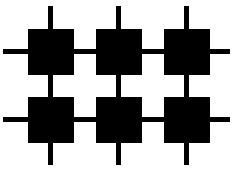
Lehrstuhl für Netzwerktheorie
und Signalverarbeitung



Forschungsberichte
Herausgeber: Prof. Dr. techn. Dr. h. c. Josef A. Nossek

Hela Jedda

Quantized Constant Envelope Transmit Signal processing

Lehrstuhl für Netzwerktheorie und Signalverarbeitung
Technische Universität München

Quantized Constant Envelope Transmit Signal Processing

Hela Jedda

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Holger Boche
Prüfer der Dissertation:

1. Prof. Dr. techn. Dr. h. c. Josef A. Nossek
2. Prof. A. Lee Swindlehurst, Ph.D.,
   University of California, Irvine/ USA

Die Dissertation wurde am 18.09.2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 22.11.2018 angenommen.

Acknowledgments

I would like to express my sincere gratitude to Professor Josef A. Nossek for giving me the opportunity to pursue a doctoral degree at the Institute of Circuit Theory and Signal Processing and supervising me with his steady support and valuable discussions.

I thank Professor A. Lee Swindlehurst for being my mentor and my second supervisor during the Hans-Fischer fellowship at the Institute for Advanced Study. I additionally would like to thank him for giving me the opportunity to spend a wonderful research stay at the University of California in Irvine during the fall term of 2016. I also thank Professor Holger Boche for serving as a chairman of the examination committee.

I thank Professor Wolfgang Utschick for welcoming me in his research group since April 2016 and providing an enjoyable atmosphere.

Many thanks to all my present and former colleagues for the pleasant and friendly work atmosphere. I would like to mention: Jawad Munir, Israa Slim, Leonardo Baltar, Dr. Michael Joham, Hans Brunner, Manuel Stein, Kilian Roth, Qing Bai, and Bernhard Lehmeyer. A special thanks goes to Amine Mezghani for his continuous help and the long constructive discussions we had together.

I also would like to thank my students for all the questions and fruitful discussions. Let me mention Oliver De Candido, Daniel Plabst, Benedikt Fesl, Ferhad Askerbeyli, Andreas Noll, Donia Ben Amor, and Ovais Bin Usman.

A very special gratitude goes to my parents and my family for their everlasting love and all the positive energy they offered me.

Finally, I thank my husband for his love and encouragement. He has been my model in diligence and discipline.

Munich, December 2018 Hela Jedda

Abstract

In this doctoral thesis, we investigate the effect of Quantized Constant Envelope (QCE) signaling at the transmitter on the performance of communication systems. Although massive Multiple-Input Multiple-Output (MIMO) systems, which are characterized by the large number of Base Station (BS) antennas, offer a significant increase in power and spectral efficiency, the hardware power consumption remains a crucial concern especially at Millimeter-Wave (mmW) frequencies. The QCE signaling at the transmitter is a promising approach to enhance the power efficiency. Therefore, we opt for this solution and develop digital signal processing techniques to mitigate the resulting performance loss. Our contribution is twofold. First, we derive the signal statistical properties of the Constant Envelope Quantizer (CEQ). To this end, we extend Price's theorem to the complex-valued case. Second, we develop new precoding techniques, linear and non-linear, to counteract additionally the CEQ distortions. Finally, we discuss the benefits of QCE systems in terms of power efficiency and their challenges in terms of spectral shaping.

Contents

1. Introduction

The remarkable increase in the number of personal mobile devices with internet access combined with the excessive emerging of bandwidth-hungry mobile applications such as multimedia communications, e.g. video streaming with High Definition Television (HDTV) and Ultra-High Definition Video (UHDV), and cloud computing, leads to an ever-increasing demand for higher data rates. However, the existing commercial standards cannot meet the ever increasing need for high-speed wireless connectivity anymore. Therefore, the next generation of mobile communication aims at increasing 1000-fold the network capacity, 10-100-fold the number of connected devices and decreasing 5-fold the latency time and the power consumption compared to 4G networks [1]. To achieve these challenging requirements, the following technologies are the subject of current research:

- massive Multiple-Input Multiple-Output (MIMO) systems, where the Base Stations (BSs) are equipped with a very large number of antennas (100 or more) that can simultaneously serve many users [2–6],
- Millimeter-Wave (mmW) communication, i.e. frequencies ranging between 30 GHz and 300 GHz, where the spectrum is less crowded and greater bandwidth is available [7–9] and
- smaller cells with ranges on the order of 10-200 m, i.e. pico- and femtocells.

First, massive MIMO systems lead to a drastic increase in the number of Radio Frequency (RF) chains at the BS and hence in the number of the wireless front-end hardware components. Second, mmW communication implies that the wireless front-end hardware components are operated at much higher frequencies. Third, reducing the cell size means that the number of cells per unit area is increased and thus results in a much more dense wireless network. Combining the three technologies means a dramatic increase in the number of RF hardware elements operating at very high frequencies per unit area. Hence, the RF power consumption per unit area alarmingly increases. While the above technologies are foreseen as key technologies for future communication systems, the increase in power consumption represents a crucial concern.

The most critical front-end elements in terms of power consumption, depending on whether the large number of antennas is situated at the receiver or at the transmitter, are the Analog-to-Digital Converters (ADCs) in the uplink scenario, and mainly the Power Amplifiers (PAs) and secondarily the Digital-to-Analog Converters (DACs) in the downlink scenario, which is the focus of this contribution. According to [10, 11], the PA is considered as the most power hungry device at the transmitter side. When the PA is run in the saturation region, i.e. the highly non-linear region, high power efficiency is achieved and hence less power is consumed [12]. However, the saturation region implies strong non-linear signal

distortions. To omit the signal distortions, while keeping the PA operate in the saturation region, the input signals should fulfill the Constant Envelope (CE) property, which leads to a unit Peak-to-Average-Power Ratio (PAPR).

To this end, polar (phase-based) DACs at the transmitter are designed to convert the discrete-time and discrete-value base-band signals into continuous-time but discrete-value (at the sample rate), i.e. discrete-phase, CE signals. The number of possible discrete phases is determined by the resolution of the DAC. Thus, the polar DAC can be considered as a Constant Envelope Quantizer (CEQ). The larger the resolution is, the more accurate the phase information at the DAC's output is, but the larger its power consumption is [13]. To further reduce the hardware power consumption, the DAC's resolution can be reduced. The use of coarsely quantized DACs is also beneficial in terms of reduced cost and circuit area and can further simplify the surrounding RF circuitry due to the relaxed linearity constraint, leading to very efficient hardware implementations. In this way, the power consumption is reduced twofold: power efficient PAs due to the CE signals and less power consuming polar DACs due to the low resolution. However, this approach can lead to non-linear distortions that degrade the system performance and have to be mitigated by the precoder design in massive Multi-User (MU) MIMO downlink systems.

1.1 Related Works

Many works have addressed the precoding poblem in the context of CE transmit signals for massive MIMO systems [14–18], where the Multi-User Interference (MUI) is minimized subject to the CE constraint. Another work [19] opts for minimizing an upper bound of the Symbol Error Ratio (SER) in the case of single-user Multiple-Input Single-Output (MISO) systems for two strategies: antenna-subset selection, where a subset of the antennas is selected for transmission, and unequal power allocation among the antennas, where the magnitude of the transmit signal at each antenna is kept constant over a transmission period but the signal magnitudes at distinct transmit antennas are not necessarily equal. The authors of [20] jointly optimize the transmit CE precoding and the constellation in order to minimize the SER in a MISO multicast system. Recent works in [21] and [22] exploit the constructive part of the MUI to design the CE precoder. The authors in [23] design a CE precoder to maximize the Signal-to-Leakage-plus-Noise Ratio (SLNR). In [24], a CE precoder is jointly designed with the receive beamforming to minimize the SER for point-to-point MIMO systems, while adopting antenna grouping for multi-stream transmission. In the above contributions, the DACs are assumed to have infinite resolution.

The contribution in [25] is an early work that addressed the precoding task with low resolution DACs at the transmitter. A linear Minimum Mean Squared Error (MMSE) precoder is designed, while quantization distortion is taken into account. This precoding design is not given in the context of coarsely Quantized Constant Envelope (QCE) signals since the DACs are not polar but cartesian (in-phase- and quadrature). However, the extreme case of 1-bit DACs in [25] represents a special case of coarsely QCE signals. Many contributions in the literature have studied this special case. They can be categorized in two groups: linear and non-linear precoders. In addition to the linear precoder in [25], we introduced in [26] another linear precoder, where the second-order statistics of the 1-bit DAC signals are computed based on Price's theorem [27]. Non-linear precoding techniques in this context were introduced in [28–34]. The non-linear methods can be classified with respect to two

design criteria: the symbol-wise Mean Squared Error (MSE) and the symbol-wise SER. In the context of the symbol-wise MSE, the authors in [29] presented a convex formulation of the problem and applied it to higher-order modulations [30]. The problem formulation is based on semidefinite relaxation and squared ℓ_∞-norm relaxation. The same optimization problem was solved more efficiently in [32] and [35].

In the context of the symbol-wise SER, we presented in [28] a precoding technique based on a minimum Bit Error Ratio (BER) criterion and made use of the box norm (ℓ_∞) to relax the 1-bit constraint. Recently, the work in [33] proposed a method to significantly improve linear precoding solutions in conjunction with 1-bit quantization by properly perturbing the linearly precoded signal for each given input signal to favorably impact the probability of correct detection. In [31] the safety margin to the decision thresholds of the received Phase-Shift Keying (PSK) symbols is maximized subject to a relaxed 1-bit constraint using linear programming for flat-fading channels and extended in [36] for frequency-selective channels. This approach is called the Maximum Safety Margin (MSM) method. The same optimization problem was solved by the Branch-and Bound algorithm in [37] for the special case of 4-PSK. The approach in [34] is based on minimizing an upper bound of the SER under the 1-bit constraint. By the use of a powerful Lemma, the problem was reformulated as a convex optimization problem, of which the solution is discrete.

To the best of our knowledge, the only works that have considered the case of coarsely QCE transmit signals are [38–41]. In [38], we proposed a symbol-wise MSE precoder based on the gradient-descent method under a strict CE constraint or a relaxed polygon constraint. In [39], the authors extended the method in [29] to fit the context of QCE transmit signals. In [40], the authors use a greedy approach for the precoder design while using the symbol-wise MSE as the design criterion. The contribution in [41] addressed the task of QCE precoding in the context of using a single common PA and separate digital phase shifters for the antenna front-ends. The optimization problem consists of designing the QCE precoder while minimizing the MUI, and the idea of constructive interference, [42, 43], is not exploited. In [44], the MSM method was extended to QCE precoding for general constellations and for flat-fading channels. The extension of the MSM method for frequency-selective channels was studied in [45] for Quadrature Amplitude Modulation (QAM) signaling only.

It is worth mentioning that the QCE precoding can be combined with appropriate pulse shaping strategies as in [46, 47] to ensure an efficient spectral confinement. In [48], it was shown that CE precoding is still power efficient even when considering time-based processing. The same investigation can be conducted for the case of QCE precoding. Here, we focus rather on the spatial design problem.

1.2 Outline and Main Contributions

The thesis is organized as follows.

- Chapter 3 is devoted to introduce the mathematical model of the CEQ. Moreover, Price's theorem is extended and applied for the CEQ to obtain its signal statistical properties under the assumption of Gaussian distributed input signals. The main results are summarized in [49]. To ensure minimal distortions, the optimal CEQ is introduced, whose statistical properties can be approximated by the Linear Covariance Approximation (LCA). Finally, a linearized model of the CEQ is derived using Bussgang's theorem.

- Chapter 4 presents the system model of the QCE MU massive MIMO system that is considered throughout the thesis. Two channel models, i.e. the independent and identically distributed (i.i.d.) and the mmW channel models, are also described. Moreover, a potential uplink system model is illustrated.
- Part I, which includes Chapter 5 and Chapter 6, addresses the linear precoding task for flat-fading and frequency-selective channels, respectively. For the precoder design, we choose the MMSE criterion. With the help of Bussgang's theorem, the CEQ can be linearized, which leads to a linear system model with an additional quantization noise. The Wiener Filter (WF) precoder that considers the quantization noise is designed under the assumption of equal and unequal power allocation at the BS antennas. The statistics of the quantization noise are computed by applying Price's theorem or the LCA. Furthermore, the MSE duality between the uplink and the downlink QCE systems is investigated to show that no MSE duality holds and only virtual or approximate duality can be achieved. Linear precoders for the virtual and approximate dual uplink systems are derived.
- Part II, which contains Chapter 7 and Chapter 8, is concerned with the non-linear precoder design for flat-fading and frequency-selective channels, respectively. A significant part of Chapter 7 is published in [44], whereas Chapter 8 is partially published in [45]. The design criterion consists of the safety margin to the decision thresholds. Maximizing the safety margin leads to decreased SER. In contrast to the linear precoding scheme, no precoding matrix is designed but the transmit vector for each given input signal and at a given channel realization is optimized. To take the QCE constraint into account, the entries of the optimized vector should belong to a relaxed convex version of the QCE constraint. The obtained optimization problem is a linear programming problem, for which there exist very efficient methods to solve.
- Chapter 9 discusses the benefits and challenges of QCE systems in terms of power efficiency and spectral shaping. We show the potential power savings of QCE systems compared to the ideal linear systems. Afterwards, we investigate the spectral regrowth in the presence of coarse quantization.

1.3 Notation

Signals, Channels, Filters

$\boldsymbol{\eta}$	Additive White Gaussian Noise (AWGN) vector
$\mathbf{F}$	Linear equalizer in the dual domain
$\mathbf{G}$	Receive matrix
$\mathbf{H}$	Channel matrix
$\mathbf{P}$	Precoding matrix
$\mathbf{r}$	Received signal vector
$\mathbf{s}$	input signal vector
$\hat{\mathbf{s}}$	Estimate of the input signal vector
$\mathbf{t}$	CEQ output signal vector
$\mathbf{T}$	Linear transmit matrix in the dual domain
$\mathbf{u}$	Received filtered signal vector
$\mathbf{x}$	CEQ input signal vector
$\mathbf{y}$	Noiseless received signal vector

Numbers and Quantities

B	Block length for the non-linear processing
δ	Safety margin to the decision thresholds at the receiver
L	Number of taps of the channel impulse response matrix
L_p	Number of taps of the linear precoder impulse response matrix
M	Number of single-antenna users
N	Number of transmit antennas at the BS
P_{tx}	Available transmit power
S	Cardinality of the input set
θ	$\frac{\pi}{S}$
T	Block length for the blind estimation at the receiver
T_{s}	Sample period
T_{sym}	Symbol period

Matrix, vector and complex number operations

$(\bullet)^{\mathrm{T}}$	Transpose of a matrix
$(\bullet)^{\mathrm{H}}$	Hermitian of a matrix
$\mathrm{tr}(\bullet)$	Trace of a matrix
$\mathrm{null}(\bullet)$	Null space of a matrix
$\mathrm{range}(\bullet)$	Range of a matrix
$(\bullet)^{*}$	Complex conjugate
$\|\bullet\|_p$	p-norm
$\Re\{\bullet\}$	Real part of a complex-valued number
$\Im\{\bullet\}$	Imaginary part of a complex-valued number
$*$	Convolution
$\circ$	Hadamard product
$\otimes$	Kronecker product
$\mathbf{0}_N$	N-dimensional zero vector
$\mathbf{1}_N$	N-dimensional all ones vector
$\mathbf{I}_N$	N-dimensional identity matrix
$\mathbf{e}_n$	n-th column of the identity matrix
$\mathbf{0}_{N,M}$	$N \times M$-dimensional zero matrix
$\mathbf{X}$	Matrix
$\mathbf{x}$	Vector
x	Scalar
x_i	i-th element of the vector $\mathbf{x}$
$\mathrm{diag}\,(\mathbf{X})$	Diagonal matrix containing the diagonal elements of $\mathbf{X}$
$\mathrm{nondiag}\,(\mathbf{X})$	$\mathbf{X} - \mathrm{diag}\,(\mathbf{X})$
$f\,[t]$	Function f evaluated at discrete instants t
$f\,(t)$	Function f evaluated at continuous instants t

Quantization

α_q	$1 - \beta_q$
β_q	Distortion factor of the CEQ
$\mathbf{d}_{\mathcal{Q}}$	Quantization noise after the Bussgang linearization of the CEQ
$\mathbf{L}_{\mathcal{Q}}$	Linear matrix between the input and output after the Bussgang linearization of the CEQ
q	Resolution in bits
Q	Number of possible discrete phases at the output of the CEQ
$\mathcal{Q}_{\mathrm{CE}}$	Element-wise CEQ
$\mathcal{Q}_{\phi}$	Element-wise phase quantization
ψ	$\frac{\pi}{Q}$
Ξ_n	Envelope magnitude at the output of the n-th CEQ
$\mathbf{\Xi}$	Diagonal matrix $\sum_{n=1}^{N} \Xi_n \mathbf{e}_n \mathbf{e}_n^{\mathrm{T}}$

Probability and Stochastics

$p_X(x)$	Probability Density Function (PDF) of the random variable X
Pr(statement)	Probability that the statement holds true
$\wedge$	Logical and
$\vee$	Logical or
$\mathrm{E}[\bullet]$	Expectation operator
σ_x^2	Variance of the signal x
$\rho_{x,y}$	Correlation factor between x and y
$\mathbf{C_{xy}}$	Covariance matrix between $\mathbf{x}$ and $\mathbf{y}$, i.e. $\mathbf{C_{xy}} = \mathrm{E}\left[\mathbf{x}\mathbf{y}^{\mathrm{H}}\right] - \mathrm{E}[\mathbf{x}]\,\mathrm{E}\left[\mathbf{y}^{\mathrm{H}}\right]$
$\mathbf{R_{xy}}$	Correlation matrix, i.e. $\mathbf{R_{xy}} = \mathrm{diag}(\mathbf{C_{xx}})^{-1/2}\mathbf{C_{xy}}\,\mathrm{diag}(\mathbf{C_{yy}})^{-1/2}$
$\mathcal{CN}_{\mathbb{C}}(\mathbf{0}, \mathbf{C})$	Circularly-symmetric complex-valued Gaussian distribution with zero-mean and covariance matrix $\mathbf{C}$

Sets

$\mathbb{C}$	Set of complex-valued numbers
$\mathbb{S}$	Input set, i.e. QAM or PSK
$\mathbb{T}$	CEQ output set
$\mathbb{X}$	CEQ input set

2. Motivation

2.1 Power Amplifiers

The PA in wireless communication systems is a device at the end of the transmit front-end to amplify the RF signal and drive it into the antenna. The amplification is in the ideal case without distortion and with 100% efficiency. However, this is not possible with real devices. Therefore, each PA is characterized by two main features, i.e. linearity and efficiency.
Linear amplification avoids introducing strong distortions to the signal and hence avoiding the higher out-of band radiations. To run the PA in the linear region, the input signal has to have a peak magnitude well below the PA output peak. This implies, however, that the supply power is not used fully and power dissipation arises.
Power efficient amplification ensures efficient power usage and hence less requirements on cooling systems at the BS. The power efficiency of the PA is defined as

$$\eta = \frac{P_{\text{out}}}{P_{\text{DC}}}, \tag{2.1}$$

where P_{out} is the output power and P_{DC} is the supply power. It describes how much percentage out of the supply power is converted to the RF power (output power). If the power efficiency is equal to 1, it means that the supply power is totally transmitted to the output. Hence, no power is consumed at the PA in the ideal case.
Both fundamental characteristics depend on the class of the PA. There are different classes of amplifiers of different theoretical efficiency values [50]; that is

- class A ($\eta = 50\%$), class AB ($\eta = 68\%$), class B ($\eta = 78.5\%$), class C ($\eta = 87\%$): behave like linear transconductors
- class D ($\eta = 100\%$), class E ($\eta = 100\%$), class S ($\eta = 100\%$), and M ($\eta = 100\%$) [51]: behave like non-linear switches with high efficiency levels.

There is always a trade-off between the linearity and the power efficiency of the PA. The linear PA is less power efficient than the non-linear PA. However, having a PA input signal of constant magnitude, the PA can be operated in the saturation region while achieving the maximal possible power efficiency and without introducing any distortion to the signal. The PA of class M shows a higher efficiency in real systems compared to other highly efficient PAs. Therefore, we suggest its use in this thesis.

2.2 DACs

The conventional DACs in wireless communication systems are the current-steering DACs thanks to their high-speed characteristic [52]. A current steering DAC consists of a number of

parallel switched current sources. Depending on the digital input signal, the switchers steer the currents to either the output or dump them to ground. The resulting weighted currents are summed at the output and converted into a voltage by a transimpedance stage.
The number of switched current sources grows exponentially with the resolution of the DAC. Therefore, a high resolution leads to a huge number of devices, increased complexity of the corresponding wiring and certainly higher hardware power consumption. Thus, low resolution ensures less power consumption and simplified circuitry.

Since we are interested in having CE signals at the PA input for power efficiency enhancement, the DAC should be designed accordingly. In CE signaling, the information is conveyed by the signal phase. Hence, instead of having two DACs for the inphase and quadrature signal parts each, the DAC should process both signals jointly to recover the phase information. Therefore, we do not talk about cartesian DACs anymore but about polar DACs that are applied in polar transmitters. The DAC can be modeled as a two-step operation. The first step consists of the quantization of the continuous-value input signal. The second step is the conversion from the discrete-time to the continuous-time representation. The quantization operation of the polar DAC can be then modeled by the CEQ, which is described in Chapter 3. Note that the cartesian DAC with resolution of one bit at each dimension is equivalent to the polar DAC with a resolution of two bits; that is only four discrete phase values can be generated at the output.

2.3 Polar Transmitters

Polar transmission can enhance simultaneously the linearity and the efficiency of wireless communication systems [53,54]. The information is conveyed by the magnitude and the phase in contrast to cartesian systems, where the information lies in the inphase and quadrature signal parts. For our study, the magnitude is constant due to the CE property. Thus, the information lies only in the phase and only phase modulation is required.
For the conversion from the digital to the analog world, Digital-to-Time Converter (DTC) are deployed in polar transmitters. The DTC is composed of an array of switchers with cascaded delay elements. The DTC input signal is an instantaneous period that is converted into a certain time delay. The time delay is obtained by turning on the corresponding switch. The number of delay elements defines the number of possible discrete phase values at the quantizer output. In other words, the resolution determines the number of delay elements. The higher the resolution is, the more delay elements we have in the converter. For less power consumption and reduced chip area we restrict our work to low resolution and therefore small number of delay elements in the DTC. Note that our derivations assume that the number of delay elements is a power of two. However, the work can be generalized for all possible numbers of delay elements.

3. On Constant Envelope Quantizers

3.1 One-Dimensional CEQ

The input of the CEQ, denoted by x, is a complex-valued signal in contrast to the quantizers with real-valued inputs that are largely studied in the literature. In the polar representation, the output signal t has a constant magnitude and a quantized version of the input signal phase as

$$\begin{aligned} t &= \mathcal{Q}_{\mathrm{CE}}(x) \\ &= \Xi\, \mathrm{e}^{\mathrm{j}\, \mathcal{Q}_\phi(\arg(x))} \\ &= \Xi \sum_{k=1}^{Q} \left(u(\arg(x) - (2k-2)\psi) - u(\arg(x) - 2k\psi)\right) \mathrm{e}^{\mathrm{j}(2k-1)\psi}, \end{aligned} \tag{3.1}$$

where Ξ is constant and denotes the magnitude of the CE, $u(\bullet)$ denotes the unit step function as

$$u(\phi) = \begin{cases} 0 & \phi < 0 \\ 1 & \phi \geq 0, \end{cases} \tag{3.2}$$

ψ is defined as

$$\psi = \frac{\pi}{Q}, \tag{3.3}$$

and $\arg(x)$ gives the phase of the complex-valued signal x. So, the information after the CEQ lies only in the phase. The phase quantizer $\mathcal{Q}_\phi(\bullet)$ is a symmetric uniform real-valued quantizer. It is characterized by its resolution q that defines the number of the discrete output phases, i.e Q, that is

$$Q = 2^q. \tag{3.4}$$

In other words, the 2π phase range is divided into Q, $\frac{2\pi}{Q}$-rotationally symmetric, sectors. The input signal that belongs to the k-th sector gets quantized (mapped) to $\Xi\, \mathrm{e}^{\mathrm{j}(2k-1)\psi}$ as shown in Fig. 3.1.

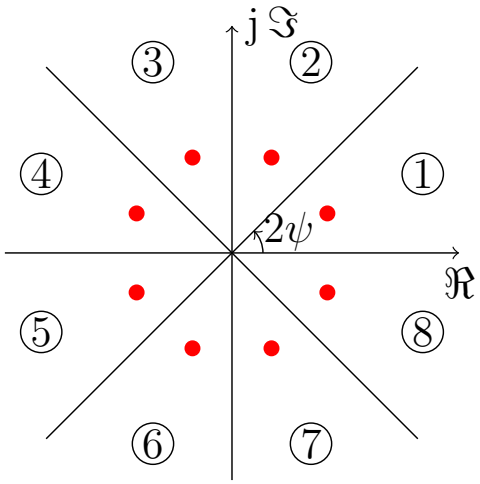


Fig. 3.1: Illustration of the CEQ output set for $q = 3$, i.e. $Q = 8$.

3.2 Multi-Dimensional CEQ

Having an N-dimensional vector $\mathbf{x}$ at the input of an N-dimensional CEQ, the output vector $\mathbf{t}$ is given by

$$\begin{aligned}\mathbf{t} &= \mathcal{Q}_{\mathrm{CE}}(\mathbf{x}) \\ &= \sum_{n=1}^{N} \mathcal{Q}_{\mathrm{CE}}(x_n)\mathbf{e}_n \\ &= \sum_{n=1}^{N} \Xi_n \, \mathrm{e}^{\mathrm{j}\,\mathcal{Q}_\phi(\arg(x_n))}\,\mathbf{e}_n,\end{aligned} \tag{3.5}$$

which corresponds to an element-wise quantization. The output vector can be compactly expressed as

$$\mathbf{t} = \boldsymbol{\Xi}\,\mathrm{e}^{\mathrm{j}\,\mathcal{Q}_\phi(\arg(\mathbf{x}))}, \tag{3.6}$$

where $\boldsymbol{\Xi}$ is a diagonal matrix containing the magnitudes for each output; that is

$$\boldsymbol{\Xi} = \sum_{n=1}^{N} \Xi_n \mathbf{e}_n \mathbf{e}_n^T. \tag{3.7}$$

3.3 Statistical Properties

In this section, we aim to find the impact of the CEQ on the statistical properties of Gaussian input signals. To this end, we consider N input signals $x_n,\ n = 1, \cdots, N$, that build the N- dimensional vector $\mathbf{x}$. The N-dimensional CEQ output vector is denoted by $\mathbf{t}$. We assume that the input signals are joint complex-valued Gaussian distributed with zero mean and variance $\sigma^2_{x_n}$, that is $x_n \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2_{x_n}),\ n = 1, \cdots, N$. Thus, the Probability Density Function (PDF) of the complex-valued Gaussian distributed random variable $x_n = r_n e^{\mathrm{j}\,\phi_n},\ n = 1, \cdots, N$,

reads as, [55, Theorem 2.4],

$$p_{X_n}(r_n, \phi_n) = \frac{1}{\pi \sigma_{x_n}^2} \, \mathrm{e}^{-\frac{r_n^2}{\sigma_{x_n}^2}} \, . \tag{3.8}$$

Additionally, the complex-valued correlation coefficient ρ_{x_i,x_j} between the input signals x_i and x_j, $i, j = 1, \cdots, N$, is defined as

$$\rho_{x_i,x_j} = \frac{\mathrm{E}\left[x_i x_j^*\right]}{\sigma_{x_i} \sigma_{x_j}} \tag{3.9}$$

$$= \frac{1}{\sigma_{x_i} \sigma_{x_j}} \int_0^{2\pi} \int_0^{2\pi} \int_0^{\infty} \int_0^{\infty} r_i^2 r_j^2 \, \mathrm{e}^{\mathrm{j}(\phi_i - \phi_j)} \, p_{X_i,X_j}(r_i, r_j, \phi_i, \phi_j) \, \mathrm{d}r_i \, \mathrm{d}r_j \, \mathrm{d}\phi_j \, \mathrm{d}\phi_i, \tag{3.10}$$

where the joint PDF for $x_i \neq x_j$ is expressed as

$$p_{X_i,X_j,i\neq j}(r_i, r_j, \phi_i, \phi_j) = \frac{1}{\pi^2 \left(1 - \left|\rho_{x_i,x_j}\right|^2\right) \sigma_{x_i}^2 \sigma_{x_j}^2} \, \mathrm{e}^{-\frac{1}{1-\left|\rho_{x_i,x_j}\right|^2} \left(\frac{r_i^2}{\sigma_{x_i}^2} + \frac{r_j^2}{\sigma_{x_j}^2} - \frac{2 r_i r_j \Re\left\{\rho_{x_i,x_j} \, e^{-\mathrm{j}(\phi_i - \phi_j)}\right\}}{\sigma_{x_i} \sigma_{x_j}} \right)}, \tag{3.11}$$

as explained in A1. Moreover, we define the input covariance matrix as

$$\begin{aligned} \mathbf{C_{xx}} &= \mathrm{E}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right] \\ &= \begin{bmatrix} \sigma_{x_1}^2 & \cdots & \mathrm{E}\left[x_1 x_N^*\right] \\ \vdots & \ddots & \\ \mathrm{E}\left[x_N x_1^*\right] & \cdots & \sigma_{x_N}^2 \end{bmatrix}. \end{aligned} \tag{3.12}$$

Consequently, the input correlation matrix reads as

$$\begin{aligned} \mathbf{R_{xx}} &= \mathrm{diag}\left(\mathbf{C_{xx}}\right)^{-1/2} \mathbf{C_{xx}} \, \mathrm{diag}\left(\mathbf{C_{xx}}\right)^{-1/2} \\ &= \begin{bmatrix} 1 & \cdots & \rho_{x_1,x_N} \\ \vdots & \ddots & \\ \rho_{x_N,x_1} & \cdots & 1 \end{bmatrix}. \end{aligned} \tag{3.13}$$

Assuming Gaussian input signals, our target consists of finding the expressions of

- the covariance matrix $\mathbf{C_{tx}} = \mathrm{E}\left[\mathbf{t}\mathbf{x}^{\mathrm{H}}\right]$ between the QCE signal vector and the unquantized signal vector, i.e. $\mathrm{E}\left[t_i x_j^*\right]$, $i, j = 1, \cdots, N$, and
- the covariance matrix $\mathbf{C_{tt}} = \mathrm{E}\left[\mathbf{t}\mathbf{t}^{\mathrm{H}}\right]$ of the QCE signal vector for a given input correlation matrix $\mathbf{R_{xx}}$, i.e. $\mathrm{E}\left[t_i t_j^*\right]$, $i, j = 1, \cdots, N$.

To this end, we introduce Price's theorem.

3.3.1 Price's Theorem

Price's theorem, first introduced in [27], consists in ordinary or partial differential equations of different orders to describe the expected value of the product of n non-linear functions of n real-valued jointly distributed Gaussian random variables. Each non-linear function depends only on one random variable. The ℓ-th derivative of the expected value with respect

to (w.r.t.) the covariances between the random variables equates the expected value of the product of the ℓ-th functions' derivatives w.r.t. the corresponding random variables. The theorem was reformulated in [56] to obtain a partial differential equation relating the derivative of the expected value of a non-linear function of two jointly distributed Gaussian random variables w.r.t. their covariance and the expected value of the second order derivative of the non-linear function w.r.t. the random variables. This relationship holds under a certain boundedness constraint given in [57]. The number of the real-valued joint Gaussian variables n determines the number of the differential equations to solve, that is $\binom{n}{2} = n(n-1)/2$. The modified version of Price's theorem in [58] extends the expression for non-linear functions of more than two real-valued variables to end up with a single ordinary differential equation involving all derivatives' orders. In [59], Price's theorem was extended for two circularly symmetric complex-valued Gaussian random variables and generalized in [60] under no circular symmetry assumption. The two latter extensions describe the complex-valued version of Price's theorem for the derivative of first order only. We could stick to the version of Price's theorem for complex-valued Gaussian random variables for the first order only given in [60]. However, we extend the theorem in [58] to obtain Price's theorem for two circularly symmetric complex-valued Gaussian random variables that involve all derivatives' orders.

Theorem 1. *Let $f(\mathbf{x}) = f(x_1, x_2)$ be a function of two joint complex-valued circularly symmetric Gaussian variables. The covariance matrix and the correlation matrix of the vector $\mathbf{x}$ are denoted by $\mathbf{C_{xx}}$ and $\mathbf{R_{xx}}$ of elements ρ_{x_i,x_j}, $i,j = 1,2$. The off-diagonal entries of $\mathbf{C_{xx}}$ are multiplied by a perturbing term υ. Then, it holds that*

$$\frac{\mathrm{d}^\ell \,\mathrm{E}\left[f(x_i,x_j)\right]}{\mathrm{d}\upsilon^\ell} = \sigma_{x_i}^\ell \sigma_{x_j}^\ell \mathrm{E}_\upsilon \Bigg[\Re\{\rho_{x_i,x_j}\}^\ell \left(\frac{\partial^{2\ell} f(x_i,x_j)}{\partial x_i^\ell \partial x_j^{*\ell}} + \frac{\partial^{2\ell} f(x_i,x_j)}{\partial x_i^{*\ell} \partial x_j^\ell} \right) + \mathrm{j}^\ell \Im\{\rho_{x_i,x_j}\}^\ell \left(\frac{\partial^{2\ell} f(x_i,x_j)}{\partial x_i^\ell \partial x_j^{*\ell}} - \frac{\partial^{2\ell} f(x_i,x_j)}{\partial x_i^{*\ell} \partial x_j^\ell} \right) \Bigg], \tag{3.14}$$

where E_υ denotes the expectation operation based on the resulting perturbed PDF.

The proof is provided in A2. ■

Price's theorem is usually an alternative way to find the analytical expression for $\mathrm{E}\left[f(x_i,x_j)\right]$, which can be obtained by integrating the first-order derivative w.r.t. υ. Therefore, we state Theorem 3.14 for the special case $\ell = 1$

$$\frac{\mathrm{d}\,\mathrm{E}\left[f(x_i,x_j)\right]}{\mathrm{d}\upsilon} = \sigma_{x_i}\sigma_{x_j}\mathrm{E}_\upsilon \left[\rho_{x_i,x_j} \frac{\partial^2 f(x_i,x_j)}{\partial x_i \partial x_j^*} + \rho_{x_i,x_j}^* \frac{\partial^2 f(x_i,x_j)}{\partial x_i^* \partial x_j} \right]. \tag{3.15}$$

The latter expression can be reformulated as

$$\begin{aligned} \mathrm{d}\,\mathrm{E}\left[f(x_i,x_j)\right] &= \sigma_{x_i}\sigma_{x_j}\mathrm{E}_\upsilon \left[\frac{\partial^2 f(x_i,x_j)}{\partial x_i \partial x_j^*} \right] \rho_{x_i,x_j}\,\mathrm{d}\upsilon + \sigma_{x_i}\sigma_{x_j}\mathrm{E}_\upsilon \left[\frac{\partial^2 f(x_i,x_j)}{\partial x_i^* \partial x_j} \right] \rho_{x_i,x_j}^*\,\mathrm{d}\upsilon \\ &= \sigma_{x_i}\sigma_{x_j}\mathrm{E}_\upsilon \left[\frac{\partial^2 f(x_i,x_j)}{\partial x_i \partial x_j^*} \right] \mathrm{d}\rho'_{x_i,x_j} + \sigma_{x_i}\sigma_{x_j}\mathrm{E}_\upsilon \left[\frac{\partial^2 f(x_i,x_j)}{\partial x_i^* \partial x_j} \right] \mathrm{d}\rho'^*_{x_i,x_j}, \end{aligned} \tag{3.16}$$

where $\rho'_{x_i,x_j} = \upsilon\rho_{x_i,x_j}$. Therefore, we obtain

$$\frac{\partial \mathrm{E}\left[f(x_i,x_j)\right]}{\partial \rho'_{x_i,x_j}} = \sigma_{x_i}\sigma_{x_j}\mathrm{E}_{\upsilon}\left[\frac{\partial^2 f(x_i,x_j)}{\partial x_i \partial x_j^*}\right], \tag{3.17}$$

$$\frac{\partial \mathrm{E}\left[f(x_i,x_j)\right]}{\partial \rho'^*_{x_i,x_j}} = \sigma_{x_i}\sigma_{x_j}\mathrm{E}_{\upsilon}\left[\frac{\partial^2 f(x_i,x_j)}{\partial x_i^* \partial x_j}\right], \tag{3.18}$$

which simplify for the case of $\upsilon = 1$ to

$$\frac{\partial \mathrm{E}\left[f(x_i,x_j)\right]}{\partial \rho_{x_i,x_j}} = \sigma_{x_i}\sigma_{x_j}\mathrm{E}\left[\frac{\partial^2 f(x_i,x_j)}{\partial x_i \partial x_j^*}\right], \tag{3.19}$$

$$\frac{\partial \mathrm{E}\left[f(x_i,x_j)\right]}{\partial \rho^*_{x_i,x_j}} = \sigma_{x_i}\sigma_{x_j}\mathrm{E}\left[\frac{\partial^2 f(x_i,x_j)}{\partial x_i^* \partial x_j}\right]. \tag{3.20}$$

The expression in (3.19) was also obtained in [59] under the circular symmetry assumption, while (3.19) and (3.20) were obtained in [60] without any circular symmetry assumption. This implies that Theorem 1 holds true for the first-order derivative for general complex-valued signals, too.

In the polar representation, (3.19) and (3.20) read as

$$\begin{aligned}&\frac{\partial \mathrm{E}\left[f(x_i,x_j)\right]}{\partial \rho_{x_i,x_j}} = \frac{1}{4}\sigma_{x_i}\sigma_{x_j}\cdot\\ &\mathrm{E}\left[\mathrm{e}^{-\mathrm{j}(\phi_i-\phi_j)}\left(\frac{\partial^2 f(x_i,x_j)}{\partial r_i \partial r_j} + \frac{1}{r_i r_j}\frac{\partial^2 f(x_i,x_j)}{\partial \phi_i \partial \phi_j} + \mathrm{j}\,\frac{1}{r_j}\frac{\partial^2 f(x_i,x_j)}{\partial r_i \partial \phi_j} - \mathrm{j}\,\frac{1}{r_i}\frac{\partial^2 f(x_i,x_j)}{\partial \phi_i \partial r_j}\right)\right]\end{aligned} \tag{3.21}$$

$$\begin{aligned}&\frac{\partial \mathrm{E}\left[f(x_i,x_j)\right]}{\partial \rho^*_{x_i,x_j}} = \frac{1}{4}\sigma_{x_i}\sigma_{x_j}\cdot\\ &\mathrm{E}\left[\mathrm{e}^{\mathrm{j}(\phi_i-\phi_j)}\left(\frac{\partial^2 f(x_i,x_j)}{\partial r_i \partial r_j} + \frac{1}{r_i r_j}\frac{\partial^2 f(x_i,x_j)}{\partial \phi_i \partial \phi_j} - \mathrm{j}\,\frac{1}{r_j}\frac{\partial^2 f(x_i,x_j)}{\partial r_i \partial \phi_j} + \mathrm{j}\,\frac{1}{r_i}\frac{\partial^2 f(x_i,x_j)}{\partial \phi_i \partial r_j}\right)\right].\end{aligned} \tag{3.22}$$

The conversion of the second order derivative of a function w.r.t. two complex-valued variables to the second order derivatives w.r.t. to their corresponding polar arguments is detailed in A3.

3.3.2 Cross-correlation between QCE Signal and Unquantized Signal

The cross-correlation factor between x_i and t_i is expressed as

$$\begin{aligned}
\mathrm{E}\left[t_i x_i^*\right] &= \int_0^{2\pi}\int_0^{\infty} \Xi_i e^{\mathrm{j}\,\mathcal{Q}_\phi(\phi_i)} r_i\,\mathrm{e}^{-\mathrm{j}\,\phi_i}\, p_{X_i}(r_i,\phi_i) r_i\,\mathrm{d}r_i\,\mathrm{d}\phi_i \\
&= \frac{1}{\pi\sigma_{x_i}^2}\int_0^{\infty}\sum_{k=1}^{Q}\int_{(2k-2)\psi}^{2k\psi} \Xi_i\,\mathrm{e}^{\mathrm{j}(2k-1)\psi}\, r_i^2 e^{-\mathrm{j}\,\phi_i}\,\mathrm{e}^{-\frac{r_i^2}{\sigma_{x_i}^2}}\,\mathrm{d}\phi_i\,\mathrm{d}r_i \\
&= \frac{1}{\pi\sigma_{x_i}^2}\int_0^{\infty} r_i^2\,\mathrm{e}^{-\frac{r_i^2}{\sigma_{x_i}^2}}\,\mathrm{d}r_i \sum_{k=1}^{Q}\Xi_i\,\mathrm{e}^{\mathrm{j}(2k-1)\psi}\int_{(2k-2)\psi}^{2k\psi}\mathrm{e}^{-\mathrm{j}\,\phi_i}\,\mathrm{d}\phi_i \\
&= \frac{1}{\pi\sigma_{x_i}^2}\frac{\sigma_{x_i}^2}{4}\sqrt{\pi\sigma_{x_i}^2}\sum_{k=1}^{Q}\Xi_i\,\mathrm{e}^{\mathrm{j}(2k-1)\psi}\left(\mathrm{e}^{\mathrm{j}\left(\frac{\pi}{2}-2k\psi\right)}-\mathrm{e}^{\mathrm{j}\left(\frac{\pi}{2}-(2k-2)\psi\right)}\right) \\
&= \frac{\sigma_{x_i}}{4\sqrt{\pi}}\sum_{k=1}^{Q}\Xi_i\left(\mathrm{e}^{\mathrm{j}\left(\frac{\pi}{2}-\psi\right)}-\mathrm{e}^{\mathrm{j}\left(\frac{\pi}{2}+\psi\right)}\right) \\
&= \frac{\Xi_i\sigma_{x_i}}{2\sqrt{\pi}}Q\sin(\psi). \qquad (3.23)
\end{aligned}$$

The computation of the cross-correlation factor between t_i and x_j with $i \neq j$ implies the evaluation of the following quadruple integral

$$\begin{aligned}
\mathrm{E}\left[t_i x_j^*\right] &= \int_0^{2\pi}\int_0^{2\pi}\int_0^{\infty}\int_0^{\infty}\Xi_i\,\mathrm{e}^{\mathrm{j}\,\mathcal{Q}_\phi(\phi_i)}\, r_j e^{-\mathrm{j}\,\phi_j} p_{X_i,X_j,i\neq j}(r_i,r_j,\phi_i,\phi_j) r_i r_j\,\mathrm{d}r_i\,\mathrm{d}r_j\,\mathrm{d}\phi_i\,\mathrm{d}\phi_j \\
&= \sum_{k=1}^{Q}\int_{(2k-2)\psi}^{2k\psi}\Xi_i\,\mathrm{e}^{\mathrm{j}(2k-1)\psi}\int_0^{2\pi}\mathrm{e}^{-\mathrm{j}\,\phi_j}\int_0^{\infty}\int_0^{\infty} r_i r_j^2 p_{X_i,X_j,i\neq j}(r_i,r_j,\phi_i,\phi_j)\,\mathrm{d}r_i\,\mathrm{d}r_j\,\mathrm{d}\phi_j\,\mathrm{d}\phi_i, \qquad (3.24)
\end{aligned}$$

which is not straightforward to calculate analytically. Therefore, we make use of Price's theorem. To this end, we compute the following derivatives

$$\begin{aligned}
\frac{\partial^2 t_i x_j^*}{\partial r_i \partial r_j} &= 0, \qquad (3.25)\\
\frac{\partial^2 t_i x_j^*}{\partial r_i \partial \phi_j} &= 0, \qquad (3.26)\\
\frac{\partial^2 t_i x_j^*}{\partial \phi_i \partial r_j} &= \Xi_i\sum_{k=1}^{Q}\left(\delta(\phi_i-(2k-2)\psi)-\delta(\phi_i-2k\psi)\right)\mathrm{e}^{\mathrm{j}((2k-1)\psi-\phi_j)}, \qquad (3.27)\\
\frac{\partial^2 t_i x_j^*}{\partial \phi_i \partial \phi_j} &= -\mathrm{j}\,\Xi_i\sum_{k=1}^{Q}\left(\delta(\phi_i-(2k-2)\psi)-\delta(\phi_i-2k\psi)\right) r_j\,\mathrm{e}^{\mathrm{j}((2k-1)\psi-\phi_j)} \\
&= -\mathrm{j}\,r_j\frac{\partial^2 t_i x_j^*}{\partial \phi_i \partial r_j}. \qquad (3.28)
\end{aligned}$$

Hence, we apply Price's theorem in the polar representation from (3.21) and (3.22). We start with the first property and we get

$$\begin{aligned}\frac{\partial \operatorname{E}\left[t_i x_j^*\right]}{\partial \rho_{x_i,x_j}} &= \sigma_{x_i}\sigma_{x_j} \operatorname{E}\left[\frac{\partial^2 t_i x_j^*}{\partial x_i \partial x_j^*}\right] \\ &= \frac{1}{4}\sigma_{x_i}\sigma_{x_j} \operatorname{E}\left[\mathrm{e}^{-\mathrm{j}(\phi_i-\phi_j)}\left(\frac{-\mathrm{j}}{r_i}\frac{\partial^2 t_i x_j^*}{\partial \phi_i \partial r_j} + \frac{1}{r_i r_j}\frac{\partial^2 t_i x_j^*}{\partial \phi_i \partial \phi_j}\right)\right] \\ &= -\mathrm{j}\,\frac{1}{2}\sigma_{x_i}\sigma_{x_j} \operatorname{E}\left[\mathrm{e}^{-\mathrm{j}(\phi_i-\phi_j)}\,\frac{1}{r_i}\frac{\partial^2 t_i x_j^*}{\partial \phi_i \partial r_j}\right] \\ &= -\mathrm{j}\,\frac{1}{2}\sigma_{x_i}\sigma_{x_j}\int_0^{2\pi}\int_0^{2\pi}\int_0^{\infty}\int_0^{\infty} \mathrm{e}^{-\mathrm{j}(\phi_i-\phi_j)}\,\frac{\partial^2 t_i x_j^*}{\partial \phi_i \partial r_j} p_{X_i,X_j,i\neq j}(r_i, r_j, \phi_i, \phi_j) r_j \,\mathrm{d}r_i \,\mathrm{d}r_j \,\mathrm{d}\phi_j \,\mathrm{d}\phi_i.\end{aligned} \tag{3.29}$$

Plugging (3.27) in (3.29) results in

$$\begin{aligned}\frac{\partial \operatorname{E}\left[t_i x_j^*\right]}{\partial \rho_{x_i,x_j}} &= -\mathrm{j}\,\frac{1}{2}\sigma_{x_i}\sigma_{x_j}\Xi_i \int_0^{2\pi} \sum_{k=1}^{Q} \mathrm{e}^{\mathrm{j}((2k-1)\psi-\phi_j)}\,\cdot \\ &\quad\left(\mathrm{e}^{-\mathrm{j}((2k-2)\psi-\phi_j)}\int_0^{\infty}\int_0^{\infty} p_{X_i,X_j,i\neq j}(r_i, r_j, (2k-2)\psi, \phi_j) r_j \,\mathrm{d}r_i \,\mathrm{d}r_j\right. \\ &\qquad\left. - \mathrm{e}^{-\mathrm{j}(2k\psi-\phi_j)}\int_0^{\infty}\int_0^{\infty} p_{X_i,X_j,i\neq j}(r_i, r_j, 2k\psi, \phi_j) r_j \,\mathrm{d}r_i \,\mathrm{d}r_j\right) \\ &= -\mathrm{j}\,\frac{1}{2}\sigma_{x_i}\sigma_{x_j}\Xi_i \int_0^{2\pi} \sum_{k=1}^{Q} \left(e^{\mathrm{j}\,\psi} w_1((2k-2)\psi - \phi_j) - e^{-\mathrm{j}\,\psi} w_1(2k\psi - \phi_j)\right) \mathrm{d}\phi_j,\end{aligned} \tag{3.30}$$

where

$$w_1(\phi_i - \phi_j) = \int_0^{\infty}\int_0^{\infty} p_{X_i,X_j,i\neq j}(r_i, r_j, \phi_i, \phi_j) r_j \,\mathrm{d}r_i \,\mathrm{d}r_j. \tag{3.31}$$

The latter double integral is computed analytically in (A2) and the obtained expression is

$$w_1(\phi_i - \phi_j) = \frac{\sqrt{1 - \left|\rho_{x_i,x_j}\right|^2}}{4\pi\sqrt{\pi}\sigma_{x_i}\left(1 - \kappa\left(\phi_i - \phi_j\right)\right)}, \tag{3.32}$$

where

$$\kappa(\phi) = \Re\{\rho_{x_i,x_j}\,\mathrm{e}^{-\mathrm{j}\,\phi}\}. \tag{3.33}$$

Hence, (3.30) reduces to

$$\begin{aligned}
\frac{\partial \,\mathrm{E}\left[t_i x_j^*\right]}{\partial \rho_{x_i,x_j}} &= -\,\mathrm{j}\,\frac{1}{2}\sigma_{x_i}\sigma_{x_j}\Xi_i \int_0^{2\pi} \sum_{k=1}^{Q} w_1(2k\psi-\phi_j)\left(e^{\mathrm{j}\,\psi}-e^{-\mathrm{j}\,\psi}\right)\mathrm{d}\phi_j \\
&= \sigma_{x_i}\sigma_{x_j}\Xi_i \sin(\psi)\int_0^{2\pi} \sum_{k=1}^{Q} w_1(2k\psi-\phi_j)\,\mathrm{d}\phi_j \\
&\overset{(a)}{=} \sigma_{x_i}\sigma_{x_j}\Xi_i \sin(\psi) Q \int_0^{2\pi} w_1(\phi)\,\mathrm{d}\phi \\
&= \frac{\sqrt{1-\left|\rho_{x_i,x_j}\right|^2}\sigma_{x_j}\Xi_i}{4\pi\sqrt{\pi}} \sin(\psi) Q \int_0^{2\pi} \frac{1}{1-\Re\{\rho_{x_i,x_j}\,\mathrm{e}^{-\mathrm{j}\,\phi}\}}\,\mathrm{d}\phi \\
&\overset{(b)}{=} \frac{\sqrt{1-\left|\rho_{x_i,x_j}\right|^2}\sigma_{x_j}\Xi_i}{4\pi\sqrt{\pi}} \sin(\psi) Q \int_0^{2\pi} \frac{1}{1-\left|\rho_{x_i,x_j}\right|\cos(\phi')}\,\mathrm{d}\phi' \\
&= \frac{\sqrt{1-\left|\rho_{x_i,x_j}\right|^2}\sigma_{x_j}\Xi_i}{4\pi\sqrt{\pi}} \sin(\psi) Q \int_0^{\pi} \frac{2}{1-\left|\rho_{x_i,x_j}\right|^2\cos^2(\phi')}\,\mathrm{d}\phi' \\
&\overset{(c)}{=} \frac{Q}{2\sqrt{\pi}} \sin(\psi)\Xi_i\sigma_{x_j},
\end{aligned} \tag{3.34}$$

where in (a) we introduce the variable $\phi = 2k\psi - \phi_j$. Since $w_1(\phi)$ is 2π-periodic, the integral operation from $(2k\psi - 2\pi)$ to $2k\psi$ is equivalent to the integration from 0 to 2π. In (b) we define $\phi' = \arg\left(\rho_{x_i,x_j}\right) - \phi$. Again the integral boundaries do not change due to the 2π-periodicity of the integrand. In (c) we use the equality

$$\int_{-1}^{1} \frac{2}{1-a^2x^2}\frac{1}{\sqrt{1-x^2}} = \frac{2\pi}{1-a^2},\ \forall a \in \mathbb{R} \quad \text{with} \quad a \le 1. \tag{3.35}$$

Additionally, we obtain

$$\begin{aligned}
\frac{\partial \,\mathrm{E}\left[t_i x_j^*\right]}{\partial \rho^*_{x_i,x_j}} &= \sigma_{x_i}\sigma_{x_j}\,\mathrm{E}\left[\frac{\partial^2 t_i x_j^*}{\partial x_i^* \partial x_j}\right] \\
&= \frac{1}{4}\sigma_{x_i}\sigma_{x_j}\,\mathrm{E}\left[\mathrm{e}^{\mathrm{j}(\phi_i-\phi_j)}\left(\frac{\mathrm{j}}{r_i}\frac{\partial^2 t_i x_j^*}{\partial\phi_i\partial r_j} + \frac{1}{r_i r_j}\frac{\partial^2 t_i x_j^*}{\partial\phi_i\partial\phi_j}\right)\right] \\
&= 0.
\end{aligned} \tag{3.36}$$

Since $\mathrm{E}\left[t_i x_j^*\right]$ depends only on ρ_{x_i,x_j} and not on $\rho^*_{x_i,x_j}$, we can get the covariance between t_i and x_j by integrating the expression in (3.34) w.r.t. ρ_{x_i,x_j} as

$$\begin{aligned}
\mathrm{E}\left[t_i x_j^*\right] &= \int_0^{\rho_{x_i,x_j}} \frac{\partial \,\mathrm{E}\left[t_i x_j^*\right]}{\partial \rho'_{x_i,x_j}}\,\mathrm{d}\rho'_{x_i,x_j} \\
&= \frac{Q}{2\sqrt{\pi}} \sin(\psi)\Xi_i\sigma_{x_j}\rho_{x_i,x_j}.
\end{aligned} \tag{3.37}$$

3.3.3 Cross-correlation between Two QCE Signals

The auto-correlation of the CEQ output t_i is given by

$$\begin{aligned}\mathrm{E}\left[|t_i|^2\right] &= \int_0^{2\pi}\int_0^{\infty} |t_i|^2 \, p_{X_i}(r_i,\phi_i) r_i \,\mathrm{d}r_i \,\mathrm{d}\phi_i \\ &= |t_i|^2 \int_0^{2\pi}\int_0^{\infty} p_{X_i}(r_i,\phi_i) r_i \,\mathrm{d}r_i \,\mathrm{d}\phi_i \\ &= |t_i|^2 = \Xi_i^2. \end{aligned} \tag{3.38}$$

However, the computation of the cross-correlation between t_i and t_j with $i \neq j$ turns to be more complicated

$$\begin{aligned}\mathrm{E}\left[t_i t_j^*\right] &= \int_0^{2\pi}\int_0^{2\pi}\int_0^{\infty}\int_0^{\infty} \Xi_i \Xi_j \, \mathrm{e}^{\mathrm{j}\left(\mathcal{Q}_\phi(\phi_i)-\mathcal{Q}_\phi(\phi_j)\right)} \, p_{X_i,X_j,i\neq j}(r_i,r_j,\phi_i,\phi_j) r_i r_j \,\mathrm{d}r_i \,\mathrm{d}r_j \,\mathrm{d}\phi_i \,\mathrm{d}\phi_j \\ &= \sum_{k=1}^{Q}\sum_{k'=1}^{Q} \int_{(2k-2)\psi}^{2k\psi} \int_{(2k'-2)\psi}^{2k'\psi} \Xi_i \Xi_j \, \mathrm{e}^{\mathrm{j}(2k-2k')\psi} \int_0^{\infty}\int_0^{\infty} r_i r_j p_{X_i,X_j,i\neq j}(r_i,r_j,\phi_i,\phi_j) \,\mathrm{d}r_i \,\mathrm{d}r_j \,\mathrm{d}\phi_j \,\mathrm{d}\phi_i. \end{aligned} \tag{3.39}$$

Therefore, we make use of Price's theorem to compute the cross-correlation between t_i and t_j with $i \neq j$. Since the information after the CEQ lies only in the phase, all derivatives w.r.t. the radii vanish and only the derivative w.r.t. to the phases is different from zero, i.e.

$$\frac{\partial^2 t_i t_j^*}{\partial r_i \partial r_j} = 0, \tag{3.40}$$

$$\frac{\partial^2 t_i t_j^*}{\partial r_i \partial \phi_j} = 0, \tag{3.41}$$

$$\frac{\partial^2 t_i t_j^*}{\partial \phi_i \partial r_j} = 0, \tag{3.42}$$

$$\begin{aligned}\frac{\partial^2 t_i t_j^*}{\partial \phi_i \partial \phi_j} = \sum_{k=1}^{Q}\sum_{k'=1}^{Q} &\left(\delta(\phi_i - (2k-2)\psi) - \delta(\phi_i - 2k\psi)\right) \cdot \\ &\left(\delta(\phi_j - (2k'-2)\psi) - \delta(\phi_j - 2k'\psi)\right) \Xi_i \Xi_j \, \mathrm{e}^{\mathrm{j}(2(k-k')\psi)}. \end{aligned} \tag{3.43}$$

Hence, we apply Price's theorem in (3.21) and (3.22). We get

$$\begin{aligned}\frac{\partial \,\mathrm{E}\left[t_i t_j^*\right]}{\partial \rho_{x_i,x_j}} &= \sigma_{x_i}\sigma_{x_j} \,\mathrm{E}\left[\frac{\partial^2 t_i t_j^*}{\partial x_i \partial x_j^*}\right] \\ &= \frac{1}{4}\sigma_{x_i}\sigma_{x_j} \,\mathrm{E}\left[\mathrm{e}^{-\mathrm{j}(\phi_i-\phi_j)} \, \frac{1}{r_i r_j} \frac{\partial^2 t_i t_j^*}{\partial \phi_i \partial \phi_j}\right] \\ &= \frac{1}{4}\sigma_{x_i}\sigma_{x_j} \int_0^{2\pi}\int_0^{2\pi}\int_0^{\infty}\int_0^{\infty} \mathrm{e}^{-\mathrm{j}(\phi_i-\phi_j)} \, \frac{\partial^2 t_i t_j^*}{\partial \phi_i \partial \phi_j} p_{X_i,X_j,i\neq j}(r_i,r_j,\phi_i,\phi_j) \,\mathrm{d}r_i \,\mathrm{d}r_j \,\mathrm{d}\phi_i \,\mathrm{d}\phi_j \end{aligned} \tag{3.44}$$

Plugging (3.43) in (3.44) leads to

$$\begin{aligned}
\frac{\partial \mathrm{E}\left[t_i t_j^*\right]}{\partial \rho_{x_i,x_j}} &= \frac{1}{4}\sigma_{x_i}\sigma_{x_j}\Xi_i\Xi_j \sum_{k=1}^{Q}\sum_{k'=1}^{Q} \mathrm{e}^{\mathrm{j}(2(k-k')\psi)} \\
&\quad \Bigg(\mathrm{e}^{-\mathrm{j}(2(k-k')\psi)} \int_0^\infty \int_0^\infty p_{X_i,X_j,i\neq j}(r_i,r_j,(2k-2)\psi,(2k'-2)\psi)\,\mathrm{d}r_i\,\mathrm{d}r_j \\
&\quad + \mathrm{e}^{-\mathrm{j}(2(k-k')\psi)} \int_0^\infty \int_0^\infty p_{X_i,X_j,i\neq j}(r_i,r_j,2k\psi,2k'\psi)\,\mathrm{d}r_i\,\mathrm{d}r_j \\
&\quad - \mathrm{e}^{-\mathrm{j}(2(k-k'-1)\psi)} \int_0^\infty \int_0^\infty p_{X_i,X_j,i\neq j}(r_i,r_j,(2k-2)\psi,2k'\psi)\,\mathrm{d}r_i\,\mathrm{d}r_j \\
&\quad - \mathrm{e}^{-\mathrm{j}(2(k-k'+1)\psi)} \int_0^\infty \int_0^\infty p_{X_i,X_j,i\neq j}(r_i,r_j,2k\psi,(2k'-2)\psi)\,\mathrm{d}r_i\,\mathrm{d}r_j \Bigg) \\
&= \frac{1}{4}\sigma_{x_i}\sigma_{x_j}\Xi_i\Xi_j \sum_{k=1}^{Q}\sum_{k'=1}^{Q} \Bigg(2w_2(2(k-k')\psi) \\
&\qquad - \mathrm{e}^{\mathrm{j}\,2\psi}\, w_2(2(k-k'-1)\psi) - \mathrm{e}^{-\mathrm{j}\,2\psi}\, w_2(2(k-k'+1)\psi) \Bigg), \quad (3.45)
\end{aligned}$$

where

$$w_2(\phi_i-\phi_j) = \int_0^\infty \int_0^\infty p_{X_i,X_j,i\neq j}(r_i,r_j,\phi_i,\phi_j)\,\mathrm{d}r_i\,\mathrm{d}r_j. \quad (3.46)$$

The latter double integral is computed analytically in (A3) and the obtained expression is

$$w_2(\phi_i-\phi_j) = \frac{1}{\pi\sigma_{x_i}\sigma_{x_j}\sqrt{1-\kappa\left(\phi_i-\phi_j\right)^2}} \left(\frac{1}{4} + \frac{1}{2\pi}\arcsin(\kappa\left(\phi_i-\phi_j\right))\right), \quad (3.47)$$

where $\kappa(\phi)$ is defined in (3.33). Thus, (3.45) can be rewritten as

$$\begin{aligned}
\frac{\partial \mathrm{E}\left[t_i t_j^*\right]}{\partial \rho_{x_i,x_j}} &= \frac{1}{4}\sigma_{x_i}\sigma_{x_j}\Xi_i\Xi_j Q \sum_{\Delta k=0}^{Q-1} \left(2w_2(2\Delta k\psi) - \mathrm{e}^{\mathrm{j}\,2\psi}\, w_2(2(\Delta k-1)\psi) - \mathrm{e}^{-\mathrm{j}\,2\psi}\, w_2(2(\Delta k+1)\psi)\right) \\
&= \frac{\Xi_i\Xi_j Q}{4\pi} \sum_{\Delta k=0}^{Q/2-1} \left(\frac{1}{\sqrt{1-\kappa(2\Delta k\psi)^2}} - \frac{\mathrm{e}^{\mathrm{j}\,2\psi}}{2\sqrt{1-\kappa(2(\Delta k-1)\psi)^2}} - \frac{\mathrm{e}^{-\mathrm{j}\,2\psi}}{2\sqrt{1-\kappa(2(\Delta k+1)\psi)^2}} \right) \\
&= \frac{\Xi_i\Xi_j Q}{4\pi} \sum_{\Delta k=0}^{Q/2-1} \frac{1}{\sqrt{1-\kappa(2\Delta k\psi)^2}} \left(1 - \frac{1}{2}\left(\mathrm{e}^{-\mathrm{j}(2\psi)} - \mathrm{e}^{\mathrm{j}(2\psi)}\right)\right) \\
&= \frac{\Xi_i\Xi_j Q}{4\pi} \left(1-\cos\left(2\psi\right)\right) \sum_{\Delta k=0}^{Q/2-1} \frac{1}{\sqrt{1-\kappa(2\Delta k\psi)^2}} \\
&= \frac{\Xi_i\Xi_j Q}{2\pi} \sin^2\left(\psi\right) \sum_{\Delta k=0}^{Q/2-1} \frac{1}{\sqrt{1-\kappa(2\Delta k\psi)^2}}. \quad (3.48)
\end{aligned}$$

Similarly, we compute $\frac{\partial \mathrm{E}\left[t_i t_j^*\right]}{\partial \rho^*_{x_i,x_j}}$ and get

$$\frac{\partial \mathrm{E}\left[t_i t_j^*\right]}{\partial \rho^*_{x_i,x_j}} = \frac{\Xi_i \Xi_j Q}{2\pi} \sin^2(\psi) \sum_{\Delta k=0}^{Q/2-1} \frac{\mathrm{e}^{\mathrm{j}\,4\Delta k\psi}}{\sqrt{1-\kappa(2\Delta k\psi)^2}}. \tag{3.49}$$

Since $\frac{\partial^2 \mathrm{E}\left[t_i t_j^*\right]}{\partial \rho_{x_i,x_j} \partial \rho^*_{x_i,x_j}} = \frac{\partial^2 \mathrm{E}\left[t_i t_j^*\right]}{\partial \rho^*_{x_i,x_j} \partial \rho_{x_i,x_j}}$, there exists a potential for the vector field built by the partial derivatives. Hence, the expression of the covariance between t_i and t_j is given by

$$\begin{aligned}
\mathrm{E}\left[t_i t_j^*\right] &= \int_0^{\rho_{x_i,x_j}} \frac{\partial \mathrm{E}\left[t_i t_j^*\right]}{\partial \rho'_{x_i,x_j}} \,\mathrm{d}\rho'_{x_i,x_j} = \int_0^{\rho^*_{x_i,x_j}} \frac{\partial \mathrm{E}\left[t_i t_j^*\right]}{\partial \rho'^*_{x_i,x_j}} \,\mathrm{d}\rho'^*_{x_i,x_j} \\
&= \int_0^{\rho_{x_i,x_j}} \frac{\Xi_i \Xi_j Q}{2\pi} \sin^2(\psi) \sum_{\Delta k=0}^{Q/2-1} \frac{1}{\sqrt{1-\kappa(2\Delta k\psi)^2}} \,\mathrm{d}\rho'_{x_i,x_j} \\
&= \frac{\Xi_i \Xi_j Q}{2\pi} \sin^2(\psi) \sum_{\Delta k=0}^{Q/2-1} \int_0^{\kappa(2\Delta k\psi)} \frac{2\,\mathrm{e}^{\mathrm{j}\,2\Delta k\psi}}{\sqrt{1-\kappa'(2\Delta k\psi)^2}} \,\mathrm{d}\kappa'(2\Delta k\psi) \\
&= \frac{\Xi_i \Xi_j Q}{\pi} \sin^2(\psi) \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \arcsin\left(\kappa(2\Delta k\psi)\right) \\
&\overset{(3.33)}{=} \frac{\Xi_i \Xi_j Q}{\pi} \sin^2(\psi) \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \arcsin\left(\Re\left\{\rho_{x_i,x_j}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\right\}\right).
\end{aligned} \tag{3.50}$$

3.3.4 Numerical Validation

In this section, we aim to check the correctness of the expressions in (3.37) and (3.50) by comparing them with numerical results. To this end, we take two sequences of length $N_x = 10^6$ of two arbitrary complex-valued Gaussian signals x_1 and x_2 of variances $\sigma^2_{x_1}$ and $\sigma^2_{x_2}$. We pass them through a two-dimensional CEQ to obtain t_1 and t_2. The input signals are correlated with a given input correlation coefficient $\rho_{\text{in}} = \rho_{x_1,x_2}$. With Monte-Carlo simulations we compute numerically the cross-correlation coefficient $\rho_{\text{cross}} = \rho_{t_1,x_2} = \mathrm{E}\left[t_1 x_2^*\right] / \Xi_1 \sigma_{x_2}$ and the output correlation coefficient, i.e. $\rho_{\text{out}} = \rho_{t_1,t_2} = \mathrm{E}\left[t_1 t_2^*\right] / \Xi_1 \Xi_2$. The obtained results (dashed lines with markers) are compared to the derived closed-form expressions (solid lines) in Fig. 3.2 for $Q = 4,\ 8,\ \infty$. Almost the same distortion behavior is observed between $Q = 16$ and $Q = \infty$. For this reason the results related to $Q = 16$ are dropped out. Since we are dealing with complex-valued correlation coefficients, we need 4 plots to study the relationship between ρ_{in} and ρ_{cross} and another 4 plots for ρ_{in} and ρ_{out}. As can be seen, the numerical simulations show the correctness of the derived expressions. Moreover, it can be deduced that the phases of the output correlation coefficient and the cross-correlation coefficient remain almost unchanged compared to the phase of the input correlation coefficient. However, most of the distortion is concerning the magnitude depending on Q.

3.3.5 In a Nutshell

In summary, applying (3.37) and (3.50) for the multi-dimensional case, the signal statistical properties of an N-dimensional CEQ of a Gaussian input signal $\mathbf{x}$ and an output signal $\mathbf{t}$

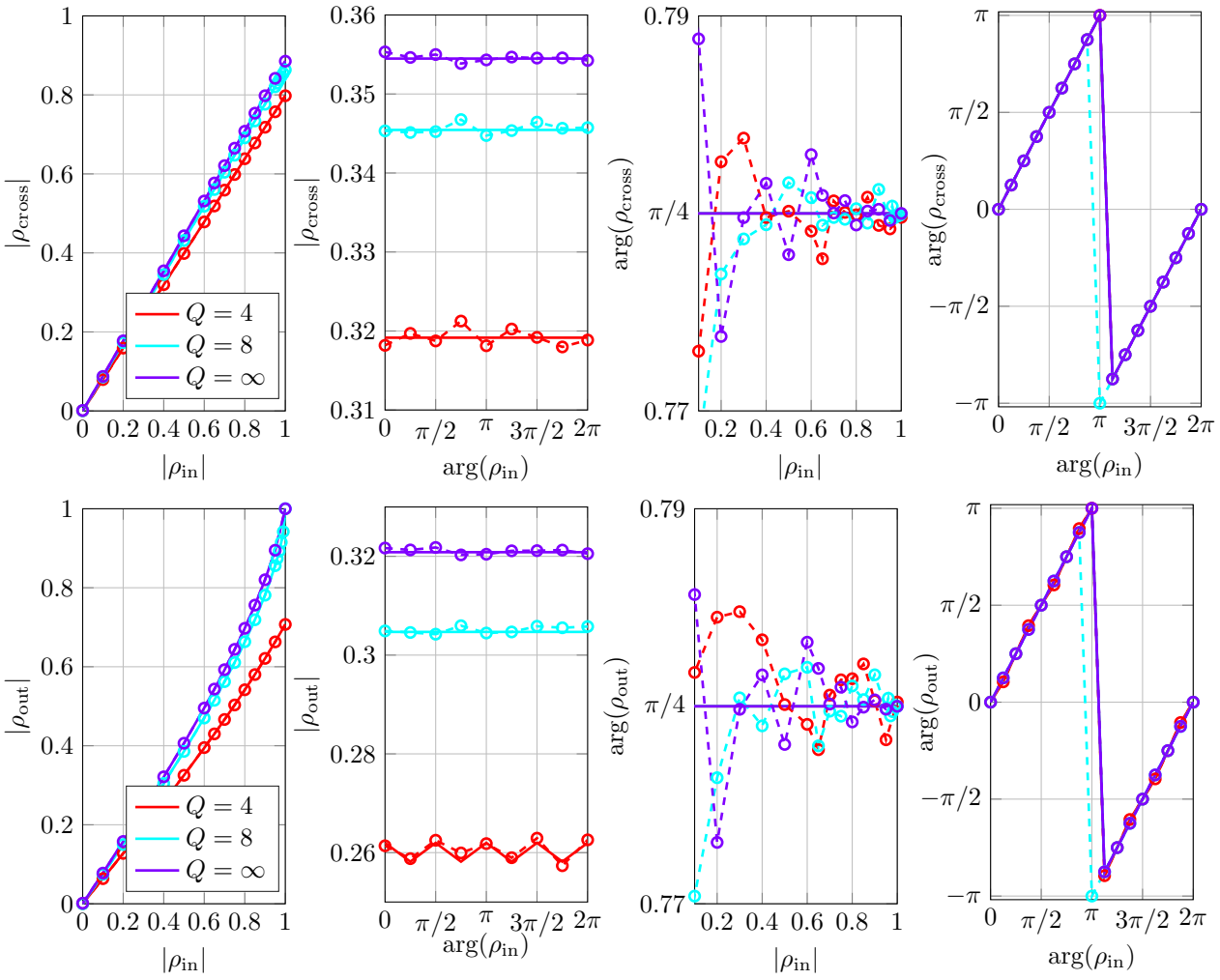


Fig. 3.2: Comparison of the derived closed-form expressions in (3.37) and (3.50) (solid lines) with the numerical results obtained by Monte-Carlo simulations (dashed lines with markers). In each figure, one dimension of ρ_{in} is fixed either $|\rho_{\text{in}}| = 0.4$ or $\arg(\rho_{\text{in}}) = \frac{\pi}{4}$,

read as

$$\mathbf{C_{tx}} = \frac{Q}{2\sqrt{\pi}} \sin(\psi) \mathbf{\Xi} \operatorname{diag} \left(\mathbf{C_{xx}} \right)^{-1/2} \mathbf{C_{xx}}, \tag{3.51}$$

and

$$\mathbf{C_{tt}} = \frac{Q}{\pi} \sin^2 (\psi) \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k \psi)} \mathbf{\Xi} \arcsin \left(\Re\{ \mathbf{R_{xx}} \, \mathrm{e}^{-\mathrm{j}\, 2\Delta k \psi} \} \right) \mathbf{\Xi}. \tag{3.52}$$

3.4 Optimal CEQ

3.4.1 One-Dimensional Optimal CEQ

In this section, the magnitude Ξ is optimized to minimize the quantization distortions introduced to the signal. In general, the one-dimensional CEQ output can be expressed as

$$t = x + \eta_q, \tag{3.53}$$

where η_q denotes the complex-valued quantization distortion term. Since we assume that the input signal $x = r \, \mathrm{e}^{\mathrm{j}\, \phi}$ is a complex-valued Gaussian distributed signal with zero mean and variance σ_x^2, we get

$$\mathrm{E}\left[\eta_q\right] = \mathrm{E}\left[t\right] - \mathrm{E}\left[x\right] = 0, \tag{3.54}$$

and

$$\mathrm{E}\left[|\eta_q|_2^2 \right] = \sigma_{\eta_q}^2 = \beta_q \sigma_x^2, \tag{3.55}$$

where β_q denotes the distortion factor of the CEQ. The task is to find the optimal Ξ that minimizes the power of the quantization distortion term; that is

$$\Xi_{\mathrm{opt}} = \arg \min_{\Xi} \sigma_{\eta_q}^2. \tag{3.56}$$

To this end, we find the expression of $\sigma^2_{\eta_q}$.

$$\begin{aligned}
\sigma^2_{\eta_q} &= \mathrm{E}\left[|t-x|_2^2\right] \\
&= \int_0^{2\pi}\int_0^{\infty} |t-x|_2^2\, p_X(r,\phi) r\,\mathrm{d}r\,\mathrm{d}\phi \\
&= \sum_{k=1}^{Q}\int_{(2k-2)\psi}^{2k\psi}\int_0^{\infty} \left|\Xi\,\mathrm{e}^{\mathrm{j}(2k-1)\psi} - r\,\mathrm{e}^{\mathrm{j}\,\phi}\right|_2^2 \frac{1}{\pi\sigma_x^2}\,\mathrm{e}^{-\frac{r^2}{\sigma_x^2}}\, r\,\mathrm{d}r\,\mathrm{d}\phi \\
&= \frac{1}{\pi\sigma_x^2}\sum_{k=1}^{Q}\int_{(2k-2)\psi}^{2k\psi}\int_0^{\infty} \left(\Xi^2 + r^2 - \Xi r\,\mathrm{e}^{\mathrm{j}((2k-1)\psi-\phi)} - \Xi r\,\mathrm{e}^{-\mathrm{j}((2k-1)\psi-\phi)}\right)\mathrm{e}^{-\frac{r^2}{\sigma_x^2}}\, r\,\mathrm{d}r\,\mathrm{d}\phi \\
&= \frac{1}{\pi\sigma_x^2}\Big(2\pi\Xi^2\int_0^{\infty}\mathrm{e}^{-\frac{r^2}{\sigma_x^2}}\, r\,\mathrm{d}r + 2\pi\int_0^{\infty}\mathrm{e}^{-\frac{r^2}{\sigma_x^2}}\, r^3\,\mathrm{d}r \\
&- \Xi\sum_{k=1}^{Q} e^{\mathrm{j}((2k-1)\psi)}\int_{(2k-2)\psi}^{2k\psi}\mathrm{e}^{-\mathrm{j}\,\phi}\int_0^{\infty}\mathrm{e}^{-\frac{r^2}{\sigma_x^2}}\, r^2\,\mathrm{d}r\,\mathrm{d}\phi \\
&- \Xi\sum_{k=1}^{Q} e^{-\mathrm{j}((2k-1)\psi)}\int_{(2k-2)\psi}^{2k\psi}\mathrm{e}^{\mathrm{j}\,\phi}\int_0^{\infty}\mathrm{e}^{-\frac{r^2}{\sigma_x^2}}\, r^2\,\mathrm{d}r\,\mathrm{d}\phi\Big) \\
&= \frac{1}{\pi\sigma_x^2}\Bigg(2\pi\Xi^2\frac{\sigma_x^2}{2} + 2\pi\frac{\sigma_x^4}{2} \\
&- \Xi\frac{\sigma_x^2}{4}\left(\sum_{k=1}^{Q} e^{\mathrm{j}((2k-1)\psi)}\int_{(2k-2)\psi}^{2k\psi}\mathrm{e}^{-\mathrm{j}\,\phi}\sqrt{\pi\sigma_x^2}\,\mathrm{d}\phi + \sum_{k=1}^{Q} e^{-\mathrm{j}((2k-1)\psi)}\int_{(2k-2)\psi}^{2k\psi}\mathrm{e}^{\mathrm{j}\,\phi}\sqrt{\pi\sigma_x^2}\,\mathrm{d}\phi\right)\Bigg) \\
&= \Xi^2 + \sigma_x^2 - Q\frac{\sigma_x}{4\sqrt{\pi}}\Xi\left(2\,\mathrm{j}\left(\mathrm{e}^{-\mathrm{j}\,\psi} - \mathrm{e}^{\mathrm{j}\,\psi}\right)\right) \\
&= \Xi^2 + \sigma_x^2 - Q\frac{\sigma_x}{\sqrt{\pi}}\Xi\sin(\psi). \qquad (3.57)
\end{aligned}$$

The optimal magnitude Ξ_{opt} is obtained by setting the derivative of $\sigma^2_{\eta_q}$ w.r.t. Ξ equal to 0; that is

$$\frac{\mathrm{d}\sigma^2_{\eta_q}}{\mathrm{d}\Xi} = 2\Xi - \frac{\sigma_x}{\sqrt{\pi}}Q\sin(\psi) = 0. \qquad (3.58)$$

Thus, the optimal magnitude is given by

$$\Xi_{\mathrm{opt}} = \frac{Q}{2\sqrt{\pi}}\sin(\psi)\sigma_x. \qquad (3.59)$$

The optimal magnitude of the CE signal is dependent on the standard deviation of the input signal x. Thus, a different optimal magnitude is obtained for a different input signal variance σ_x^2. With the use of the optimal CEQ, the following covariance factors simplify to:

$$\begin{aligned}
\mathrm{E}\left[t\eta_q^*\right] &= \mathrm{E}\left[t\left(t^* - x^*\right)\right] \\
&= \mathrm{E}\left[tt^*\right] - \mathrm{E}\left[tx^*\right] \\
&= \Xi_{\mathrm{opt}}^2 - \frac{Q}{2\sqrt{\pi}}\sin(\psi)\Xi_{\mathrm{opt}}\sigma_x \overset{(3.59)}{=} 0, \qquad (3.60)
\end{aligned}$$

q	Q	Ξ_{opt}	β_q
2	4	$\sqrt{2/\pi}$	$1-2/\pi$
3	8	0.86362402	0.2541536
4	16	0.88054342	0.2246433
5	32	0.88480399	0.2171219
∞	∞	$\sqrt{\pi/4}=0.88622693$	$1-\pi/4$

Table 3.1: Optimal step size for the CEQ and the corresponding distortion factor for unit variance inputs.

and

$$\begin{aligned}\mathrm{E}\left[x\eta_q^*\right] &= \mathrm{E}\left[(t-\eta_q)\,\eta_q^*\right]\\ &= \mathrm{E}\left[t\eta_q^*\right]-\sigma_{\eta_q}^2\\ &\overset{(3.60)}{=} -\sigma_{\eta_q}^2\\ &\overset{(3.55)}{=} -\beta_q\sigma_x^2. \end{aligned} \tag{3.61}$$

After plugging (3.59) in (3.57), the distortion factor β_q for the optimal CEQ reduces to

$$\beta_q = 1-\frac{Q^2\sin^2(\psi)}{4\pi}. \tag{3.62}$$

And we introduce the new variable α_q as

$$\alpha_q = 1-\beta_q = \frac{Q^2\sin^2(\psi)}{4\pi}. \tag{3.63}$$

The optimal values Xi_{opt} and the resulting distortion factors of the CEQ having unit variance Gaussian complex-valued inputs, i.e. $\sigma_x^2=1$, are summarized in Table 3.1 for different quantization resolutions. We notice that the values of Ξ_{opt} and β_q, when rounded up to the second digit after the decimal point, do not vary for $q\geq 4$ and hence for $Q\geq 16$. This explains why we observed the same behavior between $Q=16$ and $Q=\infty$ in Fig. 3.2.

3.4.2 Multi-Dimensional Optimal CEQ

From (3.59) and the definition in (3.63), the optimal matrix $\mathbf{\Xi}$ for an N-dimensional CEQ with input vector $\mathbf{x}$ and output vector $\mathbf{t}$ is given by

$$\mathbf{\Xi}_{\text{opt}} = \sqrt{\alpha_q}\,\mathrm{diag}(\mathbf{C_{xx}})^{1/2}. \tag{3.64}$$

3.5 Statistical Properties of the Optimal CEQ

The statistical properties of the optimal CEQ can be first obtained from Price's theorem introduced in Section 3.3 by just plugging (3.64) in (3.51) and (3.52). A second way to obtain the approximate statistical properties of the CEQ is the LCA that was applied in [25] and will be recalled below. One might ask why the LCA is only considered in the case of the optimal CEQ. The answer to this question is that the covariance factor between the output

signal t and the quantization distortion term η_q vanishes under the optimality condition of the CEQ as shown in (3.60). This fact simplifies the derivations and allows us to get closed-form approximations of the statistical properties of the CEQ.

To this end, we again assume an N-dimensional CEQ with joint Gaussian distributed input signals $x_n,\ n = 1, \cdots, N$.

3.5.1 Price's Theorem

When plugging (3.64) in (3.51) and (3.52), we get

$$\mathbf{C_{tx}} = \alpha_q \mathbf{C_{xx}}, \tag{3.65}$$

and

$$\mathbf{C_{tt}} = \frac{Q}{\pi} \sin^2(\psi) \alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \arcsin\left(\Re\{\mathbf{R_{xx}}\, \mathrm{e}^{-\mathrm{j}\, 2\Delta k\psi}\}\right) \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2}. \tag{3.66}$$

3.5.2 LCA

To find the corresponding expressions of the covariance matrices when using the LCA, we write first the general expressions

$$\mathbf{C_{tx}} = \mathbf{C_{xx}} + \mathbf{C_{\eta_q x}}, \tag{3.67}$$

and

$$\mathbf{C_{tt}} = \mathbf{C_{xx}} + \mathbf{C_{x\eta_q}} + \mathbf{C_{\eta_q x}} + \mathbf{C_{\eta_q \eta_q}}. \tag{3.68}$$

To this end, we have to find the entries of $\mathbf{C_{x\eta_q}}$ and $\mathbf{C_{\eta_q\eta_q}}$. First, the correlation factor between the input signal x_i and the quantization distortion term η_{q_j} reads as

$$\begin{aligned}
\mathrm{E}\left[x_i \eta_{q_j}^*\right] &= \mathrm{E}_{x_j}\left[\mathrm{E}\left[x_i \eta_{q_j}^* | x_j\right]\right] \\
&\overset{(a)}{=} \mathrm{E}_{x_j}\left[\mathrm{E}\left[x_i | x_j\right] \mathrm{E}\left[\eta_{q_j}^* | x_j\right]\right] \\
&\overset{(b)}{=} \mathrm{E}_{x_j}\left[\mathrm{E}\left[x_i x_j^*\right] \mathrm{E}\left[x_j x_j^*\right]^{-1} x_j\, \mathrm{E}\left[\eta_{q_j}^* | x_j\right]\right] \\
&= \mathrm{E}\left[x_i x_j^*\right] \mathrm{E}\left[x_j x_j^*\right]^{-1} \mathrm{E}\left[x_j \eta_{q_j}^*\right] \\
&\overset{(c)}{=} -\beta_q \mathrm{E}\left[x_i x_j^*\right],
\end{aligned} \tag{3.69}$$

where in (a) we use the fact that the quantization distortion term η_{q_j} does not statistically depend on the other random variables when conditioned on x_j, in (b) the Bayesian estimator is equal to the linear estimator for jointly Gaussian distributed signals x_i and x_j and (c) results from (3.61). In summary, it holds that

$$\mathbf{C_{x\eta_q}} = -\beta_q \mathbf{C_{xx}}. \tag{3.70}$$

Second, the correlation factor between two quantization distortion terms η_i and η_j, with $i \neq j$, reads as

$$\begin{aligned}
\mathrm{E}\left[\eta_{q_i}\eta_{q_j}^*\right] &= \mathrm{E}_{x_j}\left[\mathrm{E}\left[\eta_{q_i}\eta_{q_j}^*|x_j\right]\right] \\
&= \mathrm{E}_{x_j}\left[\mathrm{E}\left[\eta_{q_i}|x_j\right]\mathrm{E}\left[\eta_{q_j}^*|x_j\right]\right] \\
&\overset{(d)}{\approx} \mathrm{E}_{x_j}\left[\mathrm{E}\left[\eta_{q_i}x_j^*\right]\mathrm{E}\left[x_jx_j^*\right]^{-1}x_j\,\mathrm{E}\left[\eta_{q_j}^*|x_j\right]\right] \\
&= \mathrm{E}\left[\eta_{q_i}x_j^*\right]\mathrm{E}\left[x_jx_j^*\right]^{-1}\mathrm{E}\left[x_j\eta_{q_j}^*\right] \\
&= -\beta_q\,\mathrm{E}\left[\eta_{q_i}x_j^*\right] \\
&\overset{(3.61)}{=} -\beta_q\left(-\beta_q\,\mathrm{E}\left[x_ix_j^*\right]\right) \\
&= \beta_q^2\,\mathrm{E}\left[x_ix_j^*\right], \qquad (3.71)
\end{aligned}$$

where in (d) the Bayesian estimator is approximated with the linear estimator by assuming that η_{q_i} is Gaussian distributed. The power of the quantization distortion term is related to the input signal variance as follows

$$\mathrm{E}\left[\eta_{q_i}\eta_{q_i}^*\right] = \sigma_{\eta_{q_i}}^2 = \beta_q\sigma_{x_i}^2. \qquad (3.72)$$

Thus, we obtain

$$\mathbf{C}_{\boldsymbol{\eta_q\eta_q}} = \beta_q^2\mathbf{C}_{\mathbf{xx}} + \alpha_q\beta_q\,\mathrm{diag}\left(\mathbf{C}_{\mathbf{xx}}\right). \qquad (3.73)$$

Plugging (3.70) and (3.73) in (3.67) and (3.68), we get

$$\mathbf{C}_{\mathbf{tx}} = \alpha_q\mathbf{C}_{\mathbf{xx}}, \qquad (3.74)$$

and

$$\begin{aligned}
\mathbf{C}_{\mathbf{tt}} &= \alpha_q^2\mathbf{C}_{\mathbf{xx}} + \alpha_q\beta_q\,\mathrm{diag}\left(\mathbf{C}_{\mathbf{xx}}\right) \\
&= \alpha_q\,\mathrm{diag}\left(\mathbf{C}_{\mathbf{xx}}\right) + \alpha_q^2\,\mathrm{nondiag}\left(\mathbf{C}_{\mathbf{xx}}\right). \qquad (3.75)
\end{aligned}$$

3.5.3 Price's Theorem vs. LCA

When we compare the expressions obtained by Price's Theorem with the expressions obtained by the LCA, we observe that the only difference is in the computation of $\mathbf{C}_{\mathbf{tt}}$. To discover the relationship between both methods, we first split the expression in (3.66) into diagonal

and non-diagonal parts

$$\begin{aligned}
\mathbf{C_{tt}} &= \frac{Q}{\pi}\sin^2(\psi)\alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \arcsin\left(\Re\{\operatorname{diag}\left(\mathbf{R_{xx}}\right)\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right) \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \\
&+ \frac{Q}{\pi}\sin^2(\psi)\alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \arcsin\left(\Re\{\operatorname{nondiag}\left(\mathbf{R_{xx}}\right)\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right) \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \\
&= \frac{Q}{\pi}\sin^2(\psi)\alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \arcsin\left(\cos\left(2\Delta k\psi\right)\mathbf{I}_N\right) \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \\
&+ \frac{Q}{\pi}\sin^2(\psi)\alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \arcsin\left(\Re\{\operatorname{nondiag}\left(\mathbf{R_{xx}}\right)\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right) \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \\
&= \frac{Q}{\pi}\sin^2(\psi)\alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \left(\frac{\pi}{2} - 2\Delta k\psi\right)\mathbf{I}_N \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \\
&+ \frac{Q}{\pi}\sin^2(\psi)\alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \arcsin\left(\Re\{\operatorname{nondiag}\left(\mathbf{R_{xx}}\right)\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right) \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \\
&\overset{(e)}{=} \alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right) \\
&+ \frac{Q}{\pi}\sin^2(\psi)\alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \arcsin\left(\Re\{\operatorname{nondiag}\left(\mathbf{R_{xx}}\right)\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right) \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{1/2},
\end{aligned} \tag{3.76}$$

where in (e) we made use of the property

$$\frac{Q}{\pi}\sin^2(\psi) \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \left(\frac{\pi}{2} - 2\Delta k\psi\right) = 1. \tag{3.77}$$

Second we approximate the $\arcsin(\bullet)$ function by its first order Taylor expansion and obtain

$$\begin{aligned}
\mathbf{C_{tt}} &\approx \alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right) + \frac{Q}{\pi}\sin^2(\psi)\alpha_q \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \Re\{\operatorname{nondiag}\left(\mathbf{C_{xx}}\right)\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\} \\
&\overset{(g)}{=} \alpha_q \operatorname{diag}\left(\mathbf{C_{xx}}\right) + \alpha_q^2 \operatorname{nondiag}\left(\mathbf{C_{xx}}\right),
\end{aligned} \tag{3.78}$$

where in (g) it holds that

$$\sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \Re\{x\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\} = \frac{Q}{4}x. \tag{3.79}$$

Hence, we get the same expression as in (3.75). This implies that the approximation of the non-diagonal entries in $\mathbf{C_{tt}}$ from (3.66) based on the first order Taylor expansion of $\arcsin(\bullet)$ leads to the same expression as in (3.75) for the LCA. In other words, the LCA represents the first order Taylor expansion of Price's Theorem.

3.6 Bussgang Linearization of the CEQ

Theorem 2 (Bussgang's Theorem [61])**.** *For two Gaussian distributed signals, the crosscorrelation function taken after one of them has undergone non-linear amplitude distortion is identical, except for a factor of proportionality, to the crosscorrelation function taken before the distortion.*

According to Bussgang's Theorem [61], it holds that

$$\mathbf{C_{tx}} = \mathbf{L}_{\mathcal{Q}}\mathbf{C_{xx}}. \tag{3.80}$$

Consequently, every non-linear function with Gaussian distributed input signal can be modeled by the sum of a linear function and a distortion term that is uncorrelated with the input as

$$\mathbf{t} = \mathbf{L}_{\mathcal{Q}}\mathbf{x} + \mathbf{d}_{\mathcal{Q}}, \tag{3.81}$$

such that

$$\mathrm{E}\left[\mathbf{x}\mathbf{d}_{\mathcal{Q}}^{\mathrm{H}}\right] = \mathbf{0}. \tag{3.82}$$

This leads us to the result

$$\begin{aligned}\mathbf{L}_{\mathcal{Q}} &= \mathbf{C_{tx}}\mathbf{C_{xx}}^{-1} \\ &= \sqrt{\alpha_q}\boldsymbol{\Xi}\,\mathrm{diag}\left(\mathbf{C_{xx}}\right)^{-\frac{1}{2}}\mathbf{C_{xx}}\mathbf{C_{xx}}^{-1} \\ &= \sqrt{\alpha_q}\boldsymbol{\Xi}\,\mathrm{diag}\left(\mathbf{C_{xx}}\right)^{-\frac{1}{2}}.\end{aligned} \tag{3.83}$$

For the optimal CEQ, (3.64) can be plugged in (3.83) and the expression in (3.83) reduces to

$$\mathbf{L}_{\mathcal{Q}} = \alpha_q\mathbf{I}_N. \tag{3.84}$$

Indeed, when we plug (3.83) in (3.80) and compare the obtained expression with (3.51), we notice that Bussgang's theorem is a special case of Price' Theorem. Note that $\mathbf{d}_{\mathcal{Q}}$ is not Gaussian distributed but has an unknown distribution. However, the covariance matrix of $\mathbf{d}_{\mathcal{Q}}$ can be computed whether by applying Price's theorem or the LCA. First, with Price's theorem, we get

$$\begin{aligned}\mathbf{C}_{\mathbf{d}_{\mathcal{Q}}\mathbf{d}_{\mathcal{Q}}} &= \mathrm{E}\left[\left(\mathbf{t}-\mathbf{L}_{\mathcal{Q}}\mathbf{x}\right)\left(\mathbf{t}-\mathbf{L}_{\mathcal{Q}}\mathbf{x}\right)^{\mathrm{H}}\right] \\ &= \mathbf{C_{tt}} - \mathbf{L}_{\mathcal{Q}}\mathbf{C_{xt}} - \mathbf{C_{tx}}\mathbf{L}_{\mathcal{Q}}^{\mathrm{H}} + \mathbf{L}_{\mathcal{Q}}\mathbf{C_{xx}}\mathbf{L}_{\mathcal{Q}}^{\mathrm{H}} \\ &= \mathbf{C_{tt}} - \mathbf{L}_{\mathcal{Q}}\mathbf{C_{xx}}\mathbf{L}_{\mathcal{Q}}^{\mathrm{H}} \\ &= \mathbf{C_{tt}} - \alpha_q\boldsymbol{\Xi}\mathbf{R_{xx}}\boldsymbol{\Xi} \\ &= \frac{Q}{\pi}\sin^2(\psi)\sum_{\Delta k=0}^{Q/2-1}\mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\,\boldsymbol{\Xi}\arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}(2\Delta k\psi)}\}\right)\boldsymbol{\Xi} - \alpha_q\boldsymbol{\Xi}\mathbf{R_{xx}}\boldsymbol{\Xi} \\ &= \beta_q\boldsymbol{\Xi}^2 + \frac{Q}{\pi}\sin^2(\psi)\sum_{\Delta k=0}^{Q/2-1}\mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\,\boldsymbol{\Xi}\arcsin\left(\Re\{\mathrm{nondiag}\left(\mathbf{R_{xx}}\right)\mathrm{e}^{-\mathrm{j}(2\Delta k\psi)}\}\right)\boldsymbol{\Xi} \\ &\quad - \alpha_q\boldsymbol{\Xi}\,\mathrm{nondiag}\left(\mathbf{R_{xx}}\right)\boldsymbol{\Xi} \\ &\approx \beta_q\boldsymbol{\Xi}^2,\end{aligned} \tag{3.85}$$

where we use the first order Taylor expansion of the arcsin function. For the optimal CEQ, we get

$$\mathbf{C}_{\mathbf{d}_\mathcal{Q}\mathbf{d}_\mathcal{Q}} \approx \alpha_q \beta_q \operatorname{diag}\left(\mathbf{C}_{\mathbf{xx}}\right), \tag{3.86}$$

which matches with the result when using the LCA.

In the following, we aim at getting the covariance matrix for different time instants. When applying Price's theorem, we get

$$\begin{aligned}
\mathrm{E}\left[\mathbf{d}_\mathcal{Q}[t-t_1]\mathbf{d}_\mathcal{Q}^\mathrm{H}[t-t_2]\right] &= \mathrm{E}\left[\left(\mathbf{t}[t-t_1]-\mathbf{L}_\mathcal{Q}\mathbf{x}[t-t_1]\right)\left(\mathbf{t}[t-t_2]-\mathbf{L}_\mathcal{Q}\mathbf{x}[t-t_2]\right)^\mathrm{H}\right] \\
&= \mathrm{E}\left[\mathbf{t}[t-t_1]\mathbf{t}^\mathrm{H}[t-t_2]\right] - \mathrm{E}\left[\mathbf{t}[t-t_1]\mathbf{x}^\mathrm{H}[t-t_2]\right]\mathbf{L}_\mathcal{Q}^\mathrm{H} \\
&\quad - \mathbf{L}_\mathcal{Q}\,\mathrm{E}\left[\mathbf{x}[t-t_1]\mathbf{t}^\mathrm{H}[t-t_2]\right] + \mathbf{L}_\mathcal{Q}\,\mathrm{E}\left[\mathbf{x}[t-t_1]\mathbf{x}^\mathrm{H}[t-t_2]\right]\mathbf{L}_\mathcal{Q}^\mathrm{H} \\
&\overset{(3.80)}{=} \mathrm{E}\left[\mathbf{t}[t-t_1]\mathbf{t}^\mathrm{H}[t-t_2]\right] - \mathbf{L}_\mathcal{Q}\,\mathrm{E}\left[\mathbf{x}[t-t_1]\mathbf{x}^\mathrm{H}[t-t_2]\right]\mathbf{L}_\mathcal{Q}^\mathrm{H} \\
&\quad - \mathbf{L}_\mathcal{Q}\,\mathrm{E}\left[\mathbf{x}[t-t_1]\mathbf{x}^\mathrm{H}[t-t_2]\right]\mathbf{L}_\mathcal{Q}^\mathrm{H} + \mathbf{L}_\mathcal{Q}\,\mathrm{E}\left[\mathbf{x}[t-t_1]\mathbf{x}^\mathrm{H}[t-t_2]\right]\mathbf{L}_\mathcal{Q}^\mathrm{H} \\
&= \mathrm{E}\left[\mathbf{t}[t-t_1]\mathbf{t}^\mathrm{H}[t-t_2]\right] - \mathbf{L}_\mathcal{Q}\,\mathrm{E}\left[\mathbf{x}[t-t_1]\mathbf{x}^\mathrm{H}[t-t_2]\right]\mathbf{L}_\mathcal{Q}^\mathrm{H} \\
&= \frac{Q}{\pi}\sin^2(\psi)\sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}\,2\Delta k\psi}\,\boldsymbol{\Xi} \\
&\quad \arcsin\left(\Re\left\{\operatorname{diag}(\mathbf{C}_{\mathbf{xx}})^{-1/2}\,\mathrm{E}\left[\mathbf{x}[t-t_1]\mathbf{x}^\mathrm{H}[t-t_2]\right]\operatorname{diag}(\mathbf{C}_{\mathbf{xx}})^{-1/2}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\right\}\right) \\
&\quad \boldsymbol{\Xi} - \alpha_q\boldsymbol{\Xi}\operatorname{diag}(\mathbf{C}_{\mathbf{xx}})^{-1/2}\,\mathrm{E}\left[\mathbf{x}[t-t_1]\mathbf{x}^\mathrm{H}[t-t_2]\right]\operatorname{diag}(\mathbf{C}_{\mathbf{xx}})^{-1/2}\boldsymbol{\Xi}.
\end{aligned} \tag{3.87}$$

Again, with the first order Taylor expansion of the arcsin function, we get

$$\mathrm{E}\left[\mathbf{d}_\mathcal{Q}[t-t_1]\mathbf{d}_\mathcal{Q}^\mathrm{H}[t-t_2]\right] \approx \begin{cases} \beta_q\boldsymbol{\Xi}^2 & \text{if } t_1 = t_2, \\ \mathbf{0}_{N,N} & \text{otherwise.} \end{cases} \tag{3.88}$$

For the optimal CEQ, it holds that

$$\mathrm{E}\left[\mathbf{d}_\mathcal{Q}[t-t_1]\mathbf{d}_\mathcal{Q}^\mathrm{H}[t-t_2]\right] \approx \begin{cases} \alpha_q\beta_q \operatorname{diag}\left(\mathbf{C}_{\mathbf{xx}}\right) & \text{if } t_1 = t_2, \\ \mathbf{0}_{N,N} & \text{otherwise,} \end{cases} \tag{3.89}$$

where this approximation can be also obtained by the LCA.

4. System Model

4.1 Input-Output Relationshsip

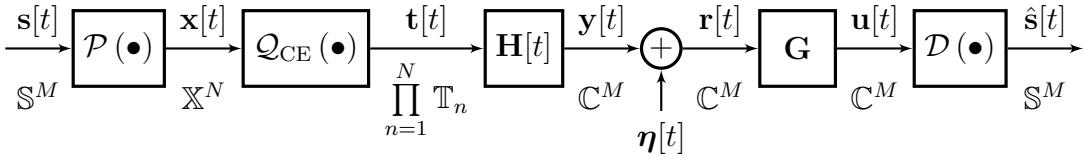


Fig. 4.1: Downlink MU-MIMO system model.

The system model shown in Fig.4.1 consists of a single-cell massive MU-MIMO downlink scenario with coarsely QCE signals at the transmitter. The BS is equipped with N antennas and serves M single-antenna users simultaneously, where $N \gg M$.
The input signal $\mathbf{s}[t] \in \mathbb{S}^M$ contains the symbols to be transmitted to each of the M users at time instant t. Each user's signal is drawn from the set $\mathbb{S}$ that represents either an S-PSK or S-QAM constellation, where S denotes the number of constellation points. The input set is detailed Section 4.3. We assume that $\mathrm{E}[\mathbf{s}[t]] = \mathbf{0}_M$, $\mathrm{E}[\mathbf{s}[t]\mathbf{s}[t]^{\mathrm{H}}] = \sigma_{\mathrm{s}}^2\mathbf{I}_M$, and $\mathrm{E}[\mathbf{s}[t]\mathbf{s}[t-\tau]^{\mathrm{H}}] = \mathbf{0}_M$, $\forall t$ and $\tau \neq t$.
The signal vector $\mathbf{s}[t]$ is precoded into the vector $\mathbf{x}[t] \in \mathbb{X}^N$ prior to the polar DACs. The set $\mathbb{X}$ can represent the set of complex numbers, i.e. $\mathbb{C}$ or a subset of it depending on the precoder design. The choice of $\mathbb{X}$ will be discussed in the following sections. The function $\mathcal{P}(\bullet)$ represents the precoding operation to reduce the distortions caused by the coarse quantization and the MUI. A brief overview about the precoding function $\mathcal{P}(\bullet)$ is given in Section 4.4. The whole thesis is devoted to explain in detail the precoder design.
The operator $\mathcal{Q}_{\mathrm{CE}}(\bullet)$, defined in Chapter 3, models the non-linear behavior of the polar DACs combined with the power allocation at the PAs as

$$\mathbf{t}[t] = \mathcal{Q}_{\mathrm{CE}}(\mathbf{x}[t]). \tag{4.1}$$

After the CEQ each entry of the transmit vector t_n, $n = 1, \cdots, N$, belongs to the set $\mathbb{T}_n$, which is defined as

$$\mathbb{T}_n = \left\{ \Xi_n \exp\left(\mathrm{j}\,(2i-1)\,\frac{\pi}{Q} \right) : i = 1, \cdots, Q \right\}, \tag{4.2}$$

where Ξ_n^2 denotes the coefficient for the power allocation at the n-th antenna that is chosen according to the transmit power constraint detailed in Section 4.5.

The signal $\mathbf{t}[t]$ is transmitted through the channel modeled by

$$\mathbf{H}[t] = \sum_{\ell=0}^{L-1} \mathbf{H}_\ell \delta[t-\ell], \tag{4.3}$$

where δ represents the Dirac function and with the (m,n)-th element of the ℓ-th channel matrix, i.e. $[\mathbf{H}_\ell]_{m,n}$, being the ℓ-th channel tap among L taps between the n-th transmit antenna and the m-th user. At the M receive antennas, an AWGN, which is denoted by the vector $\boldsymbol{\eta} \sim \mathcal{CN}_{\mathbb{C}}\left(\mathbf{0}_M, \mathbf{C}_{\boldsymbol{\eta}} = \mathbf{I}_M\right)$, perturbs the received signals

$$\mathbf{r}[t] = \mathbf{H}[t] * \mathbf{t}[t] + \boldsymbol{\eta}[t]. \tag{4.4}$$

The precoder is designed such that, without any noise, the received signals would lie in their intended decision regions and no joint receive processing is necessary. Additionally, coherent data transmission with multiple BS antennas leads to an antenna gain, which depends on the channel realization. The entries of the received signal vector $\mathbf{r}$ do not belong to the nominal decision regions of $\mathbb{S}$ but to a scaled version of them. Therefore, rescaling the received signal at each receive antenna is required to make the signal belong to the nominal decision region. The rescaling operation is modeled by the diagonal real-valued matrix $\mathbf{G}$, as follows

$$\mathbf{u}[t] = \mathbf{G}\left(\mathbf{H}[t] * \mathbf{t}[t] + \boldsymbol{\eta}[t]\right), \tag{4.5}$$

where

$$\mathbf{G} = \sum_{m=1}^{M} g_m \mathbf{e}_m \mathbf{e}_m^{\mathrm{T}}, \tag{4.6}$$

with $g_m > 0,\ m = 0, \cdots, M$. The coefficients g_m are blindly estimated at the receiver over a block of T received symbols as explained in Section 4.7. Note that no receive processing $\mathbf{G}$ is required if $\mathbb{S}$ represents the PSK constellation. Finally, based on the decision regions to which the entries of the signal $\mathbf{u}$ belong, the decision operation $\mathcal{D}(\bullet)$ produces the detected symbols $\hat{\mathbf{s}}$ at the users

$$\hat{\mathbf{s}}[t] = \mathcal{D}\left(\mathbf{G}\left(\mathbf{H}[t] * \mathbf{t}[t] + \boldsymbol{\eta}[t]\right)\right). \tag{4.7}$$

4.2 Compact Input-Output Relationship

We drop out the time index and express $\mathbf{u} = \mathbf{u}[t]$ compactly as

$$\mathbf{u} = \mathbf{G}\left(\mathbf{H}\mathbf{t} + \boldsymbol{\eta}\right), \tag{4.8}$$

where

$$\mathbf{H} = \begin{bmatrix}\mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_{L-1}\end{bmatrix}, \tag{4.9}$$

$$\mathbf{t} = \begin{bmatrix}\mathbf{t}[t]^{\mathrm{T}} & \mathbf{t}[t-1]^{\mathrm{T}} & \cdots & \mathbf{t}[t-L+1]^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} = \mathcal{Q}_{\mathrm{CE}}\left(\mathbf{x}\right), \tag{4.10}$$

$$\mathbf{x} = \begin{bmatrix}\mathbf{x}[t]^{\mathrm{T}} & \mathbf{x}[t-1]^{\mathrm{T}} & \cdots & \mathbf{x}[t-L+1]^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}, \tag{4.11}$$

and

$$\boldsymbol{\eta} = \boldsymbol{\eta}[t]. \tag{4.12}$$

4.3 Input Set

Within the scope of the thesis, the input signals can be drawn from an S-PSK constellation or an S-QAM constellation. To this end, we define both constellations. First, the S-PSK constellation is given by

$$\mathbb{S} := \{\exp(\mathrm{j}(2i-1)\theta) : i = 1, \cdots, S\}, \text{ where } \theta = \frac{\pi}{S}. \tag{4.13}$$

Second, the S-QAM constellation, where S is assumed to be a power of 4, is defined as

$$\mathbb{S} := \{\pm(2i-1) \pm \mathrm{j}(2i-1) : i = 1, \cdots, \log_4(S)\}. \tag{4.14}$$

4.4 Transmit Processing: Precoding

4.4.1 Linear Precoding

For linear precoding techniques, the precoding function reduces to a Finite Impulse Response (FIR) matrix; that is

$$\mathbf{P}[t] = \sum_{\ell'=0}^{L_p-1} \mathbf{P}_{\ell'} \delta[t-\ell'], \tag{4.15}$$

where L_p represents the number of taps of the precoding FIR filter between every antenna element and every user.

4.4.2 Non-linear Precoding

In the case of non-linear precoding, no explicit expression for $\mathcal{P}(\bullet)$ is provided. However, for a given channel realization $\mathbf{H}$, every input vector $\mathbf{s}$ is mapped to a precoded vector $\mathbf{x}$.

$$\mathbf{x} = \mathcal{P}(\mathbf{s}, \mathbf{H}). \tag{4.16}$$

4.5 Transmit Power Constraint

For an available transmit power P_{tx}, the transmit power constraint is given by

$$\mathrm{tr}(\mathbf{C_{tt}}) \leq LP_{\text{tx}}. \tag{4.17}$$

Therefore, we should compute $\mathrm{tr}\left(\mathbf{C_{tt}}\right)$. Applying (3.52), we obtain

$$\begin{aligned}
\mathrm{tr}\left(\mathbf{C_{tt}}\right) &= \frac{Q}{\pi}\sin^2\left(\psi\right)\sum_{\Delta k=0}^{Q/2-1}\mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\,\mathrm{tr}\left(\boldsymbol{\Xi}\arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\boldsymbol{\Xi}\right)\\
&= \frac{Q}{\pi}\sin^2\left(\psi\right)\sum_{\Delta k=0}^{Q/2-1}\mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\,\mathrm{tr}\left(\mathrm{diag}\left(\arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\right)\boldsymbol{\Xi}^2\right)\\
&= \frac{Q}{\pi}\sin^2\left(\psi\right)\sum_{\Delta k=0}^{Q/2-1}\mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\,\mathrm{tr}\left(\arcsin\left(\cos\left(2\Delta k\psi\right)\mathbf{I}_N\right)\boldsymbol{\Xi}^2\right)\\
&= \frac{Q}{\pi}\sin^2\left(\psi\right)\sum_{\Delta k=0}^{Q/2-1}\mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\,\mathrm{tr}\left(\left(\frac{\pi}{2}-2\Delta k\psi\right)\mathbf{I}_N\boldsymbol{\Xi}^2\right)\\
&= \frac{Q}{\pi}\sin^2\left(\psi\right)\sum_{\Delta k=0}^{Q/2-1}\left(\frac{\pi}{2}-2\Delta k\psi\right)\mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\,\mathrm{tr}\left(\boldsymbol{\Xi}^2\right)\\
&\overset{(3.77)}{=} \mathrm{tr}\left(\boldsymbol{\Xi}^2\right).
\end{aligned} \tag{4.18}$$

Hence, it must hold that

$$\mathrm{tr}\left(\boldsymbol{\Xi}^2\right) \leq LP_{\mathrm{tx}}. \tag{4.19}$$

We choose $\boldsymbol{\Xi} = \sqrt{\alpha_q}\,\mathrm{diag}(\mathbf{C_{xx}})^{1/2}$ to reduce the quantization distortions, as stated in (3.64). Consequently, the transmit power constraint can be rewritten as

$$\alpha_q\,\mathrm{tr}\left(\mathbf{C_{xx}}\right) \leq LP_{\mathrm{tx}}. \tag{4.20}$$

4.5.1 Equal Power Allocation

In the case of equal power allocation, the matrix $\boldsymbol{\Xi}$ is just a scaled identity, i.e. $\boldsymbol{\Xi} = \Xi\mathbf{I}_{NL}$. This leads to

$$\mathrm{diag}\left(\mathbf{C_{xx}}\right) = \frac{\Xi^2}{\alpha_q}\mathbf{I}_{NL} \tag{4.21}$$

due to the optimality condition of the CEQ in (3.64). The value of Ξ is found by fulfilling the power constraint in (4.19)

$$\Xi \leq \sqrt{\frac{P_{\mathrm{tx}}}{N}}. \tag{4.22}$$

To exploit the total available transmit power we choose

$$\Xi = \sqrt{\frac{P_{\mathrm{tx}}}{N}}. \tag{4.23}$$

Plugging (4.23) in (4.21), it follows that

$$\mathrm{diag}\left(\mathbf{C_{xx}}\right) = \frac{P_{\mathrm{tx}}}{\alpha_q N}\mathbf{I}_{NL}. \tag{4.24}$$

4.5.2 Unequal Power Allocation

In the case of unequal power allocation, i.e. $\boldsymbol{\Xi} \neq \Xi \mathbf{I}_N$, and for maximal exploitation of the total available transmit power, it must hold that

$$\mathrm{tr}\left(\mathbf{C}_{\mathbf{xx}}\right) = \frac{LP_{\mathrm{tx}}}{\alpha_q}. \tag{4.25}$$

4.6 Channel Model

4.6.1 The i.i.d. Channel

The first channel model that we consider throughout this thesis is the simplistic i.i.d. channel, where the entries of the matrices $\mathbf{H}_\ell$, $\ell = 0, \cdots, L-1$, are of zero means and constant variances; that is

$$\left[\mathbf{H}_\ell\right]_{m,n} = \mathcal{CN}_{\mathbb{C}}\left(0, \sigma_h^{(\ell)2}\right),\ m = 1, \cdots, M,\ n = 1, \cdots, N, \tag{4.26}$$

where it must hold that

$$\sum_{\ell=0}^{L-1} \sigma_h^{(\ell)2} = 1. \tag{4.27}$$

In the case of $L = 1$, we obtain that $\sigma_h^2 = 1$. However, in the case of $L \neq 1$, the variances are defined by the exponential power delay profiles that are given in Table 4.1 and Table 4.2 for $L = 3$ and $L = 6$, respectively.

ℓ	1	2
$10\log_{10}\left(\frac{\mathrm{E}[\|\mathbf{H}_\ell\|_{\mathrm{F}}^2]}{\mathrm{E}[\|\mathbf{H}_0\|_{\mathrm{F}}^2]}\right)$ /dB	-3	-6

Table 4.1: Exponential power delay profile with $L = 3$.

4.6.2 The mmW Sparse Channel

A more realistic channel model that takes into account the use of mmW frequencies is the mmW sparse channel, which is a clustered channel based on the extended Saleh-Valenzuela model [62]. Experiments and measurements have shown that the mmW channel can be modeled by a number of clusters N_{cl} for each user that group rays of number N_{ray} departing from the BS in similar directions [63–66]. Mathematically, we can write that

$$\mathbf{e}_m^{\mathrm{T}}\mathbf{H}_\ell = \sqrt{\frac{N}{N_{\mathrm{cl}}N_{\mathrm{ray}}}} \sum_{i=1}^{N_{\mathrm{cl}}} \sum_{j=1}^{N_{\mathrm{ray}}} \gamma_{i,j}^{(m,\ell)} \mathbf{a}\left(\phi_{i,j}^{(\ell)}\right)^{\mathrm{H}}, \tag{4.28}$$

ℓ	1	2	3	4	5
$10\log_{10}\left(\frac{\mathrm{E}[\|\mathbf{H}_\ell\|_{\mathrm{F}}^2]}{\mathrm{E}[\|\mathbf{H}_0\|_{\mathrm{F}}^2]}\right)$ /dB	-1	-9	-10	-15	-20

Table 4.2: Vehicular A (Case II) power delay profile with $L = 6$.

where $\gamma_{i,j}^{(m,\ell)}$ denotes the complex small-scale fading gain on the j-th subpath of the i-th cluster at the ℓ-th main path for the m-th user and $\mathbf{a}(\phi)$ is the vector response function of the BS antenna arrays to the angular departures; that is, when we assume a uniform linear array, we get

$$\mathbf{a}(\phi) = \frac{1}{\sqrt{N}} \left[1 \quad \mathrm{e}^{\mathrm{j}\,\pi \sin(\phi)} \quad \cdots \quad \mathrm{e}^{\mathrm{j}(N-1)\pi \sin(\phi)}\right]^{\mathrm{T}}. \tag{4.29}$$

Each i-th cluster is defined by a certain mean Angle of Departure (AOD) that is drawn from a uniform distribution. The AODs that correspond to the rays within the same cluster, i.e. $\phi_{i,j}^{(\ell)},\ j = 1, \cdots, N_{\mathrm{ray}}$, are drawn around the mean AOD from a truncated Laplacian PDF given by [67]

$$p_\phi(\phi) = \begin{cases} \frac{1}{\sqrt{2}\sigma_\phi\left(1-\mathrm{e}^{-\sqrt{2}\pi/\sigma_\phi}\right)} \mathrm{e}^{-\left|\sqrt{2}\phi/\sigma_\phi\right|} & \text{if } \phi \in [-\pi, \pi), \\ 0 & \text{else.} \end{cases} \tag{4.30}$$

In this work, we assume that

$$\sigma_\phi = 1^\circ. \tag{4.31}$$

The small-scale fading factors fulfill the following property

$$\gamma_{i,j}^{(m,\ell)} \sim \mathcal{CN}_{\mathbb{C}}\left(0, \sigma_\gamma^{(\ell)2}\right), \tag{4.32}$$

where $\sigma_\gamma^{(\ell)2}$ are chosen according to the exponential power delay profile either in Table 4.1 or in Table 4.2. Additionally, it must hold that

$$\sum_{\ell=0}^{L-1} \sigma_\gamma^{(\ell)2} = 1. \tag{4.33}$$

4.7 Receive Processing

After multiplication with the receiver coefficient g_m, the m-th scaled received signal is

$$u_m[t] = g_m r_m[t] = g_m\, \mathrm{e}_m^{\mathrm{T}}\, \mathbf{H}[t] * \mathbf{t}[t] + g_m \eta_m[t] = s_m[t-\tau] + \eta'_m[t], \tag{4.34}$$

where τ represents the time delay and $\eta'_m[t]$ denotes the deviation of $u_m[t]$ from the nominal point $s_m[t-\tau]$ due to the precoder design, the AWGN $\eta_m[t]$ and the quantization applied on the precoded vector $\mathbf{x}$. Then, we can write

$$|\Re\{r_m[t]\}| + |\mathrm{Im}\{r_m[t]\}| = g_m^{-1}\left(|\Re\{s_m[t-\tau] + \eta'_m[t]\}| + |\Im\{s_m[t-\tau] + \eta'_m[t]\}|\right) \tag{4.35}$$

$$\approx g_m^{-1}\left(|\Re\{s_m[t-\tau]\}| + |\Im\{s_m[t-\tau]\}| + \Re\{\eta'_m[t]\} + \Im\{\eta'_m[t]\}\right), \tag{4.36}$$

where the approximation is very accurate when the receiver Signal-to-Interference-Noise Ratio (SINR) is much larger than 1, which is the case for massive MIMO systems. Having zero-mean noise plus interference $\eta'_m[t]$, we get

$$\begin{aligned} \mathrm{E}[|\Re\{r_m[t]\}| + |\Im\{r_m[t]\}|] &\approx g_m^{-1}\mathrm{E}[|\Re\{s_m[t-\tau]\}| + |\Im\{s_m[t-\tau]\}|] \\ &\approx g_m^{-1}\sqrt{S}, \end{aligned} \tag{4.37}$$

when using the definition of the QAM constellation in (4.14). We recall that the receive processing is required only in the case of QAM signaling. Based on (4.37), we propose a blind estimation method to obtain the scaling factor g_m for each user prior to the decision operation; that is

$$g_m = T \cdot \frac{\sqrt{S}}{\sum_{t=1}^{T} |\Re\{r_m[t]\}| + |\Im\{r_m[t]\}|}, \tag{4.38}$$

where T is the length the received sequence. The method does not require any feedback or training from the BS nor any knowledge of the noise plus interference power at the user terminal.

4.8 Potential Dual Uplink System Model with the Optimal CEQ

As explained in Section 3.6, the CEQ can be linearized and the resulting linear system model with the optimal CEQ is shown in Fig. 4.2. The potential dual uplink system model is provided in Fig. 4.3. The MSE duality between both system models will be investigated throughout this thesis.
We have the following assumptions for the potential dual system model

$$\mathbf{F}_{\ell'} = \frac{1}{\beta}\mathbf{P}_{\ell'}^{\mathrm{H}},\ \ell' = 0, \cdots, L_p - 1, \tag{4.39}$$

$$\mathbf{T} = \beta \mathbf{C}_{\boldsymbol{\eta\eta}}^{1/2,\mathrm{H}} \mathbf{G}^{\mathrm{H}}, \tag{4.40}$$

$$\mathbf{C}_{\mathbf{ss}} = \mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}} = \sigma_s^2 \mathbf{I}_M, \tag{4.41}$$

and

$$\mathbf{C}_{\boldsymbol{\eta\eta}}^{\mathrm{UL}} = \mathbf{I}_N. \tag{4.42}$$

The transmit power constraint in (4.20) for maximal power exploitation can be rewritten as

$$\alpha_q \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1} \mathbf{P}_{\ell'} \mathbf{C}_{\mathbf{ss}} \mathbf{P}_{\ell'}^{\mathrm{H}}\right) = P_{\mathrm{tx}}. \tag{4.43}$$

Hence, the corresponding scaling factor β can be obtained by plugging (4.39) in the transmit power constraint in (4.43) as

$$\beta = \sqrt{\frac{P_{\mathrm{tx}}}{\alpha_q \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1} \mathbf{F}_{\ell'}^{\mathrm{H}} \mathbf{C}_{\mathbf{ss}} \mathbf{F}_{\ell'}\right)}}. \tag{4.44}$$

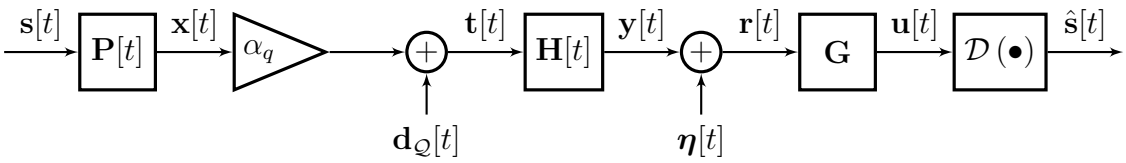


Fig. 4.2: Downlink system model with Bussgang decomposition.

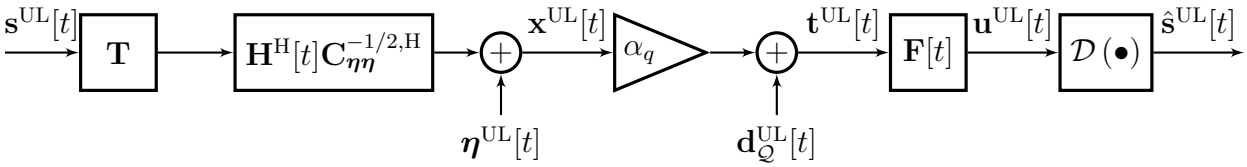


Fig. 4.3: Uplink system model with Bussgang decomposition.

Part I

Linear Transmit Signal Processing

5. Flat-Fading Channels

5.1 Input-Output Relationship

In this section, we consider linear precoding for flat-fading channels; that is $L = 1$. This implies that the precoder impluse response matrix $\mathbf{P}[t]$ reduces to the matrix $\mathbf{P}$. Accordingly, it holds that

$$\mathbf{x}[t] = \mathbf{P}\mathbf{s}[t]. \tag{5.1}$$

The time indexes are dropped out in the sequel. We recall the input-output relationship in (4.8)

$$\begin{aligned}\mathbf{u} &= \mathbf{G}\left(\mathbf{H}\mathbf{t} + \boldsymbol{\eta}\right)\\ &= \mathbf{G}\left(\mathbf{H}\mathcal{Q}_{\mathrm{CE}}\left(\mathbf{P}\mathbf{s}\right) + \boldsymbol{\eta}\right).\end{aligned} \tag{5.2}$$

5.2 Optimization Problem

The precoding task consists of finding the optimal linear precoder $\mathbf{P}_{\mathrm{opt}}$ and the diagonal positive real-valued receive filter $\mathbf{G}_{\mathrm{opt}}$ that minimize the MSE between the desired and the received signals, $\mathbf{s}$ and $\mathbf{u}$ under the transmit power constraint given in (4.20). The MMSE optimization problem is expressed as

$$\{\mathbf{P}_{\mathrm{opt}}, \mathbf{G}_{\mathrm{opt}}\} = \underset{\mathbf{P},\mathbf{G}}{\arg\min}\, \mathrm{E}\left[\|\mathbf{u} - \mathbf{s}\|_2^2\right] \qquad \text{s.t} \quad \alpha_q \,\mathrm{tr}\left(\mathbf{C}_{\mathbf{xx}}\right) \leq P_{\mathrm{tx}}. \tag{5.3}$$

5.2.1 MSE

In general, the MSE is expressed as

$$\mathrm{E}\left[\|\mathbf{u} - \mathbf{s}\|_2^2\right] = \mathrm{tr}\left(\mathbf{G}\mathbf{H}\mathbf{C}_{\mathbf{tt}}\mathbf{H}^{\mathrm{H}}\mathbf{G} + \mathbf{G}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\mathbf{G}\right) - \mathrm{tr}\left(\mathbf{G}\mathbf{H}\mathbf{C}_{\mathbf{ts}}\right) - \mathrm{tr}\left(\mathbf{C}_{\mathbf{st}}\mathbf{H}^{\mathrm{H}}\mathbf{G}\right) + \mathrm{tr}\left(\mathbf{C}_{\mathbf{ss}}\right). \tag{5.4}$$

Due to the central limit theorem [68], the entries in $\mathbf{x}$ are approximately Gaussian distributed for massive MIMO systems, where a large number of users is served by a large number of transmit antennas. Hence, Price's theorem can be applied to obtain the expressions of the covariance matrices $\mathbf{C}_{\mathbf{tt}}$ and $\mathbf{C}_{\mathbf{tx}}$ according to (3.66) and (3.65). Consequently, the MSE expression calculates to

$$\begin{aligned}\mathrm{E}\left[\|\mathbf{u} - \mathbf{s}\|_2^2\right] = \frac{Q}{\pi}\sin^2\left(\psi\right) \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\, \mathrm{tr}\left(\arcsin\left(\Re\{\mathbf{R}_{\mathbf{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\boldsymbol{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\boldsymbol{\Xi}\right)\\ + \mathrm{tr}\left(\mathbf{G}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\mathbf{G}\right) - \alpha_q\,\mathrm{tr}\left(\mathbf{G}\mathbf{H}\mathbf{P}\mathbf{C}_{\mathbf{ss}}\right) - \alpha_q\,\mathrm{tr}\left(\mathbf{C}_{\mathbf{ss}}\mathbf{P}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{G}\right) + \mathrm{tr}\left(\mathbf{C}_{\mathbf{ss}}\right).\end{aligned} \tag{5.5}$$

According to (5.1), the covariance matrix $\mathbf{C_{xx}}$ is expressed in terms of the precoder $\mathbf{P}$ as

$$\begin{aligned}\mathbf{C_{xx}} &= \mathrm{E}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right] \\ &= \mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}. \end{aligned} \tag{5.6}$$

Therefore, we obtain the following expressions

$$\mathbf{R_{xx}} = \mathrm{diag}\left(\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\right)^{-1/2}\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\,\mathrm{diag}\left(\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\right)^{-1/2}, \tag{5.7}$$

$$\mathbf{\Xi} = \mathrm{diag}\left(\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\right)^{1/2}. \tag{5.8}$$

5.2.2 Transmit Power Constraint

After plugging (5.6) in the transmit power constraint in (4.20) and for maximal exploitation of the available power, we get

$$\alpha_q\,\mathrm{tr}\left(\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\right) = P_{\mathrm{tx}}. \tag{5.9}$$

As explained in Section 4.5, we differentiate between two cases

- equal power allocation, i.e. $\mathrm{diag}\left(\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\right) = \frac{P_{\mathrm{tx}}}{\alpha_q N}\mathbf{I}_N$,

and

- unequal power allocation, i.e. $\mathrm{tr}\left(\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\right) = \frac{P_{\mathrm{tx}}}{\alpha_q}$.

5.3 Precoder Designs in the Primal Domain

5.3.1 Precoder Design Based on Gradient Projection Algorithm

Since the MSE expression in (5.5) is highly non-linear in $\mathbf{C_{xx}} = \mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}$ and thus in $\mathbf{P}$, we cannot find a closed-form expression for $\mathbf{P}_{\mathrm{opt}}$. Therefore, we use the Gradient-Projection algorithm as described in Algorithm 1 and Algorithm 2 for equal and unequal power allocation, respectively. To this end, we have to compute the derivatives of the MSE w.r.t. $\mathbf{P}$ and $\mathbf{G}$. First, the derivative of the MSE w.r.t. $\mathbf{P}$ is expressed as

$$\begin{aligned}\frac{\partial\,\mathrm{E}\left[\|\mathbf{u}-\mathbf{s}\|_2^2\right]}{\partial\mathbf{P}} &= \sum_{n=1}^{N}\sum_{m=1}^{M}\frac{\partial\,\mathrm{E}\left[\|\mathbf{u}-\mathbf{s}\|_2^2\right]}{\partial\mathbf{e}_n^{\mathrm{T}}\mathbf{P}\mathbf{e}_m}\mathbf{e}_n\mathbf{e}_m^{\mathrm{T}} \\ &= \frac{Q}{\pi}\sin^2(\psi) \\ &\quad \sum_{\Delta k=0}^{Q/2-1}\mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{\partial\,\mathrm{tr}\left(\arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\mathbf{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\mathbf{\Xi}\right)}{\partial\mathbf{e}_n^{\mathrm{T}}\mathbf{P}\mathbf{e}_m}\mathbf{e}_n\mathbf{e}_m^{\mathrm{T}} \\ &\quad - \alpha_q\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{C}_{\mathbf{ss}}^{*}. \end{aligned} \tag{5.10}$$

The challenging term in the above derivative expression is

$$\begin{aligned}\frac{\partial\,\mathrm{tr}\left(\arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\mathbf{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\mathbf{\Xi}\right)}{\partial\mathbf{e}_n^{\mathrm{T}}\mathbf{P}\mathbf{e}_m} &= \mathrm{tr}\left(\frac{\partial\arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)}{\partial\mathbf{e}_n^{\mathrm{T}}\mathbf{P}\mathbf{e}_m}\mathbf{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\mathbf{\Xi}\right) \\ &\quad + \mathrm{tr}\left(\arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\frac{\partial\mathbf{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\mathbf{\Xi}}{\partial\mathbf{e}_n^{\mathrm{T}}\mathbf{P}\mathbf{e}_m}\right). \end{aligned} \tag{5.11}$$

Algorithm 1 Gradient Projection Algorithm to obtain the QWF-Price precoder with equal power allocation.

Initialization
$\mathbf{P}_{(0)},\ \mathbf{G}_{(0)} = \mathtt{g}_{\mathtt{opt}}\left(\mathbf{P}_{(0)}\right),\ \mu = 10,\ \epsilon = 10^{-4}$ and $n = 0$
repeat
  $\mathbf{P}_{(n+1)} = \mathbf{P}_{(n)} - \mu \left(\frac{\partial \mathrm{MSE}_{(n)}}{\partial \mathbf{P}}\right)^{*}$
  $\mathbf{P}_{(n+1)} = \sqrt{\frac{P_{\mathrm{tx}}}{\alpha_q N}}\, \mathrm{diag}\left(\mathbf{P}_{(n+1)}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{(n+1)}^{\mathrm{H}}\right)^{-1/2} \mathbf{P}_{(n+1)}$ {Equal power allocation constraint}
  $\mathbf{G}_{(n+1)} = \mathtt{g}_{\mathtt{opt}}\left(\mathbf{P}_{(n+1)}\right)$
  if $\mathrm{MSE}_{(n+1)} > \mathrm{MSE}_{(n)}$ **then**
    $\mu = \mu/2$
  else
    $n = n + 1$
  end if
until $\frac{\left|\mathrm{MSE}_{(n+1)} - \mathrm{MSE}_{(n)}\right|}{\left|\mathrm{MSE}_{(n)}\right|} \leq \epsilon$

Algorithm 2 Gradient Projection Algorithm to obtain the QWF-Price precoder with unequal power allocation.

Initialization
$\mathbf{P}_{(0)},\ \mathbf{G}_{(0)} = \mathtt{g}_{\mathtt{opt}}\left(\mathbf{P}_{(0)}\right),\ \mu = 10,\ \epsilon = 10^{-4}$ and $n = 0$
repeat
  $\mathbf{P}_{(n+1)} = \mathbf{P}_{(n)} - \mu \left(\frac{\partial \mathrm{MSE}_{(n)}}{\partial \mathbf{P}}\right)^{*}$
  $\mathbf{P}_{(n+1)} = \sqrt{\frac{P_{\mathrm{tx}}}{\alpha_q}}\, \mathrm{tr}\left(\mathbf{P}_{(n+1)}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{(n+1)}^{\mathrm{H}}\right)^{-1/2} \mathbf{P}_{(n+1)}$ {Unequal power allocation constraint}
  $\mathbf{G}_{(n+1)} = \mathtt{g}_{\mathtt{opt}}\left(\mathbf{P}_{(n+1)}\right)$
  if $\mathrm{MSE}_{(n+1)} > \mathrm{MSE}_{(n)}$ **then**
    $\mu = \mu/2$
  else
    $n = n + 1$
  end if
until $\frac{\left|\mathrm{MSE}_{(n+1)} - \mathrm{MSE}_{(n)}\right|}{\left|\mathrm{MSE}_{(n)}\right|} \leq \epsilon$

According to A5, the derivative term related to $\arcsin(\bullet)$ calculates to

$$\begin{aligned}\frac{\partial \arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)}{\partial \mathbf{e}_n^{\mathrm{T}}\mathbf{P}\mathbf{e}_m} &= \operatorname{nondiag}\left(\mathbf{1}_{N,N} - \operatorname{nondiag}\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)^{\circ 2}\right)^{\circ -1/2} \circ \\ &\qquad \frac{\partial \Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}}{\partial \mathbf{e}_n^{\mathrm{T}}\mathbf{P}\mathbf{e}_m} \\ &= \frac{1}{2}\operatorname{nondiag}\left(\mathbf{1}_{N,N} - \operatorname{nondiag}\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)^{\circ 2}\right)^{\circ -1/2} \circ \\ &\Big(\operatorname{diag}(\mathbf{C_{xx}})^{-1/2}\,\mathbf{e}_n\mathbf{e}_m^{\mathrm{T}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\,\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\operatorname{diag}(\mathbf{C_{xx}})^{-1/2} \\ &+ \operatorname{diag}(\mathbf{C_{xx}})^{-1/2}\,\mathbf{P}^{*}\mathbf{C_{ss}}^{*}\,\mathrm{e}^{\mathrm{j}\,2\Delta k\psi}\,\mathbf{e}_m\mathbf{e}_n^{\mathrm{T}}\operatorname{diag}(\mathbf{C_{xx}})^{-1/2} \\ &- \left(\mathbf{e}_n^{\mathrm{T}}\mathbf{C_{xx}}\mathbf{e}_n\right)^{-3/2}\mathbf{e}_m^{\mathrm{T}}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\mathbf{e}_n\mathbf{e}_n\mathbf{e}_n^{\mathrm{T}}\Re\{\mathbf{C_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\operatorname{diag}(\mathbf{C_{xx}})^{-1/2} \\ &- \operatorname{diag}(\mathbf{C_{xx}})^{-1/2}\,\Re\{\mathbf{C_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\left(\mathbf{e}_n^{\mathrm{T}}\mathbf{C_{xx}}\mathbf{e}_n\right)^{-3/2}\mathbf{e}_m^{\mathrm{T}}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\mathbf{e}_n\mathbf{e}_n\mathbf{e}_n^{\mathrm{T}}\Big).\end{aligned} \tag{5.12}$$

Moreover, we have

$$\begin{aligned}\frac{\partial \mathbf{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\mathbf{\Xi}}{\partial \mathbf{e}_n^{\mathrm{T}}\mathbf{P}\mathbf{e}_m} &= \alpha_q\frac{1}{2}\left(\mathbf{e}_n^{\mathrm{T}}\mathbf{C_{xx}}\mathbf{e}_n\right)^{-1/2}\mathbf{e}_m^{\mathrm{T}}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\mathbf{e}_n\mathbf{e}_n\mathbf{e}_n^{\mathrm{T}}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\operatorname{diag}(\mathbf{C_{xx}})^{1/2} \\ &+ \alpha_q\frac{1}{2}\operatorname{diag}(\mathbf{C_{xx}})^{1/2}\,\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\left(\mathbf{e}_n^{\mathrm{T}}\mathbf{C_{xx}}\mathbf{e}_n\right)^{-1/2}\mathbf{e}_m^{\mathrm{T}}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\mathbf{e}_n\mathbf{e}_n\mathbf{e}_n^{\mathrm{T}}.\end{aligned} \tag{5.13}$$

Plugging (5.12) and (5.13) in (5.11) and applying the property

$$\operatorname{tr}\left(\left(\mathbf{C}\circ\mathbf{v}_1\mathbf{v}_2^{\mathrm{T}}\right)\mathbf{D}\right) = \operatorname{tr}\left(\mathbf{v}_2^{\mathrm{T}}\left(\mathbf{D}\circ\mathbf{C}^{\mathrm{T}}\right)\mathbf{v}_1\right), \tag{5.14}$$

we obtain

$$\begin{aligned}&\frac{\partial \operatorname{tr}\left(\arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\mathbf{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\mathbf{\Xi}\right)}{\partial \mathbf{e}_n^{\mathrm{T}}\mathbf{P}\mathbf{e}_m} = \\ &\frac{1}{2}\Big(\mathbf{e}_m^{\mathrm{T}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\,\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\operatorname{diag}(\mathbf{C_{xx}})^{-1/2}\,\mathbf{\Omega}\operatorname{diag}(\mathbf{C_{xx}})^{-1/2}\,\mathbf{e}_n \\ &+ \mathbf{e}_n^{\mathrm{T}}\,\mathrm{e}^{\mathrm{j}\,2\Delta k\psi}\operatorname{diag}(\mathbf{C_{xx}})^{-1/2}\,\mathbf{\Omega}\operatorname{diag}(\mathbf{C_{xx}})^{-1/2}\,\mathbf{P}^{*}\mathbf{C_{ss}}^{*}\mathbf{e}_m \\ &- \mathbf{e}_n^{\mathrm{T}}\Re\{\mathbf{C_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\operatorname{diag}(\mathbf{C_{xx}})^{-1/2}\,\mathbf{\Omega}\left(\mathbf{e}_n^{\mathrm{T}}\mathbf{C_{xx}}\mathbf{e}_n\right)^{-3/2}\mathbf{e}_m^{\mathrm{T}}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\mathbf{e}_n\mathbf{e}_n \\ &- \mathbf{e}_n^{\mathrm{T}}\mathbf{\Omega}\operatorname{diag}(\mathbf{C_{xx}})^{-1/2}\,\Re\{\mathbf{C_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\left(\mathbf{e}_n^{\mathrm{T}}\mathbf{C_{xx}}\mathbf{e}_n\right)^{-3/2}\mathbf{e}_m^{\mathrm{T}}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\mathbf{e}_n\mathbf{e}_n\Big) \\ &+ \alpha_q\frac{1}{2}\Big(\mathbf{e}_n^{\mathrm{T}}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\operatorname{diag}(\mathbf{C_{xx}})^{1/2}\arcsin(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\})\mathbf{e}_n\left(\mathbf{e}_n^{\mathrm{T}}\mathbf{C_{xx}}\mathbf{e}_n\right)^{-1/2}\mathbf{e}_n^{\mathrm{T}}\mathbf{P}^{*}\mathbf{C_{ss}}^{*}\mathbf{e}_m \\ &+ \mathbf{e}_m^{\mathrm{T}}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\mathbf{e}_n\left(\mathbf{e}_n^{\mathrm{T}}\mathbf{C_{xx}}\mathbf{e}_n\right)^{-1/2}\mathbf{e}_n^{\mathrm{T}}\arcsin(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\})\operatorname{diag}(\mathbf{C_{xx}})^{1/2}\,\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\mathbf{e}_n\Big),\end{aligned} \tag{5.15}$$

where

$$\mathbf{\Omega} = \left(\mathbf{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\mathbf{\Xi}\right)\circ\operatorname{nondiag}\left(\mathbf{1}_{N,N} - \operatorname{nondiag}\left(\Re\{\mathbf{R}_{\mathbf{xx}}^{*}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)^{\circ 2}\right)^{\circ -1/2}. \tag{5.16}$$

Thus, the derivative of the MSE w.r.t. $\mathbf{P}$ is given by

$$\begin{aligned}\frac{\partial \mathrm{E}\left[\|\mathbf{u}-\mathbf{s}\|_2^2\right]}{\partial \mathbf{P}} = \frac{1}{2}\frac{Q}{\pi}\sin^2(\psi)\sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \Bigg(& \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{-1/2}\Big(\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\,\mathbf{\Omega}^{\mathrm{T}} + \mathrm{e}^{\mathrm{j}\,2\Delta k\psi}\,\mathbf{\Omega} \\ & - \operatorname{diag}\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\mathbf{\Omega}\right) - \operatorname{diag}\left(\mathbf{\Omega}\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right) \\ & + \operatorname{diag}\left(\mathbf{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\mathbf{\Xi}\arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\right) \\ & + \operatorname{diag}\left(\arcsin\left(\Re\{\mathbf{R_{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\mathbf{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\mathbf{\Xi}\right)\Bigg) \\ & \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{-1/2}\mathbf{P}^*\mathbf{C}_{\mathbf{ss}}^*\Big) \\ & - \alpha_q\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{C}_{\mathbf{ss}}^*. \end{aligned} \tag{5.17}$$

Second, the derivative of the MSE expression in (5.5) w.r.t. $\mathbf{G}$ calculates to

$$\frac{\partial \mathrm{E}\left[\|\hat{\mathbf{u}}-\mathbf{s}\|_2^2\right]}{\partial \mathbf{G}} = 2\operatorname{diag}\left(\mathbf{H}\mathbf{C_{tt}}\mathbf{H}^{\mathrm{H}} + \mathbf{C}_{\boldsymbol{\eta\eta}}\right) - \operatorname{diag}\left(\mathbf{H}\mathbf{C_{ts}}\right) - \operatorname{diag}\left(\mathbf{H}^*\mathbf{C}_{\mathbf{ts}}^*\right). \tag{5.18}$$

Thus, the optimal filter $\mathbf{G}_{\mathrm{opt}}$ is obtained by setting (5.18) equal to a zero matrix and is given by

$$\mathbf{G}_{\mathrm{opt}} = \mathrm{g}_{\mathrm{opt}}(\mathbf{P}) = \left|\operatorname{diag}\left(\mathbf{H}\mathbf{C_{tt}}\mathbf{H}^{\mathrm{H}} + \mathbf{C}_{\boldsymbol{\eta\eta}}\right)^{-1}\operatorname{diag}\left(\Re\{\mathbf{H}\mathbf{C_{ts}}\}\right)\right|, \tag{5.19}$$

where the operator $|\bullet|$ is applied element-wise to the matrix entries.

5.3.2 Precoder Design Based on LCA

In this section, we aime at getting a closed-form approximation for the MMSE precoder by using the LCA to obtain a linear expression of the MSE. We recall the linear precoder design in [25] and apply it to the case of QCE transmit signals. The precoder aims at minimizing the MSE under a transmit power constraint, where the receive processing is assumed to be a scaled identity matrix, i.e. $\mathbf{G} = g\mathbf{I}_M$. Using the LCA, the MSE expression simplifies to

$$\begin{aligned}\mathrm{E}\left[\|\mathbf{u}-\mathbf{s}\|_2^2\right] = \operatorname{tr}\Big(& |g|^2\,\mathbf{H}\left(\alpha_q^2\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}} + \alpha_q\rho_q\operatorname{diag}(\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}})\right)\mathbf{H}^{\mathrm{H}} \\ & - g\mathbf{H}\alpha_q\mathbf{P}\mathbf{C_{ss}} - g^*\alpha_q\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}} + |g|^2\,\mathbf{C}_{\boldsymbol{\eta\eta}} + \mathbf{C_{ss}}\Big). \end{aligned} \tag{5.20}$$

The Lagrangian function is expressed by

$$\begin{aligned}\mathcal{L}(\mathbf{P}, g, \lambda) = \operatorname{tr}\Big(& |g|^2\,\mathbf{H}\left(\alpha_q^2\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}} + \alpha_q\beta_q\operatorname{diag}(\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}})\right)\mathbf{H}^{\mathrm{H}} \\ & - g\alpha_q\mathbf{H}\mathbf{P}\mathbf{C_{ss}} - g^*\alpha_q\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}} + |g|^2\,\mathbf{C}_{\boldsymbol{\eta\eta}} + \mathbf{C_{ss}}\Big) + \lambda\left(\alpha_q\operatorname{tr}\left(\mathbf{P}\mathbf{C_{ss}}\mathbf{P}^{\mathrm{H}}\right) - P_{\mathrm{tx}}\right). \end{aligned} \tag{5.21}$$

The Karush-Kuhn-Tucker (KKT) equations are then given by

$$\frac{\partial \mathcal{L}(\mathbf{P}, g, \lambda)}{\partial \mathbf{P}} = |g|^2 \left(\alpha_q^2 \mathbf{H}^{\mathrm{T}}\mathbf{H}^*\mathbf{P}^*\mathbf{C}_{\mathbf{ss}}^{\mathrm{T}} + \alpha_q\beta_q \operatorname{diag}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}^*\right)\mathbf{P}^*\mathbf{C}_{\mathbf{ss}}^{\mathrm{T}}\right) - g\alpha_q\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{T}} + \lambda\alpha_q\mathbf{P}^*\mathbf{C}_{\mathbf{ss}}^{\mathrm{T}}$$
$$= \mathbf{0}_{N,M}, \tag{5.22}$$

$$\frac{\partial \mathcal{L}(\mathbf{P}, g, \lambda)}{\partial g} = g^* \operatorname{tr}\left(\mathbf{H}\left(\alpha_q^2\mathbf{P}\mathbf{C}_{\mathbf{ss}}\mathbf{P}^{\mathrm{H}} + \alpha_q\beta_q \operatorname{diag}(\mathbf{P}\mathbf{C}_{\mathbf{ss}}\mathbf{P}^{\mathrm{H}})\mathbf{H}^{\mathrm{H}} + \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\right) - \alpha_q \operatorname{tr}(\mathbf{H}\mathbf{P}\mathbf{C}_{\mathbf{ss}}) = 0, \tag{5.23}$$

and

$$\frac{\partial \mathcal{L}(\mathbf{P}, g, \lambda)}{\partial \lambda} = \alpha_q \operatorname{tr}\left(\mathbf{P}\mathbf{C}_{\mathbf{ss}}\mathbf{P}^{\mathrm{H}}\right) - P_{\mathrm{tx}} = 0. \tag{5.24}$$

Multiplying (5.22) by $\mathbf{P}^{\mathrm{T}}$ from the left side and taking the trace leads to

$$|g|^2 \operatorname{tr}\left(\alpha_q^2\mathbf{P}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}^*\mathbf{P}^*\mathbf{C}_{\mathbf{ss}}^{\mathrm{T}} + \alpha_q\beta_q\mathbf{P}^{\mathrm{T}} \operatorname{diag}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}^*\right)\mathbf{P}^*\mathbf{C}_{\mathbf{ss}}^{\mathrm{T}}\right) - g\alpha_q \operatorname{tr}\left(\mathbf{P}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{T}}\right)$$
$$+ \lambda\alpha_q \operatorname{tr}\left(\mathbf{P}^{\mathrm{T}}\mathbf{P}^*\mathbf{C}_{\mathbf{ss}}^{\mathrm{T}}\right) = 0. \tag{5.25}$$

From (5.23), we get

$$\alpha_q \operatorname{tr}(\mathbf{H}\mathbf{P}\mathbf{C}_{\mathbf{ss}}) = g^* \operatorname{tr}\left(\mathbf{H}\left(\alpha_q^2\mathbf{P}\mathbf{C}_{\mathbf{ss}}\mathbf{P}^{\mathrm{H}} + \alpha_q\beta_q \operatorname{diag}(\mathbf{P}\mathbf{C}_{\mathbf{ss}}\mathbf{P}^{\mathrm{H}})\mathbf{H}^{\mathrm{H}} + \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\right), \tag{5.26}$$

which when inserted in (5.25) gives the expression of λ

$$\lambda = \frac{|g|^2 \operatorname{tr}(\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}})}{\alpha_q \operatorname{tr}(\mathbf{P}\mathbf{C}_{\mathbf{ss}}\mathbf{P}^{\mathrm{H}})} = \frac{|g|^2 \operatorname{tr}(\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}})}{P_{\mathrm{tx}}}, \tag{5.27}$$

where we used (5.24) and the properties $\operatorname{tr}(\mathbf{A}\operatorname{diag}(\mathbf{B})) = \operatorname{tr}(\operatorname{diag}(\mathbf{A})\mathbf{B})$ and $\operatorname{tr}(\mathbf{A}) = \operatorname{tr}(\mathbf{A})^{\mathrm{T}}$. Inserting (5.27) in (5.22) and solving it for $\mathbf{P}$, we obtain

$$\mathbf{P} = \frac{1}{g}\left(\alpha_q\mathbf{H}^{\mathrm{H}}\mathbf{H} + \beta_q \operatorname{diag}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right) + \frac{\operatorname{tr}(\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}})}{P_{\mathrm{tx}}}\mathbf{I}_N\right)^{-1}\mathbf{H}^{\mathrm{H}}. \tag{5.28}$$

The optimal g is found by satisfying (5.24) with $\mathbf{P}$ from (5.28); that is

$$g = \sqrt{\frac{\alpha_q}{P_{\mathrm{tx}}}}\sqrt{\operatorname{tr}\left(\left(\alpha_q\mathbf{H}^{\mathrm{H}}\mathbf{H} + \beta_q \operatorname{diag}(\mathbf{H}^{\mathrm{H}}\mathbf{H}) + \frac{\operatorname{tr}(\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}})}{P_{\mathrm{tx}}}\mathbf{I}_N\right)^{-2}\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}\mathbf{H}\right)}. \tag{5.29}$$

This precoder design was already derived in [25] for low resolution cartesian DACs at the transmitter. Whether we use polar or cartesian DACs, the expression of the WF precoder that takes into account the quantization distortions based on LCA turns to be the same. The coefficients α_q and β_q have to be defined accordingly.

5.4 Dual Optimization Problem

5.4.1 Does a Dual Problem Exist?

In the above sections, the MMSE precoder was designed based on Price's theorem for equal and unequal power allocation and on LCA for only unequal power allocation. With Price's

theorem, no closed-form expression of the precoder could be found but the optimal precoder is found by an iterative approach. However, with LCA an approximate expression of the optimal MMSE precoder is given in a closed-form. One might ask whether it is possible to get a closed-form expression for the exact optimal MMSE in the dual domain; that is when applying Price's theorem. To this end, we first have to find the dual uplink system model. Second, find the optimal dual filter, hopefully in closed-form expression. Finally, convert the expressions in the primal domain, i.e. downlink scenario, and obtain the expression of the MMSE precoder.

According to Fig. 4.2, the corresponding MSE expression for flat-fading channels, after dropping out the time index, is given by

$$\begin{aligned}
\mathrm{MSE}^{\mathrm{DL}} &= \mathrm{E}\left[\|\mathbf{u}-\mathbf{s}\|_2^2\right] \\
&= \alpha_q^2 \operatorname{tr}\left(\mathbf{GHPC_{ss}P^HH^HG}\right) + \operatorname{tr}\left(\mathbf{GHC_{d_Q d_Q}H^HG}\right) \\
&\quad - \alpha_q \operatorname{tr}\left(\mathbf{GHPC_{ss}}\right) - \alpha_q \operatorname{tr}\left(\mathbf{C_{ss}P^HH^HG}\right) + \operatorname{tr}\left(\mathbf{GC_{\eta\eta}G}\right) + \operatorname{tr}\left(\mathbf{C_{ss}}\right) \\
&\overset{(3.85)}{=} \alpha_q^2 \operatorname{tr}\left(\mathbf{GHPC_{ss}P^HH^HG}\right) + \alpha_q\beta_q \operatorname{tr}\left(\mathbf{GH}\operatorname{diag}\left(\mathbf{PC_{ss}P^H}\right)\mathbf{H^HG}\right) \\
&\quad + \frac{Q}{\pi}\sin^2(\psi)\sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \operatorname{tr}\left(\mathbf{GH\Xi}\arcsin\left(\Re\{\operatorname{nondiag}\left(\mathbf{R_{xx}}\right)\mathrm{e}^{-\mathrm{j}(2\Delta k\psi)}\}\right)\mathbf{\Xi H^HG}\right) \\
&\quad - \alpha_q^2 \operatorname{tr}\left(\mathbf{GH}\operatorname{nondiag}\left(\mathbf{PC_{ss}P^H}\right)\mathbf{H^HG}\right) \\
&\quad + \operatorname{tr}\left(\mathbf{GC_{\eta\eta}G}\right) - \alpha_q \operatorname{tr}\left(\mathbf{GHPC_{ss}}\right) - \alpha_q \operatorname{tr}\left(\mathbf{C_{ss}P^HH^HG}\right) + \operatorname{tr}\left(\mathbf{C_{ss}}\right) \\
&= \alpha_q \operatorname{tr}\left(\mathbf{GH}\operatorname{diag}\left(\mathbf{PC_{ss}P^H}\right)\mathbf{H^HG}\right) \\
&\quad + \frac{Q}{\pi}\sin^2(\psi)\sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \operatorname{tr}\left(\mathbf{GH\Xi}\arcsin\left(\Re\{\operatorname{nondiag}\left(\mathbf{R_{xx}}\right)\mathrm{e}^{-\mathrm{j}(2\Delta k\psi)}\}\right)\mathbf{\Xi H^HG}\right) \\
&\quad + \operatorname{tr}\left(\mathbf{GC_{\eta\eta}G}\right) - \alpha_q \operatorname{tr}\left(\mathbf{GHPC_{ss}}\right) - \alpha_q \operatorname{tr}\left(\mathbf{C_{ss}P^HH^HG}\right) + \operatorname{tr}\left(\mathbf{C_{ss}}\right). \qquad (5.30)
\end{aligned}$$

The potential dual uplink system model is illustrated in Fig. 4.3, where $\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}[t] = \mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}$ can be freely designed and is not necessarily defined by a CEQ at the receiver. The statistical properties of $\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}$ have to be chosen such that an MSE duality between both systems can be achieved. The MSE expression for the uplink system for flat-fading channels is expressed as

$$\begin{aligned}
\mathrm{MSE}^{\mathrm{UL}} &= \mathrm{E}\left[\left\|\mathbf{u}^{\mathrm{UL}}-\mathbf{s}^{\mathrm{UL}}\right\|_2^2\right] \\
&= \alpha_q^2 \operatorname{tr}\left(\mathbf{FH^HC_{\eta\eta}^{-1/2,H}TC_{ss}^{UL}T^HC_{\eta\eta}^{-1/2}HF^H}\right) + \operatorname{tr}\left(\mathbf{FC_{d_Q d_Q}^{UL}F^H}\right) + \alpha_q^2 \operatorname{tr}\left(\mathbf{FC_{\eta\eta}^{UL}F^H}\right) \\
&\quad - \alpha_q \operatorname{tr}\left(\mathbf{FH^HC_{\eta\eta}^{-1/2,H}TC_{ss}^{UL}}\right) - \alpha_q \operatorname{tr}\left(\mathbf{C_{ss}^{UL}T^HC_{\eta\eta}^{-1/2}HF^H}\right) + \operatorname{tr}\left(\mathbf{C_{ss}^{UL}}\right), \qquad (5.31)
\end{aligned}$$

where

$$\mathbf{u}^{\mathrm{UL}} = \mathbf{u}^{\mathrm{UL}}[t] \qquad (5.32)$$

$$\mathbf{s}^{\mathrm{UL}} = \mathbf{s}^{\mathrm{UL}}[t], \qquad (5.33)$$

and

$$\mathbf{F} = \mathbf{F}_0. \qquad (5.34)$$

If we assume that $\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}$ is the quantization noise that results from quantizing $\mathbf{x}^{\mathrm{UL}} = \mathbf{x}^{\mathrm{UL}}[t]$ and is uncorrelated with it and hence we use the expression in (3.85) for $\mathbf{C}^{\mathrm{UL}}_{\mathbf{d}_{\mathcal{Q}}\mathbf{d}_{\mathcal{Q}}}$, we get

$$\begin{aligned}
\mathrm{MSE}^{\mathrm{UL}} &= \alpha_q^2 \operatorname{tr}\left(\mathbf{F}\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H}\mathbf{F}^{\mathrm{H}}\right) \\
&\quad + \alpha_q\beta_q \operatorname{tr}\left(\mathbf{F}\operatorname{diag}\left(\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H} + \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{\mathrm{UL}}\right)\mathbf{F}^{\mathrm{H}}\right) \\
&\quad + \frac{Q}{\pi}\sin^2(\psi)\sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \operatorname{tr}\left(\mathbf{F}\boldsymbol{\Xi}^{\mathrm{UL}} \arcsin\left(\Re\{\operatorname{nondiag}\left(\mathbf{R}_{\mathbf{xx}}^{\mathrm{UL}}\right)\mathrm{e}^{-\mathrm{j}(2\Delta k\psi)}\}\right)\boldsymbol{\Xi}^{\mathrm{UL}}\mathbf{F}^{\mathrm{H}}\right) \\
&\quad - \alpha_q^2 \operatorname{tr}\left(\mathbf{F}\operatorname{nondiag}\left(\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}}\right)\mathbf{F}^{\mathrm{H}}\right) \\
&\quad + \alpha_q^2 \operatorname{tr}\left(\mathbf{F}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{\mathrm{UL}}\mathbf{F}^{\mathrm{H}}\right) \\
&\quad - \alpha_q \operatorname{tr}\left(\mathbf{F}\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\right) - \alpha_q \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H}\mathbf{F}^{\mathrm{H}}\right) + \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\right) \\
&\overset{(h)}{=} \alpha_q \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{H}\operatorname{diag}\left(\mathbf{F}^{\mathrm{H}}\mathbf{F}\right)\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{T}\right) \\
&\quad + \frac{Q}{\pi}\sin^2(\psi)\sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \operatorname{tr}\left(\mathbf{F}\boldsymbol{\Xi}^{\mathrm{UL}} \arcsin\left(\Re\{\operatorname{nondiag}\left(\mathbf{R}_{\mathbf{xx}}^{\mathrm{UL}}\right)\mathrm{e}^{-\mathrm{j}(2\Delta k\psi)}\}\right)\boldsymbol{\Xi}^{\mathrm{UL}}\mathbf{F}^{\mathrm{H}}\right) \\
&\quad + \alpha_q \operatorname{tr}\left(\mathbf{F}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{\mathrm{UL}}\mathbf{F}^{\mathrm{H}}\right) \\
&\quad - \alpha_q \operatorname{tr}\left(\mathbf{F}\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\right) - \alpha_q \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H}\mathbf{F}^{\mathrm{H}}\right) + \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\right), \qquad (5.35)
\end{aligned}$$

where

$$\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}} = \mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H}, \qquad (5.36)$$

$$\mathbf{R}_{\mathbf{xx}}^{\mathrm{UL}} = \operatorname{diag}\left(\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}}\right)^{-1/2}\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}}\operatorname{diag}\left(\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}}\right)^{-1/2}, \qquad (5.37)$$

$$\boldsymbol{\Xi}^{\mathrm{UL}} = \sqrt{\alpha_q}\operatorname{diag}\left(\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}}\right)^{1/2}, \qquad (5.38)$$

and in (h) we used the equality

$$\alpha_q^2 \operatorname{tr}\left(\mathbf{F}\operatorname{nondiag}\left(\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}}\right)\mathbf{F}^{\mathrm{H}}\right) = \alpha_q^2 \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{H}\operatorname{nondiag}\left(\mathbf{F}^{\mathrm{H}}\mathbf{F}\right)\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{T}\right). \qquad (5.39)$$

We aim to match both MSE expressions in (5.30) and (5.35). Applying the identities in (4.39), (4.40), (4.41) and (4.42) for the flat-fading case, i.e. $L = L_p = 1$, we get

$$\mathbf{F} = \frac{1}{\beta}\mathbf{P}^{\mathrm{H}}, \qquad (5.40)$$

$$\mathbf{T} = \beta\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{1/2,\mathrm{H}}\mathbf{G}^{\mathrm{H}}, \qquad (5.41)$$

$$\mathbf{C}_{\mathbf{ss}} = \mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}} = \sigma_s^2\mathbf{I}_M, \qquad (5.42)$$

and

$$\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{\mathrm{UL}} = \mathbf{I}_N. \qquad (5.43)$$

Thus, (5.35) calculates to

$$\begin{aligned}
\mathrm{MSE}^{\mathrm{UL}} &= \alpha_q \operatorname{tr}\left(\mathbf{G}\mathbf{H}\operatorname{diag}\left(\mathbf{P}\mathbf{C}_{\mathbf{ss}}\mathbf{P}^{\mathrm{H}}\right)\mathbf{H}^{\mathrm{H}}\mathbf{G}\right) \\
&\quad + \frac{1}{\beta^2}\frac{Q}{\pi}\sin^2(\psi)\sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \operatorname{tr}\left(\mathbf{P}^{\mathrm{H}}\boldsymbol{\Xi}^{\mathrm{UL}} \arcsin\left(\Re\{\operatorname{nondiag}\left(\mathbf{R}_{\mathbf{xx}}^{\mathrm{UL}}\right)\mathrm{e}^{-\mathrm{j}(2\Delta k\psi)}\}\right)\boldsymbol{\Xi}^{\mathrm{UL}}\mathbf{P}\right) \\
&\quad + \frac{\alpha_q}{\sigma_s^2\beta^2} \operatorname{tr}\left(\mathbf{P}\mathbf{C}_{\mathbf{ss}}\mathbf{P}^{\mathrm{H}}\right) - \alpha_q \operatorname{tr}\left(\mathbf{P}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{G}\mathbf{C}_{\mathbf{ss}}\right) - \alpha_q \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}\mathbf{G}\mathbf{H}\mathbf{P}\right) + \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}\right). \qquad (5.44)
\end{aligned}$$

As we can see, all terms of (5.30) and (5.44) match except of the second and third terms. To match the third terms, we choose β as

$$\begin{aligned}\beta &= \sqrt{\frac{\alpha_q \operatorname{tr}\left(\mathbf{P}\mathbf{C}_{\mathbf{ss}}\mathbf{P}^{\mathrm{H}}\right)}{\sigma_s^2 \operatorname{tr}\left(\mathbf{G}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\mathbf{G}\right)}} \\ &\overset{(5.9)}{=} \sqrt{\frac{P_{\mathrm{tx}}}{\sigma_s^2 \operatorname{tr}\left(\mathbf{G}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\mathbf{G}\right)}}.\end{aligned} \tag{5.45}$$

However, we cannot match the second terms due to the non-linear expressions. Thus, there exists no MSE duality if $\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}$ is the quantization noise that is uncorrelated with $\mathbf{x}^{\mathrm{UL}}$ and results from quantizing $\mathbf{x}^{\mathrm{UL}}$. The problem arises with the non-diagonal elements of $\mathbf{C}_{\mathbf{d}_{\mathcal{Q}}\mathbf{d}_{\mathcal{Q}}}^{\mathrm{UL}}$ that should be chosen in an appropriate way to match both MSE expressions.

5.4.2 Exact Dual Problem

To find the dual MSE expression, we are changing the quantization function at the receiver to replace it with a different dual non-linear function; that is the distortion term $\mathbf{d}_{\mathcal{Q}}$ has the following covariance matrix

$$\begin{aligned}\mathbf{C}_{\mathbf{d}_{\mathcal{Q}}\mathbf{d}_{\mathcal{Q}}}^{\mathrm{UL}} &= \alpha_q\beta_q \operatorname{diag}\left(\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H} + \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{\mathrm{UL}}\right) \\ &\quad - \alpha_q^2 \operatorname{nondiag}\left(\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H}\right) \\ &\quad + \frac{Q}{\pi}\sin^2(\psi)\alpha_q \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\, \mathbf{F}^{\mathrm{H}}\left(\mathbf{F}\mathbf{F}^{\mathrm{H}}\right)^{-1}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H} \operatorname{diag}\left(\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{F}\right)^{1/2} \\ &\quad \arcsin\left(\Re\{\operatorname{nondiag}\left(\operatorname{diag}\left(\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{F}\right)^{-1/2}\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{F} \operatorname{diag}\left(\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{F}\right)^{-1/2}\right)\mathrm{e}^{-\mathrm{j}(2\Delta k\psi)}\}\right) \\ &\quad \operatorname{diag}\left(\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{F}\right)^{1/2}\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\left(\mathbf{F}\mathbf{F}^{\mathrm{H}}\right)^{-1}\mathbf{F}.\end{aligned} \tag{5.46}$$

The resulting MSE expression reads as

$$\begin{aligned}\mathrm{MSE}^{\mathrm{UL}} &= \alpha_q \operatorname{tr}\left(\mathbf{F} \operatorname{diag}\left(\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H}\right)\mathbf{F}^{\mathrm{H}}\right) + \alpha_q \operatorname{tr}\left(\mathbf{F}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{\mathrm{UL}}\mathbf{F}^{\mathrm{H}}\right) \\ &\quad + \frac{Q}{\pi}\sin^2(\psi)\alpha_q \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \operatorname{tr}\Big(\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H} \operatorname{diag}\left(\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{F}\right)^{1/2} \\ &\quad \arcsin\left(\Re\{\operatorname{nondiag}\left(\operatorname{diag}\left(\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{F}\right)^{-1/2}\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{F} \operatorname{diag}\left(\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{F}\right)^{-1/2}\right)\mathrm{e}^{-\mathrm{j}(2\Delta k\psi)}\}\right) \\ &\quad \operatorname{diag}\left(\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{F}\right)^{1/2}\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\Big) \\ &\quad - \alpha_q \operatorname{tr}\left(\mathbf{F}\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\right) - \alpha_q \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H}\mathbf{F}^{\mathrm{H}}\right) + \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\right).\end{aligned} \tag{5.47}$$

By applying the equalities in (5.40), (5.41), (5.42), (5.43) and (5.45), we obtain

$$\mathrm{MSE}^{\mathrm{UL}} = \mathrm{MSE}^{\mathrm{DL}}. \tag{5.48}$$

The dual uplink system is not simply having the quantization at the receiver in the uplink scenario.

In the dual uplink system, we end up with two filters $\mathbf{F}$ and $\mathbf{T}$ to optimize. To simplify the optimization task, we assume that $\mathbf{G}$ is a scaled identity, i.e. $\mathbf{G} = g\mathbf{I}_M$ and thus from (5.41) and (5.45) we get

$$\mathbf{T} = \sqrt{\frac{P_{\text{tx}}}{\sigma_s^2 \operatorname{tr}(\mathbf{C}_{\boldsymbol{\eta\eta}})}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{1/2,\text{H}}. \tag{5.49}$$

Consequently, the transmit power is the same as in the primal domain since

$$\operatorname{tr}\left(\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\text{UL}}\mathbf{T}^{\text{H}}\right) = P_{\text{tx}}. \tag{5.50}$$

To find the dual filter $\mathbf{F}$, we should compute the derivative of MSE^{UL} w.r.t. $\mathbf{F}$. In analogy to the derivations in Section 5.3.1, we get

$$\begin{aligned}\frac{\partial \text{MSE}^{\text{UL}}}{\partial \mathbf{F}} = \frac{Q}{\pi}\sin^2(\psi)\alpha_q \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{-\mathrm{j}(2\Delta k\psi)}\, \frac{1}{2}\mathbf{C}_{\mathbf{ss}}^{\text{UL}*}\mathbf{F}^*\Big(\operatorname{diag}\left(\mathbf{F}^{\text{H}}\mathbf{C}_{\mathbf{ss}}^{\text{UL}}\mathbf{F}\right)^{-1/2}\Big(\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\,\boldsymbol{\Omega} + \mathrm{e}^{\mathrm{j}\,2\Delta k\psi}\,\boldsymbol{\Omega}^{\text{T}} \\ - \operatorname{diag}\left(\Re\{\boldsymbol{\Sigma}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\boldsymbol{\Omega}^*\right) - \operatorname{diag}\left(\boldsymbol{\Omega}^*\Re\{\boldsymbol{\Sigma}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right) \\ + \frac{P_{\text{tx}}}{\sigma_s^2 \operatorname{tr}(\mathbf{C}_{\boldsymbol{\eta\eta}})} \operatorname{diag}\left(\boldsymbol{\Lambda}\mathbf{H}^{\text{T}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2,\text{T}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2,*}\mathbf{H}^*\boldsymbol{\Lambda}\arcsin\left(\Re\{\boldsymbol{\Sigma}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\right) \\ + \frac{P_{\text{tx}}}{\sigma_s^2 \operatorname{tr}(\mathbf{C}_{\boldsymbol{\eta\eta}})} \operatorname{diag}\left(\arcsin\left(\Re\{\boldsymbol{\Sigma}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\boldsymbol{\Lambda}\mathbf{H}^{\text{T}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2,\text{T}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2,*}\mathbf{H}^*\boldsymbol{\Lambda}\right)\Big) \\ \operatorname{diag}\left(\mathbf{F}^{\text{H}}\mathbf{C}_{\mathbf{ss}}^{\text{UL}}\mathbf{F}\right)^{-1/2}\Big) \\ + \sqrt{\frac{P_{\text{tx}}}{\sigma_s^2 \operatorname{tr}(\mathbf{C}_{\boldsymbol{\eta\eta}})}}\alpha_q\mathbf{F}^*\mathbf{C}_{\boldsymbol{\eta\eta}}^{\text{UL,T}} - \alpha_q\mathbf{C}_{\mathbf{ss}}^{\text{UL,T}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2,*}\mathbf{H}^*,\end{aligned} \tag{5.51}$$

where

$$\boldsymbol{\Sigma} = \operatorname{diag}\left(\mathbf{F}^{\text{H}}\mathbf{C}_{\mathbf{ss}}^{\text{UL}}\mathbf{F}\right)^{-1/2}\mathbf{F}^{\text{H}}\mathbf{C}_{\mathbf{ss}}^{\text{UL}}\mathbf{F}\operatorname{diag}\left(\mathbf{F}^{\text{H}}\mathbf{C}_{\mathbf{ss}}^{\text{UL}}\mathbf{F}\right)^{-1/2}, \tag{5.52}$$

$$\boldsymbol{\Omega} = \left(\boldsymbol{\Lambda}\mathbf{H}^{\text{H}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2,\text{H}}\mathbf{T}\mathbf{T}^{\text{H}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2}\mathbf{H}\boldsymbol{\Lambda}\right) \circ \operatorname{nondiag}\left(\mathbf{1}_{N,N} - \operatorname{nondiag}\left(\Re\{\boldsymbol{\Sigma}^*\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)^{\circ 2}\right)^{\circ -1/2}, \tag{5.53}$$

and

$$\boldsymbol{\Lambda} = \operatorname{diag}\left(\mathbf{F}^{\text{H}}\mathbf{C}_{\mathbf{ss}}^{\text{UL}}\mathbf{F}\right)^{1/2}. \tag{5.54}$$

Unfortunately, we cannot find a closed-form expression for $\mathbf{F}$ and we have to find the optimal filter $\mathbf{F}$ again with iterative algorithms, i.e. gradient descent algorithm as in the primal problem. However, the advantage here is that no power constraint has to be considered. The gradient descent algorithm is given in Algorithm 3. After obtaining the optimal filter $\mathbf{F}$, the duality scaling factor β in (4.44) can be expressed by

$$\beta = \sqrt{\frac{P_{\text{tx}}}{\alpha_q \operatorname{tr}(\mathbf{F}^{\text{H}}\mathbf{C}_{\mathbf{ss}}^{\text{UL}}\mathbf{F})}}. \tag{5.55}$$

Hence, the dual filters $\mathbf{P}$ and $\mathbf{G}$ can be extracted by applying the equalities in (5.40) and (5.41).

Algorithm 3 Gradient Descent Algorithm to obtain the dual WFQ-Price with unequal power allocation.

Initialization
$\mathbf{F}_{(0)},\ \mu = 10,\ \epsilon = 10^{-4}$ and $n = 0$
repeat
  $\mathbf{F}_{(n+1)} = \mathbf{F}_{(n)} - \mu \left(\frac{\partial \mathrm{MSE}^{\mathrm{UL}}_{(n)}}{\partial \mathbf{F}} \right)^*$
  if $\mathrm{MSE}^{\mathrm{UL}}_{(n+1)} > \mathrm{MSE}^{\mathrm{UL}}_{(n)}$ **then**
    $\mu = \mu/2$
  else
    $n = n + 1$
  end if
until $\frac{\left|\mathrm{MSE}^{\mathrm{UL}}_{(n+1)} - \mathrm{MSE}^{\mathrm{UL}}_{(n)}\right|}{\left|\mathrm{MSE}^{\mathrm{UL}}_{(n)}\right|} \leq \epsilon$

5.4.3 Approximate Dual Problem

We still aim to get a closed-form expression for the filter $\mathbf{F}$ and hence for the precoder $\mathbf{P}$. To this end, we assume that the MSE expressions in (5.35) and in (5.30) are equal. This assumption holds true if the Bussgang decomposition is applied with the LCA; that is $\mathbf{C}_{\mathbf{d}_\mathcal{Q}\mathbf{d}_\mathcal{Q}} = \alpha_q\beta_q \operatorname{diag}\left(\mathbf{H}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}^{\mathrm{UL}}_{\mathbf{ss}}\mathbf{T}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}\mathbf{H} + \mathbf{C}^{\mathrm{UL}}_{\boldsymbol{\eta}\boldsymbol{\eta}}\right)$. The $\mathrm{MSE}^{\mathrm{UL}}$ expression is in general expressed as

$$\mathrm{MSE}^{\mathrm{UL}} = \operatorname{tr}\left(\mathbf{F}\mathbf{C}^{\mathrm{UL}}_{\mathbf{tt}}\mathbf{F}^{\mathrm{H}}\right) - \operatorname{tr}\left(\mathbf{F}\mathbf{C}^{\mathrm{UL}}_{\mathbf{ts}}\right) - \operatorname{tr}\left(\mathbf{C}^{\mathrm{UL}}_{\mathbf{st}}\mathbf{F}^{\mathrm{H}}\right) + \operatorname{tr}\left(\mathbf{C}^{\mathrm{UL}}_{\mathbf{ss}}\right). \tag{5.56}$$

The optimal filter $\mathbf{F}$ that minimizes (5.56) reads as

$$\mathbf{F} = \mathbf{C}^{\mathrm{UL}}_{\mathbf{st}}\left(\mathbf{C}^{\mathrm{UL}}_{\mathbf{tt}}\right)^{-1}. \tag{5.57}$$

To compute $\mathbf{C}^{\mathrm{UL}}_{\mathbf{tt}}$ we apply Price's theorem in (3.66). The optimal filter $\mathbf{T}$ is given in (5.49). Hence, the optimal filters in the primal domain are obtained by applying (5.40) and (5.41), where β is given in (4.44). This method was introduced in [69] for the one-bit quantization case, i.e. $Q = 4$.

5.5 Simulation Results

In this section, we compare the performance of the different linear precoding techniques that were introduced above with the ideal WF precoder. We denote the precoders by

- WF: the ideal WF, [70], where no CEQ is applied in the system.
- QWF: the Quantized WF that was introduced in [25] for cartesian DACs and was extended to the case of polar DACs in Section 5.3.2.
- QWF-$\mathrm{Apprx}_{\mathrm{Dual}}$: the approximate dual Quantized WF that applies unequal power allocation at the antennas. The resulting receive filter $\mathbf{G}$ is a scaled identity. This precoder is derived in Section 5.4.3.

- QWF-Price$_{\text{eqPA}}$: the Quantized WF based on Price's theorem that applies equal power allocation at the antennas. This precoder is detailed in Algorithm 1. The start value is the WF precoder that is projected to fulfill the equal transmit power constraint.

- QWF-Price$_{\text{neqPA}}$: the Quantized WF based on Price's theorem that applies unequal power allocation at the antennas. This precoder is detailed in Algorithm 2. The start value is the WF precoder.

- QWF-Price$_{\text{Dual}}$: the dual Quantized WF based on Price's theorem that applies unequal power allocation at the antennas. The resulting receive filter $\mathbf{G}$ is a scaled identity. This precoder is detailed in Algorithm 3. The start value is the QWF-Apprx$_{\text{Dual}}$ precoder.

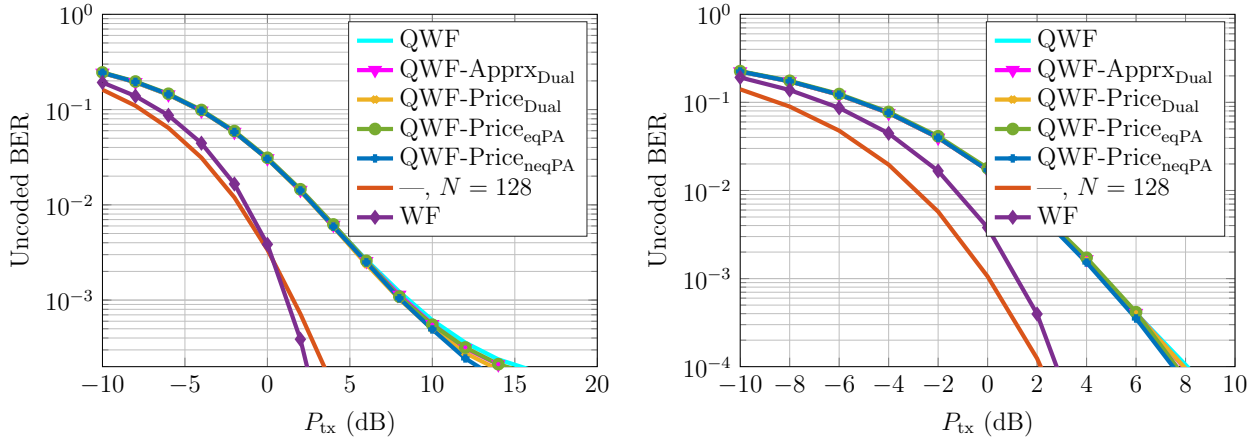
To this end, we assume a BS with $N = 64$ or $N = 128$ antennas serving $M = 8$ single-antenna users with 4-PSK or 16-QAM signals, respectively. The numerical results are obtained with Monte Carlo simulations of 100 independent flat-fading channel realizations from the i.i.d. channel model and the mmW sparse channel model described in Section 4.6. For the mmW sparse channel, we assume that $N_{\text{cl}} = 2$ and $N_{\text{ray}} = 10$. The AWGN is also i.i.d. with variance one at each antenna. The performance metric is the uncoded BER averaged over the single-antenna users. For the blind estimation of the coefficients g_m we use a block length of $T = 128$. The numerical results are plotted in Fig. 5.1, Fig. 5.2, Fig. 5.3 and Fig. 5.4.

First, it can be deduced that all proposed linear precoders perform almost the same for i.i.d. channels. There is a moderate gain compared to the QWF precoder. The loss compared to the ideal WF reduces with increased quantization resolution.

Second, the proposed precoders perform differently for mmW sparse channels. We can see that the QWF-Price$_{\text{neqPA}}$ precoder performs the best followed by QWF-Price$_{\text{eqPA}}$, QWF-Price$_{\text{Dual}}$, QWF-Apprx$_{\text{Dual}}$ and last QWF. This behavior is expected since the QWF-Price$_{\text{neqPA}}$ precoder is computed in the primal domain without any approximation of Price's theorem. Additionally, having unequal power allocation at the antennas offers more degrees of freedom for the precoder design, which explains the performance loss of QWF-Price$_{\text{eqPA}}$ compared to QWF-Price$_{\text{neqPA}}$. The performance loss of QWF-Price$_{\text{Dual}}$ can be explained by the assumption of $\mathbf{G}$ being a scaled identity, which again reduces the degrees of freedom in the precoder design. Although the QWF-Apprx$_{\text{Dual}}$ precoder is more efficient in terms of convergence time and computational complexity, it performs worse than the linear precoders based on iterative algorithms. This is due to the approximate duality that we applied for this precoder design and the assumption on $\mathbf{G}$ being a scaled identity. Last but not least, the QWF precoder performs the worst due to the LCA and the assumption on $\mathbf{G}$ being a scaled identity.

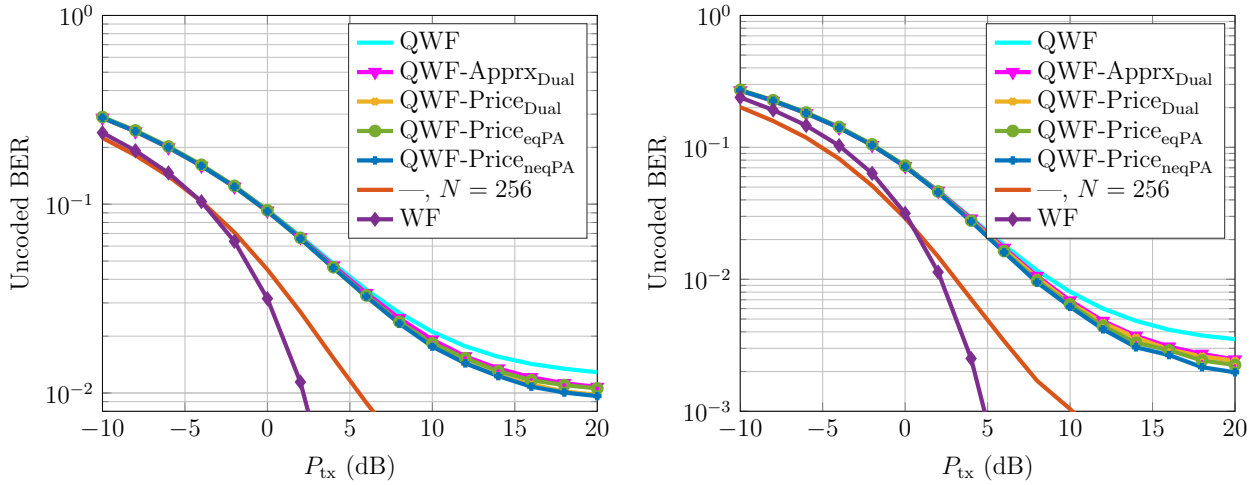
Third, doubling the number of antennas with the use of the QWF-Price$_{\text{neqPA}}$ precoder leads to almost the same BER performance as in the ideal unquantized case for $Q = 4$ and with 4-PSK signaling. With increased resolution, e.g. $Q = 8$, less than the double number of antennas in required in the quantized case to obtain the same performance as in the ideal case. For 16-QAM signaling the latter statement holds for low values of the transmit power.

In summary, the performance improvement with the iterative and more accurate linear precoding algorithms that take into account the quantization distortions is very moderate compared to the QWF that is based on approximate statistics. However, this comparison is restricted to flat-fading channels. So how does the performance comparison look like in the case of frequency-selective channels?



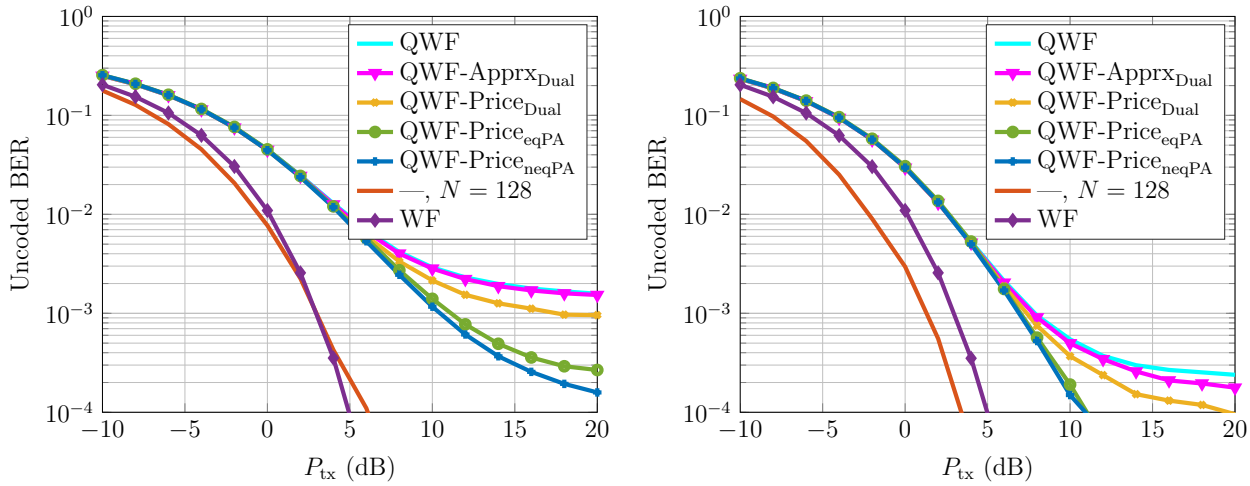
(a) $Q = 4$ (b) $Q = 8$

Fig. 5.1: Comparison between different linear precoders for 4-PSK signaling: $N = 64$, $M = 8$ with the i.i.d. channel.



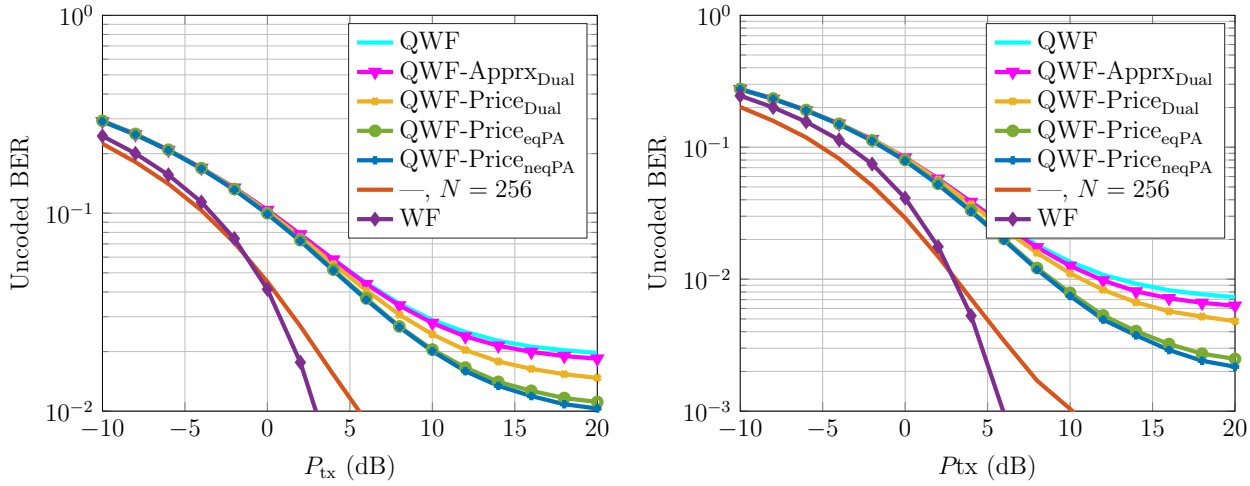
(a) $Q = 4$ (b) $Q = 8$

Fig. 5.2: Comparison between different linear precoders for 16-QAM signaling: $N = 128$, $M = 8$ with the i.i.d. channel.



(a) $Q = 4$ (b) $Q = 8$

Fig. 5.3: Comparison between different linear precoders for 4-PSK signaling: $N = 64$, $M = 8$ with the mmW sparse channel.



(a) $Q = 4$ (b) $Q = 8$

Fig. 5.4: Comparison between different linear precoders for 16-QAM signaling: $N = 128$, $M = 8$ with the mmW sparse channel.

6. Frequency-Selective Channels

6.1 Input-Output Relationship

In this section, we consider linear precoding for frequency-selective channels; that is $L > 1$. This implies that more than one precoder tap is required to mitigate the Inter-Symbol Interference (ISI); that is $L_p > 1$. Thus, the precoded signal vector is given by

$$\begin{aligned}
\mathbf{x}[t] &= \sum_{\ell'=0}^{L_p-1} \mathbf{P}_{\ell'} \mathbf{s}[t-\ell'] \\
&= \sum_{\ell'=0}^{L_p-1} \sum_{n=1}^{N} \sum_{m=1}^{M} \mathbf{e}_n \mathbf{e}_n^{\mathrm{T}} \mathbf{P}_{\ell'} \mathbf{e}_m \mathbf{e}_m^{\mathrm{T}} \mathbf{s}[t-\ell'] \\
&= \sum_{\ell'=0}^{L_p-1} \sum_{n=1}^{N} \sum_{m=1}^{M} \mathbf{e}_n \mathbf{e}_m^{\mathrm{T}} \mathbf{P}_{\ell'}^{\mathrm{T}} \mathbf{e}_n \mathbf{e}_m^{\mathrm{T}} \mathbf{s}[t-\ell'] \\
&= \sum_{n=1}^{N} \mathbf{e}_n \mathbf{p}^{\mathrm{T}} \left(\mathbf{I}_{ML_p} \otimes \mathbf{e}_n \right) \begin{bmatrix} \mathbf{s}[t] \\ \mathbf{s}[t-1] \\ \vdots \\ \mathbf{s}[t-L_p+1] \end{bmatrix},
\end{aligned} \tag{6.1}$$

where

$$\mathbf{p}^{\mathrm{T}} = \begin{bmatrix} \mathbf{e}_1^{\mathrm{T}} \mathbf{P}_1^{\mathrm{T}} & \cdots & \mathbf{e}_M^{\mathrm{T}} \mathbf{P}_1^{\mathrm{T}} & \cdots & \mathbf{e}_1^{\mathrm{T}} \mathbf{P}_{L_p-1}^{\mathrm{T}} & \cdots & \mathbf{e}_M^{\mathrm{T}} \mathbf{P}_{L_p-1}^{\mathrm{T}} \end{bmatrix}. \tag{6.2}$$

For the vector $\mathbf{x}$ defined in (4.11) we get

$$\mathbf{x} = \sum_{i=1}^{NL} \mathbf{e}_i \mathbf{p}^{\mathrm{T}} \mathbf{\Gamma}_i \begin{bmatrix} \mathbf{s}[t] \\ \vdots \\ \mathbf{s}[t-L-L_p+1] \end{bmatrix} \tag{6.3}$$

$$= \left(\mathbf{I}_{NL} \otimes \mathbf{p}^{\mathrm{T}} \right) \mathbf{\Gamma} \mathbf{s}, \tag{6.4}$$

where

$$\mathbf{\Gamma}_i = \mathbf{S} \left(\left\lceil \frac{i}{N} \right\rceil - 1, L_p, L-1 \right) \otimes \mathbf{I}_M \otimes \mathbf{e}_{\mathrm{mod}(i-1,N)+1}, \tag{6.5}$$

$$\mathbf{S}(i,j,\ell) = \left[\mathbf{0}_{j,i}, \mathbf{I}_j, \mathbf{0}_{j,\ell-i} \right] \tag{6.6}$$

$$\mathbf{\Gamma} = \sum_{i=1}^{NL} \mathbf{e}_i \otimes \mathbf{\Gamma}_i, \tag{6.7}$$

and

$$\mathbf{s} = \begin{bmatrix}\mathbf{s}[t]^{\mathrm{T}} & \mathbf{s}[t-1]^{\mathrm{T}} & \cdots & \mathbf{s}[t-L-L_p+1]^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}. \tag{6.8}$$

Thus, we recall the input-output relationship in (4.8) and use the identity in (6.4) to get

$$\begin{aligned}\mathbf{u} &= \mathbf{G}\left(\mathbf{H}\mathbf{t} + \boldsymbol{\eta}\right)\\ &= \mathbf{G}\left(\mathbf{H}\mathcal{Q}_{\mathrm{CE}}\left(\left(\mathbf{I}_{NL}\otimes\mathbf{p}^{\mathrm{T}}\right)\boldsymbol{\Gamma}\mathbf{s}\right) + \boldsymbol{\eta}\right).\end{aligned} \tag{6.9}$$

6.2 Optimization Problem

The precoding task consists of finding the optimal linear precoding vector $\mathbf{p}_{\mathrm{opt}}$ and the diagonal positive real-valued receive filter $\mathbf{G}_{\mathrm{opt}}$ that minimize the MSE between the desired signal before a time delay τ and the received signal, i.e. $\mathbf{s}[t-\tau]$ and $\mathbf{u}[t]$, under the transmit power constraint given in (4.20). The MMSE optimization problem is expressed as

$$\begin{aligned}\{\mathbf{p}_{\mathrm{opt}}, \mathbf{G}_{\mathrm{opt}}, \tau_{\mathrm{opt}}\} &= \underset{\mathbf{p},\mathbf{G},\tau}{\arg\min}\, \mathrm{E}\left[\|\mathbf{u}[t] - \mathbf{s}[t-\tau]\|_2^2\right] \qquad \text{s.t} \quad \alpha_q \,\mathrm{tr}\left(\mathbf{C}_{\mathbf{xx}}\right) \leq P_{\mathrm{tx}}\\ &= \underset{\mathbf{p},\mathbf{G},\tau}{\arg\min}\, \mathrm{E}\left[\left\|\mathbf{u} - \left(\mathbf{e}_{\tau+1}^{\mathrm{T}}\otimes\mathbf{I}_M\right)\mathbf{s}\right\|_2^2\right] \qquad \text{s.t} \quad \alpha_q \,\mathrm{tr}\left(\mathbf{C}_{\mathbf{xx}}\right) \leq P_{\mathrm{tx}}.\end{aligned} \tag{6.10}$$

6.2.1 MSE

In general, the MSE expression is given by

$$\begin{aligned}\mathrm{MSE} = {}& \mathrm{tr}\left(\mathbf{G}\mathbf{H}\mathbf{C}_{\mathbf{tt}}\mathbf{H}^{\mathrm{H}}\mathbf{G}\right) + \mathrm{tr}\left(\mathbf{G}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\mathbf{G}\right) - \mathrm{tr}\left(\mathbf{G}\mathbf{H}\mathbf{C}_{\mathbf{ts}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\right)\\ &- \mathrm{tr}\left(\left(\mathbf{e}_{\tau+1}^{\mathrm{T}}\otimes\mathbf{I}_M\right)\mathbf{C}_{\mathbf{st}}\mathbf{H}^{\mathrm{H}}\mathbf{G}\right) + \mathrm{tr}\left(\left(\mathbf{e}_{\tau+1}^{\mathrm{T}}\otimes\mathbf{I}_M\right)\mathbf{C}_{\mathbf{ss}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\right).\end{aligned} \tag{6.11}$$

In massive MIMO systems, the entries in $\mathbf{x}$ are approximately Gaussian distributed due to the central limit theorem [68]. Hence, Price's theorem can be applied to compute the covariance matrices $\mathbf{C}_{\mathbf{tt}}$ and $\mathbf{C}_{\mathbf{tx}}$ as given in (3.66) and (3.65). Consequently, the MSE expression calculates to

$$\begin{aligned}\mathrm{MSE} = {}& \frac{Q}{\pi}\sin^2(\psi)\sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\,\mathrm{tr}\left(\arcsin\left(\Re\{\mathbf{R}_{\mathbf{xx}}\,\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\}\right)\boldsymbol{\Xi}\mathbf{H}^{\mathrm{H}}\mathbf{G}^2\mathbf{H}\boldsymbol{\Xi}\right)\\ &+ \mathrm{tr}\left(\mathbf{G}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\mathbf{G}\right) - \alpha_q\,\mathrm{tr}\left(\mathbf{G}\mathbf{H}\mathbf{C}_{\mathbf{xs}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\right) - \alpha_q\,\mathrm{tr}\left(\left(\mathbf{e}_{\tau+1}^{\mathrm{T}}\otimes\mathbf{I}_M\right)\mathbf{C}_{\mathbf{sx}}\mathbf{H}^{\mathrm{H}}\mathbf{G}\right)\\ &+ \mathrm{tr}\left(\left(\mathbf{e}_{\tau+1}^{\mathrm{T}}\otimes\mathbf{I}_M\right)\mathbf{C}_{\mathbf{ss}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\right).\end{aligned} \tag{6.12}$$

According to (6.4), the covariance matrix $\mathbf{C}_{\mathbf{xx}}$ is expressed in terms of the precoder $\mathbf{p}$ as

$$\begin{aligned}\mathbf{C}_{\mathbf{xx}} &= \sum_{i=1}^{NL}\sum_{j=1}^{NL}\mathbf{p}^{\mathrm{T}}\boldsymbol{\Gamma}_i\mathbf{C}_{\mathbf{ss}}\boldsymbol{\Gamma}_j^{\mathrm{T}}\mathbf{p}^*\mathbf{e}_i\mathbf{e}_j^{\mathrm{T}} &(6.13)\\ &= \left(\mathbf{I}_{NL}\otimes\mathbf{p}^{\mathrm{T}}\right)\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{ss}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\mathbf{I}_{NL}\otimes\mathbf{p}^*\right). &(6.14)\end{aligned}$$

Therefore, we obtain the following expressions

$$\begin{aligned}\mathbf{R}_{\mathbf{xx}} = {}& \mathrm{diag}\left(\left(\mathbf{I}_{NL}\otimes\mathbf{p}^{\mathrm{T}}\right)\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{ss}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\mathbf{I}_{NL}\otimes\mathbf{p}^*\right)\right)^{-1/2}\\ &\left(\mathbf{I}_{NL}\otimes\mathbf{p}^{\mathrm{T}}\right)\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{ss}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\mathbf{I}_{NL}\otimes\mathbf{p}^*\right)\mathrm{diag}\left(\left(\mathbf{I}_{NL}\otimes\mathbf{p}^{\mathrm{T}}\right)\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{ss}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\mathbf{I}_{NL}\otimes\mathbf{p}^*\right)\right)^{-1/2},\end{aligned} \tag{6.15}$$

and

$$\boldsymbol{\Xi} = \mathrm{diag}\left(\left(\mathbf{I}_{NL}\otimes\mathbf{p}^{\mathrm{T}}\right)\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{ss}}\boldsymbol{\Gamma}^{\mathrm{T}}\left(\mathbf{I}_{NL}\otimes\mathbf{p}^*\right)\right)^{1/2}. \tag{6.16}$$

6.2.2 Transmit Power Constraint

After plugging (6.14) in the transmit power constraint in (4.20) and for maximal exploitation of the available power, we get

$$\alpha_q \operatorname{tr}\left(\left(\mathbf{I}_{NL} \otimes \mathbf{p}^{\mathrm{T}}\right) \boldsymbol{\Gamma} \mathbf{C}_{\mathbf{ss}} \boldsymbol{\Gamma}^{\mathrm{T}} \left(\mathbf{I}_{NL} \otimes \mathbf{p}^{*}\right)\right) = L P_{\mathrm{tx}}, \tag{6.17}$$

which simplifies to

$$\alpha_q \mathbf{p}^{\mathrm{T}} \sum_{i=1}^{NL} \boldsymbol{\Gamma}_i \mathbf{C}_{\mathbf{ss}} \boldsymbol{\Gamma}_i^{\mathrm{T}} \mathbf{p}^{*} = L P_{\mathrm{tx}}. \tag{6.18}$$

Based on the assumption that $\mathbf{C}_{\mathbf{ss}} = \sigma_s^2 \mathbf{I}_{M(L+L_p-1)}$ and using the identity

$$\mathbf{S}(i,j,\ell)\mathbf{S}(i,j,\ell)^{\mathrm{T}} = \mathbf{I}_j \tag{6.19}$$

that leads to $\boldsymbol{\Gamma}_i \mathbf{C}_{\mathbf{ss}} \boldsymbol{\Gamma}_i^{\mathrm{T}} = \boldsymbol{\Gamma}_{kN+i} \mathbf{C}_{\mathbf{ss}} \boldsymbol{\Gamma}_{kN+i}^{\mathrm{T}},\ k \in \mathbb{Z}+$, we can further simplify the transmit power constraint and obtain

$$\alpha_q \mathbf{p}^{\mathrm{T}} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i \mathbf{C}_{\mathbf{ss}} \boldsymbol{\Gamma}_i^{\mathrm{T}} \mathbf{p}^{*} = P_{\mathrm{tx}}. \tag{6.20}$$

6.2.2.1 Equal Power Allocation

For equal power allocation, it must holds that

$$\mathbf{p}^{\mathrm{T}} \boldsymbol{\Gamma}_i \mathbf{C}_{\mathbf{ss}} \boldsymbol{\Gamma}_i^{\mathrm{T}} \mathbf{p}^{*} = \frac{P_{\mathrm{tx}}}{\alpha_q N},\ i = 1, \cdots, N. \tag{6.21}$$

According to the definition of $\boldsymbol{\Gamma}_i$ in (6.5) and the identity in (6.19), we get

$$\boldsymbol{\Gamma}_i \mathbf{C}_{\mathbf{ss}} \boldsymbol{\Gamma}_i^{\mathrm{T}} = \sigma_s^2 \left(\mathbf{I}_{L_p M} \otimes \left(\mathbf{e}_i \mathbf{e}_i^{\mathrm{T}}\right)\right). \tag{6.22}$$

Thus, the transmit power constraint is expressed as

$$\mathbf{p}^{\mathrm{T}} \left(\mathbf{I}_{L_p M} \otimes \left(\mathbf{e}_i \mathbf{e}_i^{\mathrm{T}}\right)\right) \mathbf{p}^{*} = \frac{P_{\mathrm{tx}}}{\sigma_s^2 \alpha_q N},\ i = 1, \cdots, N. \tag{6.23}$$

6.2.2.2 Unequal Power Allocation

For unequal power allocation, the transmit power constraint is given in (6.20). From (6.22), we can conclude that

$$\sum_{i=1}^{N} \boldsymbol{\Gamma}_i \mathbf{C}_{\mathbf{ss}} \boldsymbol{\Gamma}_i^{\mathrm{T}} = \sigma_s^2 \mathbf{I}_{MNL_p}. \tag{6.24}$$

Hence, the transmit power constraint simplifies to

$$\|\mathbf{p}\|_2^2 \leq \frac{P_{\mathrm{tx}}}{\sigma_s^2 \alpha_q}. \tag{6.25}$$

6.3 Precoder Designs in the Primal Domain

6.3.1 Precoder Design Based on Gradient Projection Algorithm

Since the MSE expression in (6.12) is highly non-linear in $\mathbf{C_{xx}} = \left(\mathbf{I}_{NL} \otimes \mathbf{p}^{\mathrm{T}}\right) \mathbf{\Gamma} \mathbf{C_{ss}} \mathbf{\Gamma}^{\mathrm{T}} \left(\mathbf{I}_{NL} \otimes \mathbf{p}^{*}\right)$ and thus in $\mathbf{p}$, we cannot find a closed-form expression for $\mathbf{p}_{\mathrm{opt}}$. Therefore, we use the Gradient Projection algorithm. To this end, we have to compute the derivatives of MSE w.r.t. $\mathbf{p}$ and $\mathbf{G}$.

$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{p}} = \frac{Q}{\pi} \sin^2(\psi) \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \frac{\partial \operatorname{tr}\left(\arcsin\left(\Re\{\mathbf{R_{xx}}\, \mathrm{e}^{-\mathrm{j}\, 2\Delta k\psi}\}\right) \mathbf{\Xi} \mathbf{H}^{\mathrm{H}} \mathbf{G}^2 \mathbf{H} \mathbf{\Xi}\right)}{\partial \mathbf{p}} - \frac{Q^2}{4\pi} \sin^2(\psi) \frac{\partial \operatorname{tr}\left(\left(\mathrm{e}_{\tau+1} \otimes \mathbf{I}_M\right) \mathbf{G} \mathbf{H} \mathbf{C_{xs}}\right)}{\partial \mathbf{p}}, \tag{6.26}$$

where

$$\begin{aligned}
&\frac{\partial \operatorname{tr}\left(\arcsin\left(\Re\{\mathbf{R_{xx}}\, \mathrm{e}^{-\mathrm{j}\, 2\Delta k\psi}\}\right) \mathbf{\Xi} \mathbf{H}^{\mathrm{H}} \mathbf{G}^2 \mathbf{H} \mathbf{\Xi}\right)}{\partial \mathbf{p}} \\
&= \frac{1}{2} \sum_{i=1}^{NL} \sum_{j=1}^{NL} \mathbf{e}_i^{\mathrm{T}} \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{-1/2} \left(\mathbf{\Omega}^{\mathrm{T}}\, \mathrm{e}^{-\mathrm{j}\, 2\Delta k\psi} + \mathbf{\Omega}\, \mathrm{e}^{\mathrm{j}\, 2\Delta k\psi}\right) \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{-1/2} \mathbf{e}_j \mathbf{\Gamma}_i \mathbf{C_{ss}} \mathbf{\Gamma}_j^{\mathrm{T}} \mathbf{p}^{*} \\
&+ \frac{1}{2} \sum_{i=1}^{NL} \mathbf{e}_i^{\mathrm{T}} \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{-1/2} \mathbf{\Xi} \mathbf{H}^{\mathrm{H}} \mathbf{G}^2 \mathbf{H} \mathbf{\Xi} \arcsin\left(\Re\{\mathbf{R_{xx}}\, \mathrm{e}^{-\mathrm{j}\, 2\Delta k\psi}\}\right) \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{-1/2} \mathbf{e}_i \mathbf{\Gamma}_i \mathbf{C_{ss}} \mathbf{\Gamma}_i^{\mathrm{T}} \mathbf{p}^{*} \\
&+ \frac{1}{2} \sum_{i=1}^{NL} \mathbf{e}_i^{\mathrm{T}} \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{-1/2} \arcsin\left(\Re\{\mathbf{R_{xx}}\, \mathrm{e}^{-\mathrm{j}\, 2\Delta k\psi}\}\right) \mathbf{\Xi} \mathbf{H}^{\mathrm{H}} \mathbf{G}^2 \mathbf{H} \mathbf{\Xi} \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{-1/2} \mathbf{e}_i \mathbf{\Gamma}_i \mathbf{C_{ss}} \mathbf{\Gamma}_i^{\mathrm{T}} \mathbf{p}^{*} \\
&- \frac{1}{2} \sum_{i=1}^{NL} \mathbf{e}_i^{\mathrm{T}} \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{-1/2} \left(\Re\{\mathbf{R_{xx}}\, \mathrm{e}^{-\mathrm{j}\, 2\Delta k\psi}\} \mathbf{\Omega} + \mathbf{\Omega} \Re\{\mathbf{R_{xx}}\, \mathrm{e}^{-\mathrm{j}\, 2\Delta k\psi}\}\right) \operatorname{diag}\left(\mathbf{C_{xx}}\right)^{-1/2} \mathbf{e}_i \mathbf{\Gamma}_i \mathbf{C_{ss}} \mathbf{\Gamma}_i^{\mathrm{T}} \mathbf{p}^{*},
\end{aligned} \tag{6.27}$$

$\mathbf{\Omega}$ is defined in (5.16) and

$$\frac{\partial \operatorname{tr}\left(\left(\mathrm{e}_{\tau+1} \otimes \mathbf{I}_M\right) \mathbf{G} \mathbf{H} \mathbf{C_{xs}}\right)}{\partial \mathbf{p}} = \sum_{i=1}^{NL} \mathbf{\Gamma}_i \mathbf{C_{ss}} \left(\mathrm{e}_{\tau+1} \otimes \mathbf{I}_M\right) \mathbf{G} \mathbf{H} \mathbf{e}_i. \tag{6.28}$$

The derivative of the MSE expression in (6.12) w.r.t. $\mathbf{G}$ is expressed as

$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{G}} = 2 \operatorname{diag}\left(\mathbf{H} \mathbf{C_{tt}} \mathbf{H}^{\mathrm{H}} + \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\right) - \operatorname{diag}\left(\mathbf{H} \mathbf{C_{ts}} \left(\mathrm{e}_{\tau+1} \otimes \mathbf{I}_M\right)\right) - \operatorname{diag}\left(\mathbf{H}^{*} \mathbf{C}_{\mathbf{ts}}^{*} \left(\mathrm{e}_{\tau+1} \otimes \mathbf{I}_M\right)\right). \tag{6.29}$$

Thus, the optimal filter $\mathbf{G}_{\mathrm{opt}}$ is obtained by setting (6.29) equal to zero and is given by

$$\mathbf{G}_{\mathrm{opt}} = \mathsf{g}_{\mathsf{opt}}\left(\mathbf{p}, \tau\right) = \left| \operatorname{diag}\left(\mathbf{H} \mathbf{C_{tt}} \mathbf{H}^{\mathrm{H}} + \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\right)^{-1} \operatorname{diag}\left(\Re\{\mathbf{H} \mathbf{C_{ts}} \left(\mathrm{e}_{\tau+1} \otimes \mathbf{I}_M\right)\}\right) \right|, \tag{6.30}$$

where the $|\bullet|$ operator is applied element-wise to the matrix entries. The Gradient Projection algorithms for equal and unequal power allocation are given in Algorithm 4 and Algorithm 5, respectively. In each iteration, the MSE expression is computed for different values of τ. The value of τ that leads to the minimal MSE is the optimal τ.

Algorithm 4 Gradient Projection Algorithm to obtain the FIR-QWF-Price precoder with equal power allocation.

Initialization
$\mathbf{p}_{(0)},\ \mathbf{G}_{(0)} = \mathtt{g}_{\mathtt{opt}}\left(\mathbf{p}_{(0)}\right),\ \mu = 10$ and $n = 0$
repeat
  $\mathbf{p}_{(n+1)} = \mathbf{p}_{(n)} - \mu \left(\frac{\partial \mathrm{MSE}_{(n)}}{\partial \mathbf{p}}\right)^*$
  $\mathbf{p}_{(n+1)} = \sqrt{\frac{P_{\mathrm{tx}}}{\alpha_q N}}\, \mathrm{diag}\left(\left(\mathbf{I}_{NML_p} \otimes \mathbf{p}^{\mathrm{T}}\right) \mathbf{\Gamma}' \mathbf{C}_{\mathbf{ss}} \mathbf{\Gamma}'^{\mathrm{T}} \left(\mathbf{I}_{NML_p} \otimes \mathbf{p}^*\right)\right)^{-1/2} \mathbf{p}_{(n+1)}$,
  where $\mathbf{\Gamma}' = \sum_{i=1}^{NML_p} \mathbf{e}_i \otimes \mathbf{\Gamma}_i$ {Equal power allocation constraint}
  $\tau = \arg\min_{\tau} \mathrm{MSE}(\mathbf{p}_{(n+1)}, \tau)$
  $\mathbf{G}_{(n+1)} = \mathtt{g}_{\mathtt{opt}}\left(\mathbf{p}_{(n+1)}, \tau\right)$
  if $\mathrm{MSE}_{(n+1)} > \mathrm{MSE}_{(n)}$ **then**
    $\mu = \mu/2$
  else
    $n = n + 1$
  end if
until $\frac{\left|\mathrm{MSE}_{(n+1)} - \mathrm{MSE}_{(n)}\right|}{\mathrm{MSE}_{(n)}} \leq \epsilon$

Algorithm 5 Gradient Projection Algorithm to obtain the FIR-QWF-Price precoder with unequal power allocation.

Initialization
$\mathbf{p}_{(0)},\ \mathbf{G}_{(0)} = \mathtt{g}_{\mathtt{opt}}\left(\mathbf{p}_{(0)}\right),\ \mu = 10$ and $n = 0$
repeat
  $\mathbf{p}_{(n+1)} = \mathbf{p}_{(n)} - \mu \left(\frac{\partial \mathrm{MSE}_{(n)}}{\partial \mathbf{p}}\right)^*$
  $\mathbf{p}_{(n+1)} = \sqrt{\frac{P_{\mathrm{tx}}}{\sigma_s^2 \alpha_q}}\, \left\|\mathbf{p}_{(n+1)}\right\|_2^{-1} \mathbf{p}_{(n+1)}$ {Unequal power allocation constraint}
  $\tau = \arg\min_{\tau} \mathrm{MSE}(\mathbf{p}_{(n+1)}, \tau)$
  $\mathbf{G}_{(n+1)} = \mathtt{g}_{\mathtt{opt}}\left(\mathbf{p}_{(n+1)}, \tau\right)$
  if $\mathrm{MSE}_{(n+1)} > \mathrm{MSE}_{(n)}$ **then**
    $\mu = \mu/2$
  else
    $n = n + 1$
  end if
until $\frac{\left|\mathrm{MSE}_{(n+1)} - \mathrm{MSE}_{(n)}\right|}{\mathrm{MSE}_{(n)}} \leq \epsilon$

6.3.2 Precoder Design Based on LCA

In this section, we recall the linear precoder design in [25] and apply it to the case of QCE transmit signals for frequency-selective channels. The precoder aims at minimizing the MSE under the constraint of unequal power allocation, where the receive processing is assumed to be a scaled identity matrix, i.e. $\mathbf{G} = g\mathbf{I}_M$. Note that the case of equal power allocation will not be considered, since it must be solved with iterative algorithms and hence no benefits in terms of computational complexity are obtained by using the LCA. However, applying unequal power allocation and assuming a scaled identity processing at the receiver will lead to a closed-form expression of the precoder, as detailed in this section.

Using the LCA and the appropriate expressions in (3.74) and (3.75), we get

$$\begin{aligned}\mathrm{MSE} = \mathrm{tr}\Big(& |g|^2 \mathbf{H}\left(\alpha_q^2 \mathbf{C_{xx}} + \alpha_q\beta_q \,\mathrm{diag}(\mathbf{C_{xx}})\right)\mathbf{H}^{\mathrm{H}} \\ & - g\mathbf{H}\alpha_q \mathbf{C_{xs}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right) - g^*\alpha_q\left(\mathbf{e}_{\tau+1}^T\otimes\mathbf{I}_M\right)\mathbf{C_{sx}}\mathbf{H}^{\mathrm{H}} + |g|^2\mathbf{C}_{\boldsymbol{\eta\eta}}\Big) \\ & + \mathrm{tr}\left(\left(\mathbf{e}_{\tau+1}^{\mathrm{T}}\otimes\mathbf{I}_M\right)\mathbf{C_{ss}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\right). \end{aligned} \tag{6.31}$$

The Lagrangian function is expressed by

$$\begin{aligned}\mathcal{L}(\mathbf{P},g,\lambda) = \mathrm{tr}\Big(& |g|^2 \mathbf{H}\left(\alpha_q^2 \mathbf{C_{xx}} + \alpha_q\rho_q \,\mathrm{diag}(\mathbf{C_{xx}})\right)\mathbf{H}^{\mathrm{H}} \\ & - g\mathbf{H}\alpha_q \mathbf{C_{xs}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right) - g^*\alpha_q\left(\mathbf{e}_{\tau+1}^T\otimes\mathbf{I}_M\right)\mathbf{C_{sx}}\mathbf{H}^{\mathrm{H}} + |g|^2\mathbf{C}_{\boldsymbol{\eta\eta}}\Big) \\ & + \mathrm{tr}\left(\left(\mathbf{e}_{\tau+1}^{\mathrm{T}}\otimes\mathbf{I}_M\right)\mathbf{C_{ss}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\right) + \lambda\left(\|\mathbf{p}\|_2^2 - \frac{P_{\mathrm{tx}}}{\sigma_s^2\alpha_q}\right) \\ \overset{(5.6),(6.3)}{=} \mathrm{tr}\Big(& |g|^2 \mathbf{H}\left(\alpha_q^2\sum_{i=1}^{NL}\sum_{j=1}^{NL}\mathbf{p}^{\mathrm{T}}\boldsymbol{\Gamma}_i\mathbf{C_{ss}}\boldsymbol{\Gamma}_j^{\mathrm{T}}\mathbf{p}^*\mathbf{e}_i\mathbf{e}_j^{\mathrm{T}} + \alpha_q\beta_q\sum_{i=1}^{NL}\mathbf{p}^{\mathrm{T}}\boldsymbol{\Gamma}_i\mathbf{C_{ss}}\boldsymbol{\Gamma}_i^{\mathrm{T}}\mathbf{p}^*\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}}\right)\mathbf{H}^{\mathrm{H}} \\ & - g\mathbf{H}\alpha_q\sum_{i=1}^{NL}\mathbf{e}_i\mathbf{p}^{\mathrm{T}}\boldsymbol{\Gamma}_i\mathbf{C_{ss}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right) - g^*\alpha_q\left(\mathbf{e}_{\tau+1}^T\otimes\mathbf{I}_M\right)\mathbf{C_{ss}}\sum_{i=1}^{NL}\boldsymbol{\Gamma}_i^{\mathrm{T}}\mathbf{p}^*\mathbf{e}_i^{\mathrm{T}}\mathbf{H}^{\mathrm{H}}\Big) \\ & + \mathrm{tr}\left(|g|^2\mathbf{C}_{\boldsymbol{\eta\eta}}\right) + \mathrm{tr}\left(\left(\mathbf{e}_{\tau+1}^{\mathrm{T}}\otimes\mathbf{I}_M\right)\mathbf{C_{ss}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\right) + \lambda\left(\|\mathbf{p}\|_2^2 - \frac{P_{\mathrm{tx}}}{\sigma_s^2\alpha_q}\right). \end{aligned} \tag{6.32}$$

The KKT equations are then given by

$$\begin{aligned}\frac{\partial\mathcal{L}(\mathbf{p},g,\lambda)}{\partial\mathbf{p}} = & |g|^2\alpha_q^2\sum_{i=1}^{NL}\sum_{j=1}^{NL}\mathbf{e}_i^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}^*\mathbf{e}_j\boldsymbol{\Gamma}_i\mathbf{C_{ss}}\boldsymbol{\Gamma}_j^{\mathrm{T}}\mathbf{p}^* + |g|^2\alpha_q\beta_q\sum_{i=1}^{NL}\mathbf{e}_i^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}^*\mathbf{e}_i\boldsymbol{\Gamma}_i\mathbf{C_{ss}}\boldsymbol{\Gamma}_i^{\mathrm{T}}\mathbf{p}^* \\ & - g\alpha_q\sum_{i=1}^{NL}\boldsymbol{\Gamma}_i\mathbf{C_{ss}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\mathbf{H}\mathbf{e}_i + \lambda\mathbf{p}^* = \mathbf{0}_{NML_p\times 1}, \end{aligned} \tag{6.33}$$

$$\frac{\partial\mathcal{L}(\mathbf{p},g,\lambda)}{\partial g} = g^*\,\mathrm{tr}\left(\mathbf{H}\left(\alpha_q^2\mathbf{C_{xx}} + \alpha_q\beta_q\,\mathrm{diag}(\mathbf{C_{xx}})\right)\mathbf{H}^{\mathrm{H}} + \mathbf{C}_{\boldsymbol{\eta\eta}}\right) - \alpha_q\,\mathrm{tr}\left(\mathbf{H}\mathbf{C_{xs}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\right) = 0, \tag{6.34}$$

and

$$\frac{\partial\mathcal{L}(\mathbf{p},g,\lambda)}{\partial\lambda} = \left(\|\mathbf{p}\|_2^2 - \frac{P_{\mathrm{tx}}}{\sigma_s^2\alpha_q}\right) = 0. \tag{6.35}$$

Multiplying (6.33) by $\mathbf{p}^{\mathrm{T}}$ from the left side and taking the trace leads to

$$|g|^2 \alpha_q^2 \operatorname{tr}\left(\sum_{i=1}^{NL}\sum_{j=1}^{NL} \mathbf{e}_i^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}^{*}\mathbf{e}_j\mathbf{p}^{\mathrm{T}}\mathbf{\Gamma}_i\mathbf{C}_{\mathbf{ss}}\mathbf{\Gamma}_j^{\mathrm{T}}\mathbf{p}^{*}\right) + |g|^2 \alpha_q\beta_q \operatorname{tr}\left(\sum_{i=1}^{NL} \mathbf{e}_i^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}^{*}\mathbf{e}_i\mathbf{p}^{\mathrm{T}}\mathbf{\Gamma}_i\mathbf{C}_{\mathbf{ss}}\mathbf{\Gamma}_i^{\mathrm{T}}\mathbf{p}^{*}\right)$$
$$- g\alpha_q \operatorname{tr}\left(\sum_{i=1}^{NL} \mathbf{p}^{\mathrm{T}}\mathbf{\Gamma}_i\mathbf{C}_{\mathbf{ss}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\mathbf{H}\mathbf{e}_i\right) - \lambda\|\mathbf{p}\|_2^2 = 0. \tag{6.36}$$

From (6.34), we get

$$\begin{aligned}\alpha_q \operatorname{tr}\left(\mathbf{H}\mathbf{C}_{\mathbf{xs}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\right) &= \alpha_q \operatorname{tr}\left(\sum_{i=1}^{NL}\mathbf{H}\mathbf{e}_i\mathbf{p}^{\mathrm{T}}\mathbf{\Gamma}_i\mathbf{C}_{\mathbf{ss}}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\right)\\ &= g^{*}\operatorname{tr}\left(\mathbf{H}\left(\alpha_q^2\mathbf{C}_{\mathbf{xx}} + \alpha_q\beta_q\operatorname{diag}(\mathbf{C}_{\mathbf{xx}})\,\mathbf{H}^{\mathrm{H}} + \mathbf{C}_{\boldsymbol{\eta\eta}}\right)\right),\end{aligned} \tag{6.37}$$

which when inserted in (6.36) gives the expression of λ

$$\lambda = \frac{|g|^2\operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\eta\eta}}\right)}{\|\mathbf{p}\|_2^2} = \frac{\sigma_s^2\alpha_q\,|g|^2\operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\eta\eta}}\right)}{P_{\mathrm{tx}}}, \tag{6.38}$$

where we used (6.35). Inserting (6.38) in (6.33) and solving it for $\mathbf{p}$, we obtain

$$\mathbf{p} = \frac{1}{g}\left(\alpha_q\sum_{i=1}^{NL}\sum_{j=1}^{NL}\mathbf{e}_i^{\mathrm{T}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{e}_j\mathbf{\Gamma}_i\mathbf{C}_{\mathbf{ss}}^{*}\mathbf{\Gamma}_j^{\mathrm{T}} + \beta_q\sum_{i=1}^{NL}\mathbf{e}_i^{\mathrm{T}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{e}_i\mathbf{\Gamma}_i\mathbf{C}_{\mathbf{ss}}^{*}\mathbf{\Gamma}_i^{\mathrm{T}} + \frac{\sigma_s^2\operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\eta\eta}}\right)}{P_{\mathrm{tx}}}\mathbf{I}_{NML_p}\right)^{-1}$$
$$\sum_{i=1}^{NL}\mathbf{\Gamma}_i\mathbf{C}_{\mathbf{ss}}^{*}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\mathbf{H}^{*}\mathbf{e}_i. \tag{6.39}$$

The optimal g is found by satisfying (6.35) with $\mathbf{p}$ from (6.39); that is

$$g = \sqrt{\frac{\sigma_s^2\alpha_q}{P_{\mathrm{tx}}}}\left(\sum_{i=1}^{NL}\sum_{j=1}^{NL}\mathbf{e}_j^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{e}_{\tau+1}^{\mathrm{T}}\otimes\mathbf{I}_M\right)\mathbf{C}_{\mathbf{ss}}^{*}\mathbf{\Gamma}_j^{\mathrm{T}}\right.$$
$$\left(\alpha_q\sum_{i=1}^{NL}\sum_{j=1}^{NL}\mathbf{e}_i^{\mathrm{T}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{e}_j\mathbf{\Gamma}_i\mathbf{C}_{\mathbf{ss}}^{*}\mathbf{\Gamma}_j^{\mathrm{T}} + \beta_q\sum_{i=1}^{NL}\mathbf{e}_i^{\mathrm{T}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{e}_i\mathbf{\Gamma}_i\mathbf{C}_{\mathbf{ss}}^{*}\mathbf{\Gamma}_i^{\mathrm{T}} + \frac{\sigma_s^2\operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\eta\eta}}\right)}{P_{\mathrm{tx}}}\mathbf{I}_{NML_p}\right)^{-2}$$
$$\left.\mathbf{\Gamma}_i\mathbf{C}_{\mathbf{ss}}^{*}\left(\mathbf{e}_{\tau+1}\otimes\mathbf{I}_M\right)\mathbf{H}^{*}\mathbf{e}_i\right)^{1/2}. \tag{6.40}$$

Each value of τ determines a different precoder vector $\mathbf{p}$ and a different scalar g, which in turn determine the value of MSE. Therefore, the MSE expression is computed for different values of τ. The value of τ and the corresponding filters $\mathbf{p}$ and g that lead to the minimal MSE are the optimal solutions.

6.4 Dual Optimization Problem

6.4.1 Exact Dual Problem

To find the exact dual optimization problem, we first state the input-output relationship in the uplink system model illustrated in Fig. 4.3; that is

$$\mathbf{u}^{\mathrm{UL}}[t] = \sum_{\ell'=0}^{L_p-1} \mathbf{F}_{\ell'} \left(\alpha_q \left(\sum_{\ell=0}^{L-1} \mathbf{H}_{\ell}^{\mathrm{H}} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}} \mathbf{T} \mathbf{s}^{\mathrm{UL}}[t-\ell'-\ell] + \boldsymbol{\eta}^{\mathrm{UL}}[t-\ell'] \right) + \mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}[t-\ell'] \right). \tag{6.41}$$

Hence, the uplink MSE expression is given by

$$\begin{aligned}
\mathrm{MSE}^{\mathrm{UL}} &= \mathrm{E}\left[\left\| \hat{\mathbf{s}}^{\mathrm{UL}}[t] - \mathbf{s}^{\mathrm{UL}}[t-\tau] \right\|_2^2 \right] \\
&= \alpha_q^2 \,\mathrm{tr}\left(\sum_{\ell_1'=0}^{L_p-1} \sum_{\ell_2'=0}^{L_p-1} \sum_{\ell_1=0}^{L-1} \sum_{\ell_2=0}^{L-1} \mathbf{F}_{\ell_1'} \mathbf{H}_{\ell_1}^{\mathrm{H}} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}} \mathbf{T}\, \mathrm{E}\left[\mathbf{s}^{\mathrm{UL}}[t-\ell_1-\ell_1'] \mathbf{s}^{\mathrm{UL,H}}[t-\ell_2-\ell_2'] \right] \mathbf{T} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2} \mathbf{H}_{\ell_2} \mathbf{F}_{\ell_2'}^{\mathrm{H}} \right) \\
&+ \alpha_q^2 \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1} \mathbf{F}_{\ell'} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{\mathrm{UL}} \mathbf{F}_{\ell'}^{\mathrm{H}} \right) + \mathrm{tr}\left(\sum_{\ell_1'=0}^{L_p-1} \sum_{\ell_2'=0}^{L_p-1} \mathbf{F}_{\ell_1'} \mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}[t-\ell_1'] \mathbf{d}_{\mathcal{Q}}^{\mathrm{UL,H}}[t-\ell_2'] \right] \mathbf{F}_{\ell_2'}^{\mathrm{H}} \right) \\
&- \alpha_q \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1} \sum_{\ell=0}^{L-1} \mathbf{F}_{\ell'} \mathbf{H}_{\ell}^{\mathrm{H}} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}} \mathbf{T}\, \mathrm{E}\left[\mathbf{s}^{\mathrm{UL}}[t-\ell-\ell'] \mathbf{s}^{\mathrm{UL,H}}[t-\tau] \right] \right) \\
&- \alpha_q \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1} \sum_{\ell=0}^{L-1} \mathrm{E}\left[\mathbf{s}^{\mathrm{UL}}[t-\tau] \mathbf{s}^{\mathrm{UL,H}}[t-\ell-\ell'] \right] \mathbf{T} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2} \mathbf{H}_{\ell} \mathbf{F}_{\ell'}^{\mathrm{H}} \right) \\
&+ \mathrm{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}} \right).
\end{aligned} \tag{6.42}$$

Since it holds that $\mathrm{E}\left[\mathbf{s}[t_1]\mathbf{s}^{\mathrm{H}}[t_2]\right] = \mathbf{0}_{M,M}$, if $t_1 \neq t_2$, the uplink MSE expression recalculates to

$$\begin{aligned}
\mathrm{MSE}^{\mathrm{UL}} &= \alpha_q^2 \,\mathrm{tr}\left(\sum_{\ell_1'=0}^{L_p-1} \sum_{\ell_2'=0}^{L_p-1} \sum_{\Delta\ell=\ell_1'}^{L-1+\ell_1'} \mathbf{F}_{\ell_1'} \mathbf{H}_{\Delta\ell-\ell_1'}^{\mathrm{H}} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}} \mathbf{T} \mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}} \mathbf{T} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2} \mathbf{H}_{\Delta\ell-\ell_2'} \mathbf{F}_{\ell_2'}^{\mathrm{H}} \right) \\
&+ \alpha_q^2 \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1} \mathbf{F}_{\ell'} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{\mathrm{UL}} \mathbf{F}_{\ell'}^{\mathrm{H}} \right) + \mathrm{tr}\left(\sum_{\ell_1'=0}^{L_p-1} \sum_{\ell_2'=0}^{L_p-1} \mathbf{F}_{\ell_1'} \mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}[t-\ell_1'] \mathbf{d}_{\mathcal{Q}}^{\mathrm{UL,H}}[t-\ell_2'] \right] \mathbf{F}_{\ell_2'}^{\mathrm{H}} \right) \\
&- \alpha_q \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1} \mathbf{F}_{\ell'} \mathbf{H}_{\tau-\ell'}^{\mathrm{H}} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}} \mathbf{T} \mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}} \right) - \alpha_q \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1} \mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}} \mathbf{T} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2} \mathbf{H}_{\tau-\ell'} \mathbf{F}_{\ell'}^{\mathrm{H}} \right) \\
&+ \mathrm{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}} \right).
\end{aligned} \tag{6.43}$$

For comparison, we need to express the downlink MSE. Therefore, we state the downlink input-output relationship as

$$\mathbf{u}[t] = \sum_{\ell=0}^{L-1} \mathbf{G} \mathbf{H}_{\ell} \left(\alpha_q \sum_{\ell'=0}^{L_p-1} \mathbf{P}_{\ell'} \mathbf{s}[t-\ell-\ell'] + \mathbf{d}_{\mathcal{Q}}[t-\ell] \right) + \mathbf{G}\boldsymbol{\eta}[t]. \tag{6.44}$$

Thus, the downlink MSE expression reads as

$$\begin{aligned}\mathrm{MSE}^{\mathrm{DL}} &= \mathrm{E}\left[\|\hat{\mathbf{s}}[t]-\mathbf{s}[t-\tau]\|_2^2\right] \\ &= \alpha_q^2 \,\mathrm{tr}\left(\sum_{\ell_1'=0}^{L_p-1}\sum_{\ell_2'=0}^{L_p-1}\sum_{\Delta\ell=\ell_1'}^{L-1+\ell_1'} \mathbf{G}\mathbf{H}_{\Delta\ell-\ell_1'}\mathbf{P}_{\ell_1'}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{\ell_2'}^{\mathrm{H}}\mathbf{H}_{\Delta\ell-\ell_2'}^{\mathrm{H}}\mathbf{G}\right) \\ &+ \mathrm{tr}\left(\sum_{\ell_1=0}^{L-1}\sum_{\ell_2=0}^{L-1}\mathbf{G}\mathbf{H}_{\ell_1}\,\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}[t-\ell_1]\mathbf{d}_{\mathcal{Q}}^{\mathrm{H}}[t-\ell_2]\right]\mathbf{H}_{\ell_2}^{\mathrm{H}}\mathbf{G}\right) + \mathrm{tr}\left(\mathbf{G}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\mathbf{G}\right) \\ &- \alpha_q \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{G}\mathbf{H}_{\tau-\ell'}\mathbf{P}_{\ell'}\mathbf{C}_{\mathbf{ss}}\right) - \alpha_q \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{\ell'}^{\mathrm{H}}\mathbf{H}_{\tau-\ell'}^{\mathrm{H}}\mathbf{G}\right) + \mathrm{tr}\left(\mathbf{C}_{\mathbf{ss}}\right). \qquad (6.45)\end{aligned}$$

Section 5.4.1 proves that for flat-fading channels there is no MSE duality between the downlink system with a CEQ at the transmitter and the uplink system with a CEQ at the receiver. According to Section 5.4.3, there exists only an approximate dualtiy, when the LCA instead of Price's theorem is applied to compute the MSE expressions in the downlink and uplink scenarios. In Section 5.4.2, we modified the covariance matrix $\mathbf{C}_{\mathbf{d}_{\mathcal{Q}}\mathbf{d}_{\mathcal{Q}}}^{\mathrm{UL}}$ to achieve the exact MSE duality. Analogously, we draw the same conclusions for frequency-selective channels. To this end, we have to find the matching parts and the non-matching parts between the downlink and uplink MSE expressions.

First, we split the terms $\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}[t-\ell_1']\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL,H}}[t-\ell_2']\right]$ and $\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}[t-\ell_1']\mathbf{d}_{\mathcal{Q}}^{\mathrm{H}}[t-\ell_2']\right]$ in the uplink and the downlink systems into three terms

- $\mathrm{diag}\left(\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}[t-\ell_1']\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL,H}}[t-\ell_2']\right]\right)$ and $\mathrm{diag}\left(\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}[t-\ell_1']\mathbf{d}_{\mathcal{Q}}^{\mathrm{H}}[t-\ell_2']\right]\right),\ \ell_1'=\ell_2',$
- $\mathrm{nondiag}\left(\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}[t-\ell_1']\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL,H}}[t-\ell_2']\right]\right)$ and $\mathrm{nondiag}\left(\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}[t-\ell_1']\mathbf{d}_{\mathcal{Q}}^{\mathrm{H}}[t-\ell_2']\right]\right),\ \ell_1'=\ell_2',$

and

- $\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}[t-\ell_1']\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL,H}}[t-\ell_2']\right]$ and $\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}[t-\ell_1']\mathbf{d}_{\mathcal{Q}}^{\mathrm{H}}[t-\ell_2']\right],\ \ell_1'\neq\ell_2'.$

The above covariance matrices can be computed exactly with Price's theorem according to (3.87). Note that the LCA makes the second and third terms vanish.

Second, we compute the covariance matrices stated above according to Price's theorem and plug the resulting expressions in (6.43) and (6.45). Thus, the uplink and downlink MSE

expressions calculate to

$$\begin{aligned}
\mathrm{MSE}^{\mathrm{UL}} &= \alpha_q \,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1}\sum_{\ell=0}^{L-1} \mathbf{F}_{\ell'}\,\mathrm{diag}\left(\mathbf{H}_{\ell}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2}\mathbf{H}_{\ell}\right)\mathbf{F}_{\ell'}^{\mathrm{H}}\right) \\
&\quad + \alpha_q^2\,\mathrm{tr}\left(\sum_{\ell_1'=0}^{L_p-1}\sum_{\substack{\ell_2'=0\\ \ell_2'\neq\ell_1'}}^{L_p-1}\sum_{\Delta\ell=\ell_1'}^{L-1+\ell_1'} \mathbf{F}_{\ell_1'}\mathbf{H}_{\Delta\ell-\ell_1'}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2}\mathbf{H}_{\Delta\ell-\ell_2'}\mathbf{F}_{\ell_2'}^{\mathrm{H}}\right) \\
&\quad + \frac{Q}{\pi}\sin^2(\psi)\,\alpha_q\,\mathrm{tr}\Bigg(\sum_{\ell'=0}^{L_p-1}\mathbf{F}_{\ell'}\sum_{\Delta k=0}^{Q/2-1}\mathrm{e}^{\mathrm{j}\,2\Delta k\psi}\,\mathrm{diag}(\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}})^{1/2} \\
&\qquad\qquad \arcsin\left(\Re\left\{\mathrm{nondiag}\left(\mathbf{R}_{\mathbf{xx}}^{\mathrm{UL}}\right)\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\right\}\right)\mathrm{diag}(\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}})^{1/2}\mathbf{F}_{\ell'}^{\mathrm{H}}\Bigg) \\
&\quad + \mathrm{tr}\left(\sum_{\ell_1'=0}^{L_p-1}\sum_{\substack{\ell_2'=0\\ \ell_2'\neq\ell_1'}}^{L_p-1}\mathbf{F}_{\ell_1'}\,\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}[t-\ell_1']\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL},\mathrm{H}}[t-\ell_2']\right]\mathbf{F}_{\ell_2'}^{\mathrm{H}}\right) \\
&\quad - \alpha_q\,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{F}_{\ell'}\mathbf{H}_{\tau-\ell'}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2,\mathrm{H}}\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\right) - \alpha_q\,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}\mathbf{C}_{\boldsymbol{\eta\eta}}^{-1/2}\mathbf{H}_{\tau-\ell'}\mathbf{F}_{\ell'}^{\mathrm{H}}\right) \\
&\quad + \alpha_q\,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{F}_{\ell'}\mathbf{C}_{\boldsymbol{\eta\eta}}^{\mathrm{UL}}\mathbf{F}_{\ell'}^{\mathrm{H}}\right) + \mathrm{tr}\left(\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\right), \qquad (6.46)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{MSE}^{\mathrm{DL}} &= \alpha_q\,\mathrm{tr}\left(\sum_{\ell=0}^{L-1}\sum_{\ell'=0}^{L_p-1}\mathbf{G}\mathbf{H}_{\ell}\,\mathrm{diag}\left(\mathbf{P}_{\ell'}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{\ell'}^{\mathrm{H}}\right)\mathbf{H}_{\ell}^{\mathrm{H}}\mathbf{G}\right) \\
&\quad + \alpha_q^2\,\mathrm{tr}\left(\sum_{\ell_1'=0}^{L_p-1}\sum_{\substack{\ell_2'=0\\ \ell_2'\neq\ell_1'}}^{L_p-1}\sum_{\Delta\ell=\ell_1'}^{L-1+\ell_1'}\mathbf{G}\mathbf{H}_{\Delta\ell-\ell_1'}\mathbf{P}_{\ell_1'}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{\ell_2'}^{\mathrm{H}}\mathbf{H}_{\Delta\ell-\ell_2'}^{\mathrm{H}}\mathbf{G}\right) \\
&\quad + \frac{Q}{\pi}\sin^2(\psi)\,\alpha_q\,\mathrm{tr}\Bigg(\sum_{\ell=0}^{L-1}\mathbf{G}\mathbf{H}_{\ell}\sum_{\Delta k=0}^{Q/2-1}\mathrm{e}^{\mathrm{j}\,2\Delta k\psi}\,\mathrm{diag}(\mathbf{C}_{\mathbf{xx}})^{1/2} \\
&\qquad\qquad \arcsin\left(\Re\left\{\mathrm{nondiag}\left(\mathbf{R}_{\mathbf{xx}}\right)\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\right\}\right)\mathrm{diag}(\mathbf{C}_{\mathbf{xx}})^{1/2}\mathbf{H}_{\ell}^{\mathrm{H}}\mathbf{G}\Bigg) \\
&\quad + \mathrm{tr}\left(\sum_{\ell_1=0}^{L-1}\sum_{\ell_2=0,\ell_2\neq\ell_1}^{L-1}\mathbf{G}\mathbf{H}_{\ell_1}\,\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}[t-\ell_1]\mathbf{d}_{\mathcal{Q}}^{\mathrm{H}}[t-\ell_2]\right]\mathbf{H}_{\ell_2}^{\mathrm{H}}\mathbf{G}\right) \\
&\quad - \alpha_q\,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{G}\mathbf{H}_{\tau-\ell'}\mathbf{P}_{\ell'}\mathbf{C}_{\mathbf{ss}}\right) - \alpha_q\,\mathrm{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{\ell'}^{\mathrm{H}}\mathbf{H}_{\tau-\ell'}^{\mathrm{H}}\mathbf{G}\right) \\
&\quad + \mathrm{tr}\left(\mathbf{G}\mathbf{C}_{\boldsymbol{\eta\eta}}\mathbf{G}\right) + \mathrm{tr}\left(\mathbf{C}_{\mathbf{ss}}\right). \qquad (6.47)
\end{aligned}$$

By applying the identities in (4.39), (4.40), (4.41) and (4.42), the uplink MSE expression recalculates to

$$\begin{aligned}
\mathrm{MSE}^{\mathrm{UL}} &= \alpha_q \operatorname{tr}\left(\sum_{\ell=0}^{L-1}\sum_{\ell'=0}^{L_p-1} \mathbf{G}\mathbf{H}_\ell \operatorname{diag}\left(\mathbf{P}_{\ell'}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{\ell'}^{\mathrm{H}}\right)\mathbf{H}_\ell^{\mathrm{H}}\mathbf{G}\right) \\
&+ \alpha_q^2 \operatorname{tr}\left(\sum_{\ell_1'=0}^{L_p-1}\sum_{\substack{\ell_2'=0\\ \ell_2'\neq\ell_1'}}^{L_p-1}\sum_{\Delta\ell=\ell_1'}^{L-1+\ell_1'} \mathbf{G}\mathbf{H}_{\Delta\ell-\ell_1'}\mathbf{P}_{\ell_1'}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{\ell_2'}^{\mathrm{H}}\mathbf{H}_{\Delta\ell-\ell_2'}^{\mathrm{H}}\mathbf{G}\right) \\
&+ \frac{Q}{\pi}\sin^2(\psi)\,\frac{\alpha_q}{\beta^2}\operatorname{tr}\Bigg(\sum_{\ell'=0}^{L_p-1}\mathbf{P}_{\ell'}^{\mathrm{H}}\sum_{\Delta k=0}^{Q/2-1}\mathrm{e}^{\mathrm{j}\,2\Delta k\psi}\operatorname{diag}(\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}})^{1/2} \\
&\qquad\qquad \arcsin\left(\Re\left\{\operatorname{nondiag}\left(\mathbf{R}_{\mathbf{xx}}^{\mathrm{UL}}\right)\mathrm{e}^{-\mathrm{j}\,2\Delta k\psi}\right\}\right)\operatorname{diag}(\mathbf{C}_{\mathbf{xx}}^{\mathrm{UL}})^{1/2}\mathbf{P}_{\ell'}\Bigg) \\
&+ \frac{1}{\beta^2}\operatorname{tr}\left(\sum_{\ell_1'=0}^{L_p-1}\sum_{\substack{\ell_2'=0\\ \ell_2'\neq\ell_1'}}^{L_p-1}\mathbf{P}_{\ell_1'}^{\mathrm{H}}\,\mathrm{E}\left[\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL}}[t-\ell_1']\mathbf{d}_{\mathcal{Q}}^{\mathrm{UL,H}}[t-\ell_2']\right]\mathbf{P}_{\ell_2'}\right) \\
&- \alpha_q \operatorname{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{G}\mathbf{H}_{\tau-\ell'}\mathbf{P}_{\ell'}\mathbf{C}_{\mathbf{ss}}\right) - \alpha_q \operatorname{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{\ell'}^{\mathrm{H}}\mathbf{H}_{\tau-\ell'}^{\mathrm{H}}\mathbf{G}\right) \\
&+ \frac{\alpha_q}{\beta^2\sigma_s^2}\operatorname{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{P}_{\ell'}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{\ell'}^{\mathrm{H}}\right) + \operatorname{tr}\left(\mathbf{C}_{\mathbf{ss}}\right). \qquad (6.48)
\end{aligned}$$

By comparing (6.48) to (6.47) , we can see that all terms match except of the third, fourth and seventh terms. The seventh terms can be matched by choosing

$$\begin{aligned}
\beta &= \sqrt{\frac{\alpha_q \operatorname{tr}\left(\sum_{\ell'=0}^{L_p-1}\mathbf{P}_{\ell'}\mathbf{C}_{\mathbf{ss}}\mathbf{P}_{\ell'}^{\mathrm{H}}\right)}{\sigma_s^2\operatorname{tr}\left(\mathbf{G}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\mathbf{G}\right)}} \\
&\overset{(4.43)}{=} \sqrt{\frac{P_{\mathrm{tx}}}{\operatorname{tr}\left(\mathbf{G}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\mathbf{G}\right)}}. \qquad (6.49)
\end{aligned}$$

The obtained expression for β is the same as for flat-fading channels. In analogy to Section 5.4.2, we assume that $\mathbf{G}$ is a scaled identity, i.e. $\mathbf{G} = g\mathbf{I}_M$ and thus from (4.40) and (6.49) we get

$$\mathbf{T} = \sqrt{\frac{P_{\mathrm{tx}}}{\sigma_s^2\operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}\right)}}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{1/2,\mathrm{H}}. \qquad (6.50)$$

Hence, the transmit power constraint in the uplink system holds true; that is

$$\operatorname{tr}\left(\mathbf{T}\mathbf{C}_{\mathbf{ss}}^{\mathrm{UL}}\mathbf{T}^{\mathrm{H}}\right) = P_{\mathrm{tx}}. \qquad (6.51)$$

For exact MSE duality, the third and fourth terms in the uplink and downlink MSE expressions must additionally match.

6.4.2 Approximate Dual Problem

By using the LCA instead of Price's theorem, the third and fourth terms in the uplink and downlink MSE expressions in (6.48) and (6.47), respectively, vanish. Thus, the MSE expressions match. In other words, we can achieve approximate duality by applying the LCA for the computation of the MSE.

After considering the approximate dual uplink system, the optimal MMSE equalizer $\mathbf{F}[t]$ has to be designed. To this end, we state the compact input-output relationship as

$$\mathbf{u}^{\mathrm{UL}}[t] = \mathbf{u} = \mathbf{F}\mathbf{t}^{\mathrm{UL}}, \tag{6.52}$$

where

$$\begin{aligned}
\mathbf{F} &= \begin{bmatrix} \mathbf{F}_0 & \mathbf{F}_1 & \cdots & \mathbf{F}_{L_p-1} \end{bmatrix}, && (6.53)\\
\mathbf{t}^{\mathrm{UL}} &= \mathcal{Q}_{\mathrm{CE}}\left(\mathbf{x}^{\mathrm{UL}}\right), && (6.54)\\
\mathbf{x}^{\mathrm{UL}} &= \begin{bmatrix} \mathbf{x}[t]^{\mathrm{UL,T}} & \mathbf{x}[t-1]^{\mathrm{UL,T}} & \cdots & \mathbf{x}[t-L_p+1]^{\mathrm{UL,T}} \end{bmatrix}^{\mathrm{T}} \\
&= \mathbf{H}^{\mathrm{UL}}\mathbf{s}^{\mathrm{UL}}, && (6.55)\\
\mathbf{H}^{\mathrm{UL}} &= \sum_{\ell=0}^{L-1} \mathbf{S}(\ell, L, L_p-1) \otimes \left(\mathbf{H}_\ell^{\mathrm{H}} \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2,\mathrm{H}} \mathbf{T}\right), && (6.56)
\end{aligned}$$

and

$$\mathbf{s}^{\mathrm{UL}} = \begin{bmatrix} \mathbf{s}[t]^{\mathrm{UL,T}} & \mathbf{s}[t-1]^{\mathrm{UL,T}} & \cdots & \mathbf{s}[t-L_p-L+2]^{\mathrm{UL,T}} \end{bmatrix}^{\mathrm{T}}. \tag{6.57}$$

The uplink MSE is expressed in terms of the defined variables as

$$\begin{aligned}
\mathrm{MSE}^{\mathrm{UL}} &= \mathrm{E}\left[\left\|\mathbf{u}^{\mathrm{UL}} - \left(\mathbf{e}_{\tau+1}^{\mathrm{T}} \otimes \mathbf{I}_M\right)\mathbf{s}^{\mathrm{UL}}\right\|_2^2\right] \\
&= \mathrm{tr}\left(\mathbf{F}\mathbf{C}_{\mathbf{tt}}\mathbf{F}^{\mathrm{H}}\right) - \mathrm{tr}\left(\mathbf{F}\mathbf{C}_{\mathbf{ts}}\left(\mathbf{e}_{\tau+1} \otimes \mathbf{I}_M\right)\right) - \mathrm{tr}\left(\left(\mathbf{e}_{\tau+1}^{\mathrm{T}} \otimes \mathbf{I}_M\right)\mathbf{C}_{\mathbf{st}}\mathbf{F}^{\mathrm{H}}\right) \\
&\quad + \mathrm{tr}\left(\left(\mathbf{e}_{\tau+1}^{\mathrm{T}} \otimes \mathbf{I}_M\right)\mathbf{C}_{\mathbf{ss}}\left(\mathbf{e}_{\tau+1} \otimes \mathbf{I}_M\right)\right). \qquad (6.58)
\end{aligned}$$

Hence, the optimal MMSE dual filter is then given by

$$\mathbf{F} = \left(\mathbf{e}_{\tau+1}^{\mathrm{T}} \otimes \mathbf{I}_M\right)\mathbf{C}_{\mathbf{st}}^{\mathrm{UL}}\mathbf{C}_{\mathbf{tt}}^{\mathrm{UL},-1}. \tag{6.59}$$

First, we apply Price's theorem to compute $\mathbf{C}_{\mathbf{tt}}^{\mathrm{UL}}$ and $\mathbf{C}_{\mathbf{st}}^{\mathrm{UL}}$ and hence obtain the filter $\mathbf{F}$. The matrix $\mathbf{T}$ is given in (6.50). Second, the optimal filters in the primal domain can be computed using the identities (4.39) and (4.40). The corresponding scaling factor β in (4.44) can be expressed as

$$\beta = \sqrt{\frac{P_{\mathrm{tx}}}{\alpha_q \,\mathrm{tr}\left(\mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathbf{ss}}\mathbf{F}\right)}}. \tag{6.60}$$

6.5 Simulation Results

In this section, we compare the performance of the different linear precoding techniques that were introduced above with the ideal WF precoder while assuming full Channel State Information (CSI). We denote the precoders by

- FIR-WF: the ideal FIR-WF, [70], where no CEQ is applied in the system.
- FIR-QWF: the FIR-Quantized WF that was introduced in Section 6.3.2.
- FIR-QWF-Apprx$_{\text{Dual}}$: the approximate dual Quantized WF that applies unequal power allocation at the antennas. The resulting receive filter $\mathbf{G}$ is a scaled identity. This precoder is derived in Section 6.4.2.
- FIR-QWF-Price$_{\text{eqPA}}$: the FIR-Quantized WF based on Price's theorem that applies equal power allocation at the antennas. This precoder is detailed in Algorithm 4. The start value is the FIR-WF precoder that is projected to fulfill the equal transmit power constraint.
- FIR-QWF-Price$_{\text{neqPA}}$: the FIR-Quantized WF based on Price's theorem that applies unequal power allocation at the antennas. This precoder is detailed in Algorithm 5. The start value is the FIR-WF precoder.

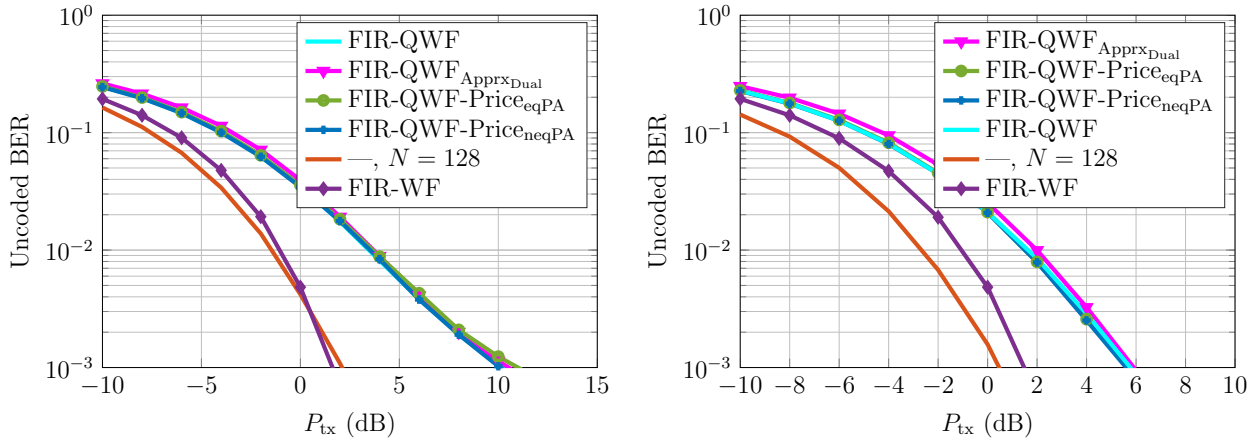
For all precoders, it holds that $L_p = L$. To this end, we assume a BS with $N = 64$ or $N = 128$ antennas serving $M = 8$ single-antenna users with 4-PSK or 16-QAM signals, respectively. The numerical results are obtained with Monte Carlo simulations of 100 independent frequency-selective channel realizations of $L = 3$ from the i.i.d. channel model and the mmW sparse channel model described in Section 4.6. For the mmWsparse channel, we assume that $N_{\text{cl}} = 2$ and $N_{\text{ray}} = 10$. The AWGN is also i.i.d. with variance one at each antenna. The performance metric is the uncoded BER averaged over the single-antenna users. For the blind estimation of the coefficients g_m we use a block length of $T = 128$. The numerical results are plotted in Fig. 6.1, Fig. 6.2, Fig. 6.3 and Fig. 6.4.

First, it can be deduced that all proposed linear precoders perform almost the same for i.i.d. channels. The loss compared to the ideal WF reduces with increased quantization resolution.

Second, the proposed precoders perform slightly different for mmW sparse channels. We can see that the QWF-Price$_{\text{neqPA}}$ and the QWF-Price$_{\text{eqPA}}$ precoders perform the best and have almost the same behavior. In the presence of the ISI, there is almost no much difference between equal and unequal power allocation at the BS antennas.

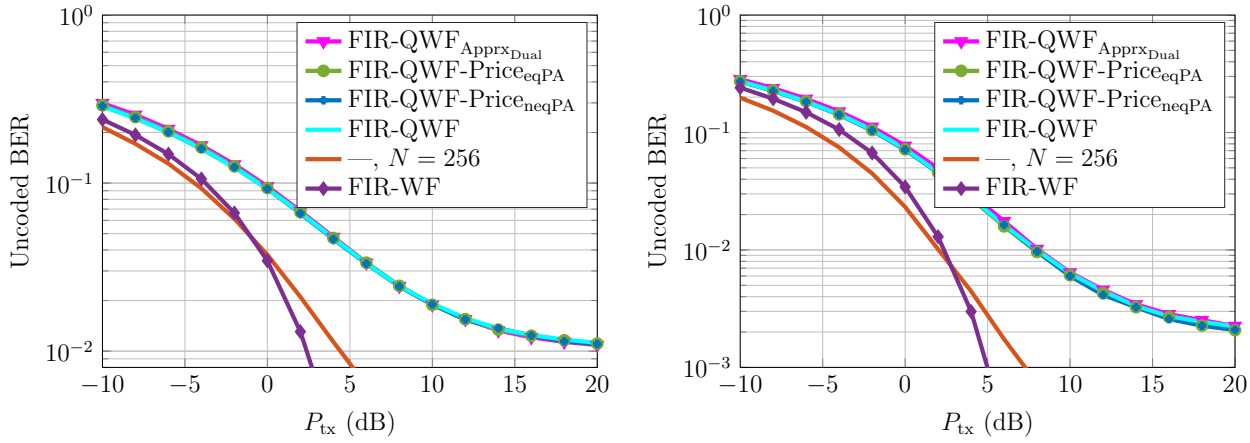
Third, as observed in the case of flat-fading channels, a QCE system requires at most the double number of antennas to obtain the same BER performance as in the ideal unquantized case.

In summary, the performance improvement between the proposed linear precoding algorithms that take into account the quantization distortions is too small and the performance gap for the same system settings to the ideal linear system is still large. Therefore, we decide to study non-linear precoding techniques with the hope that they offer a much better performance.



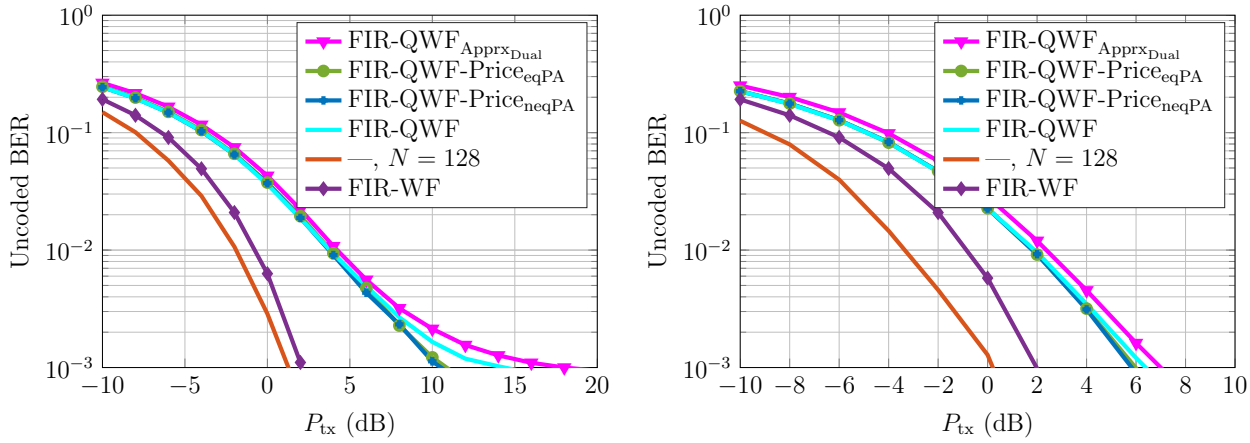
(a) $Q = 4$ (b) $Q = 8$

Fig. 6.1: Comparison between different linear precoders for 4-PSK signaling: $N = 64$, $M = 8$ with the i.i.d. channel with $L = 3$.



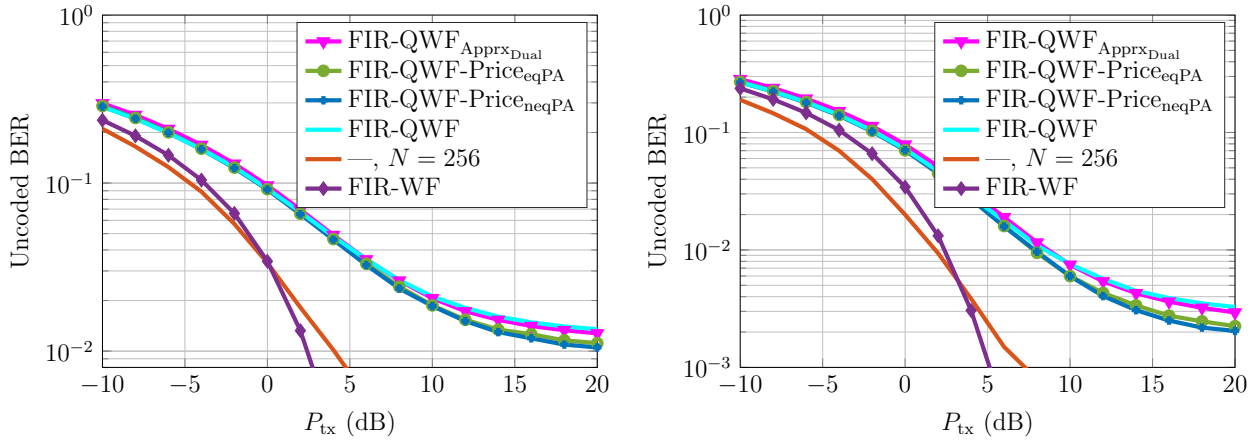
(a) $Q = 4$ (b) $Q = 8$

Fig. 6.2: Comparison between different linear precoders for 16-QAM signaling: $N = 128$, $M = 8$ with the i.i.d. channel with $L = 3$.



(a) $Q = 4$ (b) $Q = 8$

Fig. 6.3: Comparison between different linear precoders for 4-PSK signaling: $N = 64$, $M = 8$ with the mmW sparse channel with $L = 3$.



(a) $Q = 4$ (b) $Q = 8$

Fig. 6.4: Comparison between different linear precoders for 16-QAM signaling: $N = 128$, $M = 8$ with the mmW sparse channel with $L = 3$.

Part II

Non-linear Transmit Signal Processing

7. Flat-Fading Channels

As explained in Section 4.4.2, the signal vector $\mathbf{x}$ is designed rather than the precoder matrix $\mathbf{P}$. For every given input signal $\mathbf{s}$ and for each channel realization $\mathbf{H}$, the precoding task is to find

$$\mathbf{x} = \mathcal{P}(\mathbf{s}, \mathbf{H}). \tag{7.1}$$

The task consists in designing the transmit vector $\mathbf{x}$ such that $\hat{\mathbf{s}} = \mathbf{s}$ holds true with high probability to reduce the detection error probability. The symbol-wise precoder aims to mitigate all sources of distortion

- the quantization distortions
- the channel distortions and the MUI, and
- the AWGN.

Our goal is to develop a problem formulation that jointly minimizes all three distortion sources.

7.1 Transmit Power Constraint

The transmit power is allocated equally among the antennas. From (4.23), it follows that

$$\boldsymbol{\Xi} = \sqrt{\frac{P_{\text{tx}}}{N}}\mathbf{I}_N. \tag{7.2}$$

Thus, the set $\mathbb{T}_n$, $n = 1, \cdots, N$, reads as

$$\mathbb{T}_n = \mathbb{T} = \left\{ \sqrt{\frac{P_{\text{tx}}}{N}} \exp\left(\mathrm{j}\,(2i-1)\,\frac{\pi}{Q} \right) : i = 1, \cdots, Q \right\}. \tag{7.3}$$

7.2 Optimization Problem

To formulate the optimization problem, we have to consider all distortion sources and find a way to mitigate them one by one.

7.2.1 Mitigation of the Quantization Distortions

First, it is obvious that the quantization distortions can be omitted if we design the precoded vector $\mathbf{x}$ such that

$$\mathbf{x} \in \mathbb{T}^N, \tag{7.4}$$

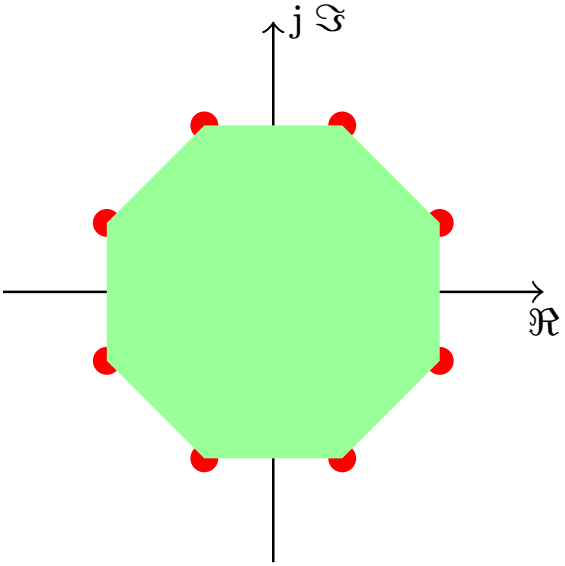


Fig. 7.1: Illustration of the relaxed polygon constraint for Q=8, © 2018 IEEE.

i.e. $\mathbb{X} = \mathbb{T}$. This would ensure that the quantizer $\mathcal{Q}_{\text{CE}}(\bullet)$ produces no distortion, and we would have an undistorted transmit signal $\mathbf{t} = \mathbf{x}$. However, the QCE constraint in (7.4) would lead to a discrete optimization problem due to the discrete nature of the set $\mathbb{T}$. To avoid this problem, we relax the discrete set $\mathbb{T}$ to the convex set $\mathbb{X}$ that represents the polygon built by the Q scaled PSK points of the set $\mathbb{T}$. Thus, the QCE constraint is relaxed to a convex constraint that we call the relaxed polygon constraint. Fig. 7.1 illustrates the relaxed polygon constraint for the case of $Q = 8$. Instead of designing $\mathbf{x} \in \mathbb{T}^N$ to completely eliminate the quantization distortions, we design $\mathbf{x} \in \mathbb{X}^N$ to minimize them.

The set $\mathbb{X}$ can be mathematically described by a set of linear inequalities. For q-bit polar DACs, i.e., where the transmitted data are constrained to be Q scaled PSK symbols, the polygon can be constructed by the intersection of $Q/4$ squares that have an angular shift of $2\pi/Q$. To this end, we define the rotation matrix $\mathbf{R}_i$ of angle $\beta_i = \frac{2\pi}{Q}(i-1)$ as

$$\mathbf{R}_i = \begin{bmatrix} \cos\beta_i & \sin\beta_i \\ -\sin\beta_i & \cos\beta_i \end{bmatrix} \otimes \mathbf{I}_N, \;\; i = 1, ..., Q/4. \tag{7.5}$$

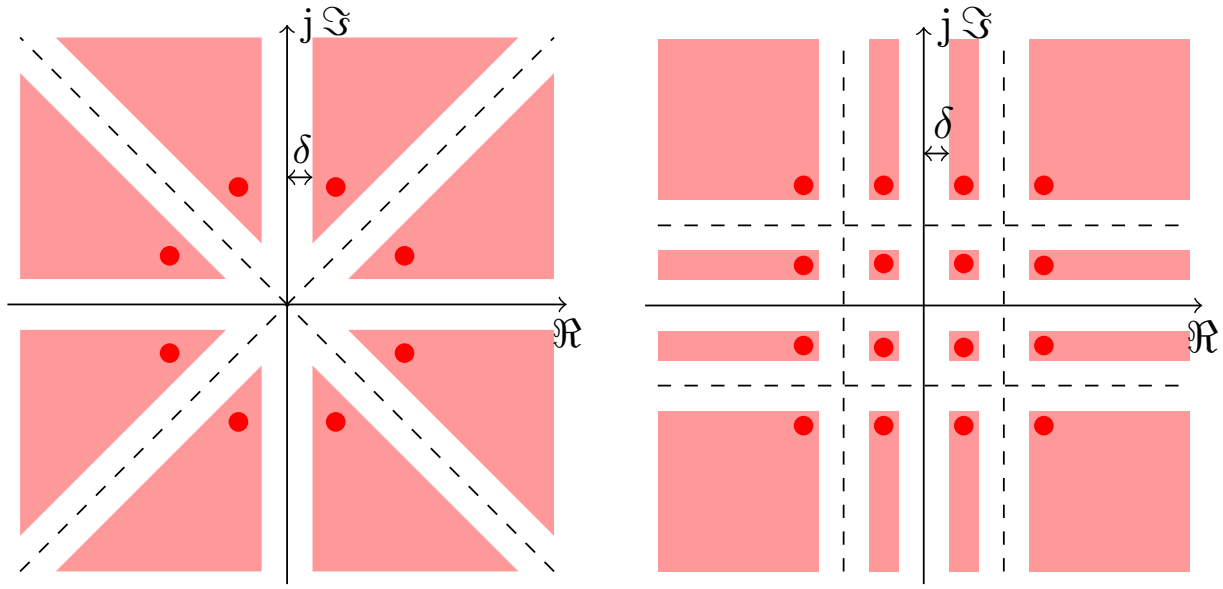
The system of inequalities that considers the feasible set, i.e. the relaxed polygon constraint, and hence relaxes the constraint in (7.4) is given by

$$\left[\mathbf{R}_1^{\mathrm{T}} \; -\mathbf{R}_1^{\mathrm{T}} \cdots \mathbf{R}_{\frac{Q}{4}}^{\mathrm{T}} \; -\mathbf{R}_{\frac{Q}{4}}^{\mathrm{T}}\right]^{\mathrm{T}} \bar{\mathbf{x}} \leq \sqrt{\frac{P_{\text{tx}}}{N}} \cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{NQ}, \tag{7.6}$$

where $\bar{\mathbf{x}} = \left[\Re\{\mathbf{x}\}^{\mathrm{T}} \;\; \Im\{\mathbf{x}\}^{\mathrm{T}}\right]^{\mathrm{T}}$. Since $\mathbf{R}_1 = \mathbf{I}_{2N}$, the first $4N$ inequalities in (7.6) define the bounds for $\bar{\mathbf{x}}$. Hence, the relaxed polygon constraint, i.e. $\mathbf{x} \in \mathbb{X}^N$, is equivalent to

$$-\sqrt{\frac{P_{\text{tx}}}{N}} \cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{2N} \leq \bar{\mathbf{x}} \leq \sqrt{\frac{P_{\text{tx}}}{N}} \cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{2N},$$

$$\text{and} \quad \underbrace{\left[\mathbf{R}_2^{\mathrm{T}} \; -\mathbf{R}_2^{\mathrm{T}} \cdots \mathbf{R}_{\frac{Q}{4}}^{\mathrm{T}} \; -\mathbf{R}_{\frac{Q}{4}}^{\mathrm{T}}\right]^{\mathrm{T}}}_{=\mathbf{E}} \bar{\mathbf{x}} \leq \sqrt{\frac{P_{\text{tx}}}{N}} \cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{N(Q-4)}. \tag{7.7}$$



(a) 8-PSK. (b) 16-QAM

Fig. 7.2: Decision regions and SRs (in red) for different constellations, © 2018 IEEE.

This reformulation leads to significant computational savings since the final optimization problem will be written as a linear program with bounded variables. As discussed in Section 7.5, it is beneficial in terms of computational complexity to have fewer inequalities.

7.2.2 Mitigation of the Channel Distortions and the MUI

Second, to minimize the channel distortions and the MUI, we look deeper into the properties of the constellations. As illustrated in Fig. 7.2a and Fig. 7.2b, each constellation is defined by thresholds that separate the distinct decision regions of the constellation points. In total, we have as many contiguous decision regions as constellation points. For each outer constellation point, the decision region has at least one infinite boundary. Therefore, we make use of the idea of constructive interference optimization [42, 43]. When the downlink channel and all users' data are known at the transmitter, the instantaneous constructive MUI can be exploited to move the received signals further from the decision thresholds [43]. In contrast to this, conventional precoding methods, e.g. MMSE, aim at minimizing the total MUI such that the received signals lie as close as possible to the nominal constellation points. Constructive interference optimization exploits the larger symbol decision regions and thus leads to a more relaxed optimization.

Each constellation symbol lies within a Symbol Region (SR) that is a downscaled version of the decision region. In contrast to the decision region, the SR has a safety margin denoted by δ that separates it from the decision thresholds. When each entry of the noiseless received signal vector $\mathbf{y}$ belongs to the correct SR and thus to the correct decision region, the channel distortions and the MUI are mitigated.

7.2.3 Mitigation of the AWGN

Finally, the safety margin δ has to be large enough such that, when $\mathbf{y}$ is perturbed by the AWGN, the received signals do not jump to unintended neighboring decision regions.

7.2.4 General Problem Formulation

In summary, the problem formulation has to take into account the relaxed QCE constraint in (7.7), the SR constraint for each received signal and maximizing the safety margin δ. Thus, the optimization problem for the symbol-wise precoder, which we call the MSM precoder, can be written in general as follows

$$\max_{\mathbf{x}} \delta \tag{7.8}$$

$$\text{s.t. } y'_m \in \text{SR}_m, \forall m \tag{7.9}$$

$$\text{and } \mathbf{x} \in \mathbb{X}^N, \tag{7.10}$$

where

$$\mathbf{y}' = \mathbf{H}\mathbf{x} \tag{7.11}$$

represents the relaxation of $\mathbf{y}$ due to the relaxation of (7.4) to (7.7). Exact expressions for the safety margin δ and the SRs as a function of $\mathbf{x}$ are provided in Section 7.3 and Section 7.4 for PSK and QAM signaling, respectively.

This problem formulation depends on the input symbol vector $\mathbf{s}$, which determines the intended SR for each received signal and on the channel $\mathbf{H}$. The optimization is run for one specific value of the transmit power, i.e. $P_{\text{tx}} = N$. Since the optimization variables depend linearly on the square-root of the transmit power $\sqrt{P_{\text{tx}}}$, it is sufficient to solve the optimization problem for one transmit power value. Once the optimal relaxed solution $\mathbf{x}$ is found, the transmit vector $\mathbf{t}$ is obtained as stated in (4.1).

7.3 Problem Formulation for PSK Signaling

7.3.1 SR for PSK Signals

In this section, we assume that the input signals s_m, $m = 1, \cdots, M$, belong to the S-PSK constellation. The set $\mathbb{S}$ in this case is defined in (4.13).

Each SR in the PSK constellation, as shown in Fig. 7.2a, is a circular sector of infinite radius and angle 2θ. To find a mathematical expression for each SR, we rotate the original coordinate system by the phase of the symbol of interest s_m, as illustrated in Fig. 7.3, and introduce a new variable z_m that represents the signal y'_m in the new coordinate system such that

$$\begin{aligned} z_m &= y'_m \frac{s_m^*}{|s_m|}, \ m = 1, \cdots, M \\ &= y'_m s_m^*, \ m = 1, \cdots, M, \end{aligned} \tag{7.12}$$

since PSK signals have unit magnitude. The m-th SR, $m = 1, \cdots, M$, can be hence described by

$$\Re\{z_m\} \geq \delta', \tag{7.13}$$

$$|\Im\{z_m\}| \leq (\Re\{z_m\} - \delta') \tan\theta, \tag{7.14}$$

where $\delta' = \frac{\delta}{\sin\theta}$. Note that the inequality in (7.13) is already fulfilled if the inequality in (7.14) is satisfied. In the vector notation, the vector $\mathbf{z}$ can be expressed as

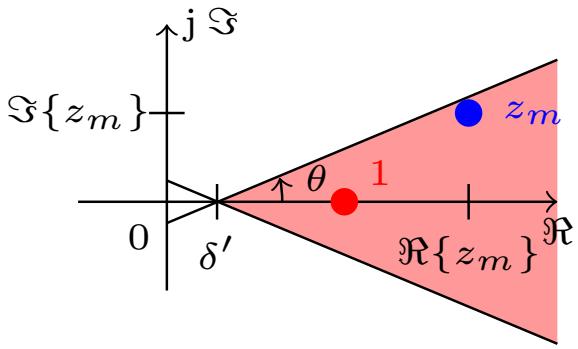


Fig. 7.3: Illustration of the PSK SR in the modified coordinate system.

$$\begin{aligned}\mathbf{z} &= \operatorname{diag}(\mathbf{s}^*)\mathbf{H}\mathbf{x} \\ &= \tilde{\mathbf{H}}\mathbf{x},\end{aligned} \tag{7.15}$$

where

$$\tilde{\mathbf{H}} = \operatorname{diag}(\mathbf{s}^*)\mathbf{H}. \tag{7.16}$$

Hence, all M SRs can be compactly described as

$$|\Im\{\mathbf{z}\}| \leq (\Re\{\mathbf{z}\} - \delta' \mathbf{1}_M)\tan\theta, \tag{7.17}$$

as also given in [43]. When using the following real-valued representation

$$\Re\{\mathbf{z}\} = \underbrace{\begin{bmatrix}\Re\{\tilde{\mathbf{H}}\} & -\Im\{\tilde{\mathbf{H}}\}\end{bmatrix}}_{=\mathbf{A}} \begin{bmatrix}\Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\}\end{bmatrix} = \mathbf{A}\bar{\mathbf{x}}, \tag{7.18}$$

$$\Im\{\mathbf{z}\} = \underbrace{\begin{bmatrix}\Im\{\tilde{\mathbf{H}}\} & \Re\{\tilde{\mathbf{H}}\}\end{bmatrix}}_{=\mathbf{B}} \begin{bmatrix}\Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\}\end{bmatrix} = \mathbf{B}\bar{\mathbf{x}}, \tag{7.19}$$

the constraint in (7.17) can be rewritten as

$$\begin{bmatrix}\mathbf{B} - \tan\theta\mathbf{A} & \frac{1}{\cos\theta}\mathbf{1}_M \\ -\mathbf{B} - \tan\theta\mathbf{A} & \frac{1}{\cos\theta}\mathbf{1}_M\end{bmatrix} \begin{bmatrix}\bar{\mathbf{x}} \\ \delta\end{bmatrix} \leq \mathbf{0}_{2M}. \tag{7.20}$$

7.3.2 Optimization Problem with the Relaxed Polygon Constraint

Finally, the optimization problem for the symbol-wise precoder with PSK signaling is obtained by combining (7.8), (7.20) and (7.7) and and is expressed for the case of $P_{\text{tx}} = N$ as

$$\begin{aligned}\max_{\mathbf{v}} \begin{bmatrix}\mathbf{0}_{2N}^{\mathrm{T}} & 1\end{bmatrix}\mathbf{v} \text{ s.t. } &\begin{bmatrix}\mathbf{B} - \tan\theta\mathbf{A} & \frac{1}{\cos\theta}\mathbf{1}_M \\ -\mathbf{B} - \tan\theta\mathbf{A} & \frac{1}{\cos\theta}\mathbf{1}_M \\ \mathbf{E} & \mathbf{0}_{N(Q-4)}\end{bmatrix}\mathbf{v} \leq \begin{bmatrix}\mathbf{0}_{2M} \\ \cos\left(\frac{\pi}{Q}\right)\mathbf{1}_{N(Q-4)}\end{bmatrix}, \\ \text{and } &\begin{bmatrix}-\cos\left(\frac{\pi}{Q}\right)\mathbf{1}_{2N} \\ 0\end{bmatrix} \leq \mathbf{v} \leq \begin{bmatrix}\cos\left(\frac{\pi}{Q}\right)\mathbf{1}_{2N} \\ \infty\end{bmatrix},\end{aligned} \tag{7.21}$$

where $\mathbf{v}^{\mathrm{T}} = \begin{bmatrix}\bar{\mathbf{x}}^{\mathrm{T}} & \delta\end{bmatrix}$. The resulting optimization problem is a linear programming problem for which there exist very efficient solving methods [71]. In order to solve the problem for different P_{tx} values, it is sufficient to scale the solution of (7.21) by $\frac{P_{\mathrm{tx}}}{N}$ due to the linearity of the problem.

When the optimization terminates, the optimal signal $\mathbf{x} \in \mathbb{X}^N$ is found. The signal $\mathbf{t}$ that goes through the channel is obtained as described in (4.1). In other words, each entry in $\mathbf{x}$ gets mapped to the corresponding CE point depending on the circular sector that it lies in.

7.3.3 Safety Margin for PSK Signals

The safety margin δ in (7.8) can be expressed for the PSK case as

$$\begin{aligned}\delta &= \min_m \delta_m \\ &= \min_m \left(\sin(\theta)\Re\{\mathbf{z}\} - \cos(\theta)\left|\Im\{\mathbf{z}\}\right|\right),\end{aligned} \tag{7.22}$$

where the operator $|\bullet|$ is applied element-wise to the entries of $\Im\{\mathbf{z}\}$. Note that an equivalent objective function was introduced in [22] in the context of continuous-phase CE precoding for PSK signaling. In [22], the strict CE constraint is relaxed to the convex unit circle, whereas in our work the QCE constraint is relaxed to the linear polygon constraint. Consequently, due to the linear objective function and the linear constraints, our optimization problem can be formulated as a linear programming problem unlike [22].

7.3.4 Interpretation of the Safety Margin δ for PSK Signals

The safety margin δ is a parameter that affects the receiver Signal-to-Noise Ratio (SNR) and the SER. These relationships will be given for the relaxed problem; that is the quantization is omitted and we consider the relaxed received signal $\mathbf{y}'$ instead of $\mathbf{y}$.

7.3.4.1 Safety Margin vs. Receiver SNR

The receiver SNR at the m-th user is given by

$$\mathrm{SNR}_m = \frac{\mathrm{E}\left[|y'_m|^2\right]}{\sigma_n^2} = \mathrm{E}\left[|y'_m|^2\right], \tag{7.23}$$

since we assume unit-variance AWGN. The expected value can be computed by averaging over N_s transmit signals. Hence, we get

$$\mathrm{SNR}_m \geq \frac{1}{N_s}\sum_{i=1}^{N_s}\left(\frac{\delta^{(i)}}{\sin\theta}\right)^2. \tag{7.24}$$

Thus, we can conclude that maximizing the safety margin δ leads in turn to maximizing the lower bound of the receive SNR at each user.

7.3.4.2 Safety Margin vs. SER

The statistics of the AWGN in the modified coordinate system do not change, since it just consists of a rotation. We denote the modified noise vector by $\tilde{\boldsymbol{\eta}}$, where $\tilde{\boldsymbol{\eta}} \sim \mathcal{CN}_{\mathbb{C}}\left(\mathbf{0}_M, \mathbf{I}_M\right)$.

The SER represents the probability that the entries of $\mathbf{z}+\tilde{\boldsymbol{\eta}}$ lie outside the corresponding decision regions. The SER of the m-th user for the t-th transmission can be then expressed as

$$\begin{aligned}
\mathrm{SER}_m^{(t)} &= \Pr\left(\Im\{z_m^{(t)}\}\geq 0\right)\left(\Pr\left(\left(\Re\{\tilde{\eta}_m+z_m^{(t)}\}\geq 0\right)\wedge\left(\Im\{\tilde{\eta}_m\}\geq\frac{\delta_m^{(t)}}{\cos(\theta)}+\Re\{\tilde{\eta}_m\}\tan(\theta)\right)\right)\right.\\
&\quad\left.+\Pr\left(\left(\Re\{\tilde{\eta}_m+z_m^{(t)}\}\geq 0\right)\wedge\left(\Im\{\tilde{\eta}_m\}\leq -2\left|\Im\{z_m\}\right|-\frac{\delta_m^{(t)}}{\cos(\theta)}-\Re\{\tilde{\eta}_m\}\tan(\theta)\right)\right)\right)\\
&\quad+\Pr\left(\Im\{z_m^{(t)}\}<0\right)\left(\Pr\left((\Re\{\tilde{\eta}_m+z_m\}\geq 0)\wedge\left(\Im\{\tilde{\eta}_m\}\leq -\frac{\delta_m^{(t)}}{\cos(\theta)}-\Re\{\tilde{\eta}_m\}\tan(\theta)\right)\right)\right.\\
&\quad\left.+\Pr\left(\left(\Re\{\tilde{\eta}_m+z_m^{(t)}\}\geq 0\right)\wedge\left(\Im\{\tilde{\eta}_m\}\geq 2\left|\Im\{z_m^{(t)}\}\right|+\frac{\delta_m^{(t)}}{\cos(\theta)}+\Re\{\tilde{\eta}_m\}\tan(\theta)\right)\right)\right)\\
&\quad+\Pr\left(\left(\Re\{\tilde{\eta}_m+z_m^{(t)}\}<0\right)\right)\\
&=1-\int_{-\Re\{z_m^{(t)}\}}^{\infty}\frac{1}{2}\left(\mathrm{erf}\left(\frac{\delta_m^{(t)}}{\cos(\theta)}+\Re\{\tilde{\eta}_m\}\tan(\theta)\right)\right.\\
&\quad\left.+\mathrm{erf}\left(2\left|\Im\{z_m^{(t)}\}\right|+\frac{\delta_m^{(t)}}{\cos(\theta)}+\Re\{\tilde{\eta}_m\}\tan(\theta)\right)\right)\frac{1}{\sqrt{\pi}}\,\mathrm{e}^{-\Re\{\tilde{\eta}_m\}^2}\;\mathrm{d}\Re\{\tilde{\eta}_m\}. \qquad (7.25)
\end{aligned}$$

Since the erf function is monotonically increasing and it holds that $\delta^{(t)}\leq\delta_m^{(t)},\forall m$ and $\Re\{z_m^{(t)}\}\geq\frac{\delta_m^{(t)}}{\sin(\theta)}$, we obtain

$$\mathrm{SER}_m^{(t)}\leq 1-\int_{-\frac{\delta^{(t)}}{\sin\theta}}^{\infty}\mathrm{erf}\left(\frac{\delta_m^{(t)}}{\cos(\theta)}+\Re\{\tilde{\eta}_m\}\tan(\theta)\right)\frac{1}{\sqrt{\pi}}\,\mathrm{e}^{-\Re\{\tilde{\eta}_m\}^2}\;\mathrm{d}\Re\{\tilde{\eta}_m\}. \qquad (7.26)$$

Thus, the SER at the m-th user averaged over a transmission block of length N_s is upper bounded by

$$\mathrm{SER}_m\leq 1-\frac{1}{N_s}\sum_{t=1}^{N_s}\int_{-\frac{\delta^{(t)}}{\sin\theta}}^{\infty}\mathrm{erf}\left(\frac{\delta^{(t)}}{\cos(\theta)}+\Re\{\tilde{\eta}_m\}\tan(\theta)\right)\frac{1}{\sqrt{\pi}}\,\mathrm{e}^{-\Re\{\tilde{\eta}_m\}^2}\;\mathrm{d}\Re\{\tilde{\eta}_m\}, \qquad (7.27)$$

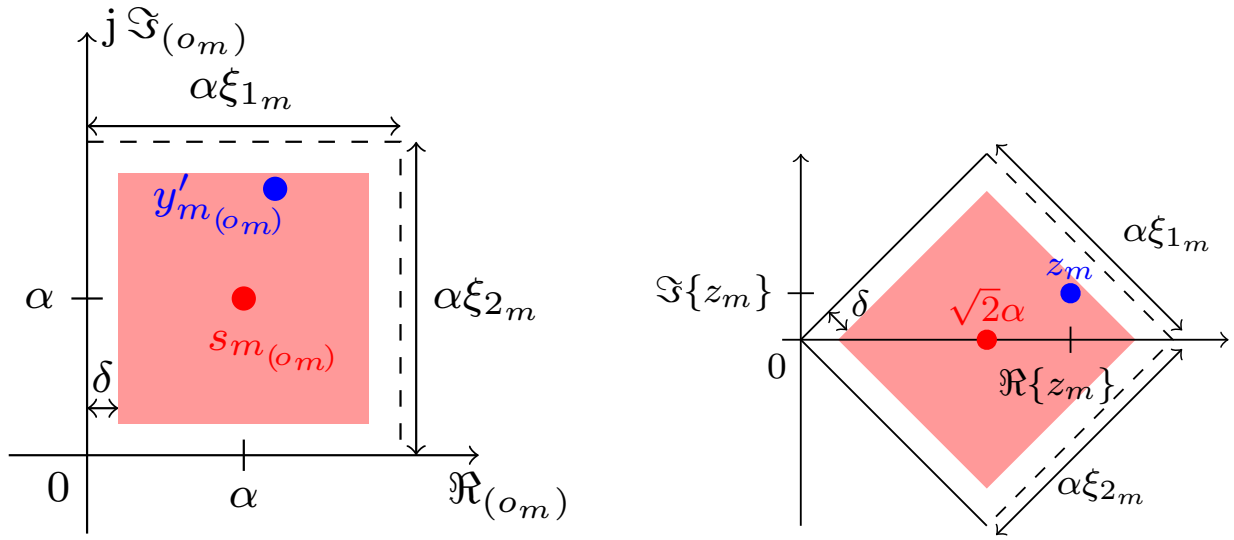
which means that maximizing δ minimizes the upper bound on the SER.

7.4 Problem Formulation for QAM Signaling

7.4.1 The Need for an Additional Degree of Freedom α

In this section, we assume that the input signals s_m, $m=1,\cdots,M$, belong to the S-QAM constellation that is defined in (4.14). As explained in Section 7.2, the safety margin δ has to be maximized such that the entries of the noiseless received signal $\mathbf{y}$ belong to the intended SRs. The SRs in turn are determined by the constellation set $\mathbb{S}$ and the safety margin δ. Hence, the maximal value is limited by the coordinates of the inner constellation points; that is

$$\delta\leq 1. \qquad (7.28)$$



(a) Shifted coordinate system for $\Re\{s_m\} > 0$ and $\Im\{s_m\} > 0$.

(b) Shifted and rotated coordinate system.

Fig. 7.4: Illustration of the QAM receiver SR in the shifted and in the shifted and rotated coordinate system : $\xi_{1/2_m} \in \{2, \infty\}$.

Independently of the available transmit power, the entries of $\mathbf{y}$ cannot have a distance to the decision thresholds larger than 1. Hence, the available transmit power cannot be exploited to the fullest. This results already in a limitation of the problem formulation.

Thanks to the receive processing $\mathbf{G}$, we can introduce an additional degree of freedom α such that the entries of the received signal $\mathbf{y}$ do not have to belong to the SRs of the set $\mathbb{S}$ but rather to a scaled version of them; that is, the QAM constellation at each receiver gets scaled by α. The receive processing $\mathbf{G}$ has to rescale the received signals to their nominal positions before the decision block. Thus, the constraint in (7.28) is replaced by

$$\delta \leq \alpha, \tag{7.29}$$

where α has to be jointly optimized with δ. Note that maximizing δ results in turn to maximizing α, which leads to a maximal exploitation of the available transmit power. Thus, the entries of the signal vector $\mathbf{x}$ will get closer to the polygon corners, which decreases the variations between $\mathbf{t}$ and $\mathbf{x}$.

The factor α denotes the expansion or shrinkage factor of the constellation at the receiver side depending on the available transmit power P_{tx}. As explained in Section 7.2, the optimization problem is formulated for the specific case, i.e. $P_{\text{tx}} = N$.

7.4.2 Scaled Symbol Region for QAM Signals

To describe the SRs for QAM signaling after considering α, we define a new coordinate system, that is a shifted and rotated version of the original coordinate system. First, the original receiver constellation system is shifted by o_m

$$o_m = \alpha\left(s_m - \operatorname{sgn}\left(s_m\right)\right), \; m = 1, \cdots, M. \tag{7.30}$$

We get the following expressions for the received and the desired signal in the intermediate coordinate system depicted in Fig. 7.4b

$$y'_{m_{(o_m)}} = y'_m - o_m \tag{7.31}$$

$$\begin{aligned} s_{m(o_m)} &= \alpha s_m - o_m \\ &\overset{(7.30)}{=} \alpha \operatorname{sgn}(s_m). \end{aligned} \tag{7.32}$$

Second, the intermediate coordinate system is rotated by the phase of the symbol of interest $s_{m(o_m)}$. So the signal z_m, which represents the received signal y'_m in the modified coordinate system, is given by

$$z_m = \frac{y'_{m_{(o_m)}} s^*_{m_{(o_m)}}}{|s_{m(o_m)}|},\ m = 1, \cdots, M. \tag{7.33}$$

The m-th SR, $m = 1, \cdots, M$, as shown in Fig. 7.4b, can be hence described by

$$\Re\{z_m\} \geq \sqrt{2}\delta, \tag{7.34}$$

$$\Re\{z_m\} \leq \sqrt{(\alpha\xi_{1_m} - \delta)^2 + (\alpha\xi_{2_m} - \delta)^2}, \tag{7.35}$$

$$|\Im\{z_m\}| \leq \left(\Re\{z_m\} - \sqrt{2}\delta\right), \tag{7.36}$$

$$\Im\{z_m\} \leq -\Re\{z_m\} + \sqrt{2}\left(\alpha\xi_{2_m} - \delta\right), \tag{7.37}$$

$$\Im\{z_m\} \geq \Re\{z_m\} - \sqrt{2}\left(\alpha\xi_{1_m} - \delta\right), . \tag{7.38}$$

Note that ξ_{1_m} and $\xi_{2_m} \in \{2, \infty\}$ depending on which constellation point the symbol of interest s_m corresponds to. If s_m is one of the outer constellation points, then at least ξ_{1_m} or ξ_{2_m} must be equal to ∞. Moreover, (7.34) and (7.35) are inherently fulfilled by (7.36), (7.37) and (7.38). In the vector notation, the vector $\mathbf{z}$ can be expressed as

$$\begin{aligned} \mathbf{z} &= \frac{1}{\sqrt{2}} \operatorname{diag}\left(\operatorname{sgn}(\mathbf{s}^*)\right)\left(\mathbf{H}\mathbf{x} - \alpha\left(\mathbf{s} - \operatorname{sgn}(\mathbf{s})\right)\right) \\ &= \hat{\mathbf{H}}\mathbf{x} - \alpha\mathbf{c}, \end{aligned} \tag{7.39, 7.40}$$

where

$$\hat{\mathbf{H}} = \frac{1}{\sqrt{2}} \operatorname{diag}\left(\operatorname{sgn}(\mathbf{s}^*)\right)\mathbf{H}, \tag{7.41}$$

$$\mathbf{c} = \frac{1}{\sqrt{2}} \operatorname{diag}\left(\operatorname{sgn}(\mathbf{s}^*)\right)\left(\mathbf{s} - \operatorname{sgn}(\mathbf{s})\right). \tag{7.42}$$

Thus, all M SRs can be compactly described by

$$|\Im\{\mathbf{z}\}| \leq \left(\Re\{\mathbf{z}\} - \sqrt{2}\delta\mathbf{1}_M\right), \tag{7.43}$$

$$\Im\{\mathbf{z}\} \leq -\Re\{\mathbf{z}\} + \sqrt{2}\left(\alpha\boldsymbol{\xi}_2 - \delta\mathbf{1}_M\right), \tag{7.44}$$

$$\Im\{\mathbf{z}\} \geq \Re\{\mathbf{z}\} - \sqrt{2}\left(\alpha\boldsymbol{\xi}_1 - \delta\mathbf{1}_M\right). \tag{7.45}$$

When using the following real-valued representation

$$\mathbf{V} = \begin{bmatrix} \Re\{\hat{\mathbf{H}}\} & -\Im\{\hat{\mathbf{H}}\} \end{bmatrix} \tag{7.46}$$

$$\mathbf{W} = \begin{bmatrix} \Im\{\hat{\mathbf{H}}\} & \Re\{\hat{\mathbf{H}}\} \end{bmatrix}, \tag{7.47}$$

the constraint in (7.45) can be rewritten as

$$\begin{bmatrix} \mathbf{W}-\mathbf{V} & \mathbf{1}_M & \Re\{\mathbf{c}\}-\Im\{\mathbf{c}\} \\ -\mathbf{W}-\mathbf{V} & \mathbf{1}_M & \Re\{\mathbf{c}\}+\Im\{\mathbf{c}\} \\ \mathbf{W}+\mathbf{V} & \mathbf{1}_M & -\Re\{\mathbf{c}\}-\Im\{\mathbf{c}\}-\sqrt{2}\boldsymbol{\xi}_2 \\ -\mathbf{W}+\mathbf{V} & \mathbf{1}_M & -\Re\{\mathbf{c}\}+\Im\{\mathbf{c}\}-\sqrt{2}\boldsymbol{\xi}_1 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}} \\ \sqrt{2}\delta \\ \alpha \end{bmatrix} \leq \mathbf{0}_{4M}. \tag{7.48}$$

By adding the first line to the fourth and the second line to the third, the latter constraint recalculates to

$$\begin{bmatrix} \mathbf{W}-\mathbf{V} & \mathbf{1}_M & \Re\{\mathbf{c}\}-\Im\{\mathbf{c}\} \\ -\mathbf{W}-\mathbf{V} & \mathbf{1}_M & \Re\{\mathbf{c}\}+\Im\{\mathbf{c}\} \\ \mathbf{0}_{M,2N} & \sqrt{2}\,\mathbf{1}_M & -\boldsymbol{\xi}_2 \\ \mathbf{0}_{M,2N} & \sqrt{2}\,\mathbf{1}_M & -\boldsymbol{\xi}_1 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}} \\ \sqrt{2}\delta \\ \alpha \end{bmatrix} \leq \mathbf{0}_{4M}. \tag{7.49}$$

7.4.3 Optimization Problem with the Relaxed Polygon Constraint

We are interested in maximizing the safety margin as presented in (7.8). In contrast to the PSK case, there is a constraint on δ in the QAM case, stated in (7.29), which is inherently fulfilled by (7.49). Combining (7.8) with the SRs constraint in (7.49) and the relaxed polygon constraint in (7.7), we get a linear programming problem for the design of the symbol-wise precoder for QAM signaling. The optimization problem for the case of $P_{\text{tx}} = N$ is given by

$$\max_{\mathbf{v}} \begin{bmatrix} \mathbf{0}_{2N}^{\mathrm{T}} & 1 & 0 \end{bmatrix} \mathbf{v} \text{ s.t. } \begin{bmatrix} \mathbf{W}-\mathbf{V} & \mathbf{1}_M & \Re\{\mathbf{c}\}-\Im\{\mathbf{c}\} \\ -\mathbf{W}-\mathbf{V} & \mathbf{1}_M & \Re\{\mathbf{c}\}+\Im\{\mathbf{c}\} \\ \mathbf{0}_{M,2N} & \sqrt{2}\,\mathbf{1}_M & -\boldsymbol{\xi}_2 \\ \mathbf{0}_{M,2N} & \sqrt{2}\,\mathbf{1}_M & -\boldsymbol{\xi}_1 \\ \mathbf{E} & \mathbf{0}_{N(Q-4)} & \mathbf{0}_{N(Q-4)} \end{bmatrix} \mathbf{v} \leq \begin{bmatrix} \mathbf{0}_{4M} \\ \cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{N(Q-4)} \end{bmatrix}$$

$$\text{and } \begin{bmatrix} -\cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{2N} \\ 0 \\ 0 \end{bmatrix} \leq \mathbf{v} \leq \begin{bmatrix} \cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{2N} \\ \infty \\ \infty \end{bmatrix}, \tag{7.50}$$

where $\mathbf{v}^{\mathrm{T}} = \begin{bmatrix} \bar{\mathbf{x}}^{\mathrm{T}} & \sqrt{2}\delta & \alpha \end{bmatrix}$. In order to solve the optimization problem for different P_{tx} values, it is sufficient to scale the optimal solution of (7.21) by $\frac{P_{\text{tx}}}{N}$ due to the linearity of the optimization problem.

Again the optimized vector $\mathbf{x} \in \mathbb{X}^N$ goes through the quantizer, as stated in (4.1), to obtain the transmit vector $\mathbf{t}$.

7.4.4 Safety Margin δ for QAM Signals

For the QAM case, the safety margin δ in (7.8) can be expressed as

$$\delta = \min_m \min \left(\frac{1}{\sqrt{2}} \left(\Re\{\mathbf{z}\} - |\Im\{\mathbf{z}\}| \right), \alpha\boldsymbol{\xi}_1 - \frac{1}{\sqrt{2}} \left(\Re\{\mathbf{z}\} - \Im\{\mathbf{z}\} \right), \alpha\boldsymbol{\xi}_2 - \frac{1}{\sqrt{2}} \left(\Re\{\mathbf{z}\} + \Im\{\mathbf{z}\} \right) \right), \tag{7.51}$$

where the operator $|\bullet|$ is applied element-wise to the entries of $\Im\{\mathbf{z}\}$.

7.4.5 Interpretation of the Safety Margin δ for QAM Signals

Again we consider the relaxed problem; that is the quantization is omitted and we consider the received signal $\mathbf{y}'$ instead of $\mathbf{y}$.

7.4.5.1 Safety Margin vs. Receiver SNR

The receiver SNR at the m-th user can be approximated by

$$\mathrm{SNR}_m \approx \frac{1}{N_s}\sum_{t=1}^{N_s}(\alpha^{(t)})^2\sigma_s^2. \tag{7.52}$$

Since $\delta \leq \alpha$, we get

$$\mathrm{SNR}_m \geq \frac{1}{N_s}\sum_{t=1}^{N_s}(\delta^{(t)})^2\sigma_s^2. \tag{7.53}$$

Thus, maximizing δ results in maximizing the lower bound of the receiver SNR.

7.4.5.2 Safety Margin vs. SER

The statistics of the AWGN in the modified coordinate system do not change, since it just consists of a rotation. The shift is applied only to the noiseless received signal $\mathbf{y}$. We denote the modified noise vector by $\tilde{\boldsymbol{\eta}}$, where $\tilde{\boldsymbol{\eta}} \sim \mathcal{CN}_{\mathbb{C}}(\mathbf{0}_M, \mathbf{I}_M)$. The SER represents the probability that the entries of $\mathbf{z} + \tilde{\boldsymbol{\eta}}$ lie outside the thresholds that are represented by the black lines in Fig. 7.4b. We consider the worst case, where the decision region is bounded from all four sides. Additionally, we assume that the distribution of the entries of $\mathbf{z}$ is uniform. Therefore, the SER of the m-th user for the t-th transmission is upper bounded by

$$\begin{aligned}
\mathrm{SER}_m^{(t)} \leq\ & \Pr\left(\left(0 \leq \Re\{z_m^{(t)}\} + \Re\{\tilde{n}_m\} \leq 2\sqrt{2}\alpha^{(t)}\right) \wedge \left(\Im\{\tilde{n}_m\} \geq \sqrt{2}\delta_m^{(t)} + \Re\{\tilde{n}_m\}\right)\right) \\
& + \Pr\left(\left(0 \leq \Re\{z_m^{(t)}\} + \Re\{\tilde{n}_m\} \leq 2\sqrt{2}\alpha^{(t)}\right) \wedge \left(\Im\{\tilde{n}_m\} \leq -\sqrt{2}\delta_m^{(t)} - \Re\{\tilde{n}_m\}\right)\right) \\
& + \Pr\left(\left(0 \leq \Re\{z_m^{(t)}\} + \Re\{\tilde{n}_m\} \leq 2\sqrt{2}\alpha^{(t)}\right) \wedge \left(\Im\{\tilde{n}_m\} \geq 2\sqrt{2}\alpha^{(t)} - \sqrt{2}\delta_m^{(t)} - \Re\{\tilde{n}_m\}\right)\right) \\
& + \Pr\left(\left(0 \leq \Re\{z_m^{(t)}\} + \Re\{\tilde{n}_m\} \leq 2\sqrt{2}\alpha^{(t)}\right) \wedge \left(\Im\{\tilde{n}_m\} \leq -2\sqrt{2}\alpha + \sqrt{2}\delta_m^{(t)} + \tilde{n}_{m,R}\right)\right) \\
& + 1 - \Pr(0 \leq \Re\{z_m^{(t)}\} + \Re\{\tilde{n}_m\} \leq 2\sqrt{2}\alpha^{(t)}) \qquad (7.54) \\
=\ & 1 + \int_{-\Re\{z_m^{(t)}\}}^{2\sqrt{2}\alpha^{(t)} - \Re\{z_m^{(t)}\}} \left(1 - \left(\mathrm{erf}\left(\sqrt{2}\delta_m^{(t)} + \Re\{\tilde{n}_m\}\right) + \mathrm{erf}\left(2\sqrt{2}\alpha^{(t)} - \sqrt{2}\delta_m^{(t)} - \Re\{\tilde{n}_m\}\right)\right)\right) \\
& \frac{1}{\sqrt{\pi}}\,\mathrm{e}^{-\Re\{\tilde{\eta}_m\}^2}\ \mathrm{d}\Re\{\tilde{\eta}_m\} \qquad (7.55)
\end{aligned}$$

Since the erf function is monotonically increasing and it holds that $\delta^{(t)} \leq \delta_m^{(t)} \leq \alpha^{(t)}, \forall m$ and $\sqrt{2}\delta_m^{(t)} \leq \Re\{z_m^{(t)}\} \leq 2\sqrt{2}\alpha^{(t)} - \sqrt{2}\delta_m^{(t)}$, we obtain

$$\begin{aligned}
\mathrm{SER}_m^{(t)} &\leq 1 + \int_{-2\sqrt{2}\alpha^{(t)} + \sqrt{2}\delta_m^{(t)}}^{2\sqrt{2}\alpha^{(t)} - \sqrt{2}\delta_m^{(t)}} \left(1 - 2\mathrm{erf}\left(\sqrt{2}\delta_m^{(t)} + \Re\{\tilde{\eta}_m\}\right)\right) \frac{1}{\sqrt{\pi}}\,\mathrm{e}^{-\Re\{\tilde{\eta}_m\}^2}\ \mathrm{d}\Re\{\tilde{\eta}_m\} \qquad (7.56) \\
&\leq 1 + \int_{-2\sqrt{2}\alpha^{(t)} + \sqrt{2}\delta^{(t)}}^{2\sqrt{2}\alpha^{(t)} - \sqrt{2}\delta^{(t)}} \left(1 - 2\mathrm{erf}\left(\sqrt{2}\delta^{(t)} + \Re\{\tilde{\eta}_m\}\right)\right) \frac{1}{\sqrt{\pi}}\,\mathrm{e}^{-\Re\{\tilde{\eta}_m\}^2}\ \mathrm{d}\Re\{\tilde{\eta}_m\}. \qquad (7.57)
\end{aligned}$$

Thus, the SER at the m-th user averaged over a transmission block of length N_s is upper bounded by

$$\mathrm{SER}_m \leq 1 + \frac{1}{N_s} \sum_{t=1}^{N_s} \int_{-2\sqrt{2}\alpha^{(t)}+\sqrt{2}\delta^{(t)}}^{2\sqrt{2}\alpha^{(t)}-\sqrt{2}\delta^{(t)}} \left(1 - 2\mathrm{erf}\left(\sqrt{2}\delta^{(t)} + \Re\{\tilde{\eta}_m\}\right)\right) \frac{1}{\sqrt{\pi}} \mathrm{e}^{-\Re\{\tilde{\eta}_m\}^2} \, \mathrm{d}\Re\{\tilde{\eta}_m\}. \tag{7.58}$$

Since erf is a monotonically increasing function, larger values of δ make the upper bound of the SER decrease. Note that larger values of δ lead inherently to larger values of α.

7.4.6 Symbol-wise Processing vs. Block-wise Processing

One might ask why we opt for symbol-wise processing and not block-wise processing. The factor α cannot be communicated to the receiver and hence has to be estimated. The estimation is based on averaging over a block of T received signals. Thus, one expects that the design of α at the transmitter has to be computed for the same block length, i.e. $B = T$. However, fixing α for a certain block length B means that B vectors $\mathbf{x}$ have to be designed jointly with a single factor α instead of having a distinct factor α for every vector $\mathbf{x}$. Additionally, the joint optimization of B vectors results in a higher-dimensional linear programming problem, where the number of inequalities is increased by a factor of B. Hence, block-wise processing not only increases the computational complexity of the problem as can be deduced from Section 7.5 but also reduces the degrees of freedom of the optimization problem at the transmitter. This leads to the entries of the vector $\mathbf{x}$ moving farther from the polygon corners, thus increasing the quantization distortions. This effect is illustrated in Fig. 7.5, where the entries of $\mathbf{e}_m^{\mathrm{T}}\mathbf{H}\mathbf{x}$, $\mathbf{e}_m^{\mathrm{T}}\mathbf{H}\mathbf{t}$ and $\frac{1}{\alpha}\mathbf{e}_m^{\mathrm{T}}\mathbf{H}\mathbf{t}$ of an arbitrary user m are obtained by transmitting 1024 16-QAM signal vectors through an i.i.d. channel of coefficients $h_{mn} \sim \mathcal{CN}_{\mathbb{C}}(0,1),\ n = 1, \cdots, N,\ m = 1, \cdots, M$, where $N = 64$, $M = 8$ and $Q = 4$. The optimization is computed for both symbol-wise processing, i.e. $B = 1$, and block-wise processing with $B = 4$. As can be deduced from the plots, block-wise processing leads to a larger safety margin with the relaxed vector $\mathbf{x}$. However, after applying the quantization this gain is lost and the symbol-wise processing is more robust against the quantization operation. This can be further explained by the results in Table 7.1, which shows $\mathrm{E}\left[\frac{\|\mathbf{t}\neq\mathbf{x}\|_1}{N}\right]$, the percentange of entries of $\mathbf{x}$ that are distorted due to the quantization and the MSE between $\mathbf{t}$ and $\mathbf{x}$. We see that increasing B significantly increases the quantization distortion. Therefore, the

	$B = 1$	$B = 4$
$\mathrm{E}\left[\frac{\|\mathbf{t}\neq\mathbf{x}\|_1}{N}\right]$	0.2176	0.4432
$\mathrm{E}\left[\|\mathbf{t}-\mathbf{x}\|_2^2\right]$	2.5458	12.6429

Table 7.1: Quantization distortion vs. B, © 2018 IEEE.

symbol-wise processing is chosen in this contribution, i.e, an optimal value of α is designed for each vector $\mathbf{x}$.

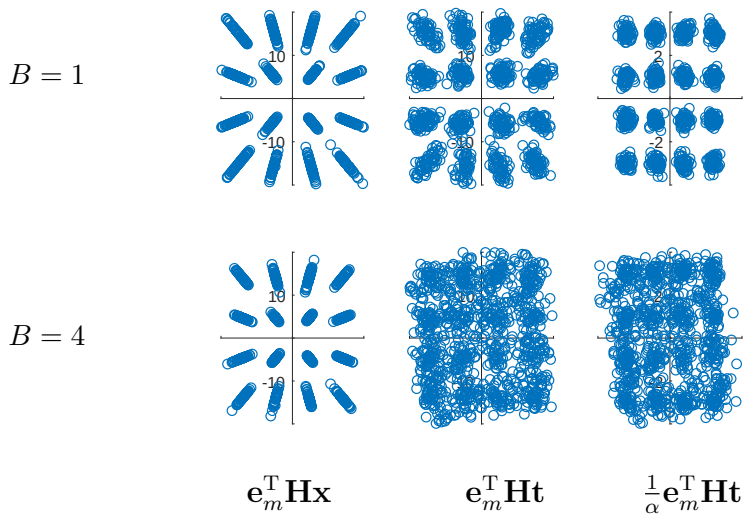


Fig. 7.5: The noiseless received symbols at one arbitrary user m for an arbitrary i.i.d. channel realization with $N = 64$, $M = 8$ and $Q = 4$, © 2018 IEEE.

7.4.7 One Joint α vs. M Distinct α's for M Users

Symbol-wise transmit processing followed by block-wise receive processing is reliable only if the obtained values of α, i.e. $\alpha^{(i)},\ i = 1, \cdots, T$, do not vary much from one vector $\mathbf{x}^{(i)}$ to another. Otherwise, estimating the mean value of α at the receiver would not be sufficient for correct detection. This explains why we choose one joint α for all users. If a different value α_m per user is chosen, this will result in more degrees of freedom and the values $\alpha_m,\ m = 1, \cdots, M$, would fluctuate much more from one vector $\mathbf{x}$ to another, which worsens the estimation result at the receiver. For a large number of users, the jointly designed α will not vary much, since the norm of the input vector $\mathbf{s}$ will not fluctuate much from one realization to another. This behavior is explained as follows.

For a given channel realization, T symbols are transmitted and the MSM method is applied T times to get in the optimal case

$$\begin{aligned} \mathbf{H}\mathbf{x}^{(1)} &= \alpha^{(1)}\mathbf{s}^{(1)} \\ &\vdots \\ \mathbf{H}\mathbf{x}^{(T)} &= \alpha^{(T)}\mathbf{s}^{(T)}. \end{aligned} \tag{7.59}$$

The factor $\alpha^{(i)}$ can be then expressed as

$$\alpha^{(i)} = \frac{\left\|\mathbf{H}\mathbf{x}^{(i)}\right\|_2}{\left\|\mathbf{s}^{(i)}\right\|_2},\ i = 1, \cdots, T \tag{7.60}$$

and hence is upper bounded by

$$\alpha^{(i)} \leq \frac{\|\mathbf{H}\|_2 \left\|\mathbf{x}^{(i)}\right\|_2}{\left\|\mathbf{s}^{(i)}\right\|_2} \leq \frac{\|\mathbf{H}\|_2 \left\|\mathbf{t}^{(i)}\right\|_2}{\left\|\mathbf{s}^{(i)}\right\|_2}. \tag{7.61}$$

Due to the constant envelope property, we get $\left\|\mathbf{t}^{(i)}\right\|_2 = \sqrt{N}, \forall i = 1, \cdots, T$. Thus, the fluctuation of the upper bound of $\alpha^{(i)}$ is determined by $\left\|\mathbf{s}^{(i)}\right\|_2$. Since the entries of the

vector $\mathbf{s}$ are i.i.d, the fluctuation of $\left\|\mathbf{s}^{(i)}\right\|_2$ from one realization to another and for $S > 4$ diminishes by increasing its dimension, i.e. the number of users M. We recall that the MSM method aims to maximize the safety margin δ that is upper bounded by α as stated in (7.29). Consequently, the MSM method will maximize α as well. As explained before, the upper bound of α fluctuates less from one signal realization to another for a large number of users and thus the optimal value of $\alpha^{(i)}$ has only small fluctuations in such cases.

To make the received signals belong to the nominal constellation, the receive processing $\mathbf{G}$, as explained in Section 4.7, is applied. For each time instant, the coefficients $g_m,\ m = 1, \cdots, M$ should be equal to the inverse of α. However, the computation of the coefficients g_m is based on blind estimation over a block length T. Therefore, the resulting coefficients g_m are ideally equal to the inverse of the mean value of α, i.e. $g_m \approx 1/\operatorname{E}[\alpha]$. Hence, the designed SR at the transmitter in the noise-free case, which is illustrated in Fig. 7.4a, is scaled by $1/\operatorname{E}[\alpha]$ at the receiver before the decision operation. For the i-th transmission, the center of the obtained nominal SR_m corresponds to $s_m^{(i)}\alpha^{(i)}/\operatorname{E}[\alpha]$. If $\alpha^{(i)}$ corresponds to $\operatorname{E}[\alpha]$, the nominal SR_m after the receive filter $\mathbf{G}$ is situated in the center of the corresponding nominal decision region, that is $s_m^{(i)}$. However, if $\alpha^{(i)}$ is smaller than $\operatorname{E}[\alpha]$, the center of the nominal SR_m shifts towards the left lower corner of the decision region of $s_m^{(i)}$ and might pass over it. Analogously, if $\alpha^{(i)}$ is larger than $\operatorname{E}[\alpha]$, the center of the nominal SR_m shifts towards the right upper corner of the decision region and might pass over it, too. To illustrate the passing over phenomenon, we introduce the nominal safety margin δ_{nom}, which is given by

$$\begin{aligned}
\delta_{\mathrm{nom}} &= \min\left(\frac{2\alpha_{\min}+\delta_{\alpha_{\min}}}{\operatorname{E}[\alpha]} - 2, 2 - \frac{2\alpha_{\max}-\delta_{\alpha_{\max}}}{\operatorname{E}[\alpha]}\right) \\
&= \min\left(\frac{\delta_{\alpha_{\min}}}{\operatorname{E}[\alpha]} - 2\frac{\operatorname{E}[\alpha]-\alpha_{\min}}{\operatorname{E}[\alpha]}, \frac{\delta_{\alpha_{\max}}}{\operatorname{E}[\alpha]} - 2\frac{\alpha_{\max}-\operatorname{E}[\alpha]}{\operatorname{E}[\alpha]}\right) \\
&\geq \frac{1}{\operatorname{E}[\alpha]}\left(\min\left(\delta_{\alpha_{\min}}, \delta_{\alpha_{\max}}\right) - 2\max\left(\operatorname{E}[\alpha]-\alpha_{\min}, \alpha_{\max}-\operatorname{E}[\alpha]\right)\right) \\
&= \frac{1}{\operatorname{E}[\alpha]}\delta_{\alpha_{\min}} - 2\Delta\alpha \\
\delta_{\mathrm{nom}} &\geq \frac{\min\limits_i \delta^{(i)}}{\sum_{i=1}^{T}\alpha^{(i)}/T} - 2\Delta\alpha,
\end{aligned} \tag{7.62}$$

where $\alpha_{\min}$, $\alpha_{\max}$, $\delta_{\alpha_{\min}}$ and $\delta_{\alpha_{\max}}$ denote the minimal and maximal value of $\alpha^{(i)}$ and their corresponding safety margins, respectively. $\Delta\alpha$ represents the maximal relative fluctuation of α. For the optimization with one single α, it is defined as

$$\Delta\alpha = \max\left(1 - \frac{\min\limits_i \alpha^{(i)}}{\sum_{i=1}^{T}\alpha^{(i)}/T}, \frac{\max\limits_i \alpha^{(i)}}{\sum_{i=1}^{T}\alpha^{(i)}/T} - 1\right), \tag{7.63}$$

whereas it is expressed as

$$\Delta\alpha = \max_m \max\left(1 - \frac{\min\limits_i \alpha_m^{(i)}}{\sum_{i=1}^{T}\alpha_m^{(i)}/T}, \frac{\max\limits_i \alpha_m^{(i)}}{\sum_{i=1}^{T}\alpha_m^{(i)}/T} - 1\right) \tag{7.64}$$

for the optimization with M distinct α's for the different users. In the case that δ_{nom} gets negative values due to the large fluctuation of α, i.e. $\Delta\alpha$, it implies that the nominal SRs overlap in the noise-free case. To ensure a large positive number for δ_{nom}, $\Delta\alpha$ must be kept as small as possible. In the best case, $\Delta\alpha$ must be equal to zero.

From (7.62), we can conclude that smaller fluctuations of α lead to larger values of δ_{nom}. This observation is justified by numerical results, where the MSM optimization is run for $T = 128$ 16-QAM symbols, for 100 i.i.d. channel realizations and for $Q = 4$. The obtained values for $\Delta\alpha$ and δ_{nom} are averaged over the channels and shown in Table 7.2 and Table 7.3. We can deduce from Table 7.2, where a common α is designed for all users, that the values

N	64		200	
M	$\Delta\alpha$	δ_{nom}	$\Delta\alpha$	δ_{nom}
2	1.3	-0.3	1.2	-0.2
8	0.5	0.2	0.4	0.3
14	0.3	0.2	0.3	0.4

Table 7.2: $\Delta\alpha$ and δ_{nom} vs. N and M: single α, © 2018 IEEE.

of α fluctuate less and hence δ_{nom} increases by increasing the number of users. However, no monotonic behavior of the fluctuations is noticed in Table 7.3, which shows relatively larger values of $\Delta\alpha$ and hence smaller values of δ_{nom} compared to Table 7.2.

7.5 Computational Complexity of MSM

7.5.1 On the Computational Complexity of General Linear Programming Problems

In this section, we study the computational complexity of the simplex method for a general linear programming problem with bounded variables in inequality form:

$$\max_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x} \text{ s.t. } \mathbf{A}\mathbf{x} \leq \mathbf{b} \quad \text{and } \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \tag{7.65}$$

where $\mathbf{c},\ \mathbf{x},\ \mathbf{l}$ and $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$.

N	64		200	
M	$\Delta\alpha$	δ_{nom}	$\Delta\alpha$	δ_{nom}
2	1.2	-0.2	1.1	-0.1
8	1.1	-0.8	0.5	0.3
14	1.6	-2	0.7	-0.4

Table 7.3: $\Delta\alpha$ and δ_{nom} vs. N and M: M α's, © 2018 IEEE.

First, we have to make sure that the entries of $\mathbf{b}$ are non-negative. To this end, we change the signs of the inequalities that correspond to negative entries in $\mathbf{b}$. So we get

$$\min_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x} \text{ s.t. } \tilde{\mathbf{A}}\mathbf{x} \gtreqless \tilde{\mathbf{b}} \quad \text{and } \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \tag{7.66}$$

where $\tilde{\mathbf{b}} \in \mathbb{R}_{+}^{m}$ and some inequalities hold with the sign $\leq$ and others with the sign $\geq$.

Second, the linear programming problem is transformed to the canonical form by introducing m slack and surplus variables denoted by $\mathbf{x}_{\mathrm{s}}$. Additionally, a artificial variables denoted by $\mathbf{x}_{\mathrm{a}}$, with $0 \leq a \leq m$, are added to set up an initial feasible solution [72]. The equivalent enlarged problem reads as

$$\min_{\bar{\mathbf{x}}} \bar{\mathbf{c}}^{\mathrm{T}}\bar{\mathbf{x}} \text{ s.t. } \bar{\mathbf{A}}\bar{\mathbf{x}} = \tilde{\mathbf{b}} \quad \text{and } \bar{\mathbf{l}} \leq \bar{\mathbf{x}} \leq \bar{\mathbf{u}}, \tag{7.67}$$

where $\bar{\mathbf{A}} = \begin{bmatrix}\tilde{\mathbf{A}} & \mathbf{A}_{\mathrm{s}} & \mathbf{I}_a\end{bmatrix} \in \mathbb{R}^{m\times(n+m+a)}$, $\bar{\mathbf{x}}^{\mathrm{T}} = \begin{bmatrix}\mathbf{x}^{\mathrm{T}} & \mathbf{x}_{\mathrm{s}}^{\mathrm{T}} & \mathbf{x}_{\mathrm{a}}^{\mathrm{T}}\end{bmatrix} \in \mathbb{R}^{n+m+a}$, $\bar{\mathbf{l}}^{\mathrm{T}} = \begin{bmatrix}\mathbf{l}^{\mathrm{T}} & \mathbf{0}_{m+a}^{\mathrm{T}}\end{bmatrix}$ and $\bar{\mathbf{u}}^{\mathrm{T}} = \begin{bmatrix}\mathbf{u}^{\mathrm{T}} & \infty\mathbf{1}_{m+a}^{\mathrm{T}}\end{bmatrix}$. The matrix $\mathbf{A}_{\mathrm{s}}$ is a diagonal matrix with entries equal to 1 or -1 depending on whether the inequality sign in (7.66) is $\leq$ or $\geq$, respectively. The number a of artificial variables is defined by the number of negative entries in $\mathbf{A}_{\mathrm{s}}$, such that the concatenation of m columns from $\begin{bmatrix}\mathbf{A}_{\mathrm{s}} & \mathbf{I}_a\end{bmatrix}$ can construct the identity matrix $\mathbf{I}_m$. For the special case $\mathbf{b} = \tilde{\mathbf{b}}$, i.e. the entries of $\mathbf{b}$ are non-negative, $\mathbf{A}_{\mathrm{s}} = \mathbf{I}_m$. Hence, no artificial variables are needed, i.e. $a = 0$.

With the use of the simplex method to solve (7.67), the number of operations (multiplication and addition pairs) on each iteration is given by, [72, p.83],

$$3m \qquad \text{or} \qquad (m+1)(n+a+1)+2m, \tag{7.68}$$

depending on whether pivoting is required or not. According to [72, p.86], in most iterations no pivoting is required and hence less computation is needed.

7.5.2 Computational Complexity of MSM for PSK Signaling

As can be seen from (7.21), there are $m = 2M + N(Q-4)$ inequalities and $n = 2N + 1$ variables. The number a of artificial variables reduces to 0, since the vector $\mathbf{b}^{\mathrm{T}} = \begin{bmatrix}\mathbf{0}_{2M}^{\mathrm{T}} & \cos\left(\frac{\pi}{Q}\right)\mathbf{1}_{N(Q-4)}^{\mathrm{T}}\end{bmatrix}$ has only non-negative entries. Thus, the number of operations (multiplication and addition pairs) on each iteration calculates in this case to

$$6M + 3(Q-4)N, \tag{7.69}$$

or

$$2N + 4MN + 8M + 2 + (Q-4)\left(2N^2 + 4N\right). \tag{7.70}$$

For massive MIMO systems, where $N \gg M$, the number of floating-point operations for each iteration is on the order of

$$\#\mathrm{FLOPs}_{\mathrm{PSK}} = \mathcal{O}\left(4MN + (Q-4)\left(2N^2 + 4N\right)\right). \tag{7.71}$$

For the special case of one-bit quantization, i.e. $Q = 4$, the computational complexity of the MSM method is linear in N and M.

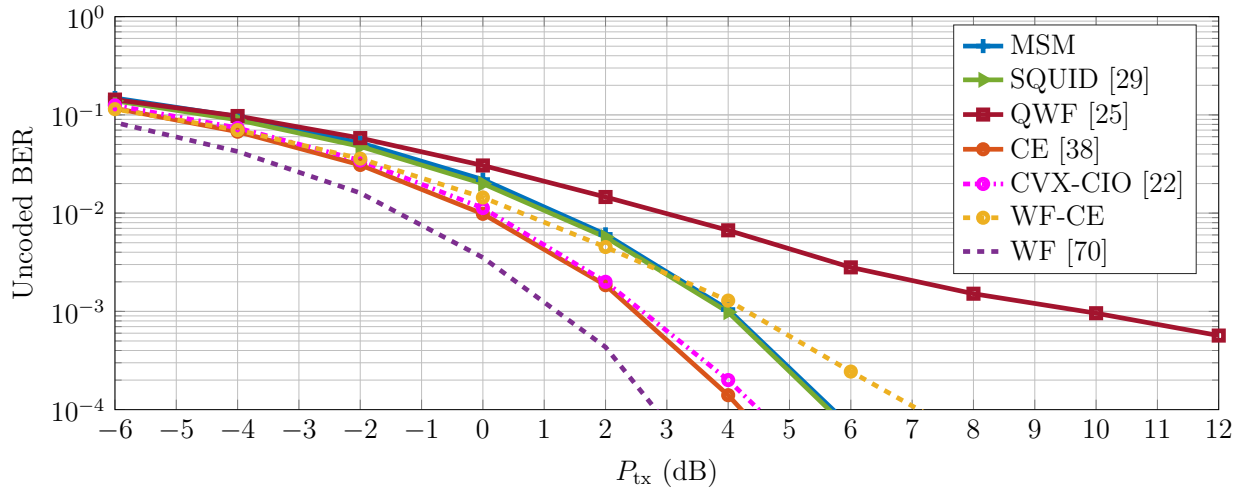


Fig. 7.6: Uncoded BER performance for a MU-MIMO system with $N = 64$ and $M = 8$ for different precoding designs and 4-PSK signaling, © 2018 IEEE.

7.5.3 Computational Complexity of MSM for QAM Signaling

From (7.50), we have $m = 4M+N\,(Q-4)$ inequalities and $n = 2N+2$ variables. The number a of artificial variables reduces to 0, since the vector $\mathbf{b}^{\mathrm{T}} = \left[\mathbf{0}_{4M+1}^{\mathrm{T}} \quad \cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{N(Q-4)}^{\mathrm{T}}\right]$ has only non-negative entries. Thus, the number of operations (multiplication and addition pairs) on each iteration calculates in this case to

$$12M + 3\,(Q-4)\,N, \tag{7.72}$$

or

$$2N + 8MN + 20M + 3 + (Q-4)\left(2N^2 + 5N\right). \tag{7.73}$$

For massive MIMO systems, the number of floating-point operations for each iteration is on the order of

$$\#\text{FLOPs}_{\text{QAM}} = \mathcal{O}\left(8MN + (Q-4)\left(2N^2 + 5N\right)\right). \tag{7.74}$$

For the special case of one-bit quantization, i.e. $Q = 4$, the complexity is linear in N and M. Note that the sparsity of $\mathbf{E}$ can be exploited by deploying the revised simplex method to reduce the number of required operations in the case of $Q > 4$ [72, p.89].

7.6 Simulation Results

For the simulations, we assume a BS with $N = 64$ antennas serving $M = 8$ single-antenna users. The channel $\mathbf{H}$ is composed of i.i.d. Gaussian random variables with zero-mean and unit variance. The numerical results are obtained with Monte Carlo simulations of 100 independent channel realizations. The additive noise is also i.i.d. with variance one at each antenna. The performance metric is the uncoded BER averaged over the single-antenna users. For the blind estimation of the coefficients g_m we use a block length of $T = 128$.

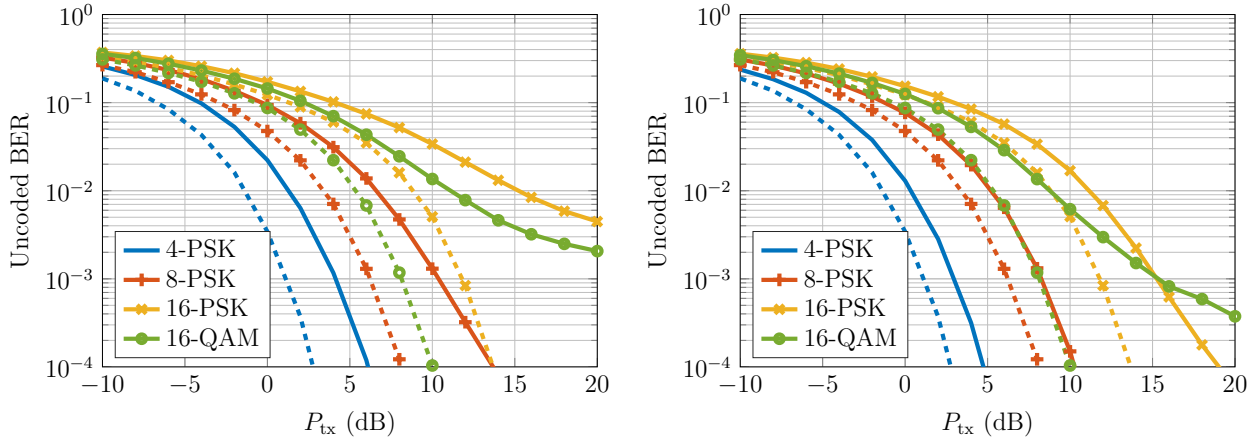
In the first simulation set, depicted in Fig. 7.6, we assume full CSI, choose QPSK modulation and compare the uncoded BER as a function of the transmit power P_{tx} for the following precoders:

- The proposed MSM method with $Q = 4$.
- The SQUID precoder presented in [29] with $Q = 4$, where the precoder design criterion is the symbol-wise MSE between $\mathbf{u}$ and $\mathbf{s}$ under a quantization constraint. The latter is equivalent to the QCE constraint for the special case $Q = 4$. The SQUID precoder is a semi-definite relaxation based algorithm.
- The quantized WF precoder denoted by "QWF" from [25] with $Q = 4$. This precoder design is based on linearizing the quantizer and considering the resulting quantization noise as additive Gaussian noise.
- The CE precoder presented in [38] denoted by "CE [38]", with $Q = \infty$, where the symbol-wise MSE between $\mathbf{y}$ and a scaled version of $\mathbf{s}$ is minimized under the CE constraint. The scaling factor that is applied to $\mathbf{s}$ is jointly optimized.
- The CE precoder from [22] denoted by CVX-CIO that aims at maximizing the constructive interference under the CE constraint.
- The WF precoder followed by the CE quantizer with $Q = \infty$ denoted by "WF-CE", and
- The WF precoder in the ideal case denoted by "The ideal WF" from [70], where neither quantization nor the CE constraint is applied to the transmit signal.

It can be seen that the CE constraint leads to a loss of almost 2 dB at a BER of 10^{-2} compared to the ideal WF and a loss of less than 1.5 dB when using the unquantized symbol-wise precoders proposed in [38] and in [22]. The one-bit quantization, which represents the QCE case of $Q = 4$, leads to more losses that depend on the precoder design. With the use of the linear precoder QWF a loss of more than 4 dB at a BER of 10^{-2} is noticed. However, the non-linear precoders MSM and SQUID improve the performance drastically and show a loss of slightly more than 2 dB compared to the ideal case at the cost of higher computational complexity. Nevertheless, the proposed MSM method is more efficient than SQUID as it is based on a purely linear programming formulation.

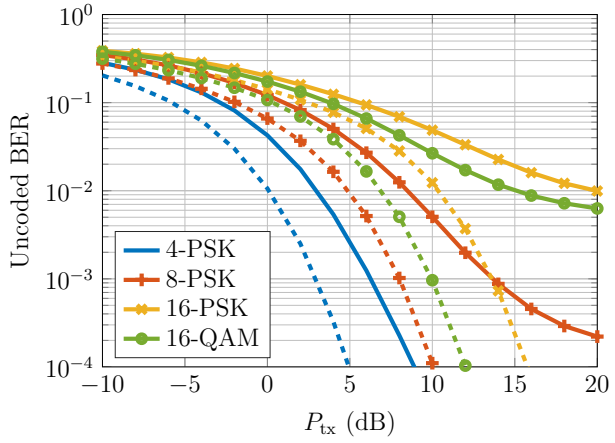
In the second simulation set, depicted in Fig. 7.7, the uncoded BER is plotted as a function of the transmit power P_{tx} using the MSM precoder for different modulation schemes, for the i.i.d. and the mmW ($N_{\text{cl}} = 2$ and $N_{\text{ray}} = 10$) channel models and for two different values of Q: $Q = 4$ and $Q = 8$. Higher values of Q are omitted since the obtained results do not differ much from the case of $Q = 8$. In addition, it is beneficial in terms of computational complexity and power consumption to keep Q as small as possible. The performance obtained by the MSM method is compared to the performance of the ideal linear system. It can be deduced from the plots that the non-linear approach improves the performance much better than the linear approaches discussed in Chapter 5. As expected, the higher the number of symbols in the modulation scheme, the higher the BER for a given P_{tx} value. However, the increase of the DAC resolution q and thus the resulting increase in Q leads to a performance improvement, which depends on the modulation scheme.

In the third simulation set, the same comparison is conducted, but with doubling the number of antennas of the quantized system. The results are plotted in Fig. 7.8. It can be concluded that the performance of the ideal system can be achieved with a QCE, which has less than the double number of antennas. When the quantization resolution is increased, less antennas are required to achieve the same performance as the one of the ideal linear system.

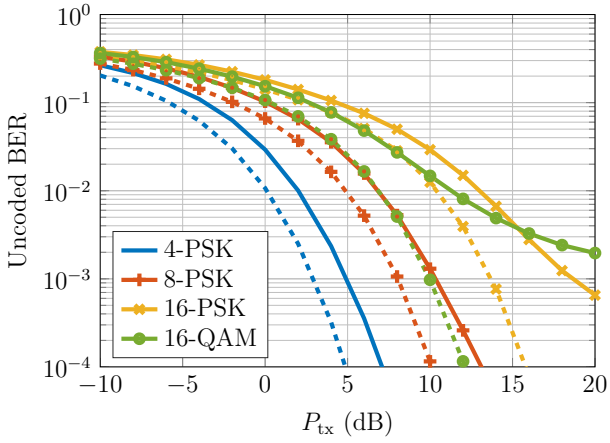


(a) i.i.d. channel model: $Q = 4$

(b) i.i.d. channel model: $Q = 8$



(c) mmW channel model: $Q = 4$

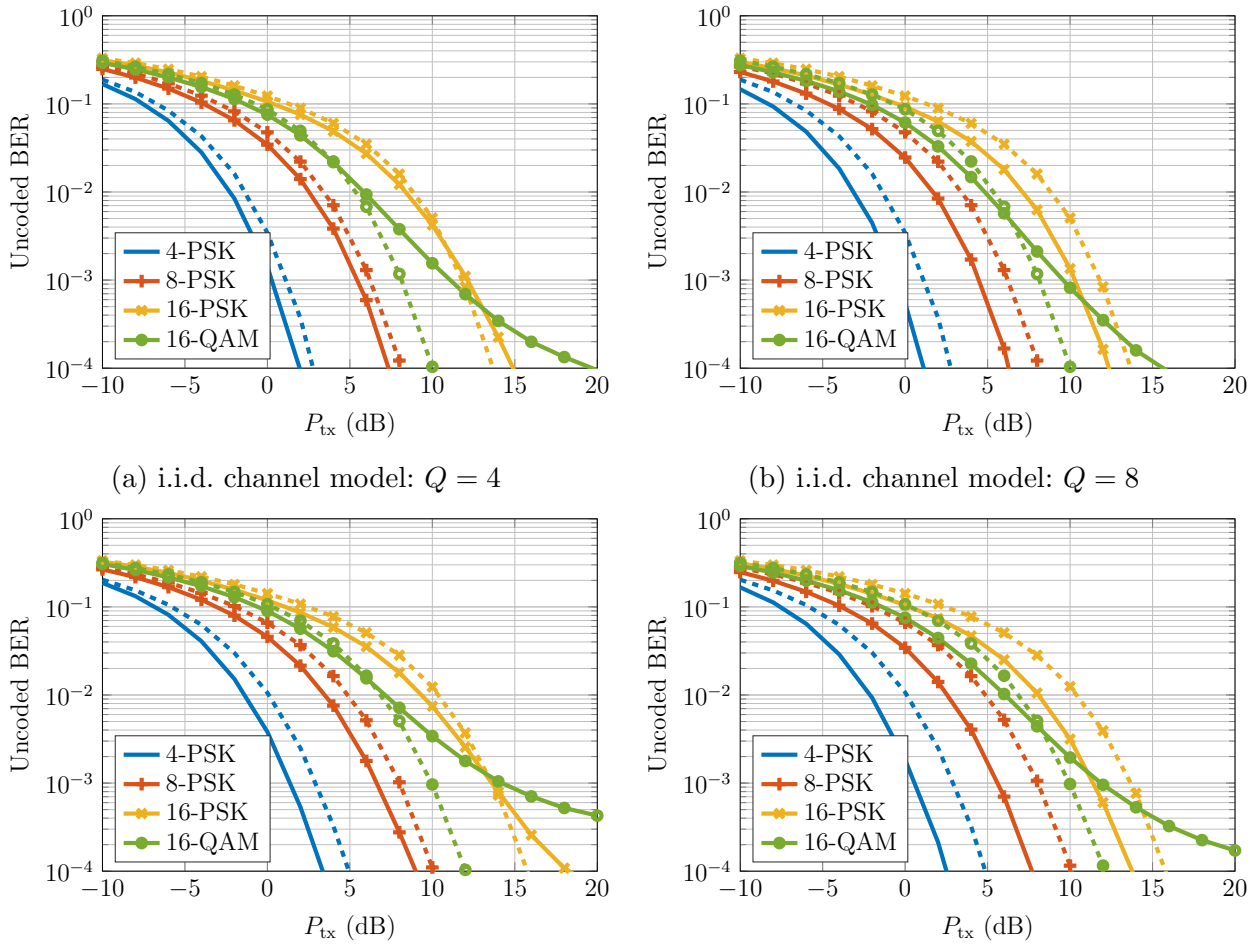


(d) mmW channel model: $Q = 8$

Fig. 7.7: Uncoded BER performance for a MU-MIMO system with $N = 64$ and $M = 8$ for different modulation schemes: MSM (solid lines), the ideal WF (dashed lines).

Since the optimization problem in [22] has some similarities with our proposed MSM, we compare the uncoded BER performance for both designs in Fig. 7.9. In our simulation, we pass the entries of $\mathbf{x}$ obtained by the CVX-CIO method through the CE quantizer to get QCE signals. Additionally, we introduce the method denoted by CVX-CIO-noCE that is the same as CVX-CIO with an instantaneous power constraint instead of the CE constraint. As can be seen from the results, CVX-CIO and CVX-CIO-noCE do not perform optimally under the constraint of QCE transmit signals. However, the loss compared to MSM reduces when the quantization resolution increases.

In the last simulation set, we counted the average number of iterations required by the MSM precoder. The results are summarized in Table 7.4, where we observe that around 50 iterations are required for all modulation schemes for $Q = 4$ and more than 100 iterations for $Q > 4$.



(a) i.i.d. channel model: $Q = 4$

(b) i.i.d. channel model: $Q = 8$

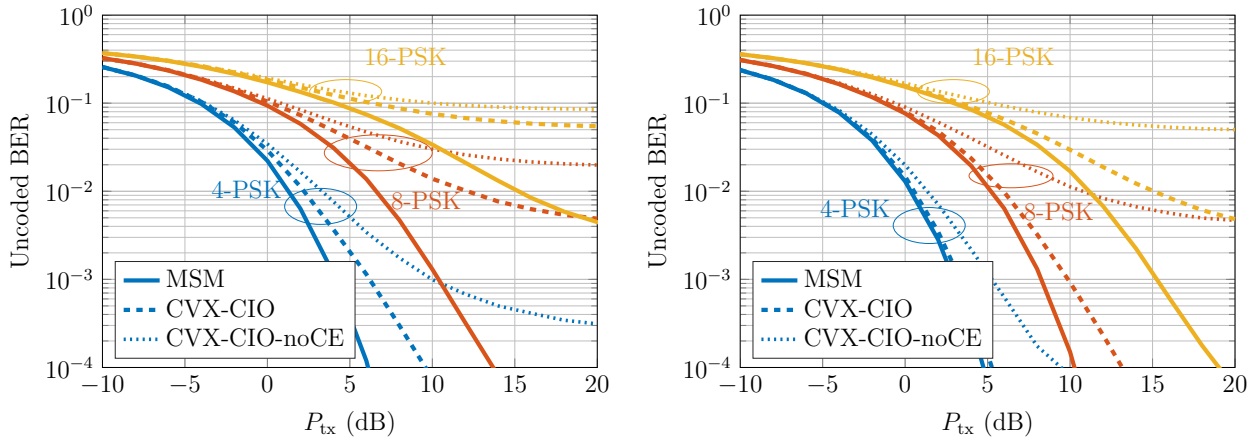
(c) mmW channel model: $Q = 4$

(d) mmW channel model: $Q = 8$

Fig. 7.8: Uncoded BER performance for a MU-MIMO system with $M = 8$ for different modulation schemes: MSM with $N = 128$ (solid lines), the ideal WF with $N = 64$ (dashed lines).

Nb. of iter	$Q = 4$	$Q = 8$	$Q = 16$
4-PSK	45.77	121.05	187.63
8-PSK	50.15	123.91	191.55
16-PSK	54.94	128.74	199.61
16-QAM	43.25	120.42	187.32
64-QAM	43.04	120.30	188.30

Table 7.4: Average number of iterations of MSM for the i.i.d. channel model with $N = 64$ and $M = 8$, © 2018 IEEE.



(a) $Q = 4$ (b) $Q = 8$

Fig. 7.9: Comparison of the uncoded BER performance between MSM, CVX-CIO from [22] and CVX-CIO-noCE for a MU-MIMO system with $N = 64$ and $M = 8$, © 2018 IEEE.

8. Frequency-Selective Channels

8.1 Input-Output Relationship

In the case of frequency-selective channels, the relaxed noiseless received signal vector is given by

$$\mathbf{y}'[t] = \mathbf{H}[t] * \mathbf{x}[t],\ \mathbf{x}[t] \in \mathbb{X}^N. \tag{8.1}$$

To mitigate the resulting ISI, multiple time instants have to be jointly considered in the optimization problem. Therefore, we consider in general signal blocks of length B. To this end, let us define the input signal block as

$$\mathbf{s}_B[t] = \begin{bmatrix}\mathbf{s}[t]^{\mathrm{T}} & \mathbf{s}[t-1]^{\mathrm{T}} & \cdots & \mathbf{s}[t-B+1]^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \tag{8.2}$$

and we aim to design the optimal transmit signal block

$$\mathbf{x}_B[t] = \begin{bmatrix}\mathbf{x}[t]^{\mathrm{T}} & \mathbf{x}[t-1]^{\mathrm{T}} & \cdots & \mathbf{x}[t-B+1]^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}. \tag{8.3}$$

The relaxed noiseless received signal block $\mathbf{y}'_B[t] = \begin{bmatrix}\mathbf{y}'[t]^{\mathrm{T}} & \mathbf{y}'[t-1]^{\mathrm{T}} & \cdots & \mathbf{y}'[t-B+1]^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$ is in general expressed as

$$\mathbf{y}'_B[t] = \mathbf{H}_B\mathbf{x}_B[t] + \mathbf{y}'_{\mathrm{IBI}}[t], \tag{8.4}$$

where $\mathbf{H}_B$ and $\mathbf{y}'_{\mathrm{IBI}}[t]$ denote the corresponding convolution matrix and the vector containing the Inter-Block Interference (IBI) due to the channel frequency-selectivity. The exact expressions depend on the block length B and are given for different block lengths in Section 8.1.1 and Section 8.1.2.

8.1.1 Symbol-wise Processing: MSM-SP

In the case of symbol-wise processing, i.e. $B = 1$, we get

$$\mathbf{y}' = \mathbf{y}'[t] = \mathbf{H}_0\mathbf{x}[t] + \sum_{\ell=1}^{L-1}\mathbf{H}_\ell\mathbf{x}[t-\ell]. \tag{8.5}$$

This leads to the expressions

$$\mathbf{H}_B = \mathbf{H}_0 \qquad \text{and} \qquad \mathbf{y}'_{\mathrm{IBI}}[t] = \sum_{\ell=1}^{L-1}\mathbf{H}_\ell\mathbf{x}[t-\ell]. \tag{8.6}$$

8.1.2 Block-wise Processing

8.1.2.1 Without Cyclic Prefix: MSM-BP

In this section, the precoding task is conducted for a block of length B, with $B \geq L$, such that we get only one interfering block. The noiseless received block signal is given by

$$\mathbf{y}'_B[t] = \mathbf{H}_B \mathbf{x}_B[t] + \mathbf{H}_{\text{IBI}} \mathbf{x}_B[t-B], \tag{8.7}$$

where

$$\mathbf{H}_B = \sum_{\ell=0}^{L-1} \Upsilon_1(B,\ell) \otimes \mathbf{H}_\ell, \tag{8.8}$$

$$\mathbf{H}_{\text{IBI}} = \sum_{\ell=0}^{L-1} \Upsilon_2(B,\ell) \otimes \mathbf{H}_\ell \tag{8.9}$$

$$\Upsilon_1(B,\ell) = \begin{bmatrix} \mathbf{0}_{B-\ell,\ell} & \mathbf{I}_{B-\ell} \\ & \mathbf{0}_{\ell,B} \end{bmatrix} \tag{8.10}$$

$$\Upsilon_2(B,\ell) = \begin{bmatrix} & \mathbf{0}_{B-\ell,B} \\ \mathbf{I}_\ell & \mathbf{0}_{\ell,B-L+\ell} \end{bmatrix}. \tag{8.11}$$

This leads to the following exact expression:

$$\mathbf{y}'_{\text{IBI}}[t] = \mathbf{H}_{\text{IBI}} \mathbf{x}_B[t-B]. \tag{8.12}$$

8.1.2.2 With Cyclic Prefix: MSM-BP-CP

To remove the IBI, we append to each block a cyclic prefix of length $L-1$ and choose a block of length $B \geq L$. Note that due to the cyclic prefix the power per useful transmit vector $\mathbf{t}[t]$ reduces to $P_{\text{tx}} \frac{B}{B+L-1}$. The resulting noiseless received vector is given by

$$\mathbf{y}'_B[t] = \mathbf{H}_{\text{circ}} \mathbf{x}_B[t], \tag{8.13}$$

where

$$\mathbf{H}_{\text{circ}} = \sum_{\ell=0}^{L-1} \left(\Upsilon_1(B,\ell) + \Upsilon_2(B,\ell)\right) \otimes \mathbf{H}_\ell \tag{8.14}$$

is the block-circular channel matrix. Thus, we get

$$\mathbf{H}_B = \mathbf{H}_{\text{circ}} \qquad \text{and} \qquad \mathbf{y}'_{\text{IBI}}[t] = \mathbf{0}_{MB}. \tag{8.15}$$

8.2 Optimization Problem

The MSM precoder that was first introduced in Chapter 7, can be reformulated for frequency-selective channels as

$$\max_{\mathbf{x}_B[t]} \delta \tag{8.16}$$

$$\text{s.t. } y'_{Bm}[t] \in \text{SR}_m, \ m = 1, \cdots, MB \tag{8.17}$$

$$\text{and } \mathbf{x}_B[t] \in \mathbb{X}^{NB}. \tag{8.18}$$

Since the mathematical expressions of the SRs depend on the constellation set $\mathbb{S}$, we introduce the optimization problem of the MSM precoder for frequency-selective channels separately for PSK signaling and QAM signaling.

8.2.1 PSK Signaling

In analogy to the derivations introduced in

$$\begin{aligned}\mathbf{z}_B &= \operatorname{diag}(\mathbf{s}_B[t]^*)\mathbf{y}'_B[t] \\ &= \tilde{\mathbf{H}}_B\mathbf{x}_B[t] + \tilde{\mathbf{y}}'_{\mathrm{IBI}},\end{aligned} \tag{8.19}$$

where

$$\tilde{\mathbf{H}}_B = \operatorname{diag}(\mathbf{s}_B[t]^*)\mathbf{H}_B \tag{8.20}$$

$$\tilde{\mathbf{y}}'_{\mathrm{IBI}} = \operatorname{diag}(\mathbf{s}_B[t]^*)\mathbf{y}'_{\mathrm{IBI}}[t]. \tag{8.21}$$

When using the following real-valued representation:

$$\Re\{\tilde{\mathbf{H}}_B\mathbf{x}_B[t]\} = \underbrace{\begin{bmatrix}\Re\{\tilde{\mathbf{H}}_B\} & -\Im\{\tilde{\mathbf{H}}_B\}\end{bmatrix}}_{=\mathbf{A}_B}\underbrace{\begin{bmatrix}\Re\{\mathbf{x}_B[t]\}\\ \Im\{\mathbf{x}_B[t]\}\end{bmatrix}}_{=\bar{\mathbf{x}}_B} = \mathbf{A}_B\bar{\mathbf{x}}_B \tag{8.22}$$

$$\Im\{\tilde{\mathbf{H}}_B\mathbf{x}_B[t]\} = \underbrace{\begin{bmatrix}\Im\{\tilde{\mathbf{H}}_B\} & \Re\{\tilde{\mathbf{H}}\}\end{bmatrix}}_{=\mathbf{B}_B}\begin{bmatrix}\Re\{\mathbf{x}_b[t]\}\\ \Im\{\mathbf{x}_B[t]\}\end{bmatrix} = \mathbf{B}_B\bar{\mathbf{x}}_B, \tag{8.23}$$

and the following definitions:

$$\tilde{\mathbf{a}} = \Re\{\tilde{\mathbf{y}}'_{\mathrm{IBI}}\} - \Im\{\tilde{\mathbf{y}}'_{\mathrm{IBI}}\} \tag{8.24}$$

$$\tilde{\mathbf{b}} = \Re\{\tilde{\mathbf{y}}'_{\mathrm{IBI}}\} + \Im\{\tilde{\mathbf{y}}'_{\mathrm{IBI}}\}, \tag{8.25}$$

the constraint in (7.17) applied on $\mathbf{z}_B$ from (8.19) can be rewritten as

$$\begin{bmatrix}\mathbf{B}_B - \tan\theta\mathbf{A}_B & \frac{1}{\cos\theta}\mathbf{1}_{MB}\\ -\mathbf{B}_B - \tan\theta\mathbf{A}_B & \frac{1}{\cos\theta}\mathbf{1}_{MB}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{x}}_B\\ \delta\end{bmatrix} \leq \begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix}. \tag{8.26}$$

Combining (8.16), (8.26) and (8.18), the optimization problem MSM for PSK signaling can be expressed for the case $P_{\mathrm{tx}} = N$ by the real-valued linear programming problem

$$\begin{aligned}\max_{\mathbf{v}_B}\begin{bmatrix}\mathbf{0}_{2NB}^{\mathrm{T}} & 1\end{bmatrix}\mathbf{v}_B \text{ s.t. } &\begin{bmatrix}\mathbf{B}_B - \tan\theta\mathbf{A}_B & \frac{1}{\cos\theta}\mathbf{1}_{MB}\\ -\mathbf{B}_B - \tan\theta\mathbf{A}_B & \frac{1}{\cos\theta}\mathbf{1}_{MB}\\ \mathbf{E}\otimes\mathbf{I}_B & \mathbf{0}_{NB(Q-4)}\end{bmatrix}\mathbf{v}_B \leq \begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\\ \cos\left(\frac{\pi}{Q}\right)\mathbf{1}_{NB(Q-4)}\end{bmatrix}\\ \text{and } &\begin{bmatrix}-\cos\left(\frac{\pi}{Q}\right)\mathbf{1}_{2NB}\\ 0\end{bmatrix} \leq \mathbf{v}_B \leq \begin{bmatrix}\cos\left(\frac{\pi}{Q}\right)\mathbf{1}_{2NB}\\ \infty\end{bmatrix},\end{aligned} \tag{8.27}$$

where $\mathbf{v}_B = \begin{bmatrix}\bar{\mathbf{x}}_B^{\mathrm{T}} & \delta\end{bmatrix}^{\mathrm{T}}$.

8.2.2 QAM Signaling

In analogy to the derivations introduced in Section 7.4, the noiseless received signal in the modified coordinate system, shifted by

$$\mathbf{o}_B = \alpha\left(\mathbf{s}_B[t] - \operatorname{sgn}\left(\mathbf{s}_B[t]\right)\right) \tag{8.28}$$

and rotated by the phases of the signals of interest in the shifted coordinate system, i.e.

$$\mathbf{s}_{B(\mathbf{o}_B)}[t] = \alpha \mathbf{s}_B[t] - \mathbf{o}_B, \tag{8.29}$$

is given by

$$\begin{aligned}\mathbf{z}_B &= \frac{1}{\sqrt{2}} \operatorname{diag}\left(\operatorname{sgn}\left(\mathbf{s}_B^*[t]\right)\right)\left(\mathbf{y}'_B[t] - \alpha\left(\mathbf{s}_B[t] - \operatorname{sgn}\left(\mathbf{s}_B[t]\right)\right)\right) \\ &= \hat{\mathbf{H}}_B \mathbf{x}_B[t] + \hat{\mathbf{y}}'_{\mathrm{IBI}} - \alpha \mathbf{c}_B,\end{aligned} \tag{8.30}$$

where

$$\hat{\mathbf{H}}_B = \frac{1}{\sqrt{2}} \operatorname{diag}\left(\operatorname{sgn}\left(\mathbf{s}_B^*[t]\right)\right) \mathbf{H}_B \tag{8.31}$$

$$\hat{\mathbf{y}}'_{\mathrm{IBI}} = \frac{1}{\sqrt{2}} \operatorname{diag}\left(\operatorname{sgn}\left(\mathbf{s}_B^*[t]\right)\right) \mathbf{y}'_{\mathrm{IBI}}[t] \tag{8.32}$$

$$\mathbf{c}_B = \frac{1}{\sqrt{2}} \operatorname{diag}\left(\operatorname{sgn}\left(\mathbf{s}_B^*[t]\right)\right)\left(\mathbf{s}_B[t] - \operatorname{sgn}\left(\mathbf{s}_B[t]\right)\right). \tag{8.33}$$

When using the following real-valued representation:

$$\Re\{\hat{\mathbf{H}}_B \mathbf{x}_B[t]\} = \underbrace{\begin{bmatrix}\Re\{\hat{\mathbf{H}}_B\} & -\Im\{\hat{\mathbf{H}}_B\}\end{bmatrix}}_{=\mathbf{V}_B} \begin{bmatrix}\Re\{\mathbf{x}_B[t]\} \\ \Im\{\mathbf{x}_B[t]\}\end{bmatrix} = \mathbf{V}_B \bar{\mathbf{x}}_B \tag{8.34}$$

$$\Im\{\hat{\mathbf{H}}_B \mathbf{x}_B[t]\} = \underbrace{\begin{bmatrix}\Im\{\hat{\mathbf{H}}_B\} & \Re\{\hat{\mathbf{H}}_B\}\end{bmatrix}}_{=\mathbf{W}_B} \begin{bmatrix}\Re\{\mathbf{x}_B[t]\} \\ \Im\{\mathbf{x}_B[t]\}\end{bmatrix} = \mathbf{W}_B \bar{\mathbf{x}}_B, \tag{8.35}$$

and the following definitions:

$$\hat{\mathbf{a}} = \Re\{\hat{\mathbf{y}}'_{\mathrm{IBI}}\} - \Im\{\hat{\mathbf{y}}'_{\mathrm{IBI}}\} \tag{8.36}$$

$$\hat{\mathbf{b}} = \Re\{\hat{\mathbf{y}}'_{\mathrm{IBI}}\} + \Im\{\hat{\mathbf{y}}'_{\mathrm{IBI}}\}, \tag{8.37}$$

the constraint in (7.9) applied on $\mathbf{z}_B$ from (8.30) can be rewritten as

$$\begin{bmatrix}\mathbf{W}_B - \mathbf{V}_B & \mathbf{1}_{MB} & \Re\{\mathbf{c}_B\} - \Im\{\mathbf{c}_B\} \\ -\mathbf{W}_B - \mathbf{V}_B & \mathbf{1}_{MB} & \Re\{\mathbf{c}_B\} + \Im\{\mathbf{c}_B\} \\ \mathbf{W}_B + \mathbf{V}_B & \mathbf{1}_{MB} & -\Re\{\mathbf{c}_B\} - \Im\{\mathbf{c}\} - \sqrt{2}\boldsymbol{\xi}_2 \\ -\mathbf{W}_B + \mathbf{V}_B & \mathbf{1}_{MB} & -\Re\{\mathbf{c}_B\} + \Im\{\mathbf{c}_B\} - \sqrt{2}\boldsymbol{\xi}_1\end{bmatrix} \begin{bmatrix}\bar{\mathbf{x}}_B \\ \sqrt{2}\delta \\ \alpha\end{bmatrix} \leq \begin{bmatrix}\hat{\mathbf{a}} \\ \hat{\mathbf{b}} \\ -\hat{\mathbf{b}} \\ -\hat{\mathbf{a}}\end{bmatrix}. \tag{8.38}$$

By adding the first line to the fourth and the second line to the third, the latter constraint recalculates to

$$\begin{bmatrix}\mathbf{W}_B - \mathbf{V}_B & \mathbf{1}_{MB} & \Re\{\mathbf{c}_B\} - \Im\{\mathbf{c}_B\} \\ -\mathbf{W}_B - \mathbf{V}_B & \mathbf{1}_{MB} & \Re\{\mathbf{c}_B\} + \Im\{\mathbf{c}_B\} \\ \mathbf{0}_{MB,2NB} & \sqrt{2}\,\mathbf{1}_{MB} & -\boldsymbol{\xi}_2 \\ \mathbf{0}_{MB,2NB} & \sqrt{2}\,\mathbf{1}_{MB} & -\boldsymbol{\xi}_1\end{bmatrix} \begin{bmatrix}\bar{\mathbf{x}}_B \\ \sqrt{2}\delta \\ \alpha\end{bmatrix} \leq \begin{bmatrix}\hat{\mathbf{a}} \\ \hat{\mathbf{b}} \\ \mathbf{0}_{MB} \\ \mathbf{0}_{MB}\end{bmatrix}. \tag{8.39}$$

Combining (8.16), (8.39) and (8.18), the optimization problem of the MSM precoder for QAM signaling can be then for $P_{\text{tx}} = N$ expressed by

$$\begin{aligned}
&\max_{\mathbf{v}_B} \begin{bmatrix} \mathbf{0}_{2NB}^{\mathrm{T}} & 1 & 0 \end{bmatrix} \mathbf{v}_B \\
&\text{s.t.} \begin{bmatrix} \mathbf{W}_B - \mathbf{V}_B & \mathbf{1}_{MB} & \Re\{\mathbf{c}_B\} - \Im\{\mathbf{c}_B\} \\ -\mathbf{W}_B - \mathbf{V}_B & \mathbf{1}_{MB} & \Re\{\mathbf{c}_B\} + \Im\{\mathbf{c}_B\} \\ \mathbf{0}_{MB,2NB} & \sqrt{2}\,\mathbf{1}_{MB} & -\boldsymbol{\xi}_2 \\ \mathbf{0}_{MB,2NB} & \sqrt{2}\,\mathbf{1}_{MB} & -\boldsymbol{\xi}_1 \\ \mathbf{E} \otimes \mathbf{I}_B & \mathbf{0}_{NB(Q-4)} & \mathbf{0}_{NB(Q-4)} \end{bmatrix} \mathbf{v}_B \leq \begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{b}} \\ \mathbf{0}_{MB} \\ \mathbf{0}_{MB} \\ \cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{NB(Q-4)} \end{bmatrix} \\
&\text{and} \begin{bmatrix} -\cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{2NB} \\ 0 \\ 0 \end{bmatrix} \leq \mathbf{v}_B \leq \begin{bmatrix} \cos\left(\frac{\pi}{Q}\right) \mathbf{1}_{2NB} \\ \infty \\ \infty \end{bmatrix},
\end{aligned} \tag{8.40}$$

where $\mathbf{v}_B^{\mathrm{T}} = \begin{bmatrix} \bar{\mathbf{x}}_B^{\mathrm{T}} & \sqrt{2}\delta & \alpha \end{bmatrix}$.

8.3 Computational Complexity

We recall the study of the computational complexity of linear programs in Section 7.5.1. Note that the sparsity of the linear program in (8.27) and (8.40) can be exploited by deploying the revised simplex method to reduce the computational complexity [72, p.89].

8.3.1 PSK

From (8.27), we have $m = 2MB + NB\,(Q-4)$ inequalities and $n = 2NB + 1$ variables. The number a of artificial variables is defined by the number of non-negative entries in $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. We introduce the ratio ε to denote the percentage of non-negative entries in $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. It holds that $a = 2MB\varepsilon$. Note that $\varepsilon = 0$ for MSM-BP-CP due to (8.15).

According to Section 7.5.1, the number of floating-point operations for each iteration, under the assumption of massive MIMO systems, is on the order of

$$\#\text{FLOPs}_{\text{PSK}} = \mathcal{O}\left(4MNB^2 + (Q-4)\left(2N^2B^2 + 2NMB^2\varepsilon\right)\right). \tag{8.41}$$

For the special case of one-bit quantization, i.e. $Q = 4$, the complexity is linear in N and M but quadratic in B. For $Q > 4$, the complexity becomes quadratic in N and B.

8.3.2 QAM

From (8.40), we have $m = 4MB + NB\,(Q-4)$ inequalities and $n = 2NB + 2$ variables. The number a of artificial variables is defined by the number of non-negative entries in $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$. We introduce the ratio ε to denote the percentage of non-negative entries in $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$. It holds that $a = 2MB\varepsilon$. Note that $\varepsilon = 0$ for MSM-BP-CP due to (8.15).

According to Section 7.5.1, the number of floating-point operations for each iteration, under the assumption of massive MIMO systems, is on the order of

$$\#\text{FLOPs}_{\text{QAM}} = \mathcal{O}\left(8MNB^2 + (Q-4)\left(2N^2B^2 + 2NMB^2\varepsilon\right)\right). \tag{8.42}$$

For the special case of one-bit quantization, i.e. $Q = 4$, the complexity is linear in N and M but quadratic in B. For $Q > 4$, the complexity becomes quadratic in N and B.

8.4 Simulation Results

In this section, we compare the performance of the different linear precoding techniques that were introduced above with the ideal WF precoder while assuming full CSI. We denote the precoders by

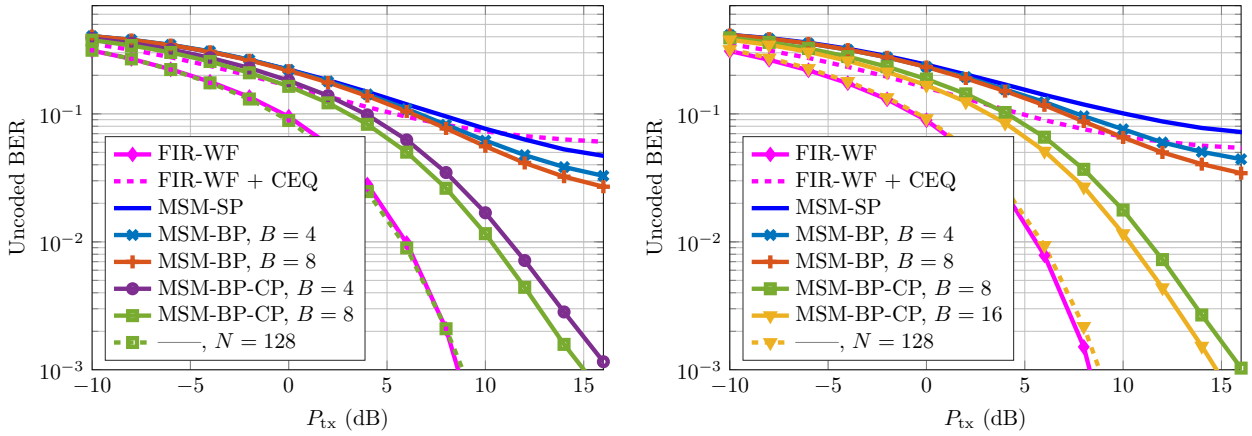
For the simulations, we consider the i.i.d. channel model with the exponential power delay profiles of $L = 3$ and $L = 6$ given in Table 4.1 and Table 4.2. Assuming full CSI, we compare the performance of the different precoding techniques that were introduced in this chapter with the ideal FIR-WF precoder of length L introduced in [73]. We denote the precoders by

- FIR-WF: the ideal FIR-WF, [70], where no CEQ is applied in the system.
- FIR-WF + CEQ: the ideal FIR-WF, that is followed by the optimal CEQ.
- MSM-SP: the MSM precoder based on symbol-wise processing as introduced in Section 8.1.1.
- MSM-BP: the MSM precoder based on block-wise processing as introduced in Section 8.1.2.1.
- MSM-BP-CP: the MSM precoder based on block-wise processing with the cyclic prefix as introduced in Section 8.1.2.2.

We assume a BS with $N = 64$ antennas serving $M = 8$ single-antenna users with 16-QAM signals. The numerical results are obtained with Monte Carlo simulations of 100 independent frequency-selective channel realizations. The AWGN is also i.i.d. with variance one at each antenna. The performance metric is the uncoded BER averaged over the single-antenna users. For the blind estimation of the coefficients g_m we use a block length of $T = 128$. The numerical results are plotted in Fig. 8.1a and Fig. 8.1b for $Q = 4$. It can be deduced, that the smaller the block length B, the worse the performance is. The MSM-SP performs the worst. When performing block-wise processing with $B = 4$ and $B = 8$, the gain in terms of the uncoded BER is moderate, whereas the addition of the cyclic prefix in the block-wise processing leads to a significant gain. This gain is due to the fact that the cyclic prefix removes the IBI, and comes with decreased computational complexity for the same block length and $Q > 4$. However, it comes at the cost of a decrease in the throughput, since $B/(B+L-1)$ is the percentage of the useful transmission. In every block transmission there are $L-1$ redundant (unuseful) symbols due to the cyclic prefix. Increasing the block length decreases the IBI and increases the throughput. However, it increases the computational complexity as discussed in Section 8.3. Thus, there is a trade-off between the performance, the throughput and the computational complexity.

It can be also noticed that the QCE system can achieve the same performance as the ideal linear system with the method MSM-BP-CP with an optimal value of B and by doubling the number of antennas.

In Fig. 8.2, the effect of increasing the resolution of the polar DACs is studied. We choose the precoder MSM-BP-CP with a block length of $B = 16$ and plot the uncoded BER as a function of the available power P_{tx} for different values of Q. As expected the performance improves for higher resolution. The performance difference between $Q = 8$ and $Q = 16$ is very



(a) $L = 3$ (b) $L = 6$

Fig. 8.1: BER performance for a MU-MIMO system with $N = 64$, $M = 8$ and $Q = 4$ with 16-QAM and the i.i.d. channel model.

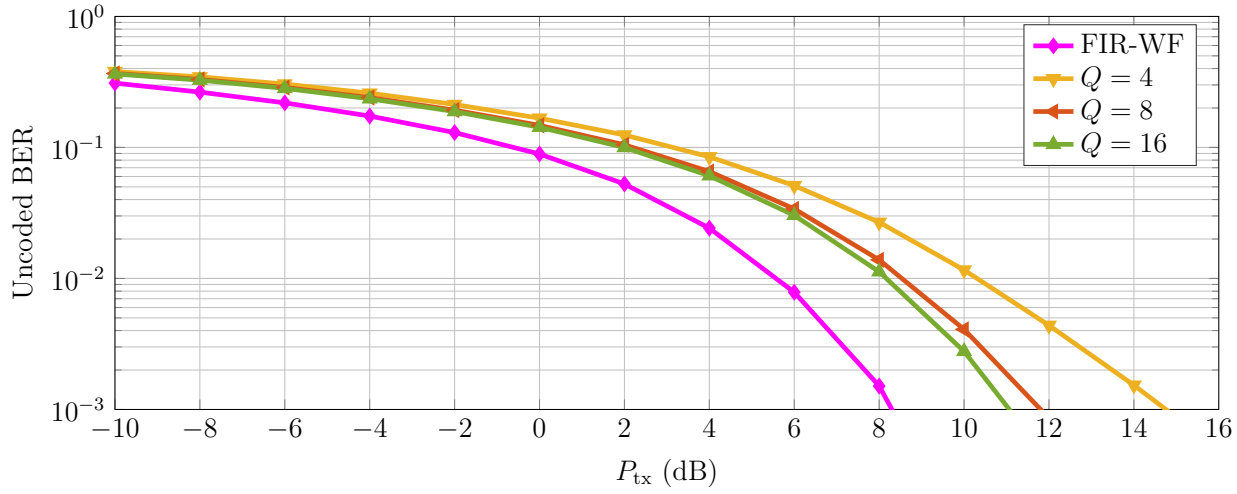


Fig. 8.2: BER performance for a MU-MIMO system with $N = 64$, $M = 8$ with 16-QAM signaling and channel exponential power delay profile with $L = 6$ for different quantization resolutions: MSM-BP-CP with $B = 16$

moderate compared to the required additional computational complexity and the resulting higher power consumption.

In summary, the proposed MSM method offers significant gains in terms of the uncoded BER for the case of flat-fading channels as well as for frequency-selective channels. It is however characterized by its high computational complexity especially for frequency-selective channels, where block-wise processing is required. This challenge can be overcome by the steady enhancement of the hardware computational power.

9. Discussion: Benefits and Challenges

Throughout this work, we are justifying the use of massive MIMO systems with CEQs by their high power efficiency. Therefore, we show in Section 9.1 a comparison between an ideal linear system and a quantized system in terms of power consumption, which proves the benefit of coarsely QCE systems. However, a strong argument againt coarse quantization at the transmitter is the spectral regrowth. Therefore, we look deeper in Section 9.2 at the obtained spectrum in the presence of coarse quantization in massive MIMO systems.

9.1 Power Efficiency

As shown in the previous chapters, the same performance of an ideal linear system, i.e. linear PAs and infinite resolution DACs, can be obtained with a QCE system having at most the double number of antennas as in the ideal case. We compare the power consumed at the PAs in both cases.

A linear PA has a given power efficiency denoted by η_{LPA}. However, only a fraction of the ideal power efficiency η_{LPA} can be achieved, if the input signal at the PA shows large PAPR, which is the case for systems without CE constraint at the PA input. In average each antenna has a transmit power of P_{tx}/N, thus the dissipated power at all PAs in the ideal system is given by

$$P_{\text{DC}_i} = N\frac{P_{\text{tx}}}{N}\frac{1}{\eta_{\text{LPA}}} = \frac{P_{\text{tx}}}{\eta_{\text{LPA}}}, \tag{9.1}$$

whereas the dissipated power in a QCE system, having the same BER performance for the same transmit power P_{tx}, is given by

$$P_{\text{DC}_Q} = 2N\frac{P_{\text{tx}}}{2N}\frac{1}{\eta_{\text{NLPA}}} = \frac{P_{\text{tx}}}{\eta_{\text{NLPA}}}, \tag{9.2}$$

where η_{NLPA} denotes the power efficiency of the PA deployed in the QCE system. Since the input to the PA in this case is of CE, there is no need to use linear PAs. Switching PAs that show high power efficiency are then suitable. A recent PA that operates at the saturation region and has a power efficiency of more than 90% is the PA class M [51]. Since QCE signals have a PAPR value equal to 1, the PA in the QCE case can be operated at its highest power efficiency.

By assuming that the linear PA achieves at most 20% of power efficiency, the ratio between

the dissipated powers in the ideal case and the QCE case is given by

$$\frac{P_{\mathrm{DC}_i}}{P_{\mathrm{DC}_Q}} = \frac{\eta_{\mathrm{NLPA}}}{\eta_{\mathrm{LPA}}} = \frac{0.9}{0.2} = 4.5. \tag{9.3}$$

This means that the power dissipated at the PAs in the ideal case is more than 4 times the power dissipated in the QCE case. In this comparison, the power consumption related to other hardware components like up-converters, is ignored. However, we justify this analysis by the fact that the PA is the most power hungry device in the BS. Note that the DAC in the quantized case is of very low resolution, which first leads to reduced power consumption of the DAC itself and second simplifies the surrounding circuitry.

9.2 Strong Non-linearities vs. Spectral Shaping

The coarse quantization is usually related to the potential deterioration of the pulse shaping in the time domain and thus the increase in bandwidth and the introduction of undesired out of band radiations in the frequency domain. This is true in the case of one transmit antenna and Single-User (SU) scenario. However, does this statement hold true for MU massive MIMO systems? In the sequel, we restrict our analysis to the case of $Q = 4$ for simplicity. The analysis for the case of $Q = 8$ is left for future work.

To check whether the BS transmits signals that respect a specified spectral mask, a measuring device is put in the far field of it and measures the Power Spectral Density (PSD) of the received signal. Therefore, we assume the measuring device as being the virtual $(M+1)$-th user and its continuous-time noiseless received signal is denoted by $y_{M+1}(t')$, which is expressed as

$$y_{M+1}(t') = \mathbf{h}_{M+1}^{\mathrm{T}} \mathbf{t}'(t'), \tag{9.4}$$

where $\mathbf{h}_{M+1}$ denotes the channel vector between the BS antennas and the measuring device. $\mathbf{h}_{M+1}$ is a random vector that is drawn from the same distribution as the user channels. Moreover, $\mathbf{t}'(t')$ denotes the continuous-time transmit signal. The vector $\mathbf{t}'(t')$ is obtained by converting the discrete-time signal $\mathbf{t}[t]$, i.e. the CEQ output signal, to the analog world. In the following, we will differentiate between two cases

- standard QAM staggering
- offset QAM staggering.

9.2.1 Standard QAM

In this case, the continuous-time signal $\mathbf{t}'(t')$ is obtained by

$$\mathbf{t}'(t') = \sum_{t=0}^{\infty} \mathbf{t}[t] g_{\mathrm{DAC}}(t' - tT_{\mathrm{s}}), \tag{9.5}$$

where T_{s} represents the sample period and $g_{\mathrm{DAC}}(t')$ is a real-valued impulse response. Since we are interested in the PSD at the measuring device, we first compute the autocorrelation

function of the signal $y_{M+1}(t')$.

$$\begin{aligned}\rho_{y_{M+1}y_{M+1}}(\tau') &= \frac{\mathrm{E}\left[y_{M+1}(t')y^*_{M+1}(t'-\tau')\right]}{\sigma^2_{y_{M+1}}} \\ &= \frac{\mathbf{h}^{\mathrm{T}}_{M+1}\,\mathrm{E}\left[\mathbf{t}'(t')\mathbf{t}'^{\mathrm{H}}(t'-\tau')\right]\mathbf{h}^*_{M+1}}{\mathbf{h}^{\mathrm{T}}_{M+1}\,\mathrm{E}\left[\mathbf{t}'(t')\mathbf{t}'^{\mathrm{H}}(t')\right]\mathbf{h}^*_{M+1}}.\end{aligned} \tag{9.6}$$

The covariance matrix $\mathrm{E}\left[\mathbf{t}'(t')\mathbf{t}'^{\mathrm{H}}(t'-\tau')\right]$ is given by

$$\begin{aligned}\mathrm{E}\left[\mathbf{t}'(t')\mathbf{t}'^{\mathrm{H}}(t'-\tau')\right] &= \sum_{k=0}^{\infty}\sum_{\ell'=-\infty}^{k} g_{\mathrm{DAC}}(t'-kT_{\mathrm{s}})g_{\mathrm{DAC}}(t'-kT_{\mathrm{s}}-(\tau'-\ell'T_{\mathrm{s}}))\,\mathrm{E}\left[\mathbf{t}[k]\mathbf{t}^{\mathrm{H}}[k-\ell']\right] \\ &= \sum_{\ell'=-\infty}^{\infty}\mathbf{C_{tt}}[\ell']\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau'-\ell'T_{\mathrm{s}}) \\ &= \left(\mathbf{C_{tt}} * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)(\tau').\end{aligned} \tag{9.7}$$

Therefore, we get

$$\rho_{y_{M+1}y_{M+1}}(\tau') = \frac{\mathbf{h}^{\mathrm{T}}_{M+1}\left(\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right) * \mathbf{C_{tt}}\right)(\tau')\mathbf{h}^*_{M+1}}{\mathbf{h}^{\mathrm{T}}_{M+1}\left(\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right) * \mathbf{C_{tt}}\right)(0)\mathbf{h}^*_{M+1}}. \tag{9.8}$$

9.2.1.1 Without Digital Pulse Shaping

If no digital pulse shaping is applied to the precoder output, then there is no time correlation for the signal $\mathbf{t}$; that is

$$\mathbf{C_{tt}}[\tau] = \begin{cases}\mathbf{C_{tt}}, & \text{if } \tau = 0, \\ \mathbf{0}_{N,N}, & \text{otherwise.}\end{cases} \tag{9.9}$$

Therefore, the autocorrelation function simplifies to

$$\begin{aligned}\rho_{y_{M+1}y_{M+1}}(\tau') &= \frac{\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau')}{\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(0)}\frac{\mathbf{h}^{\mathrm{T}}_{M+1}\mathbf{C_{tt}}\mathbf{h}^*_{M+1}}{\mathbf{h}^{\mathrm{T}}_{M+1}\mathbf{C_{tt}}\mathbf{h}^*_{M+1}} \\ &= \frac{\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau')}{\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(0)}.\end{aligned} \tag{9.10}$$

The PSD at the measuring device is then given by

$$\begin{aligned}S_{y_{M+1}y_{M+1}}(f) &= \int_{-\infty}^{\infty}\rho_{y_{M+1}y_{M+1}}(\tau')\,\mathrm{e}^{-\mathrm{j}\,2\pi f\tau'}\,\mathrm{d}\tau' \\ &= \frac{\left|G_{\mathrm{DAC}}(f)\right|^2}{\int_{-\infty}^{\infty}\left|G_{\mathrm{DAC}}(f)\right|^2\mathrm{d}f}.\end{aligned} \tag{9.11}$$

Thus, the shape of the PSD in this case is determined by the transfer function of the DAC and is independent of the number of antennas N and users M. By choosing the time impulse

response of the DAC as the rectangular function,

$$g_{\mathrm{DAC}}(t') = \mathrm{rect}\left(\frac{t'}{T_\mathrm{s}}\right) = \begin{cases} 1 & ,\text{if } |t'| < \frac{T_\mathrm{s}}{2} \\ \frac{1}{2} & ,\text{if } |t'| = \frac{T_\mathrm{s}}{2} \\ 0 & ,\text{otherwise,} \end{cases} \tag{9.12}$$

and thus the frequency transfer function

$$G_{\mathrm{DAC}}(f) = \frac{\sin(\pi f/T_\mathrm{s})}{\pi f/T_\mathrm{s}}, \tag{9.13}$$

we get the PSD depicted in Fig. 9.1 at the measuring device. Note that the sample period T_s is equal to the symbol period T_sym in the case of no digital pulse shaping.

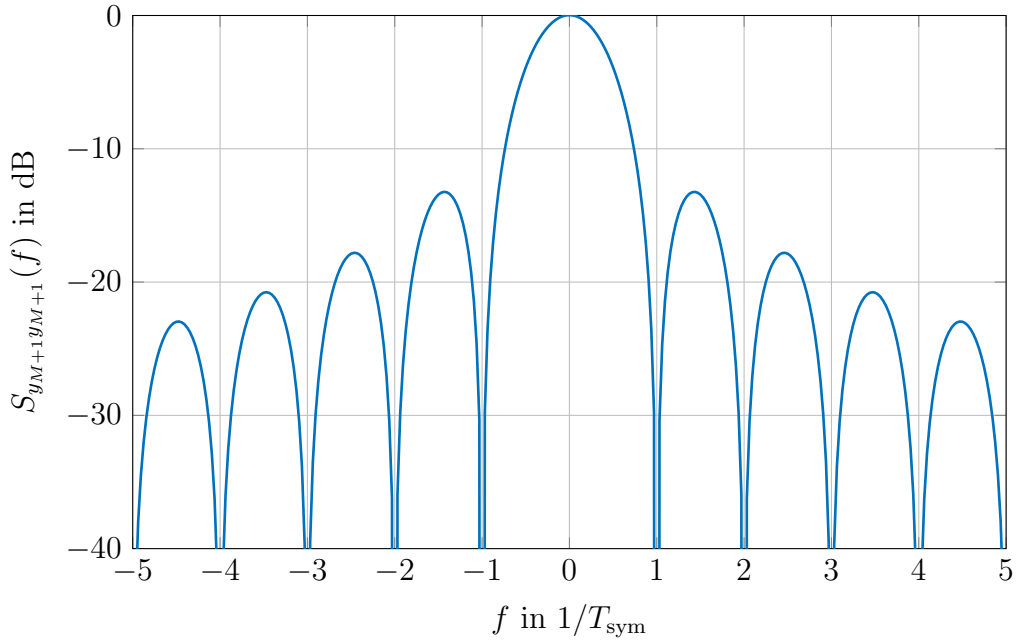


Fig. 9.1: PSD at the measuring device with the DAC as the rectangular function in standard QAM staggering and without digital pulse shaping.

9.2.1.2 With Digital Pulse Shaping

We assume that the input of the CEQ $\mathbf{x}[t]$ is given by

$$\begin{aligned} \mathbf{x}[t] &= g_{\mathrm{PS}}[t] * \mathbf{x}'[t] \\ &= \sum_{\ell=0}^{\infty} \mathbf{x}'[\ell] g_{\mathrm{PS}}[t-\ell], \end{aligned} \tag{9.14}$$

where

$$g_{\mathrm{PS}}[t] = \sum_{t=-\infty}^{\infty} g_{\mathrm{PS}}(t')\delta(t' - tT_\mathrm{s}), T_\mathrm{s} = \frac{T_\mathrm{sym}}{2}, \tag{9.15}$$

and $\mathbf{x}'[t]$ is the output of the precoder. The Discrete Time Fourier Transform (DTFT) of $g_{\mathrm{PS}}[t]$, leads to the transfer function $G_{\mathrm{PS}}(\mathrm{e}^{\mathrm{j}\,2\pi f T_{\mathrm{s}}})$, which is continuous and periodic.
The covariance matrix at the input of the CEQ calculates to

$$\begin{aligned}\mathbf{C}_{\mathbf{xx}}[t] &= \sum_{\ell'=-\infty}^{\infty} \mathbf{C}_{\mathbf{x}'\mathbf{x}'}[\ell'] \left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[t-\ell'] \\ &= \mathbf{C}_{\mathbf{x}'\mathbf{x}'} \left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[t]. \end{aligned} \tag{9.16}$$

Hence, the covariance matrix $\mathbf{R}_{\mathbf{xx}}[\tau]$ calculates to

$$\begin{aligned}\mathbf{R}_{\mathbf{xx}}[\tau] &= \mathrm{diag}(\mathbf{C}_{\mathbf{xx}})^{-1/2}\, \mathrm{E}\left[\mathbf{x}[t]\mathbf{x}^{\mathrm{H}}[t-\tau]\right] \mathrm{diag}(\mathbf{C}_{\mathbf{xx}})^{-1/2} \\ &= \frac{\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[\tau]}{\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[0]} \left(\mathrm{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2}\, \mathrm{E}\left[\mathbf{x}'\mathbf{x}'^{\mathrm{H}}\right] \mathrm{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2}\right) \\ &= \frac{\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[\tau]}{\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[0]} \mathbf{R}_{\mathbf{x}'\mathbf{x}'}. \end{aligned} \tag{9.17}$$

Consequently, the autocorrelation function of the signal $y_{M+1}(t')$ after applying Price's Theorem in (3.52) reads as

$$\begin{aligned}\rho_{y_{M+1}y_{M+1}}(\tau) =& \left(\mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \left(\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right) * \arcsin\left(\Re\left\{ \mathbf{R}_{\mathbf{xx}}\, \mathrm{e}^{-\mathrm{j}\,2\Delta k\psi} \right\} \right) \right)(0) \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \right)^{-1} \\ & \mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}} * \arcsin\left(\Re\left\{ \mathbf{R}_{\mathbf{xx}}[\tau]\, \mathrm{e}^{-\mathrm{j}\,2\Delta k\psi} \right\} \right) \right)(\tau) \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \\ =& \mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \\ & \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}} * \arcsin\left(\Re\left\{ \frac{\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)}{\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[0]} \mathbf{R}_{\mathbf{x}'\mathbf{x}'}\, \mathrm{e}^{-\mathrm{j}\,2\Delta k\psi} \right\} \right) \right)(\tau) \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \\ & \left(\mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \right. \\ & \left. \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}} * \arcsin\left(\Re\left\{ \frac{\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)}{\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[0]} \mathbf{R}_{\mathbf{x}'\mathbf{x}'}\, \mathrm{e}^{-\mathrm{j}\,2\Delta k\psi} \right\} \right) \right)(0) \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \right)^{-1}. \end{aligned} \tag{9.18}$$

The PSD is then obtained by transforming the autocorrelation function into the frequency domain. Thus, we get

$$
\begin{aligned}
S_{y_{M+1}y_{M+1}}(f) &= \int_{-\infty}^{\infty} \rho_{y_{M+1}y_{M+1}}(\tau')\, \mathrm{e}^{-\mathrm{j}\,2\pi f\tau'}\, \mathrm{d}\tau', \\
&= \Bigg(\mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)} \\
&\quad \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}} * \arcsin\left(\Re\left\{ \frac{(g_{\mathrm{PS}} * g_{\mathrm{PS}})}{(g_{\mathrm{PS}} * g_{\mathrm{PS}})\,[0]} \mathbf{R}_{\mathbf{x}'\mathbf{x}'}\, \mathrm{e}^{-\mathrm{j}\,2\Delta k\psi} \right\} \right) \right)(0) \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \Bigg)^{-1} \\
&\quad |G_{\mathrm{DAC}}(f)|^2 \sum_{\Delta k=0}^{Q/2-1} \mathrm{e}^{\mathrm{j}(2\Delta k\psi)}\, \mathrm{tr}\Big(\boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \\
&\quad \sum_{\tau=-\infty}^{\infty} \arcsin\left(\Re\left\{ \frac{(g_{\mathrm{PS}} * g_{\mathrm{PS}})\,[\tau]}{(g_{\mathrm{PS}} * g_{\mathrm{PS}})\,[0]} \mathbf{R}_{\mathbf{x}'\mathbf{x}'}\, \mathrm{e}^{-\mathrm{j}\,2\Delta k\psi} \right\} \right) \mathrm{e}^{-\mathrm{j}\,2\pi f\tau} \Big). \qquad (9.19)
\end{aligned}
$$

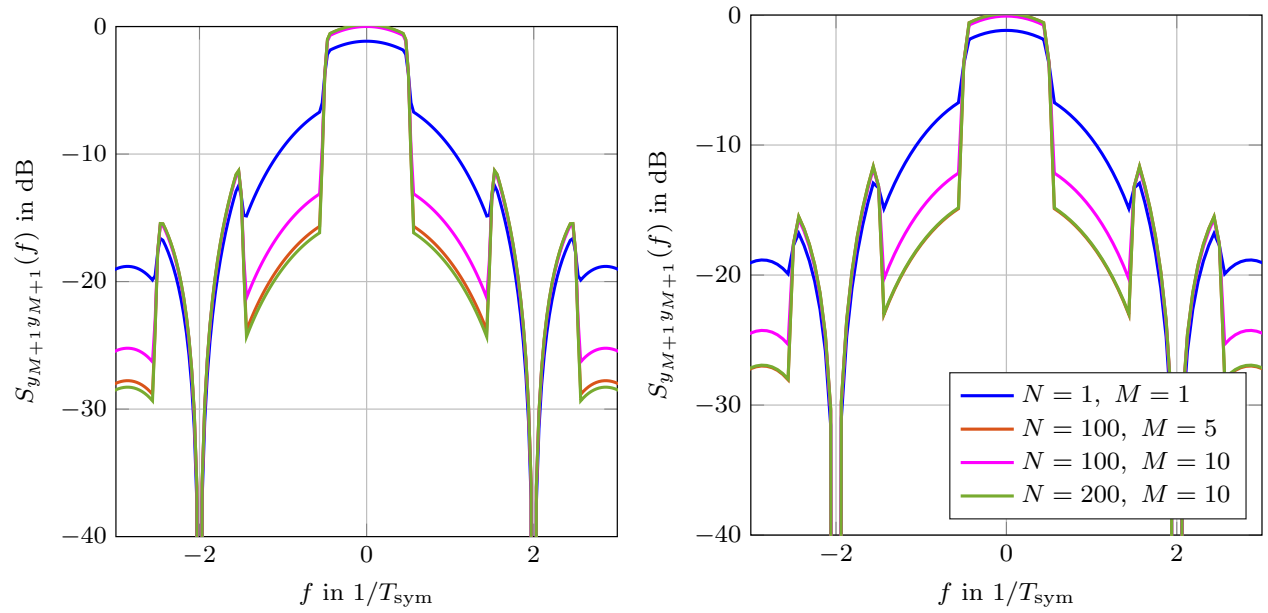
In the case that the channels of the users show small correlation, we can assume that the non-diagonal entries of $\mathbf{R}_{\mathbf{x}'\mathbf{x}'}$ are small enough to use the first order Taylor approximation of the arcsin function. According to Section 3.6, we get

$$
\begin{aligned}
S_{y_{M+1}y_{M+1}}(f) &\approx |G_{\mathrm{DAC}}(f)|^2 \\
&\quad \left(\mathbf{h}_{M+1}^{\mathrm{T}} \left(\alpha_q \frac{(g_{\mathrm{DAC}} * g_{\mathrm{DAC}} * g_{\mathrm{PS}} * g_{\mathrm{PS}})\,(0)}{(g_{\mathrm{PS}} * g_{\mathrm{PS}})\,[0]} \boldsymbol{\Xi} \mathbf{R}_{\mathbf{x}'\mathbf{x}'} \boldsymbol{\Xi} + \beta_q\, (g_{\mathrm{DAC}} * g_{\mathrm{DAC}})\,(0) \boldsymbol{\Xi}^2 \right) \mathbf{h}_{M+1}^{*} \right)^{-1} \\
&\quad \left(\mathbf{h}_{M+1}^{\mathrm{T}} \left(\alpha_q \frac{\left| G_{\mathrm{PS}}(\mathrm{e}^{\mathrm{j}\,2\pi f T_{\mathrm{s}}}) \right|^2}{(g_{\mathrm{PS}} * g_{\mathrm{PS}})\,[0]} \boldsymbol{\Xi} \mathbf{R}_{\mathbf{x}'\mathbf{x}'} \boldsymbol{\Xi} + \beta_q \boldsymbol{\Xi}^2 \right) \mathbf{h}_{M+1}^{*} \right) \\
&= |G_{\mathrm{DAC}}(f)|^2 \left(\mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \left(\left| G_{\mathrm{PS}}(\mathrm{e}^{\mathrm{j}\,2\pi f T_{\mathrm{s}}}) \right|^2 \mathbf{R}_{\mathbf{x}'\mathbf{x}'} + \beta_q/\alpha_q\, (g_{\mathrm{PS}} * g_{\mathrm{PS}})\,[0] \mathbf{I}_N \right) \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \right) \\
&\quad \Big(\mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \Big((g_{\mathrm{DAC}} * g_{\mathrm{DAC}} * g_{\mathrm{PS}} * g_{\mathrm{PS}})\,(0) \mathbf{R}_{\mathbf{x}'\mathbf{x}'} \\
&\quad + \beta_q/\alpha_q\, (g_{\mathrm{PS}} * g_{\mathrm{PS}})\,[0]\, (g_{\mathrm{DAC}} * g_{\mathrm{DAC}})\,(0) \mathbf{I}_N \Big) \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \Bigg)^{-1}. \qquad (9.20)
\end{aligned}
$$

By choosing the impulse response of the DAC as in (9.12) and the pulse shaper as the Root-Raised-Cosine (RRC) filter of roll-off factor $ro = 0.1$, we obtain the PSDs according to (9.20) depicted in Fig. 9.2 for different values of the ratio N/M, when applying linear and non-linear precoding. The PSD is measured at one of the active users. It can be observed that the spectrum is narrower for lager values of the ratio N/M. Additionally, there is a slight improvement in the spectral shape in the case of the linear precoder, which is explained by the non-equal power allocation at the antennas.

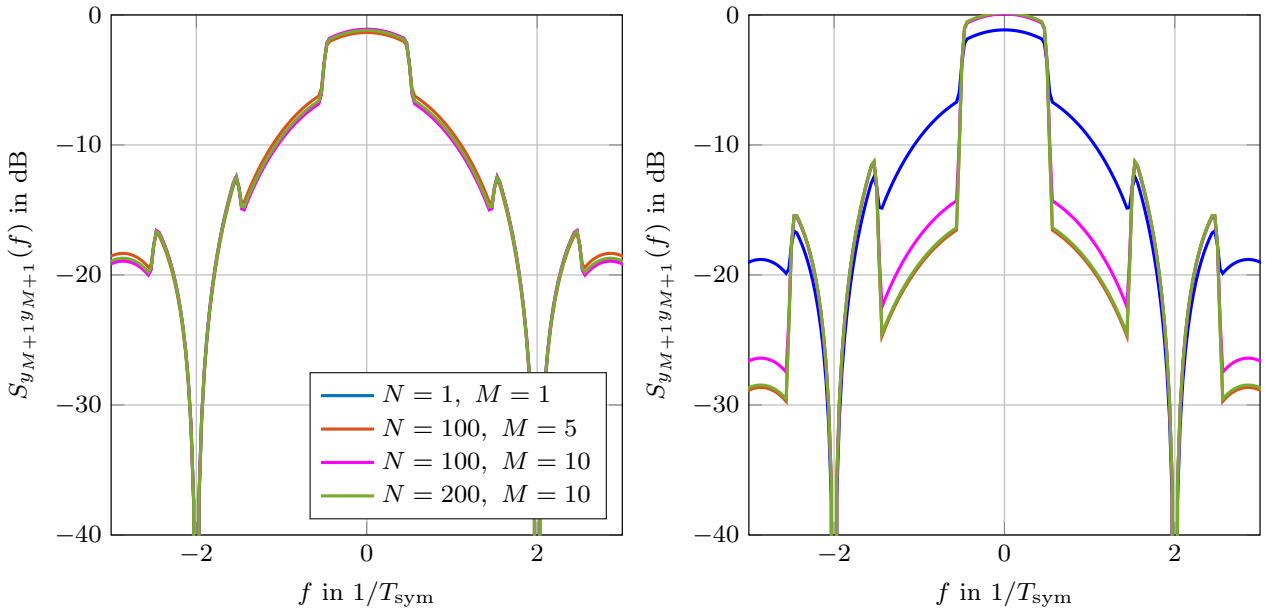
Additionally, the PSD is plotted for a random $\mathbf{h}_{M+1}$, where $\mathbf{h}_{M+1} \in \mathrm{range}\,(\mathbf{R}_{\mathbf{x}'\mathbf{x}'})$ and for the $\mathbf{h}_{M+1}$ being the eingenvector with the largest eingenvalue of $\mathbf{R}_{\mathbf{x}'\mathbf{x}'}$ in Fig. 9.3. For a random $\mathbf{h}_{M+1} \in \mathrm{range}\,(\mathbf{R}_{\mathbf{x}'\mathbf{x}'})$, there is no improvement in the spectral shape for large ratios N/M. This is due to the fact that the signals in that direction do not overlap constructively. The precoder designs the transmit signals such that they overlap constructively at the users.

When $\mathbf{h}_{M+1}$ is chosen as the eingenvector with the largest eigenvalue, we can see the best spectral shape that can be obtained, which improves with the ratio N/M.



(a) QWF-Price$_{\text{neqPA}}$ $(\text{tr}\left(\mathbf{\Xi}^2\right) = N)$ (b) MSM $(\mathbf{\Xi} = \mathbf{I}_N)$

Fig. 9.2: PSD at one of the users for an i.i.d. channel realization with the DAC as the rectangular function in standard QAM staggering and with digital pulse shaping, i.e RRC filter with roll-off factor $ro = 0.1$.



(a) Random $\mathbf{h}_{M+1} \in \text{range}\left(\mathbf{R}_{\mathbf{x}'\mathbf{x}'}\right)$ (b) $\mathbf{h}_{M+1}$: the eigenvector with the largest eigenvalue of $\mathbf{R}_{\mathbf{x}'\mathbf{x}'}$

Fig. 9.3: PSD at the measuring device for an i.i.d. channel realization with the DAC as the rectangular function in standard QAM staggering, with digital pulse shaping, i.e RRC filter with roll-off factor $ro = 0.1$ and with the MSM precoder.

Asymptotic Analysis Let us look at the obtained approximate expression of the PSD in (9.20) and assume that the precoder is close to the matched filter of the channel such that

$$\mathbf{C}_{\mathbf{x}'\mathbf{x}'} \approx \mathbf{H}^{\mathrm{H}}\mathbf{H}, \tag{9.21}$$

$$\mathbf{R}_{\mathbf{x}'\mathbf{x}'} \approx \operatorname{diag}(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1/2}\mathbf{H}^{\mathrm{H}}\mathbf{H}\operatorname{diag}(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1/2}. \tag{9.22}$$

By assuming the i.i.d. channel model described in Section 4.6.1 and a large number of antennas and users with $N \gg M$, which leads to almost orthogonal user channels that span an M-dimensional subspace, we get

$$\mathrm{E}\left[\operatorname{diag}(\mathbf{H}^{\mathrm{H}}\mathbf{H})\right] = M\mathbf{I}_N, \tag{9.23}$$

$$\mathrm{E}\left[\mathbf{h}_{M+1}^{\mathrm{T}}\mathbf{h}_{M+1}^{*}\right] = N,$$

$$0 \leq \mathrm{E}\left[\mathbf{h}_{M+1}^{\mathrm{T}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{h}_{M+1}^{*}\right] \leq N^2, \tag{9.24}$$

where the lower bound is reached, if $\mathbf{h}_{M+1}$ lies in the null space of $\mathbf{H}^{*}$, i.e. $\mathbf{h}_{M+1} \in \operatorname{null}(\mathbf{H}^{*})$, and the right limit is reached, if $\mathbf{h}_{M+1}$ lies in the range space of $\mathbf{H}^{\mathrm{T}}$, i.e. $\mathbf{h}_{M+1} \in \operatorname{range}\left(\mathbf{H}^{\mathrm{T}}\right)$. We first assume that the CEQ is optimal and hence $\boldsymbol{\Xi} = \sqrt{\alpha_q}\operatorname{diag}(\mathbf{C}_{\mathbf{xx}})^{1/2} = \sqrt{\alpha_q\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)(0)}\operatorname{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{1/2}$. Therefore, if $\mathbf{h}_{M+1} \in \operatorname{range}\left(\mathbf{H}^{\mathrm{T}}\right)$, the approximation of the PSD expression is asymptotically given by

$$\begin{aligned} S_{y_{M+1}y_{M+1}}(f) &\to |G_{\mathrm{DAC}}(f)|^2 \left(\alpha_q\left(N^2c_1 + \beta_q/\alpha_q c_2 NM\right)\right)^{-1} \\ &\quad \left(\alpha_q\left(\left|G_{\mathrm{PS}}(\mathrm{e}^{\mathrm{j}\,2\pi f T_{\mathrm{s}}})\right|^2 N^2 + \beta_q/\alpha_q c_3 NM\right)\right) \\ &\to |G_{\mathrm{DAC}}(f)|^2 \frac{\alpha_q \frac{N}{M}\left|G_{\mathrm{PS}}(\mathrm{e}^{\mathrm{j}\,2\pi f T_{\mathrm{s}}})\right|^2 + \beta_q c_3}{\alpha_q \frac{N}{M} c_1 + \beta_q c_2}, \end{aligned} \tag{9.25}$$

where

$$c_1 = \left(g_{\mathrm{PS}} * g_{\mathrm{PS}} * g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(0), \tag{9.26}$$

$$c_2 = \left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[0]\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(0), \tag{9.27}$$

$$c_3 = \left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[0]. \tag{9.28}$$

The obtained expression, under the i.i.d. channel assumption and large number of N and M, consists of two parts: the desired PSD of the pulse shaper that increases linearly with the ratio $\alpha_q N/M$ and a constant noise part due to the quantization errors. Indeed, from this approximation we can see that increasing the ratio N/M improves the spectral shape of the received signal.
In the extreme case that $\mathbf{h}_{M+1}^{\mathrm{T}}$ is orthogonal to all user channels, we get

$$\begin{aligned} S_{y_{M+1}y_{M+1}}(f) &\to |G_{\mathrm{DAC}}(f)|^2 \frac{c_3}{c_2} \\ &\to \frac{|G_{\mathrm{DAC}}(f)|^2}{\int_{-\infty}^{\infty}|G_{\mathrm{DAC}}(f)|^2\,\mathrm{d}f}. \end{aligned} \tag{9.29}$$

We see here that the digital pulse shaping vanishes and the spectrum is shaped only by the transfer function of the DAC.
Hence, this asymptotic analysis explains the behavior observed in the previous section.

9.2.2 Offset QAM

In this case, the output of the CEQ $\mathbf{t}[t]$ undergoes Offset QAM staggering. This means that the inphase and quadrature parts of the signal are alternated in a time distance of $T_\mathrm{s}/2$. Therefore, the continuous-time signal at the transmitter $\mathbf{t}'(t')$ reads as

$$\mathbf{t}'(t') = \sum_{t=0}^{\infty} \tilde{\mathbf{t}}[2t] g_{\mathrm{DAC}}(t' - 2tT') + \tilde{\mathbf{t}}[2t+1] g_{\mathrm{DAC}}(t' - (2t+1)T'), \tag{9.30}$$

where

$$\tilde{\mathbf{t}}[2t] = \Re\left\{\mathbf{t}[t]\right\}, \tag{9.31}$$

$$\tilde{\mathbf{t}}[2t+1] = \mathrm{j}\,\Im\left\{\mathbf{t}[t]\right\}. \tag{9.32}$$

We introduce two different impulse responses for the DAC. The first one is the the rectangular function introduced in (9.12)

$$g_{\mathrm{DAC}}\left(t'\right) = \mathrm{rect}\left(\frac{t'}{2T'}\right), \tag{9.33}$$

where $T' = T_\mathrm{s}/2$. The output signal $\mathbf{t}(t')$ is of constant envelope and the phase follows a rectangular time shape. The maximal difference between the phase values at subsequent time instants is equal to $\pm\pi/4$.
The second time impulse response for the DAC is the cosine impulse

$$g_{\mathrm{DAC}}\left(t'\right) = g_{\mathrm{MSK}}\left(t'\right) = \frac{1}{T'}\cos\left(\frac{\pi}{2T'}t'\right), \quad -T' \le t' \le T', \tag{9.34}$$

which leads in the considered case, i.e. $Q = 4$, to Minimum-Shift Keying (MSK) signaling at the BS antennas. The corresponding frequency transfer function is given by

$$G_{\mathrm{MSK}}\left(f\right) = \frac{4}{\pi}\frac{\cos(2\pi f T')}{1 - \left(4fT'\right)^2}. \tag{9.35}$$

Note that the output signal $\mathbf{t}(t')$ is also of constant envelope and the phase changes linearly in time.
In order to obtain the analytical expression of the autocorrelation function $\rho_{y_{M+1}y_{M+1}}(\tau')$, we have to compute the covariance matrix $\mathrm{E}\left[\mathbf{t}'(t')\mathbf{t}'^{\mathrm{H}}(t'-\tau')\right]$; that is

$$\begin{aligned}
\mathrm{E}\left[\mathbf{t}'(t')\mathbf{t}'^{\mathrm{H}}(t'-\tau')\right] &= \sum_{\ell=-\infty}^{\infty} \Re\left\{\mathbf{C}_{\mathbf{tt}}[2\ell]\right\}\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\left(\tau' - 2\ell T'\right) \\
&\quad - \mathrm{j}\sum_{\ell'=-\infty}^{\infty} \mathbf{C}_{\Re\{\mathbf{t}\}\Im\{\mathbf{t}\}}[2\ell']\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\left(\tau' - (2\ell'-1)T'\right) \\
&\quad + \mathrm{j}\sum_{\ell''=-\infty}^{\infty} \mathbf{C}_{\Im\{\mathbf{t}\}\Re\{\mathbf{t}\}}[2\ell'']\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\left(\tau' - (2\ell''+1)T'\right) \\
&= \left(\Re\left\{\mathbf{C}_{\mathbf{tt}}\right\} * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)\left(\tau'\right) \\
&\quad - \mathrm{j}\left(\mathbf{C}_{\Re\{\mathbf{t}\}\Im\{\mathbf{t}\}} * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)\left(\tau' + T'\right) \\
&\quad + \mathrm{j}\left(\mathbf{C}_{\Im\{\mathbf{t}\}\Re\{\mathbf{t}\}} * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)\left(\tau' - T'\right).
\end{aligned} \tag{9.36}$$

Therefore, we get

$$\rho_{y_{M+1}y_{M+1}}(\tau') = \left(\mathbf{h}_{M+1}^{\mathrm{T}}\left(\left(\Re\left\{\mathbf{C_{tt}}\right\} * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)(0) + \mathrm{j}\left(\Im\left\{\mathbf{C_{tt}}\right\} * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)(T')\right)\mathbf{h}_{M+1}^{*}\right)^{-1}$$
$$\mathbf{h}_{M+1}^{\mathrm{T}}\Bigg(\left(\Re\left\{\mathbf{C_{tt}}\right\} * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)(\tau') - \mathrm{j}\left(\mathbf{C}_{\Re\{\mathbf{t}\}\Im\{\mathbf{t}\}} * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)(\tau' + T')$$
$$+ \mathrm{j}\left(\mathbf{C}_{\Im\{\mathbf{t}\}\Re\{\mathbf{t}\}} * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)(\tau' - T')\Bigg)\mathbf{h}_{M+1}^{*}. \tag{9.37}$$

9.2.2.1 Without Digital Pulse Shaping

In the case of no digital pulse shaping, the autocorrelation function simplifies to

$$\rho_{y_{M+1}y_{M+1}}(\tau') = \left(\mathbf{h}_{M+1}^{\mathrm{T}}\left(\Re\left\{\mathbf{C_{tt}}\right\}\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)(0) + \mathrm{j}\,\Im\left\{\mathbf{C_{tt}}\right\}\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(T')\mathbf{h}_{M+1}^{*}\right)^{-1}$$
$$\mathbf{h}_{M+1}^{\mathrm{T}}\Bigg(\Re\left\{\mathbf{C_{tt}}\right\}\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau') - \mathrm{j}\,\mathbf{C}_{\Re\{\mathbf{t}\}\Im\{\mathbf{t}\}}\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau' + T')$$
$$+ \mathrm{j}\,\mathbf{C}_{\Im\{\mathbf{t}\}\Re\{\mathbf{t}\}}\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau' - T')\Bigg)\mathbf{h}_{M+1}^{*}. \tag{9.38}$$

Thus, the PSD can be computed as

$$S_{y_{M+1}y_{M+1}}(f) = \frac{\left|G_{\mathrm{DAC}}(f)\right|^2 \mathbf{h}_{M+1}^{\mathrm{T}}\left(\Re\left\{\mathbf{C_{tt}}\right\} + \mathrm{j}\,\mathbf{C}_{\Im\{\mathbf{t}\}\Re\{\mathbf{t}\}}\,\mathrm{e}^{-\mathrm{j}\,2\pi f T'} - \mathrm{j}\,\mathbf{C}_{\Re\{\mathbf{t}\}\Im\{\mathbf{t}\}}\,\mathrm{e}^{\mathrm{j}\,2\pi f T'}\right)\mathbf{h}_{M+1}^{*}}{\mathbf{h}_{M+1}^{\mathrm{T}}\left(\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(0)\Re\left\{\mathbf{C_{tt}}\right\} + \mathrm{j}\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(T')\Im\left\{\mathbf{C_{tt}}\right\}\right)\mathbf{h}_{M+1}^{*}}. \tag{9.39}$$

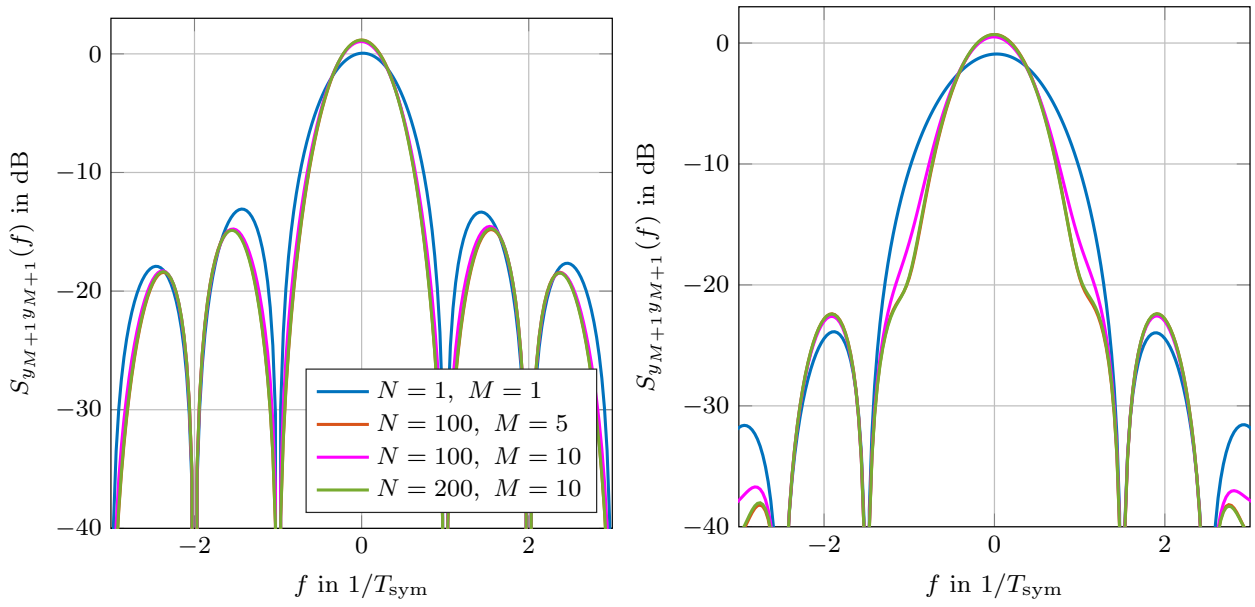
It holds that

$$\begin{aligned}\mathbf{h}_{M+1}^{\mathrm{T}}\mathbf{C}_{\Re\{\mathbf{t}\}\Im\{\mathbf{t}\}}\,\mathrm{e}^{\mathrm{j}\,2\pi f T'}\,\mathbf{h}_{M+1}^{*} &= \left(\mathbf{h}_{M+1}^{\mathrm{T}}\mathbf{C}_{\Re\{\mathbf{t}\}\Im\{\mathbf{t}\}}\,\mathrm{e}^{\mathrm{j}\,2\pi f T'}\,\mathbf{h}_{M+1}^{*}\right)^{\mathrm{T}} \\ &= \mathbf{h}_{M+1}^{\mathrm{H}}\mathbf{C}_{\Im\{\mathbf{t}\}\Re\{\mathbf{t}\}}\,\mathrm{e}^{\mathrm{j}\,2\pi f T'}\,\mathbf{h}_{M+1} \\ &= \left(\mathbf{h}_{M+1}^{\mathrm{T}}\mathbf{C}_{\Im\{\mathbf{t}\}\Re\{\mathbf{t}\}}\,\mathrm{e}^{-\mathrm{j}\,2\pi f T'}\,\mathbf{h}_{M+1}^{*}\right)^{*},\end{aligned} \tag{9.40}$$

which leads to the simplified expression

$$\begin{aligned}S_{y_{M+1}y_{M+1}}(f) &= \frac{\left|G_{\mathrm{DAC}}(f)\right|^2\left(\mathbf{h}_{M+1}^{\mathrm{T}}\Re\left\{\mathbf{C_{tt}}\right\}\mathbf{h}_{M+1}^{*} - 2\Im\left\{\mathbf{h}_{M+1}^{\mathrm{T}}\mathbf{C}_{\Im\{\mathbf{t}\}\Re\{\mathbf{t}\}}\mathbf{h}_{M+1}^{*}\,\mathrm{e}^{-\mathrm{j}\,2\pi f T'}\right\}\right)}{\mathbf{h}_{M+1}^{\mathrm{T}}\left(\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(0)\Re\left\{\mathbf{C_{tt}}\right\} + \mathrm{j}\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(T')\Im\left\{\mathbf{C_{tt}}\right\}\right)\mathbf{h}_{M+1}^{*}} \\ &\approx \left|G_{\mathrm{DAC}}(f)\right|^2\Bigg(\mathbf{h}_{M+1}^{\mathrm{T}}\mathbf{\Xi}\left(\alpha_q\Re\left\{\mathbf{R_{xx}}\right\} + \beta_q\mathbf{I}_N\right)\mathbf{\Xi}\mathbf{h}_{M+1}^{*} \\ &\quad - 2\alpha_q\Im\left\{\mathbf{h}_{M+1}^{\mathrm{T}}\mathbf{\Xi}\,\mathrm{diag}(\mathbf{C_{xx}})^{-1/2}\mathbf{C}_{\Im\{\mathbf{x}\}\Re\{\mathbf{x}\}}\,\mathrm{diag}(\mathbf{C_{xx}})^{-1/2}\mathbf{\Xi}\mathbf{h}_{M+1}^{*}\,\mathrm{e}^{-\mathrm{j}\,2\pi f T'}\right\}\Bigg) \\ &\quad \Bigg(\mathbf{h}_{M+1}^{\mathrm{T}}\mathbf{\Xi}\Big(\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(0)\left(\alpha_q\Re\left\{\mathbf{R_{xx}}\right\} + \beta_q\mathbf{I}_N\right) \\ &\quad + \mathrm{j}\,\alpha_q\left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(T')\Im\left\{\mathbf{R_{xx}}\right\}\Big)\mathbf{\Xi}\mathbf{h}_{M+1}^{*}\Bigg)^{-1}.\end{aligned} \tag{9.41}$$

In Fig. 9.4, the evaluation of the PSD expression in (9.41) is depicted for both time impulse responses for the DAC. We can observe that the spectral shape changes slightly with increasing the ratio N/M, in contrast to the case of standard QAM without pulse shaping,, where no dependency on N and M was observed.



(a) Rectangular function

(b) MSK

Fig. 9.4: PSD at one of the users for an i.i.d. channel realization with the MSM precoder in Offset QAM staggering without digital pulse shaping.

9.2.2.2 With Digital Pulse Shaping

We use the same identity as in (9.14). By applying the first order Taylor approximation of the arcsin function, we obtain

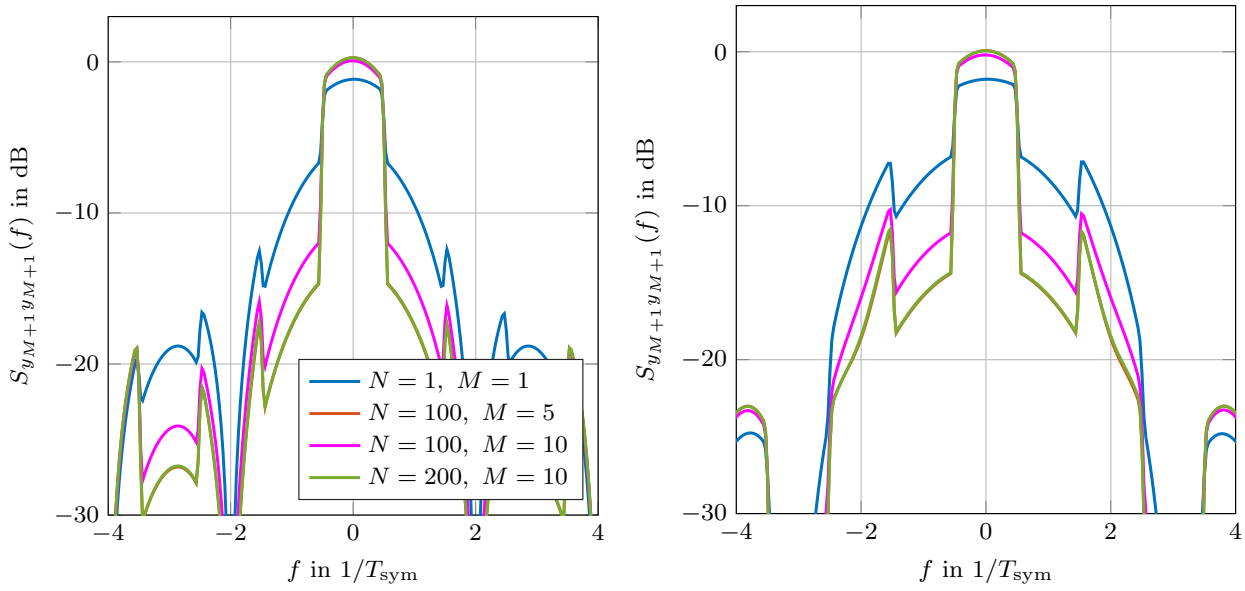
$$\begin{aligned}
\mathrm{E}\left[\mathbf{t}'(t')\mathbf{t}'^{\mathrm{H}}(t'-\tau')\right] \approx & \sum_{\ell=-\infty}^{\infty} \mathbf{L}_{\mathcal{Q}} \Re\left\{\mathbf{C}_{\mathbf{xx}}[\ell]\right\} \mathbf{L}_{\mathcal{Q}}^{\mathrm{H}} \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau' - 2\ell T_{\mathrm{s}}) \\
& + \beta_q \boldsymbol{\Xi}^2 \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau) \\
& - \mathrm{j} \sum_{\ell'=-\infty}^{\infty} \mathbf{L}_{\mathcal{Q}} \mathbf{C}_{\Re\{\mathbf{x}\}\Im\{\mathbf{x}\}}[\ell'] \mathbf{L}_{\mathcal{Q}}^{\mathrm{H}} \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau' - (2\ell'-1)T_{\mathrm{s}}) \\
& + \mathrm{j} \sum_{\ell''=-\infty}^{\infty} \mathbf{L}_{\mathcal{Q}} \mathbf{C}_{\Im\{\mathbf{x}\}\Re\{\mathbf{x}\}}[\ell''] \mathbf{L}_{\mathcal{Q}}^{\mathrm{H}} \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau' - (2\ell''+1)T_{\mathrm{s}}) \\
& = \frac{1}{\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[0]} \alpha_q \boldsymbol{\Xi} \operatorname{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2} \\
& \Bigg(\Re\left\{\mathbf{C}_{\mathbf{x}'\mathbf{x}'}\right\} \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}} * g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)(\tau) \\
& - \mathrm{j}\, \mathbf{C}_{\Re\{\mathbf{x}'\}\Im\{\mathbf{x}'\}} \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}} * g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)(\tau' + T_{\mathrm{s}}) \\
& + \mathrm{j}\, \mathbf{C}_{\Im\{\mathbf{x}'\}\Re\{\mathbf{x}'\}} \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}} * g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)(\tau' - T_{\mathrm{s}}) \Bigg) \\
& \operatorname{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2} \boldsymbol{\Xi} \\
& + \beta_q \boldsymbol{\Xi}^2 \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau'), \qquad (9.42)
\end{aligned}$$

which leads to

$$\begin{aligned}
\rho_{y_{M+1}y_{M+1}}(\tau') \approx & \Bigg(\mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \Big(\Big(\Re\left\{\mathbf{R}_{\mathbf{x}'\mathbf{x}'}\right\} \left(g_{\mathrm{PS}} * g_{\mathrm{PS}} * g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(0) \\
& + \beta_q/\alpha_q \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(0) \left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)(0) \mathbf{I}_N \Big) \\
& + \mathrm{j}\, \Im\left\{\mathbf{R}_{\mathbf{x}'\mathbf{x}'}\right\} \left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[0] * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(T') \Big) \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \Bigg)^{-1} \\
& \mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \Bigg(\Re\left\{\mathbf{R}_{\mathbf{x}'\mathbf{x}'}\right\} \left(g_{\mathrm{PS}} * g_{\mathrm{PS}} * g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau') \\
& - \mathrm{j} \operatorname{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2} \mathbf{C}_{\Re\{\mathbf{x}'\}\Im\{\mathbf{x}'\}} \operatorname{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2} \left(\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right) * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)(\tau' + T') \\
& + \mathrm{j} \operatorname{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2} \mathbf{C}_{\Im\{\mathbf{x}'\}\Re\{\mathbf{x}'\}} \operatorname{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2} \left(\left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right) * \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)\right)(\tau' - T') \Bigg) \\
& \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \\
& + \frac{\beta_q}{\alpha_q} \left(g_{\mathrm{PS}} * g_{\mathrm{PS}}\right)[0] \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}}\right)(\tau') \mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi}^2 \mathbf{h}_{M+1}^{*}. \qquad (9.43)
\end{aligned}$$

Thus, the PSD can be approximated by

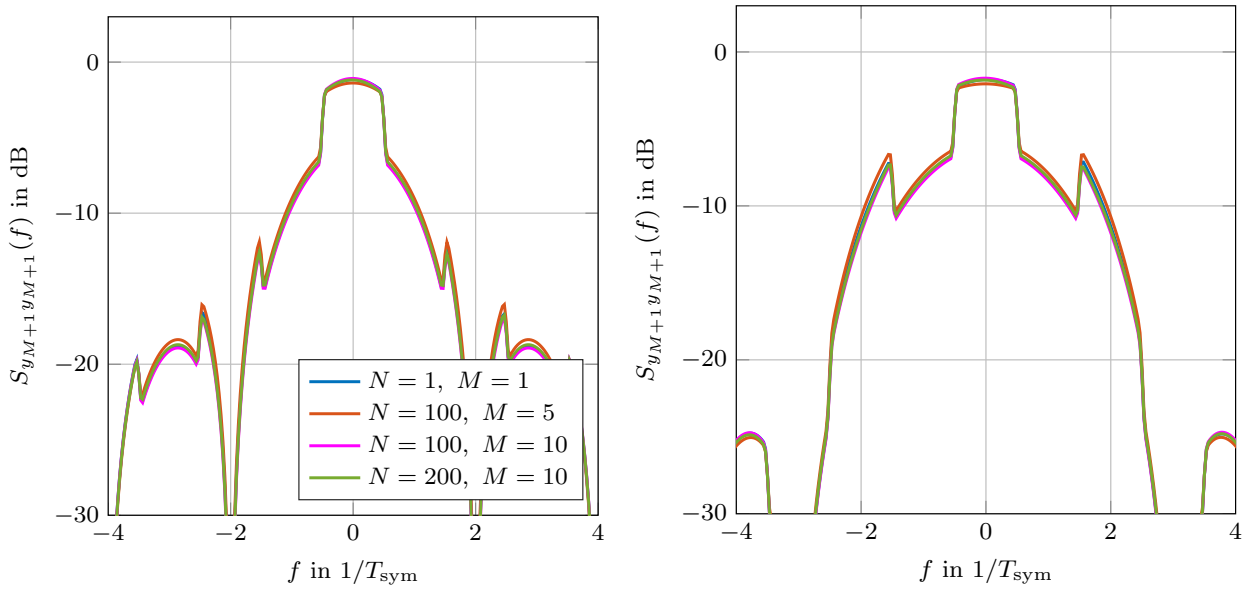
$$
\begin{aligned}
S_{y_{M+1}y_{M+1}}(f) \approx |G_{\mathrm{DAC}}(f)|^2 \Bigg(& \mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \, |G_{\mathrm{PS}}(f)|^2 \Big(\Re\{\mathbf{R}_{\mathbf{x}'\mathbf{x}'}\} \\
& - \mathrm{j}\, \mathrm{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2} \mathbf{C}_{\Re\{\mathbf{x}'\}\Im\{\mathbf{x}'\}} \, \mathrm{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2} \, \mathrm{e}^{\mathrm{j}\, 2\pi f T'} \\
& + \mathrm{j}\, \mathrm{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2} \mathbf{C}_{\Im\{\mathbf{x}'\}\Re\{\mathbf{x}'\}} \, \mathrm{diag}(\mathbf{C}_{\mathbf{x}'\mathbf{x}'})^{-1/2} \, \mathrm{e}^{-\mathrm{j}\, 2\pi f T'} \Big) \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \\
& + \frac{\beta_q}{\alpha_q} \left(g_{\mathrm{PS}} * g_{\mathrm{PS}} \right)[0] \mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi}^2 \mathbf{h}_{M+1}^{*} \Bigg) \\
& \Bigg(\mathbf{h}_{M+1}^{\mathrm{T}} \boldsymbol{\Xi} \Big(\Big(\Re\{\mathbf{R}_{\mathbf{x}'\mathbf{x}'}\} \left(g_{\mathrm{PS}} * g_{\mathrm{PS}} * g_{\mathrm{DAC}} * g_{\mathrm{DAC}} \right)(0) \\
& + \beta_q/\alpha_q \left(g_{\mathrm{DAC}} * g_{\mathrm{DAC}} \right)(0) \left(g_{\mathrm{PS}} * g_{\mathrm{PS}} \right)[0] \mathbf{I}_N \Big) \\
& + \mathrm{j}\, \Im\{\mathbf{R}_{\mathbf{x}'\mathbf{x}'}\} \left(g_{\mathrm{PS}} * g_{\mathrm{PS}} * g_{\mathrm{DAC}} * g_{\mathrm{DAC}} \right)(T') \Big) \boldsymbol{\Xi} \mathbf{h}_{M+1}^{*} \Bigg)^{-1} .
\end{aligned}
\tag{9.44}
$$



(a) Rectangular function (b) MSK

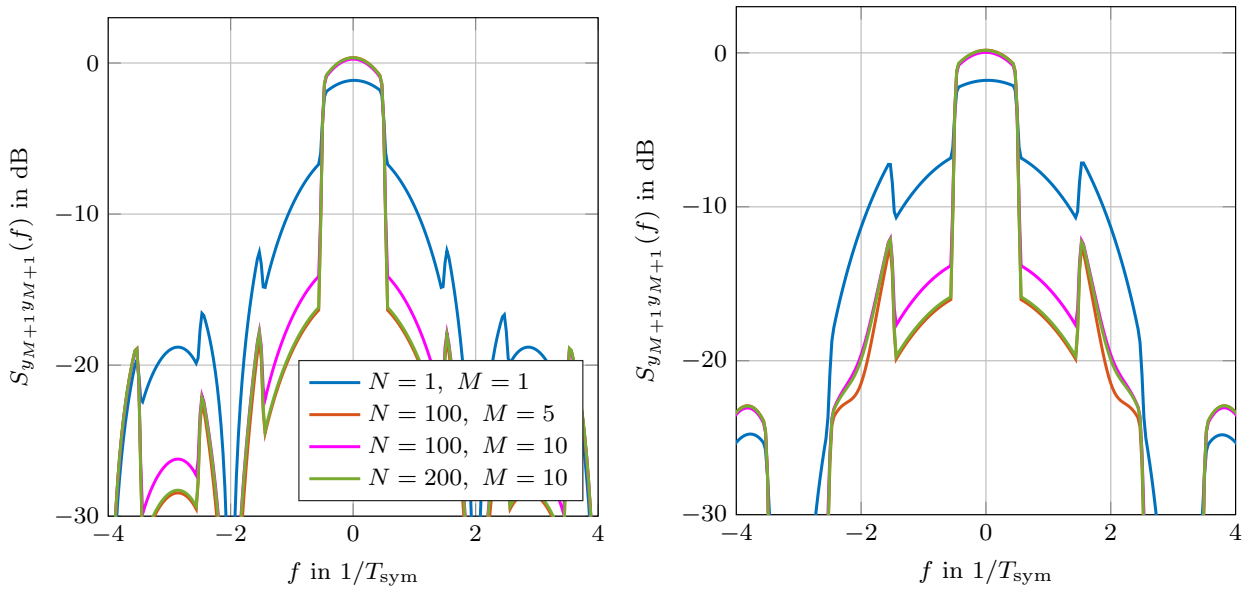
Fig. 9.5: PSD at one of the users for an i.i.d. channel realization with the MSM precoder in Offset QAM staggering with the RRC digital pulse shaping of $ro = 0.1$.

By choosing the digital pulse shaper as the RRC filter of roll-off factor $ro = 0.1$, we obtain the PSDs depicted in Fig. 9.5 for different time impulse responses for the DAC and different values of the ratio N/M. It can be again observed that the spectrum is narrower for lager values of the ratio N/M. Additionally, the spectrum with the rectangular function and in Offset-QAM staggering has less out of band radiations compared to the case of standard QAM case.
Additionally, the PSD is plotted in Fig. 9.6 for a random $\mathbf{h}_{M+1}$, where $\mathbf{h}_{M+1} \in \mathrm{range}\,(\mathbf{R}_{\mathbf{x}'\mathbf{x}'})$.



(a) Rectangular function (b) MSK

Fig. 9.6: PSD for a random $\mathbf{h}_{M+1} \in \text{range}\,(\mathbf{R}_{\mathbf{x}'\mathbf{x}'})$ for an i.i.d. channel realization with the MSM precoder in Offset QAM staggering with the RRC digital pulse shaping of $ro = 0.1$.



(a) Rectangular function (b) MSK

Fig. 9.7: PSD for $\mathbf{h}_{M+1}$ being the eigenvector with the largest eingenvalue of $\mathbf{R}_{\mathbf{x}'\mathbf{x}'}$ for an i.i.d. channel realization with the MSM precoder in Offset QAM staggering with the RRC digital pulse shaping of $ro = 0.1$.

there is no improvement in the spectral shape for large ratios N/M. This is due to the fact that the signals in that direction do not overlap constructively. The precoder designs the transmit signals such that they overlap constructively at the users.
Finally, the PSD is plotted for $\mathbf{h}_{M+1}$ being the eingenvector with the largest eingenvalue of

$\mathbf{R}_{\mathbf{x}'\mathbf{x}'}$ in Fig. 9.7. We can see the best spectral shape that can be obtained, which improves with the ratio N/M.

In summary, we have shown that the spectral shape at the measuring device, when $\mathbf{h}_{M+1}$ is one of the user channels or one of the eigenvectors with large eigenvalues of $\mathbf{R}_{\mathbf{x}'\mathbf{x}'}$, improves by increasing the ratio N/M. Indeed, the spectral shape deterioration due to the coarse quantization diminishes in massive MIMO systems and is not that dramatic as in the SU Single-Input Single-Output (SISO) system. By using the Offset-QAM staggering combined with the rectangular or MSK impulse response, we ensure that the input of the PA is of constant envelope. Therefore, only non-linear and thus highly power efficient PAs are required in the system.

We would like to remind the reader that our analysis was restricted to the case of $Q = 4$. Our goal was to show the potential of spectral shaping despite the presence of coarse quantization in massive MIMO scenarios. The optimization problem of the PSD is left for future wotk. However, we would like to share our future steps with the reader.

In order to further improve the spectral properties in QCE systems, we think that the desired pulse shaping should be considered in the precoder design. Therefore, each user' s signal should undergo a pulse shaping. Hence, the input to the precoder is not drawn from a discrete well-defined set $\mathbb{S}$ anymore but from a continuous of high resolution undefined set. We suggest to quantize the resulting continuous set to obtain a well-defined discrete set $\mathbb{S}'$ at the input of the precoder. Then, the MSM precoder should be extended to make sure that the received signals at the receiver belong to the new defined set. In other words, the new SRs should be defined that correspond to the new set $\mathbb{S}'$. Consequently, the received signals would follow the same pulse shape that is designed at the transmitter.

Another idea would be to optimize the pulse shaping after the CEQ. Therefore, to ensure the CE property, we can apply phase modulation for the transmission. Hence, we can consider the concepts of the Continuous Phase Modulation (CPM), [74], in particular the Gaussian Minimum-Shift Keying (GMSK) modulation [75]. This modulation type ensures a smooth change of the phase in time. This enhances the spectral efficiency of the transmit signals.

10. Conclusion

Throughout this thesis, we developed digital signal processing techniques for MU massive MIMO systems, whose transmit signals are restricted to be of CE. Additionally, the conversion from the digital into the analog world is deployed with very low resolution. The introduced CEQ models this behavior. Hence, we end up with coarsely QCE MU massive MIMO systems.

The choice of CE signaling at the BS antennas is motivated by having high power efficient PAs. In the case of massive MIMO systems, where the number of BS antennas is assumed to be very large, e.g. 100, the PA power consumption becomes a crucial concern especially at mmW frequencies. Therefore, we enhance the power efficiency of the communication systems by the CE constraint.
The choice of the coarse quantization is also motivated by having less power consuming DACs, whose number increases linearly with the number of the BS antennas and whose power consumption increases exponentially with the resolution. Other benefits of low resolution DACs are the reduced required chip area and the relaxed need for highly performing surrounding circuitry.
The combination of both properties leads to power efficient QCE MU massive MIMO systems.

However, this results in performance degradation, since beside to the MUI, the channel distortions and the AWGN, additional CEQ distortions are introduced to the system. Therefore, we proposed digital precoding techniques to mitigate all usual distortion sources and additionally the CEQ distortions. The proposed techniques are classified into two groups: linear and non-linear techniques.
The linear precoding designs were based on the MMSE criterion. The precoding matrix is optimized for every channel coherence time. For the matrix design, the CEQ was linearized based on Bussgang's theorem and the statistical properties of the quantization noise were first computed with Price's theorem and second approximated with the LCA. Furthermore, we have shown that there is no MSE duality between uplink and downlink in the case of QCE MIMO systems. Two linear precoder designs were proposed based on a virtual duality and an approximate duality. The simulation results showed that the proposed linear techniques are moderately better, in terms of the uncoded BER, than the well known linear WF followed by the CEQ. This observation made us think about non-linear precoding.
In the non-linear precoding approach, denoted by MSM, the transmit vector at each time instant is optimized. Therefore, we talk about symbol-wise precoding that depend on the desired received signal at the users and the channel realization. This approach is certainly more computationally complex than the linear approach. The design criterion is the safety

margin to the decision thresholds that should be maximized to minimize the SER. In order to consider the QCE constraint, the entries of the designed transmit vector should belong to a relaxed convex version of the QCE constraint. The problem could be formulated such that we obtained a linear programming problem. The simulation results showed significant gains compared to the linear approach. However, this gain comes at the cost of more computational complexity.

The main motivation of choosing the QCE is the power efficiency. We have shown that the PA power consumption of an ideal linear system is more than four times the PA power consumption of a QCE for the same uncoded BER performance.

To settle the doubts about the spectral regrowth combined with the coarse quantization, we analyzed the PSDs at the users in QCE MU massive MIMO systems. We have observed that the quantization widens the spectrum. However, this effect diminishes for larger ratios between the number of antennas and users, which is the case in massive MIMO systems. In this analysis, we did not propose an optimal solution for the time pulse shaping in order to optimize the PSD. However, we have shown that the problem of the spectral regrowth in the presence of quantization distortions is not that dramatic as in the case of SU SISO systems. This means that there is potential for optimization, which is left for future work.

Appendix

A1 Complex-Valued Gaussian Joint PDF

Let us assume that x_i and x_j with $i \neq j$, are complex-valued Gaussian random variables with zero means and positive variances $\sigma_{x_i}^2$ and $\sigma_{x_j}^2$; that is the PDFs are given by

$$p_{X_{i/j}}(x_{i/j}) = \frac{1}{\pi\sigma_{x_{i/j}}^2} \mathrm{e}^{-\frac{\left|x_{i/j}\right|^2}{\sigma_{x_{i/j}}^2}}.$$

According to [55, Theorem 2.16], the random variable x_i given x_j is a complex-valued Gaussian random variable with mean value

$$\mathrm{E}\left[x_i|x_j\right] = \frac{\mathrm{E}\left[x_i x_j^*\right]}{\sigma_{x_j}^2} x_j$$

and variance

$$\mathrm{E}\left[\left|x_i - \mathrm{E}\left[x_i|x_j\right]\right|^2 |x_j\right] = \sigma_{x_i}^2 - \frac{\mathrm{E}\left[x_i x_j^*\right]}{\sigma_{x_j}^2} \mathrm{E}\left[x_i^* x_j\right].$$

After introducing the correlation coefficient factor $\rho_{x_i,x_j} = \frac{\mathrm{E}\left[x_i x_j^*\right]}{\sigma_{x_i}\sigma_{x_j}}$, it holds that

$$x_i|x_j \sim \mathcal{N}_{\mathbb{C}}(\rho_{x_i,x_j}\frac{\sigma_{x_i}}{\sigma_{x_j}}x_j, \sigma_{x_i}^2\left(1 - \left|\rho_{x_i,x_j}\right|^2\right)).$$

Hence, the PDF of x_i given x_j is given by

$$p_{X_i|X_j}(x_i|x_j) = \frac{1}{\pi\sigma_{x_i}^2\left(1 - \left|\rho_{x_i,x_j}\right|^2\right)} \mathrm{e}^{\left(-\frac{\left|x_i - \rho_{x_i,x_j}\frac{\sigma_{x_i}}{\sigma_{x_j}}x_j\right|^2}{\sigma_{x_i}^2\left(1-\left|\rho_{x_i,x_j}\right|^2\right)}\right)}.$$

Finally, the joint PDF of x_i and x_j is obtained by

$$\begin{aligned} p_{X_i,X_j}(x_i,x_j) &= p_{X_i|X_j}(x_i|x_j)p_{X_j}(x_j) \\ &= \frac{1}{\pi^2\left(1 - \left|\rho_{x_i,x_j}\right|^2\right)\sigma_{x_i}^2\sigma_{x_j}^2} \mathrm{e}^{-\frac{1}{1-\left|\rho_{x_i,x_j}\right|^2}\left(\frac{|x_i|^2}{\sigma_{x_i}^2} + \frac{\left|x_j\right|^2}{\sigma_{x_j}^2} - \frac{2\Re\left\{\rho_{x_i,x_j}x_j x_i^*\right\}}{\sigma_{x_i}\sigma_{x_j}}\right)}. \end{aligned}$$

A2 Proof of Theorem 1

We recall the modified version of Price's theorem in [58].

Theorem 3 (Pawula's Theorem [58]). *Let $f(\mathbf{s}) = f(s_1, \cdots, s_n)$ be a function of n real-valued joint Gaussian random variables. The covariance matrix and the correlation matrix of the vector $\mathbf{s}$ are denoted by $\mathbf{C_{ss}}$ and $\mathbf{R_{ss}}$ of elements denoted by ρ_{s_i,s_j}, $i, j = 1, \cdots, n$. The off-diagonal entries of $\mathbf{C_{ss}}$ are multiplied by a perturbing term υ. Then*

$$\frac{\mathrm{d}^\ell \,\mathrm{E}\left[f(s_1, \cdots, s_n)\right]}{\mathrm{d}\upsilon^\ell} = \mathrm{E}_\upsilon \left[\sum_{i<j} \rho^\ell_{s_i,s_j} \sigma^\ell_{s_i} \sigma^\ell_{s_j} \frac{\partial^{2\ell} f(s_1, \cdots, s_n)}{\partial s_i^\ell \partial s_j^\ell} \right], \quad s_i, s_j \in \mathbb{R},\ i,\ j = 1, \cdots, n,$$

where E_υ denotes the expectation based on the resulting perturbed PDF.

Theorem 3 can be applied to two complex-valued Gaussian random variables $x_i = x_{i_R} + \mathrm{j}\, x_{i_I}$ and $x_j = x_{j_R} + \mathrm{j}\, x_{j_I}$ as follows

$$\begin{aligned}
\frac{\mathrm{d}^\ell \,\mathrm{E}\left[f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})\right]}{\mathrm{d}\upsilon^\ell} = \mathrm{E}_\upsilon \Bigg[& \rho^\ell_{x_{i_R},x_{i_I}} \sigma^\ell_{x_{i_R}} \sigma^\ell_{x_{i_I}} \frac{\partial^{2\ell} f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})}{\partial x^\ell_{i_R} \partial x^\ell_{i_I}} \\
& + \rho^\ell_{x_{i_R},x_{j_R}} \sigma^\ell_{x_{i_R}} \sigma^\ell_{x_{j_R}} \frac{\partial^{2\ell} f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})}{\partial x^\ell_{i_R} \partial x^\ell_{j_R}} \\
& + \rho^\ell_{x_{i_R},x_{j_I}} \sigma^\ell_{x_{i_R}} \sigma^\ell_{x_{j_I}} \frac{\partial^{2\ell} f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})}{\partial x^\ell_{i_R} \partial x^\ell_{j_I}} \\
& + \rho^\ell_{x_{i_I},x_{j_R}} \sigma^\ell_{x_{i_I}} \sigma^\ell_{x_{j_R}} \frac{\partial^{2\ell} f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})}{\partial x^\ell_{i_I} \partial x^\ell_{j_R}} \\
& + \rho^\ell_{x_{i_I},x_{j_I}} \sigma^\ell_{x_{i_I}} \sigma^\ell_{x_{j_I}} \frac{\partial^{2\ell} f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})}{\partial x^\ell_{i_I} \partial x^\ell_{j_I}} \\
& + \rho^\ell_{x_{j_R},x_{j_I}} \sigma^\ell_{x_{j_R}} \sigma^\ell_{x_{j_I}} \frac{\partial^{2\ell} f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})}{\partial x^\ell_{j_R} \partial x^\ell_{j_I}} \Bigg].
\end{aligned}$$

While assuming circularly symmetric distributed Gaussian signals x_i and x_j, i.e.

$$\begin{aligned}
&\sigma_{x_{i_R}} = \sigma_{x_{i_I}} = \frac{1}{\sqrt{2}} \sigma_{x_i}, && \sigma_{x_{j_R}} = \sigma_{x_{j_I}} = \frac{1}{\sqrt{2}} \sigma_{x_j}, \\
&\rho_{x_{i_R},x_{j_R}} = \rho_{x_{i_I},x_{j_I}} = \Re\{\rho_{x_i,x_j}\}, && \rho_{x_{i_I},x_{j_R}} = -\rho_{x_{i_R},x_{j_I}} = \Im\{\rho_{x_i,x_j}\}, \\
&\rho_{x_{i_R},x_{i_I}} = 0, && \rho_{x_{j_R},x_{j_I}} = 0,
\end{aligned}$$

we get

$$\begin{aligned}
\frac{\mathrm{d}^\ell \,\mathrm{E}\left[f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})\right]}{\mathrm{d}\upsilon^\ell} = & \frac{1}{2} \sigma^\ell_{x_i} \sigma^\ell_{x_j} \cdot \\
& \mathrm{E}_\upsilon \Bigg[\Re\{\rho_{x_i,x_j}\}^\ell \left(\frac{\partial^{2\ell} f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})}{\partial x^\ell_{i_R} \partial x^\ell_{j_R}} + \frac{\partial^{2\ell} f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})}{\partial x^\ell_{i_I} \partial x^\ell_{j_I}} \right) \\
& + \Im\{\rho_{x_i,x_j}\}^\ell \left(\frac{\partial^{2\ell} f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})}{\partial x^\ell_{i_I} \partial x^\ell_{j_R}} - \frac{\partial^{2\ell} f(x_{i_R}, x_{i_I}, x_{j_R}, x_{j_I})}{\partial x^\ell_{i_R} \partial x^\ell_{j_I}} \right) \Bigg].
\end{aligned} \tag{A1}$$

The following equalities

$$\frac{\partial}{\partial x_R} = \frac{\partial}{\partial x} + \frac{\partial}{\partial x^*},$$
$$\frac{\partial}{\partial x_I} = \mathrm{j}\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x^*}\right)$$

imply that

$$\begin{aligned}
\frac{\partial^2}{\partial x_{i_R}\partial x_{j_R}} &= \frac{\partial^2}{\partial x_i \partial x_j} + \frac{\partial^2}{\partial x_i \partial x_j^*} + \frac{\partial^2}{\partial x_i^* \partial x_j} + \frac{\partial^2}{\partial x_i^* \partial x_j^*},\\
\frac{\partial^2}{\partial x_{i_I}\partial x_{j_I}} &= -\frac{\partial^2}{\partial x_i \partial x_j} + \frac{\partial^2}{\partial x_i \partial x_j^*} + \frac{\partial^2}{\partial x_i^* \partial x_j} - \frac{\partial^2}{\partial x_i^* \partial x_j^*},\\
\frac{\partial^2}{\partial x_{i_I}\partial x_{j_R}} &= \mathrm{j}\left(\frac{\partial^2}{\partial x_i \partial x_j} + \frac{\partial^2}{\partial x_i \partial x_j^*} - \frac{\partial^2}{\partial x_i^* \partial x_j} - \frac{\partial^2}{\partial x_i^* \partial x_j^*}\right),\\
\frac{\partial^2}{\partial x_{i_R}\partial x_{j_I}} &= \mathrm{j}\left(\frac{\partial^2}{\partial x_i \partial x_j} - \frac{\partial^2}{\partial x_i \partial x_j^*} + \frac{\partial^2}{\partial x_i^* \partial x_j} - \frac{\partial^2}{\partial x_i^* \partial x_j^*}\right).
\end{aligned}$$

Thus, (A1) simplifies to

$$\begin{aligned}
\frac{\mathrm{d}^\ell \,\mathrm{E}\left[f(x_i, x_j)\right]}{\mathrm{d}\upsilon^\ell} = \sigma_{x_i}^\ell \sigma_{x_j}^\ell \mathrm{E}_\upsilon \Bigg[& \Re\{\rho_{x_i,x_j}\}^\ell \left(\frac{\partial^{2\ell} f(x_i, x_j)}{\partial x_i^\ell \partial x_j^{*\ell}} + \frac{\partial^{2\ell} f(x_i, x_j)}{\partial x_i^{*\ell} \partial x_j^\ell}\right)\\
& + \mathrm{j}^\ell \,\Im\{\rho_{x_i,x_j}\}^\ell \left(\frac{\partial^{2\ell} f(x_i, x_j)}{\partial x_i^\ell \partial x_j^{*\ell}} - \frac{\partial^{2\ell} f(x_i, x_j)}{\partial x_i^{*\ell} \partial x_j^\ell}\right)\Bigg].
\end{aligned}$$

A3 Second Order Derivative in the Polar Representation

$$\begin{aligned}
\frac{\partial^2 f(x_i, x_j)}{\partial x_i \partial x_j^*} &= \frac{\partial}{\partial x_i}\left(\frac{\partial f(x_i, x_j)}{\partial x_j^*}\right)\\
&= \frac{\partial}{\partial x_i}\left(\frac{\partial f(x_i, x_j)}{\partial r_j}\frac{\partial r_j}{\partial x_j^*} + \frac{\partial f(x_i, x_j)}{\partial \phi_j}\frac{\partial \phi_j}{\partial x_j^*}\right)\\
&= \frac{\partial^2 f(x_i, x_j)}{\partial r_i \partial r_j}\frac{\partial r_i}{\partial x_i}\frac{\partial r_j}{\partial x_j^*} + \frac{\partial^2 f(x_i, x_j)}{\partial \phi_i \partial r_j}\frac{\partial \phi_i}{\partial x_i}\frac{\partial r_j}{\partial x_j^*} + \frac{\partial^2 f(x_i, x_j)}{\partial r_i \partial \phi_j}\frac{\partial r_i}{\partial x_i}\frac{\partial \phi_j}{\partial x_j^*} + \frac{\partial^2 f(x_i, x_j)}{\partial \phi_i \partial \phi_j}\frac{\partial \phi_i}{\partial x_i}\frac{\partial \phi_j}{\partial x_j^*}\\
&= \begin{bmatrix} \frac{\partial r_i}{\partial x_i} & \frac{\partial \phi_i}{\partial x_i} \end{bmatrix} \begin{bmatrix} \frac{\partial^2 f(x_i, x_j)}{\partial r_i \partial r_j} & \frac{\partial^2 f(x_i, x_j)}{\partial r_i \partial \phi_j} \\ \frac{\partial^2 f(x_i, x_j)}{\partial \phi_i \partial r_j} & \frac{\partial^2 f(x_i, x_j)}{\partial \phi_i \partial \phi_j} \end{bmatrix} \begin{bmatrix} \frac{\partial r_j}{\partial x_j^*} \\ \frac{\partial \phi_j}{\partial x_j^*} \end{bmatrix}.
\end{aligned}$$

Using the properties $r = \sqrt{x x^*}$ and $\phi = \arctan\left(\frac{-\mathrm{j}(x - x^*)}{x + x^*}\right)$, we get

$$\begin{aligned}
\frac{\partial^2 f(x_i, x_j)}{\partial x_i \partial x_j^*} &= \frac{1}{2}\,\mathrm{e}^{-\mathrm{j}\,\phi_i} \begin{bmatrix} 1 & \frac{-\mathrm{j}}{r_i} \end{bmatrix} \begin{bmatrix} \frac{\partial^2 f(x_i, x_j)}{\partial r_i \partial r_j} & \frac{\partial^2 f(x_i, x_j)}{\partial r_i \partial \phi_j} \\ \frac{\partial^2 f(x_i, x_j)}{\partial \phi_i \partial r_j} & \frac{\partial^2 f(x_i, x_j)}{\partial \phi_i \partial \phi_j} \end{bmatrix} \frac{1}{2}\,\mathrm{e}^{\mathrm{j}\,\phi_j} \begin{bmatrix} 1 \\ \frac{\mathrm{j}}{r_j} \end{bmatrix}\\
&= \frac{1}{4}\,\mathrm{e}^{-\mathrm{j}(\phi_i - \phi_j)} \begin{bmatrix} 1 & \frac{-\mathrm{j}}{r_i} \end{bmatrix} \begin{bmatrix} \frac{\partial^2 f(x_i, x_j)}{\partial r_i \partial r_j} & \frac{\partial^2 f(x_i, x_j)}{\partial r_i \partial \phi_j} \\ \frac{\partial^2 f(x_i, x_j)}{\partial \phi_i \partial r_j} & \frac{\partial^2 f(x_i, x_j)}{\partial \phi_i \partial \phi_j} \end{bmatrix} \begin{bmatrix} 1 \\ \frac{\mathrm{j}}{r_j} \end{bmatrix}\\
&= \frac{1}{4}\,\mathrm{e}^{-\mathrm{j}(\phi_i - \phi_j)} \left(\frac{\partial^2 f(x_i, x_j)}{\partial r_i \partial r_j} + \frac{1}{r_i r_j}\frac{\partial^2 f(x_i, x_j)}{\partial \phi_i \partial \phi_j} + \mathrm{j}\,\frac{1}{r_j}\frac{\partial^2 f(x_i, x_j)}{\partial r_i \partial \phi_j} - \mathrm{j}\,\frac{1}{r_i}\frac{\partial^2 f(x_i, x_j)}{\partial \phi_i \partial r_j}\right).
\end{aligned}$$

A4 Special Integrals

The joint PDF of $x_i = r_i \, \mathrm{e}^{\mathrm{j}\,\phi_i}$ and $x_j = r_j \, \mathrm{e}^{\mathrm{j}\,\phi_j}$ for $i \neq j$ can be expressed in the polar coordinates as

$$p_{X_i,X_j,i\neq j}(r_i, r_j, \phi_i, \phi_j) = d \, \mathrm{e}^{-\left(a^2 r_i^2 + b^2 r_j^2 - 2abcr_i rj\right)},$$

where

$$a = \frac{1}{\sqrt{1 - \left|\rho_{x_i,x_j}\right|^2}\sigma_{x_i}}, \qquad b = \frac{1}{\sqrt{1 - \left|\rho_{x_i,x_j}\right|^2}\sigma_{x_j}}$$

$$c = \Re\{\rho_{x_i,x_j} \, \mathrm{e}^{-\mathrm{j}(\phi_i - \phi_j)}\}, \qquad d = \frac{1}{\pi^2 \left(1 - \left|\rho_{x_i,x_j}\right|^2\right) \sigma_{x_i}^2 \sigma_{x_j}^2}.$$

In this section, two different integrals as functions of the joint PDF w.r.t. the radii r_i and r_j are computed, viz.,

- $w_1(\phi_i - \phi_j) = \int_0^\infty \int_0^\infty p_{X_i,X_j,i\neq j}(r_i, r_j, \phi_i, \phi_j) r_j \, \mathrm{d}r_i \, \mathrm{d}r_j$
- $w_2(\phi_i - \phi_j) = \int_0^\infty \int_0^\infty p_{X_i,X_j,i\neq j}(r_i, r_j, \phi_i, \phi_j) \, \mathrm{d}r_i \, \mathrm{d}r_j$

The first integral calculates to

$$\begin{aligned}
w_1(\phi_i - \phi_j) &= \int_0^\infty \int_0^\infty p_{X_i,X_j,i\neq j}(r_i, r_j, \phi_i, \phi_j) r_j \, \mathrm{d}r_i \, \mathrm{d}r_j \\
&\overset{z = ar_i - bcr_j}{=} \frac{d}{a} \int_0^\infty r_j \, \mathrm{e}^{-b^2(1-c^2)r_j^2} \int_{-bcr_j}^\infty \mathrm{e}^{-z^2} \, \mathrm{d}z \, \mathrm{d}r_j \\
&= \sqrt{\frac{\pi}{4}} \frac{d}{a} \int_0^\infty r_j \, \mathrm{e}^{-b^2(1-c^2)r_j^2} \left(1 + \operatorname{erf}\left(bcr_j\right)\right) \mathrm{d}r_j \\
&\overset{x = \sqrt{2}b\sqrt{1-c^2}r_j}{=} \frac{d}{a} \frac{\pi}{\sqrt{2}b^2\left(1-c^2\right)} \frac{1}{\sqrt{2\pi}} \int_0^\infty x \, \mathrm{e}^{-\frac{x^2}{2}} \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{c}{\sqrt{1-c^2}} \frac{x}{\sqrt{2}}\right)\right) \mathrm{d}x.
\end{aligned}$$

Using the equality (1,011.2) in [76], the first integral reduces to

$$\begin{aligned}
w_1(\phi_i - \phi_j) &= \frac{\sqrt{1 - \left|\rho_{x_i,x_j}\right|^2}}{4\pi\sqrt{\pi}\sigma_{x_i}\left(1-c\right)} \\
&= \frac{\sqrt{1 - \left|\rho_{x_i,x_j}\right|^2}}{4\pi\sqrt{\pi}\sigma_{x_i}\left(1 - \Re\{\rho_{x_i,x_j} \, \mathrm{e}^{-\mathrm{j}(\phi_i - \phi_j)}\}\right)}.
\end{aligned} \tag{A2}$$

Analogously, the second integral can be computed as

$$\begin{aligned}
w_2(\phi_i - \phi_j) &= \int_0^\infty \int_0^\infty p_{X_i,X_j,i\neq j}(r_i, r_j, \phi_i, \phi_j) \, \mathrm{d}r_i \, \mathrm{d}r_j \\
&\overset{z = ar_i - bcr_j}{=} \frac{d}{a} \int_0^\infty \mathrm{e}^{-b^2(1-c^2)r_j^2} \int_{-bcr_j}^\infty \mathrm{e}^{-z^2} \, \mathrm{d}z \, \mathrm{d}r_j \\
&= \sqrt{\frac{\pi}{4}} \frac{d}{a} \int_0^\infty \mathrm{e}^{-b^2(1-c^2)r_j^2} \left(1 + \operatorname{erf}\left(bcr_j\right)\right) \mathrm{d}r_j \\
&\overset{x = \sqrt{2}b\sqrt{1-c^2}r_j}{=} \frac{d}{a} \frac{\pi}{b\sqrt{1-c^2}} \frac{1}{\sqrt{2\pi}} \int_0^\infty \mathrm{e}^{-\frac{x^2}{2}} \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{c}{\sqrt{1-c^2}} \frac{x}{\sqrt{2}}\right)\right) \mathrm{d}x.
\end{aligned}$$

Using the equality (1,010.2) in [76], we obtain

$$\begin{aligned} w_2(\phi_i - \phi_j) &= \frac{d}{a}\frac{\pi}{b\sqrt{1-c^2}}\left(\frac{1}{4} + \frac{1}{2\pi}\arcsin(c)\right) \\ &= \frac{1}{\pi\sigma_{x_i}\sigma_{x_j}\sqrt{1-\Re\{\rho_{x_i,x_j}\,\mathrm{e}^{-\mathrm{j}(\phi_i-\phi_j)}\}^2}}\left(\frac{1}{4} + \frac{1}{2\pi}\arcsin(\Re\{\rho_{x_i,x_j}\,\mathrm{e}^{-\mathrm{j}(\phi_i-\phi_j)}\})\right). \end{aligned} \tag{A3}$$

A5 Derivative of arcsin(A)

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a quadratic matrix, whose non-diagonal elements depend on a parameter b and diagonal elements are constant. Then, the derivative of $\arcsin(\mathbf{A})$ w.r.t. to b is given by

$$\begin{aligned} \frac{\partial \arcsin(\mathbf{A})}{\partial b} &= \sum_{k=1}^{N}\sum_{l=1}^{N}\frac{\partial \arcsin(a_{kl})}{\partial b}\mathbf{e}_k\mathbf{e}_l^{\mathrm{T}} \\ &= \sum_{k=1}^{N}\sum_{\substack{l=1\\ l\neq k}}^{N}\frac{1}{\sqrt{1-a_{kl}^2}}\frac{\partial a_{kl}}{\partial b}\mathbf{e}_k\mathbf{e}_l^{\mathrm{T}}, \quad \text{since } a_{kk} \text{ is constant, } k = 1,\cdots,N \\ &= \mathrm{nondiag}\left(\mathbf{1}_{N,N} - \mathrm{nondiag}\left(\mathbf{A}\right)^{\circ^2}\right)^{\circ^{-1/2}} \circ \frac{\partial \mathbf{A}}{\partial b}. \end{aligned}$$

Acronyms

ADC Analog-to-Digital Converter
AOD Angle of Departure
AWGN Additive White Gaussian Noise

BER Bit Error Ratio
BS Base Station

CE Constant Envelope
CEQ Constant Envelope Quantizer
CPM Continuous Phase Modulation
CSI Channel State Information

DAC Digital-to-Analog Converter
DTC Digital-to-Time Converter
DTFT Discrete Time Fourier Transform

FIR Finite Impulse Response

GMSK Gaussian Minimum-Shift Keying

HDTV High Definition Television

i.i.d. independent and identically distributed
IBI Inter-Block Interference
ISI Inter-Symbol Interference

KKT Karush-Kuhn-Tucker

LCA Linear Covariance Approximation

MIMO Multiple-Input Multiple-Output
MISO Multiple-Input Single-Output
MMSE Minimum Mean Squared Error
mmW Millimeter-Wave
MSE Mean Squared Error
MSK Minimum-Shift Keying
MSM Maximum Safety Margin
MU Multi-User
MUI Multi-User Interference

PA Power Amplifier
PAPR Peak-to-Average-Power Ratio
PDF Probability Density Function
PSD Power Spectral Density
PSK Phase-Shift Keying

QAM Quadrature Amplitude Modulation
QCE Quantized Constant Envelope

RF Radio Frequency
RRC Root-Raised-Cosine

SER Symbol Error Ratio
SINR Signal-to-Interference-Noise Ratio
SISO Single-Input Single-Output
SLNR Signal-to-Leakage-plus-Noise Ratio
SNR Signal-to-Noise Ratio
SR Symbol Region
SU Single-User

UHDV Ultra-High Definition Video

w.r.t. with respect to
WF Wiener Filter

List of Figures

List of Tables

Bibliography

[1] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G Mobile and Wireless Communications: The vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[2] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[4] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, 2013.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[6] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An Overview of Massive MIMO: Benefits and Challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[7] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-Wave Massive MIMO: The Next Wireless Revolution?" *IEEE Communications Magazine*, vol. 52, no. 9, pp. 56–62, 2014.

[8] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A Survey of Millimeter Wave Communications (mmWave) for 5G: Opportunities and Challenges," *Wireless Networks*, vol. 21, no. 8, pp. 2657–2676, 2015.

[9] T. S. Rappaport, *Millimeter Wave Wireless Communications*. Upper Saddle River N.J.: Prentice Hall, 2015.

[10] "EARTH PROJECT INFSO-ICT-247733 EARTH Deliverable D2.3," January 31, 2012.

[11] O. Blume, D. Zeller, and U. Barth, "Approaches to Energy Efficient Wireless Access Networks," in *Proc. 4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, 2010.

[12] P. Varahram, S. Mohammady, B. M. Ali, and N. Sulaiman, *Power Efficiency in Broadband Wireless Communications*, 1st ed. Boca Raton: CRC Press, 2014.

[13] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-Constrained Modulation Optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, 2005.

[14] S. K. Mohammed and E. G. Larsson, "Single-User Beamforming in Large-Scale MISO Systems with Per-Antenna Constant-Envelope Constraints: The Doughnut Channel," *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 3992–4005, 2012.

[15] ——, "Per-Antenna Constant Envelope Precoding for Large Multi-User MIMO Systems," *IEEE Transactions on Communications*, vol. 61, no. 3, pp. 1059–1071, 2013.

[16] ——, "Constant-Envelope Multi-User Precoding for Frequency-Selective Massive MIMO Systems," *IEEE Wireless Communications Letters*, vol. 2, no. 5, pp. 547–550, 2013.

[17] C. Mollen and E. G. Larsson, "Multiuser MIMO Precoding with Per-Antenna Continuous-Time Constant-Envelope Constraints," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2015, pp. 261–265.

[18] J.-C. Chen, "Low-Complexity Constant Envelope Precoding Using Finite Resolution Phase Shifters for Multiuser MIMO Systems With Large Antenna Arrays," *IEEE Transactions on Vehicular Technology*, p. 1, 2018.

[19] J. Pan and W.-K. Ma, "Constant Envelope Precoding for Single-User Large-Scale MISO Channels: Efficient Precoding and Optimal Designs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 982–995, 2014.

[20] S. Zhang, R. Zhang, and T. J. Lim, "MISO Multicasting With Constant Envelope Precoding," *IEEE Wireless Communications Letters*, vol. 5, no. 6, pp. 588–591, 2016.

[21] F. Liu, C. Masouros, P. V. Amadori, and H. Sun, "An Efficient Manifold Algorithm for Constructive Interference Based Constant Envelope Precoding," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1542–1546, 2017.

[22] P. V. Amadori and C. Masouros, "Constant Envelope Precoding by Interference Exploitation in Phase Shift Keying-Modulated Multiuser Transmission," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 538–550, 2017.

[23] H. Shen, W. Xu, A. Lee Swindlehurst, and C. Zhao, "Transmitter Optimization for Per-Antenna Power Constrained Multi-Antenna Downlinks: An SLNR Maximization Methodology," *IEEE Transactions on Signal Processing*, vol. 64, no. 10, pp. 2712–2725, 2016.

[24] S. Zhang, R. Zhang, and T. J. Lim, "Constant Envelope Precoding for MIMO Systems," *IEEE Transactions on Communications*, vol. 66, no. 1, pp. 149–162, 2018.

[25] A. Mezghani, R. Ghiat, and J. A. Nossek, "Transmit Processing with Low Resolution D/A-Converters," in *Proc. 16th IEEE International Conference on Electronics, Circuits and Systems - (ICECS)*, 2009, pp. 683–686.

[26] O. B. Usman, H. Jedda, A. Mezghani, and J. A. Nossek, "MMSE Precoder for Massive MIMO Using 1-Bit Quantization," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 3381–3385.

[27] R. Price, "A Useful Theorem for Nonlinear Devices Having Gaussian Inputs," *IEEE Transactions on Information Theory*, vol. 4, no. 2, pp. 69–72, 1958.

[28] H. Jedda, J. A. Nossek, and A. Mezghani, "Minimum BER Precoding in 1-Bit Massive MIMO Systems," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2016.

[29] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Nonlinear 1-Bit Precoding for Massive MU-MIMO with Higher-Order Modulation," in *Proc. 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 763–767.

[30] ——, "Quantized Precoding for Massive MU-MIMO," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 4670–4684, 2017.

[31] H. Jedda, A. Mezghani, J. A. Nossek, and A. L. Swindlehurst, "Massive MIMO Downlink 1-Bit Precoding with Linear Programming for PSK Signaling," in *Proc. 18th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2017.

[32] O. Castañeda, T. Goldstein, and C. Studer, "POKEMON: A Non-Linear Beamforming Algorithm for 1-Bit Massive MIMO," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 3464–3468.

[33] A. Swindlehurst, A. Saxena, A. Mezghani, and I. Fijalkow, "Minimum Probability-of-Error Perturbation Precoding for the One-Bit Massive MIMO Downlink," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 6483–6487.

[34] M. Shao, Q. Li, and W.-K. Ma, "One-Bit Massive MIMO Precoding via a Minimum Symbol-Error Probability Design," *arXiv:1803.04787*, Mar. 2018.

[35] O. Castañeda, S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "1-bit Massive MU-MIMO Precoding in VLSI," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 7, no. 4, pp. 508–522, 2017.

[36] H. Jedda, A. Mezghani, J. A. Nossek, and A. L. Swindlehurst, "Massive MIMO Downlink 1-Bit Precoding for Frequency Selective Channels," in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 2017.

[37] L. T. N. Landau and R. C. de Lamare, "Branch-and-Bound Precoding for Multiuser MIMO Systems with 1-Bit Quantization," *IEEE Wireless Communications Letters*, 2017.

[38] A. Noll, H. Jedda, and J. A. Nossek, "PSK Precoding in Multi-User MISO Systems," in *Proc. 21st International ITG Workshop on Smart Antennas (WSA)*. Berlin: VDE Verlag GMBH, 2017.

[39] S. Jacobsson, O. Castaneda, C. Jeon, G. Durisi, and C. Studer, "Nonlinear Precoding for Phase-Quantized Constant-Envelope Massive MU-MIMO-OFDM," in *2018 25th International Conference on Telecommunications (ICT)*. IEEE, 2018, pp. 367–372.

[40] A. Nedelcu, F. Steiner, M. Staudacher, G. Kramer, W. Zirwas, R. S. Ganesan, P. Baracca, and S. Wesemann, "Quantized Precoding for Multi-Antenna Downlink Channels with MAGIQ," in *Proc. 22nd International ITG Workshop on Smart Antennas (WSA)*. Berlin: VDE Verlag GMBH, 2018.

[41] M. Kazemi, H. Aghaeinia, and T. M. Duman, "Discrete-Phase Constant Envelope Precoding for Massive MIMO Systems," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2011–2021, 2017.

[42] 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 Communications Magazine*, vol. 51, no. 10, pp. 162–171, 2013.

[43] C. Masouros and G. Zheng, "Exploiting Known Interference as Green Signal Power for Downlink Beamforming Optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 14, pp. 3628–3640, 2015.

[44] H. Jedda, A. Mezghani, A. L. Swindlehurst, and J. A. Nossek, "Quantized Constant Envelope Precoding with PSK and QAM Signaling," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8022–8034, © 2018 IEEE.

[45] H. Jedda and J. A. Nossek, "Quantized Constant Envelope Precoding for Frequency Selective Channels," in *2018 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2018, pp. 213–217.

[46] H. Jedda, M. M. Ayub, J. Munir, A. Mezghani, and J. A. Nossek, "Power- and Spectral Efficient Communication System Design Using 1-Bit Quantization," in *Proc. International Symposium on Wireless Communication Systems (ISWCS)*, 2015, pp. 296–300.

[47] H. Jedda, A. Mezghani, and J. A. Nossek, "Spectral Shaping with Low Resolution Signals," in *Proc. 49th Asilomar Conference on Signals, Systems and Computers*, 2015, pp. 1437–1441.

[48] C. Mollen, E. G. Larsson, and T. Eriksson, "Waveforms for the Massive MIMO Downlink: Amplifier Efficiency, Distortion, and Performance," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 5050–5063, 2016.

[49] H. Jedda and J. A. Nossek, "On the Statistical Properties of Constant Envelope Quantizers," *IEEE Wireless Communications Letters*, vol. 7, no. 6, pp. 1006–1009, © 2018 IEEE.

[50] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd ed., ser. Artech House microwave library. Boston, Mass. and London: Artech House, 2006.

[51] B. Lehmeyer, A. Mezghani, and J. A. Nossek, "Electronic Amplifier for Amplifying an Input Signal," International Patent WO 2018/024 756 A1, 2018.

[52] B. Razavi, *Principles of Data Conversion System Design*. New York: IEEE Press, 1995.

[53] J. Groe, "Polar Transmitters for Wireless Communications," *IEEE Communications Magazine*, vol. 45, no. 9, pp. 58–63, 2007.

[54] M. Abolfadl Ibrahim, "The Polar Transmitter: Analysis and Algorithms," Doctoral Thesis, Universität Stuttgart, 2015. [Online]. Available: http://elib.uni-stuttgart.de/handle/11682/3647

[55] H. H. Andersen, M. Højbjerre, D. Sørensen, and P. S. Eriksen, "The Multivariate Complex Normal Distribution," in *Linear and Graphical Models*, ser. Lecture Notes in Statistics, P. Diggle, S. Fienberg, K. Krickeberg, I. Olkin, N. Wermuth, H. H. Andersen, M. Højbjerre, D. Sørensen, and P. S. Eriksen, Eds. New York, NY: Springer New York, 1995, vol. 101, pp. 15–37.

[56] E. McMahon, "An Extension of Price's theorem (Corresp.)," *IEEE Transactions on Information Theory*, vol. 10, no. 2, p. 168, 1964.

[57] A. Papoulis, "Comments on 'An Extension of Price's Theorem' by McMahon, E. L," *IEEE Transactions on Information Theory*, vol. 11, no. 1, p. 154, 1965.

[58] R. Pawula, "A Modified Version of Price's Theorem," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 285–288, 1967.

[59] W. McGee, "Circularly Complex Gaussian Noise–A price Theorem and a Mehler Expansion (Corresp.)," *IEEE Transactions on Information Theory*, vol. 15, no. 2, pp. 317–319, 1969.

[60] A. van den Bos, "Price's Theorem for Complex Variates," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 286–287, 1996.

[61] J. J. Bussgang, "Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals." [Online]. Available: http://hdl.handle.net/1721.1/4847

[62] A. Saleh and R. Valenzuela, "A Statistical Model for Indoor Multipath Propagation," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 2, pp. 128–137, 1987.

[63] L. M. Correia and P. Smulders, "Characterisation of Propagation in 60 GHz Radio Channels," *Electronics & Communication Engineering Journal*, vol. 9, no. 2, pp. 73–80, 1997.

[64] A. M. Sayeed, "Deconstructing Multiantenna Fading Channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, 2002.

[65] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and Temporal Characteristics of 60-GHz Indoor Channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 3, pp. 620–630, 2002.

[66] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[67] A. Forenza, D. J. Love, and R. W. Heath, "Simplified Spatial Correlation Models for Clustered MIMO Channels With Different Array Configurations," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, pp. 1924–1934, 2007.

[68] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed., ser. Wiley series in probability and mathematical statistics. New York: Wiley, 1970.

[69] A. Kakkavas, "Precoding Techniques for Coarsely Quantized MIMO Systems," Master's thesis, Technical University of Munich, 2015.

[70] M. Joham, W. Utschick, and J. A. Nossek, "Linear Transmit Processing in MIMO Communications Systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2700–2712, 2005.

[71] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge: Cambridge University Press, 2004.

[72] G. B. Dantzig and M. N. Thapa, *Linear Programming*, ser. Springer series in operations research. New York and London: Springer, 1997-.

[73] M. Joham, *Optimization of Linear and Nonlinear Transmit Signal Processing*. Aachen: Shaker, 2004.

[74] T. Aulin and C. Sundberg, "Continuous Phase Modulation–Part I: Full Response Signaling," *IEEE Transactions on Communications*, vol. 29, no. 3, pp. 196–209, 1981.

[75] K. Murota and K. Hirade, "GMSK Modulation for Digital Mobile Radio Telephony," *IEEE Transactions on Communications*, vol. 29, no. 7, pp. 1044–1050, 1981.

[76] D. B. Owen, "A table of Normal Integrals," *Communications in Statistics - Simulation and Computation*, vol. 9, no. 4, pp. 389–419, 2007.