SLP solution may diverge from that of the ZF. In the extreme case where $v_l = 0$ for all $l = 1, ..., 2N_{\mathrm{u}}$, there exists at least one $t_l^* \neq 0$, which can be verified as follows. The linear system $\mathbf{v} = \mathbf{Q}\mathbf{t}^* + \mathbf{p} = \mathbf{0}$ has a unique solution equal to $\mathbf{t}^* = -\mathbf{Q}^{-1}\mathbf{p} = \mathbf{A}\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s}$. Since $\mathbf{A}$ is full rank, it has an empty null space, thus $\mathbf{A}\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s} \neq \mathbf{0}$. This means that $\mathbf{t}^* \neq \mathbf{0}$ and it has at least one non-zero element. In such cases, the SLP design would be able to provide higher precoding gains compared to the ZF scheme. This case, however, is feasible only if the unique solution to the system of linear equations $\mathbf{Q}\mathbf{t}^* + \mathbf{p} = \mathbf{0}$ is non-negative, i.e., $-\mathbf{Q}^{-1}\mathbf{p} \succeq \mathbf{0}$, or equivalently $\mathbf{A}\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s} \preceq \mathbf{0}$.

4.4 Low-Complexity SLP Design

The main goal of this section is to derive low-complexity solutions for the NNLS design formulation of the SLP problem in (4.5). We first start off by reviewing some basic mathematical analysis on the NNLS problem. Let $\mathbf{t}^* = [t_1^*, ..., t_{2N_{\mathrm{u}}}^*]^{\mathrm{T}}$ denote the minimizer of (4.7). We refer to the set of indices l for which $t_l^* > 0$ as the support of $\mathbf{t}^*$, or the optimal support, i.e.,

$$\Lambda^* = \{l \mid l = 1, 2, ..., 2N_{\mathrm{u}},\ t_l^* > 0\}. \tag{4.16}$$

Given the optimal support Λ^*, the minimizer of (4.7) can be simply computed by $(\mathbf{B}_{\Lambda^*})^{\dagger}\mathbf{y}$ with appropriate zero-padding, where $\mathbf{B}_{\Lambda^*}$ denotes the matrix composed of those columns of $\mathbf{B}$ associated with the indices in Λ^*. In other words, finding Λ^* is as complex as solving (4.7) for the optimal solution. Therefore, one may attempt to solve (4.7) equivalently by perfectly identifying Λ^*. This is in fact the underlying idea behind the active set methods, where at each iteration some constraints are set to be active (i.e., zero-valued in our context), while the other constraints are used in the update equation. However, here we are interested in having an approximation of Λ^*, say $\hat{\Lambda}$, obtained in a non-iterative manner. This enables us to derive an approximate solution $\hat{\mathbf{t}}$ in an explicit form. Thereby, using (4.6), we can obtain an approximate precoded vector $\bar{\mathbf{u}}$. In the next two subsections, we aim to obtain such approximate solutions.

4.4.1 Closed-Form Approximate Solution

Using the KKT optimality conditions, a computationally-efficient solution for the SLP design (4.2) can be derived with a simple idea behind. Based on the optimal support

Λ^*, from (4.15), we have

$$v_l = \mathbf{q}_l^{\mathrm{T}}\mathbf{t}^* + p_l = 0, \quad \forall\, l \in \Lambda^*. \tag{4.17}$$

which gives a reduced system of linear equations to obtain $\mathbf{t}^*$. To approximate Λ^*, using (4.13) along with the fact that $\mathbf{Q}$ is positive definite, we obtain

$$\mathbf{p}^{\mathrm{T}}\mathbf{t}^* = \sum_{l=1}^{2N_{\mathrm{u}}} v_l\, t_l^* \leq 0, \tag{4.18}$$

where equality holds only when $\mathbf{t}^* = \mathbf{0}$. An approximation of Λ^* can be derived based on the sign of the elements in $\mathbf{p}$, i.e., $\hat{\Lambda} = \{l | v_l < 0\}$ with $|\hat{\Lambda}| = L$. Here, it is assumed that $t_l^* = 0$ (i.e., the lth constraint is active at the optimum) for those l with $v_l \geq 0$. This results in the desired reduced system of linear equations, given by

$$\mathbf{Q}_{\hat{\Lambda}}\mathbf{t}^*_{\hat{\Lambda}} + \mathbf{p} = \mathbf{0}, \tag{4.19}$$

where $\mathbf{Q}_{\hat{\Lambda}} \in \mathbb{R}^{2N_{\mathrm{t}}\times L}$ and $\mathbf{t}^* \in \mathbb{R}^{L\times 1}$ are obtained by excluding those columns and elements, respectively, in $\mathbf{Q}$ and $\mathbf{t}^*$, that are not indexed in the approximate set $\hat{\Lambda}$. This new system has $2N_{\mathrm{t}}$ linear equations but L variables, where $L \leq 2N_{\mathrm{u}}$. Hence, the reduced linear system possibly has a smaller size than the original problem. By noticing that the non-singularity of $\mathbf{Q}$ is preserved by excluding some of its columns, the unique solution to (4.19) is readily given by the following closed-form expression:

$$\mathbf{t}^*_{\hat{\Lambda}} = \max\left\{-\mathbf{Q}^{\dagger}_{\hat{\Lambda}}\mathbf{p}, \mathbf{0}\right\}, \tag{4.20}$$

where $\max\{\cdot\}$ denotes elementwise maximum, and is applied in order to guarantee the primal feasibility condition (4.10b). The entire vector $\mathbf{t}^*$ can be obtained by inserting the zero-valued elements t_l^* for $l \notin \hat{\Lambda}$ into $\mathbf{t}^*_{\hat{\Lambda}}$ as an approximate solution to (4.5). In what follows, we refer to this closed-form SLP solution as CF-SLP.

4.4.2 Improved Closed-Form Approximate Solution

Based on our experiments, the loss of the approximate solution obtained in Section 4.4.1 with respect to the optimal SLP design is unfavorably high for large values of N_{t} and N_{u}. Thus, we aim to further improve this solution by performing some intermediate steps. Our proposed method is essentially based on the following lemma from [138] which gives the sufficient conditions for nearly perfect recovery of the optimal support Λ^*. Note that here we state a modified version of this lemma according to our notations.

Lemma 8. *Let Λ be a subset of column indices of the matrix $\mathbf{B}$ with $|\Lambda| \leq 2N_{\mathrm{u}}$, and the columns associated with the indices in Λ are linearly independent. Let $\mathbf{t}^* \succeq \mathbf{0}$ be the minimizer of $\|\mathbf{y} - \mathbf{B}\mathbf{t}\|^2$. Then, Λ coincides with the support of $\mathbf{t}^*$ if*

$$\mathrm{C1}:\ \ \mathbf{B}_{\Lambda}{}^{\dagger}\mathbf{y} \succ \mathbf{0}, \quad \textit{and} \quad \mathrm{C2}:\ \ \mathbf{y}^T\mathbf{P}_{\Lambda}^{\perp}\mathbf{b}_l < \mathbf{0}, \quad \forall l \in \Lambda^{\mathrm{c}},$$

where $\mathbf{P}_{\Lambda}^{\perp}$ *is the projector onto the orthogonal complement of the column space of* $\mathbf{B}_{\Lambda}$*, denoted by* $\mathcal{R}(\mathbf{B}_{\Lambda})$*,* $\mathbf{b}_l$ *denotes the lth column of* $\mathbf{B}_{\Lambda}$*, and* $\Lambda^{\mathrm{c}} = \{1, ..., 2N_{\mathrm{u}}\} - \Lambda$*.*

Based on Lemma 8, both the conditions C1 and C2 together are sufficient for a candidate support Λ to be optimal. In fact, C1 measures if the resultant solution satisfies the positivity constraint (notice that the constraint cannot be satisfied with equality due to the definition of support), while the projection in C2 can be viewed as the deviation of $\mathbf{y}$ from the column space of $\mathbf{B}_{\Lambda}$. In other words, C1 is required to validate the columns already indexed in Λ, whereas C2 assesses the possibility of including any of the columns belonging to Λ^{c}. Armed with these two conditions, we are ready to approximately solve the NNLS problem in (4.7), as will be explained in the sequel.

First, we exploit the condition C2 to produce an initial approximation of Λ^*. Let

$$d_l \triangleq \mathbf{y}^T \mathbf{P}_{\Lambda}^{\perp} \mathbf{b}_l, \quad l = 1, ..., 2N_{\mathrm{u}},$$

where

$$\mathbf{P}_{\Lambda}^{\perp} = \mathbf{I} - \mathbf{B}_{\Lambda} \left(\mathbf{B}_{\Lambda}{}^{\mathrm{T}} \mathbf{B}_{\Lambda} \right)^{-1} \mathbf{B}_{\Lambda}{}^{\mathrm{T}}.$$

Treating the entire columns of $\mathbf{B}$ as candidate columns to be indexed in Λ, we assume $\Lambda^{\mathrm{c}} = \{1, 2, ..., 2N_{\mathrm{u}}\}$. This yields $\mathbf{P}_{\Lambda}^{\perp} = \mathbf{I}$, and hence,

$$d_l = \mathbf{y}^{\mathrm{T}} \mathbf{b}_l, \quad l = 1, ..., 2N_{\mathrm{u}}. \tag{4.21}$$

Note that the conditions in (4.21) are similarly implied from the KKT optimality conditions, as discussed in Section 4.4.1. Using the inner products (4.21), we define $\hat{\Lambda} \triangleq \{l : d_l > 0\}$ with $|\hat{\Lambda}| = L_1$, which serves as our initial approximation of Λ^*. We validate this approximation by excluding those columns in $\hat{\Lambda}$ that result in negative elements for $\mathbf{t}$, i.e.,

$$\hat{\hat{\Lambda}} \triangleq \left\{ l | l \in \hat{\Lambda}, \left[(\mathbf{B}_{\hat{\Lambda}})^{\dagger} \mathbf{y} \right]_l > 0 \right\}, \tag{4.22}$$

where $[\,\cdot\,]_l$ denotes the lth element of an input vector. It immediately follows from (4.22) that $|\hat{\hat{\Lambda}}| \triangleq L_2 \leq L_1$, which reduces the possibility of having negative elements in the final solution as a result of the additional validation step in (4.22). Our simulations indicate that in the majority of cases, $\hat{\hat{\Lambda}}$ gives a more accurate approximation of the optimal support Λ^*, compared to that given by $\hat{\Lambda}$, as we will see in Section 4.6. Note, however, that the non-negative constraints may still be violated even after the validation step in (4.22) since the remaining set of columns in $\hat{\hat{\Lambda}}$ does not necessarily guarantee that $(\mathbf{B}_{\hat{\hat{\Lambda}}})^{\dagger} \mathbf{y} \succ \mathbf{0}$. Therefore, one still needs to ignore all the negative elements in the final solution, if any. More precisely, due to the fact that $\mathcal{R}(\mathbf{B}_{\hat{\hat{\Lambda}}}) \subseteq \mathcal{R}(\mathbf{B}_{\hat{\Lambda}})$, perfect recovery of the optimal support is possible only if $\Lambda^* \subseteq \hat{\Lambda}$. In such a case, we obtain $(\mathbf{B}_{\hat{\Lambda}})^{\dagger} \mathbf{y} \succ \mathbf{0}$ and $\hat{\hat{\Lambda}}$ is the optimal support. Consequently, the approximate solution $\hat{\mathbf{t}} = [\hat{t}_1, ..., \hat{t}_{2N_{\mathrm{u}}}]^{\mathrm{T}}$ can be obtained as a zero-padded version of the vector $(\mathbf{B}_{\hat{\hat{\Lambda}}})^{\dagger} \mathbf{y}$, i.e.,

$$\hat{t}_l = \max \left\{ \left[(\mathbf{B}_{\hat{\hat{\Lambda}}})^{\dagger} \mathbf{y} \right]_l, 0 \right\}, \quad l \in \hat{\hat{\Lambda}}, \tag{4.23}$$

Algorithm 2 APGD algorithm solving the NNLS problem (4.7)

```
input : $\mathbf{B}, \mathbf{y}, n_{\max}$
output : $\bar{\mathbf{u}}$
initialize : $\mathbf{t}^{(0)} = \boldsymbol{\vartheta}^{(0)} \in \mathbb{R}_{+}^{2N_{\mathrm{u}} \times 1}$, $\boldsymbol{\Theta} = \mathbf{I} - \frac{\mathbf{B}^{\mathrm{T}}\mathbf{B}}{\|\mathbf{B}^{\mathrm{T}}\mathbf{B}\|_{\mathrm{F}}}$, $\mathbf{z} = \frac{\mathbf{B}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{B}^{\mathrm{T}}\mathbf{B}\|_{\mathrm{F}}}$, $n = 0$
set : $\psi = \frac{1-\sqrt{\kappa}}{1+\sqrt{\kappa}}$, $\kappa = \frac{\sigma_{\max}(\mathbf{B})}{\sigma_{\min}(\mathbf{B})}$, where $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ denote the maximum and minimum singular values of an input matrix, respectively.
while $n < n_{\max}$ do
    $n \leftarrow n + 1$
    $\mathbf{t}^{(n)} \leftarrow \max\left\{\boldsymbol{\Theta}\boldsymbol{\vartheta}^{(n-1)} + \mathbf{z}, \mathbf{0}\right\}$
    $\boldsymbol{\vartheta}^{(n)} \leftarrow \mathbf{t}^{(n)} + \psi\left(\mathbf{t}^{(n)} - \mathbf{t}^{(n-1)}\right)$
end while
$\bar{\mathbf{u}} \leftarrow \mathbf{y} - \mathbf{B}\mathbf{t}$
```

and $\hat{t}_l = 0$ otherwise. Having the approximate solution $\hat{\mathbf{t}}$, the corresponding precoded vector can simply be computed by replacing $\hat{\mathbf{t}}$ in (4.6). In what follows, this improved closed-form SLP solution is referred to as ICF-SLP.

4.5 Computational Complexity Analysis

In this section, we evaluate the computational complexities of the proposed CF-SLP and ICF-SLP designs and compare them with a benchmark iterative NNLS algorithm. As our benchmark for comparison, we consider the accelerated projected gradient descent (APGD) algorithm [135]. The pseudocode of the APGD algorithm solving the NNLS problem (4.5) within a limited number of iterations is given in Algorithm 2. We express the worst-case complexities of the SLP algorithms in terms of the number of floating point operations (FLOPs). For an iterative algorithm, the number of FLOPs translates to the required number of arithmetic operations until the terminating condition is met.

The main loop of the APGD algorithm is preceded by an initialization step performing two matrix multiplications and one singular value decomposition (SVD) with complexity orders of $N_{\mathrm{u}}^2 N_{\mathrm{t}}$, $N_{\mathrm{u}} N_{\mathrm{t}}$ and N_{u}^3, respectively. Within the main loop, the per-iteration complexity is dominated by a matrix multiplication of order N_{u}^2. To be more accurate, the complexity of the APGD algorithm depends also on the convergence specifications, e.g., the condition number of $\mathbf{B}$; however, we consider only those complexity terms directly relating to the problem size. On the other hand, for the ICF-SLP design, the dominant computation costs in (4.21), (4.22) and (4.23) come from $2N_{\mathrm{u}}$ vector multiplications and two matrix pseudo-inversions, resulting in computational complexities of order $N_{\mathrm{u}} N_{\mathrm{t}}$ and $N_{\mathrm{t}}(L_1^2 + L_2^2)$, respectively. Note that the CF-SLP design can also be implemented in an equivalent way using (4.21) and (4.23), and therefore, we assess the complexity of this method based on the ICF-SLP design.

TABLE 4.1: Actual complexity in FLOPs for different SLP designs.

Design	Actual Complexity (FLOPs)
APGD	$24N_{\mathrm{t}}N_{\mathrm{u}}^2 + 16N_{\mathrm{u}}^3 + 12N_{\mathrm{t}}N_{\mathrm{u}} - 2N_{\mathrm{u}}^2 - 3N_{\mathrm{u}} + (8N_{\mathrm{u}}^2 + 6N_{\mathrm{u}})(1/\sqrt{\epsilon})$
CF-SLP	$16N_{\mathrm{t}}N_{\mathrm{u}} + 10(N_{\mathrm{t}} + N_{\mathrm{u}}) + (8N_{\mathrm{t}} + 1)L_1^2 + 22L_1^3 + 2N_{\mathrm{t}}L_1$
ICF-SLP	$16N_{\mathrm{t}}N_{\mathrm{u}} + 10(N_{\mathrm{t}} + N_{\mathrm{u}}) + (8N_{\mathrm{t}} + 1)(L_1^2 + L_2^2) + 22(L_1^3 + L_2^3) + 2N_{\mathrm{t}}(L_1 + L_2)$

TABLE 4.2: Dominating complexity order for different SLP designs.

Design	Dominating Complexity Order
APGD	$N_{\mathrm{u}}^2.\, \mathcal{O}\left(N_{\mathrm{u}} + N_{\mathrm{t}}\right) + \mathcal{O}\left(N_{\mathrm{u}}^2\right)(1/\sqrt{\epsilon})$
CF-SLP	$N_{\mathrm{t}}.\, \mathcal{O}\left(N_{\mathrm{u}} + L_1^2\right) + \mathcal{O}(L_1^3)$
ICF-SLP	$N_{\mathrm{t}}.\, \mathcal{O}\left(N_{\mathrm{u}} + L_1^2 + L_2^2\right) + \mathcal{O}(L_1^3 + L_2^3)$

In Table 4.1 and 4.2, we summarize the actual complexities (in terms of the number of FLOPs) and the dominating complexity orders of different SLP designs, where by dominating complexity order, we mean the limiting order of complexity as $N_{\mathrm{t}}, N_{\mathrm{u}} \to \infty$. The number of FLOPs in Table 4.1 are calculated based on the complexities of basic matrix/vector operations provided in [139–141]. Note, further, that the reported complexity for the APGD-based design corresponds to an ϵ-optimal solution. Due to the sparsity-promoting nature of the NNLS problem [142], in practice, we usually have $L_2 \leq L_1 \ll 2N_{\mathrm{u}}$. Based on this observation and the results of Table 4.1 and 4.2, we can conclude that both CF-SLP and ICF-SLP techniques can reduce the computation cost of the SLP design compared to the case where the design problem is solved for optimality. Interestingly, even the complexity of the initialization step in the APGD algorithm without performing any iterations is higher than those of the proposed CF-SLP and ICF-SLP solutions.

4.6 Simulation Results

In this section, we provide some simulation results to evaluate and compare the performances of various approaches (with different complexities) to the SINR-constrained SLP power minimization design. We also compare the results with those of the ZF and the optimal block-level power minimization precoding schemes [1]. Our simulation setup is as follows. We consider an MU-MIMO downlink system, where all the users have equal target SINRs, i.e., $\gamma_i \triangleq \gamma$ for $i = 1, 2, ..., N_{\mathrm{u}}$. We define $N_{\mathrm{t}}/N_{\mathrm{u}} \triangleq \beta$ and assume a unit noise variance $\sigma_i^2 = 1$ at the receiver of any user $i \in \{1, 2, ..., N_{\mathrm{u}}\}$. The channel vectors $\{\mathbf{h}_i\}_{i=1}^{N_{\mathrm{u}}}$ are independently generated following a standard circularly symmetric complex Gaussian distribution as $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. The maximum number of iterations for the APGD algorithm is set to be $n_{\max} = 25$. The results are all averaged over 10^3 channel

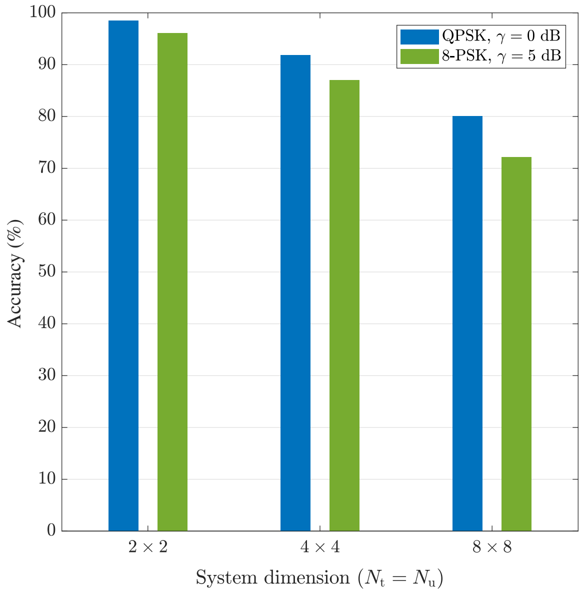
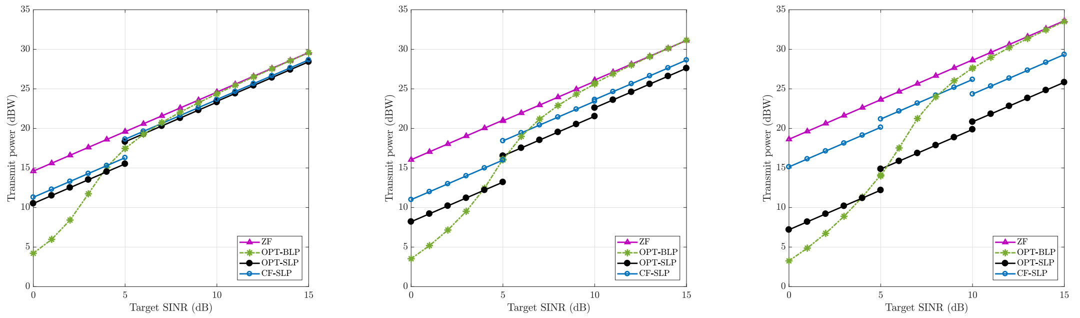


FIGURE 4.1: Accuracy of the CF-SLP solution for different system dimensions.

realizations, each consisting of 10^3 symbols. Throughout this section, we refer to the precoding schemes of interest as:

- ZF-SLP: symbol-level ZF, assuming $\mathbf{t} = \mathbf{0}$ in (4.6)
- OPT-BLP: optimal block-level precoding [1]
- OPT-SLP: optimal solution to (4.2)
- NNLS-SLP: solving (4.5) via APGD algorithm
- CF-SLP: closed-form SLP design in Section 4.4.1
- ICF-SLP: Improved closed-form SLP design in Section 4.4.2

The accuracy of the CF-SLP solution is plotted in Fig. 4.1 for two modulation schemes, namely, QPSK and 8-PSK, where the measure of accuracy is the number of correctly approximated elements in $\hat{\Lambda}$ as compared to the optimal solution. It can be seen that for a 2×2 system, the CF-SLP design can achieve an accuracy of at least 95%. This accuracy drops to $87 - 91\%$ and $70 - 80\%$, respectively, for a 4×4 and 8×8 system. However, notice that the CF-SLP design always guarantees that the DPCIR constraints are met, and therefore, it does not degrade the symbol error rate performance. The results also show that the CF-SLP technique performs more accurately for higher-order PSK modulation schemes.



(a) (b) (c)

FIGURE 4.2: Comparison of the total transmit power resulted from different precoding schemes versus target SINR with (a) $N_\mathrm{t} = N_\mathrm{u} = 4$; (b) $N_\mathrm{t} = N_\mathrm{u} = 8$; (c) $N_\mathrm{t} = N_\mathrm{u} = 16$.

We plot the total transmission power as a function of the target SINR for three system dimensions with $N_\mathrm{t} = N_\mathrm{u} = 4,\ 8$ and 16 in Fig. 4.2. In the depicted range of SINR, three different modulation schemes, namely, QPSK, optimized 8-ary [124], and 16-QAM, are respectively employed in the intervals 0-5, 5-10 and 10-15 dB. As it can be seen, in a 4×4 system, the CF-SLP design consumes almost the same amount of power as that of the OPT-SLP scheme. The loss due to the CF-SLP design's sub-optimality is not significant, with a maximum loss of 0.8 dBW for QPSK. This loss is larger for 8×8 and 16×16 systems, as demonstrated in Fig. 4.2. This can be justified as the possibility of having more incorrect approximations in $\hat{\Lambda}$ with respect to Λ grows with enlarging the problem size. Note, however, that the SLP scheme shows higher performance gains for larger system dimensions, e.g., the CF-SLP design offers gains up to 5 dBW compared to the ZF precoder for a 16×16 system. At low target SINRs, the OPT-BLP technique has the lowest transmit power among all the other precoding schemes, but in fact, this reduction in the transmit power is obtained at the cost of a degraded symbol error rate. On the other hand, for high target SINR values, the OPT-BLP scheme is equivalent to the ZF precoding.

In Fig. 4.3, we plot the total transmission power versus the target SINR for the precoding schemes of interest, where three different modulations, namely QPSK, 8-PSK and 16-QAM, are used in the SINR intervals $0-6$, $6-12$ and $12-18$ dB, respectively. The results correspond to a fully-loaded system with $N_\mathrm{t} = N_\mathrm{u}$. It can be seen that the ICF-SLP design improves the accuracy of the approximate solution by up to 3 dB, compared to its simpler counterpart, i.e., CF-SLP. Furthermore, the ICF-SLP design outperforms the NNLS-SLP method via the APGD algorithm with $n_\mathrm{max} = 25$. Our observations show that both the methods have nearly the same complexity in the considered range of N_u. Another promising observation from Fig. 4.3 is that the ICF-SLP design performs well close to the OPT-SLP scheme, but with far less computational complexity, as we will see next.

In another set of simulations for an under-loaded system with $\beta = 6/5$, we evaluate

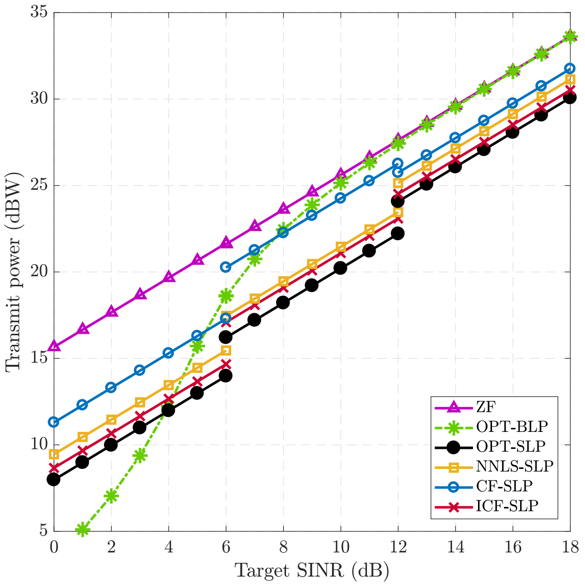


FIGURE 4.3: Transmit power versus target SINR with $N_\mathrm{t} = N_\mathrm{u} = 8$.

the performance and complexity of different approaches to the SLP optimization in (4.2). The results are shown in Fig. 4.4 with two vertical axes as a function of the number of users N_u, where the same line types and markers as those in the legend refer to the right axis but with a different color. The optimal SLP solution is obtained by solving the NNLS problem (4.5) via the "lsqnonneg" function of the MATLAB software, which is based on the Lawson and Hanson active set method. As it can be seen, the CF-SLP design's performance noticeably degrades with increasing N_u, whereas the ICF-SLP design shows a competitive performance in transmit power as compared to the OPT-SLP, even for large system dimensions. Remarkably, the optimality gap of the approximate ICF-SLP design with $N_\mathrm{u} = 100$ is just 0.15 dBW. This improvement looks more promising when we also take the design complexities into account; see Table 4.1 and 4.2. It can be verified that the time complexity results in Fig. 4.4 are in accordance with the analytical discussion in Section 4.5. Comparing the ICF-SLP and the NNLS-SLP methods, we see that the latter method's complexity grows at a higher rate, which is shown to be proportional to $\mathcal{O}(N_\mathrm{t}N_\mathrm{u}^2)$ in the limiting case. This may suggest a performance-complexity tradeoff; however, notice that with $\eta_{\max} = 25$, the dominating complexity order of the APGD algorithm in the large system limit comes from the initialization step, which is higher than the whole computation cost of the ICF-SLP design.

In Table 4.3, we compare the precoding complexity in terms of the average execution time per symbol period (the time values are computed via the relevant functions of MATLAB). As for the ZF and OPT-BLP schemes, the precoding matrix is multiplied by the users' symbol vector at every symbol period. The precoding matrix computation,

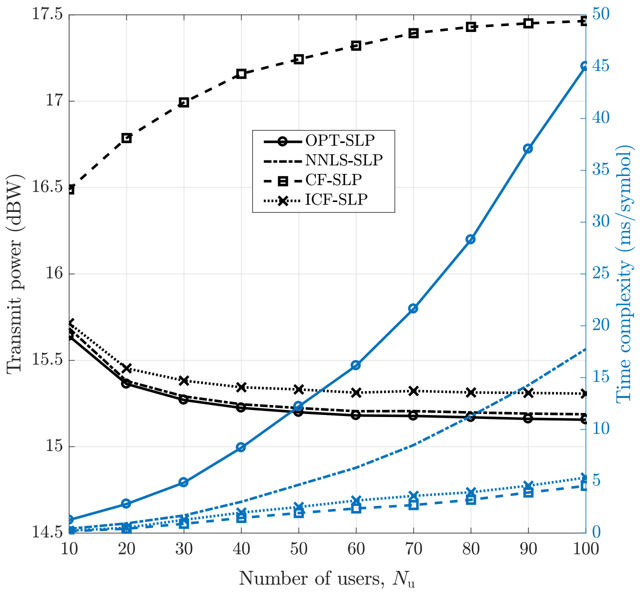


FIGURE 4.4: Transmit power and time complexity versus number of users with $\beta = 6/5$.

which is typically updated once per channel coherence block, also accounts for the per-symbol execution times (assuming 100 symbols within each coherence block). The CF-SLP method consists of computing $\mathbf{Q}$ and $\mathbf{p}$ and then solving (4.20). On the other hand, solving the convex problem (4.2) accounts for the OPT-SLP scheme's execution time. The numerical results show that the CF-SLP and ICF-SLP methods can reduce the SLP design complexity by orders of 10^3. The ICF-SLP design has a slightly increased computation time compared to the CF-SLP method, while it significantly improves the performance. Moreover, as expected, the CF-SLP and ICF-SLP execution times are larger (by orders of 10) compared to the ZF precoding but are comparable to those of the OPT-BLP scheme. These results indicate a performance-complexity tradeoff, particularly for large system dimensions.

4.7 Conclusions

Due to the high per-symbol computation cost, solving the SLP design problem for the exact solution may lead to an impractical transmitter complexity. To address this challenge, in this chapter, we proposed two computationally-efficient methods to approximately solve the SLP power minimization problem with CI and SINR constraints. This is done by first simplifying the original formulation and reformulating it as an NNLS design, and then discussing the simplified problem's optimality via the KKT conditions. The analyses helped us to derive a closed-form approximate SLP design, namely, CF-SLP. The CF-SLP design performs quite close to the optimal SLP scheme in systems

Table 4.3: Execution time of different precoding schemes.

Modulation	Dimension	Execution time (ms/symbol)				
		ZF	OPT-BLP	OPT-SLP	CF-SLP	ICF-SLP
QPSK	$(N_\mathrm{t}, N_\mathrm{u}) = (4, 4)$	0.0064	0.0249	573.8417	0.0671	0.0699
	$(N_\mathrm{t}, N_\mathrm{u}) = (8, 8)$	0.0076	0.0779	537.4583	0.1222	0.1321
	$(N_\mathrm{t}, N_\mathrm{u}) = (16, 16)$	0.0113	0.3098	595.9375	0.2388	0.2482
Optimized 8-ary	$(N_\mathrm{t}, N_\mathrm{u}) = (4, 4)$	0.0060	0.0215	588.1708	0.0633	0.0777
	$(N_\mathrm{t}, N_\mathrm{u}) = (8, 8)$	0.0080	0.0771	532.4833	0.1217	0.1360
	$(N_\mathrm{t}, N_\mathrm{u}) = (16, 16)$	0.0114	0.3627	584.1917	0.2223	0.2472
16-QAM	$(N_\mathrm{t}, N_\mathrm{u}) = (4, 4)$	0.0058	0.0194	554.3417	0.0582	0.0856
	$(N_\mathrm{t}, N_\mathrm{u}) = (8, 8)$	0.0080	0.0641	533.9958	0.1139	0.1370
	$(N_\mathrm{t}, N_\mathrm{u}) = (16, 16)$	0.0097	0.2586	518.7500	0.1760	0.2784

with a relatively small number of users. We further improved this approximate solution by applying an extra validation step to the design process and named the improved design ICF-SLP. Our numerical and simulation results indicated that this improved solution substantially reduces the loss with respect to the optimal solution, particularly in the large system regime. Furthermore, the ICF-SLP design showed competitive performance compared to the SLP solution obtained from the iterative APGD algorithm, but with reduced time complexity. In comparison with conventional block-level precoding schemes, our results show that both the CF-SLP and ICF-SLP methods outperform the ZF precoder in all scenarios and the optimal power minimizer block-level precoder at high target SINRs. According to the results, we conclude that the CF-SLP and ICF-SLP designs can successfully relieve the prohibitive computation cost of the SLP design. Further, they are promising alternatives (with a comparable complexity) for the block-level precoding schemes, especially in the high SINR regime.

Chapter 5

Computationally-Efficient Symbol-Level Precoding–Part II: Implementation

In this chapter, we develop and validate a low-complexity FPGA design for the SLP scheme in the downlink of MU-MIMO communication systems. The considered SLP design, in its original form, aims to minimize the total transmit power while satisfying the CI constraint as well as a given target SINR for each user. Such a design criterion leads us to an NNLS problem. Considering the fact that a symbol-level precoder redesigns the transmit signal specifically for any given set of users' intended symbols, solving this NNLS problem imposes a relatively high computational complexity on the system in every symbol period. To alleviate this high computation cost, we aim to reduce the per-symbol complexity of the SLP scheme by developing an approximate yet computationally-efficient closed-form solution. The proposed solution allows us to achieve a high symbol throughput in real-time implementations. The work of this chapter builds on the CF-SLP method presented in Chapter 4, and thus, the resulting design is constellation-independent which makes it appropriate for seamless handling of adaptive coding and modulation (ACM) schemes. To develop the FPGA design, we express the proposed solution in an algorithmic way and translate it to hardware description language (HDL). We then optimize the processing to accelerate the performance and generate the corresponding intellectual property (IP) core. We provide the synthesis report for the generated IP core, including performance and resource utilization estimates and interface descriptions. To validate our design, we simulate an uncoded transmission scheme over a downlink multiuser channel using the LabVIEW software, where the SLP IP core is implemented as a clock-driven logic (CDL) unit. Our simulation results show that a throughput of 100 Mega symbols per second per user can be achieved for a fully-loaded 4 × 4 system with QPSK modulation via the HDL design of the proposed approximate SLP solution. We further use the MATLAB software to produce numerical results for the conventional ZF precoding and the optimal SLP technique as benchmarks for comparison. Thereby, it is shown that the proposed low-complexity FPGA implementation offers an improvement of up to 50 percent in power efficiency compared to the

ZF precoding. Remarkably, it enjoys the same per-symbol complexity order as that of the ZF technique. We also evaluate the loss of the real-time SLP design, introduced by the algebraic approximations and arithmetic inaccuracies, with respect to the optimal scheme.

5.1 Introduction

Following our analytical discussions and derivations on low-complexity SLP design in Chapter 4, we are further interested in evaluating the possibility of using the proposed techniques in real-time applications. In the relevant literature, some studies have addressed efficient hardware demonstrations of the existing low-complexity SLP techniques, e.g., [113, 114]. Furthermore, in [50], the authors propose a computationally-efficient approximate solution to the power minimization SLP problem with strict phase constraints on the received signals and demonstrate an FPGA-accelerated design of this solution in [115], indicating that it is capable of providing a high symbol throughput in a real-time operation mode.

In this chapter, we specifically focus on the closed-form approximate solution proposed in Chapter 4, namely, CF-SLP, which is obtained for the power minimization SLP problem. Accordingly, the main contributions of this chapter are as follows:

- We further simplify this solution using some intermediate approximation steps and derive a new solution which has lower computational complexity. The approximations are mainly introduced to reduce the computation cost of the SLP design. This simplification further facilitates the design of a low-complexity algorithm operating in a real-time mode.

- We show through analytical evaluation of the computational complexity that the proposed approximate SLP solution has the same per-symbol complexity order as that of the conventional ZF precoding.

- To validate our design, we target FPGA implementation of the proposed SLP algorithm. First, we express the algorithm in C++ language and then convert it to hardware description language (HDL). The HDL implementation enables us to generate the intellectual property (IP) core targeted for a specific FPGA device. We analyze and compare two different cases: the original non-optimized HDL design and the case where the processing is optimized through function pipelining, loop unrolling and array partitioning. This indicates how optimizing the HDL design can accelerate the performance. We also provide the synthesis results for the generated IP core. In particular, the timing and latency estimates, the FPGA resource utilization ratios, and the register-transfer level (RTL) I/O ports specifications are reported.

- The synthesis and implementation results show that the proposed FPGA design is able to provide a high throughput of 100 Mega symbols per second per user for a 4×4 system with QPSK signaling. Furthermore, numerical results are obtained by simulating a multiuser downlink system in the LabVIEW and MATLAB environments and applying different precoding techniques. Our results show that the proposed low-complexity HDL implementation of the SLP algorithm substantially outperforms the ZF technique in terms of power efficiency.

The remainder of this chapter is organized as follows. In Section 5.2, we revisit the proposed closed-form solution for the power minimization SLP problem with distance-preserving CI constraints and develop another simplified SLP design algorithm. In Section 5.4, we explain the HDL design and optimization steps for the proposed algorithm and report some performance estimates for the real-time FPGA implementation. In Section 5.5, we evaluate our HDL design by presenting the results of simulation tests. Finally, we conclude the chapter in Section 5.6.

5.2 Overview of the CF-SLP design

In Chapter 4, we have shown that the DPCIR-based SLP power minimization design can be expressed in a standard NNLS form as

$$\min_{\mathbf{t}\succeq\mathbf{0}} \quad \|\mathbf{Bt}-\mathbf{y}\|^2. \tag{5.1}$$

where $\mathbf{B} \triangleq \mathbf{H}^\dagger\mathbf{A}^{-1}\mathbf{W}$ and $\mathbf{y} \triangleq -\mathbf{H}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s}$, and the same definitions as in Section 3.5 are used. Having the solution to (5.1), the optimal precoded vector is immediately given by $\mathbf{u}^* = \mathbf{Bt}^* - \mathbf{y}$. To solve the NNLS problem (5.1), in Section 4.4.1, we introduced the CF-SLP technique providing an approximate yet computationally-efficient solution. Let us describe the CF-SLP design in a more algorithmic way. This technique is composed of two steps as follows:

i. Obtain an estimate of the support as

$$\hat{\Lambda} = \left\{ l \mid l = 1, 2, ..., 2N_\mathrm{u},\ \mathbf{y}^\mathrm{T}\mathbf{b}_l \geq 0 \right\}, \tag{5.2}$$

where $\mathbf{b}_l$ denotes the lth column of $\mathbf{B}$.

ii. Let $L \triangleq |\hat{\Lambda}|$ denote the length of the estimate support set. Build a $2N_\mathrm{u} \times L$ matrix $\mathbf{B}_{\hat{\Lambda}}$ consisting of those columns in $\mathbf{B}$ that are indexed in $\hat{\Lambda}$ and let the columns of $\mathbf{B}_{\hat{\Lambda}}$ be indexed as $\mathbf{b}_l$ where $l \in \hat{\Lambda}$. Then, calculate an approximate solution by solving a reduced system of linear equations as

$$\hat{t}_l = \left\{ \left[\mathbf{B}_{\hat{\Lambda}}^\dagger \mathbf{y} \right]_l \right\}_+, \tag{5.3}$$

and $\hat{t}_l = 0$ otherwise, where $[\cdot]_l$ denotes the element that corresponds to the lth variable in $\mathbf{t}$, and operation $\{\cdot\}_+$ stands for $\max\{\cdot, 0\}$.

This approximate closed-form solution involves a matrix pseudo-inverse operation as in (5.3), which is computationally costly in practice. In the sequel, we propose an approximate alternative operation to eliminate the need for computation of this matrix pseudo-inverse.

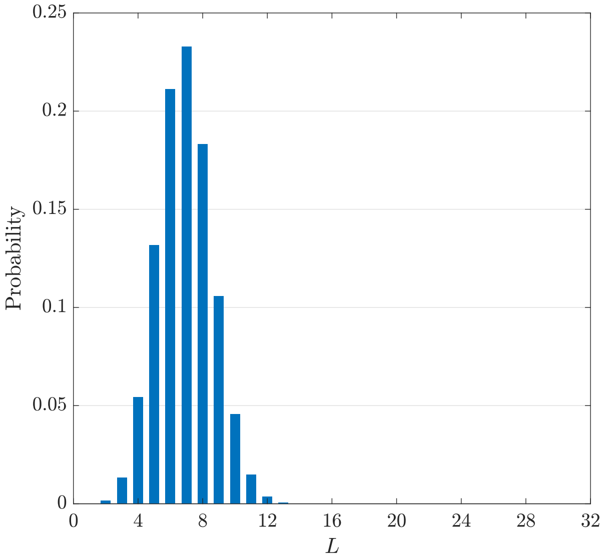


FIGURE 5.1: Experimental probability mass function of L.

5.3 Low-Complexity Implementation of CF-SLP

Our experiments show that, on average, only a few number of inner products $\mathbf{y}^{\mathrm{T}}\mathbf{b}_l$ out of a total number of $2N_{\mathrm{u}}$ are non-negative, and therefore, we usually have $L \ll 2N_{\mathrm{u}}$. Consequently, the matrix $\mathbf{B}_{\hat{\Lambda}}$ has more rows than columns. In Fig. 5.1, we support this observation by plotting the empirical probability mass function of L which is obtained by averaging the realizations of L from 10^6 trials (10^3 symbol periods over 10^3 channel realizations) in a scenario with $N_{\mathrm{t}} = N_{\mathrm{u}} = 16$. It can be seen from Fig. 5.1 that $\Pr\{L \leq 3N_{\mathrm{u}}/4\} \approx 0.99$, i.e., the length of the estimated support is, with high probability, smaller than 3/4 of the total number of elements. Based on this observation, we assume that the columns $\{\mathbf{b}_l \mid l \in \hat{\Lambda}\}$ are mutually orthogonal. Such an assumption leads us to the following approximation:

$$\left(\mathbf{B}_{\hat{\Lambda}}\mathbf{B}_{\hat{\Lambda}}^{\mathrm{T}}\right)^{-1} \approx \operatorname{diag}\left(\left\{\frac{1}{\|\mathbf{b}_l\|^2} \mid l \in \hat{\Lambda}\right\}\right). \tag{5.4}$$

As a result, the pseudo-inverse of matrix $\mathbf{B}_{\hat{\Lambda}}$ can be approximated as

$$\mathbf{B}_{\hat{\Lambda}}^{\dagger} = \mathbf{B}_{\hat{\Lambda}}^{\mathrm{T}}\left(\mathbf{B}_{\hat{\Lambda}}\mathbf{B}_{\hat{\Lambda}}^{\mathrm{T}}\right)^{-1} \approx \mathbf{B}_{\hat{\Lambda}}^{\mathrm{T}} \operatorname{diag}\left(\left\{\frac{1}{\|\mathbf{b}_l\|^2} \mid l \in \hat{\Lambda}\right\}\right). \tag{5.5}$$

Therefore, by plugging (5.5) into $\mathbf{B}_{\hat{\Lambda}}^{\dagger}\mathbf{y}$, we obtain

$$\hat{t}_l = \begin{cases} \mathbf{y}^{\mathrm{T}}\mathbf{b}_l/\|\mathbf{b}_l\|^2 & l \in \hat{\Lambda}, \\ 0 & l \notin \hat{\Lambda}. \end{cases} \tag{5.6}$$

Given the approximate solution $\hat{\mathbf{t}} = [\hat{t}_1, \hat{t}_1, ..., \hat{t}_{2N_{\mathrm{u}}}]^{\mathrm{T}}$, the vector of precoded transmit signal can be calculated as mentioned earlier. The pseudo-code of this low-complexity approximate solution is summarized in Algorithm 3. It is important to note that the non-negative constraints $\mathbf{t} \succeq \mathbf{0}$ are all satisfied by the SLP design in Algorithm 3. This implies that the approximation (5.5) does not lead to violation of the SNR constraints and the users' SNR requirements are guaranteed under the proposed approximate SLP solution.

Recall from Section 3.5 that the matrices $\mathbf{A}^{-1}$ and $\mathbf{W}$ have the following structures:

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{N_{\mathrm{u}}}^{-1} \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{N_{\mathrm{u}}} \end{bmatrix} \otimes \mathbf{I}_2.$$

To facilitate the design process, in Algorithm 3, we incorporate the diagonal elements of the matrix $\mathbf{W}$ into $\mathbf{A}^{-1}$. To do so, for any $i \in \{1, 2, ..., N_{\mathrm{u}}\}$, the matrix $\mathbf{A}_i^{-1}$ is constructed using the following criteria:

- If $\mathbf{s}_i$ is an outer constellation symbol, we obtain $\mathbf{a}_{i,1}$ and $\mathbf{a}_{i,2}$ (as defined in Section 3.3) to build $\mathbf{A}_i$, and compute the matrix $\mathbf{A}_i^{-1}$.
- If $\mathbf{s}_i$ is an inner constellation symbol, we set $\mathbf{A}_i^{-1} = \mathbf{0}$.

We further note that in Algorithm 3, a simple lookup method is used to avoid calculation of matrix $\mathbf{A}^{-1}$ at each symbol period. More precisely, given the modulation scheme, we extract all the possible realizations for a sub-matrix $\mathbf{A}_i$, calculate their inverses and store them in a lookup table, where the total number of possible realizations is equal to the modulation order. At the time of implementation, each sub-matrix $\mathbf{A}_i^{-1}$ can be read from the lookup table with respect to the given symbol $\mathbf{s}_i$, and the entire matrix $\mathbf{A}$ can be constructed accordingly. Moreover, for ease of implementation, the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}$ are incorporated into the symbol vector $\mathbf{s}$, as we will see in Section 5.4.

The proposed solution in Algorithm 3 consists of a number of loops with known and constant number of iterations, each of which includes some basic arithmetic operations, e.g., addition and multiplication. We report in Table 5.1 the actual arithmetic complexity of Algorithm 3, including the separate complexity of each computation step as well as the overall complexity, in terms of the number of floating-point operations (FLOPs). It follows from Table 5.1 that Algorithm 3 has a dominating complexity order of $\mathcal{O}(N_{\mathrm{t}}N_{\mathrm{u}})$, in the limiting case where $N_{\mathrm{t}}, N_{\mathrm{u}} \rightarrow \infty$, and therefore, it enjoys the exact

Algorithm 3 Approximate low-complexity SLP solution

```
input : $\mathbf{H}^{\dagger}, \mathbf{\Sigma}, \mathbf{\Gamma}, \mathbf{s}$
output : $\mathbf{u}$
$\mathbf{A}^{-1} \leftarrow \text{lookup}(\mathbf{s})$                ▷ Build matrix $\mathbf{A}^{-1}$
$\mathbf{B} \leftarrow \mathbf{H}^{\dagger}\mathbf{A}^{-1}$                ▷ Build matrix $\mathbf{B}$
$\mathbf{y} \leftarrow -\mathbf{H}^{\dagger}\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s}$                ▷ Build vector $\mathbf{y}$
for $l = 1$ to $2N_{\text{u}}$ do                ▷ Collect column-wise norms
    $c_l \leftarrow \mathbf{b}_l^{\text{T}}\mathbf{b}_l$
end for
for $l = 1$ to $2N_{\text{u}}$ do
    $d_l \leftarrow \mathbf{y}^{\text{T}}\mathbf{b}_l$
    if $d_l \geq 0$ then                ▷ Estimate support
        $t_l \leftarrow \mathbf{y}^{\text{T}}\mathbf{b}_l / c_l$                ▷ Compute vector $\mathbf{t}$
    else
        $t_l \leftarrow 0$
    end if
end for
$\mathbf{u}^{*} \leftarrow \mathbf{B}\mathbf{t} - \mathbf{y}$                ▷ Compute vector $\mathbf{u}$
```

same per-symbol complexity order as that of the ZF precoding technique. Based on this comparison, we state that the proposed SLP solution has low computational complexity, and hence, is suitable for real-time implementation.

TABLE 5.1: Actual arithmetic complexity of Algorithm 3.

Computation	Number of iterations	FLOPs
$\mathbf{H}^{\dagger}\mathbf{A}^{-1}$	1	$12N_{\text{t}}N_{\text{u}}$
$\mathbf{H}^{\dagger}\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{s}$	1	$8N_{\text{t}}N_{\text{u}} + 4N_{\text{u}} - 2N_{\text{t}}$
$\mathbf{b}_n^{\text{T}}\mathbf{b}_n$	$2N_{\text{u}}$	$8N_{\text{t}}N_{\text{u}} - 2N_{\text{u}}$
$\mathbf{y}^{\text{T}}\mathbf{b}_n/c_n$	$2N_{\text{u}}$	$8N_{\text{t}}N_{\text{u}}$
$\mathbf{B}\mathbf{t} - \mathbf{y}$	1	$8N_{\text{t}}N_{\text{u}}$
Overall		$44N_{\text{t}}N_{\text{u}} + 2N_{\text{u}} - 2N_{\text{t}}$

5.4 FPGA Design

To enable implementation of the proposed low-complexity SLP solution, we design the IP core using the Xilinx Vivado HLS tool. The Vivado HLS tool transforms a C specification, such as C, C++, or SystemC, into a register-transfer level (RTL) implementation that can be synthesized into Xilinx programmable devices. In this work, we have used version 2017.3 of the Xilinx Vivado HLS software and designed the IP core for Xilinx Kintex-7 xc7k410tffv900-2 FPGA part.

TABLE 5.2: Interface specifications of the IP core.

RTL port	Direction	Bit width	Protocol	Description
ap_clk	in	1	ap_ctrl_hs	Primary design clock
ap_rst_n	in	1	ap_ctrl_hs	Interface reset (active-low)
ap_start	in	1	ap_ctrl_hs	Block execution control (active-high)
ap_done	out	1	ap_ctrl_hs	Complete-transaction indicator (active-high)
ap_idle	out	1	ap_ctrl_hs	Operating/idle indicator (active-high)
ap_ready	out	1	ap_ctrl_hs	Ready-for-new-inputs indicator (active-high)
pinvH_V	in	$4BN_\mathrm{t}N_\mathrm{u}$	ap_none	Real-valued pseudo-inverse of the channel matrix
s_V_TDATA	in	$2BN_\mathrm{u}$	axis	Real-valued vector of the users' symbols
s_V_TVALID	in	1	axis	Data input valid
s_V_TREADY	out	1	axis	Data input ready
u_V_TDATA	out	$2BN_\mathrm{t}$	axis	Real-valued vector of precoded transmit signal
u_V_TVALID	out	1	axis	Data output valid
u_V_TREADY	in	1	axis	Data output ready

To generate the IP core, we have translated the algorithmic description of Algorithm 3 into C++ language. To achieve an accelerated performance and higher throughputs, we have optimized the code through many techniques, such as pipelining the functions, unrolling the loops, and partitioning the arrays. Pipelining and unrolling both improve the hardware function's performance by exploiting the parallelism between function and loop iterations. In particular, pipelining allows the operations in a function/loop to be implemented in a concurrent manner and unrolling creates multiple copies of the loop body and adjusts the loop iteration counter accordingly. These techniques have been applied to the design by adding the so-called "directives" into the C++ code. In the following, we refer to the design used in this work by applying the above techniques as the optimized HDL design. On the other hand, the original design without applying any of the above optimization techniques is referred to as the non-optimized design. Later in this section, we present the resource utilization and performance estimates for both non-optimized and optimized HDL implementations to emphasize how the design benefits from such code optimizations. We have further utilized the Vivado HLS matrix algebra library for efficient calculation of matrix multiplications. The C++ code has then been synthesized using the Vivado HLS tool, and the RTL implementation has been extracted as an IP catalog. The schematic block design of the IP core generated for a system with $(N_\mathrm{t}, N_\mathrm{u}) = (4, 4)$ is depicted in Fig. 5.2. The design takes matrix $\mathbf{H}^\dagger$ and vector $\mathbf{s}$ as data inputs to execute Algorithm 3. These two inputs are shown as ports pinvH_V and s_V_TDATA in Fig. 5.2. Note that we do not consider dedicated inputs for matrices $\mathbf{\Sigma}$ and $\mathbf{\Gamma}$, and instead, we absorb the corresponding noise variances and target SINRs into

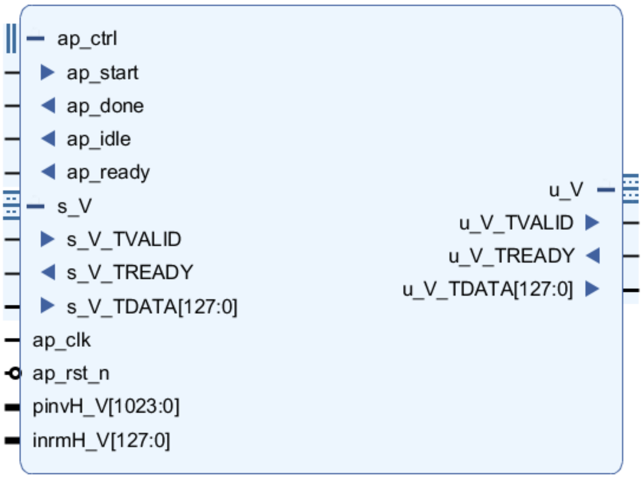


FIGURE 5.2: IP block design of the low-complexity SLP technique.

the input vector $\mathbf{s}$. The only output of the HDL design is the precoded vector $\mathbf{u}$ which is placed on port u_V_TDATA of the IP core.

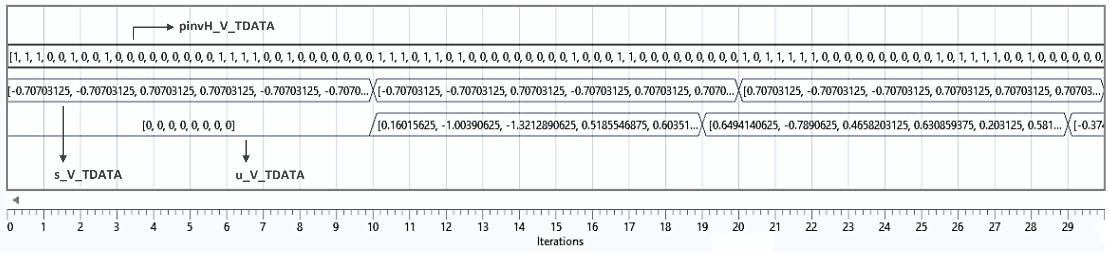
TABLE 5.3: Structure of the RTL data ports.

Data port	**Format**
s_V_TDATA	$\mathrm{Re}(s_1) \mid \mathrm{Im}(s_1) \mid \mathrm{Re}(s_2) \mid \mathrm{Im}(s_2) \mid \cdots \mid \mathrm{Re}(s_{N_\mathrm{u}}) \mid \mathrm{Im}(s_{N_\mathrm{u}})$
u_V_TDATA	$\mathrm{Re}(u_1) \mid \mathrm{Re}(u_2) \mid \cdots \mid \mathrm{Re}(u_{N_\mathrm{t}}) \mid \mathrm{Im}(u_1) \mid \mathrm{Im}(u_2) \mid \cdots \mid \mathrm{Im}(u_{N_\mathrm{t}})$

5.4.1 RTL I/O Ports Description

The designed IP core has a number of dedicated data I/O ports. In addition, a block-level I/O control handshake protocol has been added to control the RTL design independently of the data ports. We summarize the specifications and behavior of all the HDL I/O ports in Table 5.2. Note that the bit width of a data port is determined by the bit width of the fixed-point format, which is denoted by B. In this work, we adopt a 6.10 signed fixed-point format for the RTL design such that it represents the integer and fraction parts, respectively, by 5 and 10 bits, and the sign is specified by one bit. Therefore, the real and imaginary parts are allocated 16 signed bits each, resulting in a total number of 32 bits for a single complex value.

To have an efficient data transfer towards and from the IP core, we adopt an AXIS handshake protocol for the I/O data ports. The precomputed pseudo-inverse of the real-valued channel matrix feeds the input data port pinvH_V, and therefore, this port does not need a handshake signaling. The data on this port must be ready before signaling to the port s_V_TREADY. The real and imaginary parts of each element of matrix $\mathbf{H}^\dagger$



```
[1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0,
[-0.70703125, -0.70703125, 0.70703125, 0.70703125, -0.70703125, -0.7070... [-0.70703125, -0.70703125, 0.70703125, -0.70703125, 0.70703125, 0.7070... [0.70703125, -0.70703125, -0.70703125, 0.70703125, 0.70703125, 0.70703...
[0, 0, 0, 0, 0, 0, 0, 0] [0.16015625, -1.00390625, -1.3212890625, 0.5185546875, 0.60351... [0.6494140625, -0.7890625, 0.4658203125, 0.630859375, 0.203125, 0.581... [-0.37
```

Figure 5.3: Data flow of the IP core sampled in the LabVIEW environment.

are reshaped row by row into an array of length $4BN_\mathrm{t}N_\mathrm{u}$ bits. The first element of the first row starts at the the most significant bit, while the last element of the last row ends at bit 0. We further illustrate in Table 5.3 the formats of the data ports s_V_TDATA and u_V_TDATA. The s_V_TDATA port contains the elements of the symbol vector $\mathbf{s}$ in the order shown in Table 5.3, which are mapped to an array of length $2BN_\mathrm{u}$ bits. The real part of the first element starts at the most significant bit and the imaginary part of the last element ends at bit 0. The u_V_TDATA port, on the other hand, has a different format from that of the s_V_TDATA port. The imaginary parts of all the elements of the precoded vector $\mathbf{u}$ are concatenated and appended to the real parts of all the elements. The first element's real part starts with the most significant bit and the the last element's imaginary part ends at bit 0.

5.4.2 Resource Utilization and Timing Estimates

To design the IP core for the Kintex-7 xc7k410tffv900-2 FPGA device, we have set a target clock period (CP) of 10 nanosecond (ns), or equally, a 100 MHz clock rate. The estimate performance numbers, including timing and latency, produced by the C synthesis and implementation via the Vivado HLS tool are presented in Table 5.4 and Table 5.5 for both non-optimized and optimized designs, indicating that the required timing is perfectly met in both cases. In particular, the estimated timing performance after post-implementation of the IP core is shown to be 8.83 ns, which is well smaller than the target CP of the HDL design.

We further report, in Table 5.4 and Table 5.5, the latency and the initiation interval (II) estimates for the non-optimized and optimized HDL functions, where latency refers to the number of clock cycles required for the design to complete the current transaction and compute all the output values (i.e., the number of clock cycles between the input and the corresponding output), and the II is the number of clock cycles before the design can accept new input data. Comparing these two tables, we see that the non-optimized HDL design has a latency of 1493 clock cycles, whereas the optimized design can achieve a far smaller latency of 9 cycles. This significant improvement in throughput is brought by optimizing the code through, e.g., exploiting the parallelism between function and loop iterations. More precisely, the IP core has been optimized to complete a transaction in 9 cycles, which means that, upon receiving data on the s_V_TDATA port, the precoded vector is valid on the u_V_TDATA output port after 9 clock cycles. In the meantime, the IP core can accept a new input data per cycle and performs the next transactions in parallel to compute the corresponding output values. Hence, the design can produce an output every clock cycle, allowing the IP core to operate at a rate of 100 Mega symbols per second per user, as we will see in Section 5.5.

In Table 5.6 and Table 5.7, we present the estimated resource utilization on the Kintex-7 xc7k410tffv900-2 FPGA device, where the IP core is generated for two systems with $(N_\mathrm{t}, N_\mathrm{u}) = (2, 2)$ and $(N_\mathrm{t}, N_\mathrm{u}) = (4, 4)$. When comparing the non-optimized and optimized HDL designs, it can be seen that the latter design occupies more resources on the FPGA device. This originates from the well-known trade-off between area and performance in digital logic circuit design. More specifically, parallelization of functions and loops leads to higher throughputs, but it requires more resources to perform many concurrent operations. Nonetheless, the resource utilization estimates in Table 5.7 shows that the optimized design's total resource occupation is well below the available resource on this particular FPGA part for 2 × 2 and 4 × 4 systems. For larger system sizes, i.e., larger numbers of transmit antennas and users, one should either make a compromise between area and performance or use a more expensive FPGA with more available resources.

Table 5.4: Performance estimates of the non-optimized HDL design.

Timing/Clock period (ns)	
Target	10.00
C synthesis	8.63
Post-synthesis	5.18
Post-implementation	7.01
Latency (clock cycles)	
Latency	1493
Interval	1494

On the other hand, according to the utilization estimates in Table 5.7, for the 2 × 2 system, the design utilizes around 4% of the DSP blocks, 1% of the FFs, and 1% of the total LUTs that are available at this specific FPGA part, while for the 4 × 4 system, around 17% of the DSP blocks, 2% of the FFs, and 22% of the total LUTs are utilized by the design. This implies that, in general, the resource utilization ratios may not be linearly related to the system size. Roughly speaking, based on the estimates, it might be possible to support a larger system than the current design with this particular FPGA part or even use a cheaper FPGA with less available resources. For example, in the former case, the design might be able to treat several independent carriers or handle

Table 5.5: Performance estimates of the optimized HDL design.

Timing/Clock period (ns)	
Target	10.00
C synthesis	8.72
Post-synthesis	5.52
Post-implementation	8.83
Latency (clock cycles)	
Latency	9
Interval	1

larger systems on the same FPGA. Note, further, that the design's resource utilization does not depend on the constellation size (i.e., the modulation order). More precisely, having a larger constellation does not affect the design complexity but increases the size of the lookup table to form the matrix $\mathbf{A}$, as described in Section 5.2. Therefore, the same resource occupation estimates are valid also for larger signal constellations.

Table 5.6: Resource utilization of the non-optimized HDL design on the Xilinx Kintex-7 xc7k410tffv900-2 FPGA.

	2×2 system			**4×4 system**		
Resource	**DSP48E**	**FF**	**LUT**	**DSP48E**	**FF**	**LUT**
DSP	2	0	0	2	0	0
Expression	0	0	6038	0	0	11747
Instance	3	414	1522	3	6081	2248
Memory	0	26	4	0	410	37
Multiplexer	0	0	363	0	0	954
Register	0	1935	0	0	3670	0
Total	5	2375	7927	5	10161	14986
Available	1540	508400	254200	1540	508400	254200
Utilization (%)	0.3	0.5	3	0.3	1	5

Table 5.7: Resource utilization of the optimized HDL design on the Xilinx Kintex-7 xc7k410tffv900-2 FPGA.

	2 × 2 system			**4 × 4 system**		
Resource	**DSP48E**	**FF**	**LUT**	**DSP48E**	**FF**	**LUT**
DSP	68	0	0	72	0	0
Expression	0	0	1904	0	0	888
Instance	0	0	0	192	4224	54456
Memory	0	0	0	0	0	0
Multiplexer	0	0	62	0	0	62
Register	0	3590	768	0	8338	2560
Total	68	3590	2734	264	12562	57966
Available	1540	508400	254200	1540	508400	254200
Utilization (%)	4	0.7	1	17	2	22

5.4.3 Design Validation

In this subsection, we assess the performance accuracy of the designed IP core. For this purpose, we validate our design using the LabVIEW software. The generated IP core is transformed to a design block and then imported to the LabVIEW environment. The validation steps, which are performed for a $(N_{\mathrm{t}}, N_{\mathrm{u}}) = (4, 4)$ system, are described in the sequel.

The input port pinvH_V is fed with the pseudo-inverse of the channel matrix given as

$$\mathbf{H}^{\dagger} = \begin{bmatrix} 0.2880 + \mathrm{j}0.1221 & 0.1559 + \mathrm{j}0.5371 & -0.8774 - \mathrm{j}0.3437 & 0.1097 + \mathrm{j}0.3331 \\ 0.3085 + \mathrm{j}0.6187 & -0.7176 + \mathrm{j}0.0683 & 0.8212 - \mathrm{j}1.4356 & -0.6341 + \mathrm{j}0.4036 \\ 0.1790 - \mathrm{j}0.8406 & 0.6989 + \mathrm{j}0.8182 & -2.2538 + \mathrm{j}0.3444 & 0.4639 + \mathrm{j}0.6017 \\ 0.0961 - \mathrm{j}0.3560 & 0.3669 + \mathrm{j}0.3207 & -1.0475 + \mathrm{j}0.8989 & 0.6282 + \mathrm{j}0.0833 \end{bmatrix}, \tag{5.7}$$

and the symbol vector $\mathbf{s}$, taken from a normalized QPSK constellation set, is placed in order on the s_V_TDATA input port. We assume a unit noise variance and an equal target SINR of 0 dB for all the users. In the LabVIEW environment, we implement and run the imported IP core as a clock-driven logic (CDL) unit. The resulting flow of the data I/O ports is depicted in Fig. 5.3. According to the figure, it takes one iteration (clock cycle) for the IP core to read the data on input ports pinvH_V and s_V_TDATA. On the other hand, the IP core completes the current transaction after 9 cycles, and therefore, it generates the output data on the u_V_TDATA port every 10 cycles.

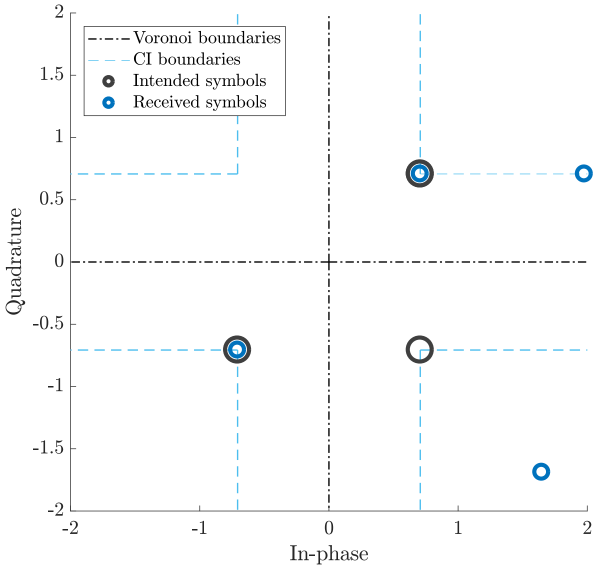


FIGURE 5.4: Intended symbols and noise-free received signals obtained by simulating the HDL design of Algorithm 3.

We particularly focus on the first transaction where the following symbol vector is placed on the data input port:

$$\text{s\_V\_TDATA} = \begin{bmatrix} -0.70703125 - \mathrm{j}0.70703125 \\ 0.70703125 - \mathrm{j}0.70703125 \\ 0.70703125 + \mathrm{j}0.70703125 \\ 0.70703125 + \mathrm{j}0.70703125 \end{bmatrix}. \tag{5.8}$$

The corresponding precoded vector generated on the output port u_V_TDATA of the IP core is

$$\text{u\_V\_TDATA} = \begin{bmatrix} 0.1601562500 + \mathrm{j}0.6035156250 \\ -1.0039062500 - \mathrm{j}1.6718750000 \\ -1.3212890625 + \mathrm{j}3.1640625000 \\ 0.5185546875 + \mathrm{j}1.0224609375 \end{bmatrix}. \tag{5.9}$$

This precoded vector is then passed through the multiuser channel $\mathbf{H}$, and eventually, the noise-free signals received by the users are plotted in Fig. 5.4. It can be seen that the received signal of each user is properly accommodated in the desired CI region. This verifies the accuracy of the designed IP core for implementation of the proposed low-complexity precoding solution.

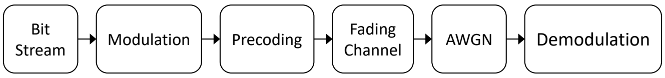


Figure 5.5: Block diagram of the simulated communication system.

5.5 Numerical and Simulation Results

In this section, we provide some simulation results to assess the performance of the proposed low-complexity approximate SLP solution implemented as an IP core. We further compare the results with those obtained from the optimal SLP solution, the closed-form SLP solution in Section 5.3, and the ZF precoding technique. Note that the optimal SLP solution refers to the solution of problem (5.1). The precoding techniques of interest in this section are referred to as:

- ZF: zero-forcing precoding
- OPT-SLP: the optimal SLP solution of problem (5.1)
- CF-SLP: closed-form SLP solution in Section 5.3
- HDL-CF-SLP: HDL implementation of Algorithm 3

The ZF, OPT-SLP and CF-SLP techniques are simulated using the MATLAB software, where a floating-point precision mode is considered by default. On the other hand, the HDL-CF-SLP technique is simulated in the LabVIEW environment, where the implementation uses fixed-point arithmetic as described in Subsection 5.4.1. As mentioned earlier in Section 5.4, to enable implementation of the HDL design in the LabVIEW environment, we have transformed the generated IP core into a design block using the Xilinx Vivado Design Suite tool, and then imported it as a CDL function into our LabVIEW simulation framework. The block diagram of the communication system, simulated in both MATLAB and LabVIEW environments, is shown in Fig. 5.5.

Our simulation setup is as follows. We consider a fully-loaded downlink MU-MIMO system with equal numbers of transmit antennas and users, i.e., $N_\mathrm{t} = N_\mathrm{u} = 4$. The BS uses QPSK signaling and an uncoded transmission scheme to communicate with the users. We assume a unit noise variance and equal target SINRs for all the users, i.e., $\sigma_i^2 = 1$ for all $i = 1, 2, ..., N_\mathrm{u}$ and $\gamma_1 = \gamma_2 = \cdots = \gamma_{N_\mathrm{u}}$. The presented plots in the following are obtained by averaging the results over 100 realizations of the Rayleigh block-fading channel matrix $\mathbf{H}$, where each realization consists of 100 symbols periods.

We show, in Fig. 5.6, the scatter plot of the users' noise-free and noisy received signals obtained from the HDL-CF-SLP technique for a target SINR of 0 dB. It can be seen that the users' noise-free received signals are properly located within the correct distance preserving CI region. As a result, the HDL implementation of our proposed approximate algorithm succeeded to satisfy the CI constraints of the SLP design problem. In the sequel, we evaluate the performance of our FPFA design in terms of average transmit power and symbol error rate (SER).

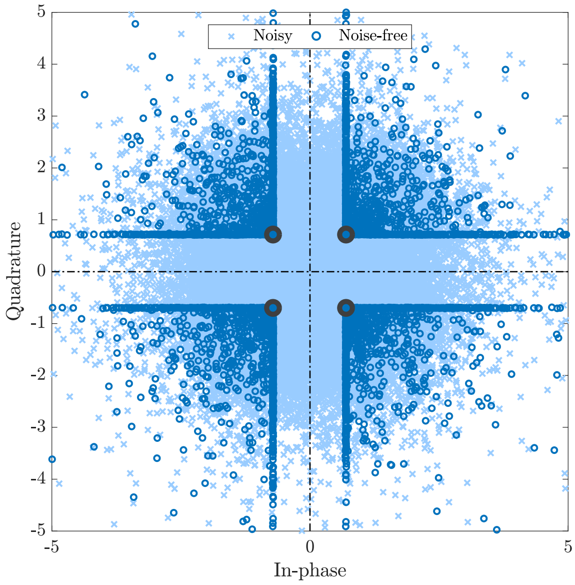
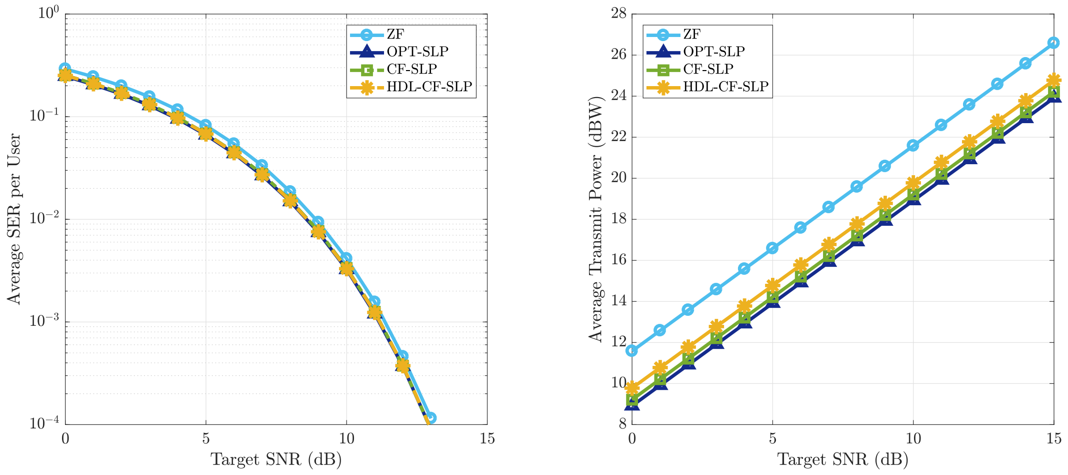


FIGURE 5.6: Scatter plot of the users' received signals at SINR = 0 dB.

In Fig. 5.7 (a), we plot the average SER of each user versus the target SINR for different precoding techniques of interest. It can be seen that all the techniques achieve almost the same SER performance, while the SLP techniques show slightly lower SER values compared to those of the ZF scheme. The reason for this lower SER is that the SLP techniques exploit the users' symbols to design the precoded vector such that it accommodates the noise-free received signal of each user in the distance-preserving CI region that corresponds to the user's intended symbol. Such a received signal has at least an equal or perhaps even an increased distance from the ML decision boundaries, which results in a higher accuracy for symbol detection at the user's receiver. It can further be seen from Fig. 5.7 (a) that the FPGA simulation of the HDL-CF-SLP technique succeeds to achieve the same SER as that of the OPT-SLP. Therefore, the loss due to the approximate solution and the HDL implementation inaccuracies is not noticeable in terms of SER performance.

The average transmit power of each precoding technique corresponding to the SER performances in Fig. 5.7 (a) is shown in Fig. 5.7 (b) versus target SINR. All the SLP techniques, including the HDL-CF-SLP implementation, consume a lower power for precoded downlink transmission, compared to the ZF scheme. In particular, the HDL-CF-SLP implementation achieves 1.9 dBW gain in transmit power against the ZF technique. On the other hand, the FPGA simulation for the HDL-CF-SLP technique shows losses of 0.5 dBW and 0.85 dBW compared to the numerical results obtained for, respectively, the CF-SLP and the OPT-SLP techniques in the MATLAB environment. The loss compared to the CF-SLP technique originates from two facts. First, the HDL-CF-SLP



(a)

(b)

FIGURE 5.7: Performance comparison of different precoding designs as a function of target SINR for a system with QPSK modulation and $N_\mathrm{t} = N_\mathrm{u} = 4$: (a) Average per-user symbol error rate; (b) Average total transmit power.

implementation, which is based on Algorithm 3, uses the approximation (5.5) to avoid the pseudo-inverse calculation in the CF-SLP solution. Second, to design the HDL for Algorithm 3, we have used a fixed-point precision due to FPGA resource limitations which could be a source of inaccuracy in the values produced by the IP core, whereas simulating the CF-SLP method via MATLAB uses floating-point arithmetic. However, one should notice that the HDL-CF-SLP implementation is designed for real-time applications on an FPGA and can provide a high throughput in practice, while the CF-SLP and the OPT-SLP techniques are not designed so. It should be further noted that the loss of the CF-SLP method compared to the OPT-SLP solution comes from the fact that the CF-SLP provides an approximate precoding solution in a two-step non-iterative way, while the OPT-SLP solution is obtained via an iterative optimization algorithm with a higher computational complexity.

Although all the precoding techniques of interest have shown comparable SER performances, they do not offer the same performance when it comes to the transmitted power. In order to incorporate these two performance measures into a single figure of merit, we define power efficiency η as the ratio between the goodput and the transmit power, i.e.,

$$\eta \triangleq \frac{\log_2(M)(1 - \mathrm{BER})}{\|\mathbf{u}\|^2}, \tag{5.10}$$

where M is the modulation order, $\|\mathbf{u}\|^2$ denotes the transmit power, and BER denotes the bit error rate which is simply obtained via dividing the SER by $\log_2(M)$.

We compare the power efficiencies of different precoding techniques in Fig. 5.8 as a

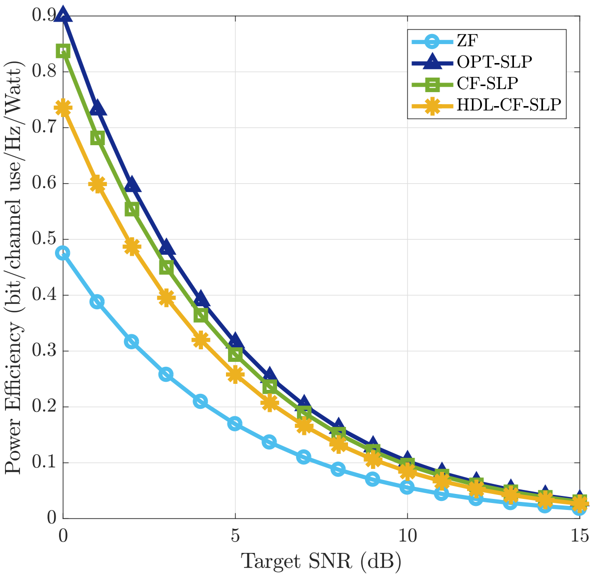


FIGURE 5.8: Power efficiency as a function of target SINR with QPSK modulation and $N_\mathrm{t} = N_\mathrm{u} = 4$.

function of target SINR. The HDL-CF-SLP implementation shows gains of up to 2 dB in power efficiency compared to the ZF scheme. When compared to the MATLAB implementation of SLP techniques, the OPT-SLP and the CF-SLP solutions outperform the HDL-CF-SLP implementation, but these techniques are not able to provide a high symbol throughput. In particular, the HDL implementation of Algorithm 3 shows at most 1 dB loss in the depicted range of target SINR, compared to the OPT-SLP technique. As mentioned earlier, this loss is due to the approximations used in deriving Algorithm 3 and also due to the adopted fixed-point precision. The latter drawback can be alleviated by increasing the bit width of the fixed-point format, but it comes with an excessive FPGA resource utilization. Furthermore, this performance loss is resulted in exchange for simplifying the design of the precoder. The simplified design enables implementation of the SLP algorithm on an actual FPGA. Our simulations in the LabVIEW environment indicate that the HDL design for Algorithm 3 allows data transmission with a high symbol throughput of 100 Mega symbols per second per user. In the considered system with $N_\mathrm{u} = 4$ users and QPSK signaling, it translates to a sum-throughput of 800 Mbps which makes the proposed FPGA design suitable for realistic wireless communication applications.

5.6 Conclusions

We developed an optimized FPGA design to enable low-complexity yet efficient implementation of SLP in a high-throughput downlink MU-MIMO system. The design is essentially based on the CF-SLP solution derived in Chapter 4. In this work, we further simplified this solution by assuming mutually orthogonal channel vectors and proposed an approximate low-complexity design algorithm that can operate in a real-time mode. We analyzed the computational complexity of the proposed design and showed that it has the same complexity order as that of the ZF precoding. We then used the Xilinx Vivado HLS tool to translate the design algorithm into an HDL code and also to optimize the design in order to achieve a low latency, and therefore, a higher throughput. The synthesis results, including performance, timing and resource utilization estimates verified the efficiency of our HDL design. The generated IP core was evaluated in a simulation environment within the LabVIEW software. The simulations for a 4×4 system with QPSK signaling showed that the HDL design of our proposed algorithm is able to operate at a symbol rate of 100 Mega symbols per second per user when deployed on a specific Xilinx FPGA part, which makes it attractive for real-time implementations. Using the MATLAB software, we further evaluated the loss of our design algorithm with respect to the optimal SLP solution, where the loss is shown to be less than 1 dB according to our numerical results. This loss is mainly due to the approximation introduced when deriving the algorithm and also due to the adopted fixed-point arithmetic for the FPGA design. Furthermore, the simulation results indicated that the proposed HDL implementation of SLP outperforms the ZF scheme in terms of power efficiency, where an improvement of up to 50 percent can be achieved.

Chapter 6

Robust Symbol-Level Precoding under System Uncertainties – Part I: Channel Uncertainty

In this chapter, we address robust design of SLP for the downlink of MU-MIMO wireless channels when imperfect CSI is available at the transmitter. In particular, we consider two well-known models for the CSI imperfection, namely, bounded uncertainty and stochastic Gaussian-distributed uncertainty. Our design objective is to minimize the total (per-symbol) transmission power subject to CI constraints as well as the users' quality-of-service requirements in terms of received SINR. Assuming bounded channel uncertainties, we obtain a convex CI constraint based on the worst-case robust analysis, whereas in the case of stochastic uncertainties, we define probabilistic CI constraints in order to achieve robustness to statistically-known CSI errors. Since these probabilistic constraints are difficult to handle, we resort to their convex approximations given in the form of tractable deterministic robust constraints. Three convex approximations are derived based on different conservatism levels, among which one is introduced as a benchmark for comparison. We show that each of our proposed approximations is tighter than the other under specific robustness settings, while both always outperform the benchmark. Using the proposed CI constraints, we formulate a robust SLP design problem as a second-order cone programming (SOCP). Extensive simulation results are provided to validate our analytical results and to make comparisons with conventional block-level robust precoding schemes. We show that the robust design of symbol-level precoder leads to improved performance in terms of energy efficiency at the cost of increasing the computational complexity by an order equal to the number of users in the large system limit, compared to the non-robust design.

6.1 Introduction

Multiuser precoding techniques typically exploit the transmit-side CSI in order to suppress/mitigate the inter-user interference, and therefore, the precoding performance relies on the accuracy of the available CSI at the transmitter. In reality, assuming perfect CSI, either statistically or instantaneously, is somewhat impractical due to various inevitable channel impairments such as imperfect channel estimation, limited (or quantized) feedback, and latency-related errors [143–145]. If accurate CSI is not available, the potential precoding gains may no longer be guaranteed as precoding techniques are generally sensitive to channel uncertainties [144]. One may expect an even more adverse effect of imperfect channel knowledge on the symbol-level precoder's performance since the promised efficiency crucially depends on the satisfaction of CI constraints to successfully accommodate each noise-free received signal in the correct CI region. To address this issue, the problem of designing a multiuser precoder that is robust to channel uncertainties becomes of practical importance.

In robust precoding design, a typical consideration is to presume the channel uncertainty exhibits some known geometric or statistical properties. The set of all possible realizations of the channel satisfying a given property is called the uncertainty region, which can be analytically modeled depending on the error source. In practical wireless systems, the transmitter typically acquires the CSI estimated by the receiver through a feedback channel or directly estimates the CSI via the uplink channel's reciprocity. In the former case, the CSI uncertainty usually originates from the induced latency, or the limited capacity of the feedback channel [146]. In contrast, in the latter case, it may be caused by the imperfections in the estimation process or by the outdated estimates due to the short coherence time of fast-varying wireless environments [147].

The channel uncertainty region is commonly considered either ellipsoidal or stochastic, or even a combination of both, e.g., see [148]. Under the ellipsoidal uncertainty model, usually, no assumption is made on the CSI error distribution; rather, the error is assumed to always lie within a bounded region. Therefore, it is sometimes referred to as bounded uncertainty. This sort of modeling, which ultimately leads to a worst-case design, is known to appropriately capture the bounded uncertainties resulted from quantization errors [149]. Further, it is adequate to deal with slow fading channels where no sufficient statistics for averaging are available. On the other hand, the stochastic uncertainty model assumes that the statistical properties of the CSI error are known. In systems performing channel estimation at either the transmitter or receiver side, such modeling is particularly suitable since the error in the estimation process can often be treated as a Gaussian random process [150].

With a particular focus on MU-MIMO downlink systems, a wide variety of robust schemes can be found in the literature on conventional multiuser precoding, addressing both bounded and stochastic uncertainty models. In this line of work, most of the existing research considers either a QoS-constrained power minimization problem or a max-min fair design with power constraints. Under norm-bounded CSI uncertainty, the QoS problem is typically constrained by the worst SINR among the users, resulting in a highly conservative design approach; see, e.g., [151–153] as some notable research in this direction. These worst-case SINR requirements can also be translated to worst-case minimum mean-square error (MMSE) constraints [154, 155]. Assuming stochastic Gaussian-distributed CSI errors, the QoS requirements are usually implied by probabilistic SINR constraints [156–159], or in terms of equivalent rate-outage probability constraints [160–162]. Implying the QoS requirements in either form, the stochastically robust schemes mostly apply the robust chance-constrained optimization techniques of [163] and [164] in order to tackle the design problem.

On the other hand, robust SLP design has been investigated in some recent work. Worst-case robust SLP approaches are proposed in [20] and [58, 63] for unsecured and secured wireless systems, respectively, aiming to design the symbol-level precoder under norm-bounded CSI errors based on the power minimization and max-min fair criteria. In [64], the authors develop an SLP design to enhance both physical-layer security against eavesdropping and the quality of legitimate transmissions in MU-MIMO wiretap systems, where the design is studied under different assumptions on the availability of CSI at the legitimate and eavesdropping channels, including a bounded CSI uncertainty model. However, it is important to note that as far as the SLP power minimization problem is concerned, the bounded uncertainty model may not yield an efficient solution. This modeling ultimately leads to a worst-case conservatism, which inherently increases the transmission power, though enhancing the users' received SINR and symbol error probability. To address stochastic channel uncertainties in the SLP design, in [116], a sphere bounding scheme is proposed for robust SLP power minimization with probabilistic CI constraints, where the probabilistic constraints are transformed into a tractable second-order cone (SOC) form and are tightened to achieve a lower SER but at the cost of a higher transmitted power. In another work published in [117], the robust SLP design problem is addressed by considering quantized transmit-side CSI. The problem is solved by decomposing the inter-user interference into predictable and unpredictable (due to the quantization error) parts, where an upper bound is derived for the latter part. Targeting CI at the receiver side, the design aligns the predictable interference to achieve much higher received power over the derived upper bound, and ultimately, lower SERs for the users. It is also worth mentioning that a precoding optimization problem with outage probability constraints based on a symbol-level approach is presented in [45], therein the goal is to achieve robustness to the receiver noise, but not to channel uncertainties.

In this work, we address the problem of robust SLP design with imperfect CSI knowledge under both bounded and stochastic uncertainty models. In the optimization problems, we aim to minimize the total transmission power under joint CI and SINR constraints, where the CI constraints are assumed to be distance-preserving. To obtain a robust formulation for the originally non-robust CI constraints, we essentially need to characterize the uncertain component in the CI inequality caused by the CSI imperfection. Our primary challenge is, however, to obtain a tractable deterministic convex approximation for the resulting robust formulation, ensuring that the desired constraint is met (with a certain probability, in the case of stochastic model) for any realization of the CSI error within the uncertainty set. In such a conservative approach, the relative tightness of the derived approximations, which (roughly speaking) measures the cost of tractability, will be of particular importance. Having robust convex constraints, the subsequent modification of the precoding design problem is straightforward due to the fact that the only uncertain part of the problem is the set of CI constraints. Accordingly, the main contributions of this chapter are listed below:

1. We propose some modifications to the CI constraints according to both bounded and stochastic uncertainty models. In the case with bounded CSI uncertainty, we derive a robust second-order cone (SOC) constraint based on a worst-case robust analysis. Under stochastic CSI errors, we redefine the CI constraints as chance-constrained inequalities, for which we derive two robust deterministic alternatives based on the notion of safe convex approximation. Both approximations are formulated as SOC constraints, and therefore, can efficiently be handled. We further obtain a third robust CI restriction as our benchmark for comparison, based on the well-known idea of sphere bounding.

2. We compare the relative tightness of the robust approximations analytically and validate our discussion through simulation results. The results indicate that the proposed robust designs provide tighter approximations than the sphere bounding method.

3. Using the proposed safe approximations for CI constraints, we case robust formulations in the form of convex second-order cone programming (SOCP) for design of the QoS-constrained symbol-level power minimization precoding. We then analyze and compare the computational complexities of the robust and non-robust precoding schemes, thereby indicating that the proposed robust approaches have higher computational cost by a limiting order of the number of users, compared to the original non-robust problem.

The rest of this chapter is organized as follows. We describe the system and uncertainty models in Section 6.2. In Section 6.3, we briefly explain the original SLP problem with non-robust CI constraints. We then define worst-case and stochastic robust formulations for the CI constraints and derive computationally tractable formulations in the form of approximate convex restrictions. We also provide analytical discussions on the approximation tightness in this section. In Section 6.4, we cast the robust SLP optimization problem and analyze the required computational complexity. Our simulation results are provided in Section 6.5. Finally, we conclude the chapter in Section 6.6.

6.2 System and Uncertainty Model

We consider an MU-MIMO wireless downlink channel where a BS, equipped with an array of N_t antennas, serves N_u single-antenna users by sending independent data streams in the same time-frequency resource block, where $N_\mathrm{u} \leq N_\mathrm{t}$. We principally consider the same transmission scheme as described in Chapter 3 under which the discrete-time baseband representation of the received signal at the receiver of the ith user is given by

$$r_i = \mathbf{h}_i \mathbf{u} + z_i, \quad i = 1, 2, ..., N_\mathrm{u}, \tag{6.1}$$

where the row vectors $\mathbf{h}_i \in \mathbb{C}^{1\times N_\mathrm{t}}$, for $i = 1, ..., N_\mathrm{u}$, denote the instantaneous frequency-flat fading channel of the ith transmit-receive antenna pair, $\mathbf{s} = [s_1, \ldots, s_{N_\mathrm{u}}]^\mathrm{T}$ collects the intended symbols for all N_u users, $\mathbf{u} = [u_1, \ldots, u_{N_\mathrm{t}}]^\mathrm{T}$ is the precoded transmit vector, and z_i denotes the zero-mean unit variance additive circularly symmetric complex Gaussian noise. In this chapter, we confine ourselves to constellation sets with unbounded (Voronoi) decision regions, including single-level modulation schemes, e.g., PSK. We further assume, without loss of generality, that identical modulation schemes are used for all the users.

While it is often assumed that all the users have perfect knowledge of their own channels (via, e.g., pilots or training sequences), the BS normally has inaccurate CSI due to several practical impairments such as imperfect channel estimation, limited (or delayed) feedback and quantization errors. Adopting a perturbation-based uncertainty model, we can model the actual channel of the ith user as

$$\mathbf{h}_i = \hat{\mathbf{h}}_i + \mathbf{e}_i, \quad i = 1, 2, ..., N_\mathrm{u}, \tag{6.2}$$

where $\hat{\mathbf{h}}_i \in \mathbb{C}^{1\times N_\mathrm{t}}$ and $\mathbf{e}_i \in \mathbb{C}^{1\times N_\mathrm{t}}$ denote the erroneous channel and the CSI error, respectively, while only $\hat{\mathbf{h}}_i$ is assumed to be known to the BS. The actual channel $\mathbf{h}_i$, the estimate channel $\hat{\mathbf{h}}_i$, and the CSI error $\mathbf{e}_i$ are assumed to be mutually uncorrelated for all $i = 1, 2, ..., N_\mathrm{u}$. To characterize the channel error vectors $\{\mathbf{e}_i\}_{i=1}^{N_\mathrm{u}}$, we consider two different models as follows.

6.2.1 Bounded Uncertainty Region

The Bounded uncertainty model assumes the actual channel $\mathbf{h}_i$ to always lie inside a sphere (in general, ellipsoid) centered at the erroneous channel $\hat{\mathbf{h}}_i$, with some known (deterministic) radius ε_i. In a formal way, it is assumed that $\mathbf{h}_i$ belongs to a spherical uncertainty set defined as

$$\mathcal{H}_i \triangleq \left\{ \mathbf{h}_i \mid \|\mathbf{h}_i - \hat{\mathbf{h}}_i\| \le \varepsilon_i \right\}, \quad i = 1, 2, ..., N_\mathrm{u}, \tag{6.3}$$

from which the ith actual channel is equally described by

$$\mathbf{h}_i = \hat{\mathbf{h}}_i + \mathbf{e}_i, \quad \|\mathbf{e}_i\| \le \varepsilon_i. \tag{6.4}$$

It immediately follows that the uncertain component of the CSI in the spherical model (6.4) has a bounded Euclidean norm. This model is particularly suitable for wireless systems with finite-rate feedback in which the CSI is acquired and quantized at the receiver and fed back to the BS [149, 165]. Note that, in this model, no assumption is made on the distribution of $\mathbf{e}_i$.

6.2.2 Stochastic Uncertainty Region

In wireless systems with imperfect channel estimation, it is commonly assumed that the BS has only knowledge of an estimate channel $\hat{\mathbf{h}}_i$, while the vector $\mathbf{e}_i$ captures the Gaussian estimation error. In this case, the ith actual channel is modeled as

$$\mathbf{h}_i = \hat{\mathbf{h}}_i + \mathbf{e}_i, \quad \mathbf{e}_i \sim \mathcal{CN}(\mathbf{0}, \xi_i^2\,\mathbf{I}), \tag{6.5}$$

where ξ_i^2 denotes the error variance, which is known to the transmitter, and generally depends on the quality of the estimate channel as well as the imperfections in the estimation process. The stochastic error model corresponds to the time-division duplex (TDD) systems, where the BS exploits the estimated uplink channel for the downlink precoding [157]. It is worth noting that the uncertainty model (6.5) may also appear in a different scenario with statistical CSI where the channel statistics are assumed to be partially known at the transmitter, e.g., the channel's mean and/or covariance is/are available; see, e.g., [156, 166, 167]. In such a scenario, one may model the statistical CSI as $\mathbf{h}_i \sim \mathcal{CN}(\hat{\mathbf{h}}_i, \xi_i^2\,\mathbf{I})$, which ultimately leads to the exact same results as presented in the sequel.

From now on, it is more convenient to use the following equivalent real-valued notations:

$$\bar{\mathbf{u}} = \begin{bmatrix} \mathrm{Re}(\mathbf{u}) \\ \mathrm{Im}(\mathbf{u}) \end{bmatrix}, \quad \mathbf{s}_i = \begin{bmatrix} \mathrm{Re}(s_i) \\ \mathrm{Im}(s_i) \end{bmatrix}, \quad i = 1, 2, ..., N_{\mathrm{u}}.$$

Besides, by defining the operator

$$\Omega(\mathbf{y}) \triangleq \begin{bmatrix} \mathrm{Re}(\mathbf{y}) & -\mathrm{Im}(\mathbf{y}) \\ \mathrm{Im}(\mathbf{y}) & \mathrm{Re}(\mathbf{y}) \end{bmatrix},$$

for any given complex vector $\mathbf{y}$, we denote

$$\mathbf{H}_i = \Omega(\mathbf{h}_i), \quad \hat{\mathbf{H}}_i = \Omega(\hat{\mathbf{h}}_i), \quad \mathbf{E}_i = \Omega(\mathbf{e}_i), \quad i = 1, 2, ..., N_{\mathrm{u}}.$$

From the real-valued notations, it is immediately apparent that

$$\mathbf{H}_i = \hat{\mathbf{H}}_i + \mathbf{E}_i, \quad i = 1, 2, ..., N_{\mathrm{u}}, \tag{6.6}$$

and

$$\mathbf{H}_i\bar{\mathbf{u}} = \begin{bmatrix} \mathrm{Re}(\mathbf{h}_i\mathbf{u}) \\ \mathrm{Im}(\mathbf{h}_i\mathbf{u}) \end{bmatrix}. \tag{6.7}$$

Note further that $\mathbf{E}_i(j,:) \sim \mathcal{N}(\mathbf{0}, \frac{1}{2}\xi_i^2\,\mathbf{I})$ for $i = 1, ..., N_{\mathrm{u}}$ and $j = 1, 2$, where $\mathbf{E}_i(j,:)$ refers to the jth row of $\mathbf{E}_i$. In the rest of this chapter, we unify the norm notations such that $\|\cdot\|$ denotes either the Frobenius norm of a matrix or the Euclidean norm of a vector, depending on the input argument. In addition, for a user $i \in \{1, 2, ..., N_{\mathrm{u}}\}$, by "received signal" we mean the noise-free received signal, i.e., $\mathbf{H}_i\bar{\mathbf{u}}$.

6.3 Robust CI Formulation with Imperfect CSI

To design the symbol-level precoder, we are particularly interested in an SINR-constrained power minimization problem. We showed in Section 3.5 that the design problem of interest can be expressed as a convex LCQP. Using the design formulation P1, by assuming imperfect CSI knowledge, we can rewrite the corresponding optimization problem as

$$\begin{aligned} \mathrm{P1}: \quad & \min_{\bar{\mathbf{u}}} \quad \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \\ & \text{s.t.} \quad \mathbf{A}_i\hat{\mathbf{H}}_i\bar{\mathbf{u}} \succeq \sigma_i\sqrt{\gamma_i}\,\mathbf{A}_i\mathbf{s}_i,\ i=1,2,...,N_{\mathrm{u}}, \end{aligned} \tag{6.8}$$

where the entry-wise inequality $\mathbf{A}_i\hat{\mathbf{H}}_i\bar{\mathbf{u}} \succeq \sigma_i\sqrt{\gamma_i}\,\mathbf{A}_i\mathbf{s}_i$ implies the CI constraint for the ith user. The SLP design in (6.8) aims to minimize the total (per-symbol) transmit power while satisfying CI constraints and given target SINRs γ_i for all N_{u} users. However, with imperfect CSI, the design constraints are no longer guaranteed by the solution to P1. To be more specific, in case $\hat{\mathbf{H}}_i \neq \mathbf{H}_i$, the region described by $\mathbf{A}_i\hat{\mathbf{H}}_i\bar{\mathbf{u}} \succeq \sigma_i\sqrt{\gamma_i}\,\mathbf{A}_i\mathbf{s}_i$ is a distorted version of the accurate CI region. Consequently, a received signal $\mathbf{H}_i\bar{\mathbf{u}}$ is no longer guaranteed to lie within the desired CI region, which may cause severe performance degradation. Further to this error-induced distortion of the CI regions, the users may not be provided with the minimum required SINRs given by the target values $\{\gamma_i\}_{i=1}^{N_{\mathrm{u}}}$. Therefore, having a robust formulation for the symbol-level precoding design problem is crucial to ensure the CI constraints as well as the minimum SINR requirements of the users for any possible CSI realization. To this end, we first reformulate the CI constraints under each uncertainty model.

We start off by restating the actual CI constraint to be met for user i, i.e.,

$$\mathbf{A}_i\mathbf{H}_i\bar{\mathbf{u}} \succeq \sigma_i\sqrt{\gamma_i}\mathbf{A}_i\mathbf{s}_i, \quad i=1,2,...,N_{\mathrm{u}},$$

By substituting (6.6) for $\mathbf{H}_i$, we have

$$\mathbf{A}_i\hat{\mathbf{H}}_i\bar{\mathbf{u}} \succeq \sigma_i\sqrt{\gamma_i}\mathbf{A}_i\mathbf{s}_i - \mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}}, \quad i=1,2,...,N_{\mathrm{u}}. \tag{6.9}$$

In the sequel, we separately consider each uncertainty model and derive robust formulations for the CI constraints. For the brevity of notation, we hereafter denote by

$$\mathbf{w}_i(\bar{\mathbf{u}}) \triangleq \sigma_i\sqrt{\gamma_i}\mathbf{A}_i\mathbf{s}_i - \mathbf{A}_i\hat{\mathbf{H}}_i\bar{\mathbf{u}}, \tag{6.10}$$

the certain part of the CI inequality (6.9) which is affine in $\bar{\mathbf{u}}$, where $\mathbf{w}_i(\bar{\mathbf{u}}) = [w_{i,1}, w_{i,2}]^{\mathrm{T}}$.

6.3.1 Worst-Case Robust Formulation

The bounded uncertainty region $\mathcal{H}_i$ can be interpreted as having all the possible error vectors inside a $2N_{\mathrm{t}}$-dimensional sphere with radius $\sqrt{2}\,\varepsilon_i$. Therefore, the robust formulation of (6.9) for the ith user can be written as

$$\mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}} \succeq \mathbf{w}_i(\bar{\mathbf{u}}),\ \forall\mathbf{E}_i : \|\mathbf{E}_i\| \leq \sqrt{2}\,\varepsilon_i, \tag{6.11}$$

which implies that (6.9) must be satisfied for all $\mathbf{E}_i$ belonging to the CSI uncertainty set. Even though the feasibility region of (6.11) is convex, this semi-infinite constraint consists of an infinite number of linear inequalities to be satisfied which is computationally intractable. In order to achieve robustness over a bounded uncertainty set as in (6.11), a common approach is to consider the design constraint in its worst case. Accordingly, letting $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}]^{\mathrm{T}}$, the worst-case formulation of (6.11) can be written as

$$\begin{bmatrix} \inf\{\mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{E}_i\bar{\mathbf{u}} : \|\mathbf{E}_i\| \leq \sqrt{2}\,\varepsilon_i\} \\ \inf\{\mathbf{a}_{i,2}^{\mathrm{T}}\mathbf{E}_i\bar{\mathbf{u}} : \|\mathbf{E}_i\| \leq \sqrt{2}\,\varepsilon_i\} \end{bmatrix} \geq \mathbf{w}_i(\bar{\mathbf{u}}). \tag{6.12}$$

In our model, the worst-case uncertainty is realized through the maximal CSI error norm, i.e., the radius of the CSI error sphere. From the definition of the spherical uncertainty set in (6.3), it can be easily shown that the entries of $\mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}}$ are bounded too. We also note that

$$\mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}} = (\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{A}_i)\,\mathrm{vec}(\mathbf{E}_i), \tag{6.13}$$

which can be simply verified using the well-known property $\mathrm{vec}(\mathbf{XYW}) = (\mathbf{W}^{\mathrm{T}} \otimes \mathbf{X})\,\mathrm{vec}(\mathbf{Y})$, for any given matrices $\mathbf{X}, \mathbf{Y}, \mathbf{W}$ with appropriate dimensions, and also the fact that $\mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}} = \mathrm{vec}(\mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}})$. It then follows that

$$\mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}} = \begin{bmatrix} (\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{a}_{i,1}^{\mathrm{T}})\,\mathrm{vec}(\mathbf{E}_i) \\ (\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{a}_{i,2}^{\mathrm{T}})\,\mathrm{vec}(\mathbf{E}_i) \end{bmatrix}. \tag{6.14}$$

Now, let us focus on the rows of the right-hand side vector in (6.14). By the Cauchy-Schwarz inequality, we have

$$(\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{a}_{i,j}^{\mathrm{T}})\mathrm{vec}(\mathbf{E}_i) \geq -\|\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{a}_{i,j}^{\mathrm{T}}\|\,\|\mathrm{vec}(\mathbf{E}_i)\|,\ j = 1, 2. \tag{6.15}$$

Using the uncertainty radius $\|\mathrm{vec}(\mathbf{E}_i)\| = \|\mathbf{E}_i\| \leq \sqrt{2}\,\varepsilon_i$, an immediate consequence of (6.15) is that $(\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{a}_{i,j}^{\mathrm{T}})\mathrm{vec}(\mathbf{E}_i)$ is bounded from below by $-\sqrt{2}\,\varepsilon_i\,\|\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{a}_{i,j}^{\mathrm{T}}\|$ for $j = 1, 2$. However, by exploiting the structure of $\mathrm{vec}(\mathbf{E}_i)$, it is possible to further obtain a tighter bound which is given by

$$\begin{aligned} \inf\left\{(\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{a}_{i,j}^{\mathrm{T}})\,\mathrm{vec}(\mathbf{E}_i) :\ \|\mathbf{E}_i\| \leq \sqrt{2}\,\varepsilon_i\right\} &= -\varepsilon_i\,\|\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{a}_{i,j}^{\mathrm{T}}\| \\ &= -\varepsilon_i\,\|\bar{\mathbf{u}}\|\,\|\mathbf{a}_{i,j}\|,\ j = 1, 2, \end{aligned} \tag{6.16}$$

where the last equality of (6.16) is derived considering the fact that $\|\mathbf{x} \otimes \mathbf{y}\| = \|\mathbf{x}\|\,\|\mathbf{y}\|$, for any two vectors $\mathbf{x}$ and $\mathbf{y}$. Finally, substituting (6.16) for the infimum in (6.12), the worst-case CI constraint for the ith user is obtained by

$$-\varepsilon_i\,\|\bar{\mathbf{u}}\| \begin{bmatrix} \|\mathbf{a}_{i,1}\| \\ \|\mathbf{a}_{i,2}\| \end{bmatrix} \succeq \mathbf{w}_i(\bar{\mathbf{u}}), \tag{6.17}$$

The CI constraint (6.17) can be equivalently expressed by two second-order cone (SOC) constraints, given in a compact form by

$$\mathrm{W}: \quad \|\bar{\mathbf{u}}\|\,\mathbf{1} \succeq \frac{-1}{\varepsilon_i}(\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}} \circ \mathbf{I})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}}). \tag{6.18}$$

In fact, the worst-case constraint W guarantees that the CI requirement for the ith user will be met in the presence of any unknown, but norm-bounded CSI error. The robust formulation (6.18) is convex and thus can efficiently be handled via off-the-shelf convex optimization algorithms [121]. It is worth mentioning that a similar worst-case robust approach has also been studied in [20] for symbol-level downlink precoding in which the CI regions coincide with the DPCIRs in the special case of PSK signaling, but characterization of the CI constraints are not identical. Nevertheless, the final robust formulations, despite being different in presentation, are based on the same idea and are basically equivalent.

6.3.2 Stochastic Robust Formulation

A stochastic robust CI constraint must satisfy (6.9) with a certain probability for any possible realization of the CSI error $\mathbf{E}_i$ within the uncertainty region. Assuming statistically-known CSI errors, the CI constraint in (6.9) turns into an uncertain inequality with the uncertainty arising from the stochastic CSI error $\mathbf{E}_i$. Although the feasible set of this uncertain inequality is always convex, the main difficulty is to efficiently check whether this convex constraint is satisfied at a given point, which is highly computationally demanding. In such a case, the deterministic constraint in (6.9) can be reformulated as a probabilistic constraint (also known as chance constraint). The chance constraint then implies that the ith user will see its received signal outside of the correct CI region only with a constrained small probability, i.e.,

$$\Pr\left\{\mathbf{A}_i\hat{\mathbf{H}}_i\bar{\mathbf{u}} \not\succeq \sigma_i\sqrt{\gamma_i}\mathbf{A}_i\mathbf{s}_i - \mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}}\right\} < \upsilon, \tag{6.19}$$

which can be equally expressed as

$$\Pr\left\{\mathbf{A}_i\hat{\mathbf{H}}_i\bar{\mathbf{u}} \succeq \sigma_i\sqrt{\gamma_i}\mathbf{A}_i\mathbf{s}_i - \mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}}\right\} \geq 1-\upsilon, \tag{6.20}$$

where $\upsilon \in (0, 1/2]$ denotes the violation probability threshold which is a system design parameter controlling the desired level of conservatism. Remark that the SINR requirement γ_i translates to an achievable target rate of $R_i = \log_2(1+\gamma_i)$, under ergodic conditions on the channel [168]. Thus, the constraint (6.20) can also be viewed as a rate-outage probability constraint, ensuring that the transmission rate R_i is achievable for the ith user with a probability of at least $1-\upsilon$. For the sake of notational simplicity, we denote the stochastic uncertain component of the CI constraint by

$$\mathbf{q}_i \triangleq \mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}} = (\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{A}_i)\,\mathrm{vec}(\mathbf{E}_i), \tag{6.21}$$

which can be simply verified using the well-known property $\mathrm{vec}(\mathbf{XYW}) = (\mathbf{W}^{\mathrm{T}} \otimes \mathbf{X})\,\mathrm{vec}(\mathbf{Y})$, for any given matrices $\mathbf{X}, \mathbf{Y}, \mathbf{W}$ with appropriate dimensions, along with the fact that $\mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}} = \mathrm{vec}(\mathbf{A}_i\mathbf{E}_i\bar{\mathbf{u}})$. Let $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}]^{\mathrm{T}}$, then using (6.21), we can write

$$\mathbf{q}_i = \begin{bmatrix} (\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{a}_{i,1}^{\mathrm{T}})\,\mathrm{vec}(\mathbf{E}_i) \\ (\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{a}_{i,2}^{\mathrm{T}})\,\mathrm{vec}(\mathbf{E}_i) \end{bmatrix} \triangleq \begin{bmatrix} q_{i,1} \\ q_{i,2} \end{bmatrix}, \tag{6.22}$$

from which it is straightforward to show that $\mathbf{q}_i$ is a possibly correlated bivariate Gaussian random variable. The chance constraint (6.20) can then be written, in a simpler form, as

$$\Pr\{\mathbf{q}_i \succeq \mathbf{w}_i(\bar{\mathbf{u}})\} \geq 1 - \upsilon, \quad i = 1, 2, ..., N_{\mathrm{u}}. \tag{6.23}$$

The constraints in (6.23) belong to chance-constrained vector inequalities, which are generally known to be computationally intractable [163], as we will also see later. In what follows, the goal is to derive equivalent deterministic expressions for (6.23). For this purpose, we first study the statistical properties of the uncertain vector $\mathbf{q}_i$.

We begin with the Gaussian error vector $\mathrm{vec}(\mathbf{E}_i)$ for which the mean and the covariance matrix are respectively given by $\mathbb{E}\{\mathrm{vec}(\mathbf{E}_i)\} = \mathbf{0}$ and

$$\mathbb{E}\left\{\mathrm{vec}(\mathbf{E}_i)\mathrm{vec}(\mathbf{E}_i)^{\mathrm{T}}\right\} = \frac{1}{2}\,\xi_i^2 \begin{bmatrix} \mathbf{I}_{2N_{\mathrm{t}}} & \mathbf{J} \\ \mathbf{J}^{\mathrm{T}} & \mathbf{I}_{2N_{\mathrm{t}}} \end{bmatrix}, \tag{6.24}$$

where

$$\mathbf{J} = \mathbf{I}_N \otimes \mathbf{J}_2,\ \mathbf{J}_2 \triangleq \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Accordingly, the mean of $\mathbf{q}_i$ can be obtained as

$$\mathbb{E}\{\mathbf{q}_i\} = \left(\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{A}_i\right)\,\mathbb{E}\left\{\mathrm{vec}(\mathbf{E}_i)\right\} = \mathbf{0}, \tag{6.25}$$

and its covariance matrix is given by

$$\begin{aligned} \mathbf{C}_i &= \mathbb{E}\{\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\} \\ &\overset{\text{(a)}}{=} (\bar{\mathbf{u}}^{\mathrm{T}} \otimes \mathbf{A}_i)\,\mathbb{E}\Big\{\mathrm{vec}(\mathbf{E}_i)\mathrm{vec}(\mathbf{E}_i)^{\mathrm{T}}\Big\}(\bar{\mathbf{u}} \otimes \mathbf{A}_i^{\mathrm{T}}) \\ &\overset{\text{(b)}}{=} \frac{1}{2}\,\xi_i^2\,\left(\bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \otimes \mathbf{A}_i\mathbf{A}_i^{\mathrm{T}}\right) = \frac{1}{2}\,\xi_i^2\,\|\bar{\mathbf{u}}\|^2\,\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}}, \end{aligned} \tag{6.26}$$

where the equality (a) can be verified using the property $(\mathbf{X} \otimes \mathbf{Y})^{\mathrm{T}} = (\mathbf{X}^{\mathrm{T}} \otimes \mathbf{Y}^{\mathrm{T}})$, for any given matrices $\mathbf{X}, \mathbf{Y}, \mathbf{W}, \mathbf{Z}$, and the equality (b) has been verified in Appendix C.1.

Remark 1. Using the fact that $\mathbf{q}_i$ has a symmetric distribution around zero, it is straightforward to verify that the chance constraint (6.23) is feasible for every $\upsilon \in (0, 1/2]$ if and only if (iff) we have $\mathbb{E}\{\mathbf{q}_i\} \succeq \mathbf{w}_i(\bar{\mathbf{u}})$. Therefore, under the assumption $\upsilon \in (0, 1/2]$, a necessary and sufficient condition for (6.23) to have a nonempty feasible region is $\mathbf{w}_i(\bar{\mathbf{u}}) \preceq \mathbf{0}$. This condition must be considered as an additional constraint for every $i \in \{1, ..., N_{\mathrm{u}}\}$ in the formulation of the robust SLP optimization problem.

Using the first two moments of $\mathbf{q}_i$, the probability in (6.23) can be precisely evaluated as the integral of the joint Gaussian probability distribution of $q_{i,1}$ and $q_{i,2}$, i.e.,

$$\begin{aligned}\Pr\{\mathbf{q}_i \succeq \mathbf{w}_i(\bar{\mathbf{u}})\} &= \Pr\left\{q_{i,1} \geq w_{i,1}, q_{i,2} \geq w_{i,2}\right\} \\ &= \int\limits_{w_{i,2}}^{\infty}\int\limits_{w_{i,1}}^{\infty} \frac{1}{2\pi\sqrt{|\mathbf{C}_i|}} \exp\left\{-\frac{1}{2}\mathbf{q}_i^{\mathrm{T}}\mathbf{C}_i^{-1}\mathbf{q}_i\right\} \mathrm{d}q_{i,1}\mathrm{d}q_{i,2}.\end{aligned} \tag{6.27}$$

However, no explicit closed-form expression is known for the integral in (6.27). It becomes even more challenging to imply the constraint (6.27) in the precoding optimization problem. In order to resolve the difficulty of finding a tractable (convex) expression for (6.27), a straightforward approach is to eliminate the (possible) correlation between the entries of $\mathbf{q}_i$ through applying a decorrelating transform. In this regard, the optimal decorrelation matrix (in the sense of minimum mean-square error) is shown in [169] to be

$$\mathbf{C}_i^{-1/2} = \frac{\sqrt{2}}{\xi_i \left\|\bar{\mathbf{u}}\right\|} \left(\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}}\right)^{-1/2}, \tag{6.28}$$

where $(\cdot)^{-1/2}$ denotes inverse square root. We recall from Section 3.3 that the 2×2 matrix $\mathbf{A}_i$ can always be formed as a non-singular matrix, which results in non-singular $\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}}$. Thus, $\mathbf{C}_i$ is positive definite and has a unique (invertible) square root. As a result, the probability expression in (6.27) can be equivalently written as

$$\begin{aligned}\Pr\left\{\mathbf{q}_i \geq \mathbf{w}_i(\bar{\mathbf{u}})\right\} &= \Pr\left\{\mathbf{C}_i^{1/2}\mathbf{C}_i^{-1/2}\mathbf{q}_i \succeq \mathbf{w}_i(\bar{\mathbf{u}})\right\} \\ &= \Pr\left\{\bar{\mathbf{q}}_i \succeq \mathbf{C}_i^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}})\right\} \\ &= \Pr\left\{\bar{\mathbf{q}}_i \succeq \bar{\mathbf{w}}_i(\bar{\mathbf{u}})\right\},\end{aligned} \tag{6.29}$$

where $\bar{\mathbf{q}}_i \triangleq \mathbf{C}_i^{-1/2}\mathbf{q}_i$ and $\bar{\mathbf{w}}_i(\bar{\mathbf{u}}) \triangleq \mathbf{C}_i^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}})$. It can be verified that $\bar{\mathbf{q}}_i$ is an uncorrelated zero-mean Gaussian random vector with unit diagonal covariance matrix, given as

$$\begin{aligned}\bar{\mathbf{C}}_i &\triangleq \mathbb{E}\left\{\bar{\mathbf{q}}_i\bar{\mathbf{q}}_i^{\mathrm{T}}\right\} \\ &= \mathbb{E}\left\{\mathbf{C}_i^{-1/2}\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\mathbf{C}_i^{-1/2}\right\} \\ &= \mathbf{C}_i^{-1/2}\mathbb{E}\left\{\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right\}\mathbf{C}_i^{-1/2} \\ &= \mathbf{C}_i^{-1/2}\mathbf{C}_i\mathbf{C}_i^{-1/2} = \mathbf{I}.\end{aligned} \tag{6.30}$$

Consequently, the chance constraint (6.23) is equivalent to

$$\Pr\left\{\bar{\mathbf{q}}_i \geq \bar{\mathbf{w}}_i(\bar{\mathbf{u}})\right\} \geq 1 - \upsilon, \tag{6.31}$$

with $\bar{\mathbf{q}}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. This probability may appear to be easily handled as it can be expressed by the product of two (complementary) error functions. In the context of convex optimization, however, we essentially need to reach a convex representation for (6.31).

This could be, in general, an intricate task since the joint probability in (6.31) does not admit a tractable convex expression. An alternative approach to tackle this intractability is to replace (6.31) with its safe tractable approximation, resulting in an efficiently computable convex constraint. Such an approximation lies within the literature of robust optimization techniques [163, 170]. Here, the term "safe" means that any feasible point for the safe approximation must be necessarily feasible also for (6.31). Therefore, in what follows, the goal is to propose computationally tractable (but possibly not equivalent) convex approximations implying the CI chance constraint (6.31).

Remark 2. Similar to Remark 1, since $\bar{\mathbf{q}}_i$ is symmetrically distributed around zero, the chance constraint (6.31) is feasible for $\upsilon \in (0, 1/2]$ iff $\mathbb{E}\{\bar{\mathbf{q}}_i\} \succeq \bar{\mathbf{w}}_i(\bar{\mathbf{u}})$, or equally iff $\bar{\mathbf{w}}_i(\bar{\mathbf{u}}) \preceq \mathbf{0}$. However, for practical modulation schemes, using the definition of $\bar{\mathbf{w}}_i(\bar{\mathbf{u}})$, one can verify that the condition $\mathbf{w}_i(\bar{\mathbf{u}}) \preceq \mathbf{0}$ is also sufficient to have $\bar{\mathbf{w}}_i(\bar{\mathbf{u}}) \preceq \mathbf{0}$.

Proposed Safe Approximation I

One may simply exploit the fact that the two random variable in $\bar{\mathbf{q}}_i$ are uncorrelated, hence independent. Consequently, denoting $\bar{\mathbf{q}}_i = [\bar{q}_{i,1}, \bar{q}_{i,2}]^{\mathrm{T}}$ and $\bar{\mathbf{w}}_i(\bar{\mathbf{u}}) = [\bar{w}_{i,1}, \bar{w}_{i,2}]^{\mathrm{T}}$, using the Gaussian cumulative distribution function, we can separate the joint probability in (6.31) as

$$\begin{aligned}\Pr\left\{\bar{\mathbf{q}}_i \geq \bar{\mathbf{w}}_i(\bar{\mathbf{u}})\right\} &= \Pr\left\{\bar{q}_{i,1} \geq \bar{w}_{i,1}\right\}\, \Pr\left\{\bar{q}_{i,2} \geq \bar{w}_{i,2}\right\} \\ &= \frac{1}{2}\mathrm{erfc}\left(\frac{\bar{w}_{i,1}}{\sqrt{2}}\right) \times \frac{1}{2}\mathrm{erfc}\left(\frac{\bar{w}_{i,2}}{\sqrt{2}}\right),\end{aligned} \tag{6.32}$$

where $\mathrm{erfc}(\cdot)$ is the complementary error function defined as $\mathrm{erfc}(y) \triangleq \frac{2}{\sqrt{\pi}} \int_y^\infty e^{-v^2} \mathrm{d}v$. Due to the decreasing monotonicity of the complementary error function, the desired probability is always bounded from below by

$$\Pr\left\{\bar{\mathbf{q}}_i \geq \bar{\mathbf{w}}_i(\bar{\mathbf{u}})\right\} \geq \frac{1}{4}\,\mathrm{erfc}^2\left(\frac{\max\{\bar{w}_{i,1}, \bar{w}_{i,2}\}}{\sqrt{2}}\right). \tag{6.33}$$

Using (6.33), in order to imply the chance constraint (6.31), it is sufficient to consider the deterministic constraint

$$\frac{1}{4}\,\mathrm{erfc}^2\left(\frac{\max\{\bar{w}_{i,1}, \bar{w}_{i,2}\}}{\sqrt{2}}\right) \geq 1 - \upsilon, \tag{6.34}$$

which can be written as

$$-\max\left[\bar{\mathbf{w}}_i(\bar{\mathbf{u}})\right] \leq \rho(\upsilon), \tag{6.35}$$

where $\rho(\upsilon) \triangleq -\sqrt{2}\,\mathrm{erfc}^{-1}\left(2\sqrt{1-\upsilon}\right)$ with $\mathrm{erfc}^{-1}(\cdot)$ denoting the inverse complementary error function, and $\max[\cdot]$ denotes elementwise maximum. It can be verified that the elementwise maximum of affine functions in (6.35) is convex; see [121, p. 80]. Therefore, replacing $\bar{\mathbf{w}}_i(\bar{\mathbf{u}})$, the conservative robust approximation (6.35) can be rewritten in the form of a convex SOC constraint as

$$\text{A1}: \quad \|\bar{\mathbf{u}}\| \leq \frac{-\sqrt{2}}{\rho(\upsilon)\,\xi_i} \max\left[(\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}})\right], \tag{6.36}$$

Note that, in general, the feasible region of A1 is a convex subset of that of (6.31). Therefore, the convex approximation A1 may not exactly imply the desired chance constraint (6.31), but any feasible solution to (6.36) is guaranteed to be feasible also for (6.31).

Proposed Safe Approximation II

Our subsequent derivation of a second safe tractable approximation for (6.31) is essentially based on the well-known Schur complement lemma and the following theorem [163, Th. 4.1].

Lemma 9. *(Schur complement) Let* $\mathbf{W}$ *be a symmetric matrix given by*

$$\mathbf{W} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y}^{\mathrm{T}} & \mathbf{Z} \end{bmatrix}. \tag{6.37}$$

Then, $\mathbf{W} \succeq 0$ *iff* $\mathbf{X} \succeq 0$ *and* $\mathbf{\Delta_X} \succeq 0$*, where* $\mathbf{\Delta_X} = \mathbf{Z} - \mathbf{Y}^{\mathrm{T}}\mathbf{X}^{-1}\mathbf{Y}$ *is the Schur complement of* $\mathbf{X}$ *in* $\mathbf{W}$*.*

Theorem 10. *Let* $\mathbf{D}_0, \mathbf{D}_1, ..., \mathbf{D}_L$ *be diagonal* $n \times n$ *matrices with* $\mathbf{D}_0 \succeq 0$*, and* $\zeta_1, ..., \zeta_L$ *be mutually independent random variables where* $\zeta_l \sim \mathcal{N}(0,1)$ *for all* $l \in \{1, ..., L\}$*. Then, the semidefinite constraint*

$$\operatorname{Arw}(\mathbf{D}_0, \mathbf{D}_1, ..., \mathbf{D}_L) \succeq 0,$$

implies, for every $\upsilon \in (0, 1/2]$*, that*

$$\Pr\left\{-\psi(\upsilon)\mathbf{D}_0 \preceq \sum_{l=1}^{L} \zeta_l \mathbf{D}_l \preceq \psi(\upsilon)\mathbf{D}_0\right\} \geq 1 - \upsilon,$$

with $\psi(\upsilon) = \operatorname{erfc}^{-1}\left(\frac{\upsilon}{2N_{\mathrm{t}}}\right)$*, where*

$$\operatorname{Arw}(\mathbf{D}_0, \mathbf{D}_1, ..., \mathbf{D}_L) \triangleq \begin{bmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_L \\ \mathbf{D}_1 & \mathbf{D}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{D}_2 & \mathbf{0} & \mathbf{D}_0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{D}_L & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_0 \end{bmatrix}.$$

We recall that our goal here is to find a tractable sufficient convex condition for the CI inequality in (6.31) to be satisfied with probability at least $1 - \upsilon$. The inequality of interest, i.e., $\bar{\mathbf{q}}_i \geq \bar{\mathbf{w}}_i(\bar{\mathbf{u}})$, can be equivalently expressed by a linear matrix inequality (LMI) as

$$\psi(\upsilon)\mathbf{D}_{0,i} + \bar{q}_{i,1}\mathbf{D}_1 + \bar{q}_{i,2}\mathbf{D}_2 \succeq 0, \tag{6.38}$$

$$\mathbf{D}_{0,i} \triangleq \frac{1}{\psi(\upsilon)}\begin{bmatrix} -\bar{w}_{i,1} & 0 \\ 0 & -\bar{w}_{i,2} \end{bmatrix}, \quad \mathbf{D}_1 \triangleq \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{D}_2 \triangleq \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

Since $\bar{q}_{i,1}$ and $\bar{q}_{i,2}$ are both symmetric in distribution and the violation probability υ is typically small, a sufficient condition for

$$\Pr\left\{\psi(\upsilon)\mathbf{D}_{0,i} + \bar{q}_{i,1}\mathbf{D}_1 + \bar{q}_{i,2}\mathbf{D}_2 \succeq 0\right\} \geq 1 - \upsilon, \tag{6.39}$$

is also sufficient for

$$\Pr\left\{-\psi(\upsilon)\mathbf{D}_{0,i} \preceq \bar{q}_{i,1}\mathbf{D}_1 + \bar{q}_{i,2}\mathbf{D}_2 \preceq \psi(\upsilon)\mathbf{D}_{0,i}\right\} \geq 1 - \upsilon. \tag{6.40}$$

A direct application of Theorem 10 with $n = 2$ and $L = 2$ implies that the chance constraint (6.40) is met if

$$\mathrm{Arw}(\mathbf{D}_{0,i}, \mathbf{D}_1, \mathbf{D}_2) \succeq 0, \tag{6.41}$$

holds true with $\psi(\upsilon) = \mathrm{erfc}^{-1}(\upsilon/4)$. Notice that a necessary condition for Theorem 10 to be valid is $\mathbf{D}_{0,i} \succeq 0$. The matrix $\mathrm{Arw}(\mathbf{D}_{0,i}, \mathbf{D}_1, \mathbf{D}_2)$ is symmetric, and further, can be partitioned as required in (6.37). As a result, using Lemma 9 with $\mathbf{X} = \mathbf{D}_{0,i}$ and $\mathbf{W} = \mathrm{Arw}(\mathbf{D}_{0,i}, \mathbf{D}_1, \mathbf{D}_2)$, it can be verified that the following implication holds:

$$\mathrm{Arw}(\mathbf{D}_{0,i}, \mathbf{D}_1, \mathbf{D}_2) \succeq 0 \implies \mathbf{D}_{0,i} \succeq 0. \tag{6.42}$$

Therefore, the safe convex constraint (6.41) sufficiently implies our desired chance constraint in (6.40). Finally, by replacing $\mathbf{D}_{0,i}$, $\mathbf{D}_1$ and $\mathbf{D}_2$ in (6.41), the safe convex approximation is obtained as the semidefinite constraint

$$\begin{bmatrix} -\frac{\bar{w}_{i,1}}{\psi(\upsilon)} & 0 & 1 & 0 & 0 & 0 \\ 0 & -\frac{\bar{w}_{i,2}}{\psi(\upsilon)} & 0 & 0 & 0 & 1 \\ 1 & 0 & -\frac{\bar{w}_{i,1}}{\psi(\upsilon)} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{\bar{w}_{i,2}}{\psi(\upsilon)} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{\bar{w}_{i,1}}{\psi(\upsilon)} & 0 \\ 0 & 1 & 0 & 0 & 0 & -\frac{\bar{w}_{i,2}}{\psi(\upsilon)} \end{bmatrix} \succeq 0. \tag{6.43}$$

It is easy to check that the LMI in (6.43) is not convex in the given form with respect to $\bar{\mathbf{u}}$. Nevertheless, it has been shown in Appendix C.2 that, using the implication $\mathbf{w}_i \preceq \mathbf{0}$ provided in Remark 1, it is possible to recast the semidefinite constraint (6.43) as an equivalent SOC constraint given by

$$\mathrm{A2}: \quad \|\bar{\mathbf{u}}\|\,\mathbf{1} \leq \frac{-\sqrt{2}}{\psi(\upsilon)\,\xi_i}(\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}}), \tag{6.44}$$

which is indeed convex in $\bar{\mathbf{u}}$, and can efficiently be handled by standard convex optimization tools [121].

Sphere Bounding Method (Benchmark)

In order to gain some insight into the proposed safe convex approximation A2, and further for comparison purposes, we also formulate a benchmark approximation based on the so-called sphere bounding method. The idea (in some sense) is borrowed from the worst-case robust design approach. More specifically, the goal is basically to find a bounded uncertainty set to which the stochastically uncertain component in (6.31) belongs with a certain probability; subsequently, the worst-case approach can be applied. The following lemma from [162] helps us to proceed with the formulation.

Lemma 11. *Let $\mathcal{K} \subset \mathbb{R}^n$ be an arbitrary set with the property $f(\mathbf{x}) \geq \mathbf{0}$ for all $\mathbf{x} \in \mathcal{K}$, where $f(\cdot)$ is in general a vector-valued function. Then, for a given $\mathbf{y} \in \mathbb{R}^n$, the restriction*

$$\Pr\left\{f(\mathbf{y}) \succeq \mathbf{0}\right\} \geq 1 - \upsilon,$$

is implied sufficiently by satisfying $\Pr\left\{\mathbf{y} \in \mathcal{K}\right\} \geq 1 - \upsilon$.

In order to imply the chance constraint (6.31), one may use the implication provided by Lemma 11 to obtain a (preferably) tight convex restriction, as long as the resulting constraint is efficiently computable. This requires to properly choose the set $\mathcal{K} \subseteq \mathbb{R}^2$ in a way that the condition

$$f(\bar{\mathbf{q}}_i) \geq \mathbf{0}, \quad f(\bar{\mathbf{q}}_i) \triangleq \bar{\mathbf{q}}_i - \bar{\mathbf{w}}_i(\bar{\mathbf{u}}), \tag{6.45}$$

is met for all $\bar{\mathbf{q}}_i \in \mathcal{K}$, while satisfying $\Pr\left\{\bar{\mathbf{q}}_i \in \mathcal{K}\right\} \geq 1 - \upsilon$. We recall that $\bar{\mathbf{q}}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and that $\bar{\mathbf{q}}_i$ has a symmetric distribution. Thus, the condition (6.45) can be equally expressed as

$$f(\bar{\mathbf{q}}_i) \leq \mathbf{0}, \quad f(\bar{\mathbf{q}}_i) \triangleq \bar{\mathbf{q}}_i + \bar{\mathbf{w}}_i(\bar{\mathbf{u}}). \tag{6.46}$$

A common convex choice for the set $\mathcal{K}$ to reach a computationally tractable formulation is the ball represented by

$$\mathcal{K} \triangleq \left\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| \leq \alpha(\upsilon)\right\}, \tag{6.47}$$

with a radius of

$$\alpha(\upsilon) = \sqrt{\Phi_2^{-1}(1 - \upsilon)},$$

where $\Phi_n^{-1}(\cdot)$ is the inverse cumulative distribution function of the central Chi-square random variable with n degrees of freedom. It is then straightforward to verify that

$$\Pr\left\{\bar{\mathbf{q}}_i \in \mathcal{K}\right\} = 1 - \upsilon, \tag{6.48}$$

from which it can be presumed that the Euclidean norm of $\bar{\mathbf{q}}_i$ is bounded by $\alpha(\upsilon)$ with

TABLE 6.1: Proposed worst-case/stochastic robust CI constraints.

Method	Robust CI constraint (for $i = 1, 2, ..., N_\mathrm{u}$)
Worst-case	W : $\|\bar{\mathbf{u}}\|\mathbf{1} \preceq \frac{-1}{\varepsilon_i}(\mathbf{A}_i\mathbf{A}_i^\mathrm{T} \circ \mathbf{I})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}}), \quad \mathbf{w}_i(\bar{\mathbf{u}}) = \sigma_i\sqrt{\gamma_i}\mathbf{A}_i\mathbf{s}_i - \mathbf{A}_i\hat{\mathbf{H}}_i\bar{\mathbf{u}}$
Safe Approx. I	A1 : $\|\bar{\mathbf{u}}\| \leq \frac{-\sqrt{2}}{\rho(\upsilon)\,\xi_i}\max\left[(\mathbf{A}_i\mathbf{A}_i^\mathrm{T})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}})\right], \quad \rho(\upsilon) = -\sqrt{2}\,\mathrm{erfc}^{-1}\big(2\sqrt{1-\upsilon}\big)$
Safe Approx. II	A2 : $\|\bar{\mathbf{u}}\|\mathbf{1} \preceq \frac{-\sqrt{2}}{\psi(\upsilon)\,\xi_i}(\mathbf{A}_i\mathbf{A}_i^\mathrm{T})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}}), \quad \psi(\upsilon) = \mathrm{erfc}^{-1}(\upsilon/4)$
Sphere Bounding	B : $\|\bar{\mathbf{u}}\|\mathbf{1} \preceq \frac{-\sqrt{2}}{\alpha(\upsilon)\,\xi_i}(\mathbf{A}_i\mathbf{A}_i^\mathrm{T})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}}), \quad \alpha(\upsilon) = \sqrt{\Phi_2^{-1}(1-\upsilon)}$

a probability of $1 - \upsilon$. As a result, the restriction

$$\alpha(\upsilon)\mathbf{1} + \bar{\mathbf{w}}_i(\bar{\mathbf{u}}) \preceq \mathbf{0}, \tag{6.49}$$

implies that (6.46) holds true for all $\bar{\mathbf{q}}_i \in \mathcal{K}$. Finally, the worst-case robust approximation (6.49) can be expressed by an SOC constraint as

$$\mathrm{B}: \quad \|\bar{\mathbf{u}}\|\,\mathbf{1} \leq \frac{-\sqrt{2}}{\alpha(\upsilon)\,\xi_i}(\mathbf{A}_i\mathbf{A}_i^\mathrm{T})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}}). \tag{6.50}$$

In particular, the convex approximation B is able to control the radius $\alpha(\upsilon)$ according to the tolerable violation probability. It can be seen by comparing (6.44) and (6.50) that A2 resembles the sphere bounding based approximation B in form. Based on this observation, the safe approximation method for $\upsilon \in (0, 1/2]$ can be considered as defining the convex set $\mathcal{K}$ as a ball with a radius different from $\alpha(\upsilon)$, and therefore, with a different level of conservatism. In the next subsection, we compare the tightness of the proposed approximations with respect to the sphere bounding approach.

6.3.3 Relative Tightness Comparison

Up until this point, we have derived deterministic tractable convex approximations that, though not exactly, sufficiently ensure the satisfaction of the robust CI constraint of interest. This tractability led us to sacrifice tightness with respect to the originally intractable chance constraint (6.31). It is therefore desirable to find the formulation provides the tightest approximation among all the other ones.

Having rather similar conic representations for the three stochastic robust CI constraints, which are summarized in Table 6.1, enables us to compare the relative tightness of the derived convex approximations. Here, we specifically define the relative tightness from the transmit power point of view according to which a convex approximation is a tighter one if it admits lower optimal transmit powers $\|\bar{\mathbf{u}}\|^2$. We use the following two lemmas in the sequel, where the proofs are straightforward.

Lemma 12. *Let* $\bar{\mathbf{u}}^*$ *be feasible to*

$$\|\bar{\mathbf{u}}\| \, \mathbf{1} \preceq \frac{-\sqrt{2}}{\beta \, \xi_i} (\mathbf{A}_i \mathbf{A}_i^{\mathrm{T}})^{-1/2} \mathbf{w}_i(\bar{\mathbf{u}}), \tag{6.51}$$

with $\beta > 0$*, and satisfy* $\bar{\mathbf{w}}_i(\bar{\mathbf{u}}^*) \leq \mathbf{0}$ *as a necessary condition. Then, it is implied that*

$$\|\bar{\mathbf{u}}^*\| \leq \frac{-\sqrt{2}}{\beta \, \xi_i} \max \left[(\mathbf{A}_i \mathbf{A}_i^{\mathrm{T}})^{-1/2} \mathbf{w}_i(\bar{\mathbf{u}}^*) \right] \tag{6.52}$$

where $\max[\,\cdot\,]$ *denotes the elementwise maximum of an input vector.*

Lemma 13. *Consider the constraint*

$$\|\bar{\mathbf{u}}\| \leq \frac{-\sqrt{2}}{\beta \, \xi_i} \max \left[(\mathbf{A}_i \mathbf{A}_i^{\mathrm{T}})^{-1/2} \mathbf{w}_i(\bar{\mathbf{u}}) \right]. \tag{6.53}$$

where $\beta > 0$*. Let* $\bar{\mathbf{u}}^*$ *be feasible to* (6.53) *with* $\beta = \beta_1 > 0$*, then for any* $\beta_1 \geq \beta_2 > 0$*, the following chain of inequalities holds:*

$$\begin{aligned} \|\bar{\mathbf{u}}^*\| &\leq \frac{-\sqrt{2}}{\beta_1 \, \xi_i} \max \left[(\mathbf{A}_i \mathbf{A}_i^{\mathrm{T}})^{-1/2} \mathbf{w}_i(\bar{\mathbf{u}}^*) \right] \\ &\leq \frac{-\sqrt{2}}{\beta_2 \, \xi_i} \max \left[(\mathbf{A}_i \mathbf{A}_i^{\mathrm{T}})^{-1/2} \mathbf{w}_i(\bar{\mathbf{u}}^*) \right], \end{aligned} \tag{6.54}$$

which implies that $\bar{\mathbf{u}}^*$ *is feasible to* (6.53) *with* $\beta = \beta_2$*.*

It follows immediately from Lemma 12 and Lemma 13 that a relative comparison of the convex approximations A1, A2 and B boils down to comparing $\rho(\upsilon)$, $\psi(\upsilon)$ and $\alpha(\upsilon)$. These three functions, however, depend on the violation probability υ, as depicted in Fig. 6.1 for $\upsilon \in (0, 1/2]$. It can be observed from Fig. 6.1 that for small values of υ below ~ 0.12, which is of high practical interest, we have $\psi(\upsilon) \leq \rho(\upsilon) \leq \alpha(\upsilon)$. This means that a feasible solution to B is also feasible for A1 and A2, i.e., the optimal transmit power $\|\bar{\mathbf{u}}^*\|^2$ obtained from A1 and A2 is no larger than that obtained from B. Therefore, the robust convex approximations A1 and A2 are tighter (hence, less conservative) than our benchmark B. In a more precise order,

$$\mathcal{F}_{\mathrm{B}} \subseteq \mathcal{F}_{\mathrm{A1}} \subseteq \mathcal{F}_{\mathrm{A2}}, \tag{6.55}$$

where $\mathcal{F}_{(\cdot)}$ denotes the feasible region. It also follows from (6.55) that A2 is tighter than A1 in this range of υ, i.e., under strict robustness settings. On the other hand, for higher values of υ up to 1/2, which can be regarded as relaxed robustness conditions (but of course might be of less importance in a real system), we have $\rho(\upsilon) \leq \psi(\upsilon) \leq \alpha(\upsilon)$. This implies that A1 provides a tighter convex approximation than A2 in the high violation probability regime, but still A2 is tighter than the benchmark approximation.

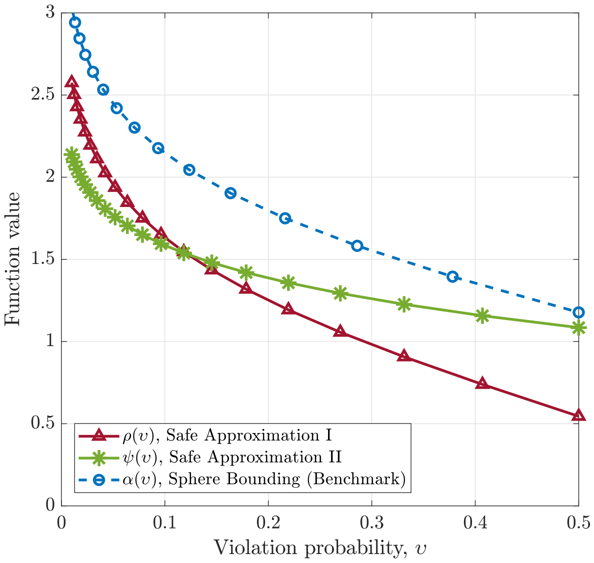


Figure 6.1: Plot of $\rho(\upsilon)$, $\alpha(\upsilon)$ and $\psi(\upsilon)$ as a function of the violation probability.

Table 6.2: Complexity comparison of the non-robust and the proposed robust SLP designs.

Design	Complexity order $\left[\times \ln(\frac{1}{\epsilon})\right]$	Dominating order [as $N_\mathrm{t}, N_\mathrm{u} \to \infty$]
P1	$2\sqrt{2N_\mathrm{u}+2}\,.\,\mathcal{O}\left((2N_\mathrm{t}+1)^3+2N_\mathrm{u}(N_\mathrm{t}+1)(2N_\mathrm{t}+1)\right)$	$\sqrt{N_\mathrm{u}}\,.\,\mathcal{O}\left(N_\mathrm{t}^3\right)\ln(\frac{1}{\epsilon})$
RP1	$2\sqrt{6N_\mathrm{u}+2}\,.\,\mathcal{O}\left((N_\mathrm{u}+1)(2N_\mathrm{t}+1)^3+2N_\mathrm{u}(2N_\mathrm{t}+1)(N_\mathrm{t}+1)\right)$	$N_\mathrm{u}\sqrt{N_\mathrm{u}}\,.\,\mathcal{O}\left(N_\mathrm{t}^3\right)\ln(\frac{1}{\epsilon})$

6.4 Robust SLP Optimization Problem

We formulate robust optimization problems for the power minimizing symbol-level precoder using the proposed robust implications of the CI constraints obtained in the previous section. First, recall the original (non-robust) formulation of the SLP design problem in (6.8). By introducing a slack variable $p \geq 0$, it is possible to recast (6.8) as

$$\begin{aligned} \mathrm{P1}: \quad & \min_{\bar{\mathbf{u}}, p \geq 0} \quad p \\ & \quad \text{s.t.} \quad \mathbf{A}_i \mathbf{H}_i \bar{\mathbf{u}} \geq \sigma_i \sqrt{\gamma_i}\, \mathbf{A}_i \mathbf{s}_i, \; i = 1, 2, ..., N_\mathrm{u}, \\ & \qquad\quad\; \bar{\mathbf{u}}^\mathrm{T} \bar{\mathbf{u}} \leq p, \end{aligned} \tag{6.56}$$

which is a more convenient form for the subsequent computational complexity discussion in this section. On the other hand, the robust counterpart of P1 can simply be expressed by replacing the actual CI constraints with any of the robust constraints W, A1, A2, or B for all the users, i.e., all N_u CI constraints must be implied through same type of convex restrictions. For example, adopting safe approximations of type A2, we can write the corresponding stochastic robust design formulation as

$$\begin{aligned}\text{RP1}:\ \min_{\bar{\mathbf{u}},p\geq 0}\quad & p\\ \text{s.t.}\quad & \|\bar{\mathbf{u}}\|\mathbf{1}\leq \frac{-\sqrt{2}}{\psi(\upsilon)\,\xi_i}(\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}}),\ i=1,2,...,N_{\mathrm{u}},\\ & \mathbf{w}_i(\bar{\mathbf{u}})\leq\mathbf{0},\ i=1,2,...,N_{\mathrm{u}},\\ & \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}}\leq p,\end{aligned}\tag{6.57}$$

The robust constraints W, A1, A2 and B, as summarized in Table 6.1, can all be represented as SOC constraints. Therefore, the robust optimization problem RP1 belongs to the class of second-order cone programming (SOCP). However, it is important to note that while the non-robust formulation P1 is always feasible, its robust counterpart RP1 may not share this property. To be more specific, there would be situations (e.g., with relatively small υ or large ε_i or ξ_i^2) in which the robust CI constraints cannot all be satisfied with a finite transmit power p, leading to a practically infeasible robust SLP design. In such cases, the intersection of all N_{u} robust CI constraints is an empty set.

6.4.1 Computational Complexity Analysis

We evaluate the computational complexity of the proposed robust design formulations based on the worst-case complexity analysis provided in [171], and compare the results with those of the original non-robust formulation. All the stochastic robust formulations are presented as SOCPs, which can efficiently be solved via interior-point methods. In general, the arithmetic complexity of a generic interior-point method entails the Newton complexity as well as per-iteration computation cost. The Newton complexity basically refers to the number of steps required to reduce the duality gap by a constant factor, while the per-iteration complexity involves finding a new search direction at each step, and is subsequently dominated by the computation cost to assemble and solve a linear system of equations.

In what follows, we briefly overview the complexity bound for an SOCP given in a generic form containing linear and (conic) quadratic constraints, to reach an ϵ-solution (i.e., an ϵ-optimal feasible solution) via a generic interior-point method. Given the SOCP

$$\begin{aligned}\underset{\mathbf{x}}{\text{minimize}}\quad & \mathbf{c}_0^{\mathrm{T}}\mathbf{x}\\ \text{s.t.}\quad & \|\mathbf{F}_k\mathbf{x}+\mathbf{b}_k\|\leq\mathbf{f}_k^{\mathrm{T}}\mathbf{x}+g_k,\ k=1,2,...,m,\\ & \mathbf{c}_j^{\mathrm{T}}\mathbf{x}\leq d_j,\ j=1,2,...,l,\end{aligned}\tag{6.58}$$

where $\mathbf{F}_k\in\mathbb{R}^{n_k\times n},\mathbf{b}_k\in\mathbb{R}^{n_k},\mathbf{f}_k\in\mathbb{R}^n,g_k\in\mathbb{R}$ for all $k=1,2,...,m$, and $\mathbf{c}_j\in\mathbb{R}^n,d_j\in\mathbb{R}$ for $j=1,...,l$, the complexity bound of an ϵ-solution is of order

$$\mathcal{C}(\epsilon)=n\sqrt{l+2m}\left(n^2+l(n+1)+\sum_{k=1}^{m}n_k^2\right).\mathcal{O}(1).\ln\left(\frac{1}{\epsilon}\right).\tag{6.59}$$

In the SOCP (6.58), n can be treated as the total number of optimization variables, and n_k determines the size of the kth cone constraint, which is related to the dimension of the kth second-order cone, for all $k = 1, 2, ..., m$. Note that this generic form of SOCP covers also the non-robust design formulation in (6.56). Based on the above analysis, we are able to evaluate the complexity of the robust SOCP design formulation (6.57), and compare it to that of its non-robust counterpart in (6.56). We also remark that

i. There are two real-valued second-order cone constraints associated with each user.

ii. The slack variables p in (6.57) can be merged into the vector $\bar{\mathbf{u}}$, increasing the ith cone's dimension by one for all $k = 1, 2, ..., m$.

Accordingly, for all the design formulations with either of the robust constraints W, A1, A2, or B, the number of variables is equal to $2N_\mathrm{t}+1$. The non-robust formulation (6.56) has $2N_\mathrm{u}+1$ linear inequalities plus one cone constraint of size $2N_\mathrm{t}+1$, while the robust design (6.57) involves $2N_\mathrm{u}$ conic constraints of size $2N_\mathrm{t}$ and one conic constraint of size $2N_\mathrm{t}+1$ which corresponds to the power constraint. In Table 6.2, we report the final computational complexity evaluations, where the dominating orders represent the largest complexity growth rate as $N_\mathrm{t}, N_\mathrm{u} \to \infty$ under the assumption $N_\mathrm{u} \leq N_\mathrm{t}$. It follows from Table 6.2 that the proposed robust formulations increase the computational complexity of SLP design by an order of $\mathcal{O}(N_\mathrm{u})$, compared to their non-robust counterpart in (6.56). Nonetheless, the increase in complexity is negligible for practical values of N_u.

6.5 Simulation Results

In this section, we present our simulation results to evaluate the performance of the proposed robust SLP techniques, and further, to validate the analytical discussions provided in earlier sections. The optimization problems have been solved using MATLAB software and SeDuMi solver [172]. The following setup is adopted in all the simulation scenarios. We consider a downlink MU-MIMO system with $N_\mathrm{t} = 6$ and $N_\mathrm{u} = 4$, employing an 8-PSK modulation scheme with uncoded transmission. For all the users $i = 1, 2, ..., N_\mathrm{u}$, we assume a unit noise variance $\sigma_i^2 = \sigma^2 = 1$ and equal SINR requirements $\gamma_i = \gamma$. The erroneous channel vectors $\{\hat{\mathbf{h}}_i\}_{i=1}^{N_\mathrm{u}}$ are randomly generated according to a zero-mean unit variance circularly symmetric complex Gaussian distribution, where the channels of any two distinct users are uncorrelated, i.e., $\mathbb{E}\{\hat{\mathbf{h}}_i^\mathrm{H}\hat{\mathbf{h}}_j\} = \mathbf{0}$ for all $i, j = 1, 2, ..., N_\mathrm{u}, i \neq j$. We consider identical uncertainty regions for all the channels, i.e., $\xi_i^2 = \xi^2$ for $i = 1, 2, ..., N_\mathrm{u}$. All the presented simulation results have been averaged over 500 fading block realizations, each consisting of 500 symbols. We evaluate the performance of the symbol-level precoded downlink transmission under bounded and stochastically-known CSI errors through various measures. The SLP designs with robust CI constraints "worst-case", "safe approximation I", "safe approximation II", and "sphere bounding" are referred to as WC-SLP, SA1-SLP, SA2-SLP and SB-SLP, respectively.

In Fig. 6.2, the transmit power performance of the proposed WC-SLP design is depicted versus target SINR γ under the bounded uncertainty model with three different

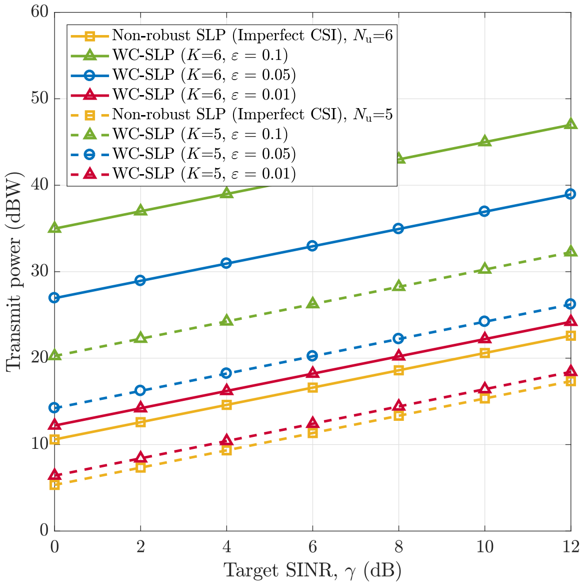


FIGURE 6.2: Average transmission power of the non-robust and the worst-case robust SLP schemes versus SINR target for a system with $N_t = 6$.

radii 0.01, 0.05 and 0.1. As it might be expected, for larger uncertainty regions, higher transmission powers are needed in order to guarantee the system/users' requirements in case of any possible realization of the bounded CSI error. Furthermore, the performance results are depicted for two system dimensions with $N_t = N_u = 6$, and $N_t = 6$ and $N_u = 5$. It follows from Fig. 6.2 that the system requires less additional power to provide robustness to bounded CSI uncertainty for a fewer number of users. For instance, in the case with $\varepsilon = 0.01$, decreasing the number of users by one results in a reduction of around 6 dBW in the average transmit power of the worst-case robust SLP. We highlight that, for PSK modulations, the WC-SLP design shows the exact same performance as that of the worst-case robust SLP scheme in [20]. However, as mentioned earlier, the SLP scheme in [20] is formulated only for constant envelope modulation schemes, whereas our proposed worst-case method does not have such a restriction and applies to a broader group of modulations.

In Fig. 6.3, a scatter plot of the noise-free received signals is illustrated for the non-robust and robust SLP schemes. The average transmission powers for the non-robust SLP scheme with erroneous CSI and the robust SA2-SLP approach are, respectively, 13.22 dBW and 15.08 dBW. It can be seen from the figure that this $\sim$ 2 dBW extra power is consumed to satisfy the CI constraints with the given violation probability, thereby providing more safety to the subsequent additive Gaussian noise. The cloud of received signals corresponding to the non-robust SLP scheme, however, shows deviations from the intended symbols towards the corresponding ML decision boundaries, which

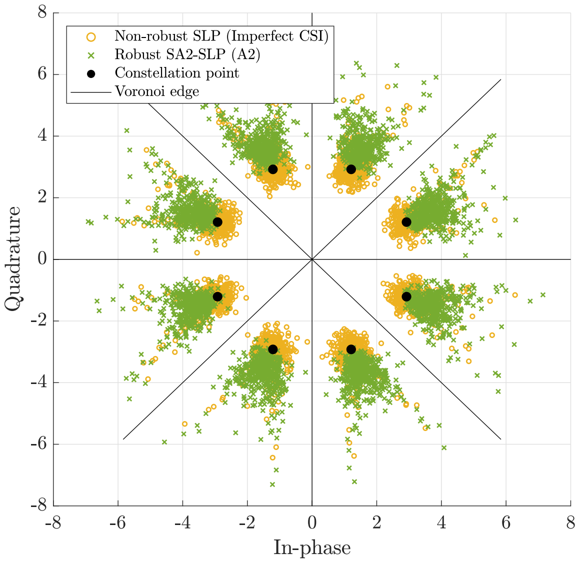


FIGURE 6.3: Scatter plot of the noise-free received signals at $\gamma = 10$ dB with a fixed channel, $\xi^2 = 0.005$ and $\upsilon = 0.05$.

may result in a higher symbol error probability (as we will see later in this section). Furthermore, the non-robust scheme may fail to satisfy the users' SINR requirements. This issue is depicted in Fig. 6.4 and 6.5, where we respectively plot the average per-user received SINR versus target SINR and the average received SINR for each user at a target value of $\gamma = 15$ dB. Given $\mathbf{H}_i$, we define the received SINR of the ith user for the SLP scheme with perfect CSI, the non-robust SLP with imperfect CSI and the stochastic robust SLP, respectively, as

$$\mathrm{SINR}_i \triangleq \frac{\mathbb{E}_{\bar{\mathbf{u}}}\left\{\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}}\right\}}{\sigma_i^2}, \tag{6.60}$$

$$\mathrm{SINR}_i \triangleq \frac{\mathbb{E}_{\bar{\mathbf{u}}}\left\{\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}}\right\}}{\sigma_i^2 + \mathbb{E}_{\bar{\mathbf{u}},\mathbf{E}_i}\left\{\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{E}_i^{\mathrm{T}}\mathbf{E}_i\bar{\mathbf{u}}\right\}}, \tag{6.61}$$

and

$$\mathrm{SINR}_i \triangleq (1-\upsilon)\left(\frac{\mathbb{E}_{\bar{\mathbf{u}}}\left\{\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}}\right\}}{\sigma_i^2}\right) + \upsilon\left(\frac{\mathbb{E}_{\bar{\mathbf{u}}}\left\{\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\bar{\mathbf{u}}\right\}}{\sigma_i^2 + \mathbb{E}_{\bar{\mathbf{u}},\mathbf{E}_i}\left\{\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{E}_i^{\mathrm{T}}\mathbf{E}_i\bar{\mathbf{u}}\right\}}\right), \tag{6.62}$$

where the expectations over $\bar{\mathbf{u}}$ and $\mathbf{E}_i$ are computed numerically. The SINR quantities in (6.60), (6.61) and (6.62) have been averaged over 1000 realizations of $\mathbf{H}_i$ to obtain the values depicted in Fig. 6.4 and Fig. 6.5. We can see from Fig. 6.4 that the given

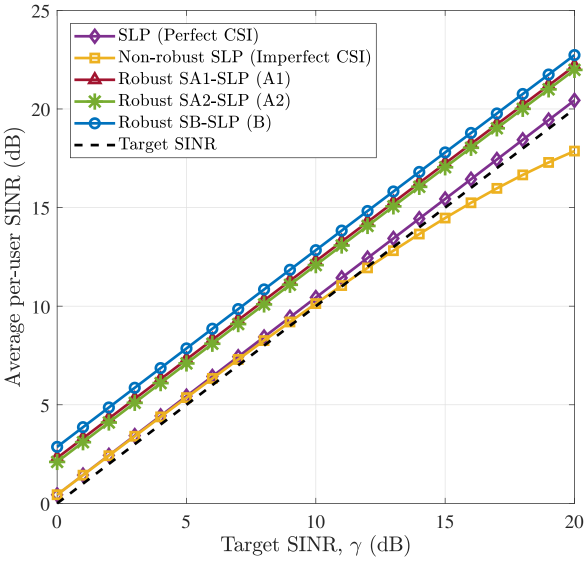


FIGURE 6.4: Average per-user received SINR versus target SINR with $\xi^2 = 0.005$ and $\upsilon = 0.05$.

target SINR is likely not to be met by all the users, particularly at high SINR values, when using the non-robust scheme. The separate bar plot of the received SINR for each user in Fig. 6.5 shows that at $\gamma = 15$ dB, the SINR requirement has not been satisfied for any of the users by employing the non-robust SLP method. On the other hand, when employing either of the robust approaches, each user's received SINR is well above the target value. This, however, means that the users are provided with higher SINRs than the required value γ, which may not be efficient in general. In a practical system design, one needs to reach a compromise based on a specific power-performance tradeoff, according to which the most efficient robust transmission scheme is preferred. We will introduce such a tradeoff and investigate the efficiencies of different approaches later in this section.

In addition to the average received SINR, we are interested in evaluating the probability with which the given target SINR of each user is met. For this purpose, we define "outage event" as a situation in which the minimum required SINR of a user can not be guaranteed. Accordingly, we define the probability of outage for user i as

$$P_{\mathrm{out},i} \triangleq \Pr\{\mathrm{SINR}_i < \gamma\}. \tag{6.63}$$

The probability of SINR outage can be equally translated to a rate-outage probability, i.e., the probability that a given target rate $\log_2(1+\gamma)$ is not achievable. This quantity is calculated over many transmissions with different channel realizations and plotted in Fig. 6.6 as a function of the target SINR γ for the non-robust/robust SLP schemes under two different scenarios with $\upsilon = 0.05$, $\xi^2 = 0.005$ and $\upsilon = 0.2$, $\xi^2 = 0.01$. Note that Fig. 6.6 shows the average probability over all the users, i.e., $\bar{P}_{\mathrm{out}} \triangleq (1/N_{\mathrm{u}}) \sum_{i=1}^{N_{\mathrm{u}}} P_{\mathrm{out},i}$.

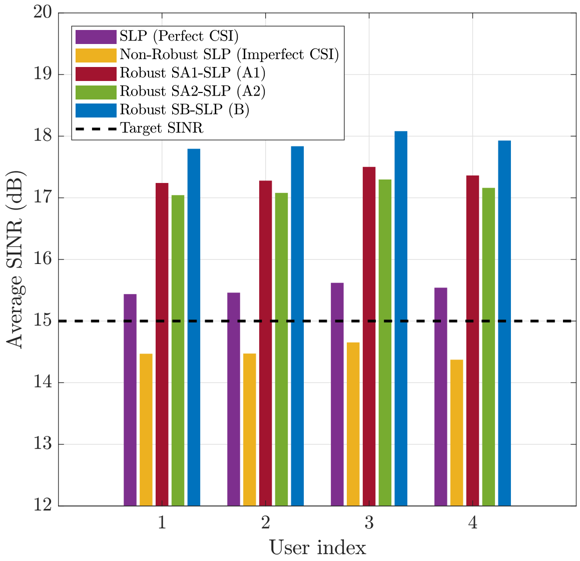


FIGURE 6.5: Average received SINR at $\gamma = 15$ dB with $\xi^2 = 0.005$ and $\upsilon = 0.05$.

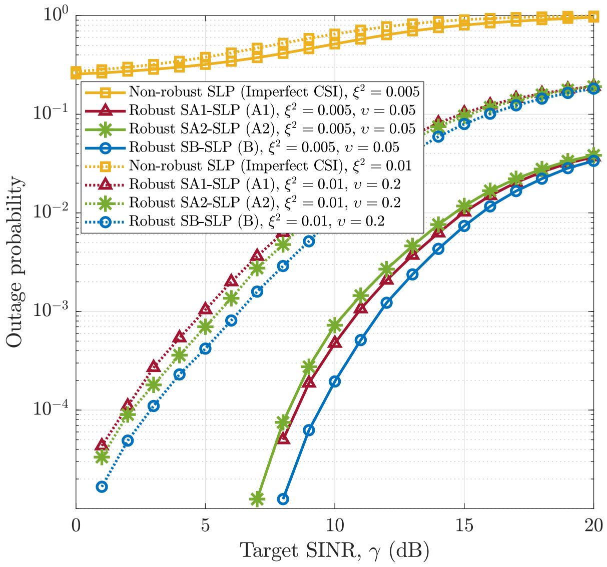


FIGURE 6.6: Probability of outage versus target SINR under two different settings with $\xi^2 = 0.005$, $\upsilon = 0.05$ and $\xi^2 = 0.01$, $\upsilon = 0.2$.

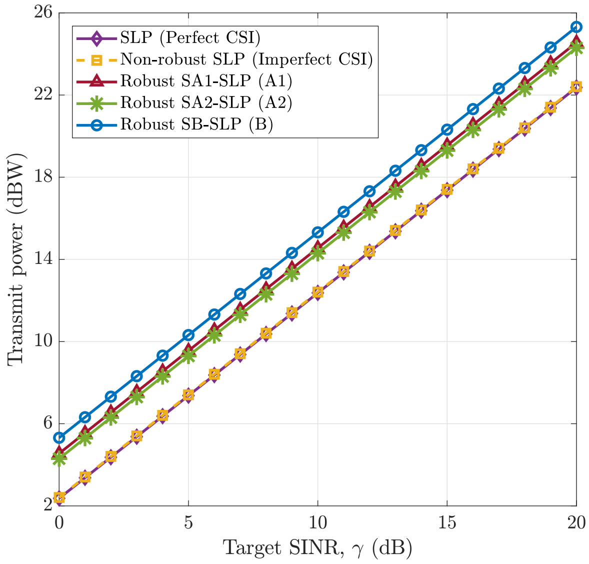


FIGURE 6.7: Average transmission power versus target SINR with $\xi^2 = 0.005$ and $\upsilon = 0.05$.

As it can be observed, the outage probability increases with γ and ξ^2. The increasing behavior of P_{out} with respect to γ can be justified from the definitions of SINR_i in (6.61) and (6.62). A larger γ results in higher transmission power, and subsequently, a greater deal of uncertainty at the receiver side (note that $\mathbf{E}_i\bar{\mathbf{u}}$ is the uncertain component at the receiver of user i). It can be seen that the conservative approach to satisfying the CI constraints taken by the robust methods can lead to significant improvement in the probability of outage compared to the non-robust scheme, i.e., the given target SINR is more probably achievable when employing a robust SLP scheme. Moreover, Fig. 6.6 shows that each of the SA1-SLP and SA2-SLP methods provides a lower probability of outage than the other under different uncertainty settings, while the benchmark SB-SLP approach achieves the lowest outage probability among all in both scenarios.

The higher received SINR and the lower outage probability provided by the robust SLP approaches are, however, achieved by consuming larger amounts of power for downlink transmission, which is inevitable to achieve the desired level of robustness. In Fig. 6.7, the average total transmit power is depicted versus target SINR, where it is shown that the robust SLP approaches require higher transmission powers than that of the non-robust scheme. A common observation from Fig. 6.5-6.7 is that among the robust SLP approaches, the more conservative method with larger transmit power results in higher average received SINR and a lower outage probability for each user.

To have a fair and meaningful comparison between the non-robust and robust SLP schemes, we need a measure that incorporates both received SINR and transmit power in evaluating the downlink performance. Similar to [173], we define “energy efficiency” as the ratio between the expected throughput and the average transmit power. Accordingly,

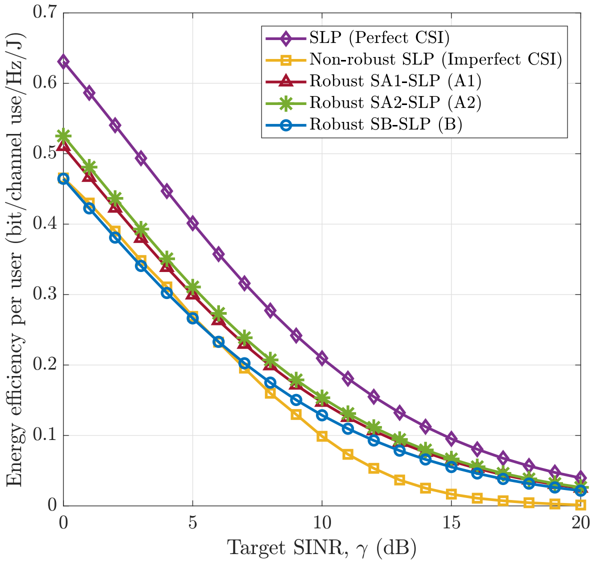


FIGURE 6.8: Energy efficiency comparison of different SLP schemes versus target SINR with $\xi^2 = 0.005$ and $\upsilon = 0.05$.

the energy efficiency for the ith user, denoted by η_i, is obtained as

$$\eta_i \triangleq \frac{(1 - P_{\text{out},i})R(\gamma)}{\|\bar{\mathbf{u}}\|^2}, \tag{6.64}$$

where $R(\gamma) = \log_2(1 + \gamma)$ refers to the achievable transmission rate corresponding to the target SINR γ. This quantity can be interpreted as the number of information bits per unit of energy that can be reliably transmitted to each user in one channel use. The average per-user energy efficiency, obtained as $\bar{\eta} \triangleq (1/N_\text{u}) \sum_{i=1}^{N_\text{u}} \eta_i$, is compared for different SLP schemes in Fig. 6.8. The results show that the proposed robust SLP designs SA1-SLP and SA2-SLP are more energy-efficient than the SLP scheme with imperfect CSI as well as the benchmark SB-SLP method. Furthermore, the SA2-SLP design is slightly more energy efficient than SA1-SLP for this particular choice of υ, as suggested by our tightness analysis in Section 6.3.3. However, we should note that this superiority is obtained in exchange for higher transmitter complexity, as discussed in Section 6.4.1.

We also plot in Fig. 6.9 the average per-user symbol error probability obtained by different SLP schemes as a function of SINR requirement γ. Having imperfect CSI, it can be seen that the non-robust and robust methods both show an error floor at high target SINRs. However, in the whole depicted range of SINR, the robust SLP approaches have lower symbol error rates compared to the non-robust scheme. Furthermore, as it might be expected, increasing ξ^2 and υ results in a degraded symbol error rate for the users. In fact, the lower symbol error rate achieved by the robust SLP methods is an advantage of introducing the (robust) CI constraints into the precoder optimization problem.

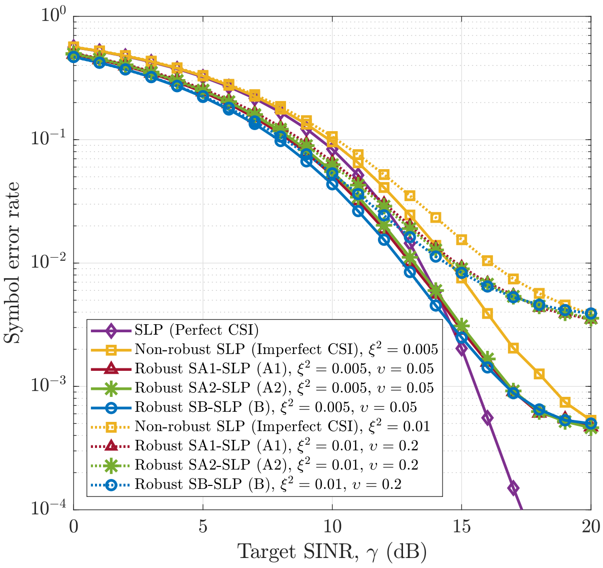


FIGURE 6.9: Average symbol error rate per user versus target SINR for two different scenarios with $\xi^2 = 0.005$, $\upsilon = 0.05$ and $\xi^2 = 0.01$, $\upsilon = 0.2$.

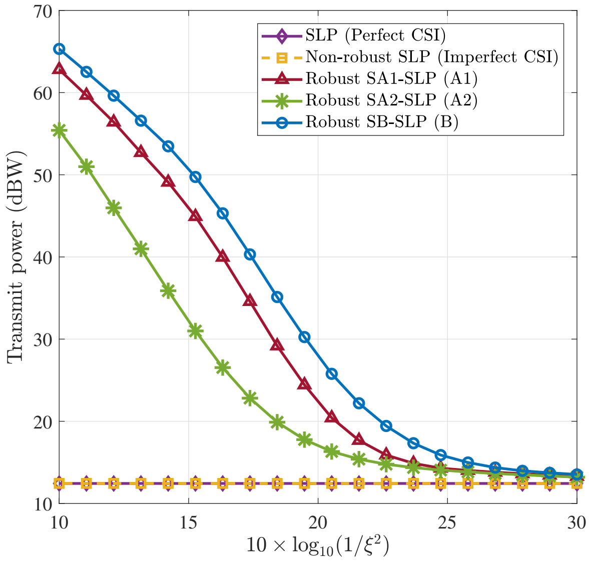


FIGURE 6.10: Average transmission power as a function of uncertainty variance with $\gamma = 10$ dB and $\upsilon = 0.05$.

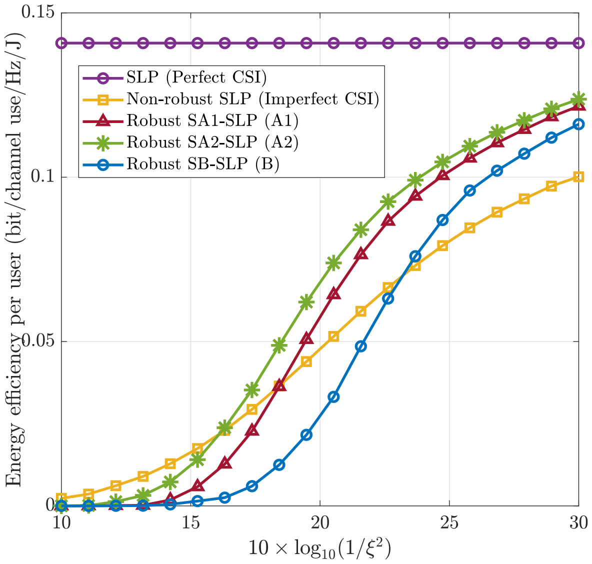


FIGURE 6.11: Per-user energy efficiency as a function of uncertainty variance with $\gamma = 10$ dB and $\upsilon = 0.05$.

In order to evaluate the effect of the environment parameter ξ^2 on the performance of the symbol-level precoded downlink transmission, in Fig. 6.10 and Fig. 6.11, we respectively plot the average transmit power and the energy efficiency versus ξ^2 in an inverse logarithmic scale. From Fig. 6.10, it can be inferred that for large noise variances, i.e., more severe uncertainty conditions, the robust SLP approaches consume relatively high powers for transmission to ensure a certain level of robustness, while the required transmit power tends to that of the case with perfect CSI as ξ^2 decreases. Fig. 6.11, on the other hand, shows that the energy efficiency under imperfect CSI has an inverse relation to ξ^2, i.e., the smaller the noise variance is, the more efficient the SLP scheme will be. This statement is true for both non-robust and robust designs. Although the non-robust scheme shows a superior energy efficiency for large values of ξ^2, the SA2-SLP design outperforms the non-robust scheme for $\xi^2 < 0.025$, i.e., $10 \log_{10}(1/\xi^2) > 16$ dB in logarithmic scale. Indeed, all the robust approaches are more energy-efficient than the non-robust case for relatively small values of the uncertainty variance, i.e., $\xi^2 < 0.005$ corresponding to $10 \log_{10}(1/\xi^2) > 23$ dB.

We mentioned earlier in Section 6.4 that the robust design RP1 might be infeasible for some values of the violation probability υ and the noise variance ξ^2. In particular, having $\upsilon \to 0$ and/or a relatively large value for ξ^2 (compared to the spectral norm of the overall channel matrix, i.e., $\|\mathbf{H}\|_2$) increases the probability of RP1 being infeasible. In a practical system, a higher rate of feasibility may be reflected in higher service availability to the users. We evaluate this issue through approximating the feasibility rate of the robust SLP approaches over several channel/error/symbol realizations, as shown

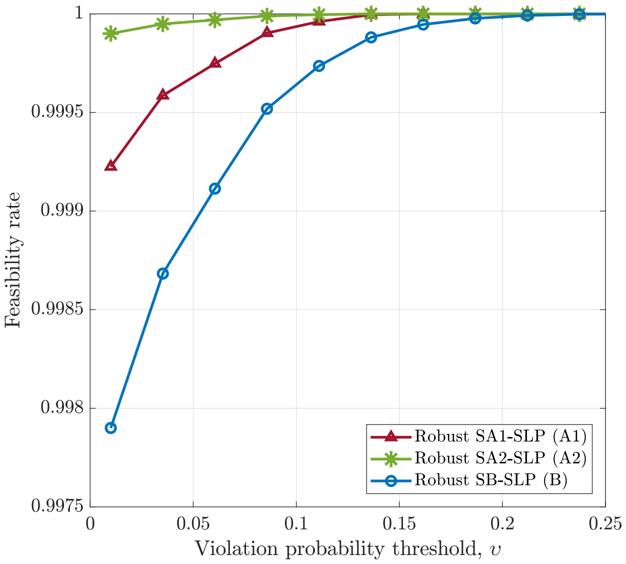


Figure 6.12: Feasibility rate as a function of violation probability at $\gamma = 10$ dB with $\xi^2 = 0.01$.

in Fig. 6.12 as a function of υ. We can see from the figure that both proposed robust SLP designs are feasible, on average, above %99 of the time even for relatively small values of υ (i.e., higher levels of conservatism). Apart from the robust design approach, this high feasibility rate is one of the advantages of the symbol-level precoder over conventional block-level techniques, mainly due to higher available degrees of freedom in designing the precoder.

Finally, we compare our results with those obtained from the robust block-level precoding scheme proposed in [159], referred to as "robust BLP", which solves a convex semidefinite programming (SDP) to minimize the average transmit power for a given target SINR γ. It is important to note that the robust BLP approach is barely feasible for large γ and ξ^2 as well as small values of υ (as we will show in Fig. 6.16). Therefore, in what follows, we present the results for some limited scenarios with sufficiently small γ and ξ^2 and large enough υ. Furthermore, we average the results obtained from the robust BLP scheme only over those realizations for which the SDP optimization problem in [159] is feasible.

The scatter plot of the noise-free received signals resulted from the block-level, and symbol-level precoding approaches of interest is shown in Fig. 6.13 for a given target SINR of $\gamma = 5$ dB. In this figure, the average transmit powers of the robust BLP, non-robust SLP and robust SA2-SLP schemes are equal to 8.16 dBW, 8.85 dBW, and 11.14 dBW, respectively. The centroids of the received signal clouds corresponding to the robust BLP approach are farther away from the original constellation points, which is an expected result of conservative precoding design. This, in turn, increases the consumed

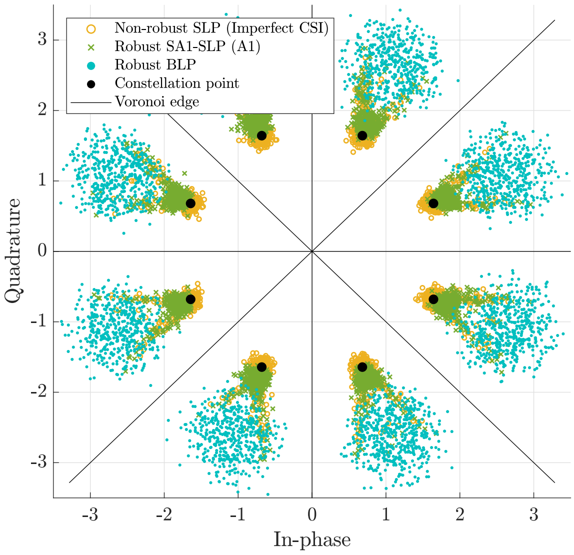


FIGURE 6.13: Scatter plot of the noise-free received signals at $\gamma = 5$ dB with a fixed channel, $\upsilon = 0.1$ and $\xi^2 = 0.001$.

transmit power and reduces the energy efficiency for high target SINR values, as we will see later.

In Fig. 6.14, we compare the energy efficiency of different non-robust/robust precoding schemes as a function of the target SINR γ. Using the robust BLP method, for given $\mathbf{h}_i$, the received SINR of the ith user is given by

$$\mathrm{SINR}_i \triangleq \frac{\mathbf{p}_i^{\mathrm{H}} \mathbf{h}_i^{\mathrm{H}} \mathbf{h}_i \mathbf{p}_i}{\sigma_i^2 + \sum_{j \neq i} \mathbf{p}_j^{\mathrm{H}} \mathbf{h}_i^{\mathrm{H}} \mathbf{h}_i \mathbf{p}_j}, \tag{6.65}$$

where $\mathbf{p}_i$ is the precoding vector that corresponds to the user i. It can be seen from Fig. 6.14 that the robust BLP scheme is more energy-efficient than all the robust SLP approaches at low target SINRs up to around 3 dB. Recall that the results are averaged only over those realizations for which the robust BLP is feasible, i.e., we do not take the infeasibility rate into account in our performance comparisons. On the contrary, for moderate-to-high SINR values, the proposed SLP approaches outperform the robust BLP scheme. Notice also that the robust BLP scheme's optimization problem was infeasible in all our trials with $\gamma \geq 14$ dB. This is mainly due to the fact that the robust BLP scheme requires an infinite transmit power (i.e., the optimization problem is practically infeasible) for target SINRs larger than a specific value. However, the feasibility of the proposed robust SLP approaches does not depend on γ. Furthermore, the energy efficiency of the precoding schemes of interest as a function of the violation probability is plotted in Fig. 6.15, where it is shown that the proposed robust SLP approaches outperform the robust BLP method for all values of $\upsilon \in (0, 1/2]$ in the considered

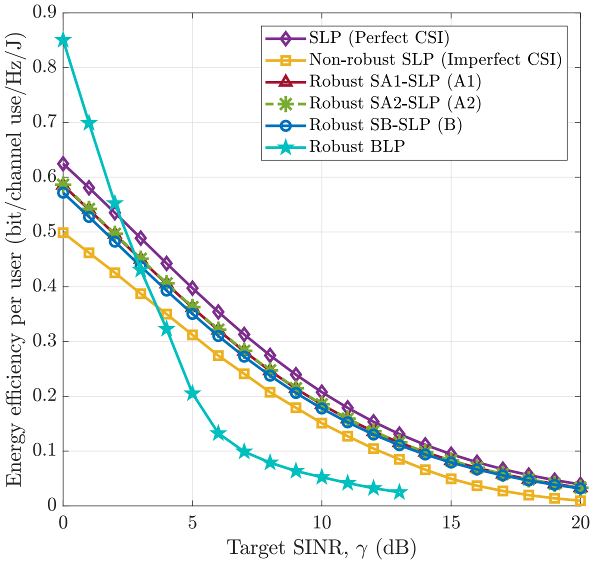


FIGURE 6.14: Energy efficiency comparison of BLP and SLP schemes versus target SINR with $\upsilon = 0.1$ and $\xi^2 = 0.001$.

setting. It further follows from Fig. 6.15 that the proposed robust SLP approaches' energy efficiency tend to that of the SLP scheme with perfect CSI as υ increases.

The feasibility rates of the robust block-level and symbol-level precoders are compared in Fig. 6.16 as a function of the uncertainty variance ξ^2 in an inverse logarithmic scale. As it can be seen, both SA1-SLP and SA2-SLP methods are feasible more than %93 of the time in the whole evaluated range of ξ^2. In particular, both our proposed robust approaches are %100 feasible for $\xi^2 < 0.015$, or $10\log_{10}(1/\xi^2) > 18$ dB. The robust BLP scheme, on the other hand, is %50 or higher feasible only for $\xi^2 < 0.003$, i.e., $10\log_{10}(1/\xi^2) > 25$ dB, while it appears to be barely feasible for uncertainty variances larger than 0.01.

It should be noted that the improved feasibility rate and energy efficiency of an SLP design compared to a block-level scheme is obtained at the cost of per-symbol optimization of the precoded signal, which may lead to higher transmitter complexity. To have an illustrative comparison of complexity, consider the robust BLP method in [159]. This method needs to solve an optimization problem with SDP and SOC constraints of dimension $2(2N_\mathrm{t}+1)(N_\mathrm{u}+1)$ and $4N_\mathrm{t}N_\mathrm{u}+1$, respectively. Roughly speaking, the worst-case complexity of finding an ϵ-optimal solution via a standard interior-point method is of order $\mathcal{O}(N_\mathrm{u}^6 N_\mathrm{t}^6)\ln(1/\epsilon)$, where such a solution has to be obtained once the CSI is updated. On the other hand, the arithmetic complexity of the proposed robust SLP approaches have been shown to be $\mathcal{O}(N_\mathrm{u}\sqrt{N_\mathrm{u}}N_\mathrm{t}^3)\ln(1/\epsilon)$; see Table 6.2. We recall that the symbol-level precoded transmit signal needs to be redesigned for every instantaneous set of users' symbols or the total number of possible symbol realizations for N_u users,

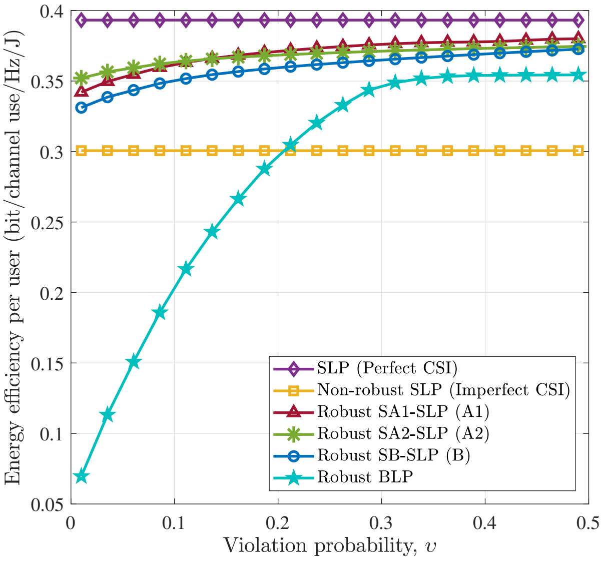


FIGURE 6.15: Per-user energy efficiency as a function of violation probability with $\gamma = 5$ dB and $\xi^2 = 0.001$.

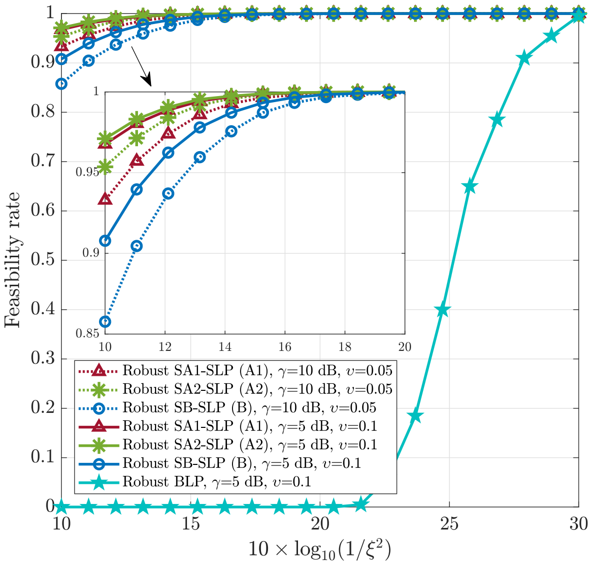


FIGURE 6.16: Feasibility rate of different robust precoding schemes as a function of uncertainty variance under two settings with $\gamma = 5$ dB, $\upsilon = 0.1$ and $\gamma = 10$ dB, $\upsilon = 0.05$.

i.e., M^{N_u} where M is the modulation order. Denoting by S the number of information symbols per a single transmitted frame, the overall (per CSI update) complexity of an SLP scheme can be approximated as $\min\{S, M^{N_\mathrm{u}}\}.\mathcal{O}(N_\mathrm{u}\sqrt{K}N_\mathrm{t}^3)\ln(1/\epsilon)$. Hence, a relative computation cost between the robust SLP and BLP methods, in the limiting case, is given by the ratio $\min\{S, M^{N_\mathrm{u}}\}/N_\mathrm{u}^4\sqrt{N_\mathrm{u}}N_\mathrm{t}^3$. In particular, for a moderate number of users and low-order modulation schemes, the computational cost of a symbol-level precoder can be alleviated by an offline optimization of the precoded signals and using a lookup table for downlink transmission [21]. Further, it might be possible to derive a low-complexity (semi closed-form) solution for the robust SLP approaches, similar to those obtained in [40,98,112] for the original SINR-constrained SLP power minimization problem, which can be the topic of future work.

6.6 Conclusions

We addressed the design problem of robust symbol-level precoded transmission in a downlink MU-MIMO system under imperfect bounded or stochastic CSI error at the BS. We considered a QoS-constrained design criterion aimed at minimizing the total (per-symbol) transmit power subject to CI constraints as well as given target SINRs. We developed robust CI constraints for each channel uncertainty model and provided the corresponding robust formulations for the SLP design problem. With bounded CSI errors, we derived a worst-case robust formulation to guarantee the users' requirements for every possible realization of the CSI error within the uncertainty region. Under the stochastic uncertainty model, we adopted a probabilistic approach to imply the optimization constraints, which led us to intractable expressions. We tackled this difficulty by deriving two computationally tractable approximate convex constraints with different levels of conservatism. A benchmark approximation was also derived based on the sphere bounding conservative method. Our analytical and simulation results showed that both the proposed robust convex approximations outperform the benchmark, while each of them is superior to the other under different robustness settings. Compared with a conventional block-level robust scheme, the proposed robust methods were shown to be more efficient at moderate-to-high target SINR values. However, a more considerable advantage of the proposed robust SLP approaches is their higher feasibility rate for wide ranges of violation probability and uncertainty variance, which is indifferent to the target SINR. We also highlight from our complexity analysis that the improved performances of the proposed robust SLP designs come with an increased computational complexity by an order of the number of users in the limiting case.

Chapter 7

Robust Symbol-Level Precoding under System Uncertainties – Part II: Design Uncertainty

This chapter addresses the optimization problem of SLP in an MU-MIMO downlink wireless system where the precoder's output is subject to partially-known distortions. In particular, we assume a linear distortion model with bounded additive noise. The original SINR-constrained SLP problem minimizing the total transmit power is first reformulated as a penalized unconstrained problem, referred to as the relaxed robust formulation. We then adopt a worst-case design approach to protect the users' intended symbols and the targeted CI with a desired level of confidence. Due to the non-convexity of the relaxed robust formulation, we propose an iterative algorithm based on the block coordinate ascent-descent method. We show through simulation results that the proposed robust design is flexible in the sense that the CI constraints can be relaxed to keep a desirable balance between achievable rate and power consumption. Remarkably, the robust formulation yields more energy-efficient solutions for appropriate choices of the penalty parameter, compared to the original SLP problem.

7.1 Introduction

As mentioned in earlier chapters, a key consideration in designing the symbol-level precoder is to properly define the CIRs based on the received signal constellation, typically with the aim of preserving (or enhancing) the detection accuracy. This type of design, however, is highly sensitive to inaccuracies in several parameters, such as the available CSI at the transmitter, the receive noise power, and any succeeding operation on the transmit signal, which is not perfectly known to the precoder. In particular, considering (non)linear distortions of the precoded signal, which falls within the third category, is the main focus of this chapter. The distorted transmit signal may reflect the effects of non-ideal elements either in the digital domain, e.g., low-resolution digital-to-analog converters (DAC), or in the RF chain, e.g., power amplifiers [174]. Furthermore, it could be an adequate model for the source-relay link over a relay channel, e.g., non-ideal feeder link in a satellite communication system [175].

There has been some effort in addressing the SLP design problem in the presence of system uncertainties. Robust symbol-level precoders under imperfect CSI are presented in [20, 58, 64, 116, 117]. Furthermore, in [45], the authors propose an SLP design with outage probability constraints to achieve robustness against the receiver noise. To the best of our knowledge, the SLP design problem under linear distortion of the precoded signal has not been addressed in the literature. In this work, by assuming a linearly distorted signal model with bounded additive distortion, we aim to design an SLP scheme such that the performance gain offered by the CI-based design is preserved. In particular, we reformulate a version of the original problem with penalized objective function and use this reformulation in a worst-case design approach. The penalty coefficient in the new formulation allows us to keep a balance between the desired level of spectral efficiency/users' symbol error probability and the consumed power.

It is worth mentioning that the problem of robust design has been widely studied in the literature for scenarios where our knowledge about the environment is subject to uncertainty [176–183]. In this chapter, we assume that our design process is subject to uncertainty, e.g., due to finite precision of the underlying design and implementation technology. This work can point the research community to address new practical challenges in robust design when the design parameters are subject to uncertainty.

The rest of this chapter is organized as follows. In Section 7.2, we describe the system and signal distortion models. After a brief overview of the original SLP problem formulation, in Section 7.3, we reformulate and discuss the worst-case design problem and present our proposed algorithm. We present the simulation results in Section 7.4. Finally, we conclude the chapter in Section 7.5.

7.2 System Model and Problem Definition

We consider an MU-MIMO downlink system with the same transmission scheme as described in Chapter 3, where an N_t-antenna BS communicates with N_u single-antenna

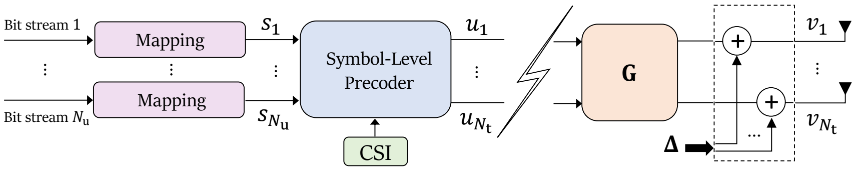


FIGURE 7.1: The considered system model where the output of the symbol-level precoder is subject to linear distortion before being transmitted to the users.

users ($N_\mathrm{u} \leq N_\mathrm{t}$) via sending independent data symbols $\{s_i\}_{i=1}^{N_\mathrm{u}}$ in the same time-frequency resource block. The BS employs a symbol-level (non-linear) precoding scheme to spatially multiplex the users' data streams in the downlink transmission, in which the $N_\mathrm{t} \times 1$ precoded signal $\mathbf{u} = [u_1, u_2, ..., u_{N_\mathrm{t}}]^\mathrm{T}$ is redesigned every symbol period by solving an optimization problem. It is further assumed that the precoded signal is subject to linear distortion before transmission, i.e., the actual $N_\mathrm{t} \times 1$ transmitted signal $\mathbf{v}$ is given by

$$\mathbf{v} = \mathbf{G}\mathbf{u} + \boldsymbol{\Delta}, \tag{7.1}$$

where $\mathbf{G} \in \mathbb{C}^{N_\mathrm{t} \times N_\mathrm{t}}$ denotes a known distortion matrix and $\boldsymbol{\Delta} \in \mathbb{C}^{N_\mathrm{t} \times 1}$ represents an additive white noise which is uncorrelated with the precoder's output $\mathbf{u}$. Such a model is particularly suitable for a relayed transmission scheme. For example, interference mitigation techniques in the forward link of a satellite communication system may take the form of on-ground precoding, i.e., the users' data streams are pre-processed at the gateway and then sent to the satellite through the feeder link [184, 185]. The received signal by the satellite (to be transmitted towards the users) can be modeled as (7.1), where $\mathbf{G}$ represents signal attenuation generated by either the atmospheric fading and/or the feed antenna radiation, and $\boldsymbol{\Delta}$ models the additive noise at the satellite's array-fed reflector. Another possible application of (7.1) could be in a massive MU-MIMO scenario where the continuous-valued precoding coefficients $\{u_j\}_{j=1}^{N_\mathrm{t}}$ are passed through low-resolution digital-to-analog converters (DAC) to be quantized in the digital domain before up-conversion via the RF chains. The non-linear quantization operation can be approximated by the additive quantization noise (AQN) model, [186, 187], which coincides with the linear distortion model in (7.1). Under the above assumptions, the baseband representation of the signal received by the ith user is given as

$$r_i = \mathbf{h}_i \mathbf{v} + z_i = \mathbf{h}_i(\mathbf{G}\mathbf{u} + \boldsymbol{\Delta}) + z_i, \quad i = 1, 2, ..., N_\mathrm{u}, \tag{7.2}$$

where $\mathbf{h}_i \in \mathbb{C}^{N_\mathrm{t} \times 1}$ contains the instantaneous fading coefficients of the quasi-static channel between the transmit antennas and the ith user, and $z_i \sim \mathcal{CN}(0, \sigma_i^2)$ models the additive thermal noise at the ith receive front-end.

To proceed, we define equivalent real-valued notations: $\bar{\mathbf{u}} \triangleq [\mathrm{Re}(\mathbf{u})^\mathrm{T}, \mathrm{Im}(\mathbf{u})^\mathrm{T}]^\mathrm{T}$, $\bar{\mathbf{v}} \triangleq [\mathrm{Re}(\mathbf{v})^\mathrm{T}, \mathrm{Im}(\mathbf{v})^\mathrm{T}]^\mathrm{T}$, $\bar{\boldsymbol{\Delta}} \triangleq [\mathrm{Re}(\boldsymbol{\Delta})^\mathrm{T}, \mathrm{Im}(\boldsymbol{\Delta})^\mathrm{T}]^\mathrm{T}$, and for all $i = 1, 2, ..., N_\mathrm{u}$, we denote $\mathbf{s}_i \triangleq [\mathrm{Re}(s_i), \mathrm{Im}(s_i)]^\mathrm{T}$, $\mathbf{H}_i \triangleq \Omega(\mathbf{h}_i)$, and $\bar{\mathbf{G}} \triangleq [\mathbf{G}_1^\mathrm{T}, ..., \mathbf{G}_{N_\mathrm{t}}^\mathrm{T}]^\mathrm{T}$ with $\mathbf{G}_j \triangleq \Omega(\mathbf{g}_j)$ and $\mathbf{g}_j$

denoting the jth row of $\mathbf{G}$ for $j = 1, 2, ..., N_\mathrm{t}$, where

$$\Omega(\mathbf{y}) \triangleq \begin{bmatrix} \mathrm{Re}(\mathbf{y}) & -\mathrm{Im}(\mathbf{y}) \\ \mathrm{Im}(\mathbf{y}) & \mathrm{Re}(\mathbf{y}) \end{bmatrix},$$

for any complex input vector $\mathbf{y}$. Using these real-valued notations, it is straightforward to verify that $\bar{\mathbf{v}} = \bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}$ holds true, and thus, the ith real-valued noise-free received signal can be represented as a 2×1 vector given by $\mathbf{H}_i\bar{\mathbf{v}} = \mathbf{H}_i(\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}})$. It is worth mentioning that the additive distortion vector $\bar{\boldsymbol{\Delta}}$, without any restriction on its distribution, is assumed to be norm-bounded, i.e., $\|\bar{\boldsymbol{\Delta}}\| \leq \varepsilon$.

Given the set of target SINRs $\{\gamma_i\}_{i=1}^{N_\mathrm{u}}$ to be achieved for all the users, our design criterion is to minimize the per-symbol total transmit power while satisfying the CI constraint and the given target SINR for each user. In our design formulation, we also need to take into account the linear distortion of the precoded vector before transmission. Therefore, by assuming the DPCIRs and using the design formulation P3, as presented in Section 3.5, we express the corresponding optimization problem as

$$\min_{\mathbf{v}, \mathbf{t} \succeq \mathbf{0}} \quad \bar{\mathbf{v}}^\mathrm{T}\bar{\mathbf{v}} \quad \text{s.t.} \quad \mathbf{H}\bar{\mathbf{v}} = \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{s} + \mathbf{A}^{-1}\mathbf{W}\mathbf{t}, \tag{7.3}$$

where in formulating (7.3), we have used the same definitions as in Section 3.5.

7.3 Worst-Case Design Formulation

We start off by casting a new optimization problem other than (7.3) by introducing the linear equality CI constraints as an ℓ_2-norm penalty into the objective function, i.e.,

$$\min_{\bar{\mathbf{v}}, \mathbf{t} \succeq \mathbf{0}} \quad \|\bar{\mathbf{v}}\|^2 + \beta\,\|\mathbf{H}\bar{\mathbf{v}} - \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{s} - \mathbf{A}^{-1}\mathbf{t}\|^2, \tag{7.4}$$

where β denotes the penalty coefficient. It is worth noting that unlike (7.3), this new formulation does not strictly impose the CI constraints. Instead, the ℓ_2-norm term in the objective function penalizes any feasible solution for which the received symbols will not exactly be located within the intended CI regions. For this reason, we refer to problem (7.4) as the relaxed SLP design. Intuitively speaking, setting larger values for β puts more emphasis on the satisfaction of CI constraints (i.e., more severely penalizes any deviation of the received symbols from the correct CI regions), but may lead to higher transmission powers. This introduces a tradeoff in choosing the penalty parameter β, where its effect on the performance will be investigated via simulation results in Section 7.4. It is also worth noting that problem (7.4) becomes equivalent to (7.3) as $\beta \to \infty$.

Replacing $\bar{\mathbf{v}}$ with $\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}$ in (7.4), we define the worst-case SLP design formulation as

$$\min_{\bar{\mathbf{u}}, \mathbf{t} \succeq \mathbf{0}} \quad \max_{\|\bar{\boldsymbol{\Delta}}\| \leq \varepsilon} \quad \|\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}\|^2 + \beta\,\|\mathbf{H}(\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}) - \boldsymbol{\Phi}(\mathbf{t})\|^2, \tag{7.5}$$

where $\boldsymbol{\Phi}(\mathbf{t}) \triangleq \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{s} + \mathbf{A}^{-1}\mathbf{t}$. The optimization problem (7.5) is non-convex, and thus, may not be amenable to a computationally efficient solution. To tackle this optimization problem, we propose a three-step iterative block coordinate ascent-descent algorithm: in the first step, the inner maximization is solved for given $\bar{\mathbf{u}}$ and $\mathbf{t} \succeq \mathbf{0}$, thereby obtaining a new value for $\bar{\boldsymbol{\Delta}}$ in a semi-closed form in terms of $\bar{\mathbf{u}}$ and $\mathbf{t}$. In the second step, the value of $\mathbf{t}$ is updated by solving a non-negative least squares (NNLS) problem, for fixed $\bar{\boldsymbol{\Delta}}$ and $\mathbf{u}$. In the third step, the value of $\bar{\mathbf{u}}$ is updated by solving a non-constrained QP, thereby obtaining the new value of $\bar{\mathbf{u}}$ in a closed form in terms of $\bar{\boldsymbol{\Delta}}$ and $\mathbf{t}$. In the sequel, we present the details of these three steps.

First step – updating $\bar{\boldsymbol{\Delta}}$: We focus on the inner maximization in (7.5), i.e.,

$$\max_{\|\bar{\boldsymbol{\Delta}}\| \le \varepsilon} \quad \|\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}\|^2 + \beta\, \|\mathbf{H}(\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}) - \boldsymbol{\Phi}(\mathbf{t})\|^2 \,. \tag{7.6}$$

Denoting the maximizer of (7.6) by $\bar{\boldsymbol{\Delta}}^*$, it is straightforward to check, by contradiction, that the norm constraint on $\bar{\boldsymbol{\Delta}}$ is active at the optimum, i.e., $\|\bar{\boldsymbol{\Delta}}^*\| = \varepsilon$. Thus, the maximization problem (7.6) is equivalent to

$$\max_{\|\bar{\boldsymbol{\Delta}}\| = \varepsilon} \quad \|\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}\|^2 + \beta\, \|\mathbf{H}(\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}) - \boldsymbol{\Phi}(\mathbf{t})\|^2. \tag{7.7}$$

In case rank$(\mathbf{H}) > 1$, no closed-form solution is known for (7.7). To tackle this problem, we start from its Lagrangian which is given by

$$\begin{aligned}\mathcal{L}(\bar{\boldsymbol{\Delta}}, \tau) = {} & \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{G}}^{\mathrm{T}}\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}^{\mathrm{T}}\bar{\boldsymbol{\Delta}} + 2\bar{\boldsymbol{\Delta}}^{\mathrm{T}}\bar{\mathbf{G}}\bar{\mathbf{u}} \\ & + \beta\,(\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}(\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}) + \beta\,\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{t})\boldsymbol{\Phi}(\mathbf{t}) \\ & - 2\beta\,\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{t})\mathbf{H}(\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\boldsymbol{\Delta}}) - \tau\left(\bar{\boldsymbol{\Delta}}^{\mathrm{T}}\bar{\boldsymbol{\Delta}} - \varepsilon^2\right),\end{aligned} \tag{7.8}$$

where τ is the Lagrange multiplier associated with the norm constraint $\|\bar{\boldsymbol{\Delta}}\| = \varepsilon$. Note that since the maximization (7.7) is a non-convex problem, the method of Lagrange multipliers yields only necessary conditions for optimality which may not be sufficient. Differentiating $\mathcal{L}(\bar{\boldsymbol{\Delta}}, \tau)$ with respect to $\bar{\boldsymbol{\Delta}}$ and equating it to zero yield

$$\bar{\boldsymbol{\Delta}}^* + \bar{\mathbf{G}}\bar{\mathbf{u}} + \beta\,\mathbf{H}^{\mathrm{T}}\mathbf{H}\bar{\boldsymbol{\Delta}}^* + \beta\,\mathbf{H}^{\mathrm{T}}\mathbf{H}\bar{\mathbf{G}}\bar{\mathbf{u}} - \beta\,\mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}(\mathbf{t}) - \mu^*\bar{\boldsymbol{\Delta}}^* = 0, \tag{7.9}$$

and therefore,

$$\bar{\boldsymbol{\Delta}}^* = -\left(\mathbf{P} - \mu^*\mathbf{I}\right)^{-1}\mathbf{H}^{\mathrm{T}}\Big(\bar{\mathbf{G}}\mathbf{H}\bar{\mathbf{u}} - \boldsymbol{\Phi}(\mathbf{t})\Big), \tag{7.10}$$

where $\mathbf{P} \triangleq \mathbf{H}^{\mathrm{T}}\mathbf{H} + (1/\beta)\mathbf{I}$ and $\mu^* \triangleq \tau^*/\beta$. The maximizer given in (7.10) must satisfy the norm constraint $\|\bar{\boldsymbol{\Delta}}^*\|^2 = \varepsilon^2$, i.e.,

$$\Big(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}(\mathbf{t})\Big)^{\mathrm{T}}\Big(\mathbf{H}^{\mathrm{T}}\mathbf{H} - \mu^*\mathbf{I}\Big)^{-2}\Big(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}(\mathbf{t})\Big) = \varepsilon^2, \tag{7.11}$$

from which one can obtain μ^*. Let us denote

$$f(\mu) \triangleq \Big(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}(\mathbf{t})\Big)^{\mathrm{T}}\Big(\mathbf{H}^{\mathrm{T}}\mathbf{H} - \mu\mathbf{I}\Big)^{-2}\Big(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}(\mathbf{t})\Big) - \varepsilon^2, \tag{7.12}$$

then μ^* is a root of $f(\mu)$. Note that no closed-form solution is known in general for $f(\mu)=0$. Nonetheless, it can be shown that function $f(\mu)$ has a finite number of roots according to the following lemma.

Lemma 14. *Let R denote the number of roots of $f(\mu)$, then R is always an even number bounded as*

$$2 \leq R \leq 2\,\mathrm{rank}(\mathbf{H}).$$

Proof. See Appendix D.1. □

Among all the roots of $f(\mu)$, the one that maximizes the objective function of (7.7) corresponds to the worst-case $\bar{\mathbf{\Delta}}$, for given $\bar{\mathbf{u}}$ and $\mathbf{t}$. The next theorem specifies the interval within which there exists a unique μ^* yielding the maximizer of (7.7).

Theorem 15. *The value of μ^* is equal the largest positive root of $f(\mu)$ and is bounded as*

$$\bar{\lambda}_{\max} < \mu^* \leq \frac{1}{\varepsilon}\left\|\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\mathbf{\Phi}(\mathbf{t})\right\| + \bar{\lambda}_{\max}, \tag{7.13}$$

with $\bar{\lambda}_{\max} \triangleq \|\mathbf{H}\|^2 + \frac{1}{\beta}$.

Proof. See Appendix D.2. □

The above theorem facilitates the possibility of searching for the intended root of $f(\mu)$ in the interval specified by (7.13) via numerical methods, e.g., a simple bisection search. Using such a numeric solution for μ^* in (7.10) yields the optimal value of $\bar{\mathbf{\Delta}}$, for given $\mathbf{u}$ and $\mathbf{t} \succeq \mathbf{0}$, in a semi-closed form

For relatively small values of ε, one can also use quite an accurate approximation for μ^* with a closed-form expression given below.

Lemma 16. *For small ε, the value of μ^* can be well approximated by*

$$\mu^* \approx 2\left(\frac{\left\|\mathbf{P}\left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\mathbf{\Phi}(\mathbf{t})\right)\right\|}{\varepsilon}\right)^{\frac{2}{3}}. \tag{7.14}$$

Proof. See Appendix D.3. □

The approximation provided by Lemma 16 is very accurate for $\varepsilon \leq 0.1$ based on our numerical observations.

Second step – updating t: For given $\bar{\mathbf{\Delta}}$ and $\mathbf{u}$, the value of $\mathbf{t}$ is updated as the solution to the following optimization problem:

$$\min_{\mathbf{t}\succeq\mathbf{0}} \quad \left\|\mathbf{H}\left(\bar{\mathbf{G}}\bar{\mathbf{u}} + \bar{\mathbf{\Delta}}\right) - \mathbf{\Phi}(\mathbf{t})\right\|^2, \tag{7.15}$$

which is a standard NNLS problem. Note, however, that using the exact solution to (7.15) in order to update $\mathbf{t}$ may result in a slow convergence rate for the iterative method [188]. One can instead update $\mathbf{t}$ by using the accelerated projected gradient descent

(APGD) algorithm [189], which provides the update by taking only one step in the steepest descent direction at the current point.

Third step – updating u: For given $\bar{\boldsymbol{\Delta}}$ and $\mathbf{t} \succeq \mathbf{0}$, the minimization over $\mathbf{u}$ is an unconstrained QP, and hence, is amenable to the following closed-form solution:

$$\bar{\mathbf{u}} = \bar{\mathbf{G}}^{-1}\mathbf{P}^{-1}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}(\mathbf{t}) - \bar{\mathbf{G}}^{-1}\bar{\boldsymbol{\Delta}}. \tag{7.16}$$

The pseudo-code of the explained block coordinate ascent-descent algorithm, including the APGD-based updating step of $\mathbf{t}$, is provided in Algorithm 4.

Algorithm 4 Block coordinate ascent-descent algorithm solving (7.5)

```
input: $\mathbf{A}, \mathbf{H}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \mathbf{s}, \varepsilon, \epsilon$
output: $\bar{\mathbf{u}}$
initialize: $\mathbf{t}^{(0)} = \boldsymbol{\vartheta}^{(0)} \in \mathbb{R}_{+}^{2N_{\mathrm{u}}\times 1}$, $\bar{\mathbf{u}}^{(0)} \in \mathbb{R}^{2N_{\mathrm{t}}\times 1}$, $n = 0$
set: $\psi = \frac{1-\sqrt{\kappa}}{1+\sqrt{\kappa}}$, $\kappa = \frac{\sigma_{\max}}{\sigma_{\min}}$, $\boldsymbol{\Theta} = \mathbf{I} - \sigma_{\min}^2(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}$, where $\sigma_{\max}$ and $\sigma_{\min}$ denote the maximum and the minimum singular value of matrix $\mathbf{A}$, respectively
while $\|\bar{\mathbf{u}}^{(n)} - \bar{\mathbf{u}}^{(n-1)}\| \leq \epsilon$ do
    $n \leftarrow n + 1$
    compute $\mu^{(n)}$ by solving $f(\mu) = 0$
    $\bar{\boldsymbol{\Delta}}^{(n)} \leftarrow -\left(\mathbf{P} - \mu^{(n)}\mathbf{I}\right)^{-1}\left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}}^{(n-1)} - \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{s} - \mathbf{A}^{-1}\mathbf{t}^{(n-1)}\right)$
    $\mathbf{t}^{(n)} \leftarrow \max\left\{\boldsymbol{\Theta}\boldsymbol{\vartheta}^{(n-1)} + \sigma_{\min}^2\mathbf{A}^{-\mathrm{T}}\left(\mathbf{H}\left(\bar{\mathbf{G}}\bar{\mathbf{u}}^{(n-1)} + \bar{\boldsymbol{\Delta}}^{(n)}\right) - \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{s}\right), \mathbf{0}\right\}$
    $\boldsymbol{\vartheta}^{(n)} \leftarrow \mathbf{t}^{(n)} + \psi\left(\mathbf{t}^{(n)} - \mathbf{t}^{(n-1)}\right)$
    $\bar{\mathbf{u}}^{(n)} \leftarrow \bar{\mathbf{G}}^{-1}\mathbf{P}^{-1}\mathbf{H}^{\mathrm{T}}\left(\boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{s} + \mathbf{A}^{-1}\mathbf{t}^{(n)}\right) - \bar{\mathbf{G}}^{-1}\bar{\boldsymbol{\Delta}}^{(n)}$
end while
```

To provide an intuition of the structure of the optimal transmit signal, let $(\bar{\boldsymbol{\Delta}}^{*}, \bar{\mathbf{u}}^{*}, \mathbf{t}^{*})$ denote the solution to (7.5). It then follows from (7.16) that

$$\bar{\mathbf{G}}\bar{\mathbf{u}}^{*} + \bar{\boldsymbol{\Delta}}^{*} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \frac{1}{\beta}\mathbf{I}\right)^{-1}\mathbf{H}^{\mathrm{T}}\left(\boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{s} + \mathbf{A}^{-1}\mathbf{t}^{*}\right), \tag{7.17}$$

i.e., the optimal worst-case robust transmit signal can simply be viewed as applying a (regularized) channel inversion to the constructively-interfered symbols, with the interference components being aligned such that the received symbols are pushed (as deep as possible) into the CI regions. Furthermore, considering the limiting case $\beta \to \infty$, in which $\mathbf{P}^{-1}\mathbf{H}^{\mathrm{T}} = \mathbf{H}^{\dagger}$, implies that for extremely large values of β, the received symbol of each UE is guaranteed to be observed within the correct CI region, even for the worst possible error realization. Note, however, that this limiting case β may cause an unaffordable transmission power.

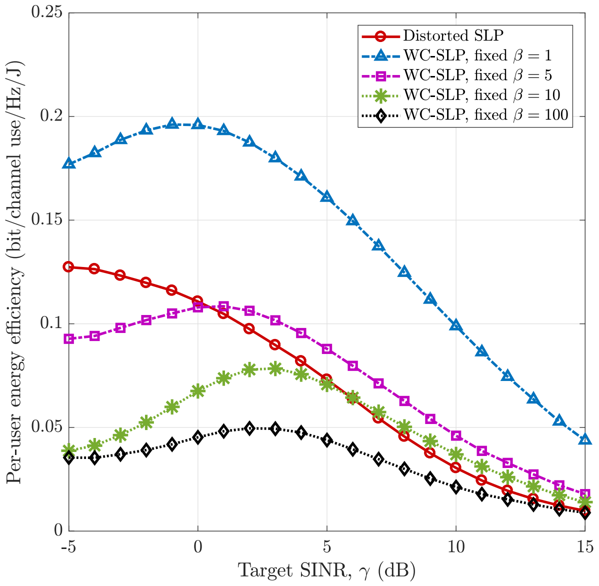


FIGURE 7.2: Energy efficiency comparison of different SLP schemes under linear distortions.

Computational Complexity

We compare the computational complexity of Algorithm 4 with the required complexity for solving the QP (7.3). Here, by complexity we mean to the number of arithmetic operations needed to reach a desired accuracy, i.e., an ϵ-optimal solution. The per-iteration complexity of the Algorithm 4 is dominated by matrix multiplications with dimension $2N_\mathrm{t} \times 2N_\mathrm{t}$ as well as matrix inversions of dimension $N_\mathrm{t} \times N_\mathrm{t}$, where both computations have a limiting order of $\mathcal{O}(N_\mathrm{t}^3)$. On the other hand, based on the worst-case complexity analysis provided in [171], the per-iteration complexity of solving the QP (7.3) via a generic interior-point method is of order $\sqrt{N_\mathrm{u}}\,\mathcal{O}(N_\mathrm{t}^3)$. Taking into account the convergence rates to reach an ϵ-optimal solution, the computation cost of the QP (7.3) is of order $\sqrt{N_\mathrm{u}}\,\mathcal{O}(N_\mathrm{t}^3)\ln(1/\epsilon)$, whereas Algorithm 4 has a lower complexity order $\mathcal{O}(N_\mathrm{t}^3)\,(1/\sqrt{\epsilon})$ and converges at a higher rate.

7.4 Simulation Results

The simulation setup is as follows. We consider a downlink MU-MIMO system with uncoded transmission, QPSK signaling, and $N_\mathrm{t} = N_\mathrm{u} = 8$. Unit noise variances $\sigma_i^2 = 1$ and equal target SINRs $\gamma_i \triangleq \gamma$ are assumed for all $i = 1, 2, ..., N_\mathrm{u}$. Assuming a Rayleigh block fading channel, the channel vectors $\{\mathbf{h}_i\}_{i=1}^{N_\mathrm{u}}$ are independently generated for each coherence block following a standard circularly symmetric complex Gaussian (CSCG) distribution, i.e., $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. All our simulation results are averaged over 500 channel coherence blocks each with 500 symbols. We refer to our proposed worst-case SLP design as WC-SLP.

The additive distortion vector $\bar{\mathbf{\Delta}}$ is randomly generated as an uncorrelated CSCG vector with standard deviation 0.1. The distortion ball radius is set to be $\varepsilon = 0.56$, which corresponds to a confidence level of 0.99, i.e., $\Pr\{\|\bar{\mathbf{\Delta}}\| > \varepsilon\} = 0.01$. We further assume $\mathbf{G} = \mathbf{I}$. In our simulations, we have defined energy efficiency as the ratio of the product of the average users' bit error rate (BER) and mutual information divided by the total consumed power, i.e., $\bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}}$. The mutual information $I(s_i; r_i)$ for the ith user can be obtained as

$$I(s_i; r_i) = \mathbb{E}_{s_i,r_i,\mathbf{H}}\left\{\log_2 \frac{P_{r_i|s_i,\mathbf{H}}(r_i|s_i,\mathbf{H})}{P_{r_i|\mathbf{H}}(r_i|\mathbf{H})}\right\}. \tag{7.18}$$

The conditional probability mass functions in (7.18) are not amenable to closed-form expressions. To tackle this difficulty, inspired by [190], we resort to empirical probability distributions obtained by generating sufficiently many channel and symbol realizations and then computing an approximation (in fact, a lower bound) for the mutual information in (7.18).

The energy efficiency performance of the WC-SLP scheme is plotted in Fig. 7.2 as a function of the users' target SINR, for different values of β. To have a benchmark for comparison, we also present the results for the SLP problem (7.3) under linear distortions, referred to as "distorted SLP". Among all the values of β shown in Fig. 7.2, choosing $\beta = 1$ results in higher energy efficiency, even compared to the distorted SLP scheme. This is a consequence of relaxing the CI constraints in the SLP design, leading to a lower transmit power in exchange for a slightly higher BER. Increasing β, on the other hand, reduces the energy efficiency of the proposed WC-SLP scheme. This can be justified by considering the limiting case $\beta \to \infty$, in which the design formulation (7.4) aims to strictly impose the CI constraints, regardless of the required transmit power. In general, a proper choice of β is application-dependent and relies on the corresponding system/user requirements. For instance, in wireless systems with strict target BERs, a larger β is preferred. On the other hand, in scenarios where transmit power is strictly limited, one may choose smaller values for β. Moreover, the value of β can be adjusted in a more sophisticated way, e.g., letting β vary as a function of the target SINR γ, which is an interesting topic for future work.

7.5 Conclusions

In this chapter, we proposed a worst-case design formulation for the QoS-constrained SLP problem minimizing the total transmit power in a scenario where the precoder's output undergoes linear distortion with bounded additive noise. First, a new problem formulation was proposed, which led us to cast the worst-case design of the distorted SLP as a min-max problem by introducing relaxed CI constraints. We then solved this problem using an iterative block coordinate ascent-descent algorithm to obtain the robust precoded signal. This algorithm iterates between finding the optimal precoded signal and the worst-case additive distortion vector. Finding the precoded signal involves solving an NNLS problem, while obtaining the worst-case distortion vector leads to a semi-closed form solution with only one scalar parameter which has to be calculated numerically. Our simulation results showed that the proposed worst-case approach can improve the SLP scheme's performance under linear distortions in terms of energy efficiency.

Chapter 8

Quantized Symbol-Level Precoding for Massive MU-MIMO Systems

In this chapter, we propose a finite-alphabet symbol-level precoding technique for massive MU-MIMO downlink systems that are equipped with finite-resolution digital-to-analog converters (DACs) of any precision. Using the idea of CI, we adopt a max-min fair design criterion which aims to maximize the minimum instantaneous received signal-to-interference-plus-noise ratio (SINR) among the users while ensuring a CI constraint for each user under the restriction that the output of the precoder is a vector with finite-alphabet discrete elements. Due to this latter restriction, the design problem is an NP-hard quadratic program with discrete variables, and hence, is difficult to solve. In this chapter, we tackle this difficulty by reformulating the problem in several steps into an equivalent continuous-domain biconvex form, including equivalent representations for discrete and binary constraints. Our final biconvex reformulation is obtained via an exact penalty approach and can efficiently be solved using a standard cyclic block coordinate descent algorithm. We evaluate the performance of the proposed finite-alphabet precoding design for DACs with different resolutions, where it is shown that employing low-resolution DACs can lead to higher power efficiencies. In particular, we focus on a setup with one-bit DACs and show through simulation results that compared to the existing schemes, the proposed design can achieve SNR gains of up to 2 dB. We further provide analytic and numerical analyses of complexity and show that our proposed algorithm is computationally-efficient as it typically needs only a few tens of iterations to converge.

8.1 Introduction

Massive MIMO is an enabling, or perhaps even indispensable, technology to deliver highly power-efficient and reliable services in future generation wireless communication networks [191, 192]. In a typical massive MU-MIMO system, the base station employs a large-scale antenna array (e.g., with hundreds of antennas) to serve a much smaller number of user equipments (UEs) via spatial multiplexing. This multitude of transmit antennas offers a large number of spatial degrees of freedom to each UE, leading to high beamforming gains and high interference management capabilities [191, 193].

When perfect transmit-side channel state information (CSIT) is non-causally available, it is well known that dirty paper coding (DPC) can achieve the sum-rate capacity of the MU-MIMO broadcast channel at an impractically high computational complexity. Alternatively, simpler practical linear precoding schemes such as matched filter (MF) or maximum ratio transmission (MRT) [2, 5], and (regularized) ZF [3, 4], have been shown to be asymptotically optimal in the large system limit [194]. Unfortunately, the benefits of such easy-to-implement precoding techniques come with a prohibitively high hardware complexity and cost as well as an extensively increased power consumption under a massive MIMO configuration. This is primarily due to the need for an ideal radio frequency (RF) chain, including highly linear power amplifiers and high-resolution digital-to-analog converter (DAC), dedicated to each antenna element. Therefore, a more limited use of RF and mixed-signal components is of practical importance to enable cost-effective implementations of large-scale MIMO systems with reasonable hardware complexity and low power consumption. Accordingly, precoding schemes need to be properly designed by taking into consideration the hardware-induced constraints.

The hardware-constrained design of multiuser precoding in a massive MU-MIMO system has been addressed in the literature via a variety of approaches, which can be broadly categorized in two groups, namely, hybrid precoding and finite-alphabet precoding. The first group includes hybrid analog-digital architectures, where a small-sized digital precoder is followed by a high-dimensional analog precoder. Such an architecture enables the possibility of using fewer RF chains, which scale with the number of multiplexed data streams rather than with the number of transmit antennas; see, e.g., [195–198]. Despite being capable of achieving the performance of fully-digital ZF precoder [199], hybrid techniques do not scale well with the number of subcarriers in wide-band systems [200]. In addition, there is still need for employing high-resolution DACs. Bearing in mind that the power consumption of a finite-resolution DAC grows exponentially with the number of resolution bits and linearly with the bandwidth [201, 202], the hybrid schemes still suffer from high power consumption as well as high complexity, losses and non-linearity of analog components. The need for high-resolution DACs may further limit the implementation of hybrid architectures. The reason is that increasing the number of RF chains beyond a certain limit (depending on the architecture) results in a lower energy-efficient implementation, as compared to its fully-digital counterpart [203].

On the other hand, the use of low-resolution DACs for each antenna element substantially reduces the amount of power consumption, simplifies the hardware design, and reduces the associated cost. This has been the motive behind introducing another family of massive MU-MIMO precoding approaches, namely, finite-precision quantized precoders. Within this line of work, there have been some efforts towards revisiting conventional linear precoding strategies, leading to what is commonly referred to as linear-quantized precoding, where the effect of quantization distortion has been taken into account for low-to-moderate resolution (up to 5 bits) DACs [200, 204–206]. These linear precoders, however, mostly suffer from an unfavorably high error floor in the moderate-to-high SNR regime [200], and perform reasonably only in systems with extremely large transmit array sizes, e.g., in the order of hundreds or more antennas. Recently, the case with one-bit DACs has become an attractive research direction due to its simplicity and the dramatic reduction it can provide in circuit power consumption and hardware cost; see, e.g., [87, 207–210]. A one-bit precoded signal further exhibits the constant modulus property, eliminating the need for highly linear power amplifiers. The enabling fact behind one-bit precoding approaches is that the severe distortion caused by one-bit DACs can be mitigated by means of proper signal processing at the transmitter, though its undesired impact is insignificant in the low-to-moderate SNR regime where massive MIMO systems will likely operate [200]. Most of the work in this direction consider non-linear precoding design based on a symbol-by-symbol approach. The superiority of these nonlinear (finite-alphabet) precoding approaches over linear-quantized precoding is demonstrated in [207].

The idea of designing the precoder in a per-symbol manner, while exploiting the inter-user interference at the UE side in a useful way, has been studied in [24], and then elaborated in [20] and [21], where the concept of CI is introduced. Referred to as symbol-level precoding (SLP), this type of precoding is based on the fact that a noise-free received signal can be decoded correctly, not necessarily when it is close enough to the intended symbol, rather, as long as it lies within the correct decision region even far away from the target symbol. This has been the underlying motivation in defining a variety of CI regions; see, e.g., [20, 21, 23, 61]. In designing a quantized precoder for massive MU-MIMO downlink, one can utilize the CI concept to achieve lower BER values for the UEs. This approach has been used in [84] and [85, 86] in order to design SLP schemes with one-bit DACs for PSK and QAM signaling. On the other hand, symbol-level precoding with low-resolution DACs is addressed in [88], where the design objective is defined based on a mean square error (MSE) criterion rather than the CI constraints.

In this paper, we propose a novel finite-alphabet CI precoding method for massive MU-MIMO downlink, where the precoded transmit signal is constrained to be chosen from a predefined codebook dictated by finite-resolution DACs. Unlike [84] and [85], our work is not restricted to PSK or QAM signaling, but considers a generic modulation scheme. Furthermore, in our design, we aim to exploit CI at the UEs' receiver side according to the distance-preserving definition of CI regions [23]. The adopted precoding design approach aims to maximize the minimum (instantaneous) SNR among all the UEs, while ensuring the UEs' symbols are received within the correct CI region. Due to the finite-alphabet domain of the design variables, the original formulation is an NP-hard problem, and thus, is difficult to solve. We deal with this difficulty through reformulating the problem in several steps. First, we reduce the problem to a binary quadratic programming through an alternative equivalent binary representation of the discrete design variable. Next, we provide an equivalent continuous-domain biconvex implication for the binary constraints. We cast our final design formulation by applying the exact penalty method, which can efficiently be solved via a standard cyclic block coordinate descent (BCD) algorithm under certain conditions which, as we prove, will be met in our design. This is different from [84] and [85], where the binary constraints are dealt with via simple convex relaxations. Note also that in [87], a similar technique is used to treat the binary constraints; however, our design objective and constraints differ from those in [87]. More precisely, the precoding design in [87] attempts to minimize the maximum (among the UEs) distance between a received signal and its corresponding target symbol up to a scaling factor, while our design aims to maximize the distance of the UE's received signals from the boundaries of the corresponding decision region via exploiting CI. In fact, the scaling factor in [87] can be viewed as a special case of CI regions with strict phase constraints. The proposed finite-alphabet symbol-level precoding approach supports DACs of any resolution. Evaluating and comparing the performance for DACs with different resolutions, we show that while using low-resolution DACs may cause a degraded bit error rate performance, it leads to a higher power efficiency. Moreover, we will show that increasing the number of resolution bits by one results in a gain of at least 0.5 dB in different scenarios. With a particular focus on the case with one-bit DACs, our simulation results further indicate an improved uncoded bit error rate performance for the proposed method, compared to the existing one-bit precoding schemes. To be more precise, depending on the system setup, SNR gains of up to 2 dB can be achieved. To assess the practicability of the proposed design, we provide an analytical analysis of computational complexity. Remarkably, for moderately-sized systems, the BCD algorithm, used to solve the proposed design problem, usually converges (with a reasonable accuracy) in a few tens of iterations, making the proposed method attractive for practical use.

The rest of this chapter is organized as follows. In Section 8.2, we describe the considered system model, including the signal model and the quantization model. In Section 8.3, we formulate the CI-based finite-alphabet symbol-level precoding problem as a discrete quadratic programming. In Section 8.4, we propose our solution to the design problem of interest, followed by an analysis of computational complexity. The spacial case of using one-bit quantized precoding is addressed in Section 8.5. We present simulation and numerical results in Section 8.6. Section 8.7 concludes the paper.

8.2 System Model

In this section, we describe the signal and quantization model considered in this chapter.

8.2.1 Signal Model

We consider a single-cell single-carrier downlink MU-MIMO wireless system where a BS, equipped with an array of N_t antennas, communicates with $N_\mathrm{u} \ll N_\mathrm{t}$ single-antenna users through multiplexing N_u independent data streams within the same time-frequency resource block. Note that the latter assumption rationalizes the use of low-resolution DACs at the BS; however, it is not strictly necessary for the subsequent derivations in this chapter. As illustrated in Fig. 8.1, we assume that the users' data symbols are spatially multiplexed at the BS via a (non-linear) symbol-level multiuser precoder so that the $N_\mathrm{t} \times 1$ complex-valued precoded signal $\mathbf{u} \triangleq [u_1, u_2, ..., u_{N_\mathrm{t}}]^\mathrm{T}$ is specifically designed every symbol period. It is further assumed that each BS's antenna is driven by one dedicated RF chain preceded by a pair of finite-resolution DACs, operating independently on the in-phase and quadrature components of a complex element of the precoded signal $\mathbf{u}$. Modeling a finite-resolution DAC as a scalar quantizer, we describe the complex-valued quantization operation as

$$u_{\mathrm{q},j} = \mathcal{Q}\Big(\mathrm{Re}\{u_j\}\Big) + \mathrm{j}\,\mathcal{Q}\Big(\mathrm{Im}\{u_j\}\Big), \quad j = 1, 2, ..., N_\mathrm{t}, \tag{8.1}$$

where $\mathbf{u}_\mathrm{q} \triangleq [u_{\mathrm{q},1}, u_{\mathrm{q},2}, ..., u_{\mathrm{q},N_\mathrm{t}}]^\mathrm{T}$ denotes the precoded signal after quantization, $\mathcal{Q}(\cdot) : \mathbb{R} \mapsto \mathbb{L}$ stands for the scalar quantization operation with $\mathbb{L}$ denoting the set of quantization levels, and $\mathrm{j} \triangleq \sqrt{-1}$. Assuming element-wise vector quantization as represented in (8.1), we require a total number of $2N_\mathrm{t}$ DACs, thereby operating independently on the real and imaginary parts of u_j for all $j \in \{1, 2, ..., N_\mathrm{t}\}$. The quantized baseband signal $\mathbf{u}_\mathrm{q}$ is then passed through N_t RF chains, which up-convert this signal to the carrier frequency. The up-converted signal is transmitted over the BS's antennas and undergoes uncorrelated quasi-static flat fading before arriving the users. Under the above described assumptions, the signal r_i received at the ith user can be modeled as

$$r_i = \sqrt{p}\,\mathbf{h}_i \mathbf{u}_\mathrm{q} + z_i, \quad i = 1, 2, ..., N_\mathrm{u}, \tag{8.2}$$

where $\mathbf{h}_i$ denotes the complex-valued $1 \times N_\mathrm{t}$ vector of the ith user's channel coefficients and z_i represents the additive noise at the ith user's receiver front-end and is modeled as a zero-mean complex Gaussian random variable with variance $\sigma_i^2/2$ per real dimension, i.e., $z_i \sim \mathcal{CN}(0, \sigma_i^2)$. Furthermore, in our design, we constrain the quantized precoded signal so as to satisfy $\|\mathbf{u}_\mathrm{q}\|^2 \leq 1$. Hence, $\sqrt{p}$ in (8.2) denotes a fixed gain ensuring a total transmission power of smaller than p. At the receiver side, it is assumed that each user employs an infinite-precision analog-to-digital converter (ADC) and is capable of detecting its target symbol via optimal single-user maximum-likelihood criterion.

For the sake of convenience, we define the following equivalent real-valued vectors:

$$\bar{\mathbf{u}} \triangleq \begin{bmatrix} \mathrm{Re}(\mathbf{u}) \\ \mathrm{Im}(\mathbf{u}) \end{bmatrix} \triangleq [\bar{u}_1, \bar{u}_2, ..., \bar{u}_{2N_\mathrm{t}}]^\mathrm{T},$$

$$\bar{\mathbf{u}}_\mathrm{q} \triangleq \begin{bmatrix} \mathrm{Re}(\mathbf{u}_\mathrm{q}) \\ \mathrm{Im}(\mathbf{u}_\mathrm{q}) \end{bmatrix} \triangleq [\bar{u}_{\mathrm{q},1}, \bar{u}_{\mathrm{q},2}, ..., \bar{u}_{\mathrm{q},2N_\mathrm{t}}]^\mathrm{T},$$

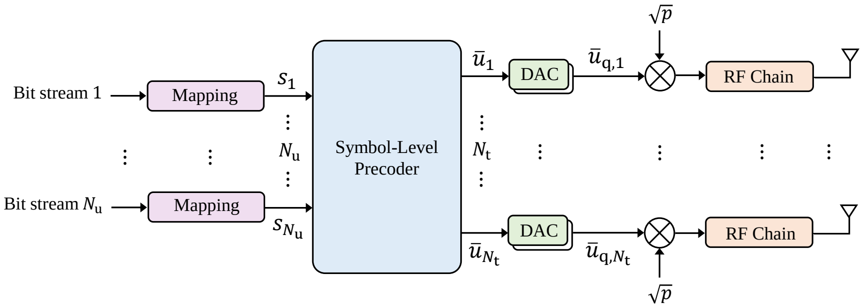


FIGURE 8.1: The considered transmission scheme with symbol-level precoding where each I/Q channel undergoes quantization via finite-resolution DACs.

Moreover, for all $i = 1, 2, ..., N_\mathrm{u}$, we denote $\mathbf{s}_i \triangleq [\mathrm{Re}(s_i), \mathrm{Im}(s_i)]^\mathrm{T}$ and

$$\mathbf{H}_i \triangleq \begin{bmatrix} \mathrm{Re}(\mathbf{h}_i) & -\mathrm{Im}(\mathbf{h}_i) \\ \mathrm{Im}(\mathbf{h}_i) & \mathrm{Re}(\mathbf{h}_i) \end{bmatrix}.$$

Using the above notations, we express the relation between the real-valued elements of the quantized and the unquantized precoded signal as

$$u_{\mathrm{q},j} \triangleq \mathcal{Q}(u_j), \quad j = 1, 2, ..., 2N_\mathrm{t}. \tag{8.3}$$

It is also worth noting that under the new real-valued notations, the power constraint to be met in our design is $\|\bar{\mathbf{u}}_\mathrm{q}\|^2 \leq 1$.

8.2.2 Quantization Model

A finite-resolution DAC can be modeled as a scalar quantizer mapping a continuous-valued input signal onto a finite discrete-valued set of possible outputs, namely, quantization levels (or reconstruction levels). Let b denote the number of resolution bits representing the quantized signal, then the total number of reconstruction levels is equal to $B \triangleq |\mathbb{L}| = 2^b$, where $\mathbb{L} \triangleq \{l_0, l_1, ..., l_{B-1}\}$ is the set of quantization levels. Let us further denote by $\mathbb{T} \triangleq \{\vartheta_0, \vartheta_1, ..., \vartheta_{B-1}, \vartheta_B\}$ the set of quantization thresholds such that $-\infty = \vartheta_0 < \vartheta_B = +\infty$. In this work, for simplicity, we consider symmetric uniform scalar quantizers, where the quantization levels $\{l_0, l_1, ..., l_{B-1}\}$ are equally and symmetrically spaced around zero, i.e.,

$$l_k = \Delta\left(k - \frac{B-1}{2}\right), \quad k = 0, 1, ..., B-1, \tag{8.4}$$

$$\vartheta_k = \Delta\left(k - \frac{B}{2}\right), \quad k = 1, 2, ..., B-1, \tag{8.5}$$

with Δ denoting the quantization step, i.e., the spacing between two consecutive reconstruction levels. Depending on the precision, the symmetric scalar quantizer uses a subset of the two sequences of quantization levels and thresholds produced by (8.4) and (8.5), i.e.,

$$\mathbb{L} \subseteq \left\{ ..., -\frac{3}{2}\Delta, -\frac{1}{2}\Delta, +\frac{1}{2}\Delta, +\frac{3}{2}\Delta, ... \right\}, \tag{8.6}$$

$$\mathbb{T} \subseteq \{..., -2\Delta, -\Delta, 0, +\Delta, +2\Delta, ...\}, \tag{8.7}$$

respectively. Given the power constraint on $\bar{\mathbf{u}}_\mathrm{q}$, we need to properly choose the quantization step Δ. Recall that the entries of $\bar{\mathbf{u}}_\mathrm{q}$ are taken from the set $\mathbb{L}$ of quantization levels. As a consequence, the following inequality always holds true:

$$\|\bar{\mathbf{u}}_\mathrm{q}\|^2 = \sum_{j=1}^{2N_\mathrm{t}} u_{\mathrm{q},j}^2 \leq 2N_\mathrm{t}\, l_{B-1}^2. \tag{8.8}$$

To guarantee the power constraint $\|\bar{\mathbf{u}}_\mathrm{q}\|^2 \leq 1$, it suffices to set $2N_\mathrm{t} l_{B-1}^2 = 1$. Hence, by replacing $l_{B-1} = (\Delta/2)(B-1)$ from (8.4), we obtain

$$\Delta = \frac{2}{(B-1)\sqrt{2N_\mathrm{t}}}. \tag{8.9}$$

In our design, we consider the value of Δ as obtained in (8.9) to implicitly enforce the desired power constraint.

8.3 Problem Formulation

The aim of this section is to formulate the SLP design problem with finite-resolution DACs. To avoid quantization distortion, the precoded signal $\bar{\mathbf{u}}$ is taken from the set $\mathbb{L}^{2N_\mathrm{t}}$ dictated by the set of finite-resolution DACs so that

$$\bar{u}_j = \bar{u}_{\mathrm{q},j} = \mathcal{Q}(\mathrm{q}_j), \quad j = 1, 2, ..., 2N_\mathrm{t}. \tag{8.10}$$

Accordingly, we aim to design a CI-based precoder which maximizes the minimum instantaneous (per-symbol) quality-of-service (QoS) level among the users, while satisfying the CI constraint for each user. As for the QoS measure, we consider the users' received SINRs. Assuming DPCIRs, we use the convex representation of CI constraint C3 as introduced in Section 3.4. Thereby, we can obtain the optimal finite-alphabet precoded signal as the solution to the following optimization problem:

$$\begin{aligned} \max_{\bar{\mathbf{u}}, \mathbf{t} \succeq \mathbf{0}} \quad & \min_{i} \; \|\mathbf{H}_i \bar{\mathbf{u}}\|^2 / \sigma_i^2 \\ \text{s.t.} \quad & \sqrt{p}\, \mathbf{H}\bar{\mathbf{u}} = \mathbf{\Sigma}\mathbf{s} + \mathbf{A}^{-1}\mathbf{W}\mathbf{t}, \\ & u_j \in \mathbb{L}, \quad j = 1, 2, ..., 2N_\mathrm{t}. \end{aligned} \tag{8.11}$$

Note that in formulating (8.11), we implicitly assumed that the user's channel coefficients $\mathbf{H}_i$ are perfectly and instantaneously known to the BS for all $i \in \{1, 2, ..., N_\mathrm{u}\}$. An illustration of the DPCIRs and their characterizing parameters and variables is shown

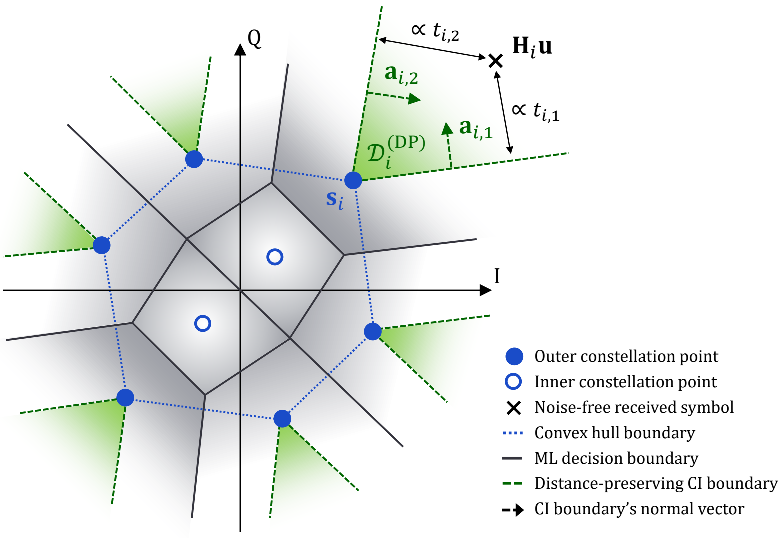

$\propto t_{i,2}$
$\mathbf{H}_i\mathbf{u}$
$\mathbf{a}_{i,2}$
$\propto t_{i,1}$
$\mathbf{a}_{i,1}$
$\mathcal{D}_i^{(\mathrm{DP})}$
$\mathbf{s}_i$
Q
I
Outer constellation point
Inner constellation point
Noise-free received symbol
Convex hull boundary
ML decision boundary
Distance-preserving CI boundary
CI boundary's normal vector

FIGURE 8.2: The DPCIRs are depicted in green color for the optimized 8-ary constellation.

in Fig. 8.2 for the optimized 8-ary constellation [124]. We have shown in Section 3.5 that maximizing the minimum SNR across the users is equivalent to maximizing $\min(\mathbf{t})$ subject to the given power constraint, where $\min(\cdot)$ denotes element-wise minimum. Note that the total power constraint has been taken into account in (8.11) and in the subsequent reformulations through the definition of the set of quantization levels $\mathbb{L}$. Introducing a slack variable γ and using the above definitions, we rewrite problem (8.11) as

$$\begin{aligned} \max_{\bar{\mathbf{u}},\mathbf{t},\gamma} \quad & \gamma \\ \text{s.t.} \quad & \sqrt{p}\,\mathbf{H}\bar{\mathbf{u}} = \boldsymbol{\Sigma}\mathbf{s} + \mathbf{A}^{-1}\mathbf{W}\mathbf{t}, \\ & u_j \in \mathbb{L}, \quad j = 1, 2, ..., 2N_\mathrm{t}, \\ & \mathbf{t} \succeq \gamma\mathbf{1}, \quad \gamma \geq 0. \end{aligned} \tag{8.12}$$

Two difficulties arise with problem (8.12) as described below:

i. The optimization problem (8.12) may have an empty feasible region, since $\bar{\mathbf{u}}$ has to be chosen from the finite set $\mathbb{L}^{2N_\mathrm{t}}$. In fact, there could be situations where one (or more) CI constraint(s) cannot be satisfied for any $\bar{\mathbf{u}} \in \mathbb{L}^{2N_\mathrm{t}}$.

ii. Due to the finite-alphabet variable $\bar{\mathbf{u}}$, problem (8.12) belongs to the class of combinatorial optimization, which is known to be difficult (in some cases, NP-hard) to solve for global optimality. To be more specific, finding the exact solution to (8.12), in the worst case, requires solving a linear programming (LP) for every single vector $\bar{\mathbf{u}} \in \mathbb{L}^{2N_\mathrm{t}}$ and then picking the best solution for $\mathbf{u}$ which results in the largest value of γ. The finite set $\mathbb{L}^{2N_\mathrm{t}}$ has a cardinality of B^{2N_t}. Keeping in mind that N_t refers to the size of a large-scale antenna array, such an approach requires solving an exponentially-growing number of LPs, which might be quite impractical.

To address the above challenges, we will take a few more steps to modify the original problem, as explained in the next section.

8.4 Quantized Symbol-Level Precoding Design

We start off by addressing the first challenge highlighted in the previous section. To avoid infeasibility, we consider a new (not necessarily equivalent) design formulation by adding soft CI constraints as a penalty term to the objective function. Doing so leads to the following problem:

$$\begin{aligned} \max_{\bar{\mathbf{u}},\mathbf{t},\gamma} \quad & \gamma - \left\| \sqrt{p}\,\mathbf{H}\bar{\mathbf{u}} - \boldsymbol{\Sigma}\mathbf{s} - \mathbf{A}^{-1}\mathbf{W}\mathbf{t} \right\|^2 \\ \text{s.t.} \quad & u_j \in \mathbb{L}, \quad j = 1, 2, ..., 2N_\mathrm{t}, \\ & \mathbf{t} \succeq \gamma\mathbf{1}, \quad \gamma \geq 0. \end{aligned} \tag{8.13}$$

We will show via simulation results in Section 8.6 that the loss due to this new formulation compared to the original problem in (8.12) is very negligible, especially in the large system limit. It is straightforward to verify that problem (8.12) is equivalent to

$$\begin{aligned} \max_{\bar{\mathbf{u}},\mathbf{t},\gamma} \quad & \gamma - \left\| \sqrt{p}\,\mathbf{H}\bar{\mathbf{u}} - \boldsymbol{\Sigma}\mathbf{s} - \mathbf{A}^{-1}\mathbf{W}(\mathbf{t} + \gamma\mathbf{1}) \right\|^2 \\ \text{s.t.} \quad & u_j \in \mathbb{L}, \quad j = 1, 2, ..., 2N_\mathrm{t}, \\ & \mathbf{t} \succeq \mathbf{0}, \quad \gamma \geq 0. \end{aligned} \tag{8.14}$$

Given $\bar{\mathbf{u}}$ and $\mathbf{t}$, the maximization problem in (8.14) can be expressed as a function of γ as

$$\max_{\gamma} \quad \gamma - \left\| \sqrt{p}\,\mathbf{H}\bar{\mathbf{u}} - \boldsymbol{\Sigma}\mathbf{s} - \mathbf{A}^{-1}\mathbf{W}(\mathbf{t} + \gamma\mathbf{1}) \right\|^2 \quad \text{s.t. } \gamma \geq 0. \tag{8.15}$$

Differentiating the objective function of (8.15) with respect to γ and equating it to zero, we can obtain a provably positive closed-form solution for γ as

$$\gamma^* = \frac{1}{2\eta} + \frac{\left(\sqrt{p}\,\mathbf{H}\bar{\mathbf{u}} - \boldsymbol{\Sigma}\mathbf{s} - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right)^\mathrm{T} \mathbf{A}^{-1}\mathbf{W}\mathbf{1}}{\eta}, \tag{8.16}$$

where $\eta \triangleq \mathbf{1}^\mathrm{T}\mathbf{W}\mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{W}\mathbf{1}$. As a result, plugging the closed-form expression for γ^* into (8.14), we can eliminate the variable γ from our design formulation. After some straightforward algebraic steps, the optimization problem (8.14) can be recast as a discrete LCQP as

$$\begin{aligned} \max_{\bar{\mathbf{u}},\mathbf{t}\succeq\mathbf{0}} \quad & \mathbf{q}^\mathrm{T}\left(\sqrt{p}\mathbf{H}\bar{\mathbf{u}} - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right) - \left\| \mathbf{Q}\left(\sqrt{p}\mathbf{H}\bar{\mathbf{u}} - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right) - \mathbf{g} \right\|^2 \\ \text{s.t.} \quad & u_j \in \mathbb{L}, \quad j = 1, 2, ..., 2N_\mathrm{t}, \end{aligned} \tag{8.17}$$

where $\mathbf{q} \triangleq (1/\eta)\mathbf{A}^{-1}\mathbf{W}\mathbf{1}$, $\mathbf{Q} \triangleq \mathbf{I} - \eta\,\mathbf{q}\mathbf{q}^{\mathrm{T}}$ and $\mathbf{g} \triangleq \mathbf{Q}\boldsymbol{\Sigma}\mathbf{s} + (1/2)\mathbf{q}$ are all non-variables. To tackle the second difficulty, arose from the discrete domain of the optimization variable $\bar{\mathbf{u}}$, we first reduce (8.17) to a binary optimization problem through an equivalent binary representation of $\bar{\mathbf{u}}$ and then propose two approaches to tackle the resulting problem, as explained in the following.

The finite-alphabet constraint on $\bar{\mathbf{u}}$ implies that, for any $j \in \{1, 2, ..., 2N_{\mathrm{t}}\}$, each element u_j must take its value from the set of quantization levels $\mathbb{L}$, as specified in (8.6). Given b, the set of levels $\mathbb{L}$ can explicitly be represented as

$$\mathbb{L} = \left\{ \frac{\Delta}{2} \sum_{n=1}^{b} 2^{n-1} v_n \mid v_n \in \{-1, +1\} \right\}, \tag{8.18}$$

where the summation $\sum_{n=1}^{b} 2^{n-1} v_n$ generates the sequence $\{...,-3,-1,+1,+3,...\}$ for different realizations of $\{v_1, ..., v_b\}$. In fact, each level in $\mathbb{L}$ corresponds to a binary realization of $\{v_1, ..., v_b\}$. For example, with $b = 3$, the quantization level $-\Delta/2$ can be represented by $\{v_1, v_2, v_3\} = \{+1, +1, -1\}$. As a result, each element u_j can be expressed as

$$u_j = \frac{\Delta}{2} \sum_{n=1}^{b} 2^{n-1} v_{j,n}, \quad j = 1, 2, ..., 2N_{\mathrm{t}}, \tag{8.19}$$

where $v_{j,n} \in \{-1, +1\}$ for all $n = 1, ..., b$ and $j = 1, 2, ..., 2N_{\mathrm{t}}$ are binary decision variables. Let $\mathbf{b} \triangleq (\Delta/2)[1, ..., 2^{b-1}]^{\mathrm{T}}$ and $\mathbf{v}_j = [v_{j,1}, ..., v_{j,b}]^{\mathrm{T}}$ denote, respectively, the vector of binary decision variables and the vector of constants. Hence, one can rewrite (8.19) as $u_j = \mathbf{v}_j^{\mathrm{T}}\mathbf{b}$. Collecting all u_j together, for $j = 1, 2, ..., 2N_{\mathrm{t}}$, vector $\bar{\mathbf{u}}$ can be represented as

$$\bar{\mathbf{u}} = \mathbf{V}\mathbf{b}, \tag{8.20}$$

where $\mathbf{V} \triangleq [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{2N_{\mathrm{t}}}]^{\mathrm{T}}$. Using the binary representation of $\bar{\mathbf{u}}$ in (8.20), the optimization problem (8.17) can be equivalently written as

$$\begin{aligned} \max_{\mathbf{V},\mathbf{t}\succeq\mathbf{0}} \quad & \mathbf{q}^{\mathrm{T}}\left(\sqrt{p}\mathbf{H}\mathbf{V}\mathbf{b} - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right) - \left\|\mathbf{Q}\left(\sqrt{p}\mathbf{H}\mathbf{V}\mathbf{b} - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right) - \mathbf{g}\right\|^2 \\ \text{s.t.} \quad & v_{j,n} \in \{-1, +1\},\ n = 1, ..., b,\ j = 1, 2, ..., 2N_{\mathrm{t}}. \end{aligned} \tag{8.21}$$

Using the fact that $\mathbf{H}\mathbf{V}\mathbf{b} = (\mathbf{b}^{\mathrm{T}} \otimes \mathbf{H})\mathrm{vec}(\mathbf{V})$, and denoting $\mathbf{x} \triangleq \mathrm{vec}(\mathbf{V})$ and $\mathbf{H}_{\mathrm{b}} \triangleq \mathbf{b}^{\mathrm{T}} \otimes \mathbf{H}$, where $\mathbf{H}_{\mathrm{b}} \in \mathbb{R}^{2N_{\mathrm{u}} \times 2bN_{\mathrm{t}}}$, we can further rewrite problem (8.21) as

$$\begin{aligned} \min_{\mathbf{x},\mathbf{t}\succeq\mathbf{0}} \quad & \mathbf{q}^{\mathrm{T}}\left(\mathbf{A}^{-1}\mathbf{W}\mathbf{t} - \sqrt{p}\,\mathbf{H}_{\mathrm{b}}\mathbf{x}\right) + \left\|\mathbf{Q}\left(\sqrt{p}\,\mathbf{H}_{\mathrm{b}}\mathbf{x} - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right) - \mathbf{g}\right\|^2 \\ \text{s.t.} \quad & x_m \in \{-1, +1\},\ m = 1, 2, ..., 2bN_{\mathrm{t}}, \end{aligned} \tag{8.22}$$

where $\mathbf{x} \triangleq [x_1, x_2, ..., x_{2bN_{\mathrm{t}}}]^{\mathrm{T}}$ is the $2bN_{\mathrm{t}} \times 1$ vector of binary decision variables. The optimization problem (8.22) belongs to the family of quadratic form minimization over binary vectors, that are known to be NP-hard in general [211]. In the sequel, we introduce two polynomial-time alternative solutions to tackle this binary optimization problem.

Solution 1: Convex Relaxation

To deal with the binary constraints on $\mathbf{x}$, one simple approach is to solve a convex relaxation of (8.22) obtained by replacing the binary constraints on the elements of $\mathbf{x}$ with appropriate box constraints. This relaxed problem can be expressed as the following standard LCQP:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{t}\succeq\mathbf{0}} \quad & \mathbf{q}^{\mathrm{T}}\left(\mathbf{A}^{-1}\mathbf{W}\mathbf{t}-\sqrt{p}\,\mathbf{H}_{\mathrm{b}}\mathbf{x}\right)+\left\|\mathbf{Q}\left(\sqrt{p}\,\mathbf{H}_{\mathrm{b}}\mathbf{x}-\mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right)-\mathbf{g}\right\|^{2} \\ \text{s.t.} \quad & -1\leq x_{m}\leq 1,\ m=1,2,...,2bN_{\mathrm{t}}. \end{aligned} \tag{8.23}$$

It is worth mentioning that solving (8.23) results in a lower bound for the objective function of the original binary problem (8.22). However, as will be shown in Section 8.6, our numerical experiments reveal that a relatively noticeable number of the element-wise box constraints are not active at the optimum of (8.23), particularly when the transmit array size N_{t} is comparable to the number of UEs, N_{u}. This implies that the optimal solution to (8.23) might not be even a feasible solution to the original problem (8.22). Consequently, the desired objective in (8.10) may not be achieved, i.e., the quantization distortion may not be fully avoided. Such an observation suggests the possibility of further improvement of the relaxation method, while considering the solution to (8.23) as our performance benchmark in Section 8.6.

Solution 2: Equivalent Biconvex Formulation

In what follows, we aim to achieve a more accurate solution with reasonable complexity via treating the binary constraints in a more sophisticated way. We use an equivalent biconvex implication of the binary constraints, given in the following lemma, which was proven in [211].

Lemma 17. *Let $\mathbf{x}$ and $\mathbf{y}$ be two real-valued vectors of equal length $2bN_{\mathrm{t}}$. Then, provided that $-\mathbf{1}\preceq\mathbf{x}\preceq\mathbf{1}$ and $\mathbf{y}^{\mathrm{T}}\mathbf{y}\leq 2bN_{\mathrm{t}}$, the condition $\mathbf{x}^{\mathrm{T}}\mathbf{y}=2bN_{\mathrm{t}}$ implies that $\mathbf{x}=\mathbf{y}$ and $x_m\in\{-1,+1\}$ for all $m=1,2,...,2bN_{\mathrm{t}}$.*

Proof. See Appendix E.1 for a shorter proof. □

As a direct consequence of Lemma 17, we further state the following corollary, which has a straightforward proof.

Corollary 18. *The binary optimization problem*

$$\min_{\mathbf{x}}\ f(\mathbf{x})\ \text{ s.t. }\ x_m\in\{-1,+1\},\ m=1,2,...,2bN_{\mathrm{t}},\ \mathbf{x}\in\Theta, \tag{8.24}$$

where $f(\cdot)$ is a (not necessarily smooth) convex function on some convex set Θ, is equivalent to

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & -1\leq x_m\leq 1,\ m=1,2,...,2bN_{\mathrm{t}}, \\ & \mathbf{x}^{\mathrm{T}}\mathbf{y}=2bN_{\mathrm{t}},\ \ \mathbf{y}^{\mathrm{T}}\mathbf{y}\leq 2bN_{\mathrm{t}},\ \ \mathbf{x}\in\Theta. \end{aligned} \tag{8.25}$$

Using Corollary 18, we are able to rewrite the binary optimization problem (8.22) in an equivalent continuous-domain form as

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y},\mathbf{t}} \quad & \mathbf{q}^{\mathrm{T}}\left(\mathbf{A}^{-1}\mathbf{W}\mathbf{t}-\sqrt{p}\,\mathbf{H}_{\mathrm{b}}\mathbf{x}\right)+\left\|\mathbf{Q}\left(\sqrt{p}\,\mathbf{H}_{\mathrm{b}}\mathbf{x}-\mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right)-\mathbf{g}\right\|^{2} \\ \text{s.t.} \quad & -1\le x_m\le 1,\ m=1,2,...,2bN_{\mathrm{t}}, \\ & \mathbf{x}^{\mathrm{T}}\mathbf{y}=2bN_{\mathrm{t}},\ \ \mathbf{y}^{\mathrm{T}}\mathbf{y}\le 2bN_{\mathrm{t}},\ \ \mathbf{t}\succeq\mathbf{0}, \end{aligned} \tag{8.26}$$

where $\mathbf{x}^{\mathrm{T}}\mathbf{y}=2bN_{\mathrm{t}}$ is referred to as the equilibrium constraint. Reformulation (8.26) is still a non-convex problem due to the biconvex equilibrium constraint; however, efficient approaches exist to solve a biconvex problem, such as the exact penalty method or the alternating direction method of multipliers (ADMM). In this work, we adopt the exact penalty method due to its simplicity. The accuracy and convergence characteristics of the exact penalty method are studied in, e.g., [211] and [212].

According to the exact penalty method, the biconvex equilibrium constraint $\mathbf{x}^{\mathrm{T}}\mathbf{y}=2bN_{\mathrm{t}}$ can alternatively be implied by adding a penalty function to the objective function. The considered penalty function consists of the difference $2bN_{\mathrm{t}}-\mathbf{x}^{\mathrm{T}}\mathbf{y}$, as a measure of deviation from the equilibrium constraint, multiplied by a non-negative penalty parameter μ. Accordingly, denoting the objective function of (8.26) by $f(\mathbf{x},\mathbf{t})$, we can write

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y},\mathbf{t}} \quad & f(\mathbf{x},\mathbf{t})+\mu\left(2bN_{\mathrm{t}}-\mathbf{x}^{\mathrm{T}}\mathbf{y}\right), \\ \text{s.t.} \quad & -1\le x_m\le 1,\ m=1,2,...,2bN_{\mathrm{t}}, \\ & \mathbf{y}^{\mathrm{T}}\mathbf{y}\le 2bN_{\mathrm{t}},\ \ \mathbf{t}\succeq\mathbf{0}, \end{aligned} \tag{8.27}$$

It should be noted that, in general, problems (8.26) and (8.27) are not equivalent. However, by monotonically increasing the penalty parameter μ in each iteration up to a certain threshold, successive solutions of the penalized problem (8.27) eventually converge to the solution of the original biconvex problem. On the other hand, given $\mathbf{t}$, it can be shown that if $f(\mathbf{x},\mathbf{t})$ is a Lipschitz continuous convex function on $-\mathbf{1}\preceq\mathbf{x}\preceq\mathbf{1}$, problem (8.27) has the same local and global minima as those of (8.26) for $\mu\ge 2L$ with L denoting the Lipschitz constant of $f(\mathbf{x},\mathbf{t})$ with respect to $\mathbf{x}$; see [211, Th. 1]. As a result, finding at least a locally optimal solution to problem (8.26) is equivalent to obtaining a local optimum of (8.27) at least. The following lemma states that the Lipschitz continuity condition holds for the convex function $f(\mathbf{x},\mathbf{t})$ on the domain $-\mathbf{1}\preceq\mathbf{x}\preceq\mathbf{1}$.

Lemma 19. *Given* $\mathbf{t}$*, function* $f(\mathbf{x},\mathbf{t})$ *is* L*-Lipschitz continuous on* $-\mathbf{1}\preceq\mathbf{x}\preceq\mathbf{1}$ *and its Lipschitz constant is given by*

$$L \triangleq 2p\sqrt{2bN_{\mathrm{t}}}\,\|\mathbf{Q}\mathbf{H}_{\mathrm{b}}\|^{2}+2\sqrt{p}\left\|\mathbf{H}_{\mathrm{b}}^{\mathrm{T}}\left(\mathbf{Q}\mathbf{A}^{-1}\mathbf{W}\mathbf{t}+\mathbf{Q}\boldsymbol{\Sigma}\mathbf{s}+\frac{1}{2}\mathbf{q}\right)\right\|. \tag{8.28}$$

Proof. See Appendix E.2. □

Finally, we note that the objective function of problem (8.27), i.e., $f(\mathbf{x},\mathbf{t})+\mu(\mathbf{x}^{\mathrm{T}}\mathbf{y}-$

$2bN_\mathrm{t}$) is a biconvex quadratic function in $\mathbf{x}$ and $\mathbf{y}$, i.e., fixing either $\mathbf{x}$ or $\mathbf{y}$ gives a convex function in the other variable. Therefore, one can use a standard block coordinate descent (BCD) algorithm in order to solve (8.27). Here, a coordinate block refers to either of the vectors $\mathbf{x}$, $\mathbf{y}$ or $\mathbf{t}$. To be more specific, the objective function $f(\mathbf{x},\mathbf{t})+\mu(\mathbf{x}^\mathrm{T}\mathbf{y}-2bN_\mathrm{t})$ can be minimized over one of these vectors while the other two are fixed, and then, repeating the same procedure for the other two blocks. The penalty parameter μ can be increased monotonically, where the Lipschitz constant L provided in Lemma 19 determines the increment limit of μ as a function of the other variables. Based on this approach, the BCD algorithm solving (8.27) performs the following steps in the kth cycle:

i. sub-problem on t: Given $\mathbf{x}$, maximizing $f(\mathbf{x},\mathbf{t})$ over $\mathbf{t}$ is equivalent to a standard LCQP. Hence, the value of $\mathbf{t}$ is updated as the solution to the following minimization problem:

$$\mathbf{t}^{(k)} = \underset{\mathbf{t}\succeq\mathbf{0}}{\operatorname{argmax}}\ \mathbf{q}^\mathrm{T}\mathbf{A}^{-1}\mathbf{W}\mathbf{t} + \left\|\mathbf{Q}\left(\sqrt{p}\,\mathbf{H}_\mathrm{b}\mathbf{x}^{(k-1)} - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right) - \mathbf{g}\right\|^2. \tag{8.29}$$

ii. sub-problem on x: Given $\mathbf{t}$ and $\mathbf{y}$, the value of $\mathbf{x}$ in the kth cycle can be updated by solving the following box-constrained quadratic program:

$$\mathbf{x}^{(k)} = \underset{-\mathbf{1}\preceq\mathbf{x}\preceq\mathbf{1}}{\operatorname{argmin}}\quad f\left(\mathbf{x},\mathbf{t}^{(k)}\right) + \mu\left(2bN_\mathrm{t} - \mathbf{x}^\mathrm{T}\mathbf{y}^{(k-1)}\right). \tag{8.30}$$

iii. sub-problem on y: The kth update of $\mathbf{y}$ can be obtained as the optimal solution to the following problem:

$$\mathbf{y}^{(k)} = \underset{\mathbf{y}}{\operatorname{argmin}}\quad \mathbf{y}^\mathrm{T}\mathbf{x}^{(k)}\quad \text{s.t.}\quad \mathbf{y}^\mathrm{T}\mathbf{y} \leq 2bN_\mathrm{t}, \tag{8.31}$$

which is a norm-constrained inner product minimization, simply admitting a closed-form solution given by

$$\mathbf{y}^{(k)} = \sqrt{2bN_\mathrm{t}}\,\frac{\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|}. \tag{8.32}$$

iv. Updating μ: The penalty parameter μ is updated in every K cycles as

$$\mu \leftarrow \min\{2L, \theta\mu\}, \tag{8.33}$$

where $\theta > 1$ is a constant design parameter and L is the update of the Lipschitz constant L which is computed by substituting $\mathbf{x}$ and $\mathbf{t}$ in (8.28) with their respective updated values. The proposed algorithm is summarized in Algorithm 5. Eventually, the solution $\mathbf{x}$ obtained from Algorithm 5 can be reshaped into matrix form using the relation $\mathbf{x} = \mathrm{vec}(\mathbf{V})$, and then substituted in (8.20) in order to achieve the precoded signal $\bar{\mathbf{u}}$.

8.4.1 Computational Complexity Analysis

The overall computation cost of solving the finite-alphabet precoding design problem (8.27) in terms of the required number of arithmetic operations, using the four-step

Algorithm 5 BCD algorithm solving (8.27)

```
input: $\mathbf{A}, \mathbf{H}, \mathbf{W}, \mathbf{\Sigma}, \mathbf{s}, b$
output: $\mathbf{x}$
initialize: $\mathbf{x}^{(0)} = \mathbf{y}^{(0)} \in \mathbb{R}^{2bN_\mathrm{t}\times 1}, \mathbf{t}^{(0)} \in \mathbb{R}^{2N_\mathrm{u}\times 1}, \mu^{(0)}, k = 0$
set: $\theta > 1$, $K \geq 1$, $\epsilon_\mathrm{o} > 0$
while $\left|2bN_\mathrm{t} - \mathbf{x}^\mathrm{T}\mathbf{y}\right| > \epsilon_\mathrm{o}$ do
    $k \leftarrow k + 1$
    compute $\mathbf{t}^{(k)}$ and $\mathbf{x}^{(k)}$ by solving (8.29) and (8.30)
    $\mathbf{y}^{(k)} \leftarrow \sqrt{2bN_\mathrm{t}}\, \mathbf{x}^{(k)}/\|\mathbf{x}^{(k)}\|$
    update $L$ via (8.28)
    $\mu \leftarrow \min\{2L, \theta\mu\}$
end while
```

BCD approach summarized in Algorithm 5, is composed of two parts: inner iterations to solve the sub-problems (8.29) and (8.30) on $\mathbf{t}$ and $\mathbf{x}$, respectively, and updating $\mathbf{y}$ using (8.32), and the outer iterations (cycles) over coordinate blocks.

The computation cost of the first part is dominated by the arithmetic complexity of solving the two sub-problems corresponding to $\mathbf{t}$ and $\mathbf{x}$. To efficiently solve (8.29) and (8.30), one may use the off-the-shelf algorithms such as (accelerated) projected/proximal gradient methods [135], or quasi-Newton approaches, e.g., L-BFGS-B [213]. In particular, for a Lipschitz smooth (not necessarily strongly) convex objective function as in (8.29) and (8.30), all the aforementioned algorithms converge at a superlinear rate of $\mathcal{O}(1/\sqrt{\epsilon_\mathrm{i}})$ to reach an ϵ_i-optimal solution. For example, using the accelerated projected gradient descent algorithm, the per-iteration complexity associated with sub-problems (8.29) and (8.30) is dominated by matrix multiplications of limiting (i.e., as $N_\mathrm{t}, N_\mathrm{u} \to \infty$) orders N_u^2 and $b^2N_\mathrm{t}^2$, respective. Therefore, in the limiting case, the computational complexity of solving both inner sub-problems with an accuracy of ϵ_i is of order

$$\mathcal{C}_\mathrm{i} = \mathcal{O}\left(b^2N_\mathrm{t}^2 + N_\mathrm{u}^2\right) . (1/\sqrt{\epsilon_\mathrm{i}}), \tag{8.34}$$

which accounts for the dominating complexity order of one cycle of the BCD algorithm.

On the other hand, given the Lipschitz continuity property of $f(\mathbf{x}, \mathbf{t})$, the BCD algorithm based on the exact penalty method is guaranteed to converge to a first-order KKT point with an accuracy of at least $2bN_\mathrm{t} - \mathbf{x}^\mathrm{T}\mathbf{y} \leq \epsilon_\mathrm{o}$ in no more than $\left\lceil \left(\ln\left(2L\sqrt{bN_\mathrm{t}}\right) - \ln\left(\mu\epsilon_\mathrm{o}\right)\right) / \ln(\theta) \right\rceil$ iterations [211], where μ is the initial value of the penalty parameter μ and $\lceil\cdot\rceil$ denotes the ceiling operation. Thus, to have a complete analysis of the complexity, we further need to evaluate the Lipschitz constant L. In Appendix E.3, we obtain an approximate upper bound on L, which is valid in the limiting case where $N_\mathrm{t}, N_\mathrm{u} \to \infty$, as

$$L \lesssim \mathcal{O}\left(p\sqrt{bN_\mathrm{t}}\right). \tag{8.35}$$

Using (8.35), the maximum number of outer iterations (cycles) required to be performed by the BCD algorithm in order to achieve an accuracy of ϵ_o can be obtained as

$$\mathcal{C}_\mathrm{o} = \mathcal{O}\left(\lceil(\ln(2p\,bN_\mathrm{t}) - \ln(\mu\epsilon_\mathrm{o}))\,/\,\ln(\theta)\rceil\right). \tag{8.36}$$

Note that, in practice, we may treat μ and θ as constants and say that convergence is achieved in $\mathcal{O}\left(\ln(p\,bN_\mathrm{t}/\epsilon_\mathrm{o})\right)$ outer iterations. Therefore, the worst-case complexity of the BCD algorithm with accelerated inner gradient steps solving the optimization problem (8.27) is of the following limiting order:

$$\mathcal{C} = \mathcal{C}_\mathrm{i}\,.\,\mathcal{C}_\mathrm{o} = \mathcal{O}\left(\left(b^2N_\mathrm{t}^2 + N_\mathrm{u}^2\right)\ln\left(pbN_\mathrm{t}/\epsilon_\mathrm{o}\right)\right).(1/\sqrt{\epsilon_\mathrm{i}}), \tag{8.37}$$

from which it follows that the arithmetic complexity of solving the proposed finite-alphabet design scales by $\mathcal{O}(b^2\ln(b))$ with the number of resolution bits b. In our simulations, however, both inner and outer optimizations converge in a few (in the order of tens of) iterations, as we will see in Section 8.6.

8.5 Special Case: One-Bit Quantized Precoding

The special case of employing one-bit DACs at the BS is of high practical importance, since it is capable of significantly reducing the power consumption at the transmitter side. In such a setup, where $b = 1$, each I/Q channel is quantized via a one-bit DAC before passing through the RF chain and deriving the antenna elements. As a result, the scalar quantization operation $\mathcal{Q}(\cdot)$ simplifies to a sign operation, thereby the quantized signal can be represented as

$$u_{\mathrm{q},j} = \mathrm{sgn}\Big(\mathrm{Re}\{u_j\}\Big) + \mathrm{j}\,\mathrm{sgn}\Big(\mathrm{Im}\{u_j\}\Big), \quad j = 1, 2, ..., N_\mathrm{t}, \tag{8.38}$$

where $\mathrm{sgn}(\cdot)$ denotes the sign function. Similarly, the real-valued representation of the quantized signal becomes $\bar{u}_{\mathrm{q},j} = \mathrm{sgn}(\bar{u}_j)$ for all $j = 1, 2, ..., 2N_\mathrm{t}$. Having $B = 2^b = 2$ bins to represent the quantizer's output, we have the following set of quantization levels:

$$\mathbb{L} = \left\{-\frac{\Delta}{2}, +\frac{\Delta}{2}\right\}, \tag{8.39}$$

located around zero, which is the only quantization threshold in this particular case, i.e., $\mathbb{T} = \{0\}$. From (8.9), by substituting $B = 2$, we obtain the quantization step for the considered symmetric uniform as

$$\Delta = \sqrt{\frac{2}{N_\mathrm{t}}}. \tag{8.40}$$

Accordingly, the design goal in the case with one-bit DACs is to have the precoded transmit signal $\bar{\mathbf{u}}$ optimized such that

$$\bar{u}_j = \mathrm{sgn}(\bar{u}_j) = \bar{u}_{\mathrm{q},j}, \quad j = 1, 2, ..., 2N_\mathrm{t}, \tag{8.41}$$

along with the other design constraints as in problem (8.11). All the theoretical discussions and derivations in Section 8.3 and 8.4 remain valid also for the one-bit case. To obtain the precoded signal $\bar{\mathbf{u}}$, one can use Algorithm 5 by considering $b = 1$. It is worth mentioning that in this case, we have $\mathbf{b} = [2^0] = 1$ while $\mathbf{V}$ is a $2N_{\mathrm{t}} \times 1$ vector, and thus $\bar{\mathbf{u}} = \mathbf{V} = \mathrm{vec}(\mathbf{V}) = \mathbf{x}$, i.e., $\bar{\mathbf{u}}$ is a binary vector by itself and no further binary representation is needed. Therefore, one can directly obtain $\bar{\mathbf{u}}$ through replacing $\mathbf{x}$ with $\bar{\mathbf{u}}$ in Algorithm 5.

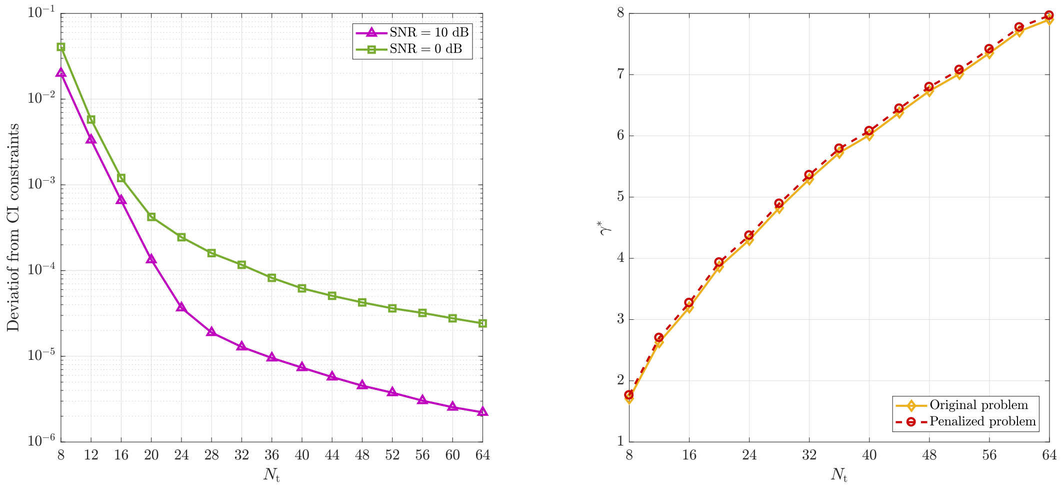
8.6 Simulation Results

The simulation setup is as follows. We consider a downlink MU-MIMO system with multiuser precoding and finite-precision quantization at the BS, where independent QPSK symbols are intended for the users. At the users' receiver sides, identical noise distributions $z_i \sim \mathcal{CN}(0, \sigma^2)$ with $\sigma^2 = 1$ are assumed, for all $i = 1, 2, ..., N_{\mathrm{u}}$. We assume a Rayleigh block fading channel, where uncorrelated vectors $\{\mathbf{h}_i\}_{i=1}^{N_{\mathrm{u}}}$ are randomly generated for each fading block following the standard circularly symmetric complex Gaussian distribution, i.e., $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. For the BCD algorithm, we set $\theta = 1.1$ to avoid overshooting and initialize the penalty parameter with a small value as $\mu^{(0)} = 0.01$ to obtain a reasonable starting point. We obtain the solutions to the inner sub-problems (8.29) and (8.30) within each cycle of the BCD algorithm using an accelerated projected gradient descent (APGD) algorithm [135]. Throughout this section, the finite-precision quantized precoding techniques of interest are referred to as:

- MSM: Maximum safety margin method of [85]
- SQUID: Squared-infinity norm Douglas-Rachford splitting technique of [207]
- RQSLP: Quantized SLP via convex relaxation as in (8.23)
- QSLP: Quantized SLP via biconvex formulation as in (8.27)

We compare the results with those obtained from the conventional matched filter (MF), ZF, and Wiener filter (WF) precoding techniques [5] with finite-precision quantized outputs. We also consider the infinite-precision WF precoding and the infinite-precision SLP, as our benchmarks. The presented results have been averaged over 100 fading block realizations, each of 100 symbols.

In Fig. 8.3, we assess the loss in optimality due to the penalized reformulation with soft CI constraints introduced in (8.13) with respect to the original problem in (8.12). Recall that the CI constraint, in its exact form, is enforced by the equality $\sqrt{p}\,\mathbf{H}\bar{\mathbf{u}} = \mathbf{\Sigma}\mathbf{s} + \mathbf{A}^{-1}\mathbf{W}\mathbf{t}$, as in problem (8.12). On the other hand, the soft CI constraints in the reformulated problem (8.13) are enforced via the penalty function $\|\sqrt{p}\,\mathbf{H}\bar{\mathbf{u}} - \mathbf{\Sigma}\mathbf{s} - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\|^2$. Hence, as a measure of deviation from the CI constraints, we consider the expression $\|\sqrt{p}\,\mathbf{H}\bar{\mathbf{u}} - \mathbf{\Sigma}\mathbf{s} - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\|^2$, which equates to zero when the CI constraints are all satisfied with equality. Note that the results shown in Fig. 8.3 have been obtained by excluding the binary constraints from problems (8.13) and (8.12). In Fig. 8.3 (a), we



(a)

(b)

FIGURE 8.3: The loss in optimality due to the penalized formulation with soft CI constraints for $N_\mathrm{u} = 4$: (a) Deviation form the CI constraints versus the number of transmit antennas N_t; (b) Comparison of γ^* obtained from the original and the reformulated problem versus N_t.

TABLE 8.1: Average percentage of inactive binary constraints in the RQSLP solution for different values of N_t with $N_\mathrm{u} = 4$ and $b = 1$.

SNR	Transmit array size (N_t)					
	8	12	16	24	32	64
0 dB	35%	26%	21%	15%	12%	7%
10 dB	40%	30%	24%	18%	15%	9%

plot this deviation as a function of N_t, for a fixed $N_\mathrm{u} = 4$ and for two different transmit SNRs. As can be seen, the deviation from CI constraints is smaller than 0.1 in the entire range of N_t, and further, it dramatically decreases with increasing N_t. Moreover, it is shown that at higher SNRs, the soft CI constraints are satisfied with more accuracy. We also plot in Fig. 8.3 (b) the optimal value of variable γ, denoted by γ^*, obtained from problems (8.13) and (8.12). This figure shows that both problems result in almost the same value for γ^*. Overall, it follows from Fig. 8.3 that the loss in optimality due to the penalized reformulation (8.13) is quite insignificant, particularly for large values of N_t.

As mentioned in Section 8.4, by simply treating the binary constraints through convex relaxation, as in problem (8.23), we may not be able to satisfy some of the binary constraints at the optimum point. In order to evaluate this issue, we report, in Table 8.1, the average percentage of inactive constraints at the optimum of problem (8.23) as a

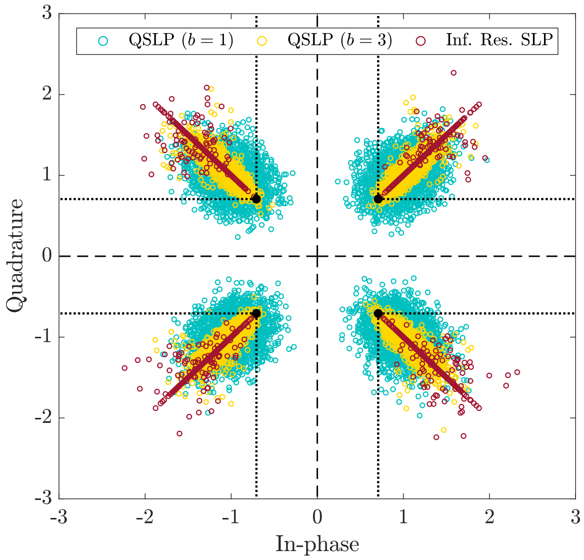


FIGURE 8.4: The users' noise-free received signal with $(N_{\mathrm{t}}, N_{\mathrm{u}}) = (16, 4)$ at a transmit SNR of 0 dB, i.e., $p/\sigma^2 = 1$.

function of the transmit array size N_{t}. By inactive constraints, we refer to those binary constraints that are not satisfied with equality. It can be seen that for a larger transmit array, more constraints become active at the optimum point. However, for smaller values of N_{t}, a noticeable percentage of the binary constraints remain inactive. For example, with $N_{\mathrm{t}} = 16$, around 8 constraints out of a total number of $2N_{\mathrm{t}} = 32$ constraints are not satisfied. This may lead to a significant performance loss, as we will see later in this section. Furthermore, it is shown in Table 8.1 that the percentage of inactive constraints becomes larger at higher SNRs. As such, one expects the bit error rate curve to show an error floor in the high SNR regime, as will be verified by the subsequent results.

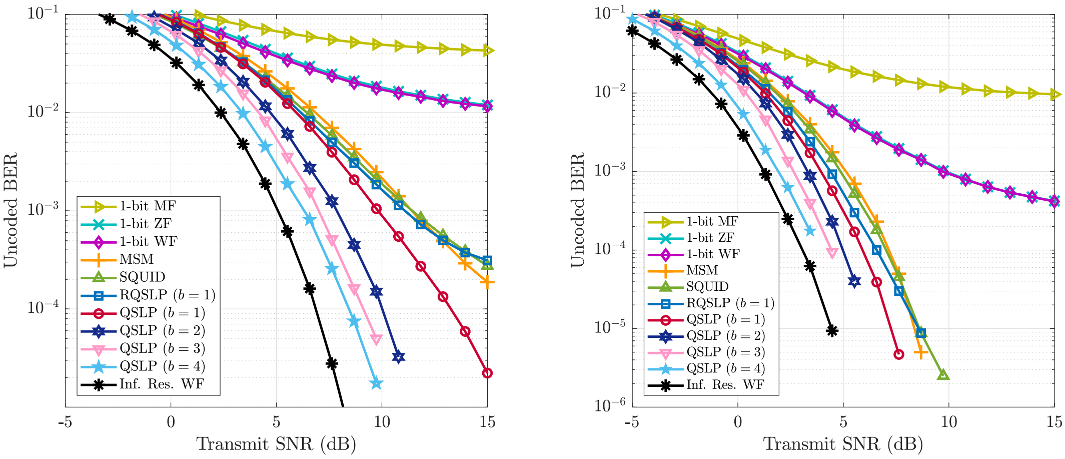
For CI-based symbol-level precoded downlink transmission, the scatter plot of the noise-free signals received at the UEs is depicted in Fig. 8.4 with different DAC resolutions, where the dashed and the dotted lines respectively represent the ML decision regions and the CI regions. The received signals, in all cases, are pushed away from the decision boundaries. This leads to an improved detection performance as we will see later in this section. Using the SLP technique with infinite-precision outputs, i.e. problem (8.22) excluding the binary constraints, the received signals are desirably located within the corresponding distance-preserving CI region. However, as it might be expected, the received signals resulted from the QSLP approach with finite-resolution DACs are more spread over the CI region. In particular, it can be seen from the figure that increasing the DACs' resolution from $b = 1$ to $b = 3$ reduces the variance of the received signal cloud, and at the same time, pushes the signals farther from the decision boundaries which can enhance the symbol error probability.

In Figs. 8.5 (a) and 8.5 (b), we compare the BER performances achieved by the precoding schemes of interest, with different DAC resolutions, versus transmit SNR, i.e., p/σ^2 for two practical systems with $(N_\text{t}, N_\text{u}) = (16, 4)$ and $(N_\text{t}, N_\text{u}) = (64, 8)$. The results obtained from the QSLP approach indicates that employing DACs with higher resolutions at the BS results in a noticeably improved BER performance at the UEs. In particular, increasing b by one bit leads to at least 0.5 dB gain at BER $= 10^{-3}$. Remarkably, the proposed QSLP approach with $b = 4$ shows an uncoded BER performance well close to that of the infinite-resolution WF precoding scheme, but with a lower hardware complexity and power consumption, especially for the larger system with $(N_\text{t}, N_\text{u}) = (64, 8)$ where the performance loss due to using DACs with $b = 4$ bits of resolution at BER $= 10^{-3}$ is less than 0.8 dB. On the other hand, for the one-bit quantized case, it can be seen that the proposed QSLP method with $b = 1$, outperforms both the MSM and the SQUID one-bit precoders. The gain is around 1 dB in the applicable range of SNR (i.e., 5-10 dB) for an uncoded QPSK signaling. Furthermore, the QSLP approach performs superior to our naive precoding formulation RQSLP, which exploits CI in the design but simply treats the one-bit constraints via convex relaxation. From Fig. 8.5 (a), we can further observe that the one-bit precoders MSM, SQUID and RQSLP all experience an error floor at high SNRs (i.e., above 15 dB). This indicates that these one-bit precoders require more degree of freedoms (either more transmit antennas or higher resolution bits) to perform well for the multiuser system with $(N_\text{t}, N_\text{u}) = (16, 4)$. Further, we plot, in Fig. 8.6, the BER performances of different precoding schemes versus the number of transmit antennas N_t, with the number of UEs fixed as $N_\text{u} = 8$. As can be seen from this figure, the BER (in logarithmic scale) decreases linearly with increasing N_t for all the schemes, but with different slopes. In particular, using the QSLP approach with higher resolution DACs results in a larger reduction slope for the BER curve as a function of N_t. Moreover, in comparison with the MSM and the SQUID one-bit precoders, the one-bit QSLP technique shows a larger reduction slope, and hence, a better BER performance, as N_t increases. As an illustrative example, for fixed $N_\text{u} = 8$, the MSM technique requires 5 more transmit antennas than that required for the one-bit QSLP precoder to achieve BER $= 10^{-3}$ at an SNR of 5 dB.

We saw from the results in Figs. 8.5 that the QSLP approach has a degraded BER performance compared to the infinite-resolution WF scheme. This degraded performance, however, is achieved using finite-resolution DACs with much lower power consumption. In order to have a fair comparison, we introduce power efficiency as a figure of merit that incorporates both aforementioned performance measures. More precisely, we define the power efficiency ρ as the ratio between the number of successfully decoded bits at the UEs and the amount of power consumption in Watts at the BS, i.e.,

$$\rho \triangleq \frac{(1 - \text{BER}) \log_2(M)}{P_\text{BS}}, \tag{8.42}$$

where M is the modulation order and P_BS denotes the power consumption at the BS. We consider the overall power dissipated by the BS's RF front-end components as the power consumption at the BS and adopt a simple model for this power as follows. The transmit RF front-end of a multi-antenna system is commonly composed of one baseband



(a)

(b)

FIGURE 8.5: Average per-user BER versus transmit SNR for an MU-MIMO system with (a) $(N_\mathrm{t}, N_\mathrm{u}) = (16, 4)$; (b) $(N_\mathrm{t}, N_\mathrm{u}) = (64, 8)$.

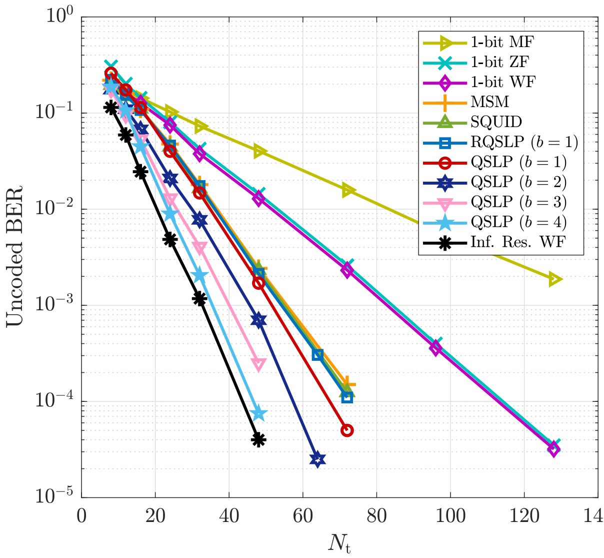


FIGURE 8.6: Average users' BER versus the number of transmit antennas N_t for a fixed number of users $N_\mathrm{u} = 8$ at a transmit SNR of 5 dB.

processor, several RF chains, each preceded by a pair of DACs (i.e., one DAC for each I/Q channel), and power amplifiers (PA). Accordingly, the BS transmit architecture requires $2N_{\mathrm{t}}$ DACs, N_{t} RF chains and PAs, and therefore, its power consumption can be modeled as

$$P_{\mathrm{FD}} = P_{\mathrm{BB}} + N_{\mathrm{t}}(2P_{\mathrm{DAC}} + P_{\mathrm{RF}} + P_{\mathrm{PA}}), \tag{8.43}$$

where P_{BB}, P_{RF}, P_{PA}, and P_{DAC} respectively denote the power consumption of the baseband processor, one single RF chain, one single PA and one single DAC. Broadly speaking, the power consumption of a DAC scales linearly in sampling rate and exponentially in the number of bits per sample (i.e., resolution bits). For DACs of binary-weighted current-steering type [214], the approximate power consumption of a single DAC is given in [215] by

$$P_{\mathrm{DAC}} = \frac{3}{2}\left(2^{b} - 1\right) \times 10^{-5} + \frac{9}{2}\, b\, F_{\mathrm{s}} \times 10^{-12}, \tag{8.44}$$

where F_{s} denotes the sampling frequency. In our simulations, we consider $F_{\mathrm{s}} = 1$ GHz and reference values of $P_{\mathrm{RF}} = 40$ mW, $P_{\mathrm{PA}} = 20$ mW, and $P_{\mathrm{BB}} = P_{\mathrm{DAC}}$, as in [203].

In Fig. 8.7, we plot the energy efficiency of the proposed QSLP approach with DACs of different resolution bits b versus transmit SNR and the number of transmit antennas, respectively. We further consider the WF scheme with moderate-resolution 8-bit DACS as our benchmark for comparison. In Fig. 8.7, the power efficiency is plotted as a function of the number of transmit antennas N_{t} for a fixed number of UEs, i.e., $N_{\mathrm{u}} = 8$. The proposed one-bit QSLP approach is shown to be the most power efficient scheme among the others. Particularly, the gain in power efficiency compared to the WF scheme with 8-bit DACs is about 0.84 bit/Watt for $N_{\mathrm{t}} = 16$. This is obviously due to the fact that a much lower amount of power – about 9 mW according to (8.44) – is consumed by one-bit DACs, as compared to a power consumption of around 76 mW for 8-bit DACs. It can further be seen from the figure that for a larger transmit antenna array, a lower power efficiency can be achieved. This can be verified via the definition of power efficiency in (8.42) as a reciprocal function of the power consumption, with the power consumption scaling linearly with N_{t}.

Following the analytic discussion on computational complexity in Section 8.4.1, we numerically evaluate the complexity of the QSLP technique in Fig. 8.8 and 8.9. For different values of DAC resolution b, the complexity in terms of the required number of outer iterations (i.e., cycles) till convergence of the BCD algorithm solving (8.27), is shown in Fig. 8.8 versus transmit SNR in linear scale, i.e., p/σ^2. The complexity analysis in Section 8.4.1 indicates that the number of cycles till convergence of the BCD algorithm scales logarithmically with b and p, i.e., with orders $\mathcal{O}\left(\ln(b)\right)$ and $\mathcal{O}\left(\ln(p)\right)$, respectively. This can be further verified from the numerical results in Fig. 8.8. Recall that in our simulations, we consider $\sigma^2 = 1$, and hence p/σ^2 refers also to the transmit power. For the special case of $b = 1$, the numbers of inner and outer iterations required for a normalized squared error of 10^{-4} are separately plotted in Fig. 8.9. It can be seen

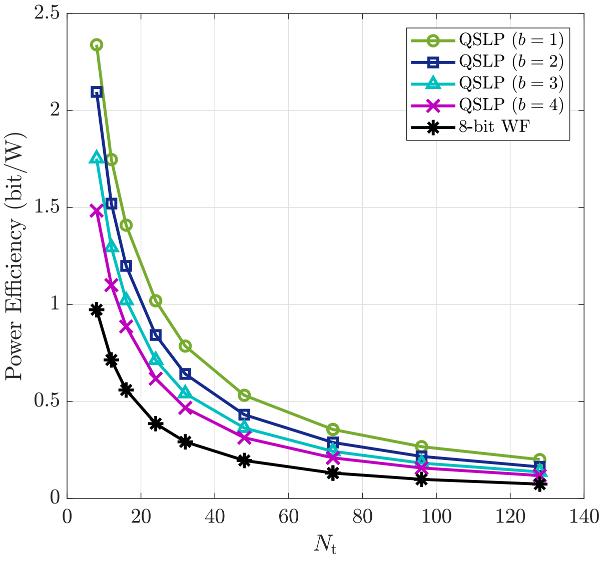


FIGURE 8.7: Power efficiency as a function of N_t for a fixed number of users $N_u = 8$ at SNR = 5 dB.

from the simulation results that all the iteration numbers grow logarithmically with transmit power p, while the number of inner iterations to solve the sub-problem on $\mathbf{t}$ shows a relatively faster growth with p. However, we remark that the update iterations on $\mathbf{t}$ are of dimension $2N_u$, whereas the dominant complexity order comes from the even larger dimension $2N_t$. For this reason, in our subsequent evaluation, only those $2N_t$-dimensional computations updating $\mathbf{x}$ are accounted for the complexity cost of the QSLP approach. To evaluate the convergence behavior of the BCD algorithm solving (8.27) as a function of system parameters, i.e., N_t and N_u, we report in Table 8.2 the average number of required iterations for different values of p/σ^2 within the effective range of SNR associated with QPSK signaling. With reference to Table 8.2, the proposed QSLP method offers a favorably fast convergence speed, in the order of tens of iterations, even for large system parameters. For instance, at $p/\sigma^2 = 3.4$ ($\approx$ 5.4 dB), the QSLP algorithm needs only $\sim$ 48 and $\sim$ 56 iterations on average to achieve those uncoded BER performances as shown in Figs. 8.5 (a) and 8.5 (b). Table 8.2, on the other hand, indicates that the complexity of QSLP (in terms of the number of iterations till convergence) scales linearly with N_t, which is an attractive feature for implementation purposes.

8.7 Conclusions

We proposed a finite-alphabet symbol-level multiuser precoding scheme for massive MU-MIMO downlink system equipped with finite-resolution DACs. To design the precoder, we adopted a power-constrained max-min fair criterion with the aim of exploiting CI

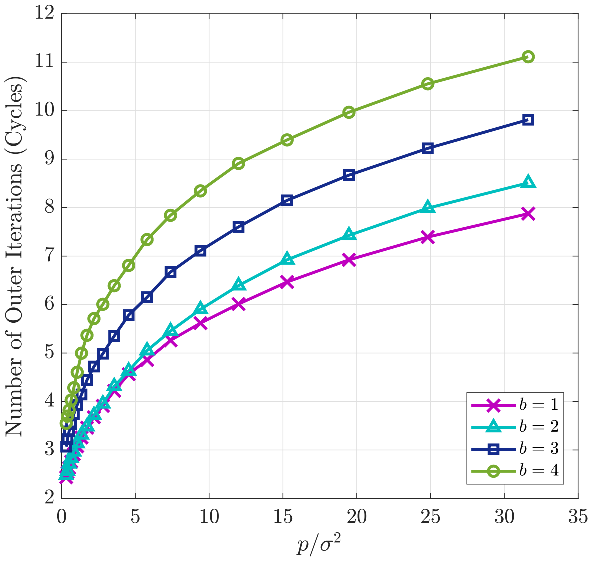


FIGURE 8.8: Average number of outer iterations of the QSLP method to reach a squared error of 10^{-4} as a function of transmit SNR with $(N_\mathrm{t}, N_\mathrm{u}) = (16, 4)$.

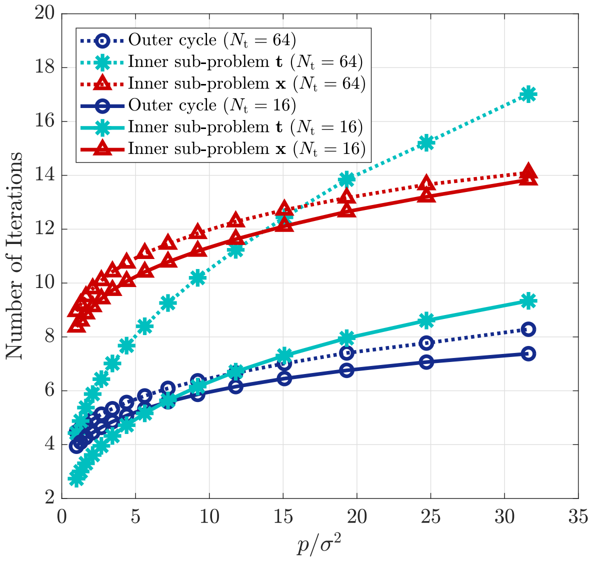


FIGURE 8.9: Average number of outer and inner iterations for the QSLP method to reach a squared error of 10^{-4} as a function of transmit SNR in linear scale.

Table 8.2: Average number of iterations with computations of dimension $2N_{\mathrm{t}}$ till convergence of the QSLP algorithm.

p/σ^2	$(N_{\mathrm{t}}, N_{\mathrm{u}})$			
	$(8,2)$	$(16,4)$	$(64,8)$	$(128,16)$
3.4 ($\approx$ 5.4 dB)	35.3	47.3	55.7	64.9
9.2 ($\approx$ 9.6 dB)	48.8	65.8	75.7	87.2

at the users. The design problem of interest, in its original form, is a discrete linearly-constrained quadratic programming whose solution requires a high computational complexity. We dealt with this issue in several steps and reformulated the problem into an equivalent continuous-domain form, which can efficiently be solved using a standard block coordinate descent algorithm. We showed by simulation results that employing DACs with higher resolutions leads to lower BERs, but at the cost of reducing the power efficiency. Focusing on the case with one-bit DACs, we observed that comparisons between our proposed quantized symbol-level precoding (namely, QSLP) technique and some other well-known one-bit precoding schemes shows a superior performance in terms of uncoded BER, with up to 2 dB gain depending on the simulation setup. Furthermore, our analytical and numerical analyses on complexity indicate that the proposed QSLP algorithm converges in a few tens of iterations in practical massive MU-MIMO systems.

Chapter 9

Hybrid Symbol-Level Precoding for mmWave MU-MIMO Systems

In this chapter, we address the SLP design problem for a millimeter wave (mmWave) downlink MU-MIMO wireless system where the transmitter is equipped with a large-scale antenna array. The high cost and power consumption associated with the massive use of radio frequency (RF) chains prohibit fully-digital implementation of the precoder. Therefore, we consider a hybrid analog-digital architecture where a small-sized baseband precoder is followed by two successive networks of analog on-off switches and variable phase shifters according to a fully-connected structure. Using the switching network allows us to implement a phase shifter selection mechanism. We jointly optimize the digital baseband precoder and the states of the switching network on a symbol-level basis, i.e., by exploiting both the CSI and the instantaneous data symbols. In contrast, the phase-shifting network is designed only based on the CSI due to practical considerations. Our approach to this joint optimization is to minimize the Euclidean distance between the optimal fully-digital and the hybrid symbol-level precoders. The phase shifter selection mechanism allows for significant power-savings in the analog precoder by switching some of the phase shifters off according to the switches' instantaneously optimized states. Our numerical results indicate that up to 50 percent of the phase shifters can be switched off, on average, in systems where the number of transmit antennas is much larger that the number of RF chains and users. We provide an analysis of energy efficiency by adopting appropriate power consumption models for the analog precoder. Accordingly, we show that the energy efficiency of precoding can substantially be improved thanks to the phase shifter selection approach, compared to the fully-digital and state-of-the-art hybrid symbol-level schemes.

9.1 Introduction

Millimeter wave (mmWave) communication has been widely accepted as a prime technology for the emerging outdoor/indoor wireless communication deployments, enabling multi-gigabit-per-second data rates thanks to the enormously available unregulated spectrum resources within 30-300 GHz band [216–218]. Communication in the mmWave band, however, suffers from an order-of-magnitude increase in the free-space path loss, higher shadow fading, and more severe penetration losses compared to the legacy lower-frequency systems [219]. On the other hand, the shorter wavelength of mmWave signals makes it possible to pack a larger number of antenna elements in the same physical dimension, allowing for large-scale spatial multiplexing and highly directional beamforming. Employing large antenna arrays, commonly known as massive MIMO, can further provide considerable beamforming gain to compensate for severe propagation losses at mmWave frequencies [197], which is indispensable to achieve high-quality communication links in mmWave systems.

In traditional MIMO systems, the convention is to perform baseband precoding fully in the digital domain, which enables modification of both the amplitudes and phases of complex signals [3, 4]. This fully-digital signal processing, however, requires one dedicated radio frequency (RF) chain per antenna element, which is challenging to implement in practical systems with large antenna arrays due to the prohibitive cost and high power consumption of mixed-signal components, especially when operating at mmWave frequencies [195]. Given mmWave massive MIMO practical constraints, the design of cost-effective low-complexity precoding implementations has become an active line of research. Various precoding schemes, mostly aimed at either simplification of or reducing the number of RF chains, have been proposed for both single-user and multiuser MIMO systems, among which we refer to analog-only beamforming using RF phase shifters [220–222], antenna (sub-set) selection [223, 224], quantized fully-digital precoding via low-resolution (especially one-bit) digital-to-analog converters (DAC) [205, 207], and hybrid analog-digital beamforming [195, 197, 225–227].

Hybrid analog-digital precoding is a cost-effective alternative to enable both multi-stream transmission and large beamforming gains via splitting the signal processing operation between the digital and analog domains. In hybrid architectures, a small-sized digital precoder is followed by a high-dimensional analog precoder which is usually implemented using RF phase shifters and/or switches [225]. Such a setup allows for employing fewer RF chains, scaling with the number of multiplexed data streams rather than the number of antennas. Specifically, in multi-user mmWave systems, the digital precoder is so designed to mitigate the inter-user interference, whereas the analog RF precoder is used to improve the antenna array gain [228]. Nevertheless, while designing the digital precoder is straightforward, the design and implementation of the analog precoder are usually nontrivial.

For large-scale multiuser mmWave systems, the design of block-level hybrid schemes where the precoding solution solely relies on the CSI, has been extensively addressed.

However, symbol-level approaches to hybrid precoding are not yet well studied. Symbol-level hybrid precoding design under mmWave hardware limitations has been addressed in some recent work [80, 81, 83]. In [80], the authors adopt a disjoint sub-optimal approach to optimize the digital and analog precoders with a focus on the analog precoder design, where different techniques are studied and compared. Power-efficient transmitter architectures, including antenna selection and analog-only, are studied for symbol-level precoding in [75], where it has been shown that the analog-only design can outperform the other schemes especially when the transmit array size is much larger than the number of UEs. An even more cost-effective hybrid structure is considered in [81] where the baseband digitally precoded signal is subject to one-bit quantization due to the use of low-cost one-bit DACs for each RF chain. This excessive constraint, however, may limit the potential gain of symbol-level baseband signal processing. The joint optimization of digital and analog symbol-level precoders is addressed in [83], where the authors exploit the symbol-based design of the phase-shifting network to achieve the performance of the fully-digital precoder. In practice, the design needs to switch between the phase states of the variable phase shifters at the symbol rate. Keeping in mind target data rates of multi-Gbps in mmWave systems, such a high phase-switching speed requirement might be prohibitive in two aspects: first, it significantly increases the power consumption in the analog circuitry, and second yet, more importantly, it might be challenging from an implementation point of view considering the current RF semiconductor technologies [229]. Among the aforementioned symbol-level precoding techniques, those proposed in [81] and [83] are more related to the scope of this work, and therefore will be considered in the chapter for comparison purposes.

Analog phase shifters and switches are two key components of the mmWave systems. A wide variety of hybrid precoding architectures are essentially based on employing either phase shifters or switches, or even a combination of both where the phase-shifting network is controlled by a preceding network of switches; see, e.g., [225] and [203] where several possible architectures are described. Employing the combination of phase-shifting and switching networks in the analog RF precoder has a two-fold advantage. On the one hand, it can provide additional degrees-of-freedom (DoF) brought by the switching network when designing the analog precoder, and on the other hand, it allows for potential power-savings through switching some of the phase shifters off. From a power consumption perspective, one further needs to take into account the excessive power consumed by the switching network. For this purpose, power consumption models such as those introduced in [203] and [205] can be used. However, roughly speaking, the excessive power consumption due to the operation of switches is relatively small compared to the power reduction in the phase-shifting network. One reason is that, in general, switches consume less power than phase shifters. Furthermore, recent advances in RF circuit design have enabled the implementation of low-power high-performance switches working at mmWave frequencies, making the switching operation even more energy-efficient; see, e.g., [230–232]. Therefore, the use of analog switches in combination with the phase-shifting network is an attractive architecture for hybrid mmWave systems. In this line of research, hybrid implementations with the so-called phase shifter selection, where a two-state on-off switch precedes each phase shifter, have been studied for conventional block-level precoding; see [233–235]. For example, in [233], it has been shown that the combination of phase shifters and switches offers noticeably higher energy efficiency compared to phase shifter-only architectures, while the spectral efficiency is almost preserved. More specifically, significant power consumption reductions are possible without sacrificing the spectral efficiency even when up to 50% of the phase shifters are turned off [234]. To the best of authors' knowledge, such an approach has not been investigated so far for hybrid symbol-level precoding.

In this work, we consider a hybrid analog-digital architecture for symbol-level precoder where the analog precoder is implemented using a network of variable phase shifters preceded by an on-off switching network of the same dimension according to a fully-connected structure. As for the analog precoder, the phase states of the phase-shifting network are designed solely based on the instantaneous CSI, i.e., they stay unchanged within the duration of one channel coherence block. On the other hand, the on-off states of the switches as well as the baseband digital precoder are jointly optimized in our design on a symbol-level basis. Our approach to this optimization is to minimize the ℓ_2-norm distance between the outputs of hybrid and optimal fully-digital symbol-level precoders. For the latter precoder, we adopt a power-constrained max-min SINR criterion subject to user-specific CI constraints, where the CI constraints are assumed to be distance-preserving, as characterized in [23]. Accordingly, the main contributions of this chapter are as follows:

1. We exploit the notion of CI along with the phase shifter selection approach in designing the hybrid precoder. The CI-based design can improve the symbol detection performance at the receiver side, while the phase shifter selection approach brings additional DoF to the design problem and further enables the reduction of dissipated power in the phase-shifting network. The use of on-off switches, however, makes our design problem an NP-hard binary optimization. We deal with this difficulty by transforming the original problem into a biconvex form using an equivalent continuous-domain implication of the binary constraints. Efficient suboptimal solutions can then be obtained via a standard block coordinate descent (BCD) algorithm.

2. We study the convergence of the proposed hybrid precoding algorithm, where it will be shown that convergence to a stationary point is guaranteed. We further analyze the required computational complexity in the large system limit. In our analysis, we consider both the Newton complexity, i.e., the number of iterations required till the BCD algorithm converges, and the per-iteration complexity. Moreover, we show via simulation results that the BCD algorithm usually converges within a few iterations for practical values of system parameters, i.e., array size, number of RF chains, and users.

3. We provide an analysis of energy efficiency, incorporating both performance and power consumption, to evaluate and compare different fully-digital/hybrid precoding architectures. For this purpose, we adopt appropriate power consumption

models to take into account the power dissipated by the transmitter's RF circuitry. According to this analysis, the phase shifter selection mechanism offers significant improvements in the energy efficiency of precoding by switching off up to 50 percent of the phase shifters.

4. Our design approach is independent of the phase-shifting precision; however, to evaluate how this affects the ultimate precoding performance, in our simulations, we consider two implementations using infinite and finite resolution phase shifters. It will be shown that implementing the phase-shifter-selection-enabled analog precoder using low-resolution phase shifters can lead to gains of tens of Mbps/Joule per user in energy efficiency, compared to the case with infinite-resolution phase shifters.

The rest of this chapter is organized as follows. In Section 9.2, we describe the adopted system, signal, and channel model. We begin Section 9.3 by designing the phase-shifting network. Then, we study the symbol-level precoding problem for fully-digital architecture. This is followed by the derivation of the proposed hybrid precoding algorithm and analyses of its convergence and computational complexity. In Section 9.4, we provide energy efficiency analysis and explain the power consumption model. Simulation results are presented and discussed in Section 9.5. Finally, we conclude the chapter in Section 9.6.

9.2 System, Signal and Channel Model

In this section, we describe the system and channel model considered in this chapter.

9.2.1 System and Signal Model

We consider a single-cell single-carrier mmWave multiuser MIMO system. The BS, which is equipped with a large-scale antenna array of N_t elements and a (typically) much smaller number of transmit RF chains, denoted by N_l, simultaneously communicates independent data streams to N_u single-antenna users, each supporting single-stream transmission. The maximum number of transmitted data streams (i.e., the maximum number of users scheduled within a transmission block) is limited by the number of available RF chains at the BS, which leads to the assumption $N_\mathrm{u} \leq N_\mathrm{l} < N_\mathrm{t}$. Due to the limited number of transmit RF chains, the fully-digital implementation of multiuser precoder is not possible, and therefore, a hybrid digital-analog architecture is employed where the digital baseband precoder is followed by the RF chains and an analog RF precoder, as shown in Fig. 9.1. It is worth noting that the baseband precoder is capable of modifying both the amplitudes and phases of the input symbols while the RF precoder adjusts only the phases of the upconverted RF signals.

Digital baseband precoder

We consider a (non-linear) symbol-level baseband precoder that calculates the digital outputs specifically for every set of the users' intended symbols. Accordingly, the discrete-time complex-valued $N_\mathrm{u} \times 1$ modulated symbol vector $\mathbf{s} = [s_1, s_2, ..., s_{N_\mathrm{u}}]^\mathrm{T}$, where $\mathbb{E}\{\mathbf{s}\mathbf{s}^\mathrm{H}\} = \mathbf{I}$, is preprocessed in the digital domain using the symbol-level precoder, resulting in the output baseband signal $\mathbf{u}_{\mathrm{BB}} \in \mathbb{C}^{N_\mathrm{l} \times 1}$. In contrast to linear precoding schemes, the nonlinear-precoded signal $\mathbf{u}_{\mathrm{BB}}$ is directly designed and may not be uniquely decomposable as a linear combination of the users' precoding vectors. The baseband precoded signal $\mathbf{u}_{\mathrm{BB}}$ is then passed through the RF chains for upconversion to the carrier frequency.

Analog RF precoder

We assume the analog precoder to be implemented following a fully-connected structure with two successive switching and phase-shifting networks of dimension $N_\mathrm{l} \times N_\mathrm{t}$ that map N_l digital outputs to N_t precoded analog signals feeding the transmit antennas; see [225, 234, 235] where similar hybrid architectures have been considered. To be more specific, each RF chain upconverts a digitally precoded signal and feeds it to a phased-array with N_t variable phase shifters, each preceded by a dedicated analog on-off switch determining whether the corresponding phase shifter is active or deactivated. The phase-shifting network outputs are then combined through N_t analog combiners before being fed to the antenna elements. Let $\mathbf{F} \in \mathbb{C}^{N_\mathrm{t} \times N_\mathrm{l}}$ and $\mathbf{T} \in \mathbb{B}$ represent the phase-shifting network and the on-off states of the switching network, respectively, where $\mathbb{B} \triangleq \left\{\mathbf{Y} \in \{0,1\}^{N_\mathrm{t} \times N_\mathrm{l}} \,|\, \mathbf{Y}\mathbf{1} \succeq \mathbf{1}, \mathbf{Y}^\mathrm{T}\mathbf{1} \succeq \mathbf{1}\right\}$. Thereby, the entire RF precoder can be represented by $\mathbf{F} \circ \mathbf{T}$. Note that the set $\mathbb{B}$ is defined such that the selection matrix $\mathbf{T}$ has no all-zero row and column, where the former case corresponds to an antenna selection scheme, and the latter case excludes an RF chain from the transmitter's analog circuitry; however, neither is of interest in this work. We further note that since $\mathbf{F}$ is implemented using analog phase shifters, each element of $\mathbf{F} \in \mathbb{C}^{N_\mathrm{t} \times N_\mathrm{l}}$ is normalized such that $|f_{k,j}| = 1/\sqrt{N_\mathrm{t}}$ with $|f_{k,j}|$ denoting the magnitude of the element in the kth row and jth column of $\mathbf{F}$.

Under the described system model, the vector collecting the baseband received signals for all N_u users is given by

$$\mathbf{r} = \sqrt{\rho}\,\mathbf{H}\left(\mathbf{F} \circ \mathbf{T}\right)\mathbf{u}_{\mathrm{BB}} + \mathbf{z}, \tag{9.1}$$

where $\mathbf{r} \in \mathbb{C}^{N_\mathrm{u} \times 1}$ is the received signal vector, ρ is the instantaneous transmit power, $\mathbf{H} \in \mathbb{C}^{N_\mathrm{u} \times N_\mathrm{t}}$ represents the mmWave multiuser channel, and $\mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}\right)$ is a circularly symmetric complex Gaussian noise vector with $\mathbf{\Sigma} \triangleq \mathrm{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_{N_\mathrm{u}}^2)$ where σ_i^2 denotes the noise variance at the receiver of the ith user, for $i = 1, 2, ..., N_\mathrm{u}$. The instantaneous total transmit power is constrained by ρ through enforcing $\|(\mathbf{F} \circ \mathbf{T})\mathbf{u}_{\mathrm{BB}}\|^2 = 1$. It is further assumed that the BS has perfect knowledge of the instantaneous channel $\mathbf{H}$. In practical wireless systems, the CSI can be estimated at the receiver via, e.g., pilots

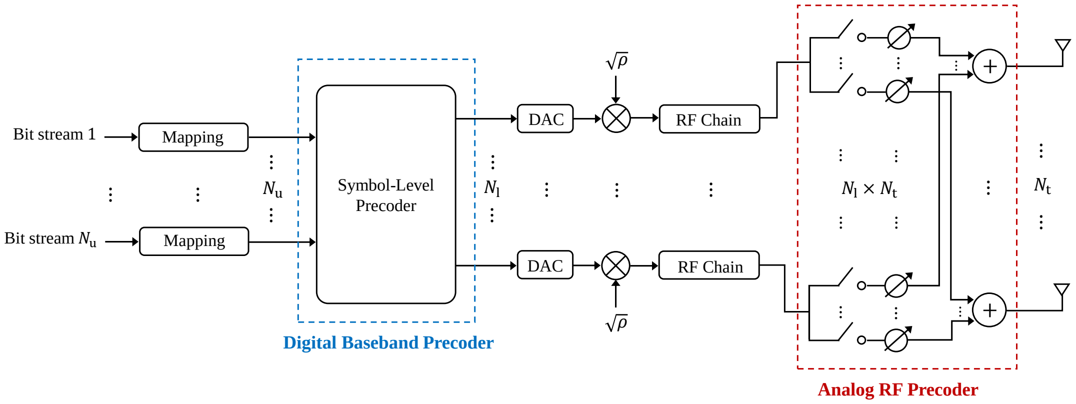


FIGURE 9.1: Schematic diagram of the considered hybrid transmitter architecture with fully-connected switching and phase-shifting networks.

or training sequences, and then fed back to the BS. Efficient mmWave channel estimation techniques that exploit the geometric nature of mmWave channels are presented in [236, 237]. At the receiver side, we assume that each user is capable of performing optimal single-user detection of the received signal using, e.g., the maximum likelihood (ML) detector.

9.2.2 Multiuser mmWave Channel Model

The mmWave propagation environment is known to feature limited multipath components. To capture this sparse scattering nature, the narrowband clustered channel modeling based on the Saleh-Valenzuela model is commonly used [238–240]. Under this model, the channel vector corresponding to a single user is a summation over the contributions of N_c scattering clusters, with each cluster contributing N_p propagation paths between the BS and the user. Assuming the same number of clusters and scatterers to be seen by each user, the narrowband mmWave channel vector for the jth user can be expressed as

$$\mathbf{h}_j^{\mathrm{H}} = \sqrt{\frac{N_t}{N_c N_p}} \sum_{i=1}^{N_c} \sum_{l=1}^{N_p} \alpha_{j,i,l}\, \mathbf{a}^{\mathrm{H}}(\phi_{j,i,l}, \theta_{j,i,l}), \tag{9.2}$$

where $\mathbf{h}_j \in \mathbb{C}^{N_t \times 1}$ such that $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_{N_u}]^{\mathrm{H}}$. For the lth path in the ith scattering cluster seen by the jth user, $\alpha_{j,i,l} \sim \mathcal{CN}(0,1)$ denotes the circularly symmetric complex Gaussian gain of the path (i.e., the small-scale fading component), $\phi_{j,i,l}$ and $\theta_{j,i,l}$ are respectively the azimuth and elevation angles of departure (AoD), and $\mathbf{a}(\phi_{j,i,l}, \theta_{j,i,l})$ represents the normalized transmit array response vector evaluated at specific azimuth and elevation angles $\phi_{j,i,l}$ and $\theta_{j,i,l}$. The array response vector further depends on the array geometry. For uniform linear array (ULA), where the antenna elements are linearly and equally spaced, the array response vector is independent of the elevation angles $\theta_{j,i,l}$

and follows the Vandermonde structure given by

$$\mathbf{a}(\phi_{j,i,l}) = \frac{1}{\sqrt{N_\mathrm{t}}} \left[1, e^{\mathrm{j}\frac{2\pi}{\lambda}d\sin(\phi_{j,i,l})}, ..., e^{\mathrm{j}(N_\mathrm{t}-1)\frac{2\pi}{\lambda}d\sin(\phi_{j,i,l})}\right]^\mathrm{T}, \tag{9.3}$$

where λ and d respectively denote the signal wavelength and the inter-element antenna spacing, and $\mathrm{j} = \sqrt{-1}$. Note that the elevation angles $\theta_{j,i,l}$ appear in the array response vector in case a uniform planar array (UPA) structure is employed [9]. The variance of the path gains $\alpha_{j,i,l}$ and the normalization constant $\sqrt{N_\mathrm{t}/(N_\mathrm{c}N_\mathrm{p})}$ are set such that $\mathbb{E}\left\{\|\mathbf{H}\|_\mathrm{F}^2\right\} = N_\mathrm{t}N_\mathrm{u}$.

9.3 Hybrid Symbol-Level Precoding Design

We start by designing the analog phase-shifting network. The matrix $\mathbf{F}$ representing the phase shifters' angles is usually considered to be solely dependent on the aggregate channel $\mathbf{H}$. Here, we adopt an analog design based on the singular value decomposition (SVD) of $\mathbf{H}$, which can be expressed as

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathrm{H}, \tag{9.4}$$

where $\mathbf{\Sigma}$ is an $N_\mathrm{u} \times N_\mathrm{t}$ rectangular diagonal matrix with the singular values on the diagonal in a descending order, and $\mathbf{U}$ and $\mathbf{V} \triangleq [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{N_\mathrm{t}}]$ are respectively $N_\mathrm{u} \times N_\mathrm{u}$ and $N_\mathrm{t} \times N_\mathrm{t}$ unitary matrices with the columns representing the left and the right singular vectors. We align the angles of the phase-shifting network to those of the first N_l right singular vectors of $\mathbf{H}$, i.e., $\{\mathbf{v}_1, ..., \mathbf{v}_{N_\mathrm{l}}\}$, with an element-wise normalization due to the constant modulus constraint of the phase shifters. Accordingly, we set

$$f_{k,j} = \frac{1}{\sqrt{N_\mathrm{t}}} e^{\mathrm{j}\varphi_{k,j}}, \quad k = 1, 2, ..., N_\mathrm{t}, \quad j = 1, 2, ..., N_\mathrm{l}, \tag{9.5}$$

where $\varphi_{k,j}$ denotes the phase of the kth element in $\mathbf{v}_j$. Aligning the angles of the phase-shifting network according to the first N_l right singular vectors of $\mathbf{H}$ enables the system to achieve larger array gains. Note that similar aligning schemes based on the SVD decomposition of the channel are used in, e.g., [81, 234]. Although infinite-resolution phase shifters are required for an accurate implementation of this approach, in practice, the use of finite-resolution phase shifters is preferred due to practical constraints of variable phase shifters, particularly in systems with large-scale antenna arrays as the number of phase shifters is proportional to the number of antenna elements. Therefore, in a more realistic implementation with discrete phase shifters, the phase states are quantized up to (typically) low bits of precision. We assume a quantization rule such that the phase of each entry of $\mathbf{F}$ is mapped to the nearest phase value in the discrete set $\{2m\pi/2^{b_\mathrm{PS}} : m = 0, 1, ..., 2^{b_\mathrm{PS}} - 1\}$. Accordingly, the quantized phase of $f_{k,j}$, denoted by $\hat{\varphi}_{k,j}$, can be obtained as

$$\hat{\varphi}_{k,j} = \frac{2\hat{m}\pi}{2^{b_{\mathrm{PS}}}}, \quad \hat{m} = \underset{m\in\{0,1,\ldots,2^{b_{\mathrm{PS}}}-1\}}{\operatorname{argmin}} \left|\varphi_{k,j} - \frac{2m\pi}{2^{b_{\mathrm{PS}}}}\right|, \tag{9.6}$$

with b_{PS} denoting the number of phase shifter's resolution bits. Although our design process is independent of the precision of the entries of $\mathbf{F}$, we investigate in Section 9.5 the performance of the proposed hybrid precoding scheme for both finite-resolution and discrete phase shifters.

Accordingly, for a given symbol vector $\mathbf{s}$, the channel matrix $\mathbf{H}$ and the phase-shifting network matrix $\mathbf{F}$, our design objective is to jointly and instantaneously (i.e., on a symbol-level basis) optimize the digitally precoded signal $\mathbf{u}$ as well as the states of the switching network, represented by $\mathbf{T}$. In our design, we utilize a baseline fully-digital precoding scheme and aim to optimize the hybrid precoder such that it performs as close as possible to the baseline scheme. To this end, we first overview the SLP design for a fully-digital architecture, which will be used as the baseline scheme in our subsequent derivation of the proposed hybrid precoding approach.

Let consider a fully-digital transmitter architecture where a dedicated RF chain drives each antenna element, i.e., $N_{\mathrm{u}} \leq N_{\mathrm{l}} = N_{\mathrm{t}}$. We assume that the symbol-level precoder is designed via a power-constrained optimization problem with a max-min fair objective subject to user-specific CI constraints. In such a scenario, we have shown in Section 3.5 that the SLP design formulation can be expressed in a convex form as

$$\begin{aligned} \max_{\bar{\mathbf{u}}_{\mathrm{FD}},\mathbf{d}\succeq\mathbf{0}} \quad & \min(\mathbf{d}) \\ \text{s.t.} \quad & \sqrt{\rho}\,\bar{\mathbf{H}}\bar{\mathbf{u}}_{\mathrm{FD}} = \bar{\mathbf{\Sigma}}\bar{\mathbf{s}} + \mathbf{A}^{-1}\mathbf{W}\mathbf{d}, \\ & \bar{\mathbf{u}}_{\mathrm{FD}}^{\mathrm{T}}\bar{\mathbf{u}}_{\mathrm{FD}} \leq 1, \end{aligned} \tag{9.7}$$

which can efficiently be solved via several off-the-shelf algorithms [121]. The optimal solution to (9.7) is, in fact, a performance upper bound that can be achieved by the symbol-level precoder when the number of BS's RF chains is equal to N_{t}. We use this optimal yet impractical solution in developing our hybrid SLP algorithm and also as a performance benchmark for comparisons in Section 9.5.

9.3.1 Hybrid Precoder with Phase Shifter Selection

We use the optimal fully-digital precoded signal to design the hybrid symbol-level precoder. More specifically, denoting by $\mathbf{u}_{\mathrm{FD}}^{\star}$ the optimal solution to (9.7), we aim to find the digital-domain precoded signal $\mathbf{u}_{\mathrm{BB}}$ and the selection matrix $\mathbf{T}$ such that the output of the hybrid precoder, i.e., $(\mathbf{F}\circ\mathbf{T})\mathbf{u}_{\mathrm{BB}}$ has a minimum Euclidean distance from $\mathbf{u}_{\mathrm{FD}}^{\star}$. The corresponding optimization problem is therefore can be written as

$$\min_{\mathbf{u}_{\mathrm{BB}},\mathbf{T}\in\mathbb{B}} \quad \left\|(\mathbf{F}\circ\mathbf{T})\mathbf{u}_{\mathrm{BB}} - \mathbf{u}_{\mathrm{FD}}^{\star}\right\|^2 \quad \text{s.t.} \quad \|(\mathbf{F}\circ\mathbf{T})\mathbf{u}_{\mathrm{BB}}\|^2 = 1. \tag{9.8}$$

To proceed, by defining $\mathbf{G} \triangleq 2\mathbf{T} - \mathbf{1}$ and $\mathbf{g} \triangleq \text{vec}(\mathbf{G})$, we recast (9.8) in an equivalent form which is more convenient for our later use, i.e.,

$$\begin{aligned}\min_{\mathbf{u}_{\text{BB}},\mathbf{g}} \quad & \left\|(\mathbf{u}_{\text{BB}}^{\text{T}} \otimes \mathbf{I}_{N_{\text{t}}})\,\text{diag}(\text{vec}(\mathbf{F}))\,\mathbf{g} + \mathbf{F}\mathbf{u}_{\text{BB}} - 2\,\mathbf{u}_{\text{FD}}^{\star}\right\|^2 \\ \text{s.t.} \quad & \frac{1}{4}\|(\mathbf{u}_{\text{BB}}^{\text{T}} \otimes \mathbf{I}_{N_{\text{t}}})\,\text{diag}(\text{vec}(\mathbf{F}))\,\mathbf{g} + \mathbf{F}\mathbf{u}_{\text{BB}}\|^2 = 1, \\ & \mathbf{g} \in \left\{\{-1,+1\}^{N_{\text{t}}N_{\text{l}}\times 1} \cap \bar{\mathbb{B}}\right\},\end{aligned} \tag{9.9}$$

where

$$\bar{\mathbb{B}} \triangleq \left\{\mathbf{y} \in \mathbb{R}^{N_{\text{t}}N_{\text{l}}\times 1} \,|\, \frac{1}{2}\left(\mathbf{1}_{N_{\text{l}}\times 1} \otimes \mathbf{I}_{N_{\text{t}}}\right)^{\text{T}}(\mathbf{y}+\mathbf{1}) \succeq \mathbf{1},\, \frac{1}{2}\left(\mathbf{1}_{N_{\text{t}}\times 1} \otimes \mathbf{I}_{N_{\text{l}}}\right)^{\text{T}}(\mathbf{y}+\mathbf{1}) \succeq \mathbf{1}\right\}.$$

The new formulation (9.9) is derived using the well-known property $\text{vec}(\mathbf{XYZ}) = (\mathbf{Z}^{\text{T}} \otimes \mathbf{X})\text{vec}(\mathbf{Y})$ for given matrices $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, along with the fact that $\mathbf{T} = (\mathbf{G} + \mathbf{1})/2$. The optimization problem (9.9) belongs to the class of minimization of quadratic forms over binary vectors (i.e., the binary constraints on the elements of $\mathbf{g}$), which is known to be NP-hard in general [211]. To tackle this difficulty, we use an equivalent biconvex implication of the binary constraints. According to Lemma 17, let $\mathbf{e}$ be a real-valued slack vector of length $N_{\text{t}}N_{\text{l}}$. Given $-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}$ and $\mathbf{e}^{\text{T}}\mathbf{e} \leq N_{\text{t}}N_{\text{l}}$, the constraint $\mathbf{g}^{\text{T}}\mathbf{e} = N_{\text{t}}N_{\text{l}}$ implies that $\mathbf{g} \in \{-1,+1\}^{N_{\text{t}}N_{\text{l}}}$. Therefore, we can rewrite problem (9.9) in an equivalent form where all the optimization variables are taken from continuous domains, i.e.,

$$\begin{aligned}\min_{\mathbf{u}_{\text{BB}},-\mathbf{1}\preceq\mathbf{g}\preceq\mathbf{1},\mathbf{e}} \quad & \left\|(\mathbf{u}_{\text{BB}}^{\text{T}} \otimes \mathbf{I}_{N_{\text{t}}})\text{diag}(\text{vec}(\mathbf{F}))\,\mathbf{g} + \mathbf{F}\mathbf{u}_{\text{BB}} - 2\,\mathbf{u}_{\text{FD}}^{\star}\right\|^2 \\ \text{s.t.} \quad & \frac{1}{4}\|(\mathbf{u}_{\text{BB}}^{\text{T}} \otimes \mathbf{I}_{N_{\text{t}}})\,\text{diag}(\text{vec}(\mathbf{F}))\,\mathbf{g} + \mathbf{F}\mathbf{u}_{\text{BB}}\|^2 = 1, \\ & \mathbf{g}^{\text{T}}\mathbf{e} = N_{\text{t}}N_{\text{l}}, \quad \mathbf{e}^{\text{T}}\mathbf{e} \leq N_{\text{t}}N_{\text{l}}, \quad \mathbf{g} \in \bar{\mathbb{B}},\end{aligned} \tag{9.10}$$

where $\mathbf{g}^{\text{T}}\mathbf{e} = N_{\text{t}}N_{\text{l}}$ is often called the equilibrium constraint. Reformulation (9.10) is still a non-convex problem due to the biconvex equilibrium constraint. We use a well known approach, namely, the exact penalty method, to efficiently solve (9.10). The interested readers are referred to [211] and [212] where studies on the accuracy and convergence characteristics of the exact penalty method are provided.

Based on the exact penalty method, the biconvex equilibrium constraint $\mathbf{g}^{\text{T}}\mathbf{e} = N_{\text{t}}N_{\text{l}}$ can be handled by adding the difference $N_{\text{t}}N_{\text{l}} - \mathbf{g}^{\text{T}}\mathbf{e}$ multiplied by $\mu > 0$ as a penalty function to the objective function, where the difference $N_{\text{t}}N_{\text{l}} - \mathbf{g}^{\text{T}}\mathbf{e}$ acts as a measure of deviation from the equilibrium constraint. Accordingly, denoting the objective function of (9.10) by $g(\mathbf{u}_{\mathrm{BB}}, \mathbf{t})$, we can write

$$\begin{aligned} \min_{\mathbf{u}_{\mathrm{BB}}, -\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}, \mathbf{e}} \quad & g(\mathbf{u}_{\mathrm{BB}}, \mathbf{g}) + \mu \left(N_{\mathrm{t}} N_{\mathrm{l}} - \mathbf{g}^{\mathrm{T}} \mathbf{e} \right) \\ \text{s.t.} \quad & \frac{1}{4} \left\| (\mathbf{u}_{\mathrm{BB}}^{\mathrm{T}} \otimes \mathbf{I}_{N_{\mathrm{t}}}) \, \mathrm{diag}(\mathrm{vec}(\mathbf{F})) \, \mathbf{g} + \mathbf{F} \mathbf{u}_{\mathrm{BB}} \right\|^2 = 1, \\ & \mathbf{e}^{\mathrm{T}} \mathbf{e} \leq N_{\mathrm{t}} N_{\mathrm{l}}, \quad \mathbf{g} \in \bar{\mathbb{B}}, \end{aligned} \tag{9.11}$$

which is our final formulation for the proposed hybrid SLP design. It is important to note that, in general, problems (9.11) and (9.10) are not equivalent. However, by monotonically increasing the penalty parameter μ in each iteration up to a certain threshold, successive solutions of the penalized problem (9.10) eventually converge to the solution of the original biconvex problem. On the other hand, for a given $\mathbf{u}_{\mathrm{BB}}$, it can be shown that if $g(\mathbf{u}_{\mathrm{BB}}, \mathbf{g})$ is an L-Lipschitz continuous convex function on $-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}$, problem (9.11) has the same local and global minima as those of (9.10) for $\mu \geq 2L$, where L denotes the Lipschitz constant of $g(\mathbf{u}_{\mathrm{BB}}, \mathbf{g})$ with respect to $\mathbf{g}$; see [211, Th. 1]. In the following lemma, we show that function $g(\mathbf{u}_{\mathrm{BB}}, \mathbf{g})$ is Lipschitz continuous on the domain $-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}$.

Lemma 20. *Let $\mathbf{u}_{\mathrm{BB}}$ be given, then $g(\mathbf{u}_{\mathrm{BB}}, \mathbf{g})$ is a Lipschitz continuous function on $-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}$ with Lipschitz constant*

$$\begin{aligned} L = {} & 2\sqrt{N_{\mathrm{t}} N_{\mathrm{l}}} \left\| (\mathbf{u}_{\mathrm{BB}}^{\mathrm{T}} \otimes \mathbf{I}_{N_{\mathrm{t}}}) \, \mathrm{diag}(\mathrm{vec}(\mathbf{F})) \right\|_{\mathrm{F}}^2 \\ & + 2 \left\| \left((\mathbf{u}_{\mathrm{BB}}^{\mathrm{T}} \otimes \mathbf{I}_{N_{\mathrm{t}}}) \, \mathrm{diag}(\mathrm{vec}(\mathbf{F})) \right)^{\mathrm{H}} \left(\mathbf{F} \mathbf{u}_{\mathrm{BB}} - 2\, \mathbf{u}_{\mathrm{FD}}^{\star} \right) \right\|. \end{aligned} \tag{9.12}$$

Proof. See Appendix F.1. □

Finally, we exploit the fact that the objective function of the minimization problem (9.11), i.e., $g(\mathbf{u}_{\mathrm{BB}}, \mathbf{g}) + \mu(N_{\mathrm{t}} N_{\mathrm{l}} - \mathbf{g}^{\mathrm{T}} \mathbf{e})$ is a biconvex quadratic function in $\mathbf{g}$ and $\mathbf{e}$, i.e., fixing either $\mathbf{g}$ or $\mathbf{e}$ gives a convex function in the other variable. As a result, we can use a standard block coordinate descent (BCD) algorithm to find at least a locally optimal solution to problem (9.10), where a coordinate block refers to either of the vectors $\mathbf{u}_{\mathrm{BB}}$, $\mathbf{g}$ or $\mathbf{e}$. To be more specific, the objective function $g(\mathbf{u}_{\mathrm{BB}}, \mathbf{g}) + \mu(N_{\mathrm{t}} N_{\mathrm{l}} - \mathbf{g}^{\mathrm{T}} \mathbf{e})$ can be minimized over one of these vectors while the other two are fixed, and then, repeating the same procedure for the other two blocks. The penalty multiplier μ should be increased monotonically every N cycles, where the Lipschitz constant L provided in Lemma 20 determines the limit for increasing μ as a function of the other variables. Accordingly, the BCD algorithm solving (9.11) performs the following steps within the nth iteration:

i. **Updating g:** Given $\mathbf{u}_{\mathrm{BB}}^{(n-1)}$ and $\mathbf{e}^{(n-1)}$, the value of $\mathbf{g}$ in the nth iteration is updated by solving the following LCQP:

$$\mathbf{g}^{(n)} = \underset{-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}, \mathbf{g} \in \bar{\mathbb{B}}}{\mathrm{argmin}} \; g\left(\mathbf{u}_{\mathrm{BB}}^{(n-1)}, \mathbf{g} \right) + \mu \left(N_{\mathrm{t}} N_{\mathrm{l}} - \mathbf{g}^{\mathrm{T}} \mathbf{e}^{(n-1)} \right), \tag{9.13}$$

ii. **Updating $\mathbf{e}$:** The value of $\mathbf{e}^{(n)}$ can be obtained as the solution to the following problem:

$$\mathbf{e}^{(n)} = \underset{\|\mathbf{e}\|^2 \le N_\mathrm{t} N_\mathrm{l}}{\operatorname{argmin}} \quad -\mathbf{e}^\mathrm{T} \mathbf{g}^{(n)}, \tag{9.14}$$

which is equivalent to a norm-constrained inner product maximization that admits a simple closed-form solution given by

$$\mathbf{e}^{(n)} = \frac{\sqrt{N_\mathrm{t} N_\mathrm{l}}}{\|\mathbf{g}^{(n)}\|}\, \mathbf{g}^{(n)}. \tag{9.15}$$

iii. **Updating $\mathbf{u}_{\mathrm{BB}}$:** Given $\mathbf{g}^{(n)}$ and $\mathbf{e}^{(n)}$, the minimization problem (9.11) is equivalent to

$$\min_{\mathbf{u}_{\mathrm{BB}}} \quad g\left(\mathbf{u}_{\mathrm{BB}}, \mathbf{g}^{(n)}\right) \quad \text{s.t.} \quad \frac{1}{4} \left\| (\mathbf{u}_{\mathrm{BB}}^\mathrm{T} \otimes \mathbf{I}_{N_\mathrm{t}})\, \mathrm{diag}(\mathrm{vec}(\mathbf{F}))\, \mathbf{g}^{(n)} + \mathbf{F}\mathbf{u}_{\mathrm{BB}} \right\|^2 = 1. \tag{9.16}$$

Using the method of Lagrange multipliers, it is straightforward to obtain the solution to (9.16) which is used as the nth update of $\mathbf{u}_{\mathrm{BB}}$ and is given by

$$\mathbf{u}_{\mathrm{BB}}^{(n)} = \frac{2\left(\left(\mathbf{F} \circ \mathbf{G}^{(n)}\right) + \mathbf{F}\right)^\dagger \mathbf{u}_{\mathrm{FD}}^\star}{\left\| \left((\mathbf{F} \circ \mathbf{G}^{(n)}) + \mathbf{F}\right) \left((\mathbf{F} \circ \mathbf{G}^{(n)}) + \mathbf{F}\right)^\dagger \mathbf{u}_{\mathrm{FD}}^\star \right\|}, \tag{9.17}$$

where $\mathrm{vec}\left(\mathbf{G}^{(n)}\right) = \mathbf{g}^{(n)}$, and $(\cdot)^\dagger$ stands for the Moore-Penrose inverse.

iv. **Updating λ:** In every N cycles, the penalty parameter μ is updated as

$$\mu^{(n)} = \min\{2L^{(n)}, \vartheta\, \mu^{(n-1)}\}, \tag{9.18}$$

where $\vartheta > 1$ is a constant design parameter and $L^{(n)}$ is the nth update of the Lipschitz constant L which is computed by substituting $\mathbf{u}_{\mathrm{BB}}^{(n)}$ in (9.12).

The pseudocode of the described BCD algorithm is presented in Algorithm 6. In what follows, we analyze the convergence behavior of this algorithm.

9.3.2 Convergence Analysis

The BCD algorithm is a successive optimization approach in which a certain approximate version of the objective function is optimized with respect to one block of variables at a time, while fixing the rest of block variables [130]. Let $h(\mathbf{u}_{\mathrm{BB}}, \mathbf{g}, \mathbf{e})$ denote the objective function of problem (9.11). As mentioned earlier in this section, by fixing two variables among $\mathbf{u}_{\mathrm{BB}}$, $\mathbf{g}$ and $\mathbf{e}$, function $h(\mathbf{u}_{\mathrm{BB}}, \mathbf{g}, \mathbf{e})$ becomes convex in the other variable. More precisely, the sub-problem (9.13) is a convex LCQP which can be solved for the optimal solution. In addition, the two sub-problems (9.14) and (9.16) are amenable to closed-form solutions, and therefore, can be solved for global optimality. This implies that, at

Algorithm 6 BCD algorithm solving (9.11)

```
input: $\mathbf{F}, \mathbf{u}_{\text{FD}}^{\star}$
output: $\mathbf{u}_{\text{BB}}, \mathbf{g}$
initialize: $\mathbf{g}^{(0)} = \mathbf{e}^{(0)} \in \mathbb{R}^{N_{\text{t}}N_{\text{l}}\times 1}, \mathbf{u}_{\text{BB}}^{(0)} \in \mathbb{R}^{N_{\text{l}}\times 1}, \mu^{(0)}, n = 0$
set: $\vartheta > 1$
while the terminating condition is met do
    $n \leftarrow n + 1$
    compute $\mathbf{g}^{(n)}$ by solving (9.13)
    $\mathbf{e}^{(n)} \leftarrow \sqrt{N_{\text{t}}N_{\text{l}}}\, \mathbf{g}^{(n)}/\|\mathbf{g}^{(n)}\|$
    compute $\mathbf{u}_{\text{BB}}^{(n)}$ using (9.17)
    obtain $L^{(n)}$ from (9.12)
    $\mu^{(n)} \leftarrow \min\{2L^{(n)}, \vartheta\, \mu^{(n-1)}\}$
end while
```

the nth iteration, we have

$$h\left(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n-1)}\right) \leq h\left(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n-1)}, \mathbf{e}^{(n-1)}\right), \tag{9.19}$$

where $\mathbf{g}^{(n)}$ denotes the nth update of $\mathbf{g}$, and $\mathbf{u}_{\text{BB}}^{(n-1)}$ and $\mathbf{e}^{(n-1)}$ denote the updates of $\mathbf{u}_{\text{BB}}$ and $\mathbf{e}$ obtained form iteration $n-1$. Similarly, we can write

$$h\left(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n)}\right) \leq h\left(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n-1)}\right), \tag{9.20}$$

and

$$h\left(\mathbf{u}_{\text{BB}}^{(n)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n)}\right) \leq h\left(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n)}\right). \tag{9.21}$$

As a result, the sequence of the objective function values after the update of each block is monotonically non-increasing, and therefore, convergence of Algorithm 6 to a stationary point (i.e., at least a local extremum) is guaranteed. We further note that the terminating condition for Algorithm 6 can be considered as

$$\left|h\left(\mathbf{u}_{\text{BB}}^{(n)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n)}\right) - h\left(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n-1)}, \mathbf{e}^{(n-1)}\right)\right| \leq \epsilon_{\text{o}}, \tag{9.22}$$

where ϵ_{o} denotes the threshold for the desired accuracy. In Fig. 9.2, we illustrate the convergence behavior of Algorithm 6 by plotting the value of the objective function $h(\mathbf{u}_{\text{BB}}, \mathbf{g}, \mathbf{e})$ versus the number of outer iterations (cycles) for phase shifters with different precision bits b_{PS}, where it is shown that the proposed algorithm converges at a favorable rate. In particular, for a desired accuracy of $\epsilon_{\text{o}} = 10^{-2}$, it can be seen that, in all cases, Algorithm 6 converges in no more than 10 iterations. It can further be seen that the algorithm shows a higher residual error for lower values of b_{PS}. This is due to the fact that discretizing the states of the phase shifters with lower number of precision bits induces

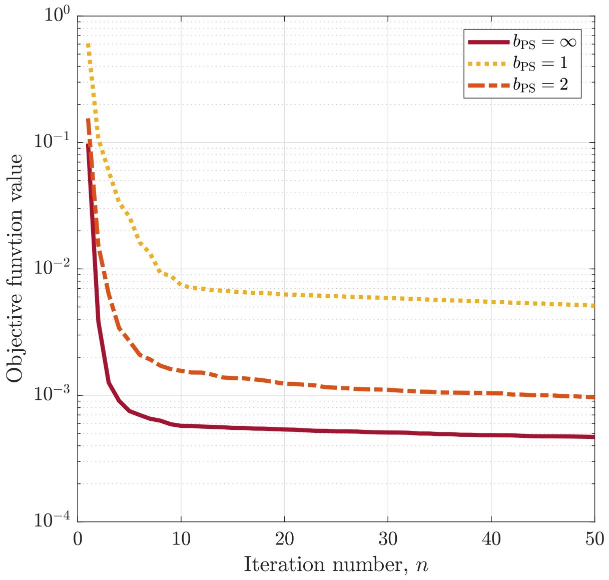


FIGURE 9.2: Convergence behavior of Algorithm 6 versus iteration number for $\mu = 10^{-4}$ and $\vartheta = 1.1$.

a greater discontinuity in the feasible region of the optimization problem, and therefore, it may not be possible to reduce the Euclidean distance between the fully-digital and the hybrid precoders beyond a certain limit.

9.3.3 Analysis of Computational Complexity

Using the four-step BCD approach summarized in Algorithm 6, the overall computation cost of solving (9.11) in terms of the required number of arithmetic operations is composed of two main parts. The first part involves inner iterations to solve the sub-problem (9.13) over $\mathbf{g}$ and updating $\mathbf{u}_{\mathrm{BB}}$ using (9.17), and the second part refers to the outer iterations (cycles) over coordinate blocks.

The computation cost of updating $\mathbf{u}_{\mathrm{BB}}$ via (9.17) is dominated by the arithmetic complexity of performing the matrix pseudo-inversion $((\mathbf{F} \circ \mathbf{G}) + \mathbf{F})^{\dagger}$, which is of order $\mathcal{O}(N_{\mathrm{t}} N_{\mathrm{l}}^{2})$ given the dimensions of $\mathbf{F}$ and $\mathbf{G}$. Furthermore, to efficiently solve (9.13), one may use the off-the-shelf algorithms such as (accelerated) projected/proximal gradient methods [135], or quasi-Newton approaches, e.g., L-BFGS-B [213]. In particular, for a Lipschitz smooth (not necessarily strongly) convex objective function as in (9.11), all the aforementioned algorithms converge superlinearly at a rate of $\mathcal{O}(1/\sqrt{\epsilon_{\mathrm{i}}})$ to reach an ϵ_{i}-optimal solution. For example, using the accelerated projected gradient descent algorithm, the per-iteration complexity associated with sub-problem (9.13) is dominated by matrix multiplications of the limiting order $N_{\mathrm{l}}^{2} N_{\mathrm{t}}^{2}$, as $N_{\mathrm{l}}, N_{\mathrm{t}} \to \infty$. Therefore, in the limiting case, the total number of operations needed to be performed in order to solve the inner sub-problem (9.13) with an accuracy of ϵ_i is of order $\mathcal{O}(N_\mathrm{l}^2 N_\mathrm{t}^2)(1/\sqrt{\epsilon_\mathrm{i}})$. Putting this together with the complexity of computing $\mathbf{u}_{\mathrm{BB}}$, every cycle of the BCD algorithm has a dominating complexity of $\mathcal{O}(N_\mathrm{t} N_\mathrm{l}^2) + \mathcal{O}(N_\mathrm{l}^2 N_\mathrm{t}^2)(1/\sqrt{\epsilon_\mathrm{i}})$.

On the other hand, the reformulation (9.11), which is obtained based on the exact penalty method, is guaranteed to converge to a first-order Karush-Kuhn-Tucker (KKT) point with an accuracy of ϵ_o in no more than $\lceil \left(\ln(2L\sqrt{N_\mathrm{t}}) - \ln(\mu^{(0)}\epsilon_\mathrm{o})\right) / \ln(\vartheta) \rceil$ iterations [211], where $\mu^{(0)}$ is the initial value of the penalty parameter μ and $\lceil \cdot \rceil$ denotes the ceiling operation. To have a complete analysis of the complexity, we further need to evaluate the constant L. From (9.12), it is straightforward to show that

$$\begin{aligned} L &\leq 2\sqrt{N_\mathrm{l} N_\mathrm{t}} \|\mathbf{u}_{\mathrm{BB}}\|^2 \|\mathbf{F}\|_\mathrm{F}^2 + 2\,\|\mathbf{u}_{\mathrm{BB}}\|^2 \|\mathbf{F}\|_\mathrm{F}^2 + 4\,\|\mathbf{u}_{\mathrm{BB}}\| \|\mathbf{F}\|_\mathrm{F} \|\mathbf{u}_{\mathrm{FD}}^{\star}\|^2 \\ &= 2N_\mathrm{l}\sqrt{N_\mathrm{t} N_\mathrm{l}}\, \|\mathbf{u}_{\mathrm{BB}}\|^2 + 2N_\mathrm{l} \|\mathbf{u}_{\mathrm{BB}}\|^2 + 4\sqrt{N_\mathrm{l}} \|\mathbf{u}_{\mathrm{BB}}\|, \end{aligned} \tag{9.23}$$

where the equality can be justified considering the definition of matrix $\mathbf{F}$ in (9.5), which yields $\|\mathbf{F}\|_\mathrm{F} = \sqrt{N_\mathrm{l}}$, along with the fact that $\|\mathbf{u}_{\mathrm{FD}}^{\star}\| = 1$; see (9.7). It can further be verified that in the large system limit where $N_\mathrm{t} \to \infty$ with $N_\mathrm{l} \ll N_\mathrm{t}$, we have $\|\mathbf{u}_{\mathrm{BB}}\| \to 1$. Therefore, one can write

$$L \leq 2N_\mathrm{l}\sqrt{N_\mathrm{t} N_\mathrm{l}} + \mathcal{O}\left(N_\mathrm{l} + \sqrt{N_\mathrm{l}}\right), \tag{9.24}$$

As a result, in the limiting case with $N_\mathrm{l} \to \infty$, we have

$$L \leq N_\mathrm{l}\,\sqrt{N_\mathrm{t} N_\mathrm{l}}\,.\,\mathcal{O}(1). \tag{9.25}$$

With the upper bound, given in (9.25), on the dominating order of L, the worst-case computational complexity of Algorithm 6 solving the design problem (9.11) with accelerated inner gradient steps can be obtained as shown in Table 9.1. In practice, however, the outer optimization usually converges in a few cycles, as we will see in Section 9.5.

For comparison purposes, the complexities of hybrid symbol-level precoding approaches proposed in [81] and [83] are reported in Table 9.1. For the hybrid scheme in [81], the reported complexity order refers to the worst-case complexity of reaching an ϵ_o-optimal solution to a linear program via the interior-point method; see [171].

9.4 Energy Efficiency Analysis

Hybrid precoding strategies predominantly focus on reducing hardware cost/complexity and power consumption by delegating part of the signal processing burden to the analog domain. In return, this may sacrifice the precoding performance, e.g., spectral efficiency, with respect to fully-digital systems. On the other hand, various hybrid implementations may differ from one another in their complexity and power consumption. In order to be able to compare different hybrid architectures and also to assess their efficiency versus the fully-digital alternative, one needs to incorporate both performance and complexity/power consumption aspects into one single figure of merit. A common choice

Table 9.1: Complexity comparison of different hybrid SLP schemes.

	Worst-case complexity
[81]	$(4N_\mathrm{l} + 2N_\mathrm{u} + 1)^{3/2}(2N_\mathrm{l} + 1)^2(1/\sqrt{\epsilon_\mathrm{o}})\,.\,\mathcal{O}(1)$
[83]	$\sqrt{N_\mathrm{t}N_\mathrm{u}}(N_\mathrm{t}^3N_\mathrm{u}^2 + N_\mathrm{t}^2N_\mathrm{u}^3) + N_\mathrm{t}(N_\mathrm{l} + 3)(1/\sqrt{\epsilon_\mathrm{o}})\,.\,\mathcal{O}(1)$
Algorithm 6	$N_\mathrm{t}N_\mathrm{l}^2\,(1 + N_\mathrm{t}(1/\sqrt{\epsilon_\mathrm{i}}))\left\lceil\Big(\ln\,\big(2\,N_\mathrm{t}N_\mathrm{l}\sqrt{N_\mathrm{l}}\big) - \ln(\mu^{(0)}\,\epsilon_\mathrm{o})\Big)\,/\,\ln(\vartheta)\right\rceil.\,\mathcal{O}(1)$
	Dominating order [as $N_\mathrm{t}, N_\mathrm{l}, N_\mathrm{u} \to \infty$]
[81]	$N_\mathrm{l}^{7/2}(1/\sqrt{\epsilon_\mathrm{o}})\,.\,\mathcal{O}(1)$
[83]	$\sqrt{N_\mathrm{t}N_\mathrm{u}}N_\mathrm{t}^3N_\mathrm{u}^2(1/\sqrt{\epsilon_\mathrm{o}})\,.\,\mathcal{O}(1)$
Algorithm 6	$N_\mathrm{l}^2N_\mathrm{t}^2\ln\,\big(N_\mathrm{t}N_\mathrm{l}\sqrt{N_\mathrm{l}}/\epsilon_\mathrm{o}\big)\,(1/\sqrt{\epsilon_\mathrm{i}})\,.\,\mathcal{O}(1)$

is energy efficiency which can simply be expressed as the ratio between spectral efficiency and power consumption. Due to the assumption of finite-alphabet signaling, we measure the spectral efficiency in bits per symbol. Thereby, the energy efficiency of the precoding scheme, in bits per Joule, is defined as the ratio between goodput and power consumption, i.e.,

$$\eta \triangleq \frac{R\,(1 - P_\mathrm{e})}{P}, \tag{9.26}$$

where $P_\mathrm{e} \triangleq 1 - (1/N_\mathrm{u})\sum_{i=1}^{N_\mathrm{u}} P_{\mathrm{e},i}$ is the average symbol error probability across all N_u users with $P_{\mathrm{e},i}$ denoting the symbol error probability for the ith user. The average per-user spectral efficiency R and the power consumption P are defined as follows.

Spectral Efficiency

Using an uncoded transmission scheme with finite-alphabet signaling, the communication rate towards the jth user can be evaluated, in terms of bits per symbol per unit bandwidth, through calculating the average mutual information between the target symbol s_i and the received signal y_i, i.e.,

$$I(s_i; y_i) = \mathbb{E}_{s_i,y_i,\mathbf{H}}\left\{\log_2\frac{\mathrm{Pr}_{y_i|s_i,\mathbf{H}}(y_i|s_i,\mathbf{H})}{\mathrm{Pr}_{y_i|\mathbf{H}}(y_i|\mathbf{H})}\right\}. \tag{9.27}$$

Assuming transmission with Nyquist rate over a double-sided bandwidth of W Hz, the maximum allowable symbol rate is W symbols per second, which results in a bit rate of $W \times I(s_i; y_i)$ for the user i. Putting this together for all N_u users, the average per-user achievable rate of the downlink channel is given by

$$R = \frac{W}{N_\mathrm{u}} \sum_{i=1}^{N_\mathrm{u}} I(s_i; y_i). \tag{9.28}$$

It should be noted that deriving closed-form expressions for the conditional probability mass functions in (9.27) is a cumbersome task. As an alternative, one can obtain experimental probability distributions over sufficiently many independent realizations of the channel and the users' symbols to approximate the mutual information $I(s_i; y_i)$ for each user $i \in \{1, 2, ..., N_\mathrm{u}\}$.

Power Consumption

The power dissipated by the BS's RF front-end components accounts for the power consumption at the BS. In the sequel, we first adopt power consumption models for typical components of an RF front-end and then specifically tailor the overall power consumption model according to each precoding architecture, namely, fully-digital and hybrid (with and without phase shifter selection). The transmit RF front-end of a multi-antenna system is commonly composed of one baseband processor, several RF chains, each preceded by a pair of DACs (i.e., one DAC for each I/Q channel), and power amplifiers (PA). The use of analog components such as dividers, combiners, switches, and/or phase shifters are limited to hybrid architectures.

As a rule of thumb, the power consumption of DAC scales linearly in sampling rate and exponentially in the number of bits per sample (i.e., resolution bits). We assume the DACs are of binary-weighted current-steering type [214], where its power consumption is approximately given in [215] as

$$P_\mathrm{DAC} = \frac{3}{2}\left(2^{b_\mathrm{DAC}} - 1\right) \times 10^{-5} + \frac{9}{2} b_\mathrm{DAC} F_\mathrm{s} \times 10^{-12}, \tag{9.29}$$

with b_DAC and F_s respectively denoting the number of precision bits and the sampling frequency.

A typical RF chain includes one mixer, one local oscillator, two low-pass filters and a baseband amplifier. We respectively denote by P_M, P_LO, P_LPF and P_BBA, the power dissipation of the RF chain components. Thereby, the power consumed by a single RF chain is equal to

$$P_\mathrm{RF} = P_\mathrm{M} + 2P_\mathrm{LO} + P_\mathrm{LPF} + P_\mathrm{BBA}. \tag{9.30}$$

In case all the RF streams are transmitted at the same frequency, it might be possible to share a single local oscillator among all the chains and divide the power consumption P_LO accordingly [203]. Further, let P_BB, P_PA, P_PS and P_SW respectively denote the power consumption of the baseband processor, a single PA, a single phase shifter and a single analog switch. Note also that, in general, the power dissipation of the RF combining network is very low [241], and thus is ignored in our modeling.

The fully-digital BS architecture requires $2N_\mathrm{t}$ DACs, and N_t RF chains and PAs, and therefore its power consumption can be modeled as

$$P_\mathrm{FD} = P_\mathrm{BB} + N_\mathrm{t}(2P_\mathrm{DAC} + P_\mathrm{RF} + P_\mathrm{PA}). \tag{9.31}$$

On the other hand, the hybrid architecture with fully-connected phase-shifting network can be implemented using $2N_\mathrm{l}$ DACs, N_l RF chains, N_t PAs, and $N_\mathrm{t} \times N_\mathrm{l}$ phase shifters. The resulting power dissipation is thus given by

$$P_\mathrm{H} = P_\mathrm{BB} + N_\mathrm{l}(2P_\mathrm{DAC} + P_\mathrm{RF}) + N_\mathrm{t}N_\mathrm{l}P_\mathrm{PS} + N_\mathrm{t}P_\mathrm{PA}. \tag{9.32}$$

To calculate the power consumption of the hybrid architecture with fully-connected networks of phase shifters and switches, i.e., with phase shifter selection, we assume the associated RF processes are turned off while a phase shifter is deactivated, and further, the phase shifter has negligible static power dissipation. Under this assumption, a deactivated phase shifter consumes no power. Denoting the average percentage of the active phase shifters at a symbol instant by β, the power consumed by the entire phase-shifting network is then $\beta N_\mathrm{t}N_\mathrm{l}P_\mathrm{PS}$. As illustrated in Fig. 9.1, the phase shifter selection mechanism is implemented through a network of $N_\mathrm{t} \times N_\mathrm{l}$ switches. Therefore, the power consumption of the hybrid precoder with phase shifter selection can be obtained as

$$P_\mathrm{HPSS} = P_\mathrm{BB} + N_\mathrm{l}(2P_\mathrm{DAC} + P_\mathrm{RF}) + N_\mathrm{t}N_\mathrm{l}(\beta P_\mathrm{PS} + P_\mathrm{SW}) + N_\mathrm{t}P_\mathrm{PA}. \tag{9.33}$$

Recall from Section 9.2 that the selection matrix $\mathbf{T}$ is constrained to has no all-zero row (column), i.e., at least one phase shifter corresponding to a specific antenna element (RF chain) must be active at a symbol instant. As a consequence, the number of active phase shifters during any symbol period is never less than $\max\{N_\mathrm{l}, N_\mathrm{t}\} = N_\mathrm{t}$, from which it follows that $1/N_\mathrm{l} < \beta \leq 1$. Our simulation results in Section 9.5 further indicate that β is usually smaller than 0.75 for the proposed hybrid symbol-level precoder in (9.11), regardless of the phase-shifting precision. This may lead to significant reductions in the power consumption of the analog phase-shifting network. It is also important to note that by employing low-power yet efficient mmWave switches, the excessive power consumption due to the switching operation can be made negligible compared with the power reduction of the phase shifters.

Using the above power consumption models with appropriate parameter selection, we will compare the power consumed by different fully-digital and hybrid architectures in Section 9.5.

9.5 Simulation Results

In this section, we present some simulation results to evaluate the performance of the proposed hybrid symbol-level precoding approach and to compare it with some other existing schemes. The simulation setup is as follows. We consider the hybrid analog-digital precoding architecture depicted in Fig. 9.1 for a downlink mmWave massive multiuser

MIMO system, performing an uncoded transmission with QPSK signaling and a carrier frequency of 60 GHz over a bandwidth of 1 GHz. We assume unit noise variances at the receivers of all the users, i.e., $\sigma_j^2 = 1, \forall j = 1, 2, ..., N_{\mathrm{u}}$. As described in Section 9.2, we adopt a geometric model for the mmWave propagation environment with $N_{\mathrm{c}} = 1$ clusters and $N_{\mathrm{p}} = 12$ scatterers between the BS and each user. For all the propagation paths, the azimuth angles of departure $\phi_{j,i,l}$ are drawn independently from a uniform distribution over $[0, 2\pi)$. To initialize Algorithm 6, we set $N = 1$ and $\vartheta = 1.1$ to avoid overshooting, and consider $\mu^{(0)} = 10^{-4}$ to have a reasonable starting point.

We consider the fully-digital Wiener filter (WF) precoding [5], and the optimal fully-digital symbol-level precoding (SLP) as our performance benchmarks, and further, provide comparisons with the block-level hybrid precoding technique PZF and its quantized variant QPZF in [227], the block-level hybrid precoding with phase shifter selection in [234], and the hybrid symbol-level precoders in [81] and [83]. Note that the application of PZF and QPZF techniques is limited only to fully-loaded systems, i.e., when $N_{\mathrm{u}} = N_{\mathrm{l}}$, and therefore, their performances have been evaluated only in the relevant scenarios. We further note that the method in [81] performs a symbol-based optimization of the digital baseband precoder subject to one-bit DACs, and adopts a CSI-only design for the phase-shifting network. On the other hand, the hybrid scheme in [83] jointly optimizes both the digital baseband precoder and the phase-shifting network on a symbol-level basis. Accordingly, we refer to the methods in [81], [83], and the proposed scheme in this work based on the adopted hybrid architecture and the precoder design approach. To summarize, throughout this section, the hybrid precoding techniques of interest are referred to as:

- Hybrid PZF: hybrid block-level precoding (BLP) based on ZF solution [227]
- Hybrid QPZF: quantized hybrid BLP based on ZF solution [227]
- Hybrid PSS BLP: hybrid BLP with phase shifter selection [234]
- Hybrid BB SLP: hybrid SLP with baseband precoder optimization [81]
- Hybrid BB+PS SLP: hybrid SLP with joint baseband precoder and phase-shifting network optimization [83]
- Hybrid BB+SW SLP: hybrid SLP with joint baseband precoder and switching network optimization (Algorithm 6)
- Hybrid BB+SW SLP-NOPSS: hybrid SLP with baseband precoder optimization and no phase shifter selection

In our simulations, the power consumption is calculated according to the model introduced in Section 9.4, in which we consider reference values of $P_{\mathrm{RF}} = 40$ mW, $P_{\mathrm{PA}} = 20$ mW, $P_{\mathrm{PS}} = 30$ mW, and $P_{\mathrm{BB}} = P_{\mathrm{DAC}}$, as in [203]. As for the power consumption of switches, it is well known that nFET switches have zero static power dissipation. On the other hand, silicon-germanium (SiGe) based switches are shown to be capable of

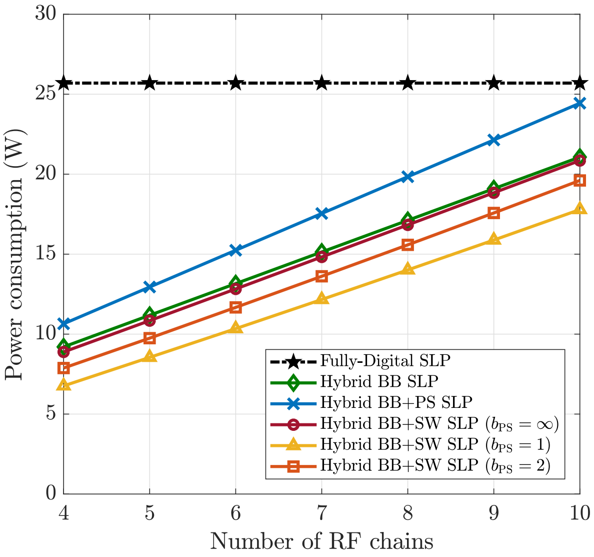


FIGURE 9.3: Power consumption of different hybrid SLP schemes as a function of N_{l} at SNR $= -5$ dB with $N_{\text{t}} = 64$ and $N_{\text{u}} = 4$.

achieving high performance while consuming powers of less than 1 mW [231]. Therefore, based on the available technology for the implementation of RF switches, a fairly conservative choice would be $P_{\text{SW}} = 1$ mW. Moreover, the power consumption of DACs is calculated via (9.29) assuming a sampling frequency of $F_{\text{s}} = 1$ GHz which should be sufficient for mmWave systems. We further assume $b_{\text{DAC}} = 12$ for those architectures employing high-resolution DACs.

In Fig. 9.3, we compare the power consumption of various hybrid precoding implementations with that of the fully-digital architecture as a function of the number of BS's RF chains N_{l}, while fixing the number of transmit antennas and users to be $N_{\text{t}} = 64$ and $N_{\text{u}} = 4$, respectively. As might be expected, the power consumption values associated with the hybrid implementations increase with N_{l}, which is a consequence of requiring more RF elements, phase shifters, and/or switches. This implies that increasing the number of RF chains beyond a certain limit makes the hybrid implementation a more power-consuming approach than the fully-digital architecture. Nevertheless, for $N_{\text{l}} \leq 10$, all the hybrid implementations consume less power than the fully-digital precoder. Remarkably, the proposed hybrid precoder in this chapter, i.e., the hybrid BB+SW SLP, offers smaller power consumption amounts, with either infinite-precision or discrete phase shifters, among the other hybrid symbol-level precoding schemes in Fig. 9.3. This is brought by the adopted phase shifter selection mechanism in implementing the hybrid precoder. In particular, the proposed hybrid precoder has the smallest power consumption with $b_{\text{PS}} = 1$ due to the large percentage of deactivated phase shifters, as we will see later in this section. Note also that the differences in power consumption of different hybrid precoding schemes in Fig. 9.3 increase with N_{l}.

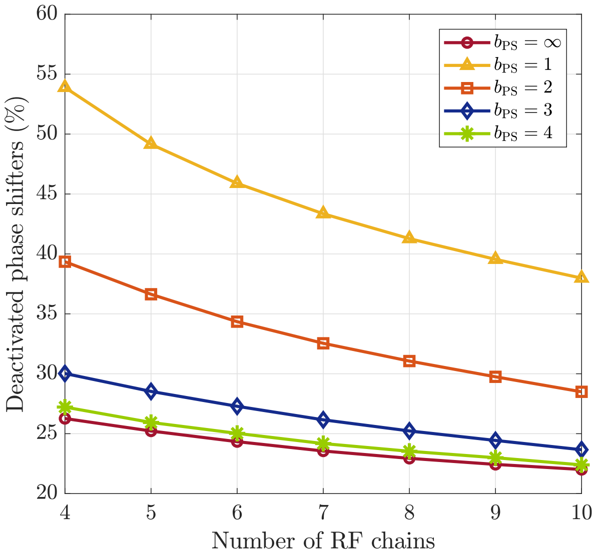


FIGURE 9.4: Average percentage of deactivated phase shifters as a function of N_l at SNR $= -5$ dB with $N_\text{t} = 64$ and $N_\text{u} = 4$.

In a scenario with $N_\text{t} = 64$ and $N_\text{u} = 4$, the percentages of deactivated phase shifters, i.e., $(1-\beta) \times 100$, for the proposed hybrid precoding approach are shown versus the number of RF chains in Fig. 9.4 for different values of b_{PS}. It follows from the results that in the case of using low-resolution phase shifters, a higher percentage of the phase shifters can be switched off, and hence more power-savings are possible. In particular, in the case where $N_\text{l} = 4$, up to 55% of the phase shifters with $b_{\text{PS}} = 1$ can be turned off, which can be roughly translated to a power reduction of $\beta N_\text{t} N_\text{l} P_{\text{PS}} \approx 4200$ mW in the phase-shifting network. It can further be observed from Fig. 9.4 that the percentage of deactivated phase shifters decreases with increasing N_l. One can justify these observations by considering that increasing the number of phase-shifting precision bits and the number of RF chains, respectively, reduces the discontinuity in the feasible region of the optimization problem (9.11) and increases the design degrees of freedom. In both cases, this enables the algorithm to achieve lower values for the objective function by activating a larger ratio of the phase shifters. The former case can also be verified from Fig. 9.2 where the residual error in the objective function is shown to be smaller for phase shifters with higher resolution bits.

We plot the average users' symbol error rate (SER) achieved by the precoding techniques of interest with either fully-digital or hybrid architecture versus the transmit SNR for a system with $(N_\text{t}, N_\text{l}, N_\text{u}) = (64, 8, 8)$ in Fig. 9.5. The proposed hybrid symbol-level precoder is evaluated for various implementations with infinite-precision and discrete phase shifters, where in the latter case we assume $b_{\text{PS}} = 1$ and $b_{\text{PS}} = 2$ bits of precision. It can be seen that, for the case with $b_{\text{PS}} = 2$, both the hybrid BB+PS SLP and the hybrid BB+SW SLP schemes are capable of performing well close to the fully-digital

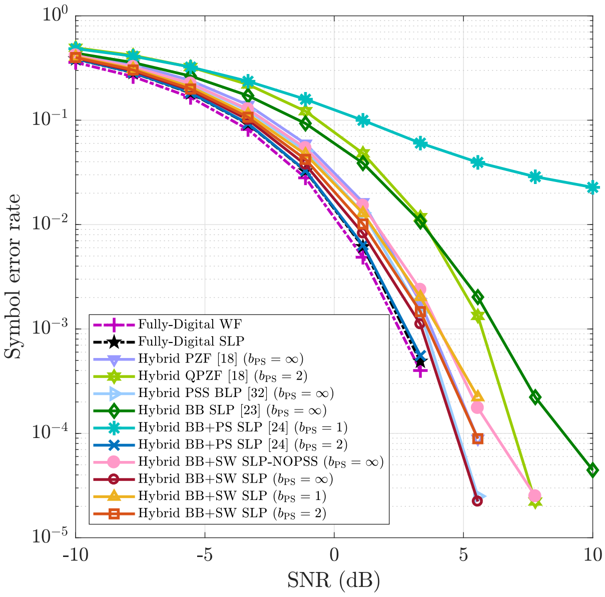


FIGURE 9.5: Average per-user symbol error rate versus transmit SNR with $(N_{\mathrm{t}}, N_{\mathrm{l}}, N_{\mathrm{u}}) = (64, 8, 8)$.

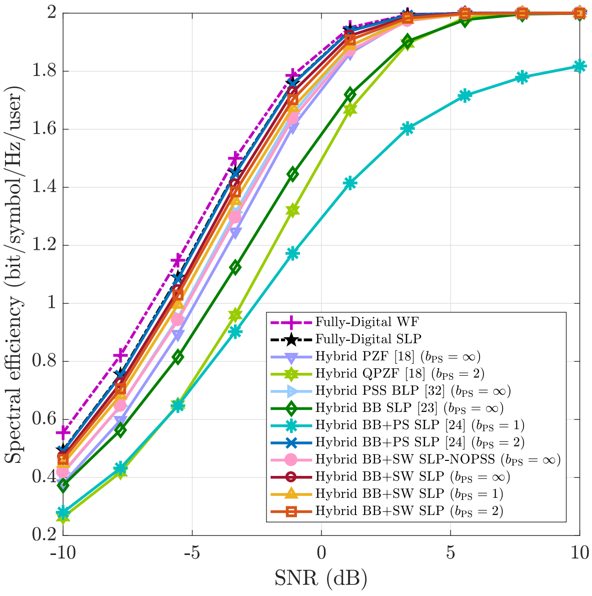


FIGURE 9.6: Per-user spectral efficiency versus transmit SNR with $(N_{\mathrm{t}}, N_{\mathrm{l}}, N_{\mathrm{u}}) = (64, 8, 8)$.

SLP, though requiring far less RF chains to process the transmitted signal. The corresponding losses at SER $= 10^{-2}$ are respectively around 0.1 dB and 0.5 dB. Using phase shifters with $b_{\mathrm{PS}} = 1$, the hybrid BB+SW SLP scheme still offers a reasonable performance with a loss smaller than 1 dB at SER $= 10^{-2}$ compared with the fully-digital SLP, as opposed to the hybrid BB+PS SLP scheme which shows a significantly deteriorated performance. It can further be seen from Fig. 9.5 that both hybrid BB+PS SLP and hybrid BB+SW SLP approaches outperform the PZF technique, which is a result of designing the precoded signal specifically for each instantaneous combination of the users' target symbols. Overall, from Fig. 9.5, it follows that the hybrid BB+PS SLP scheme offers the best SER performance compared to the other hybrid SLP schemes of interest. Nevertheless, as demonstrated in Fig. 9.3, this superior performance comes with increased power consumption.

In Fig. 9.6, the average per-user spectral efficiencies of the precoding schemes of interest are shown for the system parameter sets $(N_{\mathrm{t}}, N_{\mathrm{l}}, N_{\mathrm{u}}) = (64, 8, 8)$. As can be seen, the spectral efficiency plot follows the same relative trend as that of the SER plot. The hybrid BB+PS and BB+SW SLP schemes are more spectrally-efficient than the PZF and QPZF techniques, which is a result of the CI-based positioning of the received signals. Remarkably, the achievable spectral efficiencies by the hybrid BB+PS SLP scheme with $b_{\mathrm{PS}} = 2$ and by the hybrid BB+SW SLP scheme with either $b_{\mathrm{PS}} = 1$, $b_{\mathrm{PS}} = 2$ or $b_{\mathrm{PS}} = \infty$ are close to those of the fully-digital WF and SLP. The maximum loss with respect to the fully-digital SLP corresponds to the Hybrid BB+SW SLP scheme with $b_{\mathrm{PS}} = 1$, which is around 0.06 bps/symbol/Hz at SNR = 0 dB. On the other hand, hybrid BB+PS SLP is shown in Fig. 9.6 to be the most spectrally-efficient approach among the hybrid symbol-level precoders of interest.

Up until this point in the simulation results, we have seen that among the hybrid symbol-level precoders of interest, one approach outperforms the other in terms of either power consumption, symbol error rate, or spectral efficiency. To have an all-inclusive comparison, we use the energy efficiency measure, as defined in Section 9.4, that incorporates all the aforementioned figures of merit in evaluating the overall precoding performance. The results are shown in Fig. 9.7, where the energy efficiencies of different fully-digital and hybrid multiuser precoders are plotted as a function of the transmit SNR for a system with $(N_{\mathrm{t}}, N_{\mathrm{l}}, N_{\mathrm{u}}) = (64, 8, 8)$. As can be seen, almost all of the hybrid symbol-level precoders are more energy-efficient than the fully-digital SLP, while the proposed hybrid BB+SW SLP approach with phase shifter selection outperforms the other schemes with either infinite or finite resolution phase shifters. The most energy-efficient scheme is shown to be hybrid BB+SW SLP with $b_{\mathrm{PS}} = 1$, using which energy efficiency gains of up to 75 Mbps/Joule per user can be achieved against the fully-digital SLP. In contrast to the Hybrid BB+PS SLP scheme, employing phase shifters with lower precision bits improves the energy-efficiency of Hybrid BB+SW SLP. This is because more phase shifters can be switched off using low-precision phase shifters, which leads to larger reductions in power consumption. It is important to note that in our power consumption model, we consider the same reference value for phase shifters with any number of precision bits. This is rather a simplistic approach as, in practice, higher-resolution

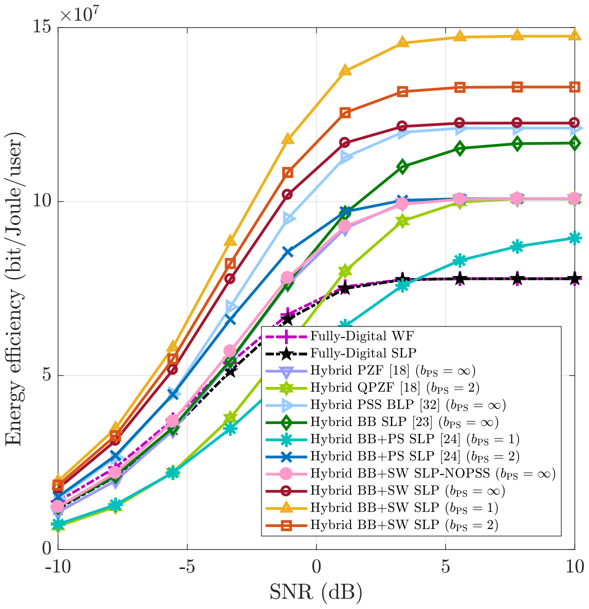


Figure 9.7: Energy efficiency versus transmit SNR with $(N_\mathrm{t}, N_\mathrm{l}, N_\mathrm{u}) = (64, 8, 8)$.

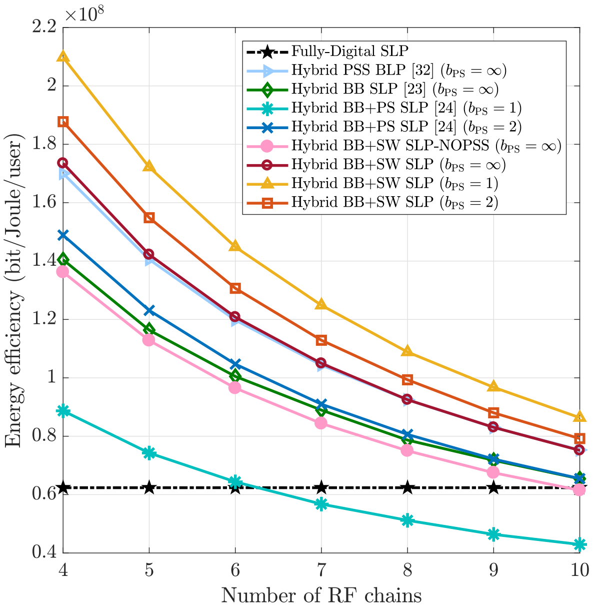


Figure 9.8: Energy efficiency at SNR $= -5$ dB as a function of N_l with $N_\mathrm{t} = 64$ and $N_\mathrm{u} = 4$.

phase shifters consume more power. In such a case, the results for power consumption and energy efficiency of the proposed hybrid precoder with low-resolution phase shifters would show an even higher gain compared to the other schemes of interest. It can further be seen from Fig. 9.7 that the proposed hybrid algorithm with $b_{\mathrm{PS}} = \infty$ outperforms the hybrid PSS BLP scheme, where both techniques employ a phase shifter selection mechanism via a switching network but on a symbol-level and block-level basis, respectively. In particular, the hybrid BB+SW technique shows higher energy efficiency gains against the hybrid PSS BLP scheme at low SNRs. We are further interested in the behavior of energy efficiency as a function of the number of available RF chains N_{l}, which is plotted in Fig. 9.8 for fixed numbers of transmit antennas $N_{\mathrm{t}} = 64$ and users $N_{\mathrm{u}} = 4$ at an SNR of -5 dB. A common trend across all the hybrid symbol-level precoders is that their energy efficiency becomes lower as N_{l} increases. This is in accordance with the power consumption results in Fig. 9.3, indicating that for a fixed number of antennas, a hybrid precoding implementation becomes less energy-efficient than its fully-digital counterpart whenever the number of RF chains exceeds an upper limit. This upper limit is shown in Fig. 9.8 to be larger for the proposed hybrid BB+SW SLP approach. On the other hand, comparing the proposed hybrid symbol-level precoder with the case where all the phase shifters are active, i.e., with no phase shifter selection, we can conclude that applying the phase shifter selection mechanism can substantially improve the energy efficiency of hybrid symbol-level precoding. The results in Fig. 9.8 shows that gains of up to 37 Mbps/Joule per user can be achieved using the hybrid BB+SW SLP method compared to its counterpart scheme without phase shifter selection.

Following the analytic complexity analysis provided in Section 9.3, we numerically evaluate the proposed hybrid SLP algorithm's computational complexity in both scenarios with infinite-resolution and discrete phase shifters. The complexity results, in terms of the required number of outer iterations (i.e., cycles) for convergence, is shown in Fig. 9.9 as a function of the number of RF chains N_{l}. However, it is important to note that the complexity of solving the inner sub-problem (9.13) is not of our interest since, as mentioned earlier, this problem is a typical linearly-constrained quadratic program which can efficiently be solved using many existing algorithms. As might be expected, the number of outer iterations until convergence of the proposed hybrid SLP algorithm increases with N_{l} in all the cases due to the corresponding growth in the problem size. On the other hand, the computation cost increases by reducing the precision of the phase shifters. Such an observation, however, is not surprising since having discrete possible phase states causes a discontinuity in the feasible region of the optimization problem, and consequently, more cycles are needed for convergence to a stationary point.

9.6 Conclusions

In this chapter, we proposed a hybrid analog-digital precoding scheme for large-scale multiuser mmWave downlink systems. The multiuser precoding operation is split between the digital and analog domains, where processing in the analog domain is carried out through fully-connected networks of switches and phase shifters. The use of on-off

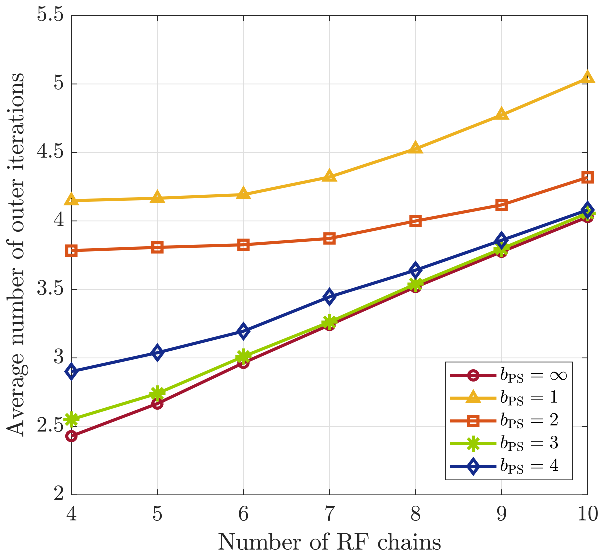


FIGURE 9.9: Average number of iterations till convergence of the proposed hybrid SLP algorithm as a function of N_{l} with $N_{\mathrm{t}} = 64$.

switches enables us to perform phase shifter selection in the analog precoder. We adopted a CSI-only design approach for the phase-shifting network, whereas the digital baseband precoder and the switching network are optimized in a symbol-level manner, i.e., by exploiting the instantaneous data symbols to enable CI at the receiver side. We formulated our design problem to minimize the Euclidean distance between the hybrid symbol-level precoder and its optimal fully-digital counterpart, where a power-constrained max-min SINR design criterion subject to CI constraints was adopted. Our design approach led us to an intractable binary optimization problem. We tackled this difficulty by transforming the original problem to an equivalent continuous-domain biconvex form, which can efficiently be solved for a sub-optimal solution via the standard block coordinate descent (BCD) algorithm. We evaluated the computational complexity of the proposed scheme, where numerical results showed that the adopted BCD algorithm needs only a few (usually less than ten) cycles to converge. To assess and compare different fully-digital/hybrid precoding schemes from both performance and power consumption points of view, we analyzed the energy efficiency by considering appropriate models for the RF elements' power dissipation. Our simulation results indicated that applying the phase shifter selection approach, up to half of the phase shifters can be switched off, allowing for reductions of multi-Watts in analog circuitry power consumption. This power consumption reduction can significantly improve precoding's energy efficiency compared to the fully-digital and state-of-the-art hybrid symbol-level techniques. Moreover, we evaluated the proposed hybrid precoding scheme with both infinite and finite precision phase shifters. It was shown that using phase shifters with lower precision bits, on the one hand, degrades the spectral efficiency, but on the other hand, allows for more power-savings due to a larger number of deactivated phase shifters, and therefore, is more energy-efficient.

Chapter 10

Concluding Remarks and Future Work

This thesis addressed several challenges in designing an SLP scheme for an MU-MIMO downlink system. In summary, we defined and formulated CI constraints for generic modulation schemes, proposed computationally-efficient solutions to the corresponding design problem, validated the design for real-time implementations, studied robust design of the precoder in the presence of channel/design uncertainty, and revisited the problem in quantized and hybrid analog-digital precoding architectures. Accordingly, the main conclusions drawn from the work carried out in this thesis and possible extensions to the current results are described in the subsequent sections.

10.1 Main Conclusions

First, we elaborated a systematic framework in Chapter 3 to describe optimal and relaxed CI restrictions as linear convex constraints which can be utilized in an SLP design problem with different objectives and requirements. This framework generalizes the definition of CIRs for modulation schemes with constellations of any given shape and order. In particular, we defined the DPCIRs and showed that these regions are optimal when the target SEP is not allowed to increase. In a more flexible setting, we considered relaxed CIRs and guaranteed the target SEP using the union bound, which led us to introduce the UBCIRs. We fully characterized the DPCIRs for a generic constellation and derived some of their properties. Using the proposed systematic description for the DPCIRs and UBCIRs, we formulated and discussed two well-known precoding optimization problems in a downlink MU-MIMO unicast channel, namely, power optimization and SINR balancing. The SINR-constrained SLP power minimization was formulated as a convex problem and studied in a realistic scenario, where a feasibility condition was obtained for this problem. Our results indicated that the DPCIR-based and UBCIR-based SLP designs can reduce the transmit power consumption without imposing additional complexity on the transmitter compared to the state-of-the-art schemes. For the more challenging and generally non-convex problem of SLP SINR balancing with a max-min fair objective, the properties of DPCIRs helped us to reformulate the problem in a convex form, which can be solved for a sub-optimal solution. To tackle this problem, we proposed two different methods, namely, SDP formulation and BCD optimization. We provided a detailed comparison of performance and complexity for the proposed methods, where it was shown that the BCD optimization based method can outperform the SDP formulation one at the cost of higher computational complexity.

It is known that solving the SLP design problem for the exact solution may lead to an impractical transmitter complexity due to high per-symbol computation cost. We addressed this challenge in chapter 4, where two computationally-efficient methods are proposed to approximately solve the SLP power minimization problem with CI and SINR constraints. This was done by first simplifying the original formulation and reformulating it as an NNLS design, and then discussing the simplified problem's optimality via the KKT conditions. The analyses helped us to derive two closed-form approximate SLP designs, namely, CF-SLP and ICF-SLP. The CF-SLP design performs quite close to the optimal SLP scheme in systems with a relatively small number of users, but shows a poor performance with increasing the system size. The ICF-SLP solution, on the other hand, substantially reduces the loss with respect to the optimal solution, particularly in the large system regime. Furthermore, the ICF-SLP design showed competitive performance compared to the SLP solution obtained from the iterative APGD algorithm, but with reduced time complexity. In comparison with conventional block-level precoding schemes, we showed that both CF-SLP and ICF-SLP methods outperform the ZF precoder in all scenarios and the optimal power minimizer block-level precoder at high target SINRs. We conclude that the CF-SLP and ICF-SLP designs can successfully relieve the prohibitive computation cost of the SLP design. Furthermore, they are promising alternatives (with a comparable complexity) for the block-level precoding schemes, especially in the high SINR regime.

To assess the potential advantages of the proposed low-complexity SLP designs in a high-throughput downlink multiuser MISO system, in Chapter 5, we developed an optimized FPGA design based on the CF-SLP solution. We further simplified this solution by assuming mutually orthogonal channel vectors and proposed an approximate low-complexity design algorithm that can operate in a real-time mode. We analyzed the computational complexity of the proposed design and showed that it has the same per-symbol complexity order as that of the ZF precoding. We then used the Xilinx Vivado HLS tool to translate the design algorithm into an HDL code and also to optimize the design in order to achieve a lower latency, and therefore, a higher throughput. The synthesis results, including performance, timing and resource utilization estimates verified the efficiency of our HDL design. The generated HDL core was evaluated in a simulation environment within the LabVIEW software. The simulations showed that the HDL design of our proposed algorithm is able to operate at a symbol rate of 100 Mega symbols per second per user when deployed on a specific Xilinx FPGA part, which makes it attractive for real-time implementations. Using the MATLAB software, we further evaluated the loss of our design algorithm with respect to the optimal SLP solution, where the loss is shown to be less than 1 dB according to our numerical results. This loss is mainly due to the used approximation in deriving the algorithm and also due to the adopted fixed-point arithmetic for the FPGA design. Furthermore, the simulation results indicated that the proposed HDL implementation of SLP outperforms the ZF scheme in terms of power efficiency, where an improvement of up to 50 percent can be achieved.

In Chapter 6 and Chapter 7, we addressed the problem of designing a robust SLP scheme in downlink MU-MIMO systems, respectively, under channel and design uncertainties. Under channel uncertainties, we assumed imperfect bounded or stochastic CSI error at the BS and considered a QoS-constrained design criterion in Chapter 6. We developed robust CI constraints for each uncertainty model and provided the corresponding robust formulations for the SLP design problem. With bounded CSI errors, we derived a worst-case robust formulation to guarantee the users' requirements for every possible realization of the CSI error within the uncertainty region. Under the stochastic uncertainty model, we adopted a probabilistic approach to enforce the CI constraints and derived two computationally tractable approximate convex restrictions with different levels of conservatism. Our results showed that both the proposed robust restrictions outperform the well-known sphere bounding method, while each of them is superior to the other under different robustness settings. Compared with a conventional block-level robust scheme, the proposed robust methods were shown to be more efficient at moderate-to-high target SINR values. However, a more considerable advantage of the proposed robust SLP approaches is their higher feasibility rate for wide ranges of violation probability and uncertainty variance, which is indifferent to the target SINR. We also showed through complexity analysis that the improved performances of the proposed robust SLP designs come with an increased computational complexity by an order of the number of users in the limiting case.

Under design uncertainties, in Chapter 7, we proposed a worst-case approach for the QoS-constrained SLP problem in a scenario where the precoder's output undergoes linear distortion with bounded additive noise. We proposed a new problem formulation which allowed us to cast the worst-case design of the distorted SLP as a min-max problem by introducing relaxed CI constraints. We solved this problem using an iterative block coordinate ascent-descent algorithm to obtain the robust precoded signal. This algorithm iterates between finding the optimal precoded signal and the worst-case additive distortion vector. Our simulation results showed that the proposed worst-case approach can improve the SLP scheme's performance under linear distortions in terms of energy efficiency.

We revisited the SLP design problem for low-cost transmitter architectures in Chapter 8 and Chapter 9, where practical limitations are given, respectively, on the resolution of DACs or the number of RF chains. In Chapter 8, we proposed a finite-alphabet SLP design for massive MU-MIMO downlink systems equipped with finite-resolution DACs. We adopted a power-constrained max-min fair design criterion with the aim of exploiting CI at the users. The design problem, in its original form, is a discrete linearly-constrained quadratic programming whose solution requires a high computational complexity. We dealt with this issue in several steps and reformulated the problem into an equivalent continuous-domain form, which can efficiently be solved using a standard block coordinate descent algorithm that converges within a few iterations. Our results showed that employing DACs with higher resolutions leads to lower BERs at the cost of reduced power efficiency. We also investigated the case with one-bit DACs, where comparisons between the proposed quantized SLP technique and some other well-known one-bit precoding schemes showed a superior performance in terms of uncoded BER, with up to 2 dB gain depending on the system setup.

Finally, in Chapter 9, we proposed a hybrid analog-digital precoding scheme for large-scale mmWave MU-MIMO downlink systems. The multiuser precoding operation was split between the digital and analog domains, where processing in the analog domain is carried out through fully-connected networks of switches and phase shifters. The use of on-off switches enabled us to perform phase shifter selection in the analog precoder. We adopted a CSI-only design approach for the phase-shifting network, whereas the digital baseband precoder and the switching network are optimized in a symbol-level manner, i.e., by exploiting the instantaneous data symbols to enable constructive interference at the receiver side. We formulated our design problem to minimize the Euclidean distance between the hybrid symbol-level precoder and its optimal fully-digital counterpart. To assess and compare different fully-digital/hybrid precoding schemes from both performance and power consumption points of view, we analyzed the energy efficiency by considering appropriate models for the RF elements' power dissipation. Our results indicated that by applying the phase shifter selection mechanism, we can reduce the power consumption of the analog circuitry through switching up to half of the phase shifters off. This reduction can significantly improve precoding's energy efficiency compared to the fully-digital and state-of-the-art hybrid SLP techniques. Moreover, we evaluated the proposed hybrid scheme with both infinite and finite precision phase shifters, where it was shown that using phase shifters with lower precision bits, on the one hand, degrades the spectral efficiency, but on the other hand, allows for more power-savings due to a larger number of deactivated phase shifters, and therefore, is more energy-efficient.

10.2 Future Work

The work carried out in this thesis can be extended in several directions. Below, we suggest some possible extensions to the current work.

1. The CI-based SLP design problem aims to find the precoded transmit signal that is optimal with respect to the given constellation set. To achieve higher CI gains by the SLP scheme, one may also optimize the constellation set while preserving the users' detection accuracy or maintaining it above a certain threshold depending on the system requirements. One may reformulate or redefine the design problem targeting mutual information while satisfying the CI and SINR constraints. In doing so, it might be helpful to study the dependence of the SLP performance on the signal constellation geometry.

2. The computationally-efficient SLP solutions of this work were proposed based on

the NNLS formulation of the design problem. It might be possible to further improve these solutions by exploiting the specific structure of the problem, e.g., the block diagonal structure of the normal matrix. In addition, these solutions were proposed only for the case where perfect transmit CSI is available. Under imperfect CSI knowledge, a possible future work could be to develop low-complexity solutions for the proposed robust SLP techniques.

3. Some practical challenges need to be further addressed in design of the robust SLP scheme under channel uncertainty, as listed below:

   – Firstly, the stochastic robust optimization problem might be infeasible for rather large CSI error variances or relatively small violation probabilities. In such cases, a trivial alternative is to use the non-robust precoder's solution; however, a more sophisticated alternative may improve the robust design's reliability. For example, one can relax the violation probability and resolve the robust optimization problem till a feasible solution is found.

   – Secondly, the proposed robust SLP techniques in this work only support single-level modulation schemes. To be more specific, they can guarantee the CI constraints (with a certain probability, in the case of stochastic uncertainty) only for outer constellation symbols, i.e., those symbols with unbounded Voronoi regions. It would be an interesting problem to extend the current scheme to a more general case where the constellation includes some inner symbols for which the DPCIRs are only the constellation symbol itself. For such symbols, the probabilistic CI constraint always has an empty feasible region, so does the robust optimization problem. In order to generalize the current scheme to the case with multi-level modulation schemes, one may define a relaxed CI region by assuming a confidence region around each inner symbol, allowing the noise-free received signal to lie within the relaxed region. This relaxation may affect the users' SER performance, but on the other hand, it may result in lower transmission powers. Hence, one also needs to carefully choose the relaxation parameter such that a certain performance level is guaranteed. In general, this might be somewhat challenging, and the design approach may need to be done analytically by taking the given system/user requirements into account.

4. An interesting extension to the proposed low-complexity FPGA design of SLP could be to estimate the amount of power consumed by the FPGA while running the IP block and compare it with the saved power at the transmitter. Another future work is to further optimize the HDL code and seek possible improvements in the algorithm's accuracy. The subsequent step is to conduct experimental validation of the proposed algorithm by deploying the HDL design on an actual FPGA.

5. It would be an interesting problem to consider a multi-carrier massive MU-MIMO system and investigate how the proposed one-bit SLP scheme scales with frequency

(i.e., with the number of subcarriers) in a wide-band system. It might be challenging to quantize several sub-carriers at the same time when using one-bit DACs.

6. The analysis on the energy efficiency of hybrid precoding schemes in this work was performed by considering only static power losses at the transmitter, i.e., the power dissipated by the analog, digital, or mixed signal components. An important challenge concerning the energy efficiency of hybrid architectures is the dynamic power losses in the system. This sort of loss refers to the power dissipated at the combiners when phase shifters' output signals are added with different phases. Accordingly, a possible extension to this work is to take the dynamic losses into account when designing the hybrid SLP scheme. On the other hand, the losses in the switching and phase-shifting networks due to the network's structure (i.e., partial or full connectivity) were not specifically modeled in our analysis; however, such losses become of concern and have to be taken into account for large networks, i.e., as the number of transmit antennas grows.

Appendix A

Appendices for Chapter 3

A.1 Proof of Lemma 2

The intersection of a finite number of closed halfspaces is an unbounded polyhedron if and only if the outward normals to the associated boundary hyperplanes lie on a single closed halfspace [242, p. 20, Theorem 4]. Accordingly, for any $\mathbf{x}_m \in \mathcal{X}$ with unbounded $\mathcal{D}_m^{(\text{ML})}$, all the outward normal vectors $-\mathbf{a}_{m,k}$ for $k \in \mathcal{J}_m$ lie on a single halfspace. Since the polyhedron $\mathcal{D}_m^{(\text{DP})}$ has the same set of outward normals $-\mathbf{a}_{m,k}$ for all $k \in \mathcal{J}_m$, it is also unbounded. An unbounded polyhedron is uniquely determined from its vertices and the directions of its infinite edges [242, p. 31, Theorem 4]. Furthermore, it is straightforward to check that $\mathbf{x}_m$ is the unique solution of $\mathbf{A}_m\mathbf{x} = \mathbf{b}_m^{(\text{ML})} + \mathbf{b}_m^{(\text{DP})}$, i.e., all the contributing hyperplanes have a common intersection point $\mathbf{x}_m$. This means that $\mathcal{D}_m^{(\text{DP})}$, which is given by the solution set of $\mathbf{A}_m\mathbf{x} \succeq \mathbf{b}_m^{(\text{ML})} + \mathbf{b}_m^{(\text{DP})}$, has a single vertex at $\mathbf{x}_m$ and two infinite edges, i.e., a polyhedral angle. In addition, since any two neighboring points share a common Voronoi edge, the two infinite edges of $\mathcal{D}_m^{(\text{DP})}$ correspond to the two neighboring points of $\mathbf{x}_m$ on $\mathbf{bd}\mathcal{X}$ (i.e., $\mathcal{S}_m \cap \mathbf{bd}\mathcal{X}$) with unbounded Voronoi regions. Each infinite edge of $\mathcal{D}_m^{(\text{DP})}$ is then parallel to a hyperplane with normal vector $\mathbf{a}_{m,k} = \mathbf{x}_m - \mathbf{x}_k$, where $\mathbf{x}_k \in \mathcal{S}_m \cap \mathbf{bd}\mathcal{X}$; therefore it is perpendicular to $\mathbf{x}_m - \mathbf{x}_k$. This completes the proof.

A.2 Proof of Lemma 3

To prove this lemma, we first state a well-known property of convex sets.

Property 2. $\mathbf{v}_o$ *is the minimum distance vector from the origin to the convex set* $\mathcal{V}$ *if and only if for any vector* $\mathbf{v} \in \mathcal{V}$ *we have* $\mathbf{v}_o^{\mathrm{T}}\mathbf{v} \geq \mathbf{v}_o^{\mathrm{T}}\mathbf{v}_o$*, with equality for* $\mathbf{v}$ *lying on the hyperplane orthogonal to* $\mathbf{v}_o$ *[243, p. 69, Theorem 1].*

For any $\mathbf{x}_m \in \mathbf{int}\mathcal{X}$, Lemma 3 holds straightforwardly as $\mathcal{D}_m^{(\text{DP})} = \mathbf{x}_m$. Therefore, in what follows we only focus on the constellation points belonging to $\mathbf{bd}\mathcal{X}$.
Sufficiency: Having $\mathbf{0} \in \mathbf{conv}\mathcal{X}$, let further assume that $\mathbf{0} \in \mathcal{X}$. This assumption, as mentioned earlier in section 3.4, does not have any impact on $\mathcal{D}_m^{(\text{DP})}$ for any $\mathbf{x}_m \in \mathbf{bd}\mathcal{X}$,

regardless of whether $\mathbf{0} \in \mathbf{bd}\mathcal{X}$ or $\mathbf{0} \in \mathbf{int}\mathcal{X}$. By substituting $\mathbf{x}_k = \mathbf{0}$ in (3.28), for all $\mathbf{x}_m \in \mathcal{X}$ we have $\|\mathbf{x}\| \geq \|\mathbf{x}_m\|$ for all $\mathbf{x} \in \mathcal{D}_m^{(\mathrm{DP})}$. This completes the proof of sufficiency.
Necessity: By contradiction, if $\mathbf{0} \notin \mathbf{conv}\mathcal{X}$, let assume a new constellation set $\tilde{\mathcal{X}}$ having all the points of $\mathcal{X}$ including the origin, i.e., $\tilde{\mathcal{X}} = \mathcal{X} \cup \{\mathbf{0}\}$, hence $\mathbf{conv}\mathcal{X} \subset \mathbf{conv}\tilde{\mathcal{X}}$. Clearly, $\mathbf{0} \in \mathbf{bd}\tilde{\mathcal{X}}$ and according to Lemma 2, there always exist exactly two constellation points on $\mathbf{bd}\tilde{\mathcal{X}}$ that $\mathbf{0}$ contributes to their DPCIRs. Let $\mathbf{x}_l$ be one of these points with $\mathcal{D}_l^{(\mathrm{DP})}$ and $\tilde{\mathcal{D}}_l^{(\mathrm{DP})}$ denoting its associated DPCIR in $\mathcal{X}$ and $\tilde{\mathcal{X}}$, respectively. We further denote by $\tilde{\mathcal{S}}_l$ the set of neighboring points of $\mathbf{x}_l$ in $\tilde{\mathcal{X}}$. Let $\mathcal{H}_l^{(\mathrm{O})} = \left\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{x}_l^{\mathrm{T}}\mathbf{x} \geq \mathbf{x}_l^{\mathrm{T}}\mathbf{x}_l\right\}$ be the distance preserving halfspace from $\mathbf{0}$ to $\mathbf{x}_l$. Since $\mathbf{0} \in \tilde{\mathcal{S}}_l$, we have $\tilde{\mathcal{D}}_l^{(\mathrm{DP})} = \mathcal{H}_l^{(\mathrm{O})} \cap \mathcal{D}_l^{(\mathrm{DP})} \neq \mathcal{D}_l^{(\mathrm{DP})}$, i.e., the halfspace $\mathcal{H}_l^{(\mathrm{O})}$ does not contain $\mathcal{D}_l^{(\mathrm{DP})}$. Hence, $\left\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{x}_l^{\mathrm{T}}\mathbf{x} = \mathbf{x}_l^{\mathrm{T}}\mathbf{x}_l\right\}$ is not a supporting hyperplane for $\mathcal{D}_l^{(\mathrm{DP})}$ at $\mathbf{x}_l$ [121, p. 51]. This implies that there exist some $\mathbf{x} \in \mathcal{D}_l^{(\mathrm{DP})}$ for which $\mathbf{x}_l^{\mathrm{T}}\mathbf{x} < \mathbf{x}_l^{\mathrm{T}}\mathbf{x}_l$. According to Property 2 (which gives a necessary and sufficient condition), $\mathbf{x}_l$ is not the minimum distance vector from the origin in $\mathcal{D}_l^{(\mathrm{DP})}$. Consequently, $\|\mathbf{x}\| \geq \|\mathbf{x}_l\|$ does not hold for some $\mathbf{x} \in \mathcal{D}_l^{(\mathrm{DP})}$ which contradicts $\|\mathbf{x}\| \geq \|\mathbf{x}_l\|$ for all $\mathbf{x} \in \mathcal{D}_l^{(\mathrm{DP})}$.

A.3 Proof of Theorem 4

To prove this theorem, we need the following lemma.

Lemma 21. *If* $\mathbf{0} \notin \mathbf{conv}\mathcal{X}$*, there exists at least one constellation point* $\mathbf{x}_l \in \mathcal{X}$ *for which for any* $\mathbf{x} \in \mathcal{D}_l^{(\mathrm{DP})}$*, we have* $\mathbf{0} \notin \mathbf{conv}\tilde{\mathcal{X}}_{\mathbf{x}_l,\mathbf{x}}$*, where* $\tilde{\mathcal{X}}_{\mathbf{x}_l,\mathbf{x}} = \mathcal{X} \cup \{\mathbf{x}\}$.

Proof. If $\mathbf{0} \notin \mathbf{conv}\mathcal{X}$, for any $\mathbf{x}_m \in \mathcal{X}$ and any $\mathbf{x} \in \mathcal{D}_m^{(\mathrm{DP})}$ with $\tilde{\mathcal{X}}_{\mathbf{x}_m,\mathbf{x}} = \mathcal{X} \cup \{\mathbf{x}\}$, we define

$$\mathcal{C}_m = \bigcup_{\mathbf{x} \in \mathcal{D}_m^{(\mathrm{DP})}} \mathbf{conv}\tilde{\mathcal{X}}_{\mathbf{x}_m,\mathbf{x}}.$$

Since $\mathbf{conv}\mathcal{X} \subseteq \mathbf{conv}\tilde{\mathcal{X}}_{\mathbf{x}_m,\mathbf{x}}$, it follows from the definition of convex hull that

$$\mathbf{conv}\mathcal{X} = \bigcap_{\mathbf{x}_m \in \mathcal{X}} \mathcal{C}_m.$$

If $\mathbf{0} \in \mathcal{C}_m$ for all $\mathbf{x}_m \in \mathcal{X}$, then $\mathbf{0} \in \mathbf{conv}\mathcal{X}$, which contradicts our assumption. As a result, there must exist at least one constellation point, say $\mathbf{x}_l$, for which $\mathcal{C}_l$ and therefore none of $\mathbf{conv}\tilde{\mathcal{X}}_{\mathbf{x}_l,\mathbf{x}}$ for $\mathbf{x} \in \mathcal{D}_l^{(\mathrm{DP})}$ contains the origin, as required.

□

Now, we can start the proof of Theorem 4 as follows.
Sufficiency: Suppose $\mathbf{0} \in \mathbf{conv}\mathcal{X}$. Assuming a constellation point $\mathbf{x}_m \in \mathcal{X}$ and its DPCIR $\mathcal{D}_m^{(\mathrm{DP})}$, let $\mathbf{y}_1$ and $\mathbf{y}_2$ be two points in $\mathcal{D}_m^{(\mathrm{DP})}$ such that $\mathbf{A}_m\mathbf{y}_1 = \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})} + \mathbf{t}_{m,1}$ and $\mathbf{A}_m\mathbf{y}_2 = \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})} + \mathbf{t}_{m,2}$ with $\{\mathbf{t}_{m,1}, \mathbf{t}_{m,2}\} \in \mathbb{R}_+^{M_m}$ and $\mathbf{t}_{m,1} \prec \mathbf{t}_{m,2}$. Let consider a new constellation $\tilde{\mathcal{X}} = \mathcal{X} \cup \{\mathbf{y}_1\}$. It is clear that $\mathbf{conv}\mathcal{X} \subseteq \mathbf{conv}\tilde{\mathcal{X}}$, and therefore

$\mathbf{0} \in \mathbf{conv}\tilde{\mathcal{X}}$. The DPCIR of $\mathbf{y}_1$ can be described as

$$\mathcal{D}_{\mathbf{y}_1}^{(\mathrm{DP})} = \left\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{A}_m\mathbf{x} = \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})} + \mathbf{t}_{m,1} + \mathbf{t}_1, \mathbf{t}_1 \in \mathbb{R}_+^{M_m}\right\}.$$

Let $\bar{\mathbf{t}} = \mathbf{t}_{m,2} - \mathbf{t}_{m,1}$, then $\mathbf{A}_m\mathbf{y}_2 = \mathbf{b}_m^{(\mathrm{ML})} + \mathbf{b}_m^{(\mathrm{DP})} + \mathbf{t}_{m,1} + \bar{\mathbf{t}}$, where $\bar{\mathbf{t}} \in \mathbb{R}_{++}^{M_m}$, which means that $\mathbf{y}_2 \in \mathcal{D}_{\mathbf{y}_1}^{(\mathrm{DP})}$. As a consequence, from Lemma 3, we have $\|\mathbf{y}_1\| < \|\mathbf{y}_2\|$ and the proof of sufficiency is complete.

Necessity: By contradiction, suppose $\mathbf{0} \notin \mathbf{conv}\mathcal{X}$. Then, based on Lemma 21, there exists a constellation point $\mathbf{x}_l$ for which $\mathbf{0} \notin \mathbf{conv}\tilde{\mathcal{X}}_{\mathbf{x}_l,\mathbf{x}}$ for all $\mathbf{x} \in \mathcal{D}_l^{(\mathrm{DP})}$. Let $\mathbf{y}_1 \in \mathcal{D}_l^{(\mathrm{DP})}$, then $\mathbf{A}_l\mathbf{y}_1 = \mathbf{b}_l^{(\mathrm{ML})} + \mathbf{b}_l^{(\mathrm{DP})} + \mathbf{t}_{l,1}$ with $\mathbf{t}_{l,1} \in \mathbb{R}_+^{M_l}$. The DPCIR of $\mathbf{y}_1$ can then be expressed as

$$\mathcal{D}_{\mathbf{y}_1}^{(\mathrm{DP})} = \left\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \mathbf{A}_l\mathbf{x} = \mathbf{b}_l^{(\mathrm{ML})} + \mathbf{b}_l^{(\mathrm{DP})} + \mathbf{t}_{l,1} + \mathbf{t}_1, \mathbf{t}_1 \in \mathbb{R}_+^{M_l}\right\}.$$

Since $\mathbf{0} \notin \mathbf{conv}\tilde{\mathcal{X}}_{\mathbf{x}_l,\mathbf{y}_1}$, it follows from Lemma 3 and Property 2 that there exists $\mathbf{y}_2 \in \mathcal{D}_{\mathbf{y}_1}^{(\mathrm{DP})}$ such that $\mathbf{A}_l\mathbf{y}_2 = \mathbf{b}_l^{(\mathrm{ML})} + \mathbf{b}_l^{(\mathrm{DP})} + \mathbf{t}_{l,1} + \bar{\mathbf{t}}, \bar{\mathbf{t}} \in \mathbb{R}_{++}^{M_l}$, for which $\|\mathbf{y}_2\| < \|\mathbf{y}_1\|$. But $\mathbf{t}_{l,1} + \bar{\mathbf{t}} = \mathbf{t}_{l,2}$ yields $\mathbf{t}_{l,2} \succ \mathbf{t}_{l,1}$, which is a contradiction. This completes the proof.

A.4 Proof of Lemma 6

To verify the equivalence of problems (3.41) and (3.43), let consider two cases. If $N_\mathrm{u} = N_\mathrm{t}$, then $\mathbf{G}$ is full rank with high probability, and hence $\mathbf{G}^\dagger = \mathbf{G}^{-1}$. As a result, the constraint $\mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma\Gamma}(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}) + \mathbf{t}$ in (3.41) gives a unique solution for any fixed $\mathbf{t}$. In this case, there exists a bijection between $\mathbf{t}$ and $\bar{\mathbf{u}}$, which implies that one of them can be obtained as a one-to-one function of the other. Therefore, optimizing $\mathbf{t}$ is equivalent to optimizing both $\bar{\mathbf{u}}$ and $\mathbf{t}$. Otherwise, if $N_\mathrm{u} < N_\mathrm{t}$, the constraint $\mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma\Gamma}(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}) + \mathbf{t}$ may have more than one solution as $\mathbf{G}$ is full row rank in this case. Thus, the LCQP in (3.41) can be written as

$$\begin{aligned}
\min_{\mathbf{t}}\ \min_{\bar{\mathbf{u}}}\quad & \bar{\mathbf{u}}^\mathrm{T}\bar{\mathbf{u}} \\
\text{s.t.}\quad & \mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma\Gamma}(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}) + \mathbf{t}, \\
& \mathbf{t}_i = \mathbf{0},\ \forall i \notin \mathcal{I}, \\
& \mathbf{t}_i \succeq \mathbf{0},\ \forall i \in \mathcal{I},
\end{aligned} \tag{A.1}$$

for which the solution to the inner minimization is given by $\bar{\mathbf{u}} = \mathbf{G}^\dagger\left(\mathbf{\Sigma\Gamma}(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}) + \mathbf{t}\right)$, which is the least-norm solution to the system of linear equations $\mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma\Gamma}(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}) + \mathbf{t}$. In addition, we can impose the constraints $\mathbf{t}_i = \mathbf{0}$ for all $i \notin \mathcal{I}$, if any exist, using a diagonal matrix $\mathbf{W}$ with a one element on the main diagonal if it corresponds to a symbol $i \in \mathcal{I}$, and zero otherwise. When computing $\mathbf{u}$, such a matrix excludes those elements of $\mathbf{t}$ that correspond to a symbol $i \notin \mathcal{I}$. Consequently, the optimal precoded vector can be expressed as $\bar{\mathbf{u}} = \mathbf{G}^\dagger\left(\mathbf{\Sigma\Gamma}(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}) + \mathbf{W}\mathbf{t}\right)$ with $\mathbf{t}$ given by $\mathbf{G}\bar{\mathbf{u}} = \mathbf{\Sigma\Gamma}(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}) + \mathbf{W}\mathbf{t}$.

A.5 Proof of Lemma 5

Let $\mathbf{t} = \mathbf{0}$, then the feasibility problem (3.39) reduces to

$$\begin{aligned} \text{find} \quad & \bar{\mathbf{u}} \\ \text{s.t.} \quad & \mathbf{G}\bar{\mathbf{u}} = \boldsymbol{\Sigma}\boldsymbol{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right), \\ & \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}} \leq p. \end{aligned} \tag{A.2}$$

Now suppose that $\bar{\mathbf{u}}_{\mathrm{o}} = \mathbf{G}^{\dagger}\boldsymbol{\Sigma}\boldsymbol{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right)$ is a (not necessarily unique) solution to the system of linear equations

$$\mathbf{G}\bar{\mathbf{u}} = \boldsymbol{\Sigma}\boldsymbol{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right). \tag{A.3}$$

In fact, $\bar{\mathbf{u}}_{\mathrm{o}}$ is equal to the solution of the zero-forcing (ZF) precoding [4] when identical target SINRs are allocated to all the users. We argue the existence of $\bar{\mathbf{u}}_{\mathrm{o}}$ as follows. In case $N_{\mathrm{u}} = N_{\mathrm{t}}$, due to the random channel matrices $\mathbf{H}_i$ for $i = 1, ..., N_{\mathrm{u}}$, matrix $\mathbf{G}$ is full rank with high probability. This means that the probability of (A.3) having more than one solution is almost zero. On the other hand, for $N_{\mathrm{u}} < N_{\mathrm{t}}$, matrix $\mathbf{G}$ is full row rank and (A.3) expresses an underdetermined system of linear equations for which $\bar{\mathbf{u}}_{\mathrm{o}}$ is the least-norm solution. Having $\bar{\mathbf{u}}_{\mathrm{o}}$ as a solution to (A.3), if $\bar{\mathbf{u}}_{\mathrm{o}}^{\mathrm{T}}\bar{\mathbf{u}}_{\mathrm{o}} \leq p$, then $\bar{\mathbf{u}}_{\mathrm{o}}$ is a feasible point for (A.2); this further implies the feasibility of problem (3.39) since this problem is a relaxed version of (A.2). Therefore,

$$\bar{\mathbf{u}}_{\mathrm{o}}^{\mathrm{T}}\bar{\mathbf{u}}_{\mathrm{o}} = \left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right)^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{\Sigma}(\mathbf{G}\mathbf{G}^{\mathrm{T}})^{\dagger}\boldsymbol{\Sigma}\boldsymbol{\Gamma}\left(\mathbf{b}^{(\mathrm{ML})} + \mathbf{b}^{(\mathrm{DP})}\right) \leq p,$$

is a sufficient condition for the feasibility problem (3.39) to have at least one solution.

Appendix **B**

Appendices for Chapter 4

B.1 Proof of Lemma 7

Sufficiency: It is clear from (4.6) that $\bar{\mathbf{u}}^*$ equals the ZF solution if and only if $\mathbf{t}^* = \mathbf{0}$. Given $\mathbf{p} \succeq \mathbf{0}$, let assume by contradiction that $\mathbf{t}^* \neq \mathbf{0}$, i.e., there exist some l such that $t_l^* > 0$, which gives $\mathbf{p}^{\mathrm{T}}\mathbf{t}^* \geq 0$. Let us rewrite the optimality condition (4.13) as $\mathbf{t}^{*\mathrm{T}}\mathbf{Q}\mathbf{t}^* + \mathbf{p}^{\mathrm{T}}\mathbf{t}^* = 0$. By definition, $\mathbf{Q}$ is symmetric and $\mathbf{Q} = (\mathbf{H}^{\dagger}\mathbf{A}^{-1})^{\mathrm{T}}\mathbf{H}^{\dagger}\mathbf{A}^{-1}$, where $\mathbf{H}^{\dagger}\mathbf{A}^{-1}$ has full column rank, with high probability, due to the random concatenated channel $\mathbf{H}$. Hence, $\mathbf{Q}$ is a positive definite matrix [119, Theorem 7.2.7], i.e., $\mathbf{t}^{*\mathrm{T}}\mathbf{Q}\mathbf{t}^* > 0$ for any $\mathbf{t}^* \neq \mathbf{0}$. This, however, yields $\mathbf{t}^{*\mathrm{T}}\mathbf{Q}\mathbf{t}^* + \mathbf{p}^{\mathrm{T}}\mathbf{t}^* > 0$ which contradicts the KKT condition (4.13). Therefore, having $\mathbf{p} \succeq \mathbf{0}$, it necessarily holds that $\mathbf{t}^* = \mathbf{0}$, as required.
Necessity: Assuming $\mathbf{t}^* = \mathbf{0}$, it immediately follows from (4.12) that $\mathbf{p} \succeq \mathbf{0}$. This completes the proof.

Appendix C

Appendices for Chapter 6

C.1 Proof of equality (b) in (6.26)

First, let $\mathbf{Q}_i \triangleq \mathbb{E}\{\text{vec}(\mathbf{E}_i)\text{vec}(\mathbf{E}_i)^{\text{T}}\}$ denote the covariance matrix of $\text{vec}(\mathbf{E}_i)$ as given in (6.24). It follows that

$$\mathbf{Q}_i = \frac{1}{2}\,\xi_i^2 \begin{bmatrix} \mathbf{I}_{N_{\text{t}}} \otimes \mathbf{I}_2 & \mathbf{I}_{N_{\text{t}}} \otimes \mathbf{J}_2 \\ \mathbf{I}_{N_{\text{t}}} \otimes \mathbf{J}_2^{\text{T}} & \mathbf{I}_{N_{\text{t}}} \otimes \mathbf{I}_2 \end{bmatrix}, \tag{C.1}$$

where we have used the facts that $(\mathbf{I}_{N_{\text{t}}} \otimes \mathbf{J}_2)^{\text{T}} = \mathbf{I}_{N_{\text{t}}} \otimes \mathbf{J}_2^{\text{T}}$ and $\mathbf{I}_{2N_{\text{t}}} = \mathbf{I}_{N_{\text{t}}} \otimes \mathbf{I}_2$. Now, the desired equality to be proven can be written as

$$(\bar{\mathbf{u}}^{\text{T}} \otimes \mathbf{A}_i)\,\mathbf{Q}_i(\bar{\mathbf{u}} \otimes \mathbf{A}_i^{\text{T}}) = \frac{1}{2}\,\xi_i^2\,(\bar{\mathbf{u}}^{\text{T}} \otimes \mathbf{A}_i)(\bar{\mathbf{u}} \otimes \mathbf{A}_i^{\text{T}}), \tag{C.2}$$

Using the property $(\bar{\mathbf{u}}^{\text{T}} \otimes \mathbf{A}_i)(\bar{\mathbf{u}} \otimes \mathbf{A}_i^{\text{T}}) = (\bar{\mathbf{u}}^{\text{T}}\bar{\mathbf{u}}) \otimes (\mathbf{A}_i\mathbf{A}_i^{\text{T}})$, equivalently, it is desired that

$$(\bar{\mathbf{u}}^{\text{T}} \otimes \mathbf{A}_i)\,\mathbf{Q}_i(\bar{\mathbf{u}} \otimes \mathbf{A}_i^{\text{T}}) = \frac{1}{2}\,\xi_i^2\,\|\bar{\mathbf{u}}\|^2(\mathbf{A}_i\mathbf{A}_i^{\text{T}}), \tag{C.3}$$

We proceed by focusing on the left-hand side of (C.3). Let us denote $(\bar{\mathbf{u}}^{\text{T}} \otimes \mathbf{A}_i)\mathbf{Q}_i(\bar{\mathbf{u}} \otimes \mathbf{A}_i^{\text{T}}) \triangleq \mathbf{G} = [g_{kj}]_{2\times 2}$ where $j, k = 1, 2$, and $\mathbf{u}_{\text{R}} \triangleq \text{Re}(\mathbf{u})$ and $\mathbf{u}_{\text{I}} \triangleq \text{Im}(\mathbf{u})$ such that $\bar{\mathbf{u}}^{\text{T}} = [\mathbf{u}_{\text{R}}^{\text{T}}, \mathbf{u}_{\text{I}}^{\text{T}}]$. Thus, considering $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}]^{\text{T}}$, we have

$$\mathbf{G} = \frac{1}{2}\,\xi_i^2 \begin{bmatrix} \mathbf{u}_{\text{R}}^{\text{T}} \otimes \mathbf{a}_{i,1}^{\text{T}} & \mathbf{u}_{\text{I}} \otimes \mathbf{a}_{i,1}^{\text{T}} \\ \mathbf{u}_{\text{R}}^{\text{T}} \otimes \mathbf{a}_{i,2}^{\text{T}} & \mathbf{u}_{\text{I}} \otimes \mathbf{a}_{i,2}^{\text{T}} \end{bmatrix} \times \begin{bmatrix} \mathbf{I}_{N_{\text{t}}} \otimes \mathbf{I}_2 & \mathbf{I}_{N_{\text{t}}} \otimes \mathbf{J}_2 \\ \mathbf{I}_{N_{\text{t}}} \otimes \mathbf{J}_2^{\text{T}} & \mathbf{I}_{N_{\text{t}}} \otimes \mathbf{I}_2 \end{bmatrix} \times \begin{bmatrix} \mathbf{u}_{\text{R}} \otimes \mathbf{a}_{i,1} & \mathbf{u}_{\text{R}} \otimes \mathbf{a}_{i,2} \\ \mathbf{u}_{\text{I}} \otimes \mathbf{a}_{i,1} & \mathbf{u}_{\text{I}} \otimes \mathbf{a}_{i,2} \end{bmatrix}. \tag{C.4}$$

For the sake of simplicity, the term $\frac{1}{2}\,\xi_i^2$ is omitted from the next equation, but it will appear in the final derivation. The matrix multiplication in the right-hand side of (C.4) can be evaluated and simplified as

$$g_{11} = \left(\mathbf{u}_{\mathrm{R}}^{\mathrm{T}}\mathbf{u}_{\mathrm{R}} + \mathbf{u}_{\mathrm{I}}^{\mathrm{T}}\mathbf{u}_{\mathrm{I}}\right) \mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{a}_{i,1} + 2\,\mathbf{u}_{\mathrm{R}}^{\mathrm{T}}\mathbf{u}_{\mathrm{I}} \otimes \mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{J}_2\mathbf{a}_{i,1}, \tag{C.5a}$$

$$g_{12} = g_{21} = \left(\mathbf{u}_{\mathrm{R}}^{\mathrm{T}}\mathbf{u}_{\mathrm{R}} + \mathbf{u}_{\mathrm{I}}^{\mathrm{T}}\mathbf{u}_{\mathrm{I}}\right) \mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{a}_{i,2} + 2\,\mathbf{u}_{\mathrm{R}}^{\mathrm{T}}\mathbf{u}_{\mathrm{I}} \otimes \left(\mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{J}_2\mathbf{a}_{i,2} + \mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{J}_2^{\mathrm{T}}\mathbf{a}_{i,2}\right), \tag{C.5b}$$

$$g_{22} = \left(\mathbf{u}_{\mathrm{R}}^{\mathrm{T}}\mathbf{u}_{\mathrm{R}} + \mathbf{u}_{\mathrm{I}}^{\mathrm{T}}\mathbf{u}_{\mathrm{I}}\right) \mathbf{a}_{i,2}^{\mathrm{T}}\mathbf{a}_{i,2} + 2\,\mathbf{u}_{\mathrm{R}}^{\mathrm{T}}\mathbf{u}_{\mathrm{I}} \otimes \mathbf{a}_{i,2}^{\mathrm{T}}\mathbf{J}_2\mathbf{a}_{i,2}, \tag{C.5c}$$

where in simplifications, we have frequently used the fact that $(\mathbf{X} \otimes \mathbf{Y})(\mathbf{W} \otimes \mathbf{Z}) = (\mathbf{XW} \otimes \mathbf{YZ})$, for any given matrices $\mathbf{X}, \mathbf{Y}, \mathbf{W}, \mathbf{Z}$ with appropriate dimensions. It is then easy to verify that $\mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{J}_2\mathbf{a}_{i,1} = \mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{J}_2^{\mathrm{T}}\mathbf{a}_{i,1} = \mathbf{0}$, and further $\mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{J}_2\mathbf{a}_{i,2} + \mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{J}_2^{\mathrm{T}}\mathbf{a}_{i,2} = \mathbf{a}_{i,1}^{\mathrm{T}}(\mathbf{J}_2 + \mathbf{J}_2^{\mathrm{T}})\mathbf{a}_{i,2} = \mathbf{0}$. Moreover, it directly follows from the definition of $\bar{\mathbf{u}}$ that $\mathbf{u}_{\mathrm{R}}^{\mathrm{T}}\mathbf{u}_{\mathrm{R}} + \mathbf{u}_{\mathrm{I}}^{\mathrm{T}}\mathbf{u}_{\mathrm{I}} = \bar{\mathbf{u}}^{\mathrm{T}}\bar{\mathbf{u}}$. Applying all these notes to (C.5a)-(C.5c), the entries of $\mathbf{G}$ are obtained as

$$g_{11} = \|\bar{\mathbf{u}}\|^2\|\mathbf{a}_{i,1}\|^2, \tag{C.6a}$$

$$g_{12} = g_{21} = \|\bar{\mathbf{u}}\|^2\,\mathbf{a}_{i,1}^{\mathrm{T}}\mathbf{a}_{i,2}, \tag{C.6b}$$

$$g_{22} = \|\bar{\mathbf{u}}\|^2\|\mathbf{a}_{i,2}\|^2. \tag{C.6c}$$

Merging the results in (C.6) yields

$$\mathbf{G} = \frac{1}{2}\,\xi_i^2\,\|\bar{\mathbf{u}}\|^2(\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}}), \tag{C.7}$$

as required.

C.2 Derivation of an equivalent SOC formulation for safe approximation II

The derivation is essentially based on Lemma 9. We denote

$$\mathbf{X} \triangleq \begin{bmatrix} -\frac{\bar{w}_{i,1}}{\psi(v)} & 0 \\ 0 & -\frac{\bar{w}_{i,2}}{\psi(v)} \end{bmatrix}, \quad \mathbf{Y} \triangleq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{Z} \triangleq \begin{bmatrix} -\frac{\bar{w}_{i,1}}{\psi(v)} & 0 & 0 & 0 \\ 0 & -\frac{\bar{w}_{i,2}}{\psi(v)} & 0 & 0 \\ 0 & 0 & -\frac{\bar{w}_{i,1}}{\psi(v)} & 0 \\ 0 & 0 & 0 & -\frac{\bar{w}_{i,2}}{\psi(v)} \end{bmatrix}.$$

Accordingly, the constraint (6.43) can be equivalently implied by the following two semidefinite restrictions:

$$\mathbf{X} \succeq 0, \tag{C.8a}$$

$$\mathbf{Z} - \mathbf{Y}^{\mathrm{T}}\mathbf{X}^{-1}\mathbf{Y} \succeq 0. \tag{C.8b}$$

The second restriction in (C.8b), after doing the matrix products and some simple algebra, can be written as

$$\begin{bmatrix} -\frac{\bar{w}_{i,1}}{\psi(v)} + \frac{\psi(v)}{\bar{w}_{i,1}} & 0 & 0 & 0 \\ 0 & -\frac{\bar{w}_{i,1}}{\psi(v)} & 0 & 0 \\ 0 & 0 & -\frac{\bar{w}_{i,2}}{\psi(v)} & 0 \\ 0 & 0 & 0 & -\frac{\bar{w}_{i,2}}{\psi(v)} + \frac{\psi(v)}{\bar{w}_{i,2}} \end{bmatrix} \succeq 0. \tag{C.9}$$

from which it is clear that (C.8b) further implies the restriction $\mathbf{X} \succeq 0$, hence it is necessary and sufficient for (6.43). We then rearrange (C.9) in a more convenient form and decompose it into two semidefinite constraints as

$$\frac{-1}{\psi(v)}\,\mathbf{D}_{\bar{\mathbf{w}}_i} \succeq 0, \tag{C.10a}$$

$$\frac{-1}{\psi(v)}\,\mathbf{D}_{\bar{\mathbf{w}}_i} + \psi(v)\,\mathbf{D}_{\bar{\mathbf{w}}_i}^{-1} \succeq 0, \tag{C.10b}$$

with $\mathbf{D}_{\bar{\mathbf{w}}_i} \triangleq \operatorname{diag}(\bar{\mathbf{w}}_i)$. Note that the restriction (C.10a) is in fact equivalent to $\mathbf{D}_{\bar{\mathbf{w}}_i} \preceq 0$ or $\bar{\mathbf{w}}_i \preceq \mathbf{0}$, which is implied by the constraint $\mathbf{w}_i \preceq \mathbf{0}$; see Remark 1. Further, note that $\operatorname{erfc}(\cdot)$ is non-negative in the interval $(0, 1]$, so is $\psi(v)$. Now, multiplying both sides of (C.10b) by $\mathbf{D}_{\bar{\mathbf{w}}_i}$, and imposing the restriction (C.10a) which changes the direction of the inequality, both of the constraints (C.10b) and (C.10a) can be simultaneously expressed by

$$\frac{-1}{\psi(v)}\,\mathbf{D}_{\bar{\mathbf{w}}_i}^2 + \psi(v)\,\mathbf{I} \preceq 0. \tag{C.11}$$

Since $\mathbf{D}_{\bar{\mathbf{w}}_i} \preceq 0$ and diagonal, from (C.11) by taking square root, we obtain

$$\frac{1}{\psi(v)}\,\mathbf{D}_{\bar{\mathbf{w}}_i} + \mathbf{I} \preceq 0, \tag{C.12}$$

which can be written in the vector form as

$$\frac{-1}{\psi(v)}\,\bar{\mathbf{w}}_i \succeq \mathbf{1}. \tag{C.13}$$

Replacing $\bar{\mathbf{w}}_i$ with $(\sqrt{2}/\xi_i \|\bar{\mathbf{u}}\|)(\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}})$, it is then routine to show that (C.13) is equivalent to

$$\|\bar{\mathbf{u}}\|\,\mathbf{1} \preceq \frac{-\sqrt{2}}{\psi(v)\,\xi_i}(\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}})^{-1/2}\mathbf{w}_i(\bar{\mathbf{u}}), \tag{C.14}$$

Appendix D

Appendices for Chapter 7

D.1 Proof of Lemma 14

Let $\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathrm{T}}$ denote the spectral decomposition of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_{2N_{\mathrm{t}}})$ is a diagonal matrix containing the eigenvalues of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ with $\lambda_j \geq 0$ denoting the jth eigenvalue for $j = 1, 2, ..., 2N_{\mathrm{t}}$, and $\mathbf{Q}$ is a unitary matrix such that $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{I}$ with columns being the corresponding eigenvectors of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$. It then immediately follows that

$$\begin{aligned}\mathbf{P} &= \mathbf{H}^{\mathrm{T}}\mathbf{H} + \frac{1}{\beta}\mathbf{I} \\ &= \mathbf{Q}\left(\boldsymbol{\Lambda} + \frac{1}{\beta}\mathbf{I}\right)\mathbf{Q}^{\mathrm{T}} \\ &= \mathbf{Q}\bar{\boldsymbol{\Lambda}}\mathbf{Q}^{\mathrm{T}},\end{aligned} \tag{D.1}$$

where $\bar{\boldsymbol{\Lambda}} \triangleq \boldsymbol{\Lambda} + (1/\beta)\mathbf{I} = (\bar{\lambda}_1, \bar{\lambda}_2, ..., \bar{\lambda}_{2N_{\mathrm{t}}})$. Replacing $\mathbf{P}$ with its spectral decomposition, we can rewrite $f(\mu)$ as

$$f(\mu) = \left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{\mathrm{T}}\mathbf{Q}\left(\bar{\boldsymbol{\Lambda}} - \mu\mathbf{I}\right)^{-2}\mathbf{Q}^{\mathrm{T}}\left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}\right) - \varepsilon^2. \tag{D.2}$$

Let $\mathbf{y} \triangleq \mathbf{Q}^{\mathrm{T}}\left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}\right)$, where $\mathbf{y} = [y_1, y_2, ..., y_{2N_{\mathrm{t}}}]^{\mathrm{T}}$. Therefore, function $f(\mu)$ can be expressed as

$$\begin{aligned}f(\mu) &= \mathbf{y}^{\mathrm{T}}(\bar{\boldsymbol{\Lambda}} - \mu\mathbf{I})^{-2}\mathbf{y} - \varepsilon^2 \\ &= \sum_{j=1}^{2N_{\mathrm{t}}}\frac{y_j^2}{(\bar{\lambda}_j - \mu)^2} - \varepsilon^2\end{aligned} \tag{D.3}$$

Notice that $f(-\infty) = f(+\infty) = -\varepsilon^2$. This, along with the fact $\lim_{\mu\to\bar{\lambda}_j} f(\mu) = +\infty$ for all $j \in \{1, 2, ..., 2N_{\mathrm{t}}\}$, implies that function $f(\mu)$ has at least two roots $\mu_1 < \bar{\lambda}_{\min}$ and

$\mu_2 > \bar{\lambda}_{\max}$, where

$$\bar{\lambda}_{\min} \triangleq \lambda_{\min}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + (1/\beta)\mathbf{I}\right) = \lambda_{\min}(\mathbf{H}^{\mathrm{T}}\mathbf{H}) + 1/\beta,$$

and

$$\bar{\lambda}_{\max} \triangleq \lambda_{\max}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + (1/\beta)\mathbf{I}\right) = \lambda_{\max}(\mathbf{H}^{\mathrm{T}}\mathbf{H}) + 1/\beta,$$

respectively denote the minimum and the maximum eigenvalue of $\mathbf{H}^{\mathrm{T}}\mathbf{H} + (1/\beta)\mathbf{I}$. Further,

$$\frac{\partial f}{\partial \mu} \triangleq f'(\mu) = \sum_{j=1}^{2N_{\mathrm{t}}} \frac{2y_j^2}{(\bar{\lambda}_j - \mu)^3}. \tag{D.4}$$

It can simply be verified that $f'(\mu) > 0$ for all $\mu < \bar{\lambda}_{\min}$ while $f'(\mu) < 0$ for $\mu > \bar{\lambda}_{\max}$. As a consequence, $f(\mu)$ has exactly one root within each interval $(-\infty, \bar{\lambda}_{\min})$ and $(\bar{\lambda}_{\max}, +\infty)$. On the other hand, we have

$$\frac{\partial^2 f}{\partial \mu^2} \triangleq f''(\mu) = \sum_{j=1}^{2N_{\mathrm{t}}} \frac{6y_i^2}{(\bar{\lambda}_j - \mu)^4}. \tag{D.5}$$

It can simply be verified that $f''(\mu) \geq 0$, with $f''(\mu) = 0$ only if $y_j^2 = 0$ for all $i = 1, 2, ..., 2N_{\mathrm{t}}$, i.e., $\mathbf{y} = \mathbf{0}$. However, we show that $\mathbf{y} \neq \mathbf{0}$ always hold true as follows. We have $\mathrm{rank}(\mathbf{Q}^{\mathrm{T}}) = \mathrm{rank}(\mathbf{H}^{\mathrm{T}}) = 2N_{\mathrm{t}}$, i.e., $\mathbf{Q}^{\mathrm{T}}$ and $\mathbf{H}^{\mathrm{T}}$ are full column rank matrices with high probability, and thus, $\mathbf{Q}^{\mathrm{T}}\left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}\right) = \mathbf{0}$ only when $\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi} = \mathbf{0}$. However, $\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi} = \mathbf{0}$ yields $\mathbf{w}^* = \mathbf{0}$ which contradicts the optimality condition $\|\mathbf{w}^*\| = \varepsilon$, and hence, $\mathbf{y} \neq \mathbf{0}$ always holds true at the optimum, and $f''(\mu) > 0$. From the definition of $f(\mu)$ in (D.2), it further follows that

$$\lim_{\mu \to \bar{\lambda}_j^+} f(\mu) = \lim_{\mu \to \bar{\lambda}_j^-} f(\mu).$$

This means that $f(\mu)$ is shaped as a parabolic function in the interval between any two consecutive $\bar{\lambda}_j$ for all $j = 1, 2, ..., 2N_{\mathrm{t}}$. Thus, $f(\mu)$ can have at most two roots corresponding to each eigenvalue $\bar{\lambda}_j$. Putting all these together along with the fact $\mathrm{rank}(\mathbf{H}^{\mathrm{T}}\mathbf{H}) = \mathrm{rank}(\mathbf{H})$, we can deduce that $f(\mu)$ always has an even number of roots bounded as $2 \leq R \leq 2\,\mathrm{rank}(\mathbf{H})$.

D.2 Proof of Theorem 15

The proof is composed of two parts. First, we show that $\mathbf{w}^*$ is unique for the largest positive root of $f(\mu)$, and then we obtain upper and lower bounds for this unique root.

Let $\mathcal{R} \triangleq \{\mu_l \,|\, l = 1, ..., 2\mathrm{rank}(\mathbf{H})\}$ denote the set of roots of $f(\mu)$, including μ_1 and μ_2 such that $\mu_1 < \bar{\lambda}_{\min}$, $\mu_2 > \bar{\lambda}_{\max}$, and $\bar{\lambda}_{\min} < \mu_l < \bar{\lambda}_{\max}$ for $l \neq 1, 2$. Note that in scenarios with an equi-power channel where, no such μ_l exist. We further denote by

$$g(\mu) \triangleq \beta\, \bar{\boldsymbol{\Delta}}^{\mathrm{T}} \mathbf{P} \bar{\boldsymbol{\Delta}} + 2\beta\, \bar{\boldsymbol{\Delta}}^{\mathrm{T}} (\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}), \tag{D.6}$$

those terms of the objective function (7.7) that depend on $\bar{\boldsymbol{\Delta}}$. In order to obtain the maximizer of (7.7), we need to evaluate and compare $g(\mu)$ among all $\mu \in \mathcal{R}$ and opt for the largest value of $g(\mu)$. For this purpose, by substituting the spectral decomposition $\mathbf{Q}\bar{\boldsymbol{\lambda}}\mathbf{Q}^{\mathrm{T}}$ for $\mathbf{P}$, replacing $\bar{\boldsymbol{\Delta}}$ from (7.10) and ignoring the constant multiplier β, we rewrite $g(\mu)$ as

$$g(\mu) = \mathbf{y}^{\mathrm{T}} \left(\bar{\boldsymbol{\Lambda}} - \mu\mathbf{I}\right)^{-2} \bar{\boldsymbol{\Lambda}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}} \left(\bar{\boldsymbol{\Lambda}} - \mu\mathbf{I}\right)^{-1} \mathbf{y}, \tag{D.7}$$

or equally,

$$g(\mu) = \sum_{j=1}^{2N_{\mathrm{t}}} \left(\frac{\bar{\lambda}_j y_j^2}{(\bar{\lambda}_j - \mu)^2} - \frac{2y_j^2}{\bar{\lambda}_j - \mu} \right), \tag{D.8}$$

where $\mathbf{y} \triangleq \mathbf{Q}^{\mathrm{T}} \left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\Phi}\right)$, with $\mathbf{y} = [y_1, y_2 ..., y_{2N_{\mathrm{t}}}]^{\mathrm{T}}$. Using the fact that

$$\sum_{j=1}^{2N_{\mathrm{t}}} \frac{y_j^2}{(\bar{\lambda}_j - \mu)^2} = \varepsilon^2, \quad \forall \mu \in \mathcal{R}, \tag{D.9}$$

we can evaluate $g(\mu)$ for $\mu \in \mathcal{R}$ as

$$\begin{aligned} g(\mu) &= \sum_{j=1}^{2N_{\mathrm{t}}} \left(\frac{\bar{\lambda}_j y_j^2}{(\bar{\lambda}_j - \mu)^2} - \frac{2y_j^2}{\bar{\lambda}_j - \mu} \right) \\ &= \sum_{j=1}^{2N_{\mathrm{t}}} \frac{(2\mu - \bar{\lambda}_j) y_j^2}{(\bar{\lambda}_j - \mu)^2} \\ &= 2\mu\varepsilon^2 - \sum_{j=1}^{2N_{\mathrm{t}}} \frac{\bar{\lambda}_j y_j^2}{(\bar{\lambda}_j - \mu)^2}. \end{aligned} \tag{D.10}$$

Since $|\bar{\lambda}_j - \mu_l| < |\bar{\lambda}_j - \mu_2|$ for all $j = 1, 2, ..., 2N_{\mathrm{t}}$ and $\mu_2 \geq \mu_l$ for all $\mu_l \in \mathcal{R}$, it readily follows from (D.10) that

$$g(\mu_2) \geq g(\mu_l), \quad \forall \mu_l \in \mathcal{R}. \tag{D.11}$$

On the other hand, considering $\mu_1 < \bar{\lambda}_{\min}$, we obtain

$$\begin{aligned} g(\mu_1) &= 2\mu_1\varepsilon^2 - \sum_{j=1}^{2N_{\mathrm{t}}} \frac{\bar{\lambda}_j y_j^2}{(\bar{\lambda}_j - \mu_1)^2} \\ &< 2\bar{\lambda}_{\min}\varepsilon^2 - \bar{\lambda}_{\min} \sum_{j=1}^{2N_{\mathrm{t}}} \frac{y_j^2}{(\bar{\lambda}_j - \mu_1)^2} = \bar{\lambda}_{\min}\varepsilon^2. \end{aligned} \tag{D.12}$$

Furthermore, evaluating $g(\mu)$ at μ_2 yields

$$\begin{aligned} g(\mu_2) &= 2\mu_2\varepsilon^2 - \sum_{j=1}^{2N_\mathrm{t}} \frac{\bar{\lambda}_j y_j^2}{(\bar{\lambda}_j - \mu_2)^2} \\ &> 2\bar{\lambda}_\mathrm{max}\varepsilon^2 - \bar{\lambda}_\mathrm{max} \sum_{j=1}^{2N_\mathrm{t}} \frac{y_j^2}{(\bar{\lambda}_j - \mu_2)^2} = \bar{\lambda}_\mathrm{max}\varepsilon^2. \end{aligned} \tag{D.13}$$

As a consequence, it always holds true that

$$g(\mu_1) < \bar{\lambda}_\mathrm{min}\varepsilon^2 \leq \bar{\lambda}_\mathrm{max}\varepsilon^2 < g(\mu_2). \tag{D.14}$$

Therefore, μ_2 is the root of $f(\mu)$ that maximizes the objective function of (7.7). This completes the proof of the first part.

Finally, we are interested in deriving lower and upper bounds on μ^*. Such bounds would be of essential use when computing μ^* through numerical methods, e.g., a bisection search. A lower bound on $\mu_2 = \mu^*$ is simply given by $\mu^* > \bar{\lambda}_\mathrm{max}$. To obtain an upper bound, recall that $\mathbf{y}^\mathrm{T}(\bar{\mathbf{\Lambda}} - \mu^*\mathbf{I})^{-2}\mathbf{y} = \varepsilon^2$, from which by applying $|\bar{\lambda}_\mathrm{max} - \mu^*| \leq |\bar{\lambda}_j - \mu^*|$ for all $j = 1, 2, ..., 2N_\mathrm{t}$, we obtain

$$\varepsilon^2 \leq \mathbf{y}^\mathrm{T}(\bar{\lambda}_\mathrm{max}\mathbf{I} - \mu^*\mathbf{I})^{-2}\mathbf{y} = \frac{\|\mathbf{y}\|^2}{(\bar{\lambda}_\mathrm{max} - \mu^*)^2}. \tag{D.15}$$

Due to the unitary invariance property, we have $\|\mathbf{y}\| = \|\mathbf{Q}^\mathrm{T}(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^\mathrm{T}\mathbf{\Phi})\| = \|\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^\mathrm{T}\mathbf{\Phi}\|$, and therefore, the inequality (D.15) yields the following upper bound:

$$\mu^* \leq \bar{\lambda}_\mathrm{max} + \frac{1}{\varepsilon}\left\|\left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^\mathrm{T}\mathbf{\Phi}\right)\right\|_2. \tag{D.16}$$

Hence, the proof is complete.

D.3 Proof of Lemma 16

Given $\bar{\mathbf{u}}$ and $\mathbf{\Phi}$, using the upper bound

$$\mu^* \leq \bar{\lambda}_\mathrm{max} + \frac{1}{\varepsilon}\left\|\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^\mathrm{T}\mathbf{\Phi}\right\|, \tag{D.17}$$

we can see that $|\mu^* - \bar{\lambda}_\mathrm{max}| \to \infty$ as $\varepsilon \to 0$ with high probability. The increasing behavior of $|\mu^* - \bar{\lambda}_\mathrm{max}|$ can also be verified as follows. The extremum point μ^* always satisfies

$$\varepsilon^2 = \sum_{j=1}^{2N_\mathrm{t}} \frac{y_j^2}{(\bar{\lambda}_j - \mu^*)^2}. \tag{D.18}$$

In case $\varepsilon \to 0$, since all the summands in (D.18) are positive, it necessarily holds true that

$$\lim_{\varepsilon \to 0} |\mu^* - \bar{\lambda}_{\max}| = \infty, \tag{D.19}$$

which yields $\mu^* \gg \bar{\lambda}_{\max}$. Recall the equation to be solved to find μ^*, given by

$$\left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\mathbf{\Phi}\right)^{\mathrm{T}} (\mathbf{P} - \mu^*\mathbf{I})^{-2} \left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\mathbf{\Phi}\right) = \varepsilon^2, \tag{D.20}$$

which can equivalently be written as

$$\left(\frac{1}{\mu^*}\right)^2 \left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\mathbf{\Phi}\right)^{\mathrm{T}} \left(\mathbf{I} - \frac{1}{\mu^*}\mathbf{P}\right)^{-2} \left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\mathbf{\Phi}\right) = \varepsilon^2, \tag{D.21}$$

As a consequence of $\mu^* \gg \bar{\lambda}_{\max}$, the scalar $1/\mu^*$ is relatively small compared to the eigenvalues of $\mathbf{I}$ and $\mathbf{P}$. Hence, the approximation $(\mathbf{I} - (1/\mu^*)\mathbf{P})^{-1} \approx \mathbf{I} + (1/\mu^*)\mathbf{P}$ might be useful [244]. Accordingly, we can write

$$\begin{aligned} \left(\mathbf{I} - \frac{1}{\mu^*}\mathbf{P}\right)^{-2} &\approx \left(\mathbf{I} + \frac{1}{\mu^*}\mathbf{P}\right)^2 \\ &= \mathbf{I} + \left(\frac{1}{\mu^*}\right)^2 \mathbf{P}^2 + \frac{2}{\mu^*}\mathbf{P} \\ &\approx \mathbf{I} + \frac{2}{\mu^*}\mathbf{P}, \end{aligned} \tag{D.22}$$

where the last approximation is obtained by ignoring the second order term which follows from $\mu^* \gg \bar{\lambda}_{\max}$. Plugging the approximation (D.21) into (D.21), we obtain

$$\left(\frac{1}{\mu^*}\right)^2 \left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\mathbf{\Phi}\right)^{\mathrm{T}} \left(\mathbf{I} + \frac{2}{\mu^*}\mathbf{P}\right) \left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\mathbf{\Phi}\right) = \varepsilon^2, \tag{D.23}$$

By replacing the spectral decomposition $\mathbf{P} = \mathbf{Q}\bar{\mathbf{\Lambda}}\mathbf{Q}^{\mathrm{T}}$ and denoting $\mathbf{y} \triangleq \mathbf{Q}^{\mathrm{T}} \left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\mathbf{\Phi}\right)$, after some straightforward algebraic steps, we can rewrite (D.23) as

$$\varepsilon^2(\mu^*)^3 - \mu^*\mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\bar{\mathbf{\Lambda}}\mathbf{y} = 0. \tag{D.24}$$

The solution to this polynomial equation can be obtained via the following lemma.

Lemma 22. *The solution to the third order polynomial equation $ax^3 + bx + c = 0$ is given by*

$$x = \frac{1}{a}\left(\frac{\theta}{18}\right)^{\frac{1}{3}} - \left(\frac{2}{3\theta}\right)^{\frac{1}{3}} b, \tag{D.25}$$

where $\theta = \sqrt{3}\sqrt{27a^4c^2 + 4a^3b^3} - 9a^2c$.

Applying Lemma 22 to the polynomial equation (D.24), we obtain

$$\mu^* = \frac{1}{3}\left(\frac{\varepsilon^2}{2\mathbf{y}^{\mathrm{T}}\bar{\mathbf{\Lambda}}\mathbf{y}}\right)^{\frac{1}{3}}\mathbf{y}^{\mathrm{T}}\mathbf{y} + 2\left(\frac{\varepsilon^2}{\mathbf{y}^{\mathrm{T}}\bar{\mathbf{\Lambda}}\mathbf{y}}\right)^{-\frac{1}{3}}. \tag{D.26}$$

Under the assumption $\varepsilon \ll 1$, we can further approximate (D.26) by a simpler expression as

$$\mu^* \approx 2\left(\frac{\mathbf{y}^{\mathrm{T}}\bar{\mathbf{\Lambda}}\mathbf{y}}{\varepsilon^2}\right)^{\frac{1}{3}}, \tag{D.27}$$

or equally,

$$\mu^* \approx 2\left(\frac{\left\|\mathbf{P}\left(\mathbf{P}\bar{\mathbf{G}}\bar{\mathbf{u}} - \mathbf{H}^{\mathrm{T}}\mathbf{\Phi}\right)\right\|}{\varepsilon}\right)^{\frac{2}{3}}, \tag{D.28}$$

as required.

Appendix E

Appendices for Chapter 8

E.1 Proof of Lemma 17

First, we prove that $x_m \in \{-1, +1\}$, for all $m = 1, 2, ..., 2bN_\mathrm{t}$. From $-\mathbf{1} \preceq \mathbf{x} \preceq \mathbf{1}$, or equally $\|\mathbf{x}\|_\infty \leq 1$, it is easy to verify that $\|\mathbf{x}\| \leq \sqrt{\|\mathbf{x}\|_1}$, using which along with $\mathbf{y}^\mathrm{T}\mathbf{y} \leq 2bN_\mathrm{t}$ and the Cauchy-Schwarz inequality, we can write the following chain of inequalities:

$$2bN_\mathrm{t} = \mathbf{x}^\mathrm{T}\mathbf{y} \leq \|\mathbf{x}\|\|\mathbf{y}\| \leq \sqrt{2bN_\mathrm{t}}\|\mathbf{x}\| \leq \sqrt{2bN_\mathrm{t}\|x\|_1}. \tag{E.1}$$

It follows from (E.1) that $\|\mathbf{x}\|_1 \geq 2bN_\mathrm{t}$. Putting $\|\mathbf{x}\|_\infty \leq 1$ and $\|\mathbf{x}\|_1 \geq 2bN_\mathrm{t}$ together, we conclude that $x_m \in \{-1, +1\}$, for all $m = 1, 2, ..., 2bN_\mathrm{t}$. This further implies that $\|\mathbf{x}\|^2 = 2bN_\mathrm{t}$. Finally, from $\mathbf{x}^\mathrm{T}\mathbf{y} = 2bN_\mathrm{t} = \|\mathbf{x}\|^2$, it is straightforward to show that $\mathbf{x} = \mathbf{y}$. Hence, the proof is complete.

E.2 Proof of Lemma 19

Given $\mathbf{t}$, function $f(\mathbf{x}, \mathbf{t})$ consists of an ℓ_2-norm function plus a linear term in $\mathbf{x}$, and hence, is continuously differentiable everywhere. Let $\mathbf{x}_1 \in \mathbb{R}^{2bN_\mathrm{t}\times 1}$ and $\mathbf{x}_2 \in \mathbb{R}^{2bN_\mathrm{t}\times 1}$ be any two distinct vector inputs to the function $f(\mathbf{x}, \mathbf{t})$ such that $-\mathbf{1} \preceq \mathbf{x}_1 \preceq \mathbf{1}$ and $-\mathbf{1} \preceq \mathbf{x}_2 \preceq \mathbf{1}$. Then, we can write

$$\begin{aligned}|f(\mathbf{x}_1, \mathbf{t}) - f(\mathbf{x}_2, \mathbf{t})| &= \Big|\mathbf{q}^\mathrm{T}\left(\mathbf{A}^{-1}\mathbf{W}\mathbf{t} - \sqrt{p}\,\mathbf{H}_\mathrm{b}\mathbf{x}_1\right) + \left\|\mathbf{Q}\left(\sqrt{p}\,\mathbf{H}_\mathrm{b}\mathbf{x}_1 - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right) - \mathbf{g}\right\|^2 \\ &\quad - \mathbf{q}^\mathrm{T}\left(\mathbf{A}^{-1}\mathbf{W}\mathbf{t} - \sqrt{p}\mathbf{H}_\mathrm{b}\mathbf{x}_2\right) + \left\|\mathbf{Q}\left(\sqrt{p}\mathbf{H}_\mathrm{b}\mathbf{x}_2 - \mathbf{A}^{-1}\mathbf{W}\mathbf{t}\right) - \mathbf{g}\right\|^2\Big| \\ &= \Big|\sqrt{p}\,\mathbf{q}^\mathrm{T}\mathbf{H}_\mathrm{b}(\mathbf{x}_2 - \mathbf{x}_1) + \|\sqrt{p}\,\mathbf{Q}\mathbf{H}_\mathrm{b}\mathbf{x}_1\|^2 - \|\sqrt{p}\,\mathbf{Q}\mathbf{H}_\mathrm{b}\mathbf{x}_2\|^2 \\ &\quad + 2\sqrt{p}\left(\mathbf{t}^\mathrm{T}\mathbf{W}^\mathrm{T}\mathbf{A}^{-T} + \mathbf{g}^\mathrm{T}\right)\mathbf{Q}\mathbf{H}_\mathrm{b}(\mathbf{x}_2 - \mathbf{x}_1)\Big|.\end{aligned} \tag{E.2}$$

Using the matrix/vector operator norm inequality, the following chain of inequalities holds true:

$$\begin{aligned} |f(\mathbf{x}_1,\mathbf{t}) - f(\mathbf{x}_2,\mathbf{t})| \; &\leq \Big| \|\sqrt{p}\,\mathbf{Q}\mathbf{H}_\mathrm{b}\mathbf{x}_1\|^2 - \|\sqrt{p}\,\mathbf{Q}\mathbf{H}_\mathrm{b}\mathbf{x}_2\|^2 \\ &\quad + \sqrt{p}\left(2\,\mathbf{t}^\mathrm{T}\mathbf{W}^\mathrm{T}\mathbf{A}^{-T}\mathbf{Q} + 2\,\mathbf{g}^\mathrm{T}\mathbf{Q} + \mathbf{q}^\mathrm{T}\right)\mathbf{H}_\mathrm{b}(\mathbf{x}_2 - \mathbf{x}_1)\Big| \\ &\overset{(a)}{\leq} \Big| \|\sqrt{p}\,\mathbf{Q}\mathbf{H}_\mathrm{b}\mathbf{x}_1\|^2 - \|\sqrt{p}\,\mathbf{Q}\mathbf{H}_\mathrm{b}\mathbf{x}_2\|^2 \Big| \\ &\quad + \left|\sqrt{p}\left(2\,\mathbf{t}^\mathrm{T}\mathbf{W}^\mathrm{T}\mathbf{A}^{-T}\mathbf{Q} + 2\,\mathbf{g}^\mathrm{T}\mathbf{Q} + \mathbf{q}^\mathrm{T}\right)\mathbf{H}_\mathrm{b}(\mathbf{x}_2 - \mathbf{x}_1)\right| \\ &\overset{(b)}{\leq} \|\sqrt{p}\,\mathbf{Q}\mathbf{H}_\mathrm{b}\mathbf{x}_1 - \sqrt{p}\,\mathbf{Q}\mathbf{H}_\mathrm{b}\mathbf{x}_2\|^2 \\ &\quad + 2\sqrt{p}\left\|\mathbf{H}_\mathrm{b}^\mathrm{T}\left(\mathbf{Q}^\mathrm{T}\mathbf{A}^{-1}\mathbf{W}\mathbf{t} + \mathbf{Q}^\mathrm{T}\mathbf{g} + \frac{1}{2}\mathbf{q}\right)\right\| \|\mathbf{x}_1 - \mathbf{x}_2\| \\ &\leq p\; \|\mathbf{Q}\mathbf{H}_\mathrm{b}\|^2 \, \|\mathbf{x}_1 - \mathbf{x}_2\|^2 \\ &\quad + 2\sqrt{p}\left\|\mathbf{H}_\mathrm{b}^\mathrm{T}\left(\mathbf{Q}\mathbf{A}^{-1}\mathbf{W}\mathbf{t} + \mathbf{Q}\boldsymbol{\Sigma}\mathbf{s} + \frac{1}{2}\mathbf{q}\right)\right\| \|\mathbf{x}_1 - \mathbf{x}_2\| \end{aligned} \tag{E.3}$$

where the inequalities (a) and (b) directly follow from the (reverse) triangle inequality. Furthermore, in deriving the last inequality, we have used the facts that $\mathbf{Q}^\mathrm{T}\mathbf{Q} = \mathbf{Q}^\mathrm{T} = \mathbf{Q}$ and $\mathbf{Q}^\mathrm{T}\mathbf{q} = \mathbf{0}$, and therefore, $\mathbf{Q}^\mathrm{T}\mathbf{g} = \mathbf{Q}\boldsymbol{\Sigma}\mathbf{s}$; these equalities can be verified using the definitions of $\mathbf{Q}$, $\mathbf{q}$ and $\mathbf{g}$. On the other hand, since $-\mathbf{1} \preceq \mathbf{x}_1 \preceq \mathbf{1}$ and $-\mathbf{1} \preceq \mathbf{x}_2 \preceq \mathbf{1}$, we always have

$$\|\mathbf{x}_1 - \mathbf{x}_2\| \leq 2\sqrt{2bN_\mathrm{t}}, \tag{E.4}$$

where equality is achieved when either $\mathbf{x}_1$ or $\mathbf{x}_2$ is equal to $\mathbf{1}$ while the other equals $-\mathbf{1}$. Using (E.4), form the last inequality in (E.3), we obtain the following upper bound:

$$\begin{aligned} |f(\mathbf{x}_1,\mathbf{t}) - f(\mathbf{x}_2,\mathbf{t})| \; &\leq 2p\,\sqrt{2bN_\mathrm{t}}\,\|\mathbf{Q}\mathbf{H}_\mathrm{b}\|^2\,\|\mathbf{x}_1 - \mathbf{x}_2\| \\ &\quad + 2\sqrt{p}\left\|\mathbf{H}_\mathrm{b}^\mathrm{T}\left(\mathbf{Q}\mathbf{A}^{-1}\mathbf{W}\mathbf{t} + \mathbf{Q}\boldsymbol{\Sigma}\mathbf{s} + \frac{1}{2}\mathbf{q}\right)\right\| \|\mathbf{x}_1 - \mathbf{x}_2\| \end{aligned} \tag{E.5}$$

It immediately follows that $|f(\mathbf{x}_1,\mathbf{t}) - f(\mathbf{x}_2,\mathbf{t})| \,/\, \|\mathbf{x}_1 - \mathbf{x}_2\|_2$ can be bounded from above by

$$L \triangleq 2p\sqrt{2bN_\mathrm{t}}\,\|\mathbf{Q}\mathbf{H}_\mathrm{b}\|^2 + 2\sqrt{p}\left\|\mathbf{H}_\mathrm{b}^\mathrm{T}\left(\mathbf{Q}\mathbf{A}^{-1}\mathbf{W}\mathbf{t} + \mathbf{Q}\boldsymbol{\Sigma}\mathbf{s} + \frac{1}{2}\mathbf{q}\right)\right\|, \tag{E.6}$$

where L is a positive real constant if $\mathbf{Q}\mathbf{H}_\mathrm{b} \neq \mathbf{0}$ and/or $\mathbf{H}_\mathrm{b}^\mathrm{T}\left(\mathbf{Q}\mathbf{A}^{-1}\mathbf{W}\mathbf{t} + \mathbf{Q}\boldsymbol{\Sigma}\mathbf{s} + \mathbf{q}/2\right) \neq \mathbf{0}$, implying that $f(\mathbf{x},\mathbf{t})$ is Lipschitz continuous on $-\mathbf{1} \preceq \mathbf{x} \preceq \mathbf{1}$ with constant L.

E.3 Derivation of an approximate upper bound on L

Let us begin from Lemma 19 where the exact value of L is provided. It then follows from (8.28) that

$$\begin{aligned} L &\leq 2p\sqrt{2bN_{\mathrm{t}}}\,\|\mathbf{Q}\|^2\,\|\mathbf{H}_{\mathrm{b}}\|^2 + 2\sqrt{p}\,\|\mathbf{H}_{\mathrm{b}}\|\,\left\|\mathbf{Q}\mathbf{A}^{-1}\mathbf{W}\mathbf{t} + \mathbf{Q}\boldsymbol{\Sigma}\mathbf{s} + \frac{1}{2}\mathbf{q}\right\| \\ &= 2p\sqrt{2bN_{\mathrm{t}}}\,\|\mathbf{Q}\|^2\,\|\mathbf{H}_{\mathrm{b}}\|^2 + \mathcal{O}\left(\|\mathbf{H}_{\mathrm{b}}\|\right), \end{aligned} \tag{E.7}$$

where the equality holds considering the fact that in the second expression, only $\|\mathbf{H}_{\mathrm{b}}\|$ scales with N_{t} while the other terms scale only with N_{u}. Thus, in the limiting case with $N_{\mathrm{t}} \to \infty$, we have

$$\begin{aligned} L &\lesssim \mathcal{O}\left(2p\sqrt{2bN_{\mathrm{t}}}\,\|\mathbf{Q}\|^2\,\|\mathbf{H}_{\mathrm{b}}\|^2\right) \\ &= \mathcal{O}\left(2p\sqrt{2bN_{\mathrm{t}}}\,\lambda_{\max}\left(\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\right)\lambda_{\max}\left(\mathbf{H}_{\mathrm{b}}\mathbf{H}_{\mathrm{b}}^{\mathrm{T}}\right)\right) \\ &= \mathcal{O}\left(2p\sqrt{2bN_{\mathrm{t}}}\,\lambda_{\max}\left(\mathbf{H}_{\mathrm{b}}\mathbf{H}_{\mathrm{b}}^{\mathrm{T}}\right)\right), \end{aligned} \tag{E.8}$$

where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of a matrix, and the last equality holds true since, by definition, $\mathbf{Q}$ is an idempotent symmetric matrix, i.e., $\mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}^2 = \mathbf{Q}$, and hence we have $\lambda_{\max}(\mathbf{Q}) = 1$. By substituting $\mathbf{b}^{\mathrm{T}} \otimes \mathbf{H}$ for $\mathbf{H}_{\mathrm{b}}$ and using some well known properties of the Kronecker product, we can write $\mathbf{H}_{\mathrm{b}}\mathbf{H}_{\mathrm{b}}^{\mathrm{T}} = (\mathbf{b}^{\mathrm{T}} \otimes \mathbf{H})(\mathbf{b}^{\mathrm{T}} \otimes \mathbf{H})^{\mathrm{T}} = (\mathbf{b}^{\mathrm{T}} \otimes \mathbf{H})(\mathbf{b} \otimes \mathbf{H}^{\mathrm{T}}) = \mathbf{b}^{\mathrm{T}}\mathbf{b} \otimes \mathbf{H}\mathbf{H}^{\mathrm{T}} = \|\mathbf{b}\|^2\mathbf{H}\mathbf{H}^{\mathrm{T}}$. As a result, we obtain

$$L \lesssim \mathcal{O}\left(2p\sqrt{2bN_{\mathrm{t}}}\,\|\mathbf{b}\|^2\lambda_{\max}\left(\mathbf{H}\mathbf{H}^{\mathrm{T}}\right)\right). \tag{E.9}$$

Recall that $\mathbf{b} = (\Delta/2)[1, ..., 2^{b-1}]^{\mathrm{T}}$ with $\Delta = 2/\left((B-1)\sqrt{2N_{\mathrm{t}}}\right)$, and therefore,

$$\|\mathbf{b}\|^2 = \frac{\Delta^2}{4}\sum_{n=1}^{b} 2^{2(n-1)} = \frac{1}{2N_{\mathrm{t}}(B-1)^2}\sum_{n'=0}^{b-1} 4^{n'}, \tag{E.10}$$

where the last equality can simply be verified using the change of variable $n-1 \to n'$. The sequence $\{4^{n'}\}_{n'=0}^{\infty}$ is a geometric series with common ration 4. For this series, the summation of the first b terms, i.e., $\sum_{n'=0}^{b-1} 4^{n'}$, is given by $(4^b-1)/3$. Replacing this in (E.10) yields

$$|\mathbf{b}\|^2 = \frac{2^b+1}{6N_{\mathrm{t}}(2^b-1)}. \tag{E.11}$$

Using (E.11), we can rewrite (E.9) as

$$L \lesssim \mathcal{O}\left(\frac{p(2^b+1)\sqrt{2b}}{3\sqrt{N_{\mathrm{t}}}(2^b-1)}\,\lambda_{\max}\left(\mathbf{H}\mathbf{H}^{\mathrm{T}}\right)\right). \tag{E.12}$$

Based on a fundamental result in random matrix theory, in the large system limit where $N_{\mathrm{t}} \to \infty$ with $N_{\mathrm{t}} \gg N_{\mathrm{u}}$, we have $\lambda_{\max}\left(\mathbf{H}\mathbf{H}^{\mathrm{T}}\right) \to 2N_{\mathrm{t}}$ with probability one [245]. Therefore, in the large system limit, an approximate upper bound on L can be obtained as

$$L \lesssim \mathcal{O}\left(\frac{2p(2^b+1)\sqrt{2bN_\mathrm{t}}}{3(2^b-1)}\right) = \mathcal{O}\left(p\sqrt{bN_\mathrm{t}}\right). \tag{E.13}$$

Appendix F

Appendices for Chapter 9

F.1 Proof of Lemma 20

Let $\mathbf{u}_{\text{BB}}$ be given, then $g(\mathbf{u}_{\text{BB}}, \mathbf{g})$ is an ℓ_2-norm function of $\mathbf{g}$, and therefore, it is continuously differentiable everywhere on $\mathbb{R}$. Suppose $\{\mathbf{g}_1, \mathbf{g}_2\} \in \mathbb{R}^{N_{\text{t}}N_{\text{l}}\times 1}$ are two inputs to $g(\mathbf{u}_{\text{BB}}, \mathbf{g})$ such that $-\mathbf{1} \preceq \{\mathbf{g}_1, \mathbf{g}_2\} \preceq \mathbf{1}$. For brevity of notation, let us further denote $\boldsymbol{\Theta} \triangleq (\mathbf{u}_{\text{BB}}^{\text{T}} \otimes \mathbf{I}_{N_{\text{t}}}) \operatorname{diag}(\operatorname{vec}(\mathbf{F}))$. We can write

$$\begin{aligned} |g(\mathbf{u}_{\text{BB}}, \mathbf{g}_1) - g(\mathbf{u}_{\text{BB}}, \mathbf{g}_2)| &= \left| \left\| \boldsymbol{\Theta}\mathbf{g}_1 + \mathbf{F}\mathbf{u}_{\text{BB}} - 2\,\mathbf{u}_{\text{FD}}^{\star} \right\|^2 - \left\| \boldsymbol{\Theta}\mathbf{g}_2 + \mathbf{F}\mathbf{u}_{\text{BB}} - 2\,\mathbf{u}_{\text{FD}}^{\star} \right\|^2 \right| \\ &= \left| \|\boldsymbol{\Theta}\mathbf{g}_1\|^2 - \|\boldsymbol{\Theta}\mathbf{g}_2\|^2 + \left(2\mathbf{u}_{\text{BB}}^{\text{H}}\mathbf{F}^{\text{H}}\boldsymbol{\Theta} - 4\,\mathbf{u}_{\text{FD}}^{\star}{}^{\text{H}}\boldsymbol{\Theta}\right)(\mathbf{g}_1 - \mathbf{g}_2) \right|. \end{aligned} \tag{F.1}$$

According to the matrix/vector operator norm inequality, the following chain of inequalities holds true:

$$\begin{aligned} |g(\mathbf{u}_{\text{BB}}, \mathbf{g}_1) - g(\mathbf{u}_{\text{BB}}, \mathbf{g}_2)| &\overset{\text{(a)}}{\leq} \left| \|\boldsymbol{\Theta}\mathbf{g}_1\|^2 - \|\boldsymbol{\Theta}\mathbf{g}_2\|^2 \right| + \left| \left(2\mathbf{u}_{\text{BB}}^{\text{H}}\mathbf{F}^{\text{H}}\boldsymbol{\Theta} - 4\mathbf{u}_{\text{FD}}^{\star}{}^{\text{H}}\boldsymbol{\Theta}\right)(\mathbf{g}_1 - \mathbf{g}_2) \right| \\ &\overset{\text{(b)}}{\leq} \|\boldsymbol{\Theta}\mathbf{g}_1 - \boldsymbol{\Theta}\mathbf{g}_2\|^2 + \left\| 2\boldsymbol{\Theta}^{\text{H}}\mathbf{F}\mathbf{u}_{\text{BB}} - 4\boldsymbol{\Theta}^{\text{H}}\mathbf{u}_{\text{FD}}^{\star} \right\| \|\mathbf{g}_1 - \mathbf{g}_2\| \\ &\leq \|\boldsymbol{\Theta}\|_{\text{F}}^2 \|\mathbf{g}_1 - \mathbf{g}_2\|^2 + \left\| 2\boldsymbol{\Theta}^{\text{H}}\mathbf{F}\mathbf{u}_{\text{BB}} - 4\boldsymbol{\Theta}^{\text{H}}\mathbf{u}_{\text{FD}}^{\star} \right\| \|\mathbf{g}_1 - \mathbf{g}_2\|, \end{aligned} \tag{F.2}$$

where inequalities (a) and (b) follow from the (reverse) triangle inequality. As $-\mathbf{1} \preceq \{\mathbf{g}_1, \mathbf{g}_2\} \preceq \mathbf{1}$, we always have

$$\|\mathbf{g}_1 - \mathbf{g}_2\| \leq 2\sqrt{N_{\text{t}}N_{\text{l}}}, \tag{F.3}$$

where equality is achieved in case either $\mathbf{g}_1$ or $\mathbf{g}_2$ is equal to $\mathbf{1}$ while the other equals $-\mathbf{1}$. Using (F.3), form the last inequality in (F.2), we obtain the following upper bound:

$$|g(\mathbf{u}_{\text{BB}}, \mathbf{g}_1) - g(\mathbf{u}_{\text{BB}}, \mathbf{g}_2)| \leq \left(2\sqrt{N_\text{t}N_\text{l}}\|\mathbf{\Theta}\|_\text{F}^2 + \left\|2\mathbf{\Theta}^\text{H}\mathbf{F}\mathbf{u}_{\text{BB}} - 4\,\mathbf{\Theta}^\text{H}\mathbf{u}_{\text{FD}}^\star\right\|\right)\|\mathbf{g}_1 - \mathbf{g}_2\|. \tag{F.4}$$

It immediately follows that

$$\frac{|g(\mathbf{u}_{\text{BB}}, \mathbf{g}_1) - g(\mathbf{u}_{\text{BB}}, \mathbf{g}_2)|}{\|\mathbf{g}_1 - \mathbf{g}_2\|} \leq 2\sqrt{N_\text{t}N_\text{l}}\|\mathbf{\Theta}\|_\text{F}^2 + 2\left\|\mathbf{\Theta}^\text{H}(\mathbf{F}\mathbf{u}_{\text{BB}} - 2\,\mathbf{u}_{\text{FD}}^\star)\right\| \triangleq L, \tag{F.5}$$

where L is a positive real constant in case $\mathbf{\Theta} \neq \mathbf{0}$ and/or $\mathbf{F}\mathbf{u}_{\text{BB}} - 2\,\mathbf{u}_{\text{FD}}^\star \neq \mathbf{0}$. This implies the Lipschitz continuity property for the function $g(\mathbf{u}_{\text{BB}}, \mathbf{g})$ on $-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}$.

Bibliography

[1] M. Bengtsson and B. Ottersten, *Handbook of Antennas in Wireless Communications*, 2001, ch. Optimal and suboptimal transmit beamforming.

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

[3] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[4] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.

[5] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[6] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, May 2010.

[7] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.

[8] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[9] Y. F. Liu, Y. H. Dai, and Z. Q. Luo, "Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1142–1157, Mar. 2011.

[10] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[11] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[12] O. Tervo, L. N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574–5588, Oct. 2015.

[13] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *1999 IEEE 49th Vehicular Technology Conference (Cat. No.99CH36363)*, vol. 1, Jul. 1999, pp. 851–856 vol.1.

[14] M. Schubert and H. Boche, "Iterative multiuser uplink and downlink beamforming under SINR constraints," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2324–2334, Jul. 2005.

[15] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wirel. Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[16] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate maximization in multi-antenna broadcast channels with linear preprocessing," *IEEE Trans. Wirel. Commun.*, vol. 5, no. 9, pp. 2338–2342, Sep. 2006.

[17] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO wireless communications.* Cambridge university press, 2007.

[18] C. Masouros and E. Alsusa, "Dynamic linear precoding for the exploitation of known interference in MIMO broadcast systems," *IEEE Trans. Wirel. Commun.*, vol. 8, no. 3, pp. 1396–1404, 2009.

[19] C. Masouros, "Correlation rotation linear precoding for MIMO broadcast communications," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 252–262, Jan. 2011.

[20] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3628–3640, Jul. 2015.

[21] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Constructive multiuser interference in symbol level precoding for the MISO downlink channel," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2239–2252, May 2015.

[22] C. Masouros, T. Ratnarajah, M. Sellathurai, C. B. Papadias, and A. K. Shukla, "Known interference in the cellular downlink: a performance limiting factor or a source of green signal power?" *IEEE Commun. Mag.*, vol. 51, no. 10, pp. 162–171, Oct. 2013.

[23] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Constructive interference for generic constellations," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 586–590, Apr. 2018.

[24] C. Masouros and E. Alsusa, "Soft linear precoding for the downlink of DS/CDMA communication systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 203–215, Jan. 2010.

[25] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Energy-efficient symbol-level precoding in multiuser MISO based on relaxed detection region," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 5, pp. 3755–3767, May 2016.

[26] ——, "Symbol-level multiuser MISO precoding for multi-level adaptive modulation," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 8, pp. 5511–5524, Aug. 2017.

[27] A. Li, C. Masouros, X. Liao, Y. Li, and B. Vucetic, "Multiplexing more data streams in the MU-MISO downlink by interference exploitation precoding," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

[28] A. Li, C. Masouros, B. Vucetic, Y. Li, and A. Lee Swindlehurst, "Interference exploitation precoding for multi-level modulations: Closed-form solutions," *IEEE Trans. Commun.*, pp. 1–1, 2020.

[29] M. Alodeh, D. Spano, A. Kalantari, C. Tsinos, D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Symbol-level and multicast precoding for multiuser multiantenna downlink: A state-of-the-art, classification and challenges," *IEEE Commun. Surveys Tutorials*, pp. 1–1, 2018.

[30] A. Li, D. Spano, J. Krivochiza, S. Domouchtsidis, C. G. Tsinos, C. Masouros, S. Chatzinotas, Y. Li, B. Vucetic, and B. Ottersten, "A tutorial on interference exploitation via symbol-level precoding: Overview, state-of-the-art and future directions," *IEEE Commun. Surveys and Tutorials*, vol. 22, no. 2, pp. 796–839, 2020.

[31] C. Masouros and E. Alsusa, "A novel transmitter-based selective-precoding technique for DS/CDMA systems," in *2007 IEEE International Conference on Communications*, 2007, pp. 2829–2834.

[32] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Interference optimization for transmit power reduction in Tomlinson-Harashima precoded MIMO downlinks," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2470–2481, 2012.

[33] A. Garcia-Rodriguez and C. Masouros, “Power-efficient Tomlinson-Harashima precoding for the downlink of multi-user MISO systems,” *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 1884–1896, 2014.

[34] C. Masouros, M. Sellathurai, and T. Ratnarajah, “Vector perturbation based on symbol scaling for limited feedback MISO downlinks,” *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 562–571, Feb. 2014.

[35] D. Kwon, W. Yeo, and D. K. Kim, “A new precoding scheme for constructive superposition of interfering signals in multiuser MIMO systems,” *IEEE Commun. Lett.*, vol. 18, no. 11, pp. 2047–2050, 2014.

[36] M. Alodeh, S. Chatzinotas, and B. Ottersten, “A multicast approach for constructive interference precoding in MISO downlink channel,” in *IEEE International Symposium on Information Theory*, 2014, pp. 2534–2538.

[37] C. Masouros and G. Zheng, “Power efficient downlink beamforming optimization by exploiting interference,” in *IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6.

[38] M. Alodeh, S. Chatzinotas, and B. Ottersten, “Energy efficient symbol-level precoding in multiuser MISO channels,” in *IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 36–40.

[39] K. L. Law and C. Masouros, “Symbol error rate minimization precoding for interference exploitation,” *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5718–5731, 2018.

[40] A. Li and C. Masouros, “Interference exploitation precoding made practical: Optimal closed-form solutions for PSK modulations,” *IEEE Trans. Wirel. Commun.*, vol. 17, no. 11, pp. 7661–7676, 2018.

[41] A. Li, C. Masouros, Y. Li, and B. Vucetic, “Interference exploitation precoding for multi-level modulations,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4679–4683.

[42] A. Li and C. Masouros, “Exploiting constructive mutual coupling in P2P MIMO by analog-digital phase alignment,” *IEEE Trans. Wirel. Commun.*, vol. 16, no. 3, pp. 1948–1962, 2017.

[43] M. Alodeh, S. Chatzinotas, and B. Ottersten, “Constructive interference through symbol level precoding for multi-level modulation,” in *IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6.

[44] A. Li and C. Masouros, “Mutual coupling exploitation for point-to-point MIMO by constructive interference,” in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[45] K. L. Law and C. Masouros, “Constructive interference exploitation for downlink beamforming based on noise robustness and outage probability,” in *2016 IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, Mar. 2016, pp. 3291–3295.

[46] A. Kalantari, C. Tsinos, M. Soltanalian, S. Chatzinotas, W. Ma, and B. Ottersten, “MIMO directional modulation M-QAM precoding for transceivers performance enhancement,” in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2017, pp. 1–5.

[47] D. Spano, M. Alodeh, S. Chatzinotas, and B. Ottersten, “Per-antenna power minimization in symbol-level precoding,” in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.

[48] Y. I. Choi, J. W. Lee, C. G. Kang, and M. Rim, “Constructive multi-user interference for symbol-level link adaptation: MMSE approach,” in *2017 IEEE Globecom Workshops (GC Wkshps)*, 2017, pp. 1–6.

[49] . T. Demir and T. E. Tuncer, “A new beamformer design method for multi-group multicasting by enforcing constructive interference,” in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 632–636.

[50] J. Krivochiza, J. Merlano-Duncan, Stefano Andrenacci, S. Chatzinotas, and B. Ottersten, “Closed-form solution for computationally efficient symbol-level precoding,” in *IEEE Global Commun. Conf. (GLOBECOM)*, Sep. 2018.

[51] C. Masouros and T. Ratnarajah, “Interference as a source of green signal power in cognitive relay assisted co-existing MIMO wireless transmissions,” *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 525–536, 2012.

[52] M. Alodeh, S. Chatzinotas, and B. Ottersten, “Symbol based precoding in the downlink of cognitive MISO channel,” in *International Conference on Cognitive Radio Oriented Wireless Networks.* Springer, 2015, pp. 370–380.

[53] F. A. Khan, C. Masouros, and T. Ratnarajah, “Interference-driven linear precoding in multiuser MISO downlink cognitive radio network,” *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2531–2543, 2012.

[54] K. L. Law, C. Masouros, and M. Pesavento, “Transmit precoding for interference exploitation in the underlay cognitive radio Z-channel,” *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3617–3631, 2017.

[55] ——, “Bivariate probabilistic constrained programming for interference exploitation in the cognitive radio,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 3390–3394.

[56] A. Li and C. Masouros, “Constructive interference beamforming for cooperative dual-hop MIMO relay systems - invited paper,” in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, 2018, pp. 1–5.

[57] S. Timotheou, G. Zheng, C. Masouros, and I. Krikidis, “Exploiting constructive interference for simultaneous wireless information and power transfer in multiuser downlink systems,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1772–1784, 2016.

[58] M. R. A. Khandaker, C. Masouros, K. Wong, and S. Timotheou, “Secure SWIPT by exploiting constructive interference and artificial noise,” *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1326–1340, Feb. 2019.

[59] G. Zheng, C. Masouros, I. Krikidis, and S. Timotheou, “Exploring green interference power for wireless information and energy transfer in the MISO downlink,” in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 149–153.

[60] S. Timotheou, G. Zheng, C. Masouros, and I. Krikidis, “Symbol-level precoding in MISO broadcast channels for SWIPT systems,” in *2016 23rd International Conference on Telecommunications (ICT)*, 2016, pp. 1–5.

[61] A. Kalantari, M. Soltanalian, S. Maleki, S. Chatzinotas, and B. Ottersten, “Directional modulation via symbol-level precoding: A way to enhance security,” *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 8, pp. 1478–1493, Dec. 2016.

[62] ——, “Secure M-PSK communication via directional modulation,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 3481–3485.

[63] M. R. A. Khandaker, C. Masouros, and K. Wong, “Constructive interference based secure precoding: A new dimension in physical layer security,” *IEEE Trans. Inform. Forensics and Security*, vol. 13, no. 9, pp. 2256–2268, Sep. 2018.

[64] R. Liu, M. Li, Q. Liu, and A. L. Swindlehurst, “Secure symbol-level precoding in MU-MISO wiretap systems,” *IEEE Trans. Inform. Forensics and Security*, vol. 15, pp. 3359–3373, 2020.

[65] M. T. Kabir, M. R. A. Khandaker, and C. Masouros, “Reducing self-interference in full duplex transmission by interference exploitation,” in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[66] ——, “Interference exploitation in full-duplex communications: Trading interference power for both uplink and downlink power savings,” *IEEE Trans. Wirel. Commun.*, vol. 17, no. 12, pp. 8314–8329, 2018.

[67] ——, “Robust energy harvesting FD transmission: Interference suppression versus exploitation,” *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1866–1869, 2018.

[68] Z. Wei and C. Masouros, “Device-centric distributed antenna transmission: Secure precoding and antenna selection with interference exploitation,” *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2293–2308, 2020.

[69] D. Spano, M. Alodeh, S. Chatzinotas, and B. Ottersten, "Faster-than-Nyquist signaling through spatio-temporal symbol-level precoding for the multiuser MISO downlink channel," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 9, pp. 5915–5928, 2018.

[70] M. Alodeh, D. Spano, S. Chatzinotas, and B. Ottersten, "Faster-than-nyquist spatiotemporal symbol-level precoding in the downlink of multiuser MISO channels," in *IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2017, pp. 3779–3783.

[71] D. Spano, S. Chatzinotas, and B. Ottersten, "Sequential spatia-temporal symbol-level precoding enabling faster-than-Nyquist signaling for multi-user MISO systems," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 827–831.

[72] P. V. Amadori and C. Masouros, "Constructive interference based constant envelope precoding," in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2016, pp. 1–5.

[73] ——, "Constant envelope precoding by interference exploitation in phase shift keying-modulated multiuser transmission," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 1, pp. 538–550, 2017.

[74] F. Liu, C. Masouros, P. V. Amadori, and H. Sun, "An efficient manifold algorithm for constructive interference based constant envelope precoding," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1542–1546, Oct. 2017.

[75] S. Domouchtsidis, C. G. Tsinos, S. Chatzinotas, and B. Ottersten, "Symbol-level precoding for low complexity transmitter architectures in large-scale antenna array systems," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 2, pp. 852–863, 2019.

[76] P. V. Amadori and C. Masouros, "Power efficient massive MU-MIMO via antenna selection for constructive interference optimization," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 1607–1612.

[77] ——, "Interference-driven antenna selection for massive multiuser MIMO," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 5944–5958, 2016.

[78] ——, "A mixed-integer programming approach to interference exploitation for massive-MIMO," in *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2018, pp. 107–112.

[79] ——, "Large scale antenna selection and precoding for interference exploitation," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4529–4542, 2017.

[80] G. Hegde, C. Masouros, and M. Pesavento, "Analog beamformer design for interference exploitation based hybrid beamforming," in *IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2018, pp. 109–113.

[81] A. Li, C. Masouros, and F. Liu, "Hybrid analog-digital precoding for interference exploitation," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 812–816.

[82] G. Hegde, C. Masouros, and M. Pesavento, "Interference exploitation-based hybrid precoding with robustness against phase errors," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 7, pp. 3683–3696, 2019.

[83] R. Liu, H. Li, and M. Li, "Symbol-level hybrid precoding in mmWave multiuser MISO systems," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1636–1639, Sep. 2019.

[84] A. Li, C. Masouros, F. Liu, and A. L. Swindlehurst, "Massive MIMO 1-bit DAC transmission: A low-complexity symbol scaling approach," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 11, pp. 7559–7575, Nov. 2018.

[85] H. Jedda, A. Mezghani, A. L. Swindlehurst, and J. A. Nossek, "Quantized constant envelope precoding with psk and qam signaling," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 12, pp. 8022–8034, Dec. 2018.

[86] H. Jedda, A. Mezghani, J. A. Nossek, and A. L. Swindlehurst, "Massive MIMO downlink 1-bit precoding with linear programming for PSK signaling," in *2017 IEEE 18th Int. Workshop Signal Process. Advances Wirel. Commun. (SPAWC)*, Jul. 2017, pp. 1–5.

[87] M. Shao, Q. Li, and W. Ma, "One-bit massive MIMO precoding via a minimum symbol-error probability design," in *IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Apr. 2018, pp. 3579–3583.

[88] C. G. Tsinos, A. Kalantari, S. Chatzinotas, and B. Ottersten, "Symbol-level precoding with low resolution DACs for large-scale array MU-MIMO systems," in *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5.

[89] A. Li, C. Masouros, and A. L. Swindlehurst, "1-bit massive MIMO downlink based on constructive interference," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 927–931.

[90] A. Li, F. Liu, C. Masouros, Y. Li, and B. Vucetic, "Interference exploitation 1-bit massive MIMO precoding: A partial branch-and-bound solution with near-optimal performance," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 5, pp. 3474–3489, 2020.

[91] D. Spano, M. Alodeh, S. Chatzinotas, and B. Ottersten, "Symbol-level precoding for the nonlinear multiuser MISO downlink channel," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1331–1345, 2018.

[92] D. Spano, M. Alodeh, S. Chatzinotas, J. Krause, and B. Ottersten, "Spatial PAPR reduction in symbol-level precoding for the multi-beam satellite downlink," in *2017*

IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2017, pp. 1–5.

[93] D. Spano, M. Alodeh, S. Chatzinotas, and B. Ottersten, "PAPR minimization through spatio-temporal symbol-level precoding for the non-linear multi-user MISO channel," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 3599–3603.

[94] D. Spano, M. Alodeh, S. Chatzinotas, and B. Ottersten, "Per-antenna power minimization in symbol-level precoding," in *IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[95] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Symbol-level precoding design based on distance preserving constructive interference regions," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5817–5832, Nov. 2018.

[96] ——, "Symbol-level precoding design for max-min SINR in multiuser MISO broadcast channels," in *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2018, pp. 1–5.

[97] ——, "Power minimizer symbol-level precoding: A closed-form suboptimal solution," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1730–1734, Nov. 2018.

[98] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "An approximate solution for symbol-level multiuser precoding using support recovery," in *2019 IEEE 20th Int. Workshop on Signal Process. Advances in Wirel. Commun. (SPAWC)*, Jul. 2019, pp. 1–5.

[99] A. Haqiqatnejad, F. Kayhan, J. Krivochiza, J. Merlano Duncan, S. Chatzinotas, and B. Ottersten, "Design optimization for low-complexity FPGA implementation of symbol-level multiuser precoding," 2021.

[100] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Robust sinr-constrained symbol-level multiuser precoding with imperfect channel knowledge," *IEEE Trans. Signal Process.*, vol. 68, pp. 1837–1852, 2020.

[101] ——, "Robust design of power minimizing symbol-level precoder under channel uncertainty," in *2018 IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[102] A. Haqiqatnejad, S. Shahbazpanahi, and B. Ottersten, "A worst-case performance optimization based design approach to robust symbol-level precoding for downlink MU-MIMO," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019, pp. 1–5.

[103] A. Haqiqatnejad, F. Kayhan, S. ShahbazPanahi, and B. Ottersten, "Finite-alphabet symbol-level multiuser precoding for massive MU-MIMO downlink," 2020.

[104] A. Haqiqatnejad, F. Kayhan, S. Shahbazpanahi, and B. Ottersten, “One-bit quantized constructive interference based precoding for massive multiuser MIMO downlink,” in *IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[105] A. Li and C. Masouros, “Exploiting constructive mutual coupling in P2P MIMO by analog-digital phase alignment,” *IEEE Trans. Wirel. Commun.*, vol. 16, no. 3, pp. 1948–1962, Mar. 2017.

[106] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, “Energy-efficient hybrid symbol-level precoding for large-scale mmWave multiuser MIMO systems,” 2021.

[107] ——, “Energy-efficient hybrid symbol-level precoding via phase shifter selection in mmWave MU-MIMO] SYSTEMS, year=2020, volume=, number=, pages=1-6, doi=,” in *IEEE Global Communications Conference (GLOBECOM)*.

[108] C. Masouros, M. Sellathurai, and T. Ratnarajah, “Interference optimization for transmit power reduction in Tomlinson-Harashima precoded MIMO downlinks,” *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2470–2481, 2012.

[109] A. Garcia-Rodriguez and C. Masouros, “Power-efficient Tomlinson-Harashima precoding for the downlink of multi-user MISO systems,” *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 1884–1896, 2014.

[110] P. V. Amadori and C. Masouros, “Constant envelope precoding by interference exploitation in phase shift keying-modulated multiuser transmission,” *IEEE Trans. Wirel. Commun.*, vol. 16, no. 1, pp. 538–550, Jan. 2017.

[111] Y. Liu and W.-K. Ma, “Symbol-level precoding is symbol-perturbed zf when energy efficiency is sought,” *arXiv:1803.05094 [cs.IT]*, Mar. 2018.

[112] J. Krivochiza, A. Kalantari, S. Chatzinotas, and B. Ottersten, “Low complexity symbol-level design for linear precoding systems,” in *Symposium on Information Theory and Signal Processing in the Benelux*, Mar. 2016.

[113] J. C. Merlano-Duncan, J. Krivochiza, S. Andrenacci, S. Chatzinotas, and B. Ottersten, “Computationally efficient symbol-level precoding communications demonstrator,” in *2017 IEEE 28th Annual Int. Symp. Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, 2017, pp. 1–5.

[114] J. Duncan, J. Krivochiza, S. Andrenacci, S. Chatzinotas, and B. Ottersten, “Hardware demonstration of precoded communications in multi-beam UHTS systems,” in *36th Int. Commun. Satellite Syst. Conf. (ICSSC 2018)*, 2018, pp. 1–5.

[115] J. Krivochiza, J. Merlano Duncan, S. Andrenacci, S. Chatzinotas, and B. Ottersten, “FPGA acceleration for computationally efficient symbol-level precoding in multi-user multi-antenna communication systems,” *IEEE Access*, vol. 7, pp. 15 509–15 520, 2019.

[116] Y. You and G. Lv, “Sphere bounding scheme for probabilistic robust constructive interference precoding in MISO downlink transmission,” *arXiv preprint arXiv:1903.04740*, 2019.

[117] D. Kwon, H. S. Kang, and D. K. Kim, “Robust interference exploitation-based precoding scheme with quantized CSIT,” *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 780–783, 2016.

[118] P. Zetterberg and B. Ottersten, “The spectrum efficiency of a base station antenna array system for spatially selective transmission,” *IEEE Trans. Veh. Technol.*, vol. 44, no. 3, pp. 651–660, 1995.

[119] R. A. Horn and C. R. amnson, *Matrix analysis.* Cambridge university press, 1990.

[120] M. Bengtsson and B. Ottersten, “Optimal downlink beamforming using semidefinite optimization,” in *37th Annual Allerton Conference on Communication, Control, and Computing*, 1999, pp. 987–996.

[121] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge Univ. Press, 2004.

[122] T. H. Chang, Z. Q. Luo, and C. Y. Chi, “Approximation bounds for semidefinite relaxation of max-min-fair multicast transmit beamforming problem,” *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3932–3943, Aug. 2008.

[123] F. Shu, X. Wu, J. Li, R. Chen, and B. Vucetic, “Robust synthesis scheme for secure multi-beam directional modulation in broadcasting systems,” *IEEE Access*, vol. 4, pp. 6614–6623, 2016.

[124] F. Kayhan and G. Montorsi, “Constellation design for memoryless phase noise channels,” *IEEE Trans. Wirel. Commun.*, vol. 13, no. 5, pp. 2874–2883, May 2014.

[125] A. Goldsmith, *Wireless Communications.* Cambridge Univ. Press, 2005.

[126] F. Aurenhammer and R. Klein, *Handbook of computational geometry*, 2000, ch. Voronoi diagrams, pp. 201–290.

[127] M. X. Goemans, “Linear programming and polyhedral combinatorics,” 2009.

[128] R. M. Freund, “Solution methods for quadratic optimization,” 2004.

[129] R. Bro and S. De Jong, “A fast non-negativity-constrained least squares algorithm,” *J. Chemometrics: A Journal of the Chemometrics Society*, vol. 11, no. 5, pp. 393–401, 1997.

[130] D. P. Bertsekas, *Nonlinear Programming.* Athena scientific Belmont, 1999.

[131] *Digital Video Broadcasting (DVB) Part 2: DVB-S2 Extensions (DVB-S2X).* ETSI EN Std. 302 307-2 V1.1.1, 2014.

[132] (2017, Mar.) CVX: MATLAB software for disciplined convex programming. [Online]. Available: http://cvxr.com/cvx

[133] C. Lawson and R. Hanson, *Solving Least Squares Problems.* Society for Industrial and Applied Mathematics, 1995.

[134] R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 11, no. 5, pp. 393–401, 1997.

[135] R. A. Polyak, "Projected gradient method for non-negative least square," *Contemp Math*, vol. 636, pp. 167–179, 2015.

[136] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate O(1/k^2)," in *Dokl. Akad. Nauk SSSR*, vol. 269, 1983, pp. 543–547.

[137] N. Parikh, S. Boyd *et al.*, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[138] Y. Itoh, M. F. Duarte, and M. Parente, "Perfect recovery conditions for non-negative sparse modeling," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 69–80, Jan. 2017.

[139] G. H. Golub and C. F. Van Loan, "An analysis of the total least squares problem," *SIAM journal on numerical analysis*, vol. 17, no. 6, pp. 883–893, 1980.

[140] L. N. Trefethen and D. Bau III, *Numerical linear algebra.* Siam, 1997, vol. 50.

[141] R. Hunger, *Floating point operations in matrix-vector calculus.* Munich University of Technology, Inst. for Circuit Theory and Signal . . . , 2005.

[142] M. Slawski and M. Hein, "Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization," *Electronic J. Statistics*, vol. 7, pp. 7661–7676, 2013.

[143] D. J. Love, R. W. Heath, W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[144] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[145] T. Weber, A. Sklavos, and M. Meurer, "Imperfect channel-state information in MIMO transmission," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 543–552, Mar. 2006.

[146] B. Hassibi and B. M. Hochwald, “How much training is needed in multiple-antenna wireless links?” *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[147] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: Information-theoretic and communications aspects,” *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[148] M. Payaro, A. Pascual-Iserte, and M. A. Lagunas, “Robust power allocation designs for multiuser and multiantenna downlink communication systems through convex optimization,” *IEEE J. Sel. Areas in Commun.*, vol. 25, no. 7, pp. 1390–1401, Sep. 2007.

[149] A. Pascual-Iserte, D. P. Palomar, A. I. Perez-Neira, and M. A. Lagunas, “A robust maximin approach for MIMO communications with imperfect channel state information based on convex optimization,” *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 346–360, Jan. 2006.

[150] H. V. Poor, *An introduction to signal detection and estimation.* Springer Science & Business Media, 2013.

[151] I. Wajid, M. Pesavento, Y. C. Eldar, and D. Ciochina, “Robust downlink beamforming with partial channel state information for conventional and cognitive radio networks,” *IEEE Trans. Signal Process.*, vol. 61, no. 14, pp. 3656–3670, Jul. 2013.

[152] M. B. Shenouda and T. N. Davidson, “Convex conic formulations of robust downlink precoder designs with quality of service constraints,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 714–724, Dec. 2007.

[153] A. Abdel-Samad, T. N. Davidson, and A. B. Gershman, “Robust transmit eigen beamforming based on imperfect channel state information,” *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1596–1609, May 2006.

[154] M. B. Shenouda and T. N. Davidson, “Nonlinear and linear broadcasting with QoS requirements: Tractable approaches for bounded channel uncertainties,” *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1936–1947, May. 2009.

[155] N. Vucic and H. Boche, “Robust QoS-constrained optimization of downlink multiuser MISO systems,” *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 714–725, Feb. 2009.

[156] X. Zhang, D. P. Palomar, and B. Ottersten, “Statistically robust design of linear MIMO transceivers,” *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.

[157] N. Vucic and H. Boche, “A tractable method for chance-constrained power control in downlink multiuser MISO systems with channel uncertainty,” *IEEE Signal Process. Lett.*, vol. 16, no. 5, pp. 346–349, May 2009.

[158] M. B. Shenouda and T. N. Davidson, “Probabilistically-constrained approaches to the design of the multiple antenna downlink,” in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, Oct. 2008, pp. 1120–1124.

[159] M. Botros Shenouda and T. N. Davidson, “Probabilistically-constrained approaches to the design of the multiple antenna downlink,” in *2008 42nd Asilomar Conf. Signals, Systems and Computers*, Oct. 2008, pp. 1120–1124.

[160] B. K. Chalise, S. Shahbazpanahi, A. Czylwik, and A. B. Gershman, “Robust downlink beamforming based on outage probability specifications,” *IEEE Trans. Wirel. Commun.*, vol. 6, no. 10, pp. 3498–3503, Oct. 2007.

[161] B. K. Chalise and A. Czylwik, “Robust uplink beamforming based upon minimum outage probability criterion,” in *Global Telecommun. Conf., 2004. GLOBECOM '04. IEEE*, vol. 6, Nov. 2004, pp. 3974–3978 Vol.6.

[162] K. Wang, A. M. So, T. Chang, W. Ma, and C. Chi, “Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization,” *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690–5705, Nov. 2014.

[163] A. Ben-Taly and A. Nemirovskiz, “On safe tractable approximations of chance constrained linear matrix inequalities,” *Mathematics of Operations Research*, vol. 34, no. 1, pp. 1–25, Feb. 2009.

[164] D. Bertsimas and M. Sim, “Tractable approximations to robust conic optimization problems,” *Mathematical programming*, vol. 107, no. 1-2, pp. 5–36, 2006.

[165] N. Jindal, “MIMO broadcast channels with finite-rate feedback,” *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[166] D. P. Palomar, “A unified framework for communications through MIMO channels,” Ph.D. dissertation, Technical University of Catalonia (UPC), 2003.

[167] T. Yoo and A. Goldsmith, “Capacity and power allocation for fading MIMO channels with channel estimation error,” *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[168] D. Tse and P. Viswanath, *Fundamentals of wireless communication.* Cambridge university press, 2005.

[169] A. L. Agnan Kessy and K. Strimmer, “Optimal whitening and decorrelation,” *The American statistician*, pp. 1–6, Dec. 2016.

[170] D. Bertsimas and M. Sim, “Tractable approximations to robust conic optimization problems,” *Mathematics of Operations Research*, vol. 107, no. 1-2, pp. 5–36, Jun. 2006.

[171] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization.* Siam, 2001, vol. 2.

[172] J. F. Sturm, “Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones,” *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[173] E. V. Belmega and S. Lasaulce, “Energy-efficient precoding for multiple-antenna terminals,” *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 329–340, Jan. 2011.

[174] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An overview of signal processing techniques for millimeter wave MIMO systems,” *IEEE J. Sel. Topics in Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[175] A. I. Perez-Neira, M. A. Vazquez, M. R. B. Shankar, S. Maleki, and S. Chatzinotas, “Signal processing for high-throughput satellites: Challenges in new interference-limited scenarios,” *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 112–131, Jul. 2019.

[176] M. Biguesh, S. Shahbazpanahi, and A. B. Gershman, “Robust downlink power control in wireless cellular systems,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2004, pp. 261–272, Dec. 2004.

[177] B. K. Chalise, S. Shahbazpanahi, A. Czylwik, and A. B. Gershman, “Robust downlink beamforming based on outage probability specifications,” vol. 6, no. 10, pp. 3498–3503, Oct. 2007.

[178] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, “Robust adaptive beamforming for general-rank signal models,” vol. 51, no. 9, pp. 2257–2269, Sep. 2003.

[179] S. Vorobyov, “Robust cdma multiuser detectors: Probability-constrained versus the worst-case-based design,” vol. 15, pp. 273 –276, Nov. 2008.

[180] S. Vorobyov, H. Chen, and A. Gershman, “On the relationship between robust minimum variance beamformers with probabilistic and worst-case distortionless response constraints,” vol. 56, pp. 5719 –5724, Nov. 2008.

[181] S. Shahbazpanahi and A. B. Gershman, “Robust blind multiuser detection for synchronous cdma systems using worst-case performance optimization,” vol. 3, no. 6, pp. 2232–2245, Nov 2004.

[182] K. Zarifi, S. Shahbazpanahi, A. B. Gershman, and Zhi-Quan Luo, “Robust blind multiuser detection based on the worst-case performance optimization of the mmse receiver,” vol. 53, no. 1, pp. 295–305, Jan 2005.

[183] Yue Rong, S. Shahbazpanahi, and A. B. Gershman, “Robust linear receivers for space-time block coded multiaccess MIMO systems with imperfect channel state information,” vol. 53, no. 8, pp. 3081–3090, Aug 2005.

[184] C. M. Jesús Arnau, Bertrand Devillers and A. Pérez-Neira, “Performance study of multiuser interference mitigation schemes for hybrid broadband multibeam satellite architectures,” *EURASIP J. Wirel. Commun. Netw.*, vol. 2012, no. 1, p. 132, Apr. 2012.

[185] V. Joroughi, M. A. Vazquez, and A. I. Perez-Neira, “Generalized multicast multibeam precoding for satellite communications,” *IEEE Trans. Wirel. Commun.*, vol. 16, no. 2, pp. 952–966, Feb. 2017.

[186] A. Gersho and R. M. Gray, *Vector quantization and signal compression.* Springer Science & Business Media, 2012, vol. 159.

[187] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, “Robust predictive quantization: Analysis and design via convex optimization,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 618–632, Dec 2007.

[188] D. R. Hunter and K. Lange, “A tutorial on mm algorithms,” *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[189] Y. Sun, P. Babu, and D. P. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[190] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, “Throughput analysis of massive MIMO uplink with low-resolution ADCs,” *IEEE Trans. Wirel. Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.

[191] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[192] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[193] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[194] H. Yang and T. L. Marzetta, “Performance of conjugate and zero-forcing beamforming in large-scale antenna systems,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[195] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Trans. Wirel. Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.

[196] F. Sohrabi and W. Yu, “Hybrid digital and analog beamforming design for large-scale antenna arrays,” *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[197] A. Alkhateeb, G. Leus, and R. W. Heath, “Limited feedback hybrid precoding for multi-user millimeter wave systems,” *IEEE Trans. Wirel. Commun.*, vol. 14, no. 11, pp. 6481–6494, 2015.

[198] S. Han, C. I, Z. Xu, and C. Rowell, “Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G,” *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[199] L. Liang, W. Xu, and X. Dong, “Low-complexity hybrid precoding in massive multiuser MIMO systems,” *IEEE Wirel. Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[200] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, “Analysis of one-bit quantized precoding for the multiuser massive MIMO downlink,” *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4624–4634, Sep. 2017.

[201] R. H. Walden, “Analog-to-digital converter survey and analysis,” *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[202] C. Svensson, S. Andersson, and P. Bogner, “On the power consumption of analog to digital converters,” in *2006 NORCHIP*, Nov. 2006, pp. 49–52.

[203] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, “Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?” *Ieee Access*, vol. 4, pp. 247–267, 2016.

[204] A. Mezghani, R. Ghiat, and J. A. Nossek, “Transmit processing with low resolution D/A-converters,” in *2009 16th IEEE Int. Conf. Electronics, Circuits and Syst. (ICECS)*, Dec. 2009, pp. 683–686.

[205] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, “Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs,” *IEEE J. Sel. Topics in Signal Process.*, vol. 12, no. 2, pp. 298–312, May 2018.

[206] Y. Li, C. Tao, A. Lee Swindlehurst, A. Mezghani, and L. Liu, “Downlink achievable rate analysis in massive MIMO systems with one-bit DACs,” *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1669–1672, Jul. 2017.

[207] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, “Quantized precoding for massive MU-MIMO,” *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, Nov. 2017.

[208] F. Sohrabi, Y. Liu, and W. Yu, “One-bit precoding and constellation range design for massive MIMO with QAM signaling,” *IEEE J. Sel. Topics in Signal Process.*, vol. 12, no. 3, pp. 557–570, Jun. 2018.

[209] A. Swindlehurst, A. Saxena, A. Mezghani, and I. Fijalkow, “Minimum probability-of-error perturbation precoding for the one-bit massive MIMO downlink,” in *2017 IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2017, pp. 6483–6487.

[210] O. B. Usman, H. Jedda, A. Mezghani, and J. A. Nossek, “MMSE precoder for massive MU-MIMO using 1-bit quantization,” in *IEEE Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2016, pp. 3381–3385.

[211] G. Yuan and B. Ghanem, “Binary optimization via mathematical programming with equilibrium constraints,” *arXiv preprint:1608.04425*, 2016.

[212] X. Hu and D. Ralph, “Convergence of a penalty method for mathematical programming with complementarity constraints,” *Journal of Optimization Theory and Applications*, vol. 123, no. 2, pp. 365–390, 2004.

[213] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, “A limited memory algorithm for bound constrained optimization,” *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.

[214] M. Gustavsson, J. J. Wikner, and N. Tan, *CMOS data converters for communications.* Springer Science & Business Media, 2000, vol. 543.

[215] S. Cui, A. J. Goldsmith, and A. Bahai, “Energy-constrained modulation optimization,” *IEEE Trans. Wirel. Commun.*, vol. 4, no. 5, pp. 2349–2360, 2005.

[216] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!” *IEEE Access*, vol. 1, pp. 335–349, 2013.

[217] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.

[218] S. K. Yong and C.-C. Chong, “An overview of multiGigabit wireless through millimeter wave technology: Potentials and technical challenges,” *EURASIP journal on wireless communications and networking*, vol. 1, no. 1, 2007.

[219] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, “Cellular architecture and key technologies for 5G wireless communication networks,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, 2014.

[220] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, “Millimeter wave beamforming for wireless backhaul and access in small cell networks,” *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, 2013.

[221] A. M. Sayeed and V. Raghavan, “Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 156–166, 2007.

[222] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, “The capacity optimality of beam steering in large millimeter wave MIMO systems,” in *IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2012, pp. 100–104.

[223] X. Gao, O. Edfors, J. Liu, and F. Tufvesson, “Antenna selection in measured massive MIMO channels using convex optimization,” in *IEEE GLOBECOM Workshops.* IEEE, 2013, pp. 129–134.

[224] B. M. Lee, J. Choi, J. Bang, and B.-C. Kang, “An energy efficient antenna selection for large scale green MIMO systems,” in *IEEE International Symposium on Circuits and Systems (ISCAS).* IEEE, 2013, pp. 950–953.

[225] S. S. Ioushua and Y. C. Eldar, “A family of hybrid analog–digital beamforming methods for massive MIMO systems,” *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3243–3257, June 2019.

[226] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, “Hybrid beamforming for massive MIMO: A survey,” *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, 2017.

[227] L. Liang, W. Xu, and X. Dong, “Low-complexity hybrid precoding in massive multiuser MIMO systems,” *IEEE Wirel. Commun. Lett.*, vol. 3, no. 6, pp. 653–656, 2014.

[228] J.-C. Chen, “Hybrid beamforming with discrete phase shifters for millimeter-wave massive MIMO systems,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7604–7608, 2017.

[229] L. A. Belov, S. M. Smolskiy, and V. N. Kochemasov, *Handbook of RF, microwave, and millimeter-wave components.* Artech house, 2012.

[230] D. Liu, U. Pfeiffer, J. Grzyb, and B. Gaucher, *Advanced millimeter-wave technologies: antennas, packaging and circuits.* John Wiley & Sons, 2009.

[231] R. L. Schmid, P. Song, C. T. Coen, A. Ç. Ulusoy, and J. D. Cressler, “On the analysis and design of low-loss single-pole double-throw W-band switches utilizing saturated SiGe HBTs,” *IEEE Trans. Microw. Theory Tech.*, vol. 62, no. 11, pp. 2755–2767, 2014.

[232] I. Kallfass, S. Diebold, H. Massler, S. Koch, M. Seelmann-Eggebert, and A. Leuther, “Multiple-throw millimeter-wave FET switches for frequencies from 60 up to 120 GHz,” in *2008 38th European Microwave Conference.* IEEE, 2008, pp. 1453–1456.

[233] S. Payami, N. Mysore Balasubramanya, C. Masouros, and M. Sellathurai, “Phase shifters versus switches: An energy efficiency perspective on hybrid beamforming,” *IEEE Wirel. Commun. Lett.*, vol. 8, no. 1, pp. 13–16, 2019.

[234] S. Payami, M. Ghoraishi, and M. Dianati, “Hybrid beamforming for large antenna arrays with phase shifter selection,” *IEEE Trans. Wirel. Commun.*, vol. 15, no. 11, pp. 7258–7271, Nov. 2016.

[235] Y. Jiang, Y. Feng, and M. K. Varanasi, “Hybrid beamforming for massive MIMO: A unified solution for both phase shifter and switch networks,” in *2018 10th Int. Conf. Wirel. Commun. and Signal Process. (WCSP)*, 2018, pp. 1–5.

[236] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” *IEEE J. Sel. Topics in Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.

[237] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, “Compressed channel sensing: A new approach to estimating sparse multipath channels,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[238] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter wave wireless communications.* Pearson Education, 2015.

[239] H. Xu, V. Kukshya, and T. S. Rappaport, “Spatial and temporal characteristics of 60-GHz indoor channels,” *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, 2002.

[240] P. F. M. Smulders and L. M. Correia, “Characterisation of propagation in 60 GHz radio channels,” *Electronics Communication Engineering Journal*, vol. 9, no. 2, pp. 73–80, April 1997.

[241] A. Natarajan, S. K. Reynolds, M.-D. Tsai, S. T. Nicolson, J.-H. C. Zhan, D. G. Kam, D. Liu, Y.-L. O. Huang, A. Valdes-Garcia, and B. A. Floyd, “A fully-integrated 16-element phased-array receiver in SiGe BiCMOS for 60-GHz communications,” *IEEE Journal of Solid-State Circuits*, vol. 46, no. 5, pp. 1059–1075, 2011.

[242] A. D. Alexandrov, *Convex Polyhedra.* Springer, 2005.

[243] D. G. Luenberger, *Optimization by Vector Space Methods.* New York: Wiley, 1988.

[244] K. B. Petersen and M. S. Pedersen, “The matrix cookbook,” 2012.

[245] A. M. Tulino, S. Verdú *et al.*, “Random matrix theory and wireless communications,” *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.