Technische Universität München

Linear and Non-linear Estimation Methods for Single Carrier and Multicarrier Coarsely Quantized MIMO Systems

Master's Thesis

Francisco Javier García Gómez

Supervisor: M. Sc. Jawad Munir
Examiner: Univ.-Prof. i.R. Dr. techn. Dr. h. c. Josef A. Nossek
Date: 20.08.2016

Department of Electrical Engineering and Information Technology
Fachgebiet Methoden der Signalverarbeitung
Univ.-Prof. Dr.-Ing. Wolfgang Utschick

Fachgebiet Methoden der Signalverarbeitung
Technische Universität München
D-80290 München
Germany

http://www.msv.ei.tum.de

Technische Universität München

Linear and Non-linear Estimation Methods for Single Carrier and Multicarrier Coarsely Quantized MIMO Systems

Master's Thesis

Francisco Javier García Gómez

Supervisor: M. Sc. Jawad Munir
Examiner: Univ.-Prof. i.R. Dr. techn. Dr. h. c. Josef A. Nossek
Date: 20.08.2016

Department of Electrical Engineering and Information Technology
Fachgebiet Methoden der Signalverarbeitung
Univ.-Prof. Dr.-Ing. Wolfgang Utschick

Abstract

This thesis deals with channel estimation and equalization in frequency-selective Multiple Input-Multiple Output (MIMO) channels with 1-bit quantization at the receiver. Most current literature on quantized MIMO focuses on flat-fading channels, which motivates the topic for this thesis. The study intends to be as generic as possible, and therefore assumes no special structure of the channel. Furthermore, no joint processing or Channel State Information (CSI) is assumed at the transmitter, making the results valid also for multiple-access MIMO channels.

Three different communication schemes are explored: Orthogonal Frequency Division Multiplexing (OFDM), single-carrier with cyclic prefix (SC), and single carrier without cyclic prefix (NCP). Equivalent problem models are derived for the all the settings. A Cramér-Rao Lower Bound (CRLB) is derived for the estimation problem, as well as for the joint estimation of channel and data. The existing nonlinear algorithms Expectation Maximization (EM) and Generalized Approximate Message Passing (GAMP), as well as a linear estimator based on the Bussgang theorem, are adapted to the problems of estimation an detection. An optimal A-Posteriori Probability (APP) Soft Input-Soft Output (SISO) equalizer is designed, which has high complexity but is good as a performance benchmark. A turbo approach for Joint Channel and Data Estimation (JCD) is also explored.

Through simulations, all the methods and settings are compared. The advantage of SC over OFDM is shown. NCP turns out to be the best option at low SNR, but saturates early at higher SNR. The nonlinear algorithms exhibit better performance than the linear ones at the cost of higher computational complexity. An analysis of the required number of pilots for channel estimation is carried out. Higher order modulations are also explored, and PSK is compared to QAM for modulation orders of 8 and 16. The JCD scheme is shown to considerably improve the results for both estimation and detection.

Contents

Bibliography 67

Chapter 1

Introduction

The fifth generation (5G) of mobile communications is expected to increase spectral and energy efficiency by several orders of magnitude [1]. A number of new technologies are being considered to fulfill this requirement. One of them is the use of Multiple Input-Multiple Output (MIMO) systems with large numbers of antennas [2], which is expected to increase array gain without bound. Higher frequency bands [3], in the order of 6-100 GHz are also being considered. They would allow for larger bandwidth and lower interference.

These changes place stringent requirements on the receiver-side analog-to-digital converters (ADCs). Due to the high frequency and large bandwidth, the ADCs need to operate at high sampling rate, which leads to high power consumption. This problem is further increased in a Massive MIMO setting, which requires a large amount of ADCs.

The power consumption of the ADCs grows exponentially with the number of bits, as shown in [4]. Therefore, the use of low-resolution (1 to 3 bits) ADCs has been proposed as a way to address the power consumption problem. In this paper, we focus on the 1-bit quantization case.

Various aspects of the 1-bit quantized MIMO channel have been analyzed in recent work. An analysis in [5] and [6] shows that the capacity loss due to the 1-bit quantization can be compensated for by increasing the number of receive antennas.

Different algorithms for channel estimation are discussed in [7]. They obtain good results but are iterative and nonlinear. Furthermore, convergence is not guaranteed, especially if the channel taps are not i.i.d. Gaussian, which is the case in practical scenarios.

A linear MMSE receiver for equalization of quantized MIMO channels is proposed in [8], and an iterative nonlinear one in [9]. The nonlinear equalizer achieves better BER performance at the cost of increased computational complexity.

Finally, the problem of joint channel and data estimation (JCD) is treated in [10]. The authors show that this approach greatly improves the results, and requires fewer pilots.

One of the main shortcomings of the mentioned contributions is that all of them consider only flat-fading channels. With unquantized systems, this assumption can be justified by the use of multi-carrier modulations, such as Orthogonal Frequency Division Multiplexing

(OFDM), because then the channel in each subcarrier is flat. However, in the quantized case, the subcarriers can no longer be separated without loss. We note that OFDM is still attractive in this case because it allows for uplink multiplexing of different users' streams.

To the best of our knowledge, very little work has been done on quantized frequency-selective MIMO channels or on the loss incurred by using multi-carrier modulations. A model for channel estimation and equalization in quantized MIMO OFDM systems is proposed in [11]. However, it relies on convex optimization algorithms, which, for the large-dimensional problems at hand, have high computational cost.

In this thesis, we develop a model for channel estimation and for equalization of three different kinds of 1-bit quantized MIMO frequency-selective systems: OFDM, single-carrier without cyclic prefix, and single carrier with cyclic prefix. We derive a Cramér-Rao lower bound for both the separate and the joint (JCD) estimation problem.

We adapt the existing nonlinear iterative algorithms Expectation Maximization (EM) and Generalized Approximate Message Passing (GAMP) to solve these estimation problems in the minimum mean square error (MMSE) sense. Additionally, we develop a linear estimator based on the Bussgang theorem, which greatly reduces the computational complexity in the OFDM case because it allows for per-subcarrier equalization.

An optimal A-posteriori Probability Soft-Input-Soft-Output (APP SISO) equalizer is also developed in this thesis. It has prohibitive computational complexity, but is useful as a benchmark for the equalization task. Finally, we adapt a turbo-like Joint Channel and Data estimator [12] to our setting.

Through simulations, we then compare the performance of all the estimation methods in different scenarios. We show that the nonlinear methods perform notably better than the linear Bussgang estimator, and that JCD also greatly increases performance. The single-carrier systems with cyclic prefix shown to outperform OFDM, and the systems without cyclic prefix are best at low SNR and worst at high SNR. Finally, we show that 8-PSK is better suited to the the 1-bit quantized system than 8-QAM, but for higher order modulations again QAM is the best option.

We note that all our analysis does not assume any joint processing or Channel State Information (CSI) at the transmitter. Therefore, our findings are also applicable to Multi-User MIMO uplink channels, by considering each transmit antenna (or group of them) as a separate user.

Chapter 2

MIMO with 1-bit Quantization at the receiver: System Model

In this chapter, we will derive system models for channel estimation and equalization in both multi-carrier and single-carrier 1-bit quantized MIMO schemes. We will find out that the four problems can be expressed with the same model, and therefore the same techniques can be applied to solve all of them.

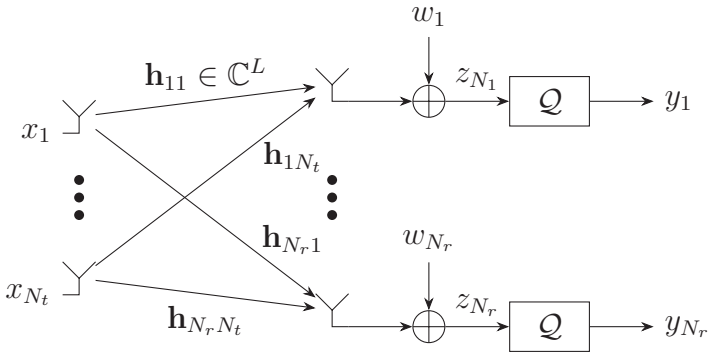

Figure 2.1: MIMO system with 1-bit quantization

We consider the system depicted in Fig. 2.1. It has N_r receive antennas and N_t transmit antennas. Each pair of transmit antenna $n_t \in \{1, \ldots, N_t\}$ and receive antenna $n_r \in \{1, \ldots, N_r\}$ has an arbitrary, frequency-selective channel impulse response of L taps, denoted by $\mathbf{h}_{n_r n_t} \in \mathbb{C}^L$.

The quantizer blocks $\mathcal{Q}$ apply the 1-bit quantization function separately to the real and imaginary parts of their input:

$$\mathcal{Q}(z) = \operatorname{sign}(\Re\{z\}) + j \operatorname{sign}(\Im\{z\}). \tag{2.1}$$

The signal after each quantizer is then:

$$y_{n_r}[n] = \mathcal{Q}\left(\sum_{n_t=0}^{N_t} h_{n_r n_t}[n] \star x_{n_t}[n] + w_{n_r}[n]\right), \tag{2.2}$$

where $x_{n_t}[n]$ is the transmitted signal at antenna $n_t \in \{1, \ldots, N_t\}$; y_{n_r} is the RX quantizer output at antenna $n_r \in \{1, \ldots, N_r\}$, and $w_{n_r}[n]$ is the additive Gaussian noise at receive antenna n_r. We will assume that this noise is temporally and spatially uncorrelated, but we will allow different noise variances at each antenna:

$$\sigma_{n_r}^2 = \mathrm{E}\left\{|w_{n_r}[n]|^2\right\}. \tag{2.3}$$

2.1 Multi-Carrier Scheme (OFDM)

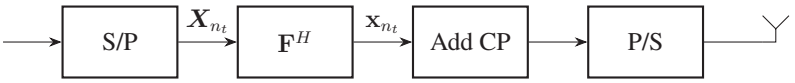

Figure 2.2: Quantized MIMO OFDM system: Block diagram at each transmit antenna

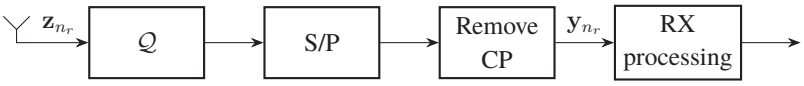

Figure 2.3: Quantized MIMO OFDM system: Block diagram at each RX antenna

We consider the use of Orthogonal Frequency Division Multiplexing (OFDM) [13] in the quantized MIMO system. Fig. 2.2 depicts the block diagram of the transmitter at each antenna $n_t \in \{1, \ldots, N_t\}$, which has the same structure as in the unquantized case. The data is arranged in blocks of N symbols $\boldsymbol{\mathcal{X}}_{n_t} \in \mathbb{C}^N$, where N is the number of OFDM subcarriers. An Inverse Fast Fourier Transform (IFFT) is applied to each block, and then a cyclic prefix (CP) of length L (equal to the length of the impulse response of the channel) is added. This makes the equivalent channel convolution matrix $\mathbf{H}_{n_r n_t} \in \mathbb{C}^{N \times N}$ between each pair of antennas $(n_r, n_t) \in \{1, \ldots N_r\} \times \{1, \ldots, N_t\}$ circulant:

$$[\mathbf{H}_{n_r n_t}]_{ij} = h_{n_r n_t}\left[(i-j) \mod N\right], \quad i, j \in \{1, \ldots, N\}. \tag{2.4}$$

Consider the transmission of M OFDM blocks, horizontally stacked in a matrix $\mathbf{X}_{n_t \cdot \cdot} \in \mathbb{C}^{N \times M}$. After quantization and removal of the cyclic prefix at the receiver, the signal $\mathbf{Y}_{n_r \cdot \cdot} \in \mathbb{C}^{N \times M}$ at receive antenna n_r, from (2.2), is:

$$\mathbf{Y}_{n_r \cdot \cdot} = \mathcal{Q}\left(\mathbf{H}_{n_r n_t}\mathbf{F}^H\mathbf{X}_{n_t \cdot \cdot} + \mathbf{W}_{n_r \cdot \cdot}\right), \tag{2.5}$$

where $\mathbf{F} \in \mathbb{C}^{N \times N}$ is an N-point unitary DFT matrix (whose entries have magnitude $1/\sqrt{N}$), and $\mathbf{W}_{n_r \cdot \cdot} \in \mathbb{C}^{N \times M}$ contains i.i.d. additive complex Gaussian noise with variance $\sigma_{n_r}^2$.

Due to the circulant property of $\mathbf{H}_{n_r n_t}$, it is diagonalized by Fourier matrices:

$$\mathbf{F}\mathbf{H}_{n_r n_t}\mathbf{F}^H = \mathbf{\Lambda}_{n_r n_t} = \mathrm{diag}\left\{\mathbf{F}_{N \times L}\mathbf{h}_{n_r n_t}\right\}. \tag{2.6}$$

The matrix $\mathbf{F}_{N \times L} = \sqrt{N}\mathbf{F}_{(1:N)(1:L)}$ contains its L first columns, with their entries normalized to magnitude 1. This allows rewriting (2.5) as:

$$\mathbf{Y}_{n_r \cdot \cdot} = \mathcal{Q}\left(\sum_{n_t=1}^{N_t} \mathbf{F}^H\mathbf{\Lambda}_{n_r n_t}\mathbf{X}_{n_t \cdot \cdot} + \mathbf{W}_{n_r}\right), \tag{2.7}$$

2.1.1 OFDM channel estimation

Consider the channel estimation stage, in which T OFDM blocks are sent as pilots. The matrices $\mathbf{X}_{n_t\cdot\cdot}$ in (2.7), now have dimensions $N \times T$. If we vectorize (see Appendix A) the signal $\mathbf{Y}_{n_r\cdot\cdot}$ at each receive antenna, and use $\mathbf{\Lambda}_{n_r n_t} = \mathrm{diag}\left\{\mathbf{F}_{N\times L}\mathbf{h}_{n_r n_t}\right\}$, we can express (2.7) as:

$$\mathbf{y}_{n_r} = \mathcal{Q}\left(\sum_{n_t=1}^{N_t}\left(\boldsymbol{\mathcal{X}}^T_{n_t\cdot\cdot}\diamond\mathbf{F}^H\right)\mathbf{F}_{N\times L}\mathbf{h}_{n_r n_t} + \mathbf{w}_{n_r}\right), \tag{2.8}$$

where $\mathbf{y}_{n_r} \triangleq \mathrm{vec}\left(\mathbf{Y}_{n_r\cdot\cdot}\right)$, and $\mathbf{w}_{n_r} \triangleq \mathrm{vec}\left(\mathbf{W}_{n_r}\right)$, and $\diamond$ denotes the Khatri-Rao product, which is a column-wise Kronecker product defined in Appendix A. Here, we have used the following property of the vectorization operator:

$$\mathrm{vec}\left(\mathbf{A}\mathrm{diag}\left\{\mathbf{b}\right\}\mathbf{C}\right) = \left(\mathbf{C}^T\diamond\mathbf{A}\right)\mathbf{b}. \tag{2.9}$$

Now, we can express the summation over n_t in (2.8) as a matrix-vector product. We define the vector $\mathbf{h}_{n_r} \in \mathbb{C}^{LN_t}$ as the vertical stacking of the channel impulse responses corresponding to receive antenna n_r for all n_t:

$$\mathbf{h}_{n_r} \triangleq \begin{pmatrix}\mathbf{h}_{n_r 1}\\ \mathbf{h}_{n_r 2}\\ \vdots\\ \mathbf{h}_{n_r N_t}\end{pmatrix}. \tag{2.10}$$

Furthermore, we define the matrix $\mathbf{A} \in \mathbb{C}^{NT\times LN_t}$ as:

$$\mathbf{A} = \left[\ \left(\boldsymbol{X}^T_{1\cdot\cdot}\diamond\mathbf{F}^H\right)\mathbf{F}_{N\times L}\ \middle|\ \left(\boldsymbol{X}^T_{2\cdot\cdot}\diamond\mathbf{F}^H\right)\mathbf{F}_{N\times L}\ \middle|\ \cdots\ \middle|\ \left(\boldsymbol{X}^T_{N_t\cdot\cdot}\diamond\mathbf{F}^H\right)\mathbf{F}_{N\times L}\ \right]. \tag{2.11}$$

We can now write (2.8) as:

$$\mathbf{y}_{n_r} = \mathcal{Q}\left(\mathbf{A}\mathbf{h}_{n_r} + \mathbf{w}_{n_r}\right), \quad n_r \in \{1,\ldots,N_r\}. \tag{2.12}$$

2.1.2 OFDM data equalization

Starting from (2.7), we now consider the data transmission stage, where M OFDM blocks are transmitted. The matrices $\mathbf{X}_{n_t\cdot\cdot}$ in (2.7) have dimensions $N \times M$.

The conversion to a matrix-vector product is straightforward in this case. Let us define a vector $\mathbf{x}_m \in \mathbb{C}^{M\times 1}$ that vertically stacks all the data symbols corresponding to block $m \in \{1,\ldots,M\}$:

$$\mathbf{x}_m = \begin{pmatrix}\mathbf{x}_{1m\cdot}\\ \vdots\\ \mathbf{x}_{N_t m\cdot}\end{pmatrix}, \tag{2.13}$$

where $\mathbf{x}_{n_t m\cdot} \in \mathbb{C}^{N\times 1}$ denotes the m-th column of $\mathbf{X}_{n_t\cdot\cdot}$, with $n_t \in \{1,\ldots,N_t\}$. The sensing matrix $\mathbf{A} \in \mathbb{C}^{NN_r\times NN_t}$ is defined as:

$$\mathbf{A} = \begin{pmatrix}\mathbf{F}^H\mathbf{\Lambda}_{11} & \cdots & \mathbf{F}^H\mathbf{\Lambda}_{1N_t}\\ \vdots & \ddots & \vdots\\ \mathbf{F}^H\mathbf{\Lambda}_{N_r 1} & \cdots & \mathbf{F}^H\mathbf{\Lambda}_{N_r N_t}\end{pmatrix}, \tag{2.14}$$

and the observation vector $\mathbf{y}_m \in \mathbb{C}^{NN_r \times 1}$ vertically stacks the received symbols for block m:

$$\mathbf{y}_m = \begin{pmatrix} \mathbf{y}_{1m\cdot} \\ \vdots \\ \mathbf{y}_{N_r m\cdot} \end{pmatrix}, \tag{2.15}$$

where $\mathbf{y}_{n_r m\cdot} \in \mathbb{C}^{N \times 1}$ is the m-th column of $\mathbf{Y}_{n_r \cdot\cdot}$. With these definitions, the model for data equalization is given by:

$$\mathbf{y}_m = \mathcal{Q}\left(\mathbf{A}\mathbf{x}_m + \mathbf{w}_m\right), \quad m \in \{1, \ldots, M\}. \tag{2.16}$$

where $\mathbf{w}_m$ is defined in the same way as $\mathbf{y}_m$, and contains uncorrelated complex Gaussian samples with variance $\sigma^2_{n_r nm}$.

2.2 Single-Carrier with Cyclic Prefix

We now consider the single-carrier (SC) case. To prevent interference between blocks of data, we will first work with a system that also uses a cyclic prefix (CP). The block size is assumed to be N (equal to the number of OFDM subcarriers), and the cyclic prefix has length L. With these assumptions, the received signal at antenna $n_r \in \{1, \ldots, N_r\}$ can be written as:

$$\mathbf{Y}_{n_r \cdot\cdot} = \mathcal{Q}\left(\sum_{n_t=1}^{N_t} \mathbf{H}_{n_r n_t} \mathbf{X}_{n_t \cdot\cdot} + \mathbf{W}_{n_r}\right), \tag{2.17}$$

where $\mathbf{X}_{n_t \cdot\cdot} \in \mathbb{C}^{N \times M}$ horizontally stacks M blocks of transmitted symbols, and $\mathbf{H}_{n_r n_t} \in \mathbb{C}^{N \times N}$ is defined in (2.4).

2.2.1 SC Channel Estimation

For channel estimation, each transmit antenna sends T blocks as pilots. We denote the pilot vector at transmit antenna $n_t \in \{1, \ldots, N_t\}$ and block $t \in \{1, \ldots, T\}$ as $\mathbf{x}_{n_t t\cdot} \in \mathbb{C}^{N \times 1}$. We further define the partial circulant convolution matrix $\mathbf{X}_{n_t t} \in \mathbb{C}^{N \times L}$ in the following way:

$$[\mathbf{X}_{n_t t}]_{n\ell} \triangleq \mathbf{x}_{n_t t\cdot}[(n - \ell) \mod N], \tag{2.18}$$

i.e. the first L columns of a circulant matrix whose first column is $\mathbf{x}_{n_t t\cdot}$. With this definition, we can express the channel estimation problem for single-carrier with the same model of the previous problems:

$$\mathbf{y}_{n_r} = \mathcal{Q}\left(\mathbf{A}\mathbf{h}_{n_r} + \mathbf{w}_{n_r}\right), \tag{2.19}$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{X}_{11} & \cdots & \mathbf{X}_{N_t 1} \\ \vdots & \ddots & \vdots \\ \mathbf{X}_{1T} & \cdots & \mathbf{X}_{N_t T} \end{pmatrix} \in \mathbb{C}^{NT \times LNt}, \tag{2.20}$$

and $\mathbf{h}_{n_r}$ is given by (2.10). Furthermore, $\mathbf{y}_{n_r} = \mathrm{vec}\left(\mathbf{Y}_{n_r \cdot\cdot}\right)$.

2.2.2 SC Data Equalization

The problem formulation for data equalization in the SC case is straightforward from (2.17):

$$\mathbf{y}_m = \mathcal{Q}\left(\mathbf{A}\mathbf{x}_m + \mathbf{w}_m\right), \tag{2.21}$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{H}_{11} & \cdots & \mathbf{H}_{1N_t} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{N_r1} & \cdots & \mathbf{H}_{N_rN_t} \end{pmatrix}, \tag{2.22}$$

and the vectors $\mathbf{y}_m$, $\mathbf{x}_m$, and $\mathbf{w}_m$ are defined in Section 2.1.2.

2.3 Single carrier without cyclic prefix (NCP)

The use of a cyclic prefix comes at the cost of a loss in throughput. In this thesis, we also consider a system without cyclic prefix. This requires interference cancellation in the equalization stage, which results in performance degradation. The trade-off between these two effects will be evaluated through simulations in Chapter 5.

Consider the transmission of T_{ncp} pilots without cyclic prefix from each antenna. The symbol at antenna $n_t \in \{1, \ldots, N_t\}$ and time $t \in \{1, \ldots, T_{\text{ncp}}\}$ is denoted by $x_{n_t}[t]$. The received sequence at antenna $n_r \in \{1, \ldots, N_r\}$ is given by:

$$y_{n_r}[t] = \mathcal{Q}\left(\sum_{n_t=1}^{N_t} h_{n_rn_t}[t] \star x_{n_t}[t] + w_{n_r}[t]\right). \tag{2.23}$$

2.3.1 NCP channel estimation

The channel estimation problem without cyclic prefix can be expressed in a similar way as with the cyclic prefix case. Let us stack all the channel vectors for receive antenna $n_r \in \{1, \ldots, N_r\}$ in a vector $\mathbf{h}_{n_r} \in \mathbb{C}^{LN_t}$, as in (2.10).

For each transmit antenna $n_t \in \{1, \ldots, N_t\}$, we arrange all the T transmitted pilots in the convolution matrix $\mathbf{X}_{n_t} \in \mathbb{C}^{(T_{\text{ncp}}+L-1)\times L}$:

$$[\mathbf{X}_{n_t}]_{n\ell} = \begin{cases} x_{n_t}[t-\ell], & \text{if } 0 \le t-\ell < N \\ 0, & \text{otherwise,} \end{cases} \tag{2.24}$$

with $\ell \in \{0, \ldots, L-1\}$, and $t \in \{0, \ldots, T_{\text{ncp}} + L - 2\}$. If we define the sensing matrix $\mathbf{A} \in \mathbb{C}^{(T_{\text{ncp}}+L-1)\times LN_t}$ as:

$$\mathbf{A} = \begin{pmatrix} \mathbf{X}_1 & \cdots & \mathbf{X}_{N_t} \end{pmatrix}, \tag{2.25}$$

we can yet again express our problem as a quantized matrix-vector multiplication for each receive antenna:

$$\mathbf{y}_{n_r} = \mathcal{Q}\left(\mathbf{A}\mathbf{h}_{n_r} + \mathbf{w}_{n_r}\right), \tag{2.26}$$

where the vector $\mathbf{y}_{n_r} \in \mathbb{C}^{T_{\text{ncp}}+L-1}$ contains the received symbols at each receive antenna n_r. The noise vector $\mathbf{w}_{n_r} \in \mathbb{C}^{T_{\text{ncp}}+L-1}$ contains the corresponding noise.

2.3.2 NCP data equalization

Theoretically, a data equalization model with a convolution matrix for all the received symbols could be set up. However, the solution to this problem would require the inversion of a very large matrix, which is not practical, especially in a quantized system where the number of receive antennas needs to be high. Therefore, we develop a model that allows processing the received data in blocks, by removing the interference from the previous block.

In order to have a similar computational complexity to that of the OFDM and SC models, we will use the same block size of N symbols. The transmitted symbols at antenna n_t and block $m \in \{0, \dots, M-1\}$ are denoted by $\mathbf{x}_{n_t m} \in \mathbb{C}^N$, such that:

$$x_{n_t m}[n] = x_{n_t}[mN + n]. \tag{2.27}$$

Similarly, the received signal at antenna n_r is divided into blocks $\mathbf{y}_{n_r m} \in \mathbb{C}^N$:

$$y_{n_r m}[n] = y_{n_r}[mN + n]. \tag{2.28}$$

With these definitions, the received blocks $\mathbf{y}_{n_r,m}$ depends mostly on the corresponding transmitted blocks $\mathbf{x}_{n_t,m}$, but the first samples of $\mathbf{y}_{n_r,m}$ have interference from the previous transmitted block $\mathbf{x}_{n_t,(m-1)}$. Consider the full channel convolution matrix $\mathbf{H}_{n_r n_t} \in \mathbb{C}^{N\times(N+L-1)}$ for the antenna pair (n_r, n_t):

$$[\mathbf{H}_{n_r n_t}]_{n_1 n_2} = \begin{cases} \mathbf{h}_{n_r n_t}[n_2 - n_1], & \text{if } 0 \le n_2 - n_1 < L \\ 0, & \text{otherwise,} \end{cases} \tag{2.29}$$

with $n_1 \in \{0, \dots, N-1\}$ and $n_2 \in \{0, \dots, N+L-2\}$. Let us split this matrix horizontally into two parts: a square matrix $\overline{\mathbf{H}}_{n_r n_t} \in \mathbb{C}^{N\times N}$ from the left and the remaining rectangular matrix $\tilde{\mathbf{H}}_{n_r n_t} \in \mathbb{C}^{N\times(L-1)}$:

$$\mathbf{H}_{n_r n_t} = \begin{pmatrix} \overline{\mathbf{H}}_{n_r n_t} & \tilde{\mathbf{H}}_{n_r n_t} \end{pmatrix}. \tag{2.30}$$

The m-th received block at antenna n_r is then given by:

$$\overline{\mathbf{y}}_{n_r m} = \mathcal{Q}\left(\sum_{n_t=1}^{N_t} \left(\overline{\mathbf{H}}_{n_r n_t} \overline{\mathbf{x}}_{n_t,m} + \tilde{\mathbf{H}}_{n_r n_t} \overline{\mathbf{x}}_{n_t,(m-1)}[0:(L-1)] \right) + \overline{\mathbf{w}}_{n_r m} \right), \tag{2.31}$$

where $\overline{\mathbf{y}}_{n_r m}$ is the reversed version of $\mathbf{y}_{n_r m}$, and $\overline{\mathbf{x}}_{n_t m}$ is the reversed version of $\mathbf{x}_{n_t m}$. Now, assume that there is a guard period before the transmission of the first block. Then, the term $\tilde{\mathbf{H}}_{n_r n_t} \overline{\mathbf{x}}_{n_t,(m-1)}[0:(L-1)]$ will be $\mathbf{0}$ for $m = 0$. For $m > 0$, this term can be calculated by using the estimated transmitted block $\hat{\overline{\mathbf{x}}}_{n_t,(m-1)}$ at time $m-1$, which will already have been computed. Therefore, this term is always known at the receiver.

With this information, we can write the NCP data equalization problem in a form that is slightly different from what was obtained for the other problems, but that requires minimal modifications to the algorithms:

$$\mathbf{y}_m = \mathcal{Q}\left(\mathbf{A}\mathbf{x}_m + \mathbf{b}_m + \mathbf{w}_m\right), \tag{2.32}$$

where $\mathbf{A} \in \mathbb{C}^{NN_r \times NN_t}$ is given by:

$$\mathbf{A} = \begin{pmatrix} \overline{\mathbf{H}}_{11} & \cdots & \overline{\mathbf{H}}_{1N_t} \\ \vdots & \ddots & \vdots \\ \overline{\mathbf{H}}_{N_r1} & \cdots & \overline{\mathbf{H}}_{N_rN_t} \end{pmatrix}. \tag{2.33}$$

The vectors $\mathbf{y}_m \in \mathbb{C}^{NN_r}$ vertically stacks the reversed signals at all receive antennas:

$$\mathbf{y}_m = \begin{pmatrix} \overline{\mathbf{y}}_{1m} \\ \vdots \\ \overline{\mathbf{y}}_{N_rm} \end{pmatrix}. \tag{2.34}$$

The vector $\mathbf{w}_m \in \mathbb{C}^{NN_r}$ contains Gaussian noise and is defined in the same way as $\mathbf{y}_m$ Similarly, the vector $\mathbf{x}_m \in \mathbb{C}^{NN_t}$ stacks the reversed transmit signals at block m:

$$\mathbf{x}_m = \begin{pmatrix} \overline{\mathbf{x}}_{1m} \\ \vdots \\ \overline{\mathbf{x}}_{N_tm} \end{pmatrix}. \tag{2.35}$$

Finally, the *known* vector $\mathbf{b}_m \in \mathbb{C}^{NN_r}$ is given by:

$$\mathbf{b}_m = \begin{pmatrix} \sum_{n_t=1}^{N_t} \tilde{\mathbf{H}}_{1n_t} \overline{\mathbf{x}}_{n_t,(m-1)}[0:(L-1)] \\ \vdots \\ \sum_{n_t=1}^{N_t} \tilde{\mathbf{H}}_{N_rn_t} \overline{\mathbf{x}}_{n_t,(m-1)}[0:(L-1)] \end{pmatrix}. \tag{2.36}$$

Chapter 3

Cramér-Rao Lower Bound for the Quantized Estimation Problem

Before going into the algorithms for quantized data estimation, in this section we will derive the Cramér-Rao Lower Bound (CRLB) of the problems. The CRLB gives a minimum covariance matrix of an estimation problem, in the following sense: Consider an estimation problem in which a vector of parameters $\boldsymbol{\theta}$ needs to be estimated from a vector of observations $\mathbf{y}$. If the conditional probability of the observation vector given the parameters is $p_{\mathbf{y}\,|\,\boldsymbol{\theta}}\left(\mathbf{y}\,|\,\boldsymbol{\theta}\right)$, then the covariance matrix of an unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ verifies that:

$$\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}} \triangleq \mathrm{E}\left\{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{H}\right\} \succeq \mathbf{I}^{-1}\left(\boldsymbol{\theta}\right), \tag{3.1}$$

where $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A}-\mathbf{B}$ is positive semidefinite, and $\mathbf{I}\left(\boldsymbol{\theta}\right)$ is the Fisher information matrix, whose (k,p)-th entry is given by:

$$i_{kp}\left(\boldsymbol{\theta}\right) = \mathrm{E}_{\mathbf{y}}\left\{-\frac{\partial^2}{\partial\theta_k \partial\theta_p^*}\ln p_{\mathbf{y}\,|\,\boldsymbol{\theta}}\left(\mathbf{y}\,|\,\boldsymbol{\theta}\right)\right\}. \tag{3.2}$$

3.1 Cramér-Rao Lower Bound for Independent Estimation

All the estimation and equalization problem formulations derived in Chapter 2 (namely (2.12), (2.16), (2.19), (2.21), (2.26), and (2.32)) have the following form:

$$\overline{\mathbf{y}} = \mathcal{Q}\left(\mathbf{A}\overline{\mathbf{h}} + \overline{\mathbf{b}} + \overline{\mathbf{w}}\right), \tag{3.3}$$

where $\mathbf{A}$ is a matrix, $\overline{\mathbf{b}}$ is a *known* vector (which in most of the cases is 0), $\overline{\mathbf{h}}$ is the parameter vector to estimate, $\overline{\mathbf{y}}$ is the observation vector, and $\overline{\mathbf{w}}$ contains Gaussian noise. Let us make the problem real-valued by applying the following transformations:

$$\mathbf{h} = \begin{pmatrix} \Re\left\{\overline{\mathbf{h}}\right\} \\ \Im\left\{\overline{\mathbf{h}}\right\} \end{pmatrix}; \qquad \tilde{\mathbf{A}} = \begin{pmatrix} \Re\left\{\mathbf{A}\right\} & -\Im\left\{\mathbf{A}\right\} \\ \Im\left\{\mathbf{A}\right\} & \Re\left\{\mathbf{A}\right\} \end{pmatrix}; \tag{3.4}$$

and defining $\mathbf{b}$, $\mathbf{y}$ and $\mathbf{w}$ in a similar way as $\mathbf{h}$. The problem then becomes:

$$\mathbf{y} = \mathcal{Q}\left(\tilde{\mathbf{A}}\mathbf{h} + \mathbf{b} + \mathbf{w}\right). \tag{3.5}$$

The Cramér-Rao Lower Bound states that the covariance matrix of any unbiased estimator $\hat{\mathbf{h}}$ of $\mathbf{h}$ satisfies:

$$\mathrm{E}\left\{\left(\hat{\mathbf{h}} - \mathbf{h}\right)\left(\hat{\mathbf{h}} - \mathbf{h}\right)^{H}\right\} \succeq \left(\mathbf{I}(\mathbf{h})\right)^{-1}, \tag{3.6}$$

where $\mathbf{I}(\mathbf{h})$ is the Fisher information matrix:

$$\mathbf{I}\left(\mathbf{h}\right) = \mathrm{E}\left\{-\frac{\partial^2}{\partial \mathbf{h}^H \partial \mathbf{h}} \ln p_{\mathbf{y}\,|\,\mathbf{h}}\left(\mathbf{y}\,|\,\mathbf{h}\right)\right\}, \tag{3.7}$$

where the expectation is taken over $\mathbf{y}\,|\,\mathbf{h}$.

Let us denote the dimensions of $\tilde{\mathbf{A}}$ by K and P, so that $\tilde{\mathbf{A}} \in \mathbb{C}^{K\times P}$ (for the channel estimation problem, we have $K = 2NT$ and $P = 2LN_t$, while for data equalization in the cyclic prefix case, the values are $K = 2NN_r$ and $P = 2NN_t$).

Consider the case in which $\mathcal{Q}\left(\cdot\right)$ applies 1-bit quantization (2.1) and $\mathbf{w}$ has uncorrelated Gaussian samples with variances $\sigma_k^2, k \in \{1, \ldots, K\}$. Then, each quantization process is independent of the others, and we can write:

$$p_{\mathbf{y}\,|\,\mathbf{h}}\left(\mathbf{y}\,|\,\mathbf{h}\right) = \prod_{k=1}^{K} p_{y_k\,|\,\mathbf{h}}\left(y_k\,|\,\mathbf{h}\right). \tag{3.8}$$

From (3.5), we have:

$$y_k = \mathcal{Q}\left(\sum_{p=1}^{P} \tilde{a}_{kp} h_p + b_k + w_p\right). \tag{3.9}$$

As $\mathcal{Q}$ is the sign operator, we have:

$$p_{y_k\,|\,\mathbf{h}}\left(1\,|\,\mathbf{h}\right) = \Pr\left[w_k \geq -\sum_{p=1}^{P} \tilde{a}_{kp} h_p - b_k\right], \tag{3.10}$$

$$p_{y_k\,|\,\mathbf{h}}\left(-1\,|\,\mathbf{h}\right) = \Pr\left[w_k < -\sum_{p=1}^{P} \tilde{a}_{kp} h_p - b_k\right]. \tag{3.11}$$

Due to the symmetry of the Gaussian distribution, we have $\Pr\left[w_k < c\right] = \Phi\left(c/\sigma_k\right)$, and $\Pr\left[w_k \geq c\right] = 1 - \Phi\left(c/\sigma_k\right) = \Phi\left(-c/\sigma_k\right)$, where $\Phi\left(x\right) \triangleq \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-v^2/2}\,\mathrm{d}v$ denotes the standard cumulative Gaussian distribution function. This, together with (3.8), (3.10) and (3.11), allows us to express the conditional probability of the observation given the parameter as:

$$p_{\mathbf{y}\,|\,\mathbf{h}}\left(\mathbf{y}\,|\,\mathbf{h}\right) = \prod_{k=1}^{K} \Phi\left(\frac{y_k\left(\sum_{p=1}^{P} \tilde{a}_{kp} h_p + b_k\right)}{\sigma_k}\right) = \prod_{k=1}^{K} \Phi\left(\eta_k\right), \tag{3.12}$$

where we define η_k as the content of the parenthesis to simplify notation.

Now we can compute the Fisher information matrix according to (3.7). The (p_1, p_2)-th element is given by:

$$\begin{aligned} i_{p_1p_2}(\mathbf{h}) &= \mathrm{E}\left\{-\frac{\partial^2}{\partial h_{p_2}\partial h_{p_1}}\sum_{k=1}^{K}\ln\Phi\left(\eta_k\right)\right\} \\ &= \mathrm{E}\left\{-\frac{\partial}{\partial h_{p_2}}\sum_{k=1}^{K}\frac{y_k}{\sigma_k}\tilde{a}_{kp_1}\frac{\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}\right\} \\ &= \mathrm{E}\left\{\sum_{k=1}^{K}\frac{1}{\sigma_k^2}\tilde{a}_{kp_1}\tilde{a}_{kp_2}\left(\frac{\eta_k\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}+\left(\frac{\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}\right)^2\right)\right\} \end{aligned} \tag{3.13}$$

$$\mathbf{I}\left(\mathbf{h}\right) = \mathrm{E}\left\{\tilde{\mathbf{A}}^T\mathrm{diag}\left\{\frac{1}{\sigma_k^2}\left(\frac{\eta_k\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}+\left(\frac{\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}\right)^2\right)\right\}_{k=1}^{K}\tilde{\mathbf{A}}\right\}, \tag{3.14}$$

where $\phi\left(x\right) \triangleq \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is the standard Gaussian density function. Now, we take the expectation over $\mathbf{y} \mid \mathbf{h}$:

$$\begin{aligned} &\mathbf{I}(\mathbf{h}) \\ &= \tilde{\mathbf{A}}^T\mathrm{diag}\left\{\frac{1}{\sigma_k^2}\sum_{\mathbf{y}\in\{-1,1\}^K}\left[\left(\frac{\eta_k\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}+\left(\frac{\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}\right)^2\right)\prod_{k'=1}^{K}\Phi\left(\eta_{k'}\right)\right]\right\}_{k=1}^{K}\tilde{\mathbf{A}}. \end{aligned} \tag{3.15}$$

Note that $\left(\frac{\eta_k\phi(\eta_k)}{\Phi(\eta_k)}+\left(\frac{\phi(\eta_k)}{\Phi(\eta_k)}\right)^2\right)$ depends on y_k (through η_k), but not on any other component of $\mathbf{y}$. Therefore, we reorganize the summation $\sum_{\mathbf{y}\in\{-1,1\}^K} \equiv \sum_{y_1\in\{-1,1\}}\sum_{y_2\in\{-1,1\}}\cdots\sum_{y_K\in\{-1,1\}}$ as follows:

$$\begin{aligned} &\sum_{\mathbf{y}\in\{-1,1\}^K}\left[\left(\frac{\eta_k\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}+\left(\frac{\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}\right)^2\right)\prod_{k'=1}^{K}\Phi\left(\eta_{k'}\right)\right] \\ =&\sum_{y_k\in\{-1,1\}}\left(\frac{\eta_k\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}+\left(\frac{\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}\right)^2\right)\Phi\left(\eta_k\right)\sum_{y_1\in\{-1,1\}}\overset{\cancel{y_k}}{\cdots}\sum_{y_K\in\{-1,1\}}\prod_{k'\neq k}\Phi\left(\eta_{k'}\right) \\ =&\sum_{y_k\in\{-1,1\}}\left(\frac{\eta_k\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}+\left(\frac{\phi\left(\eta_k\right)}{\Phi\left(\eta_k\right)}\right)^2\right)\Phi\left(\eta_k\right)\prod_{k'\neq k}\sum_{y_{k'}\in\{-1,1\}}\Phi\left(\eta_{k'}\right) \\ =&\frac{\left(\phi\left(\mu_k\right)\right)^2}{\Phi\left(\mu_k\right)\left(1-\Phi\left(\mu_k\right)\right)}. \end{aligned} \tag{3.16}$$

where

$$\mu_k = \frac{1}{\sigma_k}\sum_{p=1}^{P}\tilde{a}_{kp}h_p + b_k = \left.\eta_k\right|_{y_k=1}, \tag{3.17}$$

and $\overset{\cancel{y_k}}{\cdots}$ means that the stacked summations in the right half of the second line of (3.16) go over all components of $\mathbf{y}$ except y_k. In the last step, we have used the fact that

$\Phi(\eta_k) + \Phi(-\eta_k) = 1$ (which cancels out the whole product of sums in the right side of the third line), and that $\phi(\eta_k) = \phi(-\eta_k)$ (which removes the first term in the parenthesis).

Plugging (3.16) into (3.15), we obtain the final result for the real-valued Fisher Information Matrix for the 1-bit quantized estimation problem:

$$\begin{aligned}\tilde{\mathbf{I}}(\boldsymbol{\theta}) &= \tilde{\mathbf{A}}^T \mathrm{diag}\left\{\frac{1}{\sigma_k^2}\frac{\phi(\mu_k)^2}{\Phi(\mu_k)(1-\Phi(\mu_k))}\right\}_{k=1}^{K}\tilde{\mathbf{A}} \\ &= \begin{pmatrix} \left[\tilde{\mathbf{I}}(\boldsymbol{\theta})\right]_{\Re\Re} & \left[\tilde{\mathbf{I}}(\boldsymbol{\theta})\right]_{\Re\Im} \\ \left[\tilde{\mathbf{I}}(\boldsymbol{\theta})\right]_{\Im\Re} & \left[\tilde{\mathbf{I}}(\boldsymbol{\theta})\right]_{\Im\Im} \end{pmatrix},\end{aligned} \tag{3.18}$$

where μ_k is given by (3.17). Using $\theta_\Re = \Re\{\theta\} = (\theta + \theta^*)/2$ and $\theta_\Im = \Im\{\theta\} = (\theta - \theta^*)/(2j)$, we can derive the following identities:

$$\frac{\partial f(\theta)}{\partial \theta} = \frac{1}{2}\frac{\partial f(\theta)}{\partial \theta_\Re} - \frac{j}{2}\frac{\partial f(\theta)}{\partial \theta_\Im}, \tag{3.19}$$

$$\frac{\partial f(\theta)}{\partial \theta^*} = \frac{1}{2}\frac{\partial f(\theta)}{\partial \theta_\Re} + \frac{j}{2}\frac{\partial f(\theta)}{\partial \theta_\Im}, \tag{3.20}$$

$$f_{\theta_1\theta_2^*} = \frac{1}{4}\left(f_{\theta_{1,\Re}\theta_{2,\Re}} + f_{\theta_{1,\Im}\theta_{2,\Im}}\right) + \frac{j}{4}\left(f_{\theta_{1,\Re}\theta_{2,\Im}} - f_{\theta_{1,\Im}\theta_{2,\Re}}\right). \tag{3.21}$$

In (3.21), the subscripts denote derivatives with respect to the corresponding variable. Using this identity, we can transform the real-valued Fisher Information Matrix to the complex-valued one:

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{4}\left(\left[\tilde{\mathbf{I}}(\boldsymbol{\theta})\right]_{\Re\Re} + \left[\tilde{\mathbf{I}}(\boldsymbol{\theta})\right]_{\Im\Im}\right) + \frac{j}{4}\left(\left[\tilde{\mathbf{I}}(\boldsymbol{\theta})\right]_{\Re\Im} - \left[\tilde{\mathbf{I}}(\boldsymbol{\theta})\right]_{\Im\Re}\right). \tag{3.22}$$

The Cramér-Rao Lower Bound is then simply the inverse of the Fisher Information Matrix:

$$\mathbf{C}_{\hat{\mathbf{h}}\hat{\mathbf{h}}} \succeq \left(\mathbf{I}(\boldsymbol{\theta})\right)^{-1}. \tag{3.23}$$

As a final note, the ratio $\frac{\phi(\mu_k)^2}{\Phi(\mu_k)(1-\Phi(\mu_k))}$ is numerically unstable if $|\mu_k|$ is large. This can be overcome by using the following approximation, which follows from the fact that $\phi(x)/\left(-x\Phi(x)\right) \to 1$ when $x \to -\infty$:

$$\frac{\left(\phi(\mu_k)\right)^2}{\Phi(\mu_k)\left(1-\Phi(\mu_k)\right)} \underset{|\mu_k|\to\infty}{\sim} |\mu_k|\,\phi(\mu_k). \tag{3.24}$$

3.2 Cramér-Rao Lower Bound for Joint Channel and Data (JCD) Estimation

In this thesis, we also consider the joint estimation of channel and data, which can considerably improve performance with respect to independent estimation. In this section, we derive the Cramér-Rao bound for the joint problem, which will turn out to be lower than that of the separate problem.

3.2.1 JCD CRLB for Single-Carrier with Cyclic Prefix (SC)

Let us first consider a system with single carrier and cyclic prefix, in which M_p known blocks of size N are transmitted as pilots, and M_d blocks are transmitted as data. We will denote by $x_{n_t,m,n}$ the symbol transmitted by antenna $n_t \in \{1,\dots,N_t\}$, at block $m \in \{0,\dots,M_p+M_d-1\}$, at position $n \in \{0,\dots,N-1\}$ within the block. We assume that $m \in \{0,\dots,M_p-1\}$ corresponds to the pilot blocks, and $m \in \{M_p,\dots,M_p+M_d-1\}$ corresponds to the data blocks.

Likewise, we denote by $h_{n_r,n_t,\ell}, \ell \in \{0,\dots,L-1\}$ the ℓ-th channel tap of the channel vector $\mathbf{h}_{n_r n_t}$ from transmit antenna n_t to receive antenna n_r. We write $y_{n_r,m,n}$ for the received symbol at antenna n_r, block m and position n, and $w_{n_r,n,m}$ for the noise at the same location. With these definitions, we can write the following equation for the system:

$$y_{n_r,m,n} = \mathcal{Q}\left(\sum_{n_t=1}^{N_t}\sum_{\ell=0}^{L-1} h_{n_r,n_t\ell} x_{n_r,m,(n-l)\bmod N} + w_{n_r,m,n}\right). \tag{3.25}$$

Now, we can apply the same reasoning as we did to derive (3.12) from (3.5). Due to independence of the noise samples, the conditional probability of the observations $\boldsymbol{\mathcal{Y}}$ given the parameters $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{X}}_d$ can be expressed as the product of the marginals:

$$p_{\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{X}}_d}\left(\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{X}}_d\right) = \prod_{n_r=1}^{N_r}\prod_{m=0}^{M-1}\prod_{n=0}^{N-1} p_{y_{n_r,m,n}\,|\,\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{X}}_d}\left(y_{n_r,m,n}\,|\,\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{X}}_d\right), \tag{3.26}$$

where $M = M_p + M_d$. Here, the calligraphic letters $\boldsymbol{\mathcal{Y}}$ given the parameters $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{X}}_d$ denote respectively the lists of all received samples, all channel taps and all transmitted data symbols. By applying the same idea as in (3.10)-(3.12), we arrive at the following expression:

$$p_{\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{X}}_d}\left(\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{X}}_d\right) = \prod_{n_r=1}^{N_r}\prod_{m=0}^{M-1}\prod_{n=0}^{N-1} \Phi\left(\frac{y_{\Re,n_r,m,n} z_{\Re,n_r,m,n}}{\sigma_{n_r}/\sqrt{2}}\right)\Phi\left(\frac{y_{\Im,n_r,m,n} z_{\Im,n_r,m,n}}{\sigma_{n_r}/\sqrt{2}}\right), \tag{3.27}$$

where

$$z_{\Re,n_r,m,n} = \sum_{n_t=1}^{N_t}\sum_{\ell=0}^{L-1}\left(h_{\Re,n_r,n_t,\ell} x_{\Re,n_t,m,(n-\ell)\bmod N} - h_{\Im,n_r,n_t,\ell} x_{\Im,n_t,m,(n-\ell)\bmod N}\right) \tag{3.28}$$

and

$$z_{\Im,n_r,m,n} = \sum_{n_t=1}^{N_t}\sum_{\ell=0}^{L-1}\left(h_{\Re,n_r,n_t,\ell} x_{\Im,n_t,m,(n-\ell)\bmod N} + h_{\Im,n_r,n_t,\ell} x_{\Re,n_t,m,(n-\ell)\bmod N}\right). \tag{3.29}$$

The subscripts $\Re$ and $\Im$ denote, respectively, the real and imaginary part (for example, $y_{\Re,n_r,m,n} \triangleq \Re\{y_{n_r,m,n}\}$).

Let us define the parameter vector $\boldsymbol{\theta}$ as:

$$\boldsymbol{\theta} = \begin{pmatrix} \Re\{\mathbf{h}\} \\ \Re\{\mathbf{x}_d\} \\ \Im\{\mathbf{h}\} \\ \Im\{\mathbf{x}_d\} \end{pmatrix} \in \mathbb{R}^{2N_rN_tL+2N_tM_dN}, \tag{3.30}$$

where $\mathbf{h} \in \mathbb{C}^{LN_tN_r}$ contains all channel taps $h_{n_rn_t\ell}$ in ascending order of the indices $\ell \in \{0,\ldots,L-1\}$, $n_t \in \{1,\ldots,N_t\}$, $n_r \in \{1,\ldots,N_r\}$, where the ℓ varies fastest and n_r varies slowest. Likewise, $\mathbf{x}_d \in \mathbb{C}^{NN_tM}$ contains all the data symbols $x_{n_t,m,n}$ for index $n \in \{0,\ldots,N-1\}$ varying fastest, followed by $n_t \in \{1,\ldots,N_t\}$, and with $m \in \{M_p,\ldots,M_p+M_d-1\}$ changing slowest.

The Fisher Information Matrix is then defined as:

$$\mathbf{I}(\boldsymbol{\theta}) = \mathrm{E}\left\{-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\ln p_{\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\theta}}(\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\theta})\right\}, \tag{3.31}$$

where $p_{\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\theta}}(\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\theta})$ is obtained from (3.27).

A detailed derivation of the solution to (3.31) can be found in Appendix B. The resulting values for all the entries of $\mathbf{I}(\boldsymbol{\theta})$ are given in closed form in the following, where $P(\boldsymbol{\mathcal{Y}}\,|\,\theta) = \ln p_{\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\theta}}(\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\theta})$. In these expressions, the subscripts $\mathcal{P}$ and $\mathcal{T}$ can take the value $\Re$ or $\Im$, making the formulas valid for both the real and imaginary parts of the taps. The expression $\overline{\mathcal{P}}$ denotes the opposite of $\mathcal{P}$ (i.e., $\overline{\mathcal{P}} = \Im$ if $\mathcal{P} = \Re$, and $\overline{\mathcal{P}} = \Re$ if $\mathcal{P} = \Im$). The same applies for $\mathcal{T}$ and $\overline{\mathcal{T}}$. Furthermore, we define the following variables:

$$s_{\mathcal{P}} = \begin{cases} -1, & \text{if } \mathcal{P} = \Im \\ +1, & \text{if } \mathcal{P} = \Re, \end{cases} \tag{3.32}$$

$$\mu_{\mathcal{P},n_r,m,n} = \frac{\sqrt{2}}{\sigma_{n_r}} z_{\mathcal{P},n_r,m,n}, \tag{3.33}$$

(where $z_{\Re,n_r,m,n}$ and $z_{\Im,n_r,m,n}$ are defined in (3.28) and (3.29)),

$$\Psi_{\mathcal{P},n_r,m,n} = \frac{\left(\phi\left(\mu_{\mathcal{P},n_r,n,m}\right)\right)^2}{\Phi\left(\mu_{\mathcal{P},n_r,n,m}\right)\left(1-\Phi\left(\mu_{\mathcal{P},n_r,n,m}\right)\right)}. \tag{3.34}$$

With all these definitions, the entries of the Fisher information matrix are given by:

$$\begin{aligned} &\mathrm{E}\left\{-\frac{\partial^2 P(\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\theta})}{\partial h_{\mathcal{P}n_r,n_t,\ell}\partial h_{\mathcal{T}n_r',n_t',\ell'}}\right\} \\ &\qquad = \delta[n_r - n_r']\frac{2}{\sigma_{n_r}}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \\ &\qquad \Big[s_{\mathcal{P}}s_{\mathcal{T}}x_{\mathcal{P},n_t,m,(n-\ell)\bmod N}x_{\mathcal{T},n_t',m,(n-\ell')\bmod N}\Psi_{\Re,n_r,m,n} \\ &\qquad\quad + x_{\overline{\mathcal{P}},n_t,m,(n-\ell)\bmod N}x_{\overline{\mathcal{T}},n_t',m,(n-\ell')\bmod N}\Psi_{\Im,n_r,m,n}\Big], \end{aligned} \tag{3.35}$$

$$\mathrm{E}\left\{-\frac{\partial^2 P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\theta}\right)}{\partial x_{\mathcal{P}n_t,m,n}\partial x_{\mathcal{T}n'_t,m',n'}}\right\} = \delta\left[m-m'\right]\sum_{n_r=1}^{N_r}\frac{2}{\sigma_{n_r}}\sum_{n''=\max\{n,n'\}}^{(\min\{n+L-1,n'+L-1\})\mathrm{mod}N}\Big[s_{\mathcal{P}}s_{\mathcal{T}}h_{\mathcal{P},n_r,n_t,(n''-n)\mathrm{mod}N}h_{\mathcal{T},n_r,n'_t,(n''-n')\mathrm{mod}N}\Psi_{\Re,n_r,m,n''} + h_{\overline{\mathcal{P}},n_r,n_t,(n''-n)\mathrm{mod}N}h_{\overline{\mathcal{T}},n_r,n'_t,(n''-n')\mathrm{mod}N}\Psi_{\Im,n_r,m,n''}\Big], \quad (3.36)$$

$$\mathrm{E}\left\{-\frac{\partial^2 P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\theta}\right)}{\partial h_{\mathcal{P}n_r,n_t,\ell}\partial x_{\mathcal{T}n'_t,m,n}}\right\} = \frac{2}{\sigma_{n_r}}\sum_{n''=n}^{(n+L-1)\mathrm{mod}N}\Big[s_{\mathcal{P}}s_{\mathcal{T}}x_{\mathcal{P},n_t,m,(n''-\ell)\mathrm{mod}N}h_{\mathcal{T},n_r,n'_t,(n''-n)\mathrm{mod}N}\Psi_{\Re,n_r,m,n''} + x_{\overline{\mathcal{P}},n_t,m,(n''-\ell)\mathrm{mod}N}h_{\overline{\mathcal{T}},n_r,n'_t,(n''-n)\mathrm{mod}N}\Psi_{\Im,n_r,m,n''}\Big], \quad (3.37)$$

where $\delta[n]$ is the Kronecker delta, which is 1 if its argument is 0, and 0 otherwise.

Equations (3.35) to (3.37) give the complete Fisher information matrix $\mathbf{I}\left(\boldsymbol{\theta}\right)$. The CRLB is then obtained, as before, by applying (3.22) to write it in the complex-valued form and then inverting the result.

3.2.2 JCD CRLB for OFDM

The CRLB for joint channel and data estimation in the OFDM case can be easily calculated by computing the IFFT of the pilot and data blocks, and applying the formulas for the Single-Carrier with Cyclic Prefix case. Therefore, we do not include an additional derivation for this problem.

3.2.3 JCD CRLB for Single Carrier Without Cyclic Prefix (NCP)

The Fisher Information Matrix for the NCP case has a slightly different form than for the other problems, as here there are no transmission blocks. Consider a scenario in which M_{np} pilot symbols and M_{nd} data symbols are sent from each transmit antenna. We denote by $x_{p,n_t,m_{np}}$ the pilot symbol at antenna $n_t \in \{i,\ldots,N_t\}$ and time instant $m_{np} \in \{0,\ldots,M_{np}-1\}$, while the data symbol at antenna n_t and time instant $m_{nd} \in \{0,\ldots,M_{nd}-1\}$. We make the following two important assumptions in our model:

- There is a long enough guard interval between the pilots and the data, so that they do not interfere with each other.
- Only the first M_{np} samples of the channel output are considered for channel estimation, while the trailing $L-1$ samples are discarded (otherwise, the channel estimation matrix becomes ill-conditioned).

With these assumptions, the entries for the Fisher Information Matrix in the NCP case can be obtained by setting $N = 1$ in the SC case (3.35)-(3.37), and applying some changes in the indices of the summations to account for the lack of a cyclic prefix. We now define:

$$\mu_{p,\Re,n_r,m} = \frac{\sqrt{2}}{\sigma_{n_r}} \sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \left(h_{\Re,n_r,n_t,\ell} x_{p,\Re,n_t,m-\ell} - h_{\Im,n_r,n_t,\ell} x_{p,\Im,n_t,m-\ell} \right) \quad (3.38)$$

and

$$\mu_{p,\Im,n_r,m} = \frac{\sqrt{2}}{\sigma_{n_r}} \sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \left(h_{\Re,n_r,n_t,\ell} x_{p,\Im,n_t,m-\ell} + h_{\Im,n_r,n_t,\ell} x_{p,\Re,n_t,m-\ell} \right), \quad (3.39)$$

for $m \in \{0, \ldots, M_{np} - 1\}$. Similarly, we define

$$\mu_{d,\Re,n_r,m} = \frac{\sqrt{2}}{\sigma_{n_r}} \sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \left(h_{\Re,n_r,n_t,\ell} x_{d,\Re,n_t,m-\ell} - h_{\Im,n_r,n_t,\ell} x_{d,\Im,n_t,m-\ell} \right) \quad (3.40)$$

and

$$\mu_{d,\Im,n_r,m} = \frac{\sqrt{2}}{\sigma_{n_r}} \sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \left(h_{\Re,n_r,n_t,\ell} x_{d,\Im,n_t,m-\ell} + h_{\Im,n_r,n_t,\ell} x_{d,\Re,n_t,m-\ell} \right), \quad (3.41)$$

for $m \in \{0, \ldots, M_{nd} + L - 1\}$. We further define:

$$\Psi_{p,\mathcal{T},n_r,m} = \frac{\left(\phi \left(\mu_{p,\mathcal{T},n_r,m} \right) \right)^2}{\Phi \left(\mu_{p,\mathcal{T},n_r,m} \right) \left(1 - \Phi \left(\mu_{p,\mathcal{T},n_r,m} \right) \right)}, \quad (3.42)$$

and

$$\Psi_{d,\mathcal{T},n_r,m} = \frac{\left(\phi \left(\mu_{d,\mathcal{T},n_r,m} \right) \right)^2}{\Phi \left(\mu_{d,\mathcal{T},n_r,m} \right) \left(1 - \Phi \left(\mu_{d,\mathcal{T},n_r,m} \right) \right)}. \quad (3.43)$$

With these definitions, the entries of the Fisher Information Matrix for the NCP case are given by (note that, in the derivatives with respect to two channel taps, two summations are needed to take into account the dependence on the pilots and the data):

$$\mathrm{E}\left\{-\frac{\partial^2 P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\theta}\right)}{\partial h_{\mathcal{P}n_r,n_t,\ell}\partial h_{\mathcal{T}n'_r,n'_t,\ell'}}\right\}$$
$$= \delta\left[n_r - n'_r\right]\frac{2}{\sigma_{n_r}}\left\{\sum_{m=0}^{M_p-1}\right.$$
$$\left[s_{\mathcal{P}}s_{\mathcal{T}}x_{p,\mathcal{P},n_t,(m-\ell)}x_{p,\mathcal{T},n'_t,(m-\ell')}\Psi_{p,\Re,n_r,m}\right.$$
$$\left.+x_{p,\overline{\mathcal{P}},n_t,(m-\ell)}x_{p,\overline{\mathcal{T}},n'_t,(m-\ell')}\Psi_{p,\Im,n_r,m}\right]$$
$$+\sum_{m=0}^{M_d+L-1}$$
$$\left[s_{\mathcal{P}}s_{\mathcal{T}}x_{d,\mathcal{P},n_t,(m-\ell)}x_{d,\mathcal{T},n'_t,(m-\ell')}\Psi_{d,\Re,n_r,m}\right.$$
$$\left.\left.+x_{d,\overline{\mathcal{P}},n_t,(m-\ell)}x_{d,\overline{\mathcal{T}},n'_t,(m-\ell')}\Psi_{d,\Im,n_r,m}\right]\right\}, \quad (3.44)$$

$$\mathrm{E}\left\{-\frac{\partial^2 P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\theta}\right)}{\partial x_{d,\mathcal{P}n_t,m}\partial x_{d,\mathcal{T}n'_t,m'}}\right\}$$
$$= \sum_{n_r=1}^{N_r}\frac{2}{\sigma_{n_r}}\sum_{m''=\max\{m,m'\}}^{(\min\{m+L-1,m'+L-1\})}$$
$$\left[s_{\mathcal{P}}s_{\mathcal{T}}h_{\mathcal{P},n_r,n_t,(m''-m)}h_{\mathcal{T},n_r,n'_t,(m''-m')}\Psi_{d,\Re,n_r,m''}\right.$$
$$\left.+h_{\overline{\mathcal{P}},n_r,n_t,(m''-m)}h_{\overline{\mathcal{T}},n_r,n'_t,(m''-m')}\Psi_{d,\Im,n_r,m''}\right], \quad (3.45)$$

$$\mathrm{E}\left\{-\frac{\partial^2 P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\theta}\right)}{\partial h_{\mathcal{P}n_r,n_t,\ell}\partial x_{d,\mathcal{T}n'_t,m}}\right\}$$
$$= \frac{2}{\sigma_{n_r}}\sum_{m''=m}^{(m+L-1)}$$
$$\left[s_{\mathcal{P}}s_{\mathcal{T}}x_{d,\mathcal{P},n_t,(m''-\ell)}h_{\mathcal{T},n_r,n'_t,(m''-m)}\Psi_{d,\Re,n_r,m''}\right.$$
$$\left.+x_{d,\overline{\mathcal{P}},n_t,(m''-\ell)}h_{\overline{\mathcal{T}},n_r,n'_t,(m''-m)}\Psi_{d,\Im,n_r,m''}\right]. \quad (3.46)$$

The Cramér-Rao Lower Bound for the NCP case can then be calculated by inverting the Fisher Information Matrix $\mathbf{I}\left(\boldsymbol{\theta}\right) \in \mathbb{R}^{2N_rN_tL+2N_tM_d \times 2N_rN_tL+2N_tM_d}$, whose entries are given by (3.44)-(3.46), arranged in any convenient order (as long as each row contains all the entries where the first derivative is with respect to the same parameter, each column contains all the terms with second derivative w.r.t. the same parameter, and the ordering of the parameters along rows and columns is the same).

Chapter 4

Algorithms for Parameter Estimation with Quantized Observations

In this chapter, we will present algorithms that solve the quantized estimation problems (2.12), (2.16), (2.19), (2.21), (2.26), and (2.32), which can all be expressed with the following model:

$$\mathbf{y} = \mathcal{Q}\left(\mathbf{A}\mathbf{h} + \mathbf{b} + \mathbf{w}\right), \tag{4.1}$$

where $\mathbf{A} \in \mathbb{C}^{K \times P}$ is a known coupling matrix, $\mathbf{b} \in \mathbb{C}^{K}$ is a known offset vector, $\mathbf{h} \in \mathbb{C}^{P}$ is the unknown parameter vector that is to be estimated, and $\mathbf{w} \in \mathbb{C}^{K}$ is a vector of circularly symmetric independent Gaussian noise, with variances $\sigma_k^2, k \in \{1, \ldots, K\}$.

In this chapter, we focus on independent channel and data estimation, i.e. the channel is estimated first, and then the data is equalized using the obtained channel estimate. In the next chapter, we will deal with joint estimation of both parameters, which will turn out to give better results at an increased computational complexity.

There are two types of estimation algorithms for the model in (4.1): nonlinear methods and linear methods:

- *Nonlinear methods* take into account the inherent nonlinearity of the problem to and estimate $\mathbf{h}$. As it is not possible to drive a closed-form expression for this case, the nonlinear methods are iterative. They achieve better (more accurate) performance than the linear approaches, but suffer from higher computational complexity and are also not guaranteed to converge. In this thesis, we consider two nonlinear algorithms: Expectation Maximization (EM) and Generalized Approximate Message Passing (GAMP).

- *Linear methods* approximate the problem (4.1) with a linear model, and solve it using algorithms designed for this type of models. These methods are faster and more robust, but have lower accuracy due to the model mismatch. The linear method considered in this thesis is the Bussgang linearization.

Additionally, we derived a soft-input soft-output (SISO) a-posteriori probability (APP) equalizer. This equalizer is optimal when it converges, but its computational complexity is prohibitive. We will use it in the simulations in a scenario with reduced complexity to evaluate the performance gap between the practical algorithms and optimal equalization.

4.1 Expectation Maximization (EM) Algorithm

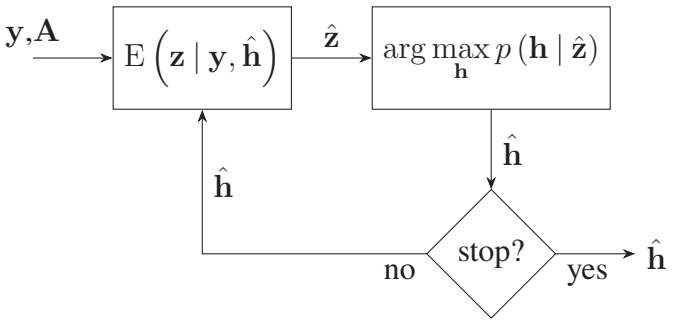

Figure 4.1: Expectation Maximization-MMSE (EM-MMSE) algorithm

The Expectation Maximization (EM) approach [14] is a nonlinear iterative method that alternately applies two steps at each iteration i:

- *Expectation step*: this step obtains the expected value of the unquantized observations $\mathbf{z} = \mathbf{A}\mathbf{h} + \mathbf{w}$, given the quantized output $\mathbf{y}$ and the current estimate $\hat{\mathbf{h}}^{(i-1)}$:

$$\hat{\mathbf{z}}^{(i)} = \mathrm{E}\left(\mathbf{z} \mid \mathbf{y}, \hat{\mathbf{h}}^{(i-1)}\right). \tag{4.2}$$

- *Maximization step*: once the expectation step obtained the expected unquantized observation $\hat{\mathbf{z}}^{(i)}$, the maximization step applies a linear Maximum Likelihood (ML) or Maximum A-Posteriori Probability (MAP) criterion to obtain the corresponding estimated input $\hat{\mathbf{h}}^{(i)}$:

$$\hat{\mathbf{h}}^{(i)} = \arg\max_{\mathbf{h}} p_{\mathbf{h}\,|\,\mathbf{z}}\left(\mathbf{h} \mid \hat{\mathbf{z}}^{(i)}\right) \tag{4.3}$$

After the two steps, the new estimate $\mathbf{h}^{(i)}$ is used as input for the Expectation Step in the next iteration, as depicted in Figure 4.1.

The two steps are treated in more detail in the following.

4.1.1 Expectation step:

The expectation step obtains the expected unquantized observation $\hat{\mathbf{z}}^{(i)}$ given the previous estimate of the parameter $\hat{\mathbf{h}}^{(i-1)}$ and the quantized observation $\mathbf{y}$. In the following, we drop the iteration index i for clarity. Using Bayes rule, we obtain the distribution of

$\mathbf{z} \mid \mathbf{y}, \hat{\mathbf{h}}$ as:

$$p_{\mathbf{z} \mid \mathbf{y}, \hat{\mathbf{h}}}\left(\mathbf{z} \mid \mathbf{y}, \hat{\mathbf{h}}\right) = \frac{p_{\mathbf{z} \mid \hat{\mathbf{h}}}(\mathbf{z} \mid \hat{\mathbf{h}}) p_{\mathbf{y} \mid \mathbf{z}, \hat{\mathbf{h}}}(\mathbf{y} \mid \mathbf{z}, \hat{\mathbf{h}})}{p_{\mathbf{y} \mid \hat{\mathbf{h}}}(\mathbf{y} \mid \hat{\mathbf{h}})} = \frac{p_{\mathbf{w}}\left(\mathbf{z} - \mathbf{A}\hat{\mathbf{h}} - \mathbf{b}\right) \mathbf{1}\left\{\mathbf{z} \in \mathcal{Q}^{-1}(\mathbf{y})\right\}}{\int\limits_{\mathbf{w} \in \mathcal{Q}^{-1}(\mathbf{y}) - \mathbf{A}\hat{\mathbf{h}} - \mathbf{b}} p_{\mathbf{w}}(\mathbf{w}) \mathrm{d}\mathbf{w}}, \tag{4.4}$$

where $\mathbf{1}\{s\}$ is an indicator function with value 1 if s is true, and 0 otherwise. The set $\mathcal{Q}^{-1}(\mathbf{y})$ is defined as:

$$\mathcal{Q}^{-1}(\mathbf{y}) \triangleq \left\{\mathbf{z} \in \mathbb{C}^K : \mathcal{Q}(\mathbf{z}) = \mathbf{y}\right\}, \tag{4.5}$$

and $\mathcal{Q}^{-1}(\mathbf{y}) - \mathbf{A}\hat{\mathbf{h}} - \mathbf{b}$ is the translation of $\mathcal{Q}^{-1}(\mathbf{y})$ by $-\mathbf{A}\hat{\mathbf{h}} - \mathbf{b}$.

Now, we apply the expectation operator to (4.4):

$$\mathrm{E}\left(\mathbf{z} \mid \mathbf{y}, \hat{\mathbf{h}}\right) = \frac{\int\limits_{\mathbf{z} \in \mathcal{Q}^{-1}(\mathbf{y})} \mathbf{z}\, p_{\mathbf{w}}\left(\mathbf{z} - \mathbf{A}\hat{\mathbf{h}} - \mathbf{b}\right) \mathrm{d}\mathbf{z}}{\int\limits_{\mathbf{w} \in \mathcal{Q}^{-1}(\mathbf{y}) - \mathbf{A}\hat{\mathbf{h}} - \mathbf{b}} p_{\mathbf{w}}(\mathbf{w}) \mathrm{d}\mathbf{w}} = \mathbf{A}\hat{\mathbf{h}} + \frac{\int\limits_{\mathbf{w} \in \mathcal{Q}^{-1}(\mathbf{y}) - \mathbf{A}\hat{\mathbf{h}} - \mathbf{b}} \mathbf{w} p_{\mathbf{w}}(\mathbf{w}) \mathrm{d}\mathbf{w}}{\int\limits_{\mathbf{w} \in \mathcal{Q}^{-1}(\mathbf{y}) - \mathbf{A}\hat{\mathbf{h}} - \mathbf{b}} p_{\mathbf{w}}(\mathbf{w}) \mathrm{d}\mathbf{w}} = \mathbf{A}\hat{\mathbf{h}} + \hat{\mathbf{w}}. \tag{4.6}$$

Assuming that the noise $\mathbf{w}$ is Gaussian and uncorrelated with variances $\sigma_k^2, k \in \{1, \ldots, K\}$, we have:

$$p_{\mathbf{w}}(\mathbf{w}) = \prod_{k=1}^{K} \left(\frac{1}{\sigma_k \sqrt{\pi}} e^{-\frac{\Re\{w_k\}^2}{\sigma_k^2}} \frac{1}{\sigma_k \sqrt{\pi}} e^{-\frac{\Im\{w_k\}^2}{\sigma_k^2}} \right), \tag{4.7}$$

where w_k denotes the k-th element of $\mathbf{w}$.

The integral in the numerator of (4.6) is vector-valued. Note that $p_{\mathbf{w}}(\mathbf{w})$ is separable. Therefore, for the k-th component of the numerator, all dimensions will cancel out except for the k-th one. The k-th component where the components of $\hat{\mathbf{w}}$ are given by:

$$\hat{w_k} = \frac{\int\limits_{l_{\Re,k}}^{u_{\Re,k}} \frac{w}{\sigma_k \sqrt{\pi}} e^{-\frac{-w^2}{\sigma_k^2}} \mathrm{d}w}{\int\limits_{l_{\Re,k}}^{u_{\Re,k}} \frac{1}{\sigma_k \sqrt{\pi}} e^{-\frac{-w^2}{\sigma_k^2}} \mathrm{d}w} + j \frac{\int\limits_{l_{\Im,k}}^{u_{\Im,k}} \frac{w}{\sigma_k \sqrt{\pi}} e^{-\frac{-w^2}{\sigma_k^2}} \mathrm{d}w}{\int\limits_{l_{\Im,k}}^{u_{\Im,k}} \frac{1}{\sigma_k \sqrt{\pi}} e^{-\frac{-w^2}{\sigma_k^2}} \mathrm{d}w}, \tag{4.8}$$

where the bounds of the integrals are given by:

$$l_{\Re,k} = \begin{cases} -\infty, & \text{if } \Re\{y_k\} = -1 \\ -\Re\left\{b_k + \sum_{p=1}^{P} a_{kp} h_p\right\}, & \text{if } \Re\{y_k\} = 1, \end{cases} \tag{4.9}$$

$$u_{\Re,k} = \begin{cases} -\Re\left\{b_k + \sum_{p=1}^{P} a_{kp} h_p\right\}, & \text{if } \Re\{y_k\} = -1 \\ \infty, & \text{if } \Re\{y_k\} = 1, \end{cases} \tag{4.10}$$

and the bounds in the imaginary part $l_{\Im,k}$ and $u_{\Im,k}$ have the same expressions with all $\Re\{\cdot\}$ operators replaced by $\Im\{\cdot\}$. The denominators in (4.8) are integrals of even functions, which have the property that $\int_{-c}^{\infty} e(x)\,\mathrm{d}x = \int_{-\infty}^{c} e(x)\,\mathrm{d}x$. On the other hand, the integrands in the numerators are odd functions, for which $\int_{-c}^{\infty} o(x)\,\mathrm{d}x = -\int_{-\infty}^{c} o(x)\,\mathrm{d}x$. This allows us to rewrite (4.8) as:

$$\hat{w_k} = \frac{\Re\{y_k\} \int\limits_{-\infty}^{d_{\Re,k}} \frac{w}{\sigma_k\sqrt{\pi}} e^{-\frac{-w^2}{\sigma_k^2}}\,\mathrm{d}w}{\int\limits_{-\infty}^{d_{\Re,k}} \frac{1}{\sigma_k\sqrt{\pi}} e^{-\frac{-w^2}{\sigma_k^2}}\,\mathrm{d}w} + j\frac{\Im\{y_k\} \int\limits_{-\infty}^{d_{\Im,k}} \frac{w}{\sigma_k\sqrt{\pi}} e^{-\frac{-w^2}{\sigma_k^2}}\,\mathrm{d}w}{\int\limits_{-\infty}^{d_{\Im,k}} \frac{1}{\sigma_k\sqrt{\pi}} e^{-\frac{-w^2}{\sigma_k^2}}\,\mathrm{d}w}, \tag{4.11}$$

where

$$d_{\Re,k} = \Re\{y_k\}\,\Re\left\{b_k + \sum_{p=1}^{P} a_{kp}h_p\right\}; \qquad d_{\Im,k} = \Im\{y_k\}\,\Im\left\{b_k + \sum_{p=1}^{P} a_{kp}h_p\right\}. \tag{4.12}$$

By solving the integrals in (4.11), we obtain the final expression for the Expectation Step of the EM algorithm, which in full can be described as:

$$\hat{\mathbf{z}}^{(i)} = \mathbf{A}\hat{\mathbf{h}}^{(i-1)} + \mathbf{b} + \hat{\mathbf{w}}, \tag{4.13}$$

where the components of $\hat{\mathbf{w}}$ are given by:

$$\hat{w_k} = \frac{\sigma_k}{\sqrt{2}}\left(\frac{\Re\{y_k\}\,\phi(\eta_{\Re,k})}{\Phi(\eta_{\Re,k})} + j\frac{\Im\{y_k\}\,\phi(\eta_{\Im,k})}{\Phi(\eta_{\Im,k})}\right), \tag{4.14}$$

where

$$\eta_{\Re,k} = \frac{\Re\{y_k\}\,\Re\left\{b_k + \sum_{p=1}^{P} a_{kp}h_p\right\}}{\sigma_k/\sqrt{2}}; \quad \eta_{\Im,k} = \frac{\Im\{y_k\}\,\Im\left\{b_k + \sum_{p=1}^{P} a_{kp}h_p\right\}}{\sigma_k/\sqrt{2}}. \tag{4.15}$$

4.1.1.1 Covariance matrix of the Expectation Step

For completeness, in this section we will also derive the covariance matrix $\mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}} \triangleq \mathrm{E}\left\{(\mathbf{z}-\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^H \mid \mathbf{y}, \hat{\mathbf{h}}\right\}$ of the Expectation Step estimator in (4.13). Recall that $\hat{\mathbf{z}} = \mathbf{A}\hat{\mathbf{h}} + \mathbf{b} + \hat{\mathbf{w}}$. Then, we have $\mathbf{z} - \hat{\mathbf{z}} = \left(\mathbf{A}\hat{\mathbf{h}} + \mathbf{b} + \mathbf{w}\right) - \left(\mathbf{A}\hat{\mathbf{h}} + \mathbf{b} + \hat{\mathbf{w}}\right) = \mathbf{w} - \hat{\mathbf{w}}$, and therefore:

$$\mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}} = \mathrm{E}\left\{\mathbf{w}\mathbf{w}^H \mid \mathbf{y}, \hat{\mathbf{h}}\right\} - \hat{\mathbf{w}}\hat{\mathbf{w}}^H. \tag{4.16}$$

Using the same reasoning as for $\hat{\mathbf{w}}$ (4.6), the first term is computed as follows:

$$\mathbf{R}_{\hat{\mathbf{w}}\hat{\mathbf{w}}} \triangleq \mathrm{E}\left\{\mathbf{w}\mathbf{w}^H \mid \mathbf{y}, \hat{\mathbf{h}}\right\} = \frac{\int\limits_{\mathbf{w}\in\mathcal{Q}^{-1}(\mathbf{y})-\mathbf{A}\hat{\mathbf{h}}} \mathbf{w}\mathbf{w}^H p_{\mathbf{w}}(\mathbf{w})\,\mathrm{d}\mathbf{w}}{\int\limits_{\mathbf{w}\in\mathcal{Q}^{-1}(\mathbf{y})-\mathbf{A}\hat{\mathbf{h}}} p_{\mathbf{w}}(\mathbf{w})\mathrm{d}\mathbf{w}}. \tag{4.17}$$

Equations (4.16), (4.17) and (4.6) already give a general expression for $\mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}}$, valid for any function $\mathcal{Q}(\cdot)$. For the 1-bit quantization case, let us compute $\mathbf{R}_{\hat{\mathbf{w}}\hat{\mathbf{w}}}$ element-wise:

$$\begin{aligned}[\mathbf{R}_{\hat{\mathbf{w}}\hat{\mathbf{w}}}]_{k_1k_2} &= \frac{\int\limits_{\mathbf{w}\in\mathcal{Q}^{-1}(\mathbf{y})-\mathbf{A}\hat{\mathbf{h}}-\mathbf{b}} w_{k_1}w_{k_2}^* p_{\mathbf{w}}(\mathbf{w})\,\mathrm{d}\mathbf{w}}{\int\limits_{\mathbf{w}\in\mathcal{Q}^{-1}(\mathbf{y})-\mathbf{A}\hat{\mathbf{h}}-\mathbf{b}} p_{\mathbf{w}}(\mathbf{w})\mathrm{d}\mathbf{w}} \\ &= \frac{\iint\limits_{w_{k_1}}\iint\limits_{w_{k_2}} w_{k_1}w_{k_2}^* p_{w_{k_1}}(w_{k_1})\,p_{w_{k_2}}(w_{k_2})\;\mathrm{d}w_{k_2}\,\mathrm{d}w_{k_1}}{\iint\limits_{w_{k_1}}\iint\limits_{w_{k_2}} p_{w_{k_1}}(w_{k_1})\,p_{w_{k_2}}(w_{k_2})\;\mathrm{d}w_{k_2}\,\mathrm{d}w_{k_1}}\end{aligned} \tag{4.18}$$

where the integrals span the values of w_{k_1} and w_{k_2} that belong to $\mathcal{Q}^{-1}(\mathbf{y})-\mathbf{A}\hat{\mathbf{h}}-\mathbf{b}$. We can distinguish two cases here:

- $k_1 \neq k_2$: in this case, the integrals over w_{k_1} and w_{k_2} are separable and the result simplifies to:

$$\begin{aligned}[\mathbf{R}_{\hat{\mathbf{w}}\hat{\mathbf{w}}}]_{k_1k_2} &= \frac{\iint\limits_{w_{k_1}} w_{k_1}p_{w_{k_1}}(w_{k_1})\;\mathrm{d}w_{k_1}}{\iint\limits_{w_{k_1}} p_{w_{k_1}}(w_{k_1})\;\mathrm{d}w_{k_1}}\left(\frac{\iint\limits_{w_{k_2}} w_{k_2}p_{w_{k_2}}(w_{k_2})\;\mathrm{d}w_{k_2}}{\iint\limits_{w_{k_2}} p_{w_{k_2}}(w_{k_2})\;\mathrm{d}w_{k_2}}\right)^* \\ &= \hat{w}_{k_1}\hat{w}_{k_2}^*,\end{aligned} \tag{4.19}$$

  where $\hat{w}_m$ is given by (4.14). Note that, from (4.16), this result means that the off-diagonal terms of $\mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}}$ will cancel out, which means that the estimator covariance is a diagonal matrix.

- $k_1 = k_2 = k$: in this case, (4.17) reduces to:

$$\begin{aligned}[\mathbf{R}_{\hat{\mathbf{w}}\hat{\mathbf{w}}}]_{kk} &= \frac{\iint\limits_{w_k} |w_k|^2 p_{w_k}(w_k)\;\mathrm{d}w_k}{\iint\limits_{w_k} p_{w_k}(w_k)\;\mathrm{d}w_k} \\ &= \frac{\int\limits_{-\infty}^{d_{\Re,k}} \frac{w_{\Re}^2}{\sigma_k\sqrt{\pi}} e^{-\frac{w_{\Re}}{\sigma_k^2}}\;\mathrm{d}w_{\Re}}{\Phi(\eta_{\Re,k})} + \frac{\int\limits_{-\infty}^{d_{\Im,k}} \frac{w_{\Im}^2}{\sigma_k\sqrt{\pi}} e^{-\frac{w_{\Im}}{\sigma_k^2}}\;\mathrm{d}w_{\Im}}{\Phi(\eta_{\Im,k})}, \\ &= \sigma_k^2\left(1-\frac{1}{2}\frac{\eta_{\Re,k}\phi(\eta_{\Re,k})}{\Phi(\eta_{\Re,k})}-\frac{1}{2}\frac{\eta_{\Im,k}\phi(\eta_{\Im,k})}{\Phi(\eta_{\Im,k})}\right),\end{aligned} \tag{4.20}$$

where $d_{\Re,k}$ and $d_{\Im,k}$ are given by (4.12), and $\eta_{\Re,k}$ and $\eta_{\Im,k}$ are defined in (4.15).

Finally, plugging (4.19) and (4.20) into (4.16), we obtain the final result for the estimator covariance matrix:

$$\mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}} = \mathrm{diag}\left\{\sigma_k^2\left(1-\frac{1}{2}\frac{\eta_{\Re,k}\phi(\eta_{\Re,k})}{\Phi(\eta_{\Re,k})}-\frac{1}{2}\frac{\eta_{\Im,k}\phi(\eta_{\Im,k})}{\Phi(\eta_{\Im,k})}-\frac{1}{2}\left(\frac{\phi(\eta_{\Re,k})}{\Phi(\eta_{\Re,k})}\right)^2-\frac{1}{2}\left(\frac{\phi(\eta_{\Im,k})}{\Phi(\eta_{\Im,k})}\right)^2\right)\right\}_{k=1}^{M}, \tag{4.21}$$

with $\eta_{\Re,k}$ and $\eta_{\Im,k}$ given by (4.15).

The ratio $\phi(\eta)/\Phi(\eta)$ is numerically unstable when η is very negative. To prevent this problem in both (4.14) and (4.21), the following approximation can be used:

$$\frac{\phi(\eta)}{\Phi(\eta)} \underset{\eta\to-\infty}{\sim} -\eta. \tag{4.22}$$

4.1.2 Maximization Step

In the maximization step, the obtained $\hat{\mathbf{z}}^{(i)} \triangleq \mathrm{E}\left\{\mathbf{z} \mid \mathbf{y}, \hat{\mathbf{h}}^{(i-1)}\right\}$ (4.14) is used as observation vector in an unquantized problem:

$$\hat{\mathbf{z}}^{(i)} = \mathbf{A}\mathbf{h} + \mathbf{b} + \mathbf{w}. \tag{4.23}$$

In the literature [7], a maximum likelihood (ML) estimator is used to solve this problem:

$$\hat{\mathbf{h}}^{(i)} = \left(\mathbf{A}^H \mathbf{R}_{\mathbf{ww}}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^H \mathbf{R}_{\mathbf{ww}}^{-1} \left(\hat{\mathbf{z}}^{(i)} - \mathbf{b}\right). \tag{4.24}$$

We propose an alternative to this approach, which consists of using an MMSE estimator:

$$\hat{\mathbf{h}}^{(i)} = \left(\mathbf{A}^H \mathbf{R}_{\mathbf{ww}}^{-1} \mathbf{A} + \mathbf{R}_{\mathbf{hh}}^{-1}\right)^{-1} \mathbf{A}^H \mathbf{R}_{\mathbf{ww}}^{-1} \left(\hat{\mathbf{z}}^{(i)} - \mathbf{b}\right); \tag{4.25}$$

This solution (EM-MMSE) gives better performance, as it takes into account prior information. The program flow of the full EM-MMSE method is graphically depicted in Fig. 4.1, and an implementation in pseudo-code is given in Algorithm 4.1.

Algorithm 4.1 Expectation Maximization (EM)

Input: $\mathbf{A}$, $\mathbf{b}$, $\mathbf{y}$
  Initialize:
  $\hat{\mathbf{z}}^{(0)} = \mathbf{y}$
  $\hat{\mathbf{h}}^{(0)}$ from (4.24) (ML) or (4.25) (MMSE)
  $i = 1$
  while $i \leq i_{\max}$ and $\left\|\hat{\mathbf{h}}^{(i)} - \hat{\mathbf{h}}^{(i-1)}\right\|_F^2 \geq \epsilon \left\|\mathbf{h}^{(i)}\right\|_F^2$ **do**
    $\hat{\mathbf{z}} = \mathrm{E}\left(\mathbf{z} \mid \mathbf{y}, \hat{\mathbf{h}}^{(i-1)}\right)$ from (4.13)
    $\hat{\mathbf{h}}^{(i)}$ from (4.24) (ML) or (4.25) (MMSE)
    $i = i + 1$
  end while
Output: $\hat{\mathbf{h}}^{(i)}$

4.2 Generalized Approximate Message Passing (GAMP)

The Generalized Approximate Message Passing method, developed in [15], can also be applied to the quantized estimation problem. This is a very general method that estimates intermediate signals on a coupled channel given the input and the output. As depicted in Figure 4.2, an input $\mathbf{x} \in \mathbb{C}^P$ with a known prior $p_{\mathbf{x}}$ goes successively through a linear transformation $\mathbf{z} = \mathbf{A}\mathbf{x} \in \mathbb{C}^K$ and through the output channel $p_{\mathbf{y}\,|\,\mathbf{z}}$, giving the observed

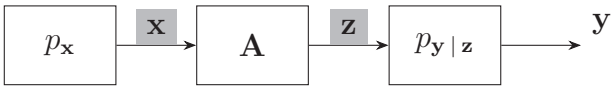

Figure 4.2: Problem formulation of GAMP: the unknown signals to estimate are shaded in gray

output $\mathbf{y}$. The GAMP method estimates the input signal $\mathbf{x}$ and the intermediate signal $\mathbf{z}$ from the knowledge of $\mathbf{A} \in \mathbb{C}^{K\times P}$, $\mathbf{y}$, $p_{\mathbf{x}\,|\,\mathbf{q}}$, and $p_{\mathbf{y}\,|\,\mathbf{z}}$ by using a loopy belief propagation approach. GAMP can only be used if the priors $p_{\mathbf{x}}(\mathbf{x})$ and $p_{\mathbf{y}\,|\,\mathbf{z}}(\mathbf{y}\mid\mathbf{z})$ are separable (i.e. independent across samples of the vectors), which is the case in our problem. The details of the algorithm are developed in [15], and a listing is provided in Algorithm 4.2 (where $\odot$ denotes elementwise product).

Algorithm 4.2 Generalized Approximate Message Passing (GAMP)

Input: $\mathbf{A}$, $\mathbf{b}$, $\mathbf{y}$, $p_{\mathbf{x}}$, and $p_{\mathbf{y}\,|\,\mathbf{z}}$
  Compute: $\mathbf{A}_2 = |\mathbf{A}|^2$ (elementwise)
  Initialize: $i = 0$, $\mathbf{s} = \mathbf{0}_{P\times 1}$, $\mathbf{x}$, $\boldsymbol{\tau}^x$
  while $i < i_{\max}$ **and** $\|\mathbf{x} - \mathbf{x}(i-1)\|_2^2 \geq \epsilon \|\mathbf{x}\|_2^2$ **do**
    $\mathbf{x}_{\text{old}} = \mathbf{x}$
    Output linear step:
    $\boldsymbol{\tau}^p = \mathbf{A}_2\boldsymbol{\tau}^x$
    $\mathbf{p} = \mathbf{A}\mathbf{x} - \frac{1}{2}\boldsymbol{\tau}^p \odot \mathbf{s}$
    $\mathbf{z} = \mathbf{A}\mathbf{x}$
    Output nonlinear step:
    for $\ell = 1 : K$ **do**
      $s_\ell = g_{out}(i, p_\ell, y_\ell, \tau_\ell^p)$
      $\tau_\ell^s = -\frac{\partial}{\partial p} g_{out}(i, p_\ell, y_\ell, \tau_\ell^p)$
    end for
    Input linear step:
    $\boldsymbol{\tau}^r = 4/\left(\mathbf{A}_2^H\boldsymbol{\tau}^s\right)$ (elementwise inverse)
    $\mathbf{r} = \mathbf{x} + \frac{1}{2}\boldsymbol{\tau}^r \odot \left(\mathbf{A}^H\mathbf{s}\right)$
    Input nonlinear step:
    for $\ell = 1 : P$ **do**
      $x_\ell = g_{in}(i, r_\ell, \tau_\ell^r)$
      $\tau_\ell^x = \tau^r \frac{\partial}{\partial r} g_{in}(i, r_\ell, \tau_\ell^r)$
    end for
    Increment loop index: $i = i + 1$
  end while
Output: $\mathbf{x}$, $\mathbf{z}$

The scalar input estimation function g_{in} is defined as:

$$g_{in}(i, r, \tau^p) = \mathrm{E}\{x \mid r\}, \tag{4.26}$$

where the corresponding known prior $p_{\mathbf{x}}(\mathbf{x})$ is used, and r is defined as:

$$r = x + v; \quad v \sim \mathcal{N}\left(0, \tau^r\right). \tag{4.27}$$

The output estimation function g_{out} is defined as:

$$g_{out}\left(i, p, y, \tau^p\right) = \frac{1}{\tau^p}\left(\mathrm{E}\left\{z \mid p, y\right\} - p\right) = \frac{1}{\tau^p}\mathrm{E}\left\{u \mid p, y\right\}, \tag{4.28}$$

where the corresponding known $p_{\mathbf{y} \mid \mathbf{z}}$ is used, and u is defined such that:

$$z = p + u; \qquad u \sim \mathcal{N}\left(0, \tau^p\right). \tag{4.29}$$

4.2.1 Scalar estimation functions for GAMP

The nonlinear steps of GAMP are elementwise independent, and therefore we will derive them for an individual sample x_i (input) or y_j (output). We drop the sample index for clarity.

4.2.1.1 Gaussian Input Step

For channel estimation, the input is assumed to be Gaussian uncorrelated with variance σ_x^2 (possibly different for each sample): $x \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma_x^2\right)$. Using Bayes' Rule, we obtain:

$$\begin{aligned} p_{x \mid r}\left(x \mid r\right) &= \frac{p_x\left(x\right) p_{r \mid x}\left(r \mid x\right)}{p_r\left(r\right)} \qquad (4.30) \\ &= \frac{\frac{1}{\sigma_x^2 \pi} e^{-\frac{|x|^2}{\sigma_x^2}} \frac{1}{\pi \tau^r} e^{-\frac{|r-x|^2}{\tau^r}}}{\frac{1}{\pi\left(\sigma_x^2 + \tau^r\right)} e^{-\frac{|r|^2}{\left(\sigma_x^2 + \tau^r\right)}}} \\ &= \frac{1}{\pi \sigma_{x \mid r}^2} e^{-\frac{\left|x - \mu_{x \mid r}\right|^2}{\sigma_{x \mid r}^2}}, \qquad (4.31) \end{aligned}$$

where

$$\mu_{x \mid r} = \frac{\sigma_x^2 r}{\sigma_x^2 + \tau^r}, \tag{4.32}$$

$$\sigma_{x \mid r}^2 == \frac{\sigma_x^2 \tau_r}{\sigma_x^2 + \tau^r}. \tag{4.33}$$

The PDF in (4.31) is Gaussian with mean $\mu_{x \mid r}$ and variance $\sigma_{x \mid r}^2$. From (4.26), g_{in} is equal to $\mu_{x \mid r}$, which results in the input nonlinear functions:

$$g_{in}\left(i, r, \tau^r\right) = \frac{\sigma_x^2}{\sigma_x^2 + \tau^r} r, \tag{4.34}$$

$$\tau^r \frac{\partial}{\partial r} g_{in}\left(i, r, \tau^r\right) = \frac{\sigma_x^2 \tau^r}{\sigma_x^2 + \tau^r}, \tag{4.35}$$

4.2.1.2 Constellation Input Step

In the equalization problems, the constellation of the input x is known. Let us denote the constellation points by $\overline{x}_a, a \in \{1, \ldots, A\}$, where A is the constellation order. The probability of $\overline{x}_a$ is denoted by P_a. By applying (4.27) and (4.30), we obtain:

$$p_{x \mid q,r}\left(\overline{x}_a \mid q, r\right) = \frac{P_a \frac{1}{\tau^r \pi} e^{-\frac{|r-\overline{x}_a|^2}{\tau^r}}}{\sum_{a=1}^{A} P_a \frac{1}{\tau^r \pi} e^{-\frac{|r-\overline{x}_a|^2}{\tau^r}}}, \tag{4.36}$$

The expectation $\mathrm{E}\left\{x \mid q, r\right\}$ is computed by averaging over x, yielding:

$$g_{in}\left(i, r, q, \tau^r\right) = \frac{\sum_{a=1}^{A} P_a \overline{x}_a e^{-\frac{|r-\overline{x}_a|^2}{\tau^r}}}{\sum_{a=1}^{A} P_a e^{-\frac{|r-\overline{x}_a|^2}{\tau^r}}}, \tag{4.37}$$

$$\begin{aligned} \tau^r \frac{\partial}{\partial r} g_{in}\left(i, r, q, \tau^r\right) & \\ = \frac{\sum_{a=1}^{A} P_a \left|\overline{x}_a\right|^2 e^{-\frac{|r-\overline{x}_a|^2}{\tau^r}}}{\sum_{a=1}^{A} P_a e^{-\frac{|r-\overline{x}_a|^2}{\tau^r}}} & - \left|g_{in}\left(i, r, q, \tau^r\right)\right|^2, \end{aligned} \tag{4.38}$$

where $\overline{x}_a, a \in \{1, \ldots, A\}$ are the constellation symbols, and P_a are their corresponding probabilities.

4.2.1.3 Quantized Output Step

In our 1-bit quantized case, the real and imaginary parts of the problem are independent. Therefore, the expectation can be taken separately for the two components, and we only derive the result for the real part. Recall the definition of p and u from (4.29). We can write:

$$y = \mathcal{Q}\left(z + b + w\right) = \mathcal{Q}\left(p + u + b + w\right), \tag{4.39}$$

where b is known, and $w \sim \mathcal{N}\left(0, \sigma_w^2\right)$ is uncorrelated with p and u. Now, we can use Bayes' rule again to obtain the joint PDF of $u, w \mid p, y$

$$\begin{aligned} p_{u,w \mid p,y}\left(u, w \mid p, y\right) &= \frac{p_{u,w \mid p}\left(u, w \mid p\right) p_{y \mid u,w,p}\left(y \mid u, w, p\right)}{p_{y \mid p}\left(y \mid p\right)} \\ &= \frac{\frac{1}{2\pi\sigma_w\sqrt{\tau^p}} e^{-\frac{u^2}{2\tau^p} - \frac{w^2}{2\sigma_w^2}} \mathbf{1}\left\{\left(p + b + u + w\right) y \geq 0\right\}}{\Phi\left(\frac{yp}{\sqrt{\tau^p + \sigma_w^2}}\right)}. \end{aligned} \tag{4.40}$$

Now, we marginalize over w and average over u:

$$\mathrm{E}\left\{u \mid p, y\right\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u p_{u,w \mid p,y}\left(u, w \mid p, y\right) \, \mathrm{d}u \, \mathrm{d}w. \tag{4.41}$$

By reversing the order of the integrals and appropriately expressing the indicator function, (4.41) can be written as:

$$\begin{aligned}&\mathrm{E}\left\{u \mid p, y\right\}\\ &\qquad = \frac{1}{\Phi\left(\eta\right)} \int_{-\infty}^{\infty} \frac{1}{\sigma_w \sqrt{2\pi}} e^{-\frac{w^2}{2\sigma_w^2}} y \int_{-p-b-w}^{\infty} \frac{u}{\sqrt{2\pi\tau^p}} e^{-\frac{u^2}{2\tau^p}} \,\mathrm{d}u \,\mathrm{d}w, \end{aligned} \tag{4.42}$$

where $\eta = yp/\sqrt{\tau^p + \sigma_w^2}$. The integral along u has the limits corresponding to $y = 1$ (see the indicator function $\mathbf{1}\left\{\ldots\right\}$ in (4.40)). For the case $y = -1$, we have used the property that the integrand $f(u)$ is an odd function, and therefore $\int_{-p-w}^{\infty} f(u)\mathrm{d}u = \int_{-p-w}^{-\infty} f(u)\mathrm{d}u = -\int_{-\infty}^{-p-w} f(u)\,\mathrm{d}u$, which accounts for the pre-multiplying term y. The solution to (4.42) is:

$$\begin{aligned}\mathrm{E}\left\{u \mid p, y\right\} &= \frac{1}{\Phi\left(\eta\right)} \int_{-\infty}^{\infty} \frac{1}{\sigma_w \sqrt{2\pi}} e^{-\frac{w^2}{2\sigma_w^2}} y \frac{\sqrt{\tau^p}}{\sqrt{2\pi}} e^{-\frac{(p+b+w)^2}{2\tau^p}} \,\mathrm{d}w\\ &= \frac{y\tau^p}{\sqrt{\sigma_w^2 + \tau^p}} \frac{\phi\left(\eta\right)}{\Phi\left(\eta\right)}, \end{aligned} \tag{4.43}$$

which, plugged into (4.28), gives the output nonlinear step functions:

$$g_{out}\left(i, p, y, \tau^p\right) = \frac{y}{\sqrt{\sigma_w^2 + \tau^p}} \frac{\phi\left(\eta\right)}{\Phi\left(\eta\right)} \tag{4.44}$$

$$-\frac{\partial}{\partial p} g_{out}\left(i, p, y, \tau^p\right) = \frac{1}{\sigma_w^2 + \tau^p} \left(\eta \frac{\phi\left(\eta\right)}{\Phi\left(\eta\right)} + \left(\frac{\phi\left(\eta\right)}{\Phi\left(\eta\right)}\right)^2\right), \tag{4.45}$$

where

$$\eta = \frac{y\left(p + b\right)}{\sqrt{\sigma_w^2 + \tau^p}}. \tag{4.46}$$

The scalar estimation function g_{out} gives a conditional expectation of z, while its negative derivative $-g'_{out}$ is its conditional variance. Furthermore, the real and imaginary priors $p_{\mathbf{y} \mid \mathbf{z}}\left(\mathbf{y} \mid \mathbf{z}\right)$ are independent. Hence, the expressions to go back to the complex domain are:

$$g_{out}\left(i, p, y, \tau^p\right) = g_{out}\left(i, \Re\left\{p\right\}, \Re\left\{y\right\}, \frac{\tau^p}{2}\right) + j \cdot g_{out}\left(i, \Im\left\{p\right\}, \Im\left\{y\right\}, \frac{\tau^p}{2}\right), \tag{4.47}$$

$$\begin{aligned}&-\frac{\partial}{\partial p} g_{out}\left(i, p, y, \tau^p\right)\\ &= -\frac{\partial}{\partial \Re\left\{p\right\}} g_{out}\left(i, \Re\left\{p\right\}, \Re\left\{y\right\}, \frac{\tau^p}{2}\right) - \frac{\partial}{\partial \Im\left\{p\right\}} g_{out}\left(i, \Im\left\{p\right\}, \Im\left\{y\right\}, \frac{\tau^p}{2}\right). \end{aligned} \tag{4.48}$$

4.3 Subcarrier-Wise Estimation with Bussgang Theorem

Both EM and GAMP have very high complexity, and are not practical for Massive MIMO scenarios, or for high numbers of subcarriers. In this section, a linear estimator based on the Bussgang theorem is proposed. This theorem [16] states that a nonlinear

distortion of a Gaussian signal can be expressed as a linear transformation plus uncorrelated noise. Mathematically, (4.1) can be modeled as:

$$\mathbf{y} - \mathbf{b} = \mathbf{K}\mathbf{z} + \mathbf{e}, \tag{4.49}$$

and $\mathbf{K} \in \mathbb{C}^{K \times K}$ can be chosen such that:

$$\mathrm{E}\left\{\mathbf{z}\mathbf{e}^H\right\} = \mathbf{0}. \tag{4.50}$$

Using $\mathbf{e} = \mathbf{y} - \mathbf{b} - \mathbf{K}\mathbf{z}$, it is easy to derive $\mathbf{K}$ and the covariance matrix of the quantization noise, $\mathbf{R}_{\mathbf{ee}}$:

$$\mathbf{K} = \mathbf{R}_{\mathbf{yz}}\mathbf{R}_{\mathbf{zz}}^{-1}; \tag{4.51}$$

$$\mathbf{R}_{\mathbf{ee}} = \mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{yz}}\mathbf{R}_{\mathbf{zz}}^{-1}\mathbf{R}_{\mathbf{zy}}. \tag{4.52}$$

For our problem, we have:

$$\mathbf{R}_{\mathbf{zz}} = \mathbf{A}\mathbf{R}_{\mathbf{hh}}\mathbf{A}^H + \mathbf{R}_{\mathbf{ww}}. \tag{4.53}$$

The Bussgang gain and noise covariance are easily adapted from the results in [17] (note the additional factors of $\sqrt{2}$ or 2 with respect to the reference due to the signals being complex-valued):

$$\mathbf{K} = \frac{2}{\sqrt{\pi}}\mathrm{diag}\left\{\mathbf{R}_{\mathbf{zz}}\right\}^{-1/2}; \tag{4.54}$$

$$\mathbf{R}_{\mathbf{ee}} = \mathbf{R}_{\mathbf{yy}} - \frac{4}{\pi}\mathrm{diag}\left\{\mathbf{R}_{\mathbf{zz}}\right\}^{-1/2}\mathbf{R}_{\mathbf{zz}}\mathrm{diag}\left\{\mathbf{R}_{\mathbf{zz}}\right\}^{-1/2}, \tag{4.55}$$

where

$$\mathbf{R}_{\mathbf{yy}} = \frac{4}{\pi}\arcsin\left(\mathrm{diag}\left\{\mathbf{R}_{\mathbf{zz}}\right\}^{-1/2}\mathbf{R}_{\mathbf{zz}}\mathrm{diag}\left\{\mathbf{R}_{\mathbf{zz}}\right\}^{-1/2}\right). \tag{4.56}$$

Using (4.49), we can now model our quantized system (4.1) as an unquantized one:

$$\mathbf{y} - \mathbf{b} = \mathbf{B}\mathbf{h} + \boldsymbol{\eta}, \tag{4.57}$$

where $\mathbf{B} = \mathbf{K}\mathbf{A}$, and $\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \mathbf{K}\mathbf{R}_{\mathbf{ww}}\mathbf{K}^H + \mathbf{R}_{\mathbf{ee}}$. Note that the quantization noise $\mathbf{e}$ is *not* Gaussian, and therefore this approach is suboptimal. If $\mathbf{R}_{\mathbf{zz}}$ is assumed to be diagonal (which holds if $\mathbf{R}_{\mathbf{hh}}$ is diagonal, the pilots are orthogonal and the number of transmit antennas is large), the problem can be decoupled and $\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ reduces to:

$$\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \frac{4}{\pi}\mathrm{diag}\left\{\mathbf{R}_{\mathbf{zz}}\right\}^{-1}\mathrm{diag}\left\{\mathbf{R}_{\mathbf{ww}}\right\} + 2\left(1 - \frac{2}{\pi}\right)\mathbf{I}_M. \tag{4.58}$$

If, additionally, $\mathbf{R}_{\mathbf{zz}}$ and $\mathbf{R}_{\mathbf{ww}}$ are scaled identities (which, if diagonality is already assumed, only requires that the noise and pilots do not change their variance over time), then the problem simplifies even further. In this case, the Bussgang decomposition reduces to a scalar factor and i.i.d. noise, and from (2.7), we have:

$$\mathbf{Y}_{n_r\cdot\cdot} = \mathcal{Q}\left(\sum_{n_t=1}^{N_t}\mathbf{F}^H\boldsymbol{\Lambda}_{n_r n_t}\mathbf{X}_{n_t\cdot\cdot} + \mathbf{W}_{n_r}\right), \tag{4.59}$$

$$\mathbf{Y}_{n_r\cdot\cdot} = k\sum_{n_t=1}^{N_t}\mathbf{F}^H\mathbf{\Lambda}_{n_rn_t}\mathbf{X}_{n_t\cdot\cdot} + \tilde{\mathbf{W}}_{n_r}, \tag{4.60}$$

with $k = \frac{2}{\sigma_z\sqrt{\pi}}$, and $\sigma_{\tilde{w}}^2 = \frac{4}{\pi}\frac{\sigma_w^2}{\sigma_z^2} + 2\left(1-\frac{2}{\pi}\right)$. This allows to use standard OFDM techniques: apply an FFT to $\mathbf{Y}$, and then estimate $\mathbf{H}$ subcarrier-wise:

$$\boldsymbol{\mathcal{Y}}_{n_r\cdot\cdot} = \mathbf{F}\mathbf{Y}_{n_r\cdot\cdot} = k\sum_{n_t=1}^{N_t}\mathrm{diag}\left\{\mathcal{H}_{n_rn_t\cdot}\right\}\mathbf{X}_{n_t\cdot\cdot} + \tilde{\boldsymbol{\mathcal{W}}}_{n_r\cdot\cdot}, \tag{4.61}$$

$$\boldsymbol{\mathcal{Y}}_{\cdot\cdot n} = k\mathbf{H}_{\cdot\cdot n}\mathbf{X}_{\cdot\cdot n} + \tilde{\boldsymbol{\mathcal{W}}}_{\cdot\cdot n}, \tag{4.62}$$

where $\boldsymbol{\mathcal{Y}}_{\cdot\cdot n} \in \mathbb{C}^{N_r\times T}$, $\mathbf{H}_{\cdot\cdot n} \in \mathbb{C}^{N_r\times N_t}$, $\mathbf{X}_{\cdot\cdot n} \in \mathbb{C}^{N_t\times T}$, and $\tilde{\boldsymbol{\mathcal{W}}}_{\cdot\cdot n} \in \mathbb{C}^{N_r\times T}$ are respectively the frequency-domain observations, channel, pilots and noise at subcarrier n. Then, the frequency-domain channel estimation at each subcarrier can be done, as usual, with BLUE (recall that the noise is assumed to be i.i.d. for this case):

$$\hat{\boldsymbol{\mathcal{H}}}_{\cdot\cdot n} = \frac{1}{k}\boldsymbol{\mathcal{Y}}_{\cdot\cdot n}\mathbf{X}_{\cdot\cdot n}^H\left(\mathbf{X}_{\cdot\cdot n}\mathbf{X}_{\cdot\cdot n}^H\right)^{-1}, \tag{4.63}$$

or MMSE:

$$\hat{\boldsymbol{\mathcal{H}}}_{\cdot\cdot n} = \frac{1}{k}\boldsymbol{\mathcal{Y}}_{\cdot\cdot n}\mathbf{X}_{\cdot\cdot n}^H\left(\mathbf{X}_{\cdot\cdot n}\mathbf{X}_{\cdot\cdot n}^H + \frac{\sigma_{\tilde{w}}^2}{\sigma_h^2}\mathbf{I}_{N_t}\right)^{-1}. \tag{4.64}$$

4.4 A-Posteriori-Probability Soft-Input-Soft-Output equalizer

In this section, we present an A-Posteriori-Probability (APP) Soft-Input-Soft-Output (SISO) equalizer, based on the Message Passing (MP) algorithm introduced in [18].

Consider a modulation scheme with B bits per symbol, and $Q = 2^B$ constellation points. Let us denote the sequence of transmitted bits in one block by $\mathbf{m} \in \mathbb{Z}_2^{BP\times 1}$, where $\mathbb{Z}_2 \triangleq \{0,1\}$. Note that $\mathbf{m}$ denotes the input to the modulator, with channel coding already applied. We further define the modulation function $\mathcal{M}(\cdot)$ as the function that transforms the input bits to the transmitted constellation symbols:

$$\begin{aligned}\mathcal{M}: \mathbb{Z}_2^{BP} &\rightarrow \mathbb{C}^P \\ \mathbf{m} &\rightarrow \mathcal{M}(\mathbf{m}) = \begin{pmatrix}\mathcal{M}\left(\mathbf{m}_{1:B}\right)\\ \mathcal{M}\left(\mathbf{m}_{(B+1):(2B)}\right)\\ \vdots \\ \mathcal{M}\left(\mathbf{m}_{((P-1)B+1):(PB)}\right)\end{pmatrix}.\end{aligned} \tag{4.65}$$

With this definition, and using the appropriate section of Chapter 2, we can write the following problem model:

$$\mathbf{y} = \mathcal{Q}\left(\mathbf{A}\mathcal{M}(\mathbf{m}) + \mathbf{b} + \mathbf{w}\right), \tag{4.66}$$

with $\mathbf{A} \in \mathbb{C}^{K\times P}$.

We define the log-likelihood function $\ell(m)$ of a bit m as:

$$\ell(m) = \log\frac{\Pr[m=0]}{\Pr[m=1]}. \tag{4.67}$$

We denote $D = BP$, and define the vector $\boldsymbol{\ell}^{\text{in}} \in \mathbb{R}^{D\times 1}$ of a-priori log-likelihood ratios of the encoded message bits $\mathbf{m}$:

$$\ell_d^{\text{in}} = \ell\left(m_d\right), \qquad d \in \{1, \ldots, D\}. \tag{4.68}$$

The vector $\boldsymbol{\ell}^{\text{out}}$ of output log-likelihood ratios is an updated version of $\boldsymbol{\ell}^{\text{in}}$, which includes the information in $\mathbf{y}$ given by (4.66):

$$\ell_d^{\text{out}} = \ell\left(m_d \mid \mathbf{y}\right), \qquad d \in \{1, \ldots, D\}. \tag{4.69}$$

Figure 4.3: A-posteriori Probability Soft-Input-Soft-Output equalizer

The proposed SISO APP equalizer is depicted in Figure 4.3. From the observation vector $\mathbf{y}$ and the log-likelihood ratios of the input bits, $\boldsymbol{\ell}^{\text{in}}$, the equalizer obtains the updated log-likelihood ratios of the input bits, $\boldsymbol{\ell}^{\text{out}}$. These log-likelihood ratios can then be used in the channel decoder to construct a fully soft scheme.

4.4.1 Message Passing algorithm

The proposed APP SISO block is based on the Message Passing algorithm [18]. This algorithm is useful for marginalizing high-dimensional probability density functions (PDFs) without incurring exponential complexity. To apply the method, first we need to factorize the PDF into several factor functions which depend on few variables. From Bayes' Rule we know that:

$$p_{\mathbf{m}\,|\,\mathbf{y}}\left(\mathbf{m} \mid \mathbf{y}\right) = \frac{p_{\mathbf{m}}\left(\mathbf{m}\right) p_{\mathbf{y}\,|\,\mathbf{m}}\left(\mathbf{y} \mid \mathbf{m}\right)}{p_{\mathbf{y}}\left(\mathbf{y}\right)}. \tag{4.70}$$

As $p_{\mathbf{y}}\left(\mathbf{y}\right)$ does not depend on $\mathbf{m}$, we can factorize $p_{\mathbf{m}\,|\,\mathbf{y}}\left(\mathbf{m} \mid \mathbf{y}\right)$ as:

$$p_{\mathbf{m}\,|\,\mathbf{y}}\left(\mathbf{m} \mid \mathbf{y}\right) = \frac{1}{p_{\mathbf{y}}\left(\mathbf{y}\right)} \prod_{d=1}^{D} p_{m_d}\left(m_d\right) \prod_{k=1}^{K} p_{y_k\,|\,\mathbf{m}}\left(y_k \mid \mathbf{m}\right), \tag{4.71}$$

where

$$p_{y_k\,|\,\mathbf{m}}\left(y_k \mid \mathbf{m}\right) = \Phi\left(\frac{\Re\left\{y_k\right\}\Re\left\{z_k\right\}}{\sigma_k/\sqrt{2}}\right)\Phi\left(\frac{\Im\left\{y_k\right\}\Im\left\{z_k\right\}}{\sigma_k/\sqrt{2}}\right), \tag{4.72}$$

and

$$z_k = b_k + \sum_{p=1}^{P} a_{kp}\mathcal{M}\left(\mathbf{m}_{((p-1)B+1):(pB)}\right). \tag{4.73}$$

The use of Message Passing only makes sense if each one of the $p_{y_k\,|\,\mathbf{m}}\left(y_k \mid \mathbf{m}\right)$ depend only on a few m_d. From (4.73), we see that this happens only when each one of the rows of the matrix $\mathbf{A}$ has only a few nonzero elements. In the SC (2.21) and NCP (2.32) cases, this is fulfilled when $L \ll N$, because $\mathbf{A}$ is defined in terms of convolution matrices.

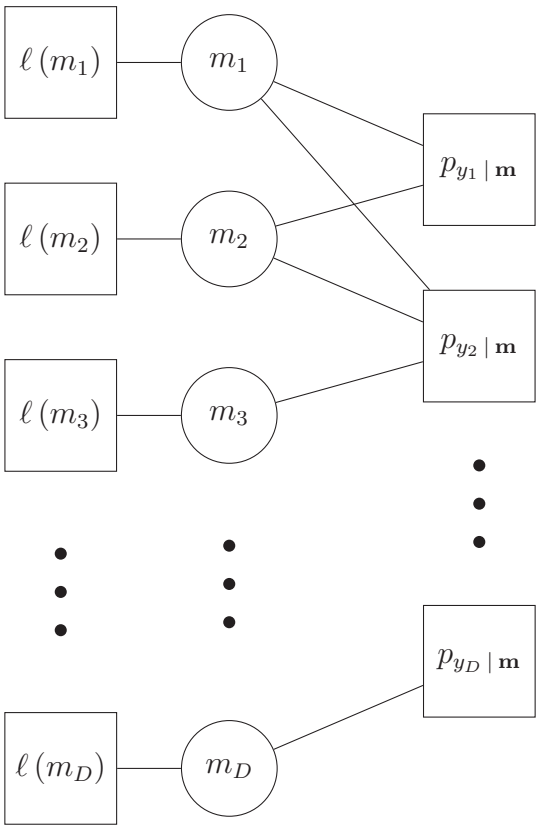

Figure 4.4: Factor graph for the Message Passing algorithm

The factor graph for (4.71) is depicted in Figure 4.4. The Message Passing algorithm uses a strategy called Belief Propagation (BP), in which both the variable nodes (circles in the figure) and the factor nodes (squares in the figure) iteratively send messages to their neighbors.

The variable-to-factor messages at iteration i are denoted by $\mu^{(i)}_{d\to k}$, where d is the source variable node and k is the destination factor node. The content of the message is the LLR of the corresponding bit m_d, given the information received from all the other factors in the previous iteration:

$$\mu^{(i)}_{d\to k} = \ell\left(m_d \,\middle|\, \left\{\mu^{(i-1)}_{d\leftarrow k'}\right\}_{k'\in\mathcal{K}_d\setminus k}\right), \tag{4.74}$$

where $\mu^{(i)}_{d\leftarrow k}$ denotes the message sent by factor k to variable d at iteration i. The set $\mathcal{K}_d$ is the set of neighbors of variable node d. In our problem, the variable-to-factor messages (4.74) become:

$$\mu^{(i)}_{d\to k} = \ell^{\text{in}}_d + \sum_{k'\in\mathcal{K}_d\setminus k} \mu^{(i-1)}_{d\to k'}. \tag{4.75}$$

The factor-to-variable messages $\mu^{(i)}_{d\leftarrow k}$ ae defined as the LLR of the corresponding bit m_d, given the messages received from all other variables and the information contained in

the observation y_k:

$$\begin{aligned}\mu_{d\leftarrow k}^{(i)} &= \ell\left(m_d \mid y_k, \left\{\mu_{d'\to k}^{(i)}\right\}_{d'\in\mathcal{D}_k\setminus d}\right) \\ &= \log\frac{p_{m_d\mid y_k,\left\{\mu_{d'\to k}^{(i)}\right\}_{d'\in\mathcal{D}_k\setminus d}}\left(0\mid y_k, \left\{\mu_{d'\to k}^{(i)}\right\}_{d'\in\mathcal{D}_k\setminus d}\right)}{p_{m_d\mid y_k,\left\{\mu_{d'\to k}^{(i)}\right\}_{d'\in\mathcal{D}_k\setminus d}}\left(1\mid y_k, \left\{\mu_{d'\to k}^{(i)}\right\}_{d'\in\mathcal{D}_k\setminus d}\right)},\end{aligned} \tag{4.76}$$

where $\mathcal{D}_k$ is the set of neighbors of factor node k. To compute $\mu_{d\leftarrow k}^{(i)}$, we first transform the received variable-to-node messages to probabilities. From (4.67), we obtain:

$$p_{d\to k}^{(i)}(1) \triangleq p_{m_d\mid\mu_{d\to k}^{(i)}}\left(1\mid\mu_{d\to k}^{(i)}\right) = \frac{1}{1+e^{\mu_{d\to k}^{(i)}}}, \tag{4.77}$$

$$p_{d\to k}^{(i)}(0) \triangleq p_{m_d\mid\mu_{d\to k}^{(i)}}\left(0\mid\mu_{d\to k}^{(i)}\right) = \frac{e^{\mu_{d\to k}^{(i)}}}{1+e^{\mu_{d\to k}^{(i)}}}. \tag{4.78}$$

This allows us to express (4.76) as:

$$\mu_{d\leftarrow k}^{(i)} = \log\frac{\sum\limits_{\{m_{d'}\}_{d'\in\mathcal{D}_k\setminus d}\in\mathbb{Z}_2^{|\mathcal{D}_k|-1}}\cdots\sum\prod_{d'\in\mathcal{D}_k\setminus d} p_{d\to k}^{(i)}(m_{d'})\, p_{y_k\mid m_d,\{m_{d'}\}_{d\in\mathcal{D}_k\setminus d}}\left(y_k\mid 0,\{m_{d'}\}_{d\in\mathcal{D}_k\setminus d}\right)}{\sum\limits_{\{m_{d'}\}_{d'\in\mathcal{D}_k\setminus d}\in\mathbb{Z}_2^{|\mathcal{D}_k|-1}}\cdots\sum\prod_{d'\in\mathcal{D}_k\setminus d} p_{d\to k}^{(i)}(m_{d'})\, p_{y_k\mid m_d,\{m_{d'}\}_{d\in\mathcal{D}_k\setminus d}}\left(y_k\mid 1,\{m_{d'}\}_{d\in\mathcal{D}_k\setminus d}\right)}, \tag{4.79}$$

where, due to the zeros in $\mathbf{A}$, we have $p_{y_k\mid m_d,\{m_{d'}\}_{d\in\mathcal{D}_k\setminus d}}\left(y_k\mid m_d,\{m_{d'}\}_{d\in\mathcal{D}_k\setminus d}\right) = p_{y_k\mid\mathbf{m}}(y_k\mid\mathbf{m})$, which can be computed from (4.72) and (4.73) by appropriately changing the range of the summation that calculates z_k. The notation $\sum\limits_{\{m_{d'}\}_{d'\in\mathcal{D}_k\setminus d}\in\mathbb{Z}_2^{|\mathcal{D}_k|-1}}\cdots\sum$ represents the sum over all $2^{|\mathcal{D}_k|-1}$ possible combinations of values of $\{m_{d'}\}_{d'\in\mathcal{D}_k\setminus d}$, where each $m_{d'}$ can take the value 0 or 1.

The Message Passing algorithm applies (4.74) and (4.79) in each iteration, until the change in $\mu_{d\leftarrow k}^{(i)}$ is small enough. The output marginal LLRs of the bits, $\boldsymbol{\ell}^{\text{out}}$, are then computed as:

$$\ell_d^{\text{out}} = \ell_d^{\text{in}} + \sum_{k\in\mathcal{K}_d}\mu_{d\leftarrow k}^{(i)}. \tag{4.80}$$

A full listing of the algorithm is given in Algorithm 4.3.

4.5 Computational complexity of the presented algorithms

In this section, we compare the computational complexity of the presented algorithms, measured as the number of complex multiplications. We define dimensions K, P, R such that $\mathbf{A}\in\mathbb{C}^{K\times P}$, and (4.1) is solved R times ($R = N_r$ for estimation and $R = M$ for equalization).

Algorithm 4.3 APP SISO equalizer

Input: $\boldsymbol{\ell}^{\text{in}}, \mathbf{y}, \mathbf{A}, \mathbf{b}$
  Initialize: $i = 0$, $\mu_{d\leftarrow k}^{(-1)} = 0$
  while $i < i_{\max}$ **and** $\sum_{d,k} \left(\mu_{d\leftarrow k}^{(i)} - \mu_{d\leftarrow k}^{(i-1)}\right)^2 < \epsilon \sum_{d,k} \left(\mu_{d\leftarrow k}^{(i)}\right)^2$ **do**
    for $d = 1$ **to** D **do**
      for $k = 1$ **to** K **do**
        $\mu_{d\to k}^{(i)}$ from (4.74)
      end for
    end for
    for $k = 1$ **to** K **do**
      for $d = 1$ **to** D **do**
        $\mu_{d\leftarrow k}^{(i)}$ from (4.79)
      end for
    end for
    $i = i + 1$
  end while
  for $d = 1$ **to** D **do**
    $\ell_d^{\text{out}} = \ell_d^{\text{in}} + \mu_{d\leftarrow k}^{(i-1)}$
  end for
Output: $\boldsymbol{\ell}^{\text{out}}$

4.5.1 Computational Complexity of EM

- **Expectation step:**

  The expectation step computes $\mathrm{E}\left\{\mathbf{z} \mid \mathbf{y}, \hat{\mathbf{h}}^{(i)}\right\}$. This amounts to R computations of (4.13), each one having a complexity dominated by the product $\mathbf{A}\hat{\mathbf{h}}$, yielding $\mathcal{O}(KPR)$.

- **Maximization step:**

  This is an MMSE solution that multiplies $\mathbf{B} = \left(\mathbf{A}^H \mathbf{R}_{\mathbf{ww}} \mathbf{A} + \mathbf{R}_{\mathbf{hh}}\right)^{-1} \mathbf{A}\mathbf{R}_{\mathbf{ww}} \in \mathbb{C}^{P\times K}$ by the expectation $\hat{\mathbf{z}} \in \mathbb{C}^{K\times 1}$. Note that $\mathbf{B}$ only varies when the channel changes, and thus it only needs to be recalculated once in each channel coherence period. The maximization step amounts only to a matrix-vector multiplication $\mathbf{By}$, which is done R times. The complexity of this step is: $\mathcal{O}(KPR)$.

These two steps are done for I iterations, until the algorithm converges. This results in an overall complexity of EM of:

$$T_{\text{EM}} = \mathcal{O}(2IKPR). \tag{4.81}$$

4.5.2 Computational Complexity of GAMP

The scalar activation functions g_{in} and g_{out} and their derivatives (4.34)-(4.45) all have linear complexity in either the size of the input (P) or that of the output (K). Therefore,

the most computationally expensive step of each iteration of GAMP are the matrix-vector multiplications involving $\mathbf{A}$ and $\mathbf{A}_2$. Specifically, two multiplications with $\mathbf{A}$ and two with $\mathbf{A}_2$ are applied, each one with complexity KP. The algorithm needs to be applied R times. Assuming it runs for I iterations, the overall complexity of GAMP-MMSE is:

$$T_{\mathrm{GAMP-MMSE}} = \mathcal{O}\left(4IKPR\right). \tag{4.82}$$

4.5.3 Computational Complexity of the Bussgang estimator

In the single-carrier case, the linear estimator needs to compute an MMSE solution with the whole matrix $\mathbf{KA} \in \mathbb{C}^{K\times P}$ and observation $\mathbf{Y} \in \mathbb{C}^{K\times P}$. Again, note that the computation of the matrices $\mathbf{KA}$ and $\mathbf{R}_{\boldsymbol{\eta\eta}}$ only needs to be performed once every channel realization, and the same applies to the MMSE multiplier matrix $\mathbf{G} = \left(\mathbf{A}^H\mathbf{K}^H\mathbf{R}_{\boldsymbol{\eta\eta}}^{-1}\mathbf{KA} + \mathbf{R}_{\mathbf{hh}}\right)^{-1}\mathbf{A}^H\mathbf{K}^H$ The complexity of the Bussgang estimator then reduces to a matrix-vector multiplication of $\mathbf{G} \in \mathbb{C}^{P\times K}$ with $\mathbf{y}$, which is done R times

$$T_{\mathrm{Buss.-SC}} = \mathcal{O}\left(KPR\right). \tag{4.83}$$

In an OFDM system, the Bussgang estimator allows subcarrier-wise equalization, which amounts to N MMSE calculations of (4.64), where again the matrix inversion only needs to be performed once per channel realization. This gives a complexity of:

$$T_{\mathrm{Buss.-MC,CE}} = \mathcal{O}\left(N_r N_t N T\right) \tag{4.84}$$

for channel estimation, and

$$T_{\mathrm{Buss.-MC,EQ}} = \mathcal{O}\left(N_r N_t N M\right) \tag{4.85}$$

for equalization.

4.5.4 Computational complexity of the APP SISO equalizer

The dominant step in the complexity of the APP SISO equalizer is the computation of the factor-to-variable messages $\mu_{d\leftarrow k}^{(i)}$ in (4.79). In the single-carrier case (recall that the APP equalizer is not useful for OFDM systems), the rows of $\mathbf{A}$ have LN_t nonzero elements. Thus, all factor nodes have $|\mathcal{D}_k| = BLN_t$ neighbors, whre B is the number of bits per constellation symbol. The summation in the numerator of (4.79) has then 2^{BLN_t-1} terms. For each term, BLN_t multiplications are performed to compute $\prod_{d'\in\mathcal{D}_k\setminus d} p_{d\to k}^{(i)}\left(m_{d'}\right)$, and another LN_t multiplications are needed to calculate $p_{y_k \mid m_d,\{m_{d'}\}_{d\in\mathcal{D}_k\setminus d}}$ from (4.72). This is done twice (once for the numerator and once for the denominator), which means that each factor-to-variable message has complexity:

$$T_{\mathrm{APP},\mu_{d\leftarrow k}^{(i)}} = \mathcal{O}\left(BLN_t 2^{BLN_t-1}\right). \tag{4.86}$$

The priors $\prod_{d'\in\mathcal{D}_k\setminus d} p_{d\to k}^{(i)}\left(m_{d'}\right)$ depend weakly on d. Only one of the $p_{d\to k}^{(i)}\left(m_{d'}\right)$ is missing in the product each time, which allows computation by calculating the full product

and dividing by the missing $p^{(i)}_{d\to k}(m_d)$). This lets us compute all the priors with only double complexity compared to that of calculating the full product once. Additionally, $p_{y_k \mid m_d, \{m_{d'}\}_{d\in\mathcal{D}_k\setminus d}}$ has the same values in different positions for every d. which means it only needs to be calculated for one value of d. Therefore, the calculation of $\mu^{(i)}_{d\leftarrow k}$ for one k and all d has double complexity compared to that for only one d:

$$T_{\mathrm{APP},\left\{\mu^{(i)}_{d\leftarrow k}\right\}_{d\in\mathcal{D}_k}} = \mathcal{O}\left(BLN_t 2^{BLN_t}\right). \tag{4.87}$$

The messages still have to be computed for all $k \in \{1,\ldots,NN_r\}$, at each one of the I iterations, and for each one of the M symbol blocks. The overall complexity is, as expected, prohibitive:

$$T_{\mathrm{APP}} = \mathcal{O}\left(IN_rN_tNLMB2^{BLN_t}\right). \tag{4.88}$$

However, as mentioned before, the APP SISO equalizer is an optimal detector when it converges, and is useful as a benchmark for the performance of the practical algorithms in simulations.

All the complexity results for channel estimation and equalization are summarized in Table 4.1, where MC stands for multi-carrier (OFDM), and B is the number of bits per symbol.

Table 4.1: Computational complexity of the presented algorithms

Algorithm	**Estimation**	**Equalization**
EM	$2IN_rN_tNLT$	$2IN_rN_tN^2M$
GAMP	$4IN_rN_tNLT$	$4IN_rN_tN^2M$
Bussgang (SC)	N_rN_tNLT	$N_rN_tN^2M$
Bussgang (MC)	N_rN_tNT	N_rN_tNM
APP SISO	–	$IN_rN_tNLMB2^{BLN_t}$

Chapter 5

Algorithms for Joint Channel and Data Estimation (JCD)

In the previous chapter, we introduced several linear and nonlinear algorithms that can be used for both channel estimation and data equalization in the quantized MIMO problem. A receiver can use one of the algorithms for channel estimation, and then use the obtained channel estimate to equalize the data using the same algorithm or another one. However this approach can be improved if channel estimation and equalization are performed jointly. In this chapter, an iterative, turbo-like technique introduced in [19] for this purpose is discussed. Additionally, we derive closed-form expressions for the intermediate estimation variances of the algorithm for our frequency-selective system in the OFDM, SC and NCP settings

5.1 Iterative turbo-based JCD estimator

Consider a quantized MIMO system of either one of the OFDM, SC or NCP types. For the Joint Channel and Data (JCD) estimation problem, we denote the number of known pilot blocks by T and the number of data blocks by M. The pilots corresponding to antenna $n_t \in \{1, \ldots, N_t\}$ are denoted by $\mathbf{X}_{n_t,\{1,\ldots,T\}} \in \mathbb{C}^{N\times T}$, while $\mathbf{X}_{n_t,\{T+1,\ldots,T+M\}} \in \mathbb{C}^{N\times M}$ represents the data blocks transmitted by the same antenna (during this section and for the NCP case, set $N = 1$).

The program flow of the turbo-based JCD estimator is depicted in Figure 5.1. At each iteration i, two steps are applied alternately: channel estimation and equalization. This procedure repeats until the algorithm converges. It will be seen in the simulation results that this gives better accuracy than successive estimation of channel and data.

5.1.1 Channel estimation stage

The channel estimation step of the JCD turbo-based approach uses as input the quantized observations $\{\mathbf{Y}_{n_r\cdot\cdot}\}_{n_r=1}^{N_r}$, the known pilots $\left\{\mathbf{X}_{n_t,\{1,\ldots,T\},\cdot}\right\}_{n_t=1}^{N_t}$, estimated data

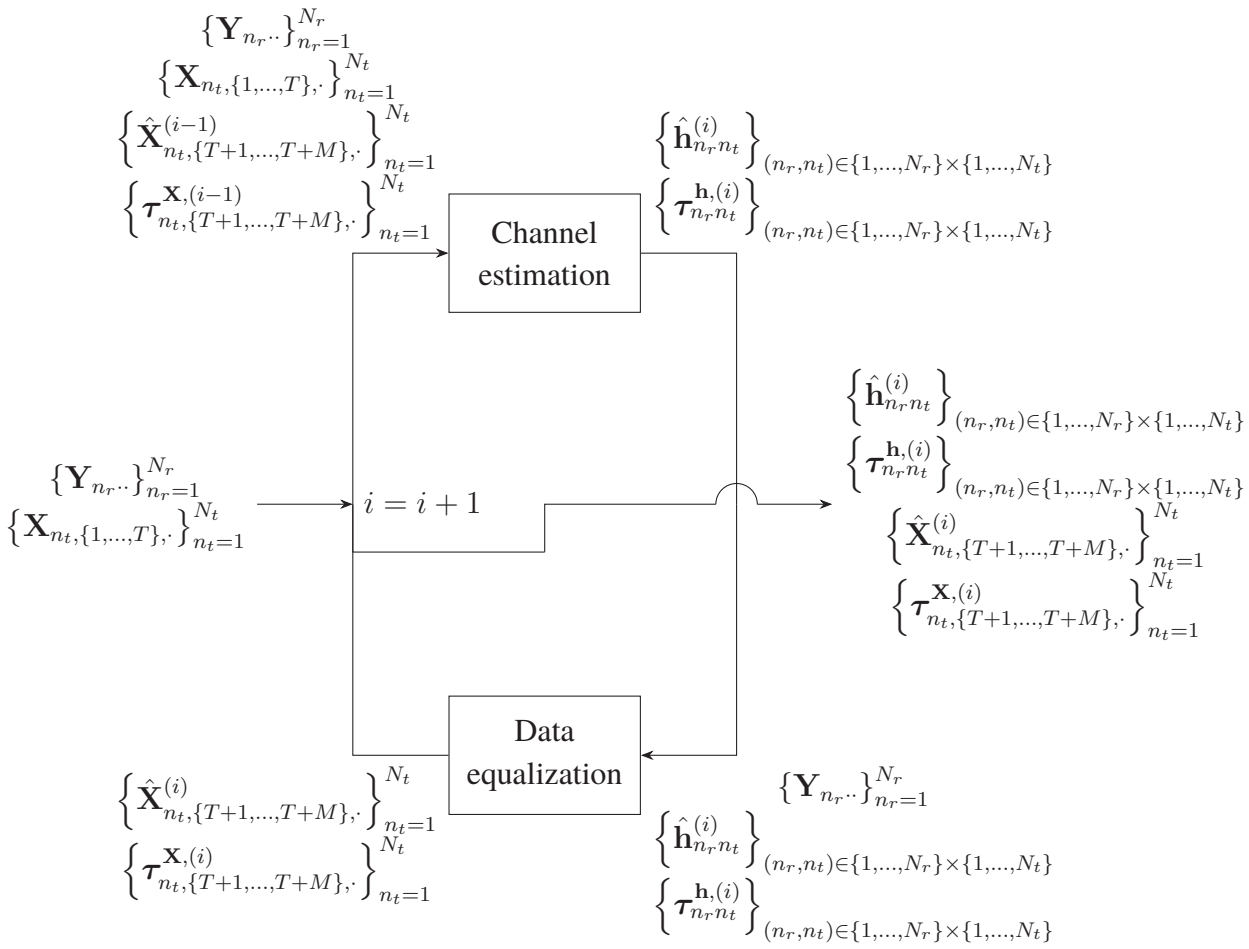

Figure 5.1: Turbo-based JCD estimator

symbols from the previous iteration, $\left\{\hat{\mathbf{X}}^{(i-1)}_{n_t,\{T+1,\ldots,T+M\},\cdot}\right\}_{n_t=1}^{N_t}$, and their estimated variances $\left\{\boldsymbol{\tau}^{\mathbf{X},(i-1)}_{n_t,\{T+1,\ldots,T+M\},\cdot}\right\}_{n_t=1}^{N_t}$.

This block performs channel estimation using both the known pilots and the estimated data symbols. For this purpose, a large pilot block is constructed by concatenating the actual pilots and the data estimates:

$$\mathbf{X}^{(i)}_{n_t\cdot\cdot} = \left[\begin{array}{cc} \mathbf{X}_{n_t,\{1,\ldots,T\},\cdot} & \hat{\mathbf{X}}_{n_t,\{T+1,\ldots,T+M\},\cdot} \end{array}\right] \in \mathbb{C}^{N\times(T+M)}. \tag{5.1}$$

Then, depending on the system type (OFDM, SC or NCP), the matrices $\mathbf{X}^{(i)}_{n_t\cdot\cdot}$ and the observations $\mathbf{X}^{(i)}_{n_t\cdot\cdot}$ are used to construct the quantized matrix-vector channel estimation model:

$$\mathbf{y}_{n_r} = \mathcal{Q}\left(\mathbf{A}\mathbf{h}_{n_r} + \mathbf{w}_{n_r}\right). \tag{5.2}$$

The procedure to construct the model for all systems is explained in Chapter 2. The matrix $\mathbf{A} \in \mathbb{C}^{N(T+M)\times LN_t}$ is not known exactly, due to the variance in the estimation of $\left\{\hat{\mathbf{X}}^{(i-1)}_{n_t,\{T+1,\ldots,T+M\},\cdot}\right\}_{n_t=1}^{N_t}$. To improve the estimation performance, this variance is also taken into account. Let us denote the estimated value of $\mathbf{A}$ by $\hat{\mathbf{A}}$, and define the perturbation $\tilde{\mathbf{A}}$ such that:

$$\mathbf{A} = \hat{\mathbf{A}} + \tilde{\mathbf{A}}. \tag{5.3}$$

We can now rewrite (5.2) as:

$$\mathbf{y}_{n_r} = \mathcal{Q}\left(\hat{\mathbf{A}}\mathbf{h}_{n_r} + \left(\tilde{\mathbf{A}}\mathbf{h}_{n_r} + \mathbf{w}_{n_r}\right)\right) = \mathcal{Q}\left(\hat{\mathbf{A}}\mathbf{h}_{n_r} + \boldsymbol{\eta}\right). \tag{5.4}$$

The noise term is now $\boldsymbol{\eta} = \tilde{\mathbf{A}}\mathbf{h}_{n_r} + \mathbf{w}_{n_r}$. We define $K = N(T+M)$, $P = LN_t$, and denote the variance of the p-th component of $\mathbf{h}_{n_r}$ by $\sigma^2_{h_p}$. As the quantized estimation algorithms in Chapter 4 assume uncorrelated noise, we will neglect the cross-correlations of the noise terms and compute only the diagonal terms. The variance of the k-th component of $\boldsymbol{\eta}$ is given by:

$$\begin{aligned}\sigma^2_{\eta_k} &= \begin{cases} \sigma^2_{w_k}, & \text{if } k \le NT, \\ \sigma^2_{w_k} + \mathrm{E}\left\{\sum_{p=1}^{P} |\tilde{a}_{kp}|^2 \sigma^2_{h_p}\right\}, & \text{if } k > NT \end{cases} \\ &= \begin{cases} \sigma^2_{w_k}, & \text{if } k \le NT, \\ \sigma^2_{w_k} + \sum_{p=1}^{P} \sigma^2_{a_{kp}}\sigma^2_{h_p}, & \text{if } k > NT, \end{cases}\end{aligned} \tag{5.5}$$

where the case distinction is done because, when $k \le NT$, the sample corresponds to a pilot symbol and hence there is no error in a_{kp}. Additionally, independence between a_{kp} and h_p was assumed. This variance is used in the estimation algorithms of Chapter 4 instead of the noise variance σ^2_w. The value of $\sigma^2_{a_{kp}}$ is obtained from the respective definition of the matrix $\mathbf{A}$. In the following, we give expressions for $\sigma^2_{n_k}$ for the OFDM, SC and NCP cases:

5.1.1.1 Perturbation in the OFDM case

In the OFDM case, the matrix $\mathbf{A}$ is defined as (2.11):

$$\mathbf{A} = \left[\left(\mathbf{X}_{1..}^T \diamond \mathbf{F}^H\right) \mathbf{F}_{N\times L} \;\middle|\; \left(\mathbf{X}_{2..}^T \diamond \mathbf{F}^H\right) \mathbf{F}_{N\times L} \;\middle|\; \cdots \;\middle|\; \left(\mathbf{X}_{N_t..}^T \diamond \mathbf{F}^H\right) \mathbf{F}_{N\times L} \right]. \tag{5.6}$$

Let us denote the entry of $\mathbf{A}$ corresponding to subcarrier n, symbol m channel tap ℓ, and transmit antenna n_t by $a_{n,m;\ell,n_t}$, such that $a_{n,m;\ell,n_t} = a_{n+(m-1)N,\ell+(n_t-1)L}$. From (5.6), we have:

$$a_{n,m;\ell,n_t} = \frac{1}{\sqrt{N}} \sum_{n'=0}^{N-1} x_{n_t,m,n'} e^{j\frac{2\pi}{N}nn'} e^{-j\frac{2\pi}{N}n'\ell}. \tag{5.7}$$

Plugging this into (5.5) yields:

$$\begin{aligned}
&\sigma^2_{\eta_{n_r,m,n}} \\
&= \sigma^2_{w_{n_r,m,n}} + \tfrac{1}{N}\mathrm{E}\left\{ \sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \left|\tilde{a}_{n,m;\ell,n_t}\right|^2 \sigma^2_{h_{n_r,n_t,\ell}} \right\} \\
&= \begin{cases} \sigma^2_{w_{n_r,m,n}}, & \text{if } m \le T \\ \sigma^2_{w_{n_r,m,n}} + \dfrac{1}{N} \displaystyle\sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \sum_{n'=0}^{N-1} \sum_{n''=0}^{N-1} & \\ \qquad \mathrm{E}\left\{ \tilde{x}_{n_t,m-T,n'} \tilde{x}^*_{n_t,m-T,n''} \right\} e^{j\frac{2\pi}{N}(n-\ell)(n'-n'')} \sigma^2_{h_{n_r,n_t,\ell}}, & \text{if } m > T \end{cases}
\end{aligned} \tag{5.8}$$

where $\tilde{x}_{n_t,m,n} = x_{n_t,m,n} - \hat{x}^{(i)}_{n_t,m,n}$ is the estimation error in $x_{n_t,m,n}$ We neglect any cross-correlation of the samples in $\mathbf{X}_{n_t..}$. Therefore, the only terms in the summation in (5.8) that are nonzero correspond to $n' = n''$. The equivalent noise variance for channel estimation is then:

$$\sigma^2_{\eta_{n_r,m,n}} = \begin{cases} \sigma^2_{w_{n_r,m,n}}, & \text{if } m \le T, \\ \sigma^2_{w_{n_r,m,n}} + \frac{1}{N} \displaystyle\sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \sum_{n'=0}^{N-1} \tau^x_{n_t,m,n'} \sigma^2_{h_{n_r,n_t,\ell}}, & \text{if } m > T. \end{cases} \tag{5.9}$$

If the channel variance is equal for all taps, (5.9) reduces to:

$$\sigma^2_{\eta_{n_r,m,n}} = \begin{cases} \sigma^2_{w_{n_r,m,n}}, & \text{if } m \le T, \\ \sigma^2_{w_{n_r,m,n}} + \sigma^2_h \frac{L}{N} \displaystyle\sum_{n_t=1}^{N_t} \sum_{n'=0}^{N-1} \tau^x_{n_t,m,n'}, & \text{if } m > T. \end{cases} \tag{5.10}$$

5.1.1.2 Perturbation in the SC and NCP cases

In the SC case, the entries of the matrix $\mathbf{A}$ are given by (see (2.20)):

$$a_{n,m;\ell,n_t} = x_{n_t,m,(n-\ell) \bmod N}. \tag{5.11}$$

Substituting this in (5.5) gives:

$$\sigma^2_{\eta_{n_r,m,n}} = \begin{cases} \sigma^2_{w_{n_r,m,n}}, & \text{if } m \le T, \\ \sigma^2_{w_{n_r,m,n}} + \displaystyle\sum_{\ell=0}^{L-1} \sum_{n_t=1}^{N_t} \tau^x_{n_t,m,(n-\ell) \bmod N} \sigma^2_{h_{n_r,n_t,\ell}}, & \text{if } m > T, \end{cases} \tag{5.12}$$

which, if the variance of the channel taps is constant, simplifies to:

$$\sigma^2_{\eta_{n_r,m,n}} = \begin{cases} \sigma^2_{w_{n_r,m,n}}, & \text{if } m \leq T, \\ \sigma^2_{w_{n_r,m,n}} + \sigma^2_h L \sum\limits_{n_t=1}^{N_t} \tau^x_{n_t,m,(n-\ell) \bmod N}, & \text{if } m > T. \end{cases} \tag{5.13}$$

For the NCP case, the expressions are obtained in the same way and turn out to be:

$$\sigma^2_{\eta_{n_r,m}} = \begin{cases} \sigma^2_{w_{n_r,m}}, & \text{if } m \leq T, \\ \sigma^2_{w_{n_r,m}} + \sum\limits_{\ell=0}^{L-1} \sum\limits_{n_t=1}^{N_t} \tau^x_{n_t,(m-\ell)} \sigma^2_{h_{n_r,n_t,\ell}}, & \text{if } m > T, \end{cases} \tag{5.14}$$

for the different variance case, and

$$\sigma^2_{\eta_{n_r,m}} = \begin{cases} \sigma^2_{w_{n_r,m}}, & \text{if } m \leq T, \\ \sigma^2_{w_{n_r,m}} + \sigma^2_h L \sum\limits_{n_t=1}^{N_t} \tau^x_{n_t,(m-\ell)}, & \text{if } m > T. \end{cases} \tag{5.15}$$

for the equal variance case.

5.1.2 Data equalization stage

The second step in each iteration is the data equalization stage. Here, the previously estimated channel $\left\{\hat{\mathbf{h}}^{(i)}_{n_r n_t}\right\}_{(n_r,n_t)\in\{1,\ldots,N_r\}\times\{1,\ldots,N_t\}}$, the estimation variance $\{\mathbf{Y}_{n_r\cdot\cdot}\}_{n_r=1}^{N_r}$, and the quantized observations $\left\{\boldsymbol{\tau}^{\mathbf{h},(i)}_{n_r n_t}\right\}_{(n_r,n_t)\in\{1,\ldots,N_r\}\times\{1,\ldots,N_t\}}$ are used to obtain a new estimate of the data symbols $\left\{\hat{\mathbf{X}}^{(i)}_{n_t,\{T+1,\ldots,T+M\},\cdot}\right\}_{n_t=1}^{N_t}$, and their estimation variance $\left\{\boldsymbol{\tau}^{\mathbf{X},(i)}_{n_t,\{T+1,\ldots,T+M\},\cdot}\right\}_{n_t=1}^{N_t}$. For this purpose, the appropriate matrix-vector model from Chapter 2 and one of the algorithms from Chapter 5 are used.

Similarly to the channel estimation case, the perturbation of the matrix $\mathbf{A} \in \mathbb{C}^{NN_r \times NN_t}$ can be included in the noise term as follows:

$$\mathbf{y}_m = \mathcal{Q}\left(\hat{\mathbf{A}}\mathbf{x}_m + \mathbf{b}_m + \tilde{\mathbf{A}}\mathbf{x}_m + \mathbf{w}_m\right) = \mathcal{Q}\left(\hat{\mathbf{A}}\mathbf{x}_m + \mathbf{b}_m + \boldsymbol{\eta}\right), \tag{5.16}$$

where the noise term is $\boldsymbol{\eta} = \tilde{\mathbf{A}}\mathbf{x}_m + \mathbf{w}_m$. Again the matrix $\mathbf{A} = \hat{\mathbf{A}} + \tilde{\mathbf{A}}$ is split into the estimated value $\hat{\mathbf{A}}$ and the error $\tilde{\mathbf{A}}$.

The variance of the noise term is now:

$$\sigma^2_{\eta_{n_r,m,n}} = \sigma^2_{w_{n_r,m,n}} + \sum_{n'=0}^{N-1} \sum_{n_t=1}^{N_t} \sigma^2_{a_{n,n_r;n',n_t}} \sigma^2_{x_{n_t,m,n'}}, \tag{5.17}$$

where $\sigma^2_{a_{n,n_r;n',n_t}}$ is the variance of $a_{n,n_r;n',n_t} = a_{n+(n_r-1)N,n'+(n_t-1)N}$.

In the following, we give expressions for the equivalent noise variance (5.17) for the OFDM, SC and NCP cases.

5.1.2.1 Perturbation in the OFDM case

From (2.16), we can write $a_{n,n_r;n',n_t}$ in the OFDM case as:

$$a_{n,n_r;n',n_t} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}nn'} \sum_{\ell=0}^{L-1} e^{-j\frac{2\pi}{N}n'\ell} h_{n_r,n_t,\ell}, \tag{5.18}$$

which, plugged into (5.17), yields:

$$\sigma^2_{\eta_{n_r,m,m}} = \sigma^2_{w_{n_r,m,n}} + \frac{1}{N} \sum_{n'=0}^{N-1} \sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \sum_{\ell'=0}^{L-1} e^{j\frac{2\pi}{N}n'(\ell'-\ell)} \mathrm{E}\left\{h_{n_r,n_t,\ell} h^*_{n_r,n_t,\ell'}\right\} \sigma^2_{x_{n_t,m,n'}}. \tag{5.19}$$

Again, we neglect the cross-correlations among the channel taps, which reduces (5.19) to:

$$\sigma^2_{\eta_{n_r,m,m}} = \sigma^2_{w_{n_r,m,n}} + \frac{1}{N} \sum_{n'=0}^{N-1} \sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \tau^h_{n_r,n_t,\ell} \sigma^2_{x_{n_t,m,n'}}. \tag{5.20}$$

Finally, if all symbols have the same variance σ^2_x, the result is:

$$\sigma^2_{\eta_{n_r,m,m}} = \sigma^2_{w_{n_r,m,n}} + \sigma^2_x \sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \tau^h_{n_r,n_t,\ell}. \tag{5.21}$$

5.1.2.2 Perturbation in the SC and NCP cases

In the SC case, from (2.21), we have:

$$a_{n,n_r;n',n_t} = h_{n_r,n_t,(n-n')\mathrm{mod}N}. \tag{5.22}$$

With the help of (5.17), we obtain:

$$\sigma^2_{\eta_{n_r,m,n}} = \sigma^2_{w_{n_r,m,n}} + \sum_{n'=n-\ell+1}^{n} \sum_{n_t=1}^{N_t} \tau^h_{n_r,n_t,(n-n')\mathrm{mod}N} \sigma^2_{x_{n_t,m,n'}}, \tag{5.23}$$

and, if all symbols have the same variance,

$$\sigma^2_{\eta_{n_r,m,n}} = \sigma^2_{w_{n_r,m,n}} + \sigma^2_x \sum_{n'=n-\ell+1}^{n} \sum_{n_t=1}^{N_t} \tau^h_{n_r,n_t,(n-n')\mathrm{mod}N}. \tag{5.24}$$

The variances in the NCP case have the same expressions as (5.23) and (5.24), but removing the $\mathrm{mod}N$ operations.

Chapter 6

Simulation Results

In this chapter, simulation results will be given to compare the algorithms presented in this thesis. The performance gap between linear and nonlinear techniques will be evaluated in different settings. The advantages of Joint Channel and Data Estimation (JCD) will be shown. The performance with higher order modulations and with non-Gaussian channel models will also be evaluated.

Except for the APP estimator test, all the simulations use a channel impulse response length of $L = 4$ taps. The block size is $N = 32$ for OFDM and SC with cyclic prefix. In the NCP case, the additional data slots available due to the lack of a cyclic prefix are also included, giving a block size of $N_{\mathrm{NCP}} = 36$.

In order to have a realistic comparison, all the simulations in this chapter use a punctured convolutional channel code. The rate of the code is $R_{\mathrm{cc,cp}} = 3/4$ for systems with cyclic prefix. To compare systems with the same throughput, the code rate in the scheme without cyclic prefix is $R_{\mathrm{cc,ncp}} = R_{\mathrm{cc,cp}} N/N_{\mathrm{ncp}} = 2/3$.

Two figures of merit have been used to evaluate the performance of the algorithms. In the channel estimation problem, the Normalized Mean Square Error (NMSE) is defined as:

$$\mathrm{NMSE}_{\mathbf{H}} = \frac{1}{N_{\mathrm{sim}} N_r N_t L \sigma_h^2} \sum_{n_{\mathrm{sim}}=1}^{N_{\mathrm{sim}}} \sum_{n_r=1}^{N_r} \sum_{n_t=1}^{N_t} \sum_{\ell=1}^{L} \left| \hat{h}_{n_r,n_t,\ell}^{(n_{\mathrm{sim}})} - h_{n_r,n_t,\ell}^{(n_{\mathrm{sim}})} \right|^2 . \tag{6.1}$$

In the equalization problems, the coded Bit Error Rate (BER), is defined as the bit error rate of the uncoded message in a coded system:

$$\mathrm{BER} = \frac{1}{N_{\mathrm{sim}} N_{\mathrm{bits}}} \sum_{n_{\mathrm{sim}}=1}^{N_{\mathrm{sim}}} \sum_{n_{\mathrm{bits}}=1}^{N_{\mathrm{bits}}} \left| \hat{m}_{n_{\mathrm{bits}}}^{(n_{\mathrm{sim}})} - m_{n_{\mathrm{bits}}}^{(n_{\mathrm{sim}})} \right| , \tag{6.2}$$

where $N_{\mathrm{bits}} = N_t MN * R_{\mathrm{cc,cp}}$ is the number of bits sent in each channel coherence period, and $m_{n_{\mathrm{bits}}}^{(n_{\mathrm{sim}})}$ is the bit sent in position $n_{\mathrm{bits}} \in \{1, \ldots, N_{\mathrm{bits}}\}$ and channel realization $n_{\mathrm{sim}} \in \{1, \ldots, N_{\mathrm{sim}}\}$. The estimate of this bit at the receiver is $\hat{m}_{n_{\mathrm{bits}}}^{(n_{\mathrm{sim}})}$.

The bit energy to noise spectral density E_b/N_0 throughout this chapter is defined at the transmitter:

$$\frac{E_b}{N_0} = \frac{P_t}{N_t \sigma_w^2} \frac{1}{B R_{\mathrm{cc,ncp}}}, \tag{6.3}$$

where B is the number of bits per constellation symbol, and σ_w^2 is the average noise variance. Note that the channel variance σ_h^2 is always set to 1, in order to make the comparisons fair.

Most of the simulations share a similar setting, to make comparisons easier. The default parameters are given in Table 6.1. All experiments use these parameters unless otherwise stated. Note that, with the default setting, the number of data bits sent is $N_{\text{bits}} = N_{\text{sim}} B N_t M N R_{\text{cc,cp}} \approx 2.5 \cdot 10^7$.

Table 6.1: Common system parameters

Parameter	Symbol	Value
Number of TX antennas	N_t	2
Number of RX antennas	N_r	10
Modulation		4-QAM
Channel variance	σ_h^2	1
Noise variance	σ_w^2	1
Number of pilot blocks	T	4
Block size (CP)	N	32
Block size (NCP)	N_{ncp}	36
CC rate (CP)	$R_{\text{cc,cp}}$	$3/4$
CC rate (NCP)	$R_{\text{cc,cp}}$	$2/3$
Number of pilot blocks	T	4
Channel coherence time	M	64
Number of channel realizations	N_{sim}	4096

The simulations compare the following methods: Expectation Maximization (EM, Section 4.1), Generalized Approximate Message Passing (GAMP, Section 4.2), and the linear Bussgang estimator (Section 4.3), and the result obtained if quantization is completely ignored, and a standard MMSE estimator is applied (denoted as *Ignoring*) in the plot legends.

6.1 Channel Estimation Performance

The first experiment aimed at comparing the performance of the individual quantized estimation methods when applied to the channel estimation problem. For this purpose, a system with the parameters given in Table 6.1 was simulated, and the channel estimation NMSE (6.1) was plotted against E_b/N_0 for the different methods. The Cramér-Rao Lower Bound from Section 3.1 was also plotted as a benchmark.

Figure 6.1 compares a multi-carrier (OFDM, blue solid curves) system with a cyclic prefix single-carrier (SC, green dashed curves) one. We observe that, in NMSE terms, both OFDM and SC exhibit a similar performance. A typical characteristic of quantized systems is the performance saturation at a certain finite SNR, which can be seen in the graph. The nonlinear methods (EM and GAMP, represented by square and inverted triangle markers respectively) saturate at a considerable lower value of NMSE, and therefore have an unavoidable benefit with respect to the linear approach. The use of Bussgang Theorem

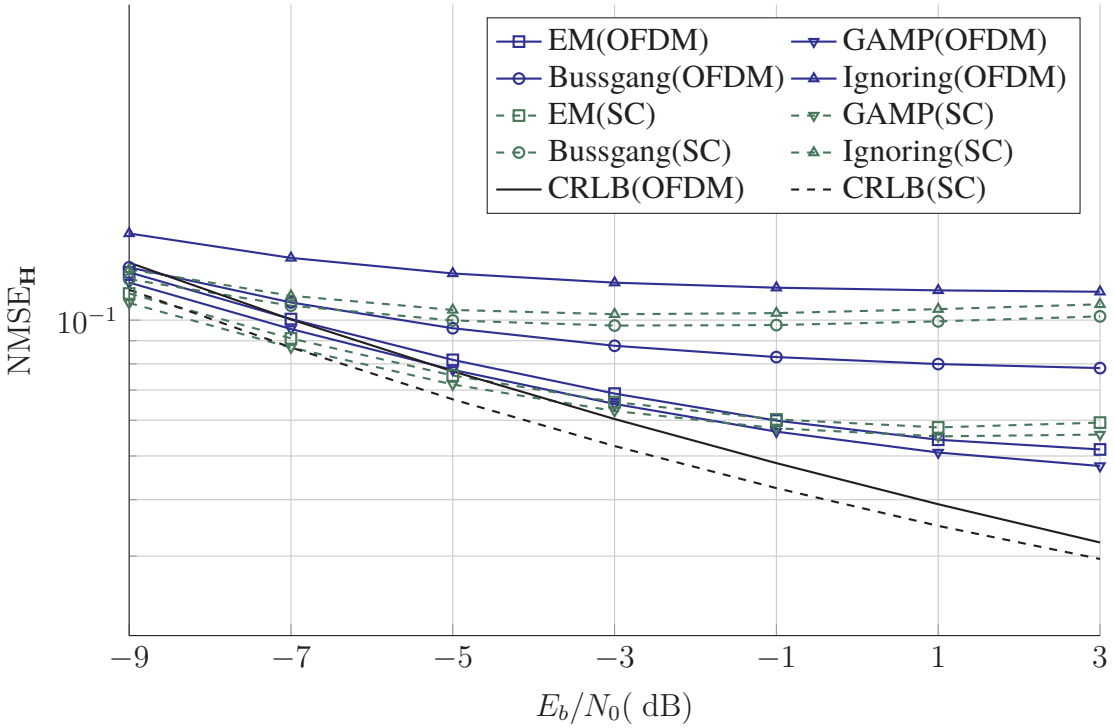

Figure 6.1: Comparison of channel estimation NMSE for OFDM and SC schemes (simulation parameters in Table 6.1)

for linear estimation (round markers) as opposed to the naive "Ignoring" method (triangle markers), is more beneficial in the OFDM case than in SC. This is because the transmitted OFDM block is closer to a Gaussian distribution, as required by the theorem.

Note also that, at low SNR, the Cramér-Rao Lower Bound (CRLB, derived in Section 3.1, represented by black curves without markers) is achieved by the nonlinear methods. However, no method reaches it at high SNR, which might mean that there is room for improvement in this region.

In Figure 6.2, the same SC system is compared to a single-carrier system without cyclic prefix (NCP, solid red curves). Recall that the comparison is done with equal throughput, by adding more channel code redundancy to the NCP case. The same observations as with the previous curve can be made, and additionally we can see that NCP achieves better performance than SC in the low SNR regime, but becomes worse at high SNR. This is a behavior that will be present in all NCP simulations.

6.2 Equalization Performance

In this section, experiments to compare the performance of the presented methods for equalization were carried out. For this purpose, perfect Channel State Information (CSI) was assumed, and the equalization methods were compared in terms of BER.

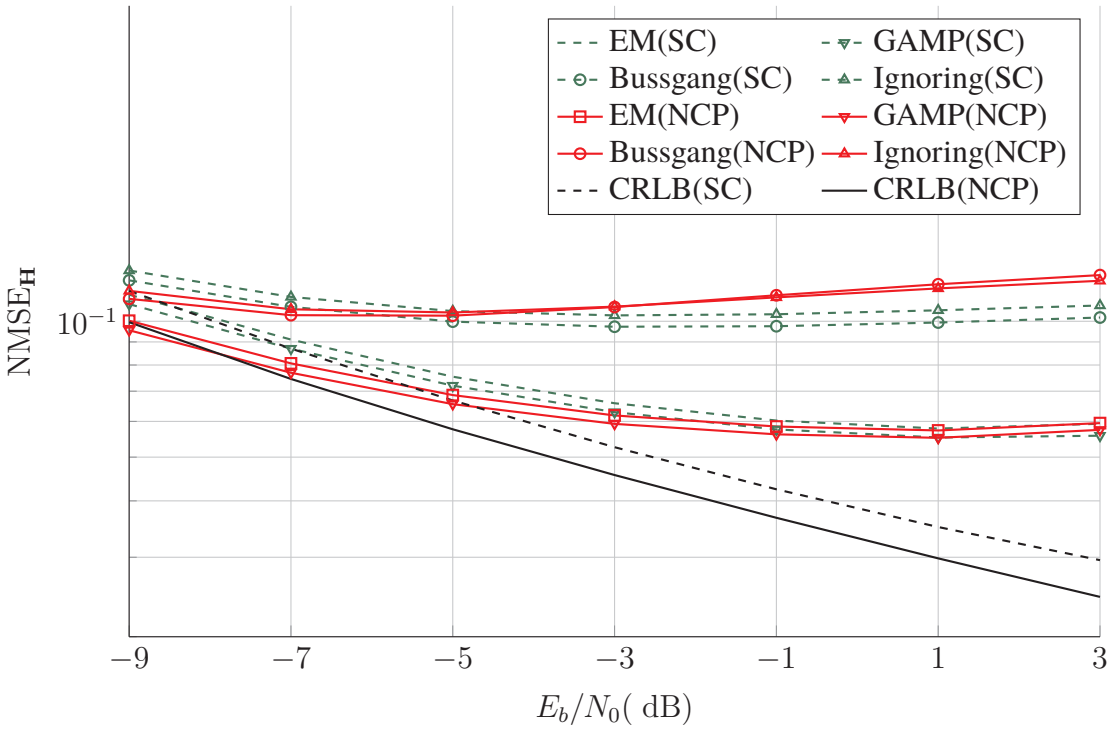

Figure 6.2: Comparison of channel estimation NMSE for SC and NCP schemes (simulation parameters in Table 6.1)

Table 6.2: System parameters for the APP SISO comparison (Figure 6.3)

Parameter	**Symbol**	**Value**
Number of TX antennas	N_t	2
Number of RX antennas	N_r	10
Modulation		4-QAM
Channel variance	σ_h^2	1
Noise variance	σ_w^2	1
Block size (CP)	N	32
CC rate (CP)	$R_{\text{cc,cp}}$	$3/4$
Number of pilot blocks	T	2
Channel coherence time	M	32
Number of channel realizations	N_{sim}	4096

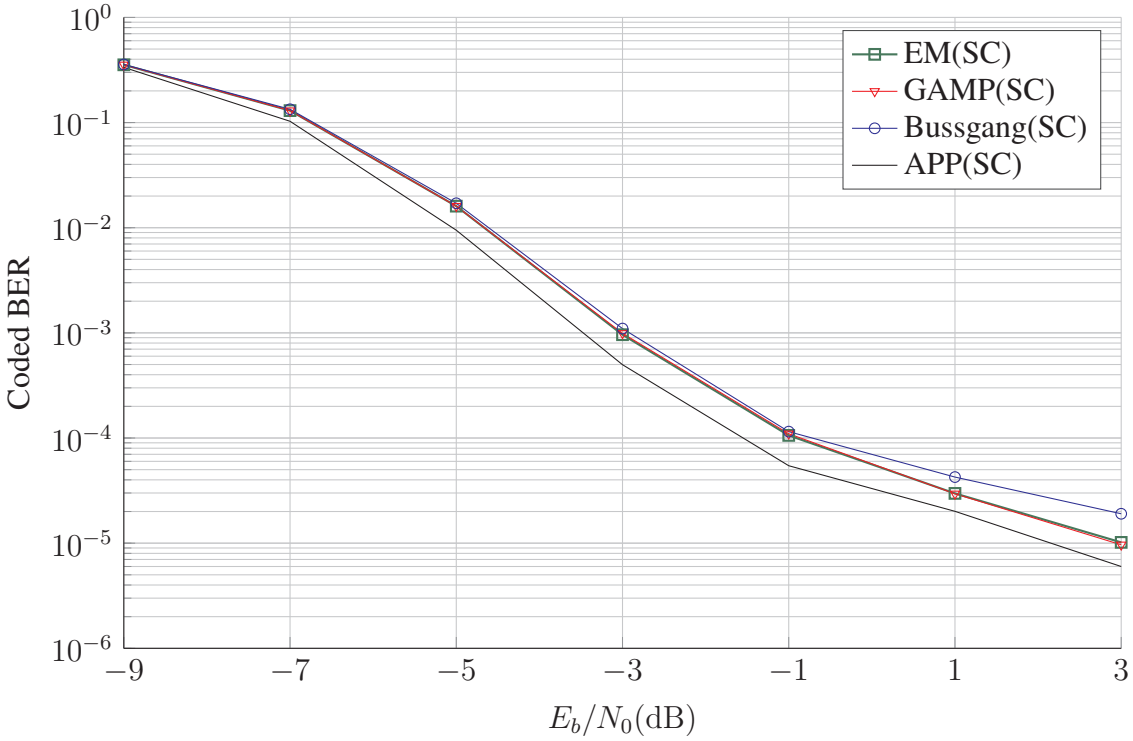

Figure 6.3: Comparison of equalization methods to optimal APP decoding for SC schemes and perfect CSI. Simulation parameters given in Table 6.2

6.2.1 Comparison to optimal APP decoding

First, the performance gap between the practical methods and the optimal A-posteriori Probability Soft-Input-Soft-Output (APP SISO) decoder from Section 4.4 was evaluated. Due to the prohibitive complexity of APP SISO, the simulations were carried out in a simplified setting, given in Table 6.2. Only the Single-Carrier (SC) case was tested, as the algorithm is not useful for OFDM. The channel length was reduced to $L = 2$, and the channel coherence time to $M = 32$. This makes $BLN_t = 4$, which keeps runtime at acceptable levels.

Figure 4.3 shows the coded BER results. In our simplified scenario, we observe that the performance of linear and nonlinear methods is similar, and the use on nonlinear algorithms would not be justified (this will not be the case in the standard scenarios). A gap of 0.5 dB is observed between the practical methods and the APP SISO block (black curve without markers). The practical methods are therefore close enough to optimal performance, at least in this simplified setting.

6.2.2 Comparison of practical algorithms

Once we have established that the practical equalization methods are good enough, we present now a rigorous comparison of their performance in coded BER terms, as was done for channel estimation. Again, we assume perfect CSI and use the parameters of our standard scenario (Table 6.1).

The results for SC vs OFDM are plotted in Figure 6.4. Some observations here are

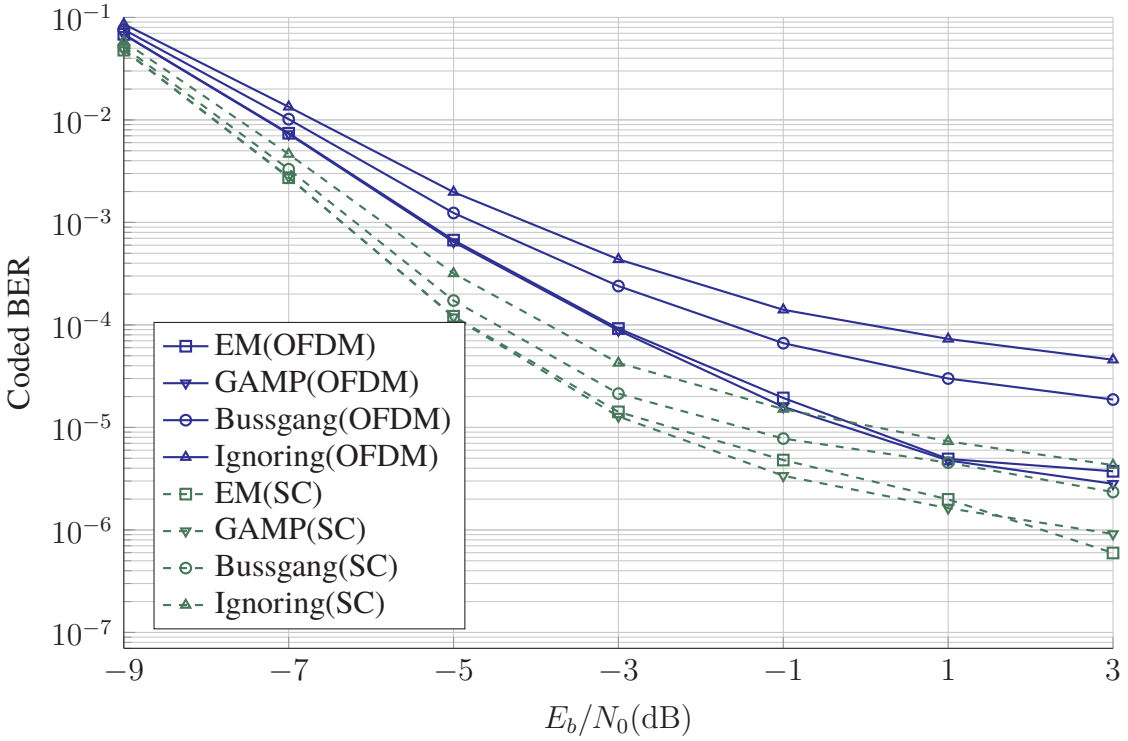

Figure 6.4: Comparison of equalization techniques with perfect CSI (OFDM vs SC). Simulation parameters given in Table 6.1

common to what we could see in the channel estimation test. The nonlinear algorithms EM (square markers) and GAMP (inverted triangle makers) obtain considerably better performance than the linear methods at a higher computational complexity.

Within the nonlinear methods, EM and GAMP have similar performance. In general, GAMP can obtain a slightly better performance due to the fact that it takes the prior into account. However, GAMP also has divergence issues more frequently, which forces the use of strong damping. For this reason, many more iterations are required for convergence than in the EM case, making GAMP the slowest method.

Within the linear methods, again the Bussgang theorem (round markers) brings a higher benefit in the OFDM case than in the SC case, because the input to the quantizer has a distribution closer to Gaussian.

However, in this equalization stage we can observe that the single-carrier system (green dashed curves) performs considerably better than the OFDM one (blue solid lines). The gap is about 2 dB at $\mathrm{BER} = 10^{-4}$). This might be because the matrix $\mathbf{A}$ is sparse in the single-carrier case, and therefore each symbol suffers interference only from LN_t other symbols, instead of NN_t. For the linear estimators, the reduced complexity due to subcarrier-wise equalization and the possibility to multiplex channels in the frequency domain might still make OFDM schemes attractive.

The comparison between single-carrier schemes with and without cyclic prefix is given in Figure 6.5. Here, we observe that, again, the system without cyclic prefix (NCP, red solid curves) outperforms the one with cyclic prefix (SC, green dashed curves) only in the low SNR regime. For high SNR, NCP saturates faster and SC has a better performance. In the

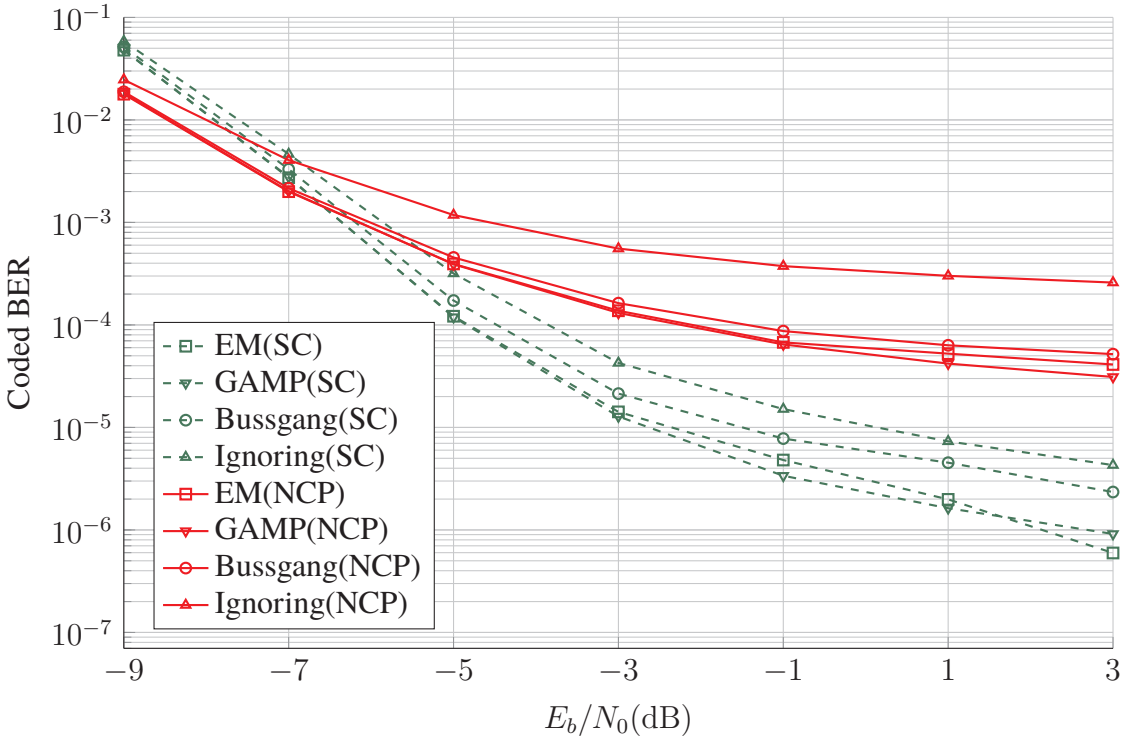

Figure 6.5: Comparison of equalization techniques with perfect CSI (SC vs NCP). Simulation parameters given in Table 6.1

equalization problem, this can be explained with the following argument. The NCP system has lower BER at low SNR due to the additional coding redundancy allowed by the lack of cyclic prefix. When the SNR increases, the system becomes limited by interference, and the interference from previous blocks saturates the NCP system earlier than the SC one.

Note that, for this scenario, the crossing point of the SC and NCP curves is at a very low SNR and high BER, making the NCP system unattractive. However, the position of the crossing point depends on the scenario parameters, so the NCP system might still be attractive in a different simulation setting.

Another effect that can be appreciated in Figure 6.5 is that the difference between linear and nonlinear methods becomes very small in the NCP system. For this kind of scheme, the increased complexity of the nonlinear methods is not worth the marginal improvement in performance. Again, the interference might be an explanation for this. In the NCP case, the symbols of one block are estimated by the corresponding algorithm, and then their estimates are input to the algorithms as "known" interference in the next block. The error incurred by doing this will translate linearly to the output if a linear method is used. However, the same error will degrade performance in an unpredictable and probably more harmful way if the method is not linear.

6.3 Full system with sequential estimation

In the next experiment a full system with sequential estimation was simulated. This means that the receiver uses one of the algorithms for channel estimation, and then the same

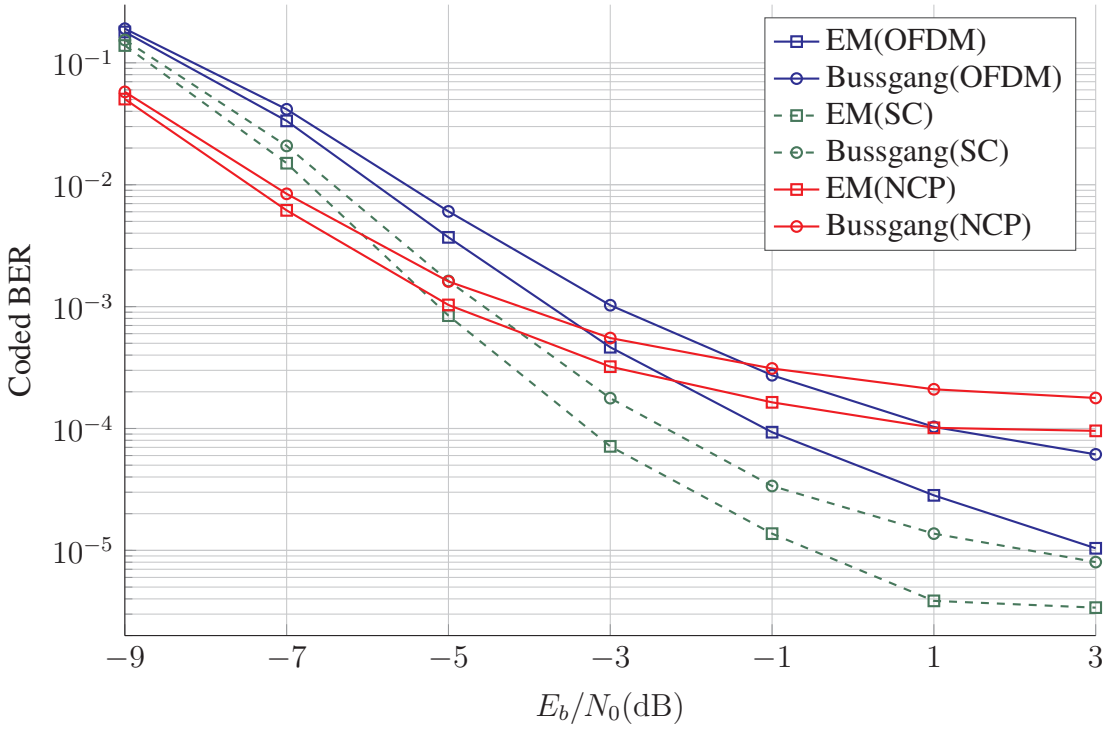

Figure 6.6: Full system with sequential estimation (same method for channel estimation and equalization). Simulation parameters given in Table 6.1

algorithm for equalization with the obtained channel estimate. The simulation parameters are once again those from Table 6.1.

The results in Figure 6.6 show the same trend as usual. Single carrier with cyclic prefix (SC, green dashed curves) outperforms OFDM (blue solid curves), and the system without cyclic prefix (NCP, red solid curve) has the best performance at low SNR but saturates earlier and is the worst option at high SNR. Again, the difference between the linear and nonlinear methods is smaller in the NCP case, due to the inter-block interference.

6.4 Number of pilots

The next experiment aimed at evaluating the required number of pilot blocks (a pilot block has N pilot symbols) to obtain an accurate channel estimate. In the single-carrier with cyclic prefix (SC) setting, the E_b/N_0 was fixed at -3 dB. A system with EM equalization and perfect CSI was compared to a system which uses EM in both channel estimation and equalization. The same was done for a system that uses Bussgang instead of EM.

The curves for BER vs number of pilot blocks T of these four settings are plotted in Figure 6.7. Of course, in the perfect CSI case (lines without markers), there are no pilots and the performance is just a horizontal line. The imperfect CSI cases (lines with markers) tend to their corresponding perfect CSI lines as the number of pilot blocks increases. It is seen in the graph that the imperfect CSI performance saturates with about 6 to 8 pilot blocks. The advantage of the nonlinear method EM (square marker, dashed green line) with respect to the linear Bussgang is clear, as it allows to reduce the number of pilots by

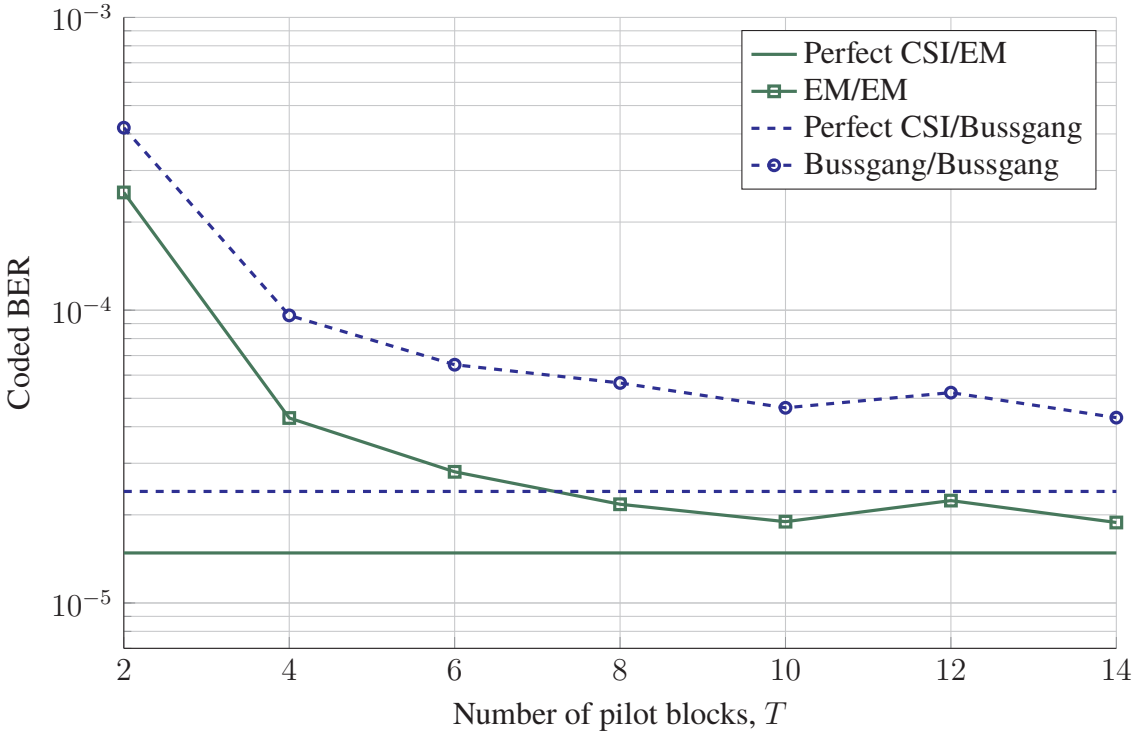

Figure 6.7: BER vs number of pilot blocks of size N required (SC system). The E_b/N_0 is set to -3 dB, the other simulation parameters are given in Table 6.1

$25 - 50\%$ for the same performance.

6.5 Higher order modulation

6.5.1 Constellation shape

In this experiment, a system with higher order modulation was tested. In order to use high order modulation in a 1-bit quantized system, it is not enough to have very high SNR. This is because the performance saturates at a certain finite SNR, as seen in the previous experiments. For high order modulation, the number of receive antennas N_r needs to be increased. This provides more measurements at the receiver, each of which narrows down the probable location of the transmitted data symbols.

The aim of this experiment was to decide on a constellation design for higher order modulation. For this purpose, PSK and QAM constellations were compared. The graphs in Figure 6.8 correspond to 3 bits per symbol (8-PSK vs 8-QAM). The simulation setting is the one given in Table 6.1, except now the number of receive antennas is $N_r = 24$.

Channel estimation and equalization are done with the same method (EM for the solid green curves and Bussgang for the dashed blue curves). The round markers correspond to PSK, and the square markers to QAM. For the case with 3 bits per symbol, it is seen that 8-PSK outperforms 8-QAM, especially in the linear case (about 5 dB performance gap at $\mathrm{BER} = 10^{-4}$ in the linear case, and only 1 dB in the nonlinear case). This is because 1-bit quantization is more harmful to the information contained in the amplitude than to that

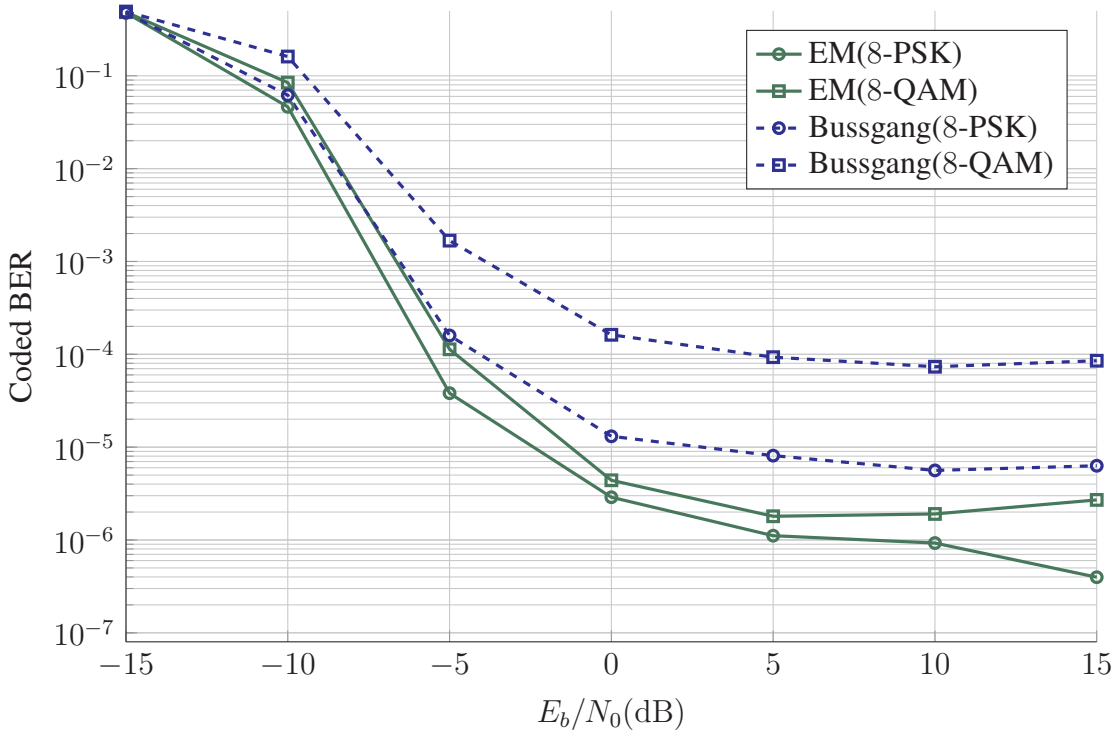

Figure 6.8: Comparison between 8-PSK and 8-QAM for full EM and Bussgang SC systems. The number of receive antennas is $N_r = 24$; all other simulation parameters are given by Table 6.1.

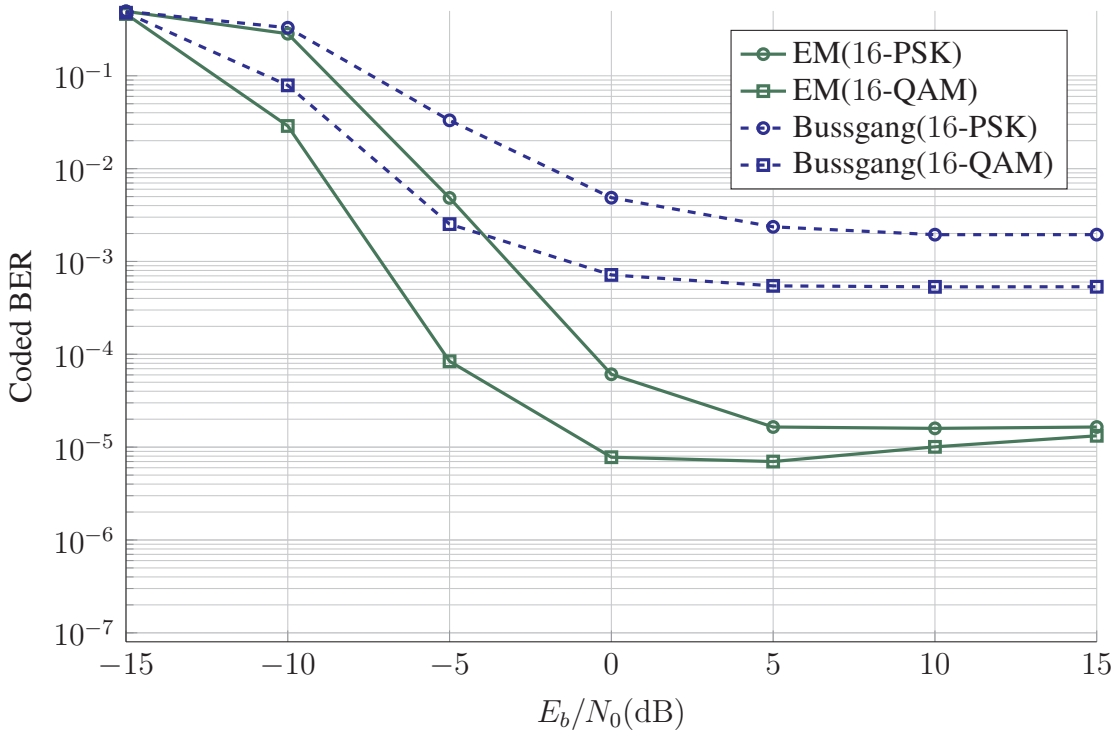

Figure 6.9: Comparison between 16-PSK and 16-QAM for full EM and Bussgang SC systems. The number of receive antennas is $N_r = 36$; all other simulation parameters are given by Table 6.1.

contained in the phase. Only the dithering effect can be used to distinguish between two amplitudes.

However, if the number of bits per symbol increases to 4 (and the number of receive antennas to $N_r = 36$, we can see in Figure 6.9 that the positions are reversed. The symbols in 16-PSK are too close to each other, and 16-QAM now has better performance. For even higher modulation orders, QAM is still the best option.

Another important conclusion from these simulations is that the difference between the linear and nonlinear methods increases considerably for higher order modulations. For 16-QAM, for example, the BER at the saturation point of EM is about 100 times lower than that of Bussgang. For 8-QAM, it is only 50 times lower, and for 4-QAM, only 2 (see Figure 6.6). The effect of model mismatch is more severe when the detection requirements become stricter.

6.5.2 Required number of antennas

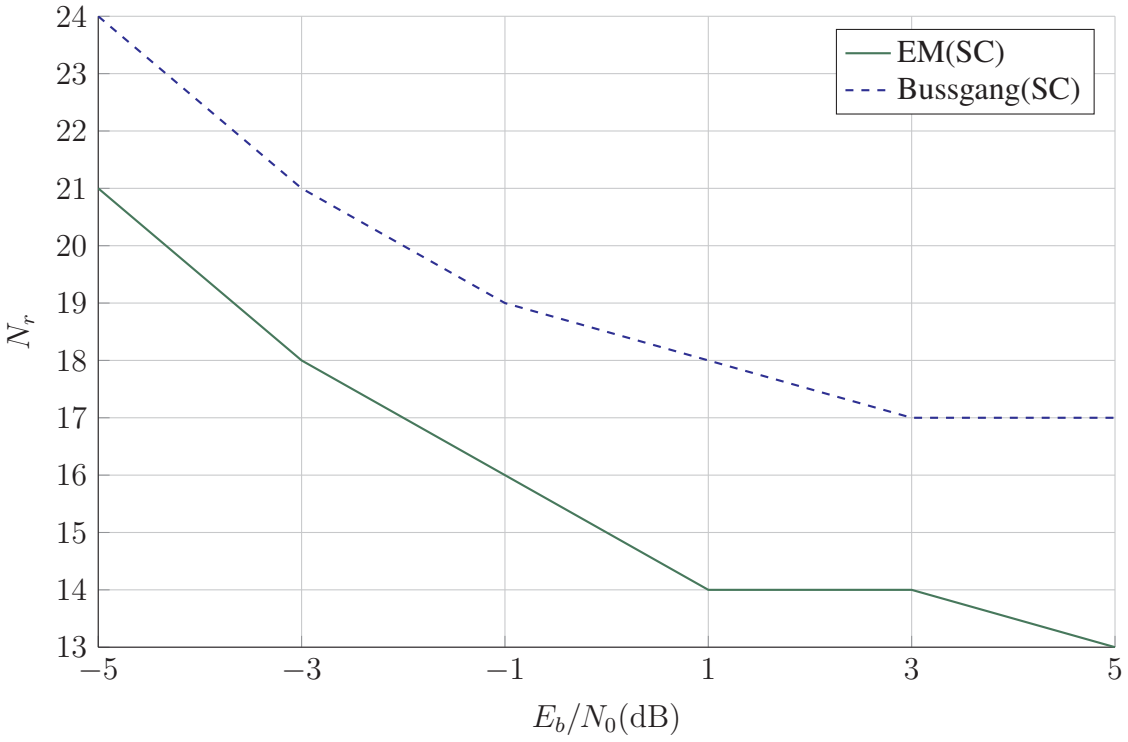

Figure 6.10: Required number of receive antennas N_r to achieve $\mathrm{BER} = 10^{-4}$ in an SC system with 8-PSK modulation. Simulation parameters given in Table 6.1

In this experiment, the gain in number of antennas obtained by using nonlinear algorithms was evaluated. A system using single carrier with cyclic prefix was simulated (except for N_r, all the other simulation parameters are given in Table 6.1). For different values of E_b/N_0, the number of receive antennas N_r required to achieve $\mathrm{BER} = 10^{-4}$ with $N_t = 2$ was obtained through simulations. The results are plotted in Figure 6.10 for 8-PSK modulation, and in Figure 6.11 for 16-QAM. It is seen that the nonlinear method (EM) allows a reduction of the number of receive antennas of about 15% in the 8-PSK case,

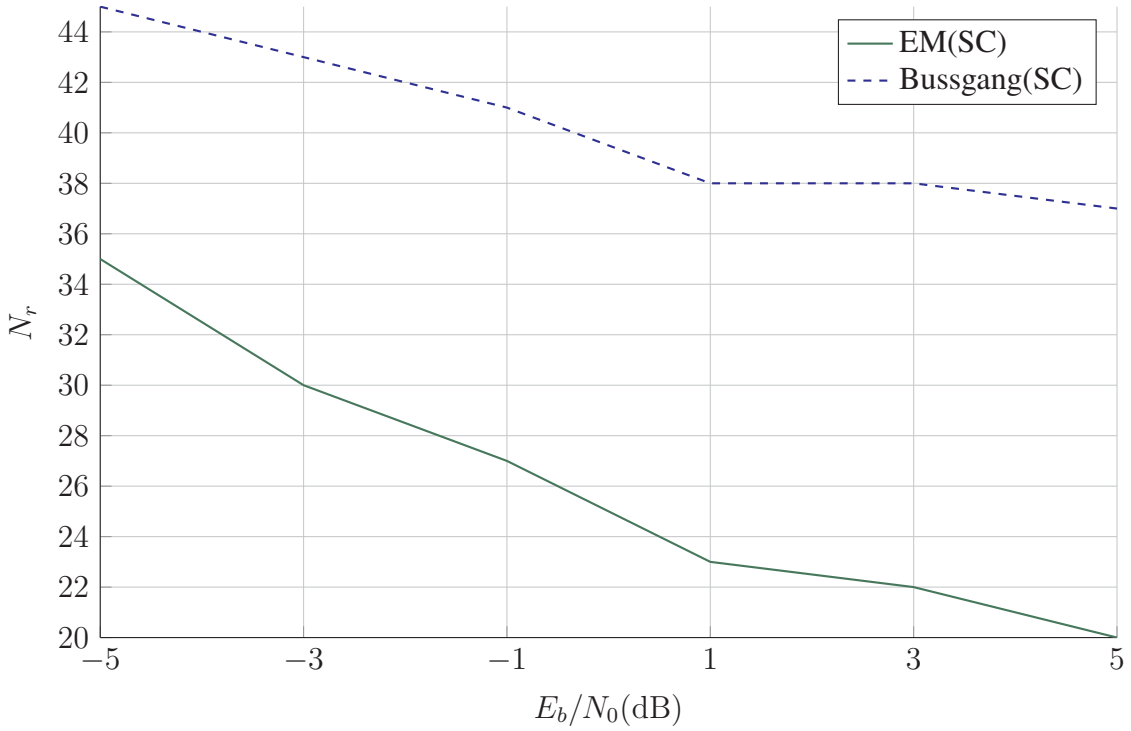

Figure 6.11: Required number of receive antennas N_r to achieve $\mathrm{BER} = 10^{-4}$ in an SC system with 16-QAM modulation. Simulation parameters given in Table 6.1

and a of as much as 50% for 16-QAM. with respect to the linear Bussgang method. This reduction results in less power consumption and space constraints, and might be worth the additional complexity, especially when the modulation order is 16 or more.

6.6 Joint Channel and Data Estimation (JCD)

The last experiment evaluates the benefit of the turbo-like Joint Channel and Data Estimation (JCD) scheme explained in Chapter 5. A system with the parameters given in Table 6.1 is simulated, and sequential estimation and equalization is compared to the joint scheme.

Figure 6.12 gives the channel estimation NMSE result for OFDM and SC, and Figure 6.13 gives the data BER result. The solid curves, which correspond to sequential estimation, are the same as those in Figure 6.6. The dashed curves correspond to the turbo-like JCD scheme. In the NMSE curves, we see that JCD greatly improves the channel estimation accuracy, by a factor of about 4.

It can be observed in Figure 6.13 that the JCD systems consistently achieve a 3 times lower BER than the equivalent sequential ones. Furthermore, the turbo-like approach in JCD converges usually in about 3 iterations, which makes it attractive given the important gain it brings. The results for the JCD system without cyclic prefix (NCP), given in the red curves with unfilled markers of Figure 6.14, also exhibit a similar gain.

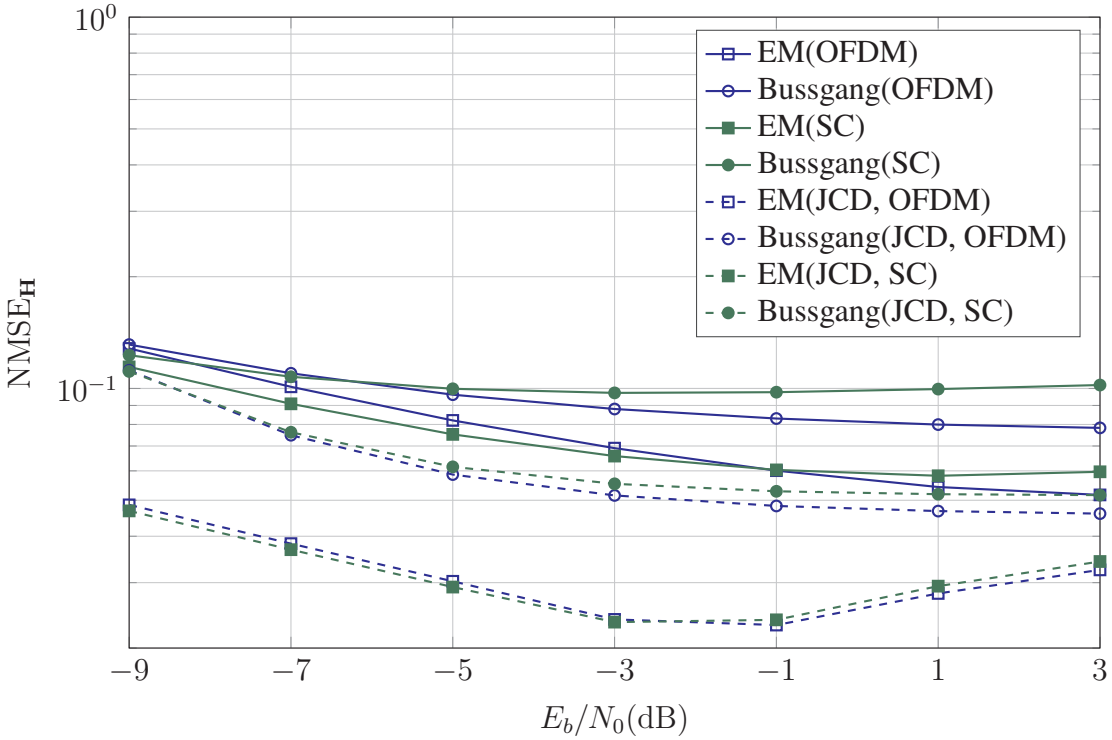

Figure 6.12: NMSE comparison of sequential estimation vs JCD, in the OFDM and SC cases. Simulation parameters given in Table 6.1

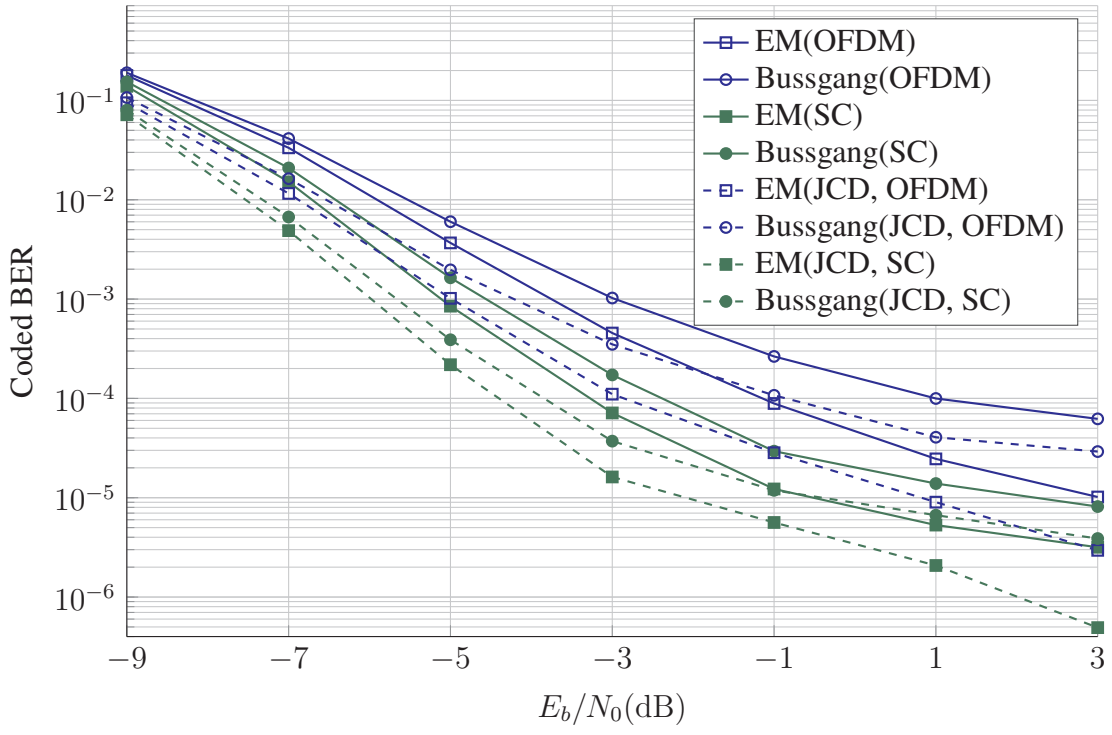

Figure 6.13: BER comparison of sequential estimation vs JCD, in the OFDM and SC cases. Simulation parameters given in Table 6.1

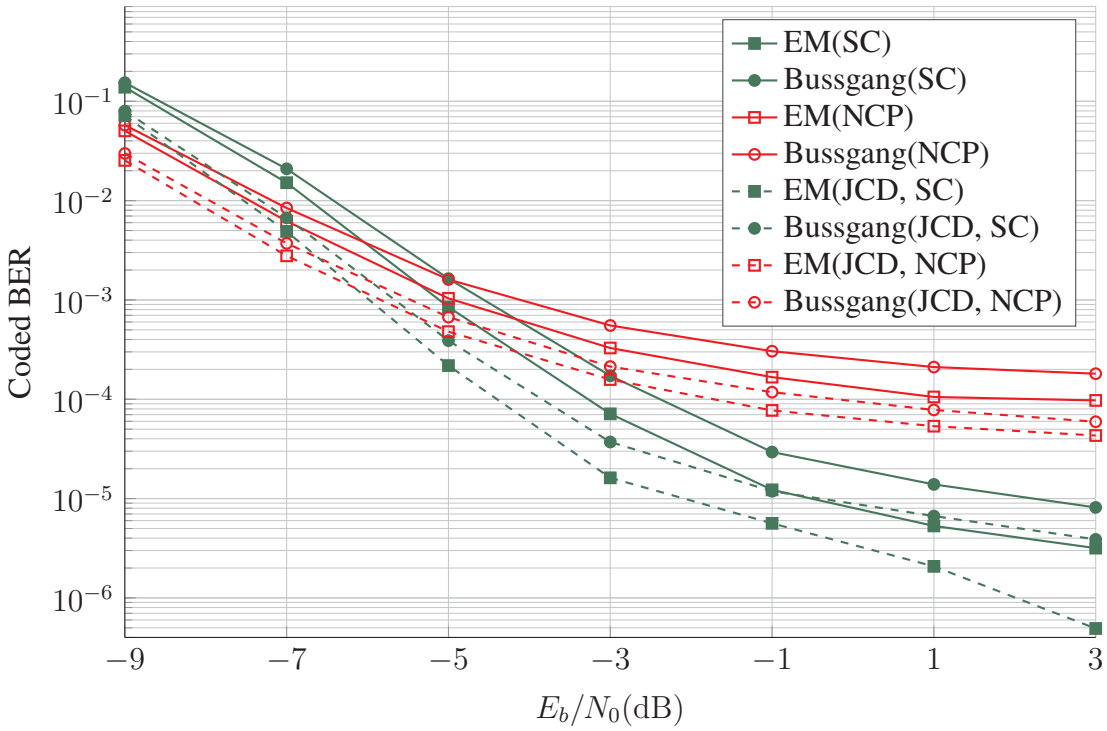

Figure 6.14: BER comparison of sequential estimation vs JCD, in the SC and NCP cases. Simulation parameters given in Table 6.1

Chapter 7

Conclusion

In this thesis, the problem of channel estimation and equalization in 1-bit quantized Multiple Input-Multiple Output (MIMO) frequency-selective channels was addressed. The topic was chosen because most of the literature on quantized MIMO focuses on flat-fading channels, and very few works are available that deal with multi-carrier systems.

Quantized matrix-vector system models were derived for both the channel estimation and equalization problems in three different kinds of systems: Orthogonal Frequency Division Multiplexing (OFDM), single-carrier with cyclic prefix (SC), and single-carrier without cyclic prefix (NCP).

A Cramér-Rao Lower Bound (CRLB) was derived for the estimation problems. Two non-linear quantized estimation algorithms: Expectation Maximization (EM) and Generalized Approximate Message Passing (GAMP), as well as a linear method based on the Bussgang theorem, were adapted to the derived system model. The nonlinear methods are iterative and have higher computational complexity, but perform considerably better. An optimal A-posteriori Probability Soft-Input-Soft-Output equalizer was also derived. It has prohibitive complexity but is good as a performance benchmark.

A turbo-like Joint Channel and Data estimation scheme (JCD) was also adapted to the problem and tested.

The simulation results show that single-carrier systems with cyclic prefix always outperform OFDM ones. The systems without cyclic prefix have a worse saturation point at high SNR, but are the best ones at low SNR. The nonlinear algorithms turn out to give considerable gains with respect to the linear estimator, and allow reduction of the number of pilot blocks or of receive antennas.

The results show that 8-PSK is more suited to 1-bit quantized systems than 8-QAM, but for modulation orders of 16 and above, QAM is again the best constellation type. Finally, the turbo-like approach (JCD) was shown to greatly improve the estimation and detection accuracy, at the cost of about 3 to 5 times slower execution.

Many topics are open for future work in this area. The application of a Bayesian JCD estimator, the use of Filter Bank Multi-Carrier (FBMC) schemes, and the design of transmission and coding strategies are just a few of them.

Appendix A

Matrix Operators

In this Appendix, we define some of the less common matrix operators that have been used throughout the thesis.

A.1 Vectorization

The vectorization operator, $\mathrm{vec}\,(\cdot)$, trasforms a matrix into a vector by stacking its columns vertically. Given a matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$:

$$\mathrm{vec}\,(\mathbf{A}) = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_N \end{pmatrix} \in \mathbb{C}^{MN}, \tag{A.1}$$

where $\mathbf{a}_n, n \in \{1, \ldots, N\}$ denotes the n-th column of $\mathbf{A}$.

A.2 Khatri-Rao product

The Khatri-Rao product of two matrices with the same number of columns, $\mathbf{A} \in \mathbb{C}^{M\times N}$ and $\mathbf{B} \in \mathbb{C}^{P\times N}$ is a new matrix $\mathbf{A} \diamond \mathbf{B} \in \mathbb{C}^{MP\times N}$, with a definition that can be interpreted as a column-wise Kronecker product:

$$\mathbf{A} \diamond \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{b}_1 & a_{12}\mathbf{b}_2 & \cdots & a_{1N}\mathbf{b}_N \\ a_{21}\mathbf{b}_1 & a_{22}\mathbf{b}_2 & \cdots & a_{2N}\mathbf{b}_N \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}\mathbf{b}_2 & a_{M2}\mathbf{b}_2 & \cdots & a_{MN}\mathbf{b}_N \end{pmatrix}, \tag{A.2}$$

where $\mathbf{b}_n$ denotes the n-th column of $\mathbf{B}$.

Appendix B

Derivation of the Fisher Information Matrix for Joint Channel and Data Estimation

In this appendix, the entries if the Fisher information matrix for Joint Channel and Data Estimation (JCD), given in (3.35)-(3.37) are derived in more detail. We start with the log likelihood $P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\theta}\right) = \ln p_{\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{X}}_d}\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{X}}_d\right)$, obtainable from (3.27):

$$P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{X}}_d\right) = \sum_{n_r=1}^{N_r} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[\ln \Phi\left(\frac{y_{\Re,n_r,m,n} z_{\Re,n_r,m,n}}{\sigma_{n_r}/\sqrt{2}} \right) + \ln \Phi\left(\frac{y_{\Im,n_r,m,n} z_{\Im,n_r,m,n}}{\sigma_{n_r}/\sqrt{2}} \right) \right], \quad \text{(B.1)}$$

where

$$z_{\Re,n_r,m,n} = \sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \left(h_{\Re,n_r,n_t,\ell} x_{\Re,n_t,m,(n-\ell) \bmod N} - h_{\Im,n_r,n_t,\ell} x_{\Im,n_t,m,(n-\ell) \bmod N} \right) \quad \text{(B.2)}$$

and

$$z_{\Im,n_r,m,n} = \sum_{n_t=1}^{N_t} \sum_{\ell=0}^{L-1} \left(h_{\Re,n_r,n_t,\ell} x_{\Im,n_t,m,(n-\ell) \bmod N} + h_{\Im,n_r,n_t,\ell} x_{\Re,n_t,m,(n-\ell) \bmod N} \right). \quad \text{(B.3)}$$

Using the identities:

$$\frac{\mathrm{d}\Phi\left(x\right)}{\mathrm{d}x} = \phi\left(x\right); \qquad \frac{\mathrm{d}\phi\left(x\right)}{\mathrm{d}x} = -x\phi\left(x\right), \quad \text{(B.4)}$$

we can obtain the first derivatives of $P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{X}}_d\right)$ with respect to each channel tap and each data symbol. Let us define:

$$\eta_{\mathcal{P},n_r,m,n} = \frac{\sqrt{2}}{\sigma_{n_r}} y_{\mathcal{P},n_r,m,n} z_{\mathcal{P},n_r,m,n}, \quad \text{(B.5)}$$

where $\mathcal{P}$ can mean either $\Re$ or $\Im$. Then,the first derivatives of (B.1) yield:

$$\frac{\partial P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{X}}_d\right)}{\partial h_{\mathcal{P},n_r,n_t,\ell}} = \frac{\sqrt{2}}{\sigma_{n_r}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[\frac{\phi\left(\eta_{\Re,n_r,m,n}\right)}{\Phi\left(\eta_{\Re,n_r,m,n}\right)} y_{\Re,n_r,m,n} s_{\mathcal{P}} x_{\mathcal{P},n_t,m,(n-\ell) \bmod N} + \frac{\phi\left(\eta_{\Im,n_r,m,n}\right)}{\Phi\left(\eta_{\Im,n_r,m,n}\right)} y_{\Im,n_r,m,n} x_{\overline{\mathcal{P}},n_t,m,(n-\ell) \bmod N} \right], \quad \text{(B.6)}$$

$$\frac{\partial P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{X}}_d\right)}{\partial x_{\mathcal{P},n_t,m,n}} = \sum_{n_r=1}^{N_r} \frac{\sqrt{2}}{\sigma_{n_r}} \sum_{n''=n}^{(n+L-1) \bmod N} \left[\frac{\phi\left(\eta_{\Re,n_r,m,n''}\right)}{\Phi\left(\eta_{\Re,n_r,m,n''}\right)} y_{\Re,n_r,m,n''} s_{\mathcal{P}} h_{\mathcal{P},n_r,n_t,(n''-n) \bmod N} + \frac{\phi\left(\eta_{\Im,n_r,m,n''}\right)}{\Phi\left(\eta_{\Im,n_r,m,n''}\right)} y_{\Im,n_r,m,n''} h_{\overline{\mathcal{P}},n_r,n_t,(n''-n) \bmod N} \right]. \quad \text{(B.7)}$$

Now, let us define:

$$\psi\left(\eta\right) = -\frac{\partial}{\partial \eta} \frac{\phi\left(\eta\right)}{\Phi\left(\eta\right)} = \frac{\eta \phi\left(\eta\right)}{\Phi\left(\eta\right)} + \left(\frac{\phi\left(\eta\right)}{\Phi\left(\eta\right)} \right)^2 \quad \text{(B.8)}$$

Then, we have the following identities:

$$-\frac{\partial}{\partial x} \frac{\phi\left(\eta\right)}{\Phi\left(\eta\right)} = \psi(\eta) \frac{\partial \eta}{\partial x}, \quad \text{(B.9)}$$

$$-\frac{\partial}{\partial x} \frac{\phi\left(\eta\right)}{\Phi\left(\eta\right)} x = \psi(\eta) \frac{\partial \eta}{\partial x} - \frac{\phi\left(\eta\right)}{\Phi\left(\eta\right)}. \quad \text{(B.10)}$$

The negative second derivatives, from (B.6) and (B.7), are then given by:

$$-\frac{\partial^2}{\partial h_{\mathcal{P},n_r,n_t,\ell} \partial h_{\mathcal{T},n_r',n_t',\ell'}} = \delta[n_r - n_r'] \frac{2}{\sigma_{n_r}^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[\psi\left(\eta_{\Re,n_r,m,n}\right) s_{\mathcal{P}} s_{\mathcal{T}} x_{\mathcal{P},n_t,m,(n-\ell) \bmod N} x_{\mathcal{T},n_t',m,(n-\ell') \bmod N} + \psi\left(\eta_{\Im,n_r,m,n}\right) x_{\overline{\mathcal{P}},n_t,m,(n-\ell) \bmod N} x_{\overline{\mathcal{T}},n_t',m,(n-\ell') \bmod N} \right], \quad \text{(B.11)}$$

$$-\frac{\partial^2 P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{X}}_d\right)}{\partial x_{\mathcal{P},n_t,m,n} \partial x_{\mathcal{T},n_t',m',n'}} = \delta[m - m'] \sum_{n_r=1}^{N_r} \frac{2}{\sigma_{n_r}^2} \sum_{n''=\max\{n,n'\}}^{\min\{n+L-1,n'+L-1\} \bmod N} \left[\psi\left(\eta_{\Re,n_r,m,n''}\right) s_{\mathcal{P}} s_{\mathcal{T}} h_{\mathcal{P},n_r,n_t,(n''-n) \bmod N} h_{\mathcal{T},n_r,n_t',(n''-n') \bmod N} + \psi\left(\eta_{\Im,n_r,m,n''}\right) h_{\overline{\mathcal{P}},n_r,n_t,(n''-n) \bmod N} h_{\overline{\mathcal{T}},n_r,n_t',(n''-n') \bmod N} \right]. \quad \text{(B.12)}$$

$$
\begin{aligned}
-\frac{\partial^2 P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{X}}_d\right)}{\partial h_{\mathcal{P},n_r,n_t,\ell} \partial x_{\mathcal{T},n_t',m,n}} = \Bigg\{ & \frac{2}{\sigma_{n_r}^2} \sum_{n''=n}^{n+L-1 \bmod N} \\
& \Big[\psi\left(\eta_{\Re,n_r,m,n''}\right) s_{\mathcal{P}} s_{\mathcal{T}} x_{\mathcal{P},n_t,m,(n''-\ell) \bmod N} h_{\mathcal{T},n_r,n_t',(n''-n) \bmod N} \\
& + \psi\left(\eta_{\Im,n_r,m,n''}\right) x_{\overline{\mathcal{P}},n_t,m,(n''-\ell) \bmod N} h_{\overline{\mathcal{T}},n_r,n_t',(n''-n) \bmod N} \Big] \Bigg\} \\
& - \delta[n_t - n_t'] s_{\mathcal{P}} \frac{\sqrt{2}}{\sigma_{n_r}} \frac{\phi\left(\eta_{(\mathcal{P}\cdot\mathcal{T}),n_r,m,(n+\ell) \bmod N}\right)}{\Phi\left(\eta_{(\mathcal{P}\cdot\mathcal{T}),n_r,m,(n+\ell) \bmod N}\right)} y_{(\mathcal{P}\cdot\mathcal{T}),n_r,m,(n+\ell) \bmod N}. \quad \text{(B.13)}
\end{aligned}
$$

In (B.13), the operation $\mathcal{P} \cdot \mathcal{T}$ is defined as:

$$
\mathcal{P} \cdot \mathcal{T} = \begin{cases} \Re, & \text{if } \mathcal{P} = \mathcal{T}, \\ \Im, & \text{if } \mathcal{P} \neq \mathcal{T}. \end{cases} \quad \text{(B.14)}
$$

The last summand in (B.13) accounts for the case in which $x_{\mathcal{T},n_t',m,n} = x_{\mathcal{P},n_t,m,(n''-\ell) \bmod N}$. When this happens, the derivative of (B.6) with respect to $x_{\mathcal{T},n_t',m,n}$ involves (B.10), whose last term corresponds to the last summand in (B.13). All the other summands in (B.11)-(B.13) are obtained by using (B.9).

Now only one step remains, which consists of taking the expectation of (B.11)-(B.13) with respect to $\boldsymbol{\mathcal{Y}}$. Note that (B.11) and (B.12) only depend on $\boldsymbol{\mathcal{Y}}$ through the terms $\psi\left(\eta_{\Re,n_r,m,n}\right)$ and $\psi\left(\eta_{\Im,n_r,m,n}\right)$. Each summand in the summation depends only on one entry of $\boldsymbol{\mathcal{Y}}$, and therefore the expectation can be taken only over the corresponding marginal distribution, derived in Chapter 3 (3.27):

$$
p_{y_{\mathcal{P},n_r,m,n} \mid \boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{X}}_d}\left(y_{\mathcal{P},n_r,m,n} \mid \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{X}}_d\right) = \Phi\left(\eta_{\mathcal{P},n_r,m,n}\right). \quad \text{(B.15)}
$$

The expectation of $\psi\left(\eta_{\Re,n_r,m,n}\right)$ over $y_{\mathcal{P},n_r,m,n} \mid \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{X}}_d$ then yields:

$$
\begin{aligned}
\mathrm{E}\left\{\psi\left(\eta_{\Re,n_r,m,n}\right)\right\} &= \sum_{y \in \{-1,1\}} \Phi\left(\eta_{\mathcal{P},n_r,m,n}\right) \psi\left(\eta_{\Re,n_r,m,n}\right) \\
&= -\mu_{\mathcal{P},n_r,m,n} \phi\left(-\mu_{\mathcal{P},n_r,m,n}\right) + \frac{\left(\phi\left(-\mu_{\mathcal{P},n_r,m,n}\right)\right)^2}{\Phi\left(-\mu_{\mathcal{P},n_r,m,n}\right)} \\
&+ \mu_{\mathcal{P},n_r,m,n} \phi\left(\mu_{\mathcal{P},n_r,m,n}\right) + \frac{\left(\phi\left(\mu_{\mathcal{P},n_r,m,n}\right)\right)^2}{\Phi\left(\mu_{\mathcal{P},n_r,m,n}\right)} \\
&= \frac{\left(\phi\left(\mu_{\mathcal{P},n_r,m,n}\right)\right)^2}{\Phi\left(\mu_{\mathcal{P},n_r,m,n}\right)\left(1 - \Phi\left(\mu_{\mathcal{P},n_r,m,n}\right)\right)} \\
&= \Psi_{\mathcal{P},n_r,m,n}, \quad \text{(B.16)}
\end{aligned}
$$

where

$$
\mu_{\mathcal{P},n_r,m,n} = \frac{\sqrt{2}}{\sigma_{n_r}} z_{\mathcal{P},n_r,m,n} = \eta_{\mathcal{P},n_r,m,n}\big|_{y_{\mathcal{P},n_r,m,n}=1}. \quad \text{(B.17)}
$$

The definition of $\Psi_{\mathcal{P},n_r,m,n}$ in (B.16) is the same as in Chapter 3 (3.34). To arrive at the final results, the identities $\phi(-x) = \phi(x)$ and $\Phi(-x) = 1 - \Phi(x)$ were used. Note furthermore that:

$$
\mathrm{E}\left\{\frac{\phi\left(\eta_{(\mathcal{P}),n_r,m,n}\right)}{\Phi\left(\eta_{(\mathcal{P}),n_r,m,n}\right)} y_{(\mathcal{P}),n_r,m,n}\right\} = -\phi\left(-\mu_{(\mathcal{P}),n_r,m,n}\right) + \phi\left(\mu_{(\mathcal{P}),n_r,m,n}\right) = 0, \quad \text{(B.18)}
$$

which cancels out the last summand in (B.13). Plugging in (B.16) and (B.18) into (B.11)-(B.13) gives the final result for the entries of the Fisher Information Matrix:

$$\begin{aligned}
&\mathrm{E}\left\{-\frac{\partial^2 P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\theta}\right)}{\partial h_{\mathcal{P}n_r,n_t,\ell}\partial h_{\mathcal{T}n_r',n_t',\ell'}}\right\} \\
&\qquad = \delta\left[n_r - n_r'\right]\frac{2}{\sigma_{n_r}}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \\
&\qquad\quad \Big[s_{\mathcal{P}}s_{\mathcal{T}}x_{\mathcal{P},n_t,m,(n-\ell)\bmod N}x_{\mathcal{T},n_t',m,(n-\ell')\bmod N}\Psi_{\Re,n_r,m,n} \\
&\qquad\qquad + x_{\overline{\mathcal{P}},n_t,m,(n-\ell)\bmod N}x_{\overline{\mathcal{T}},n_t',m,(n-\ell')\bmod N}\Psi_{\Im,n_r,m,n}\Big], \qquad \text{(B.19)}
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}\left\{-\frac{\partial^2 P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\theta}\right)}{\partial x_{\mathcal{P}n_t,m,n}\partial x_{\mathcal{T}n_t',m',n'}}\right\} \\
&\qquad = \delta\left[m - m'\right]\sum_{n_r=1}^{N_r}\frac{2}{\sigma_{n_r}}\sum_{n''=\max\{n,n'\}}^{(\min\{n+L-1,n'+L-1\})\bmod N} \\
&\qquad\quad \Big[s_{\mathcal{P}}s_{\mathcal{T}}h_{\mathcal{P},n_r,n_t,(n''-n)\bmod N}h_{\mathcal{T},n_r,n_t',(n''-n')\bmod N}\Psi_{\Re,n_r,m,n''} \\
&\qquad\qquad + h_{\overline{\mathcal{P}},n_r,n_t,(n''-n)\bmod N}h_{\overline{\mathcal{T}},n_r,n_t',(n''-n')\bmod N}\Psi_{\Im,n_r,m,n''}\Big], \qquad \text{(B.20)}
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}\left\{-\frac{\partial^2 P\left(\boldsymbol{\mathcal{Y}} \mid \boldsymbol{\theta}\right)}{\partial h_{\mathcal{P}n_r,n_t,\ell}\partial x_{\mathcal{T}n_t',m,n}}\right\} \\
&\qquad = \frac{2}{\sigma_{n_r}}\sum_{n''=n}^{(n+L-1)\bmod N} \\
&\qquad\quad \Big[s_{\mathcal{P}}s_{\mathcal{T}}x_{\mathcal{P},n_t,m,(n''-\ell)\bmod N}h_{\mathcal{T},n_r,n_t',(n''-n)\bmod N}\Psi_{\Re,n_r,m,n''} \\
&\qquad\qquad + x_{\overline{\mathcal{P}},n_t,m,(n''-\ell)\bmod N}h_{\overline{\mathcal{T}},n_r,n_t',(n''-n)\bmod N}\Psi_{\Im,n_r,m,n''}\Big]. \qquad \text{(B.21)}
\end{aligned}$$

Bibliography

[1] C. X. Wang, F. Haider, X. Gao, X. H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5g wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, February 2014.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[3] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5g," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.

[4] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, Apr 1999.

[5] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized mimo systems with transmitter channel state information," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5498–5512, Oct 2015.

[6] K. Roth, J. Garcia, J. Munir, M. Faerber, and J. A. Nossek, "Channel capacity comparison of different system concepts for mmWave," in *WSA 2016; 20th International ITG Workshop on Smart Antennas*, March 2016, pp. 1–7.

[7] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath, "Channel estimation in millimeter wave mimo systems with one-bit quantization," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 957–961.

[8] A. Mezghani, M. Koufi, and J. Nossek, "A modified mmse receiver for quantized mimo systems," in *ITG/IEEE Workshop on Smart Antennas*, February 2007, vienna, Austria.

[9] A. Mezghani, M. S. Khoufi, and J. A. Nossek, "Spatial mimo decision feedback equalizer operating on quantized data," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 2893–2896.

[10] C. Wen, C. Wang, S. Jin, K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision adcs," *CoRR*, vol. abs/1507.07766, 2015. [Online]. Available: http://arxiv.org/abs/1507.07766

[11] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *CoRR*, vol. abs/1509.07928, 2015. [Online]. Available: http://arxiv.org/abs/1509.07928

[12] A. Mezghani, M. Rouatbi, and J. A. Nossek, "An iterative receiver for quantized mimo systems," in *2012 16th IEEE Mediterranean Electrotechnical Conference*, March 2012, pp. 1049–1052.

[13] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Artech House, 2000.

[14] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, 1977.

[15] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, July 2011, pp. 2168–2172.

[16] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted gaussian signals," Massachusetts Institute of Technology. Research Laboratory of Electronics, Tech Report, Mar 1952.

[17] F. Wendler, M. Stein, A. Mezghani, and J. A. Nossek, "Quantization-loss reduction for 1-bit boc positioning," in *2013 International Technical Meeting of The Institute of Navigation*, . I. T. M. of The Institute of Navigation, Ed., January 2013.

[18] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[19] F. Steiner, A. Mezghani, L. Swindlehurst, J. A. Nossek, and W. Utschick, "Turbo-like joint data-and-channel estimation in quantized massive mimo systems," in *WSA 2016; 20th International ITG Workshop on Smart Antennas*, March 2016, pp. 1–5.