DIPLOMARBEIT

OPTIMAL RECEIVER DESIGN FOR PILOT-ASSISTED COMMUNICATION SYSTEMS

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Diplomingenieurs

unter der Leitung von

Ao. Univ.-Prof. Dipl.-Ing. Dr. Gerald Matz

Projektass. Dipl.-Ing. Peter Fertl

Univ.Ass. Dipl.-Ing. Clemens Novak

Institut für Nachrichtentechnik und Hochfrequenztechnik

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik

von

Himanshu Jha
Schluckergasse 9/3/9
1130 Wien

Wien, im Juni 2008 ____________________

— To my wife Radha —

Abstract

This thesis deals with *optimal receiver* design for pilot-assisted communication systems with imperfectly available channel state information (CSI). In conventional receiver design the channel is assumed to be perfectly known. The maximum likelihood (ML) decision metric is then derived under this assumption. But acquiring perfect knowledge of the channel poses a fundamental problem for the receiver. In practice, the receiver performs channel estimation using techniques like *least squares* (LS) or *minimum mean square error* (MMSE) estimation. This is accomplished by sending pilots, which are perfectly known by the receiver. Due to the limited number of pilots, the channel estimation is imperfect. The so called *mismatched receiver* replaces the true channel by its noisy estimate in the metric originally designed for a perfectly known channel. The resulting mismatch leads to performance degradation in terms of bit error rate (BER).

In this thesis, we pursue a more advanced approach to designing a receiver by utilizing statistics of the channel and its estimation error to derive a so-called *modified* ML metric. The metric obtained by this method is better suited to the presence of channel estimation errors. This concept is applied in deriving a *modified receiver* for an iterative system architecture based on the bit-interleaved coded modulation scheme with iterative decoding (BICM-ID). Numerical simulations using an i.i.d Rayleigh block fading channel model show the superior performance of the *modified receiver* in terms of BER.

We further extend the idea of utilizing the channel statistics to correlated channel models and derive an *optimum* maximum likelihood metric for a non-iterative system architecture. The resulting *optimum receiver* performs sequence detection without prior channel estimation, because the received pilots are directly incorporated into the metric. We also provide low-complexity implementation of the *optimum* metric. Numerical simulations based on orthogonal frequency division multiplexing (OFDM) and autoregressive (AR) channel models show that the *optimum receiver* outperforms the *mismatched receiver* in terms of BER. The *optimum receiver* further is observed to be less sensitive to the number of pilots used.

Kurzfassung

Diese Diplomarbeit beschäftigt sich mit dem optimalen Empfängerdesign für pilotgestützte Kommunikationssysteme mit imperfekter Kanalzustandsinformation. Im konventionellen Empfängerdesign wird angenommen, dass der Kanal perfekt bekannt ist. Die Maximum-Likelihood (ML) Entscheidungsmetrik wird dann unter dieser Annahme hergeleitet. Ein perfektes Wissen über den Kanal zu erlangen, stellt ein fundamentales Problem für den Empfänger dar. In der Praxis, schätzt der Empfänger den Kanal mithilfe von Techniken wie die Methode der kleinsten Quadrate oder Minimierung des mittleren quadratischen Fehlers. Dies wird durch das Senden von Piloten, die dem Empfänger vollständig bekannt sind, erreicht. Wegen der limitierten Anzahl von Piloten ist die Kanalschätzung imperfekt. Der sogenannte fehlangepasste Empfänger verwendet die ursprünglich für einen perfekt bekannten Kanal entworfene Metrik und ersetzt darin den echten Kanal durch dessen verrauschten Schätzwert. Die dadurch entstandene Fehlanpassung führt zu einer Leistungsverminderung hinsichtlich der Bitfehlerrate.

In dieser Diplomarbeit verwenden wir einem fortgeschrittenen Ansatz, um einen Empfänger zu entwerfen, in dem die Statistik des Kanals und dessen Schätzfehlers ausgenützt wird, um eine sogenannte modifizierte ML Metrik herzuleiten. Die Metrik, die durch diese Methode ermittelt wird, ist besser an die Anwesenheit von Kanalschätzfehlern angepasst. Dieses Konzept wird bei der Herleitung eines modifizierten Empfängers für eine iterative Systemarchitektur, die auf Bit-ineinander Kodierte Modulation mit iterativer Decodierung (BICM-ID) basiert, angewendet. Numerische Simulationen, die ein unabhängig und identisch verteiltes Rayleigh Blockschwund Kanalmodell verwenden, zeigen eine überlegene Leistung des modifizierten Empfängers bezüglich der Bitfehlerrate.

Ferner erweitern wir die Idee der Ausnützung von Kanalstatistik auf korrelierte Kanalmodelle und leiten damit eine optimale ML Metrik für eine nicht iterative Systemarchitektur her. Der dadurch entstandene optimale Empfänger führt eine Sequenzdetektion ohne vorausgehende Kanalschätzung aus. Wir stellen eine Implementierung der optimalen Metrik mit geringerer Komplexität bereit. Numerische Simulationen, basierend auf orthogonale frequenzgeteilte Multiplexing (OFDM) und autoregressive (AR) Kanalmodelle zeigen, wie der optimale Empfänger den fehlangepassten Empfänger hinsichtlich der Bitfehlerrate übertrifft. Zusätzlich beobachten wir beim optimalen Empfänger eine geringere Empfindlichkeit gegenüber der Anzahl der verwendeten Piloten.

Acknowledgements

It is my pleasure to express my sincere thanks to all who have contributed to the completion of this thesis in varoius ways.

I am very thankful and want to show my utmost respect to Prof. Dipl.-Ing. Dr. techn. Gerald Matz for supervising my work, introducing me to optimal receiver design and most importantly for motivating me throughout the duration of this thesis.

I want to thank and express my heartfelt gratitude to Dipl.-Ing. Clemens Novak and Dipl.-Ing. Peter Fertl for sharing their knowledge with me in countless discussions and for the proof reading of this thesis. I am very thankful for all the time and effort they have invested in helping me.

I especially want to thank my parents Dr. M.N. Jha and Dr.(Mrs) Shashi Jha for being the best parents a son could ever wish for. I also want to thank the whole Gratz family and specially my parents-in-law Dr. Ewald Gratz and Vimala Gratz for their encouraging love and support.

I most sincerely want to thank my brother-in-law Axel Krystufek and my sister Dr. Mag. Vedhu Krystufek for helping me come to this beautiful country called Austria. Thus, making my studies possible. I want to thank them from the bottom of my heart for their kind love and support.

I want to thank my wife Radha for so wonderfully encouraging me, giving me self confidence and diligently caring for me. All her immeasurable efforts and patience have played a major role in the completion of this thesis.

Finally and most importantly, I want to thank H. H. Shri Mataji Nirmala Devi for everything.

Contents

1

Introduction

"[. . .] they did not know it was impossible so they did it." —— Samuel Langhorne Clemens

IN conventional receiver design it is common to use maximum likelihood (ML) detection based on the well know Euclidean distance metric [1] [2]. There is a however a fundamental shortcoming in using this design. The ML detector based on this metric assumes a perfect knowledge of the fading channel at the receiver. However, this assumption cannot be fulfilled in practical scenarios. The so-called *mismatched receiver* has to perform an estimation of the fading channel prior to detection. Channel estimation is facilitated by multiplexing pilots in the transmitted symbol sequence [3], [4], which are fully known to the receiver. It is to be noted that pilot symbols are not information bearing symbols. The accuracy of channel estimation depends on the number of pilots and their mean symbol energy [3]. Using a noisy channel estimate in conventional ML detection causes a mismatch leading to performance degradation [5], [6].

An innovative method to counteract this problem is to utilize the statistics of the channel and its estimation error in the derivation of a ML metric for a *modified receiver* [2], [5]. We study this detection approach and apply the *modified receiver* to iterative system architectures. The iterative system we use is based on the bit-interleaved coded modulation scheme with bit-wise soft decision iterative decoding (BICM-ID) [7], [8]. The BICM-ID receiver consists of two main sub-blocks, i.e., a demodulator and a decoder which exchange soft information about the transmitted bits in an iterative manner [9]. We consider a Rayleigh block fading channel model for this system.

For a non-iterative system architecture, we take the channel correlation into account while deriving the ML metric of the *optimum receiver* for hard decision sequence detection. Moreover, the channel estimation for this system does not need to be performed explicitly as the received pilots are incorporated in the ML metric. For this system, we consider an un-coded transmission over Rayleigh fast fading channel models. We conclude this decision metric to be optimal for ML sequence detection.

A short overview of the thesis is given below.

- **Chapter 2:** Introduction to state of the art research work in the field of optimal receiver design. We consider both iterative and non-iterative receivers.

- **Chapter 3:** Describes the iterative receiver based on BICM-ID. We describe the system model, the pilot-assisted channel estimation and the demodulation/decoding procedure.

- **Chapter 4:** Deals with the *modified receiver* design and its implementation in the BICM-ID system. We derive the *modified metric* in this chapter and discuss its performance with numerical simulations.

- **Chapter 5:** Deals with the derivation of the *optimal receiver* design for use in a non-iterative system. We describe the system model, pilot-assisted channel estimation and the derivation of the *optimum metric*. The performance is discussed with the help of numerical simulations.

- **Chapter 6:** Concludes the thesis with a summary and gives the outlook for further research.

2

State of the Art

"[. . .] Engineering is the professional art of applying science to the optimum conversion of natural resources to the benefit of man." —— Ralph J. Smith

In this chapter we discuss current research work that investigates optimum receiver design if only imperfect CSI is available at the receiver. In particular we look at the work in [2], which derives a iterative receiver for a BICM multiband OFDM system and [5], which derives a non-iterative receiver for a Multiple input multiple output (MIMO) antenna system using space-time codes. Although they discuss different setups, they employ the same fundamental concept.

2.1 Receiver Architectures

Much of the State of the art research work deals with the problem of imperfect CSI and methods to reduce the performance degradation caused by channel estimation errors (CEE). The following receiver design approaches will be analyzed throughout the work. The *ideal receiver* works with the assumption of a perfectly known channel. The *mismatched receiver* performs channel estimation with the ML or the MMSE criterion prior to decoding, and uses this estimate in the metric of the ideal receiver. In [2] a *modified receiver* is derived which not only estimates the channel but also utilizes the statistics of the channel and its estimation error in deriving the likelihood function used for ML detection, thereby adapting the ML decision metric to the presence of CEE. In [5] an *optimum receiver* which uses the channel statistics is derived. It performs joint ML detection without explicitly having to estimate the channel. These receivers can be implemented both in an iterative and also in a non-iterative fashion. In any case the channel estimation is performed with the help of pilots embedded in the transmit sequence. In the following we will see examples of these receivers in different communication setups.

2.1.1 Iterative Receivers

For iterative receivers we will have a look at the work done by Sadough et al. in [2]. The fundamental task playing a key role in designing a *modified receiver* is the derivation of the respective maximum likelihood metrics, which depends on the choice of the system model. But the principle concept of utilizing the statistics of the channel and its estimation error remains the same.

In [2] a BICM-ID multiband OFDM (BICM-ID MB-OFDM) communication system is considered. MB-OFDM is known to be a spectrally efficient technique for high data rates in short range ultra wideband (UWB) applications [10]. In MB-OFDM it is assumed that the channel is time invariant during the transmission of an entire frame. Here, the system model is basically the same as that of a BICM-ID system (later described in Chapter 3) with some additional signal processing blocks to incorporate

OFDM modulation. At the transmitter, standard OFDM modulation is applied after the symbol mapper with an inverse fast fourier transformation (IFFT), along with a cyclic prefix (CP) and a guard interval. At the receiver side, a fast fourier transformation (FFT) is performed for the demodulation of the OFDM signal. An OFDM with M subcarriers allows to convert the channel into M parallel Rayleigh distributed flat fading subchannels [11], [2]. For any subband, the system can be described with an equivalent baseband model as

$$\mathbf{y}^{\mathrm{d}} = \mathbf{H}\mathbf{x}^{\mathrm{d}} + \mathbf{w}^{\mathrm{d}}, \tag{2.1}$$

where the received data symbols and transmitted data symbols are $M \times 1$ vectors represented by $\mathbf{y}^{\mathrm{d}} = [y_1^{\mathrm{d}} \quad \cdots \quad y_M^{\mathrm{d}}]$ and $\mathbf{x}^{\mathrm{d}} = [x_1^{\mathrm{d}} \quad \cdots \quad x_M^{\mathrm{d}}]$, respectively. The additive noise is assumed to be *zero mean circular symmetric complex Gaussian* (ZMCSCG), i.e., $\mathbf{w}^{\mathrm{d}} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_M)$. Here, σ_w^2 denotes the noise variance and $\mathbf{I}_M$ is a $M \times M$ identity matrix. The channel $\mathbf{H}$ is a diagonal matrix with the diagonal elements equal to the vector $\mathbf{h} = [h_1 \quad \cdots \quad h_M]$ containing the FFT coefficients.

At the receiver the ML based detector assumes perfectly known channel state information, which in practice cannot be fulfilled. Instead of the exact channel an imperfect channel estimate is used. Thus, the metric of the *ideal receiver* used in ML detection is suboptimal and the receiver employing this metric is denoted as the *mismatched receiver*. However, the mismatched metric is not able to cope up with CEE. The presence of CEE will influence the reliability of the transmission and its capacity [12]. The channel estimation of the k-th fading coefficient h_k, where $k \in \{1, \cdots, M\}$ is done by inserting pilot symbols $\mathbf{x}^{\mathrm{p}} = [x_1^{\mathrm{p}} \quad \cdots \quad x_P^{\mathrm{p}}]$ in the transmit sequence. The pilot symbols are assumed constant modulus with energy E_p. The pilots and the data symbols are assumed to observe the same channel. The received pilot sequence is given by

$$\mathbf{y}^{\mathrm{p}} = h_k \mathbf{x}^{\mathrm{p}} + \mathbf{w}^{\mathrm{p}}. \tag{2.2}$$

The $P \times 1$ noise vector $\mathbf{w}^{\mathrm{p}}$ is distributed as $\mathbf{w}^{\mathrm{p}} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_P)$. The channel estimation can now be done using a ML procedure, which is achieved by maximizing the pdf

$f(\mathbf{y}^{\mathrm{p}}|h_k, \mathbf{x}^{\mathrm{p}})$ to give

$$\hat{h}_{k,\,\mathrm{ML}} = \frac{(\mathbf{x}^{\mathrm{p}})^H \mathbf{y}^{\mathrm{p}}}{PE_p} = h_k + \epsilon_k, \tag{2.3}$$

where PE_p is the total pilot power and $\epsilon_k = \dfrac{(\mathbf{x}^{\mathrm{p}})^H \mathbf{y}^{\mathrm{p}}}{PE_p}$ is the estimation error with a complex Gaussian distribution $\epsilon_k \sim \mathcal{CN}(0, \dfrac{\sigma_w^2}{PE_p})$.

We will now have a look at the ML metrics for the different receiver designs mentioned earlier. The system model in (2.1) written in a component-wise form is given by

$$y_k^{\mathrm{d}} = h_k x_k^{\mathrm{d}} + w_k.$$

In the case of the *ideal receiver*, the ML metric can be derived by maximizing the pdf $f(y_k^{\mathrm{d}}|h_k, x_k^{\mathrm{d}})$ under the assumption that h_k is fully known. The pdf $f(y_k^{\mathrm{d}}|h_k, x_k^{\mathrm{d}})$ is distributed as $\mathcal{CN}(h_k x_k^{\mathrm{d}}, \sigma_w^2)$. This results in the following ML decision rule given by

$$\begin{aligned} \hat{x}_{k,\,\mathrm{ML}} &= \underset{x_k^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmax}}\, f(y_k^{\mathrm{d}}|h_k, x_k^{\mathrm{d}}) \\ &= \underset{x_k^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmin}} \left(|y_k^{\mathrm{d}} - h_k x_k^{\mathrm{d}}|^2\right), \end{aligned} \tag{2.4}$$

where $\mathcal{A}$ denotes the set containing all possible realizations of x_k^{d}. As explained before, the *mismatched receiver* uses the channel estimate (2.3) in the metric derived for the ideal receiver by replacing h_k with $\hat{h}_k$ in (2.4). We get

$$\hat{x}_{k,\,\mathrm{ML}} = \underset{x_k^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmin}} \left(|y_k^{\mathrm{d}} - \hat{h}_k x_k^{\mathrm{d}}|^2\right). \tag{2.5}$$

The *modified receiver* uses the pdf $f(y_k^{\mathrm{d}}|\hat{h}_k, x_k^{\mathrm{d}})$ of the received symbols given the channel estimate. This can be calculated by the integral [2]

$$\begin{aligned} f(y_k^{\mathrm{d}}|\hat{h}_k, x_k^{\mathrm{d}}) &= \int_{h_k} f(y_k^{\mathrm{d}}, h_k|\hat{h}_k, x_k^{\mathrm{d}}) dh_k \\ &= \int_{h_k} f(y_k^{\mathrm{d}}|h_k, x_k^{\mathrm{d}}) f(h_k|\hat{h_k}) dh_k, \end{aligned} \tag{2.6}$$

For the calculation of (2.6) cf. [2]. The ML decision rule for the *modified receiver* is given by

$$\begin{aligned}
\hat{x}_{k,\,\mathrm{ML}} &= \underset{x_k^{\mathrm{d}}\in\mathcal{A}}{\operatorname{argmax}}\, f(y_k^{\mathrm{d}}|\hat{h}_k, x_k^{\mathrm{d}}) \\
&= \underset{x_k^{\mathrm{d}}\in\mathcal{A}}{\operatorname{argmax}}\, \frac{\exp\left(-|y_k^{\mathrm{d}} - \rho\hat{h}_k x_k^{\mathrm{d}}|^2\right)}{\pi\left(\sigma_w^2 + (1-\rho)|x_k^{\mathrm{d}}|^2\right)} \\
&= \underset{x_k^{\mathrm{d}}\in\mathcal{A}}{\operatorname{argmin}}\, \ln\left(\sigma_w^2 + (1-\rho)|x_k^{\mathrm{d}}|^2\right) + \frac{|y_k^{\mathrm{d}} - \rho\hat{h}_k x_k^{\mathrm{d}}|^2}{\left(\sigma_w^2 + (1-\rho)|x_k^{\mathrm{d}}|^2\right)}.
\end{aligned} \tag{2.7}$$

where $\rho = \dfrac{PE_p}{PE_p + \sigma_w^2}$ is a constant. It can be easily seen that the modified receiver metric in the case of perfectly available CSI, i.e., $\rho \longrightarrow 1$ and $\hat{h}_k \longrightarrow h_k$ reduces to the metric of an *ideal receiver* in (2.4). Although the system model was based on OFDM scheme, the metric in (2.7) can still be used for ML detection in single carrier frequency selective fading scenarios [2].

The BICM-ID works with soft information using log likelihood ratios and the pdfs used in (2.4), (2.5), (2.7) are the likelihood functions, which are used in the iterative decoding process of the respective receivers (cf. Chapter 3).

The authors in [2] analyzed the performance gain obtained by using the *modified* ML metric through numerical simulations.

The *modified receiver* was shown to outperform the *mismatched receiver* especially when fewer number of pilots were used for channel estimation. The SNR required to achieve a BER $=10^{-3}$ was about 1.5 dB less if the *modified receiver* is used instead of the *mismatched receiver* (cf. [2]). For a larger number of pilots the performances of both the receivers comes very close.

Sadough et al. extend the concept of *modified* ML decoding to a MIMO-OFDM system with BICM-ID over a frequency selective Rayleigh fading channel (cf. [13]). The derived *modified* ML metric for this system has a similar form to the one in (2.7). Thus, we only discuss the numerical results presented in [13]. The SNR required to achieve a BER $= 10^{-5}$ is reduced by 1.5 dB if the *modified* ML metric is used instead of the *mismatched* ML metric (cf. [13]). The performance loss of the *mismatched receiver* becomes insignificant for a larger number of pilots.

We will see in the next subsection that the concepts for deriving the *modified receiver* can also be applied to non-iterative systems. The concept of an *optimum receiver* is also introduced.

2.1.2 Non-iterative Receivers

The non-iterative system we discuss, is based on work [5] by Taricco et al. We now consider the non-iterative narrow band multiple-input multiple-output (MIMO) communication system using standard space-time codes given in [5]. Space-time codes are designed for transmit diversity systems. The basic idea of a transmit diversity scheme is to transmit symbols in a redundant fashion spread spatially over the antennas and also in time [1]. Antenna diversity is a practical technique for reducing the effects of multipath fading, hereby improving error performance of the wireless communication system (cf. [14]). The most famous example for a space-time code is the so called Alamouti scheme [14] used in systems with two transmit antennas and one receive antenna, yielding a diversity order of two. The Alamouti scheme can also be extended to multiple antennas at the receiver and has been shown in [15].

The channel model for a MIMO system with t transmit and r receive antennas with $r \geq t$, can be given by the linear equation

$$\mathbf{Y}^{\mathrm{d}} = \mathbf{H}\mathbf{X}^{\mathrm{d}} + \mathbf{W}.$$

The space-time code χ used here, has a block length N. The transmitted data symbol matrix $\mathbf{X}^{\mathrm{d}} = [\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_N]$ is a $t \times N$ matrix and the received data symbol is represented by the $r \times N$ matrix $\mathbf{Y}$. The channel itself is denoted by a $r \times t$ complex random matrix $\mathbf{H}$. The elements of $\mathbf{H}$ are i.i.d *zero-mean circularly symmetric complex Gaussian* (ZMCSCG) random variables with unit variance and the elements of the noise matrix $\mathbf{W}$ are i.i.d ZMCSCG random variables with variance σ_w^2 and $\mathbb{E}\{\mathbf{W}\mathbf{W}^H\} = N\sigma_w^2\mathbf{I}_r$, where $\mathbf{I}_r$ is a $r \times r$ identity matrix. The average symbol energy for a MIMO based system is calculated using the *Frobenius* norm, which is actually a standard norm used for MIMO systems

$$E_{\mathrm{d}} = \frac{1}{tN}\mathbb{E}\{\|\mathbf{X}^{\mathrm{d}}\|_F^2\},$$

where $\|\mathbf{X}^{\mathrm{d}}\|_F$ represents the *Frobenius* norm and is defined as, $\|\mathbf{X}^{\mathrm{d}}\|_F \triangleq (\sum_{i,j} |\mathbf{X}^{\mathrm{d}}_{\mathbf{i,j}}|^2)^{1/2}$. The channel matrix $\mathbf{H}$ is assumed to be constant during the transmission of the whole codeword and also statistically independent of $\mathbf{X}^{\mathrm{d}}$ and $\mathbf{W}$.

Generally, for study and design purposes the channel matrix is often assumed to be perfectly known. However in a reality, this assumption is incorrect. The receiver has to estimate the channel matrix which is made possible by sending pilot symbols at the beginning of each frame. These pilots and their positions are known at the receiver. If the transmitter sends a $t \times P$ pilot matrix $\mathbf{X}^{\mathrm{p}}$ and a $t \times N$ data matrix $\mathbf{X}^{\mathrm{d}}$, the receiver observes the pilots and data symbols given by the following equations as

$$\begin{aligned} \mathbf{Y}^{\mathrm{p}} &= \mathbf{H}\mathbf{X}^{\mathrm{p}} + \mathbf{W} \\ \mathbf{Y}^{\mathrm{d}} &= \mathbf{H}\mathbf{X}^{\mathrm{d}} + \mathbf{W}. \end{aligned}$$

The average pilot symbol energy can be calculated by the *Frobenius* norm in the same way as for the data symbols and is given by

$$E_{\mathrm{p}} = \frac{1}{tN}\mathbb{E}\{\|\mathbf{X}^{\mathrm{p}}\|_F^2\}.$$

The estimation of the channel matrix $\mathbf{H}$ can be done by using a maximum likelihood (ML) or a minimum mean squared error (MMSE) estimator.

The ML estimate of $\mathbf{H}$ is obtained by maximizing $f(\mathbf{Y}^{\mathrm{p}}|\mathbf{H}, \mathbf{X}^{\mathrm{p}})$, which leads to minimizing $\|\mathbf{Y}^{\mathrm{p}} - \mathbf{H}\mathbf{X}^{\mathrm{p}}\|^2$ with respect to $\mathbf{H}$ yielding:

$$\hat{\mathbf{H}}_{\mathrm{ML}} = \mathbf{Y}^{\mathrm{p}}\mathbf{X}^{\mathrm{p}}\,(\mathbf{X}^{\mathrm{p}}(\mathbf{X}^{\mathrm{p}})^H)^{-1} = \mathbf{H} + \mathbf{E}$$

where the channel estimation error matrix $\mathbf{E} = \mathbf{W}\mathbf{X}^{\mathrm{p}}\,(\mathbf{X}^{\mathrm{p}}(\mathbf{X}^{\mathrm{p}})^H)^{-1}$.

The MMSE channel estimate is obtained by the linear transformation $\mathbf{Y}^{\mathrm{p}}\mathbf{A}$, where $\mathbf{A}$ is a $P \times t$ matrix minimizing the mean square error $\mathbb{E}\{\|\mathbf{Y}^{\mathrm{p}}\mathbf{A} - \mathbf{H}\|^2\}$ yielding:

$$\hat{\mathbf{H}}_{\mathrm{MMSE}} = \mathbf{Y}^{\mathrm{p}}(\mathbf{X}^{\mathrm{p}})^H\,(\sigma_w^2\mathbf{I}_t + \mathbf{X}^{\mathrm{p}}(\mathbf{X}^{\mathrm{p}})^H)^{-1}$$

The receivers can be differentiated in the way they utilize the channel estimate $\hat{\mathbf{H}}$, which can be done in various ways. Firstly the receiver estimates the channel $\mathbf{H}$ by observing the pilot matrix sent in the transmission and the resulting estimate is used in the decision metrics of the corresponding receivers. Note that different estimators will in general produce different ML receivers [5].

Mismatched receiver: The *mismatched receiver* uses the estimated channel matrix in the metric designed under the assumption that perfect channel state information is available. This results in the following metric

$$\hat{\mathbf{X}}_{\text{mis, ML}} = \underset{\mathbf{X}^{\text{d}}}{\operatorname{argmin}} \|\mathbf{Y}^{\text{d}} - \hat{\mathbf{H}}\mathbf{X}^{\text{d}}\|^2. \tag{2.8}$$

The mismatched receiver depends on the channel estimator used, as $\hat{\mathbf{H}}$ depends on the channel estimation criterion.

Modified receiver: If the receiver, apart from estimating the channel, conditions the likelihood function in its ML decision rule on the channel estimate, we get

$$\hat{\mathbf{X}}_{\text{mod, ML}} = \underset{\mathbf{X}^{\text{d}}}{\operatorname{argmax}}\, f(\mathbf{Y}^{\text{d}}|\mathbf{X}^{\text{d}}, \hat{\mathbf{H}}). \tag{2.9}$$

This receiver utilizes not only the channel estimate but also the statistical distribution of the channel estimation error. It may seem that the ML metrics would depend on the channel estimator, but it turns out to be independent of the linear channel estimator used [5]. This receiver is termed as a *modified receiver*.

Optimum receiver: It is also possible for the receiver to avoid estimating the channel explicitly if it jointly detects the transmitted symbol matrix $\mathbf{X}^{\text{d}}$ by processing $\mathbf{Y}^{\text{d}}$, $\mathbf{Y}^{\text{p}}$ and $\mathbf{X}^{\text{p}}$ at the same time. The ML decision rule for this receiver is given by

$$\hat{\mathbf{X}}_{\text{opt, ML}} = \underset{\mathbf{X}^{\text{d}}}{\operatorname{argmax}}\, f(\mathbf{Y}^{\text{d}}, \mathbf{Y}^{\text{p}}|\mathbf{X}^{\text{d}}, \mathbf{X}^{\text{p}}). \tag{2.10}$$

The receiver using this detection method is termed as an *optimum receiver* in [5]. It focuses on directly detecting the data symbols.

The pdfs (likelihood functions) and the corresponding ML metrics of receivers in are derived in [5]. Here, we only discuss their final expressions.

For the *modified receiver* in (2.9) the pdf $f(\mathbf{Y}^{\mathrm{d}}|\mathbf{X}^{\mathrm{d}}, \hat{\mathbf{H}})$ is given by[1]

$$f(\mathbf{Y}^{\mathrm{d}}|\mathbf{X}^{\mathrm{d}}, \hat{\mathbf{H}}) = \frac{\exp\left(\mathrm{tr}\left[-(\mathbf{Y}^{\mathrm{d}} - \rho\hat{\mathbf{H}}\mathbf{X}^{\mathrm{d}})[\sigma_w^2\mathbf{I}_N + (1-\rho)(\mathbf{X}^{\mathrm{d}})^H\mathbf{X}^{\mathrm{d}}]^{-1}(\mathbf{Y}^{\mathrm{d}} - \rho\hat{\mathbf{H}}\mathbf{X}^{\mathrm{d}})^H\right]\right)}{\det\left(\pi[\sigma_w^2\mathbf{I}_N + (1-\rho)(\mathbf{X}^{\mathrm{d}})^H\mathbf{X}^{\mathrm{d}}]\right)^r}, \quad (2.11)$$

where ρ is a constant defined as $\rho \triangleq \dfrac{PE_p}{\sigma_w^2 + PE_p}$. The corresponding ML metric is calculated by taking logarithm of (2.11) and ignoring terms independent of $\mathbf{X}^{\mathrm{d}}$. The *modified* ML metric is given by

$$\hat{\mathbf{X}}_{\text{mod, ML}} = \underset{\mathbf{X}^{\mathrm{d}}}{\operatorname{argmin}} \quad r\sigma_w^2 \ln\det\left[\mathbf{I}_N + \frac{(1-\rho)(\mathbf{X}^{\mathrm{d}})^H\mathbf{X}^{\mathrm{d}}}{\sigma_w^2}\right] + \mathrm{tr}\left((\mathbf{Y}^{\mathrm{d}} - \rho\hat{\mathbf{H}}\mathbf{X}^{\mathrm{d}})\left[\mathbf{I}_N + \frac{(1-\rho)(\mathbf{X}^{\mathrm{d}})^H\mathbf{X}^{\mathrm{d}}}{\sigma_w^2}\right]^{-1}(\mathbf{Y}^{\mathrm{d}} - \rho\hat{\mathbf{H}}\mathbf{X}^{\mathrm{d}})^H\right). \quad (2.12)$$

For the sake of simplicity, the metric to be minimized in (2.12) is denoted by $\xi_{\text{mod}}(\mathbf{X}^{\mathrm{d}})$. The *modified* metric in (2.12) yields interesting results for following extreme cases

$$\xi_{\text{mod}}(\mathbf{X}^{\mathrm{d}}) = \begin{cases} \|\mathbf{Y}^{\mathrm{d}} - \hat{\mathbf{H}}\mathbf{X}^{\mathrm{d}}\|^2, & \rho \longrightarrow 1; \\ r\sigma_w^2 \ln\det\left[\mathbf{I}_N + \dfrac{(\mathbf{X}^{\mathrm{d}})^H\mathbf{X}^{\mathrm{d}}}{\sigma_w^2}\right] + & \\ \mathrm{tr}\left(\mathbf{Y}^{\mathrm{d}}\left[\mathbf{I}_N + \dfrac{(\mathbf{X}^{\mathrm{d}})^H\mathbf{X}^{\mathrm{d}}}{\sigma_w^2}\right]^{-1}\mathbf{Y}^{\mathrm{d}\,H}\right), & \rho \longrightarrow 0. \end{cases}$$

The first case for $\rho \longrightarrow 1$ corresponds to $\dfrac{PE_p}{\sigma_w^2} \longrightarrow \infty$, i.e., perfect channel state information is available. The ML metric automatically reduces to that of an *ideal*

[1] The function tr(A) is equivalent to trace(A)

receiver. The second case, where $\rho \longrightarrow 0$, i.e., $\frac{PE_p}{\sigma_w^2} \longrightarrow 0$ corresponds to the situation where no channel state information is available. This is visible by the fact that $\hat{\mathbf{H}}$ is missing in the second expression. An important fact to be noted is that the modified metric suits itself to the quality of the available CSI. This actually results from the fact that the distribution of the channel estimation error was incorporated while calculating this metric. These results are only valid under the conditional independence of the received vector components, which was assumed in the channel model beforehand.

Finally the ML metrics for the *optimum receiver* is calculated by maximizing the pdf $f(\mathbf{Y}^{\mathrm{d}}, \mathbf{Y}^{\mathrm{p}}|\mathbf{X}^{\mathrm{d}}, \mathbf{X}^{\mathrm{p}})$ avoiding the need to involve any prior estimation procedure of the channel matrix. It is assumed that that joint pdf of $\mathbf{H}$ and $\mathbf{W}$ is known. After taking the logarithm of $f(\mathbf{Y}^{\mathrm{d}}, \mathbf{Y}^{\mathrm{p}}|\mathbf{X}^{\mathrm{d}}, \mathbf{X}^{\mathrm{p}})$ the resulting *optimum* ML metric is given by the expression (cf. [5])

$$\begin{aligned}\xi_{\mathrm{opt}}(\mathbf{X}^{\mathrm{d}}) \quad &= \quad r \ln\det\left[\mathbf{I}_t + \frac{\mathbf{X}^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}})^H + \mathbf{X}^{\mathrm{p}}(\mathbf{X}^{\mathrm{p}})^H}{\sigma_w^2}\right] \quad + \\ \mathrm{tr}&\left((\mathbf{Y}^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}})^H + \mathbf{Y}^{\mathrm{p}}(\mathbf{X}^{\mathrm{p}})^H)\left[\mathbf{I}_t + \frac{\mathbf{X}^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}})^H + \mathbf{X}^{\mathrm{p}}(\mathbf{X}^{\mathrm{p}})^H}{\sigma_w^2}\right]^{-1} \frac{\mathbf{X}^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}})^H + \mathbf{X}^{\mathrm{p}}(\mathbf{X}^{\mathrm{p}})^H}{(\sigma_w^2)^2}\right).\end{aligned} \tag{2.13}$$

The results for the *optimum receiver* are only valid under the assumption that $\mathbf{H}$ and $\mathbf{W}$ have i.i.d elements distributed as $\mathcal{CN}(0,1)$ and $\mathcal{CN}(0,\sigma_w^2)$, respectively. Otherwise the metric is not optimum. For the special case of orthogonal pilot matrix, i.e., $\mathbf{X}^{\mathrm{p}}(\mathbf{X}^{\mathrm{p}})^H = PE_p\mathbf{I}_t$ the *optimum* metric becomes equivalent to the *modified* metric.

Moreover, the result in (2.12) shows a great similarity to the metric derived in [13] for a MIMO-OFDM with BICM-ID, even though the latter was calculated for an iterative setup with a different system model (cf. [13], [5]). We have thus seen, that for both iterative and non-iterative receiver designs the basic task of calculating a *modified* or *optimum* metric is fundamentally the same.

2.1.2.1 Iterative Metric Computation

In order to implement the *optimum* metric in (2.13) using a sequential decoding algorithm e.g., a *Viterbi* decoder, it is necessary to compute the metric recursively. The authors in [5] have been able to find such an iterative computational algorithm for the *optimum* metric which is discussed here. For that cause the transmitted and received matrices can be split as $\mathbf{X}^{\mathrm{d}} = (\mathbf{X}^{\mathrm{d}-}, \mathbf{x}^{\mathrm{d}})$ and $\mathbf{Y}^{\mathrm{d}} = (\mathbf{Y}^{\mathrm{d}-}, \mathbf{y}^{\mathrm{d}})$, where $\mathbf{x}^{\mathrm{d}}$ and $\mathbf{y}^{\mathrm{d}}$ are column vectors. The objective is to calculate the metric increment defined as

$$\Delta\xi_{\mathrm{opt}}(\mathbf{x}^{\mathrm{d}}; \mathbf{X}^{\mathrm{d}-}) \triangleq \xi_{\mathrm{opt}}(\mathbf{X}^{\mathrm{d}}) - \xi_{\mathrm{opt}}(\mathbf{X}^{\mathrm{d}-}). \tag{2.14}$$

The necessary functions for further calculations are defined as

$$\begin{aligned} \boldsymbol{\Xi}(\mathbf{X}^{\mathrm{d}}) &\triangleq \frac{\mathbf{X}^{\mathrm{d}}(\mathbf{Y}^{\mathrm{d}})^H + \mathbf{X}^{\mathrm{p}}(\mathbf{Y}^{\mathrm{p}})^H}{\sigma_w^2} \\ \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}}) &\triangleq \left(\mathbf{I}_t + \frac{\mathbf{X}^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}})^H + \mathbf{X}^{\mathrm{p}}(\mathbf{X}^{\mathrm{p}})^H}{\sigma_w^2}\right)^{-1}. \end{aligned} \tag{2.15}$$

With (2.15) the *optimum* metric can be rewritten as

$$\xi_{\mathrm{opt}}(\mathbf{X}^{\mathrm{d}}) = -r \ln\det \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}}) - \mathrm{tr}\left(\boldsymbol{\Xi}(\mathbf{X}^{\mathrm{d}})^H \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}}) \boldsymbol{\Xi}(\mathbf{X}^{\mathrm{d}})\right). \tag{2.16}$$

To calculate the metric increment defined in (2.14), the functions in (2.15) have to be split up into components dependent on $\mathbf{X}^{\mathrm{d}-}$ and $\mathbf{x}^{\mathrm{d}}$ and are given as

$$\begin{cases} \boldsymbol{\Xi}(\mathbf{X}^{\mathrm{d}}) = \boldsymbol{\Xi}(\mathbf{X}^{\mathrm{d}-}) + \dfrac{\mathbf{x}^{\mathrm{d}}(\mathbf{y}^{\mathrm{d}})^H}{\sigma_w^2} \\ \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}}) = \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}-}) - \left(\dfrac{\boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}-})\mathbf{x}^{\mathrm{d}}(\mathbf{x}^{\mathrm{d}})^H \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}-})}{\sigma_w^2 + (\mathbf{x}^{\mathrm{d}})^H \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}-})\mathbf{x}^{\mathrm{d}}}\right) \\ \ln\det \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}}) = \ln\det \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}-}) - \ln\left(1 + \dfrac{(\mathbf{x}^{\mathrm{d}})^H \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}-})\mathbf{x}^{\mathrm{d}}}{\sigma_w^2}\right). \end{cases} \tag{2.17}$$

Using the results in (2.17), the metric increment turns out to be

$$\begin{aligned} \Delta\xi_{\mathrm{opt}}(\mathbf{x}^{\mathrm{d}}; \mathbf{X}^{\mathrm{d}-}) &= \ln\left(1 + \frac{(\mathbf{x}^{\mathrm{d}})^H \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}-})\mathbf{x}^{\mathrm{d}}}{\sigma_w^2}\right) \quad + \\ &\mathrm{tr}\left(\boldsymbol{\Xi}(\mathbf{X}^{\mathrm{d}-})^H \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}-}) \boldsymbol{\Xi}(\mathbf{X}^{\mathrm{d}-}) - \boldsymbol{\Xi}(\mathbf{X}^{\mathrm{d}})^H \boldsymbol{\Lambda}(\mathbf{X}^{\mathrm{d}}) \boldsymbol{\Xi}(\mathbf{X}^{\mathrm{d}})\right). \end{aligned} \tag{2.18}$$

Equation (2.18) can be used to calculate the branch metrics for a *Viterbi* decoder. It is to be noted that the branch metric is independent of the code word length N. The iterative metric algorithm is practical for implementing the *optimum* metric, because it reduces the computational complexity and also the processing time. The extensive calculation procedure for deriving the branch metrics is in [5].

The numerical results presented by the authors in [5] were based on two different trellis space time codes (STC-1 and STC-2) (cf. [5] for details).

The performance was measured in terms of frame error rate (FER). Moreover, the number of pilots needed to achieve optimum performance by the *mismatched* and *modified* receivers was found. With $t = r = 2$ the gaps of the *mismatched* and *modified* receivers to the *ideal receiver* with perfect CSI were measured in terms of SNR. A few results are summarized here.

The *modified receiver* achieves a 0.3 dB gap using STC-1 at a FER $= 10^{-2}$ with $P = 4$ pilots and a 0.45 dB gap using STC-2 at a FER $= 10^{-2}$ with $P = 4$ pilots.

The *mismatched receiver* achieves 1 dB gap using STC-1 at a FER $= 10^{-2}$ with $P = 16$ pilots and a 1.1 dB gap using STC-2 at a FER $= 10^{-2}$ with $P = 16$ pilots.

It was concluded in [5] that the *mismatched receiver* attains its optimum performance with $P = 16$ pilots (11% pilot overhead), while the *modified receiver* attains its optimum performance for $P = 4$ pilots (3% pilot overhead). The overall gain achieved by using the *modified receiver* is about 0.7 dB [5].

3

Bit-interleaved coded modulation

"[. . .] Everything should be made as simple as possible, but not one bit simpler." —— ALBERT EINSTEIN

IN this chapter we take a detailed look at the bit-interleaved coded modulation scheme employing soft decision iterative decoding (BICM-ID) [16], [2]. The BICM-ID scheme builds the foundation for the implementation of iterative receivers in Chapter 4. Bit-interleaving helps to achieve a large diversity order and provides protection against burst errors [16]. ML detection in BICM requires splitting of the sequence probabilities into bit probabilities. This is only possible by assuming the encoded bits to be independent. Hence, BICM is suboptimal in performance [16]. But by using iterative demodulation and decoding the performance loss can be reduced [9]. The advantage of iterative decoding lies in the reduction of computational complexity because joint demodulation and decoding is too complicated to implement [17], [18]. We only investigate the BICM-ID scheme for single antenna systems in this thesis.

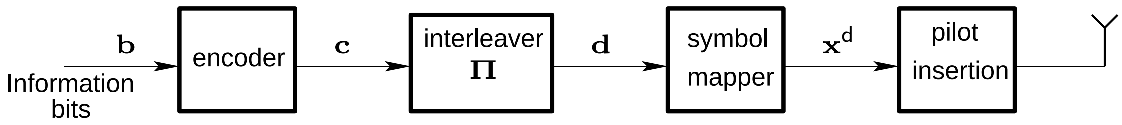


Figure 3.1: BICM Transmitter

3.1 System Model

The BICM transmitter consists of the blocks shown in Fig. 3.1. These blocks are described as follows:

In the BICM scheme an information bit sequence of length K, i.e., $\mathbf{b} = [b_1 \quad \cdots \quad b_K]^T$ where $b_k \in \{0, 1\}$ and $k \in \{1, \cdots, K\}$ is first encoded with an error-correcting code $\mathbb{C}$ to form a sequence $\mathbf{c} = [c_1 \quad \cdots \quad c_L]^T$ of coded bits c_l, where $l \in \{1, \cdots, L\}$ and $\mathbf{c} \in \mathbb{C}$. The length $L \geq K$ of the coded bit sequence depends on the code. Note that the indices taken for the information bits k and for the coded bits l are different. This is done in order to emphasize the fact that the given number of information bits usually corresponds to a different number of encoded bits. In general, the rate of the code is given by $R = K/L$. For example consider a convolutional code of rate $R = \dfrac{1}{2}$. The length of the coded bit sequence is then given by $L = |\mathbf{c}| = K/R = 2K$ resulting in the coded bit sequence $\mathbf{c} = [c_1 \quad \cdots \quad c_{2K}]^T$. As there is a one to one correspondence between the bit sequences and codewords, the cardinality of the set of valid codewords is given by $|\mathbb{C}| = 2^K$ for the code rate $R = \dfrac{1}{2}$.

The coded bit sequence $\mathbf{c}$ is permuted via a pseudo-random interleaver $\mathbf{\Pi}$, which is considered to be ideal [8]. The correspondence between the coded bit sequences and the interleaved bit sequences is represented by the permutation function $\mathbf{\Pi}$, giving us the equation $\mathbf{d} = \mathbf{\Pi}(\mathbf{c})$. The interleaver permutates the coded bit sequence $\mathbf{c} = [c_1 \quad \cdots \quad c_L]^T$ into an interleaved bit sequence $\mathbf{d} = [d_1 \quad \cdots \quad d_L]^T$, which has the same cardinality as the coded bit sequence $\mathbf{c}$. The interleaver helps to protect the transmission against burst errors [19]. These errors can cause too many consecutive bits to be overwritten in a single codeword, making it impossible to decode it correctly. Error correcting codes can correct only a limited number of errors in one single codeword. Interleaving solves this problem and distributes the errors over many codewords. This way, the error correcting code is able to work more efficiently, as there are less number of errors per codeword. Moreover, an ideal interleaver introduces statistical independency between the coded bits.

The symbol alphabet is represented by $\mathcal{A}$ and contains all the possible symbols. For the chosen modulation type the cardinality of the symbol alphabet is given by $M_a = |\mathcal{A}| = 2^m$, where m is the number of bits assigned per symbol. For example, 16QAM has an alphabet size of 16 resulting in $m = 4$ bits/symbol.

The deinterleaved bit sequence $\mathbf{d}$ is broken into $n \in \{1, \cdots, N\}$ subsequences of m bits each. The n-th subsequence $\mathbf{d}_n = [d_n \quad d_{n+1} \quad \cdots \quad d_{n+m-1}]^T$ is mapped onto complex higher level non-binary symbols to give the symbol vector $\mathbf{x}^{\mathrm{d}} = [x_1^{\mathrm{d}} \cdots \quad x_N^{\mathrm{d}}]^T$, where $x_n^{\mathrm{d}} \in \mathcal{A}$. The mean power of the data symbols is given by $E_d = \mathbb{E}\{|x_n^{\mathrm{d}}|^2\}$. The relation between the number of subsequences and mapped symbols N is defined as $L = N\,m$.

Finally, the pilot symbols are inserted at the beginning of each data block in order to facilitate estimation of the channel at the receiver. The pilots and their positions are fully know at the receiver. It is possible to estimate the channel state information using linear estimation techniques like least squares (LS) estimation, minimum mean square error (MMSE) estimation, see Section 3.2.

The $N \times 1$ transmitted data symbol vector $\mathbf{x}^{\mathrm{d}}$ and the $P \times 1$ inserted pilot symbol vector $\mathbf{x}^{\mathrm{p}}$ are defined as

$$\mathbf{x}^{\mathrm{d}} = \begin{bmatrix} x_1^{\mathrm{d}} & \cdots & x_N^{\mathrm{d}} \end{bmatrix}^T \tag{3.1}$$

$$\mathbf{x}^{\mathrm{p}} = \begin{bmatrix} x_1^{\mathrm{p}} & \cdots & x_P^{\mathrm{p}} \end{bmatrix}^T . \tag{3.2}$$

We assume pilot symbols with equal energy, i.e., $|x_1^{\mathrm{p}}|^2 = |x_2^{\mathrm{p}}|^2 = \cdots = |x_P^{\mathrm{p}}|^2 = E_p$. The system model for the data and pilot symbol vectors is described by the following linear equations

$$\mathbf{y}^{\mathrm{d}} \triangleq h\mathbf{x}^{\mathrm{d}} + \mathbf{w}^{\mathrm{d}} \tag{3.3}$$

$$\mathbf{y}^{\mathrm{p}} \triangleq h\mathbf{x}^{\mathrm{p}} + \mathbf{w}^{\mathrm{p}}. \tag{3.4}$$

The $N \times 1$ vector $\mathbf{y}^{\mathrm{d}}$ represents the received data symbol and the $P \times 1$ vector $\mathbf{y}^{\mathrm{p}}$ represents the received pilot symbols. We define a *complex Gaussian* distribution with mean μ and variance σ^2 as $\mathcal{CN}(\mu, \sigma^2)$. The $N \times 1$ vector $\mathbf{w}^{\mathrm{d}}$ denotes additive noise with a *zero-mean circular symmetric complex Gaussian* (ZMCSCG) distribution $\mathbf{w}^{\mathrm{d}} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_N)$ and the $P \times 1$ vector $\mathbf{w}^{\mathrm{p}}$ also has a ZMCSCG distribution $\mathbf{w}^{\mathrm{p}} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_P)$, where σ_w^2 denotes the noise variance. Here, $\mathbf{I}_N$ and $\mathbf{I}_P$ are identity matrices of dimension $N \times N$ and $P \times P$, respectively. We consider a Rayleigh flat fading channel model [1]. It is to be noted in (3.3) and (3.4) that the pilot and data symbol vectors observe the same channel h, which is typical for a block fading scenario. The channel only changes for each block and can thus be represented by a single coefficient with a ZMCSCG distribution as $h \sim \mathcal{CN}(0, \sigma_h^2)$, where σ_h^2 denotes the channel variance.

3.2 Pilot Assisted Channel Estimation

For this thesis we consider linear channel estimation techniques based on pilots.

The channel estimator observes the received pilot symbol sequence $\mathbf{y}^{\mathrm{p}}$. The pilots $\mathbf{x}^{\mathrm{p}}$ are known at the receiver. From the definition of the system model for pilots given in (3.4), we know that $\mathbf{y}^{\mathrm{p}} = h\mathbf{x}^{\mathrm{p}} + \mathbf{w}^{\mathrm{p}}$. We will use this equation to derive an estimate of the channel.

Two well known fundamental techniques for linear estimation [20] are

- Least squares (LS) estimation
- Minimum mean square error (MMSE) estimation.

A significant property of LS estimation method is that no statistical assumptions are made regarding the parameters which are estimated. The basic idea in the least squares approach is to minimize the squared difference between the observed pilot symbol sequence $\mathbf{y}^{\mathrm{p}}$ and the assumed sequence $h\mathbf{x}^{\mathrm{p}}$ with respect to h. We define the *Euclidean* length of a complex valued vector $\boldsymbol{\chi} = [\chi_1 \quad \chi_2 \quad \cdots \quad \chi_N]^T$ by

$$||\boldsymbol{\chi}|| = \sqrt{\sum_{i=1}^{N} \chi_i^2} = \sqrt{\boldsymbol{\chi}^H \boldsymbol{\chi}}. \tag{3.5}$$

The LS estimate $\hat{h}$ is found by minimizing the *least squares* error which is given by

$$\begin{aligned} \epsilon_{\mathrm{LS}}(h) &= (\mathbf{y}^{\mathrm{p}} - \tilde{\mathbf{x}})^H (\mathbf{y}^{\mathrm{p}} - \tilde{\mathbf{x}}) \\ &= (\mathbf{y}^{\mathrm{p}} - h\mathbf{x}^{\mathrm{p}})^H (\mathbf{y}^{\mathrm{p}} - h\mathbf{x}^{\mathrm{p}}) \\ &= \|\mathbf{y}^{\mathrm{p}} - h\mathbf{x}^{\mathrm{p}}\|^2, \end{aligned} \tag{3.6}$$

The function $\epsilon_{\mathrm{LS}}(h)$ in (3.6) is a quadratic function of h. To minimize the *least squares* error, the gradient of (3.6) with respect to h can be calculated as

$$\frac{\partial \epsilon_{\mathrm{LS}}(h)}{\partial h} = -2(\mathbf{x}^{\mathrm{p}})^H \mathbf{y}^{\mathrm{p}} + 2(\mathbf{x}^{\mathrm{p}})^H \mathbf{x}^{\mathrm{p}} h. \tag{3.7}$$

By setting $\dfrac{\partial \epsilon_{\mathrm{LS}}(h)}{\partial h} = 0$ from (3.7) we get the LS estimate given by

$$\hat{h}_{\mathrm{LS}} = ((\mathbf{x}^{\mathrm{p}})^H \mathbf{x}^{\mathrm{p}})^{-1} (\mathbf{x}^{\mathrm{p}})^H \mathbf{y}^{\mathrm{p}}. \tag{3.8}$$

The corresponding least square error can be calculated by inserting (3.8) in (3.6) and is given by

$$\begin{aligned} \epsilon_{\mathrm{LS}}(h) \,|_{h=\hat{h}} &= \epsilon_{\mathrm{LS}}(\hat{h}) \\ &= (\mathbf{y}^{\mathrm{p}})^H (\mathbf{I} - \mathbf{x}^{\mathrm{p}}((\mathbf{x}^{\mathrm{p}})^H \mathbf{x}^{\mathrm{p}})^{-1} (\mathbf{x}^{\mathrm{p}})^H) \mathbf{y}^{\mathrm{p}}. \end{aligned} \tag{3.9}$$

From (3.6) it is obvious that the LS approach strives to minimize the Euclidean distance between the observed signal vector $\mathbf{y}^{\mathrm{p}}$ to the assumed signal vector $h\mathbf{x}^{\mathrm{p}}$. A geometrical interpretation of the LS method is given in [20]. An important result of this interpretation is that the LS error lies orthogonal to signal space spanned by the pilots vectors.

The process of minimizing the *least squares* error is also equivalent to maximizing the likelihood function $f(\mathbf{y}^{\mathrm{p}}|h, \mathbf{x}^{\mathrm{p}})$ with respect to h and yields the same results as in

(3.8). With the number of pilots P and pilot energy E_p as defined in the system model in Section 3.1, the LS estimate of h from (3.8) is given by

$$\hat{h}_{\mathrm{LS}} = \frac{1}{PE_p} (\mathbf{x}^{\mathrm{p}})^H \mathbf{y}^{\mathrm{p}} = h + \tilde{w}, \tag{3.10}$$

where $\tilde{w} = \frac{1}{PE_p} (\mathbf{x}^{\mathrm{p}})^H \mathbf{w}^{\mathrm{p}}$ is the channel estimation error (CEE). The product $(\mathbf{x}^{\mathrm{p}})^H \mathbf{x}^{\mathrm{p}}$ in (3.8) reduces to $\|\mathbf{x}^{\mathrm{p}}\|^2 = PE_p$, for the equal energy assumption we made in the system model. The noise, as we know from the system model is Gaussian distributed as $\mathbf{w}^{\mathrm{p}} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_P)$. As a consequence $\tilde{w}$ is also Gaussian distributed but with a different variance, i.e., $\tilde{w} \sim \mathcal{CN}(0, \sigma_{\tilde{w}}^2)$. The error variance is given by

$$\sigma_{\tilde{w}}^2 \triangleq \frac{\sigma_w^2}{PE_p}. \tag{3.11}$$

The statistics of the channel estimation error can be exploited to enhance the performance of the system. Such metrics are derived and discussed in Chapter 4.

The MMSE estimator is based on the Bayesian approach [20]. Unlike LS estimation the Bayesian approach does not assume the channel to be an unknown deterministic constant. Instead, the statistic of h as defined in the system model in Section 3.1 is taken into account. With this method, prior knowledge of the channel statistics can be utilized to improve the estimation accuracy.

We will first define the *mean square error* (MSE) for h as

$$\epsilon_{\mathrm{MSE}}(\hat{h}) = \mathbb{E}\{(h - \hat{h})^2\}, \tag{3.12}$$

where the channel coefficient h has a given pdf as described in the system model model 3.1, i.e., $h \sim \mathcal{CN}(0, \sigma_h^2)$. The problem now is to find $\hat{h}$ that minimizes the mean square error defined in (3.12). The expectation operator $\mathbb{E}$ is with respect to the joint probability density function $f(\mathbf{y}^{\mathrm{p}}, h)$, with the observations $\mathbf{y}^{\mathrm{p}}$ defined in the system model (3.4).

As the channel h is Gaussian distributed, the MMSE estimator reduces to a linear MMSE (LMMSE) estimator (cf. [20]). Thus, $\hat{h}$ can be expressed as a linear function of the observations $\mathbf{y}^{\mathrm{p}}$ as

$$\hat{h} = \mathbf{a}^T \mathbf{y}^{\mathrm{p}}. \tag{3.13}$$

Equation (3.13) shows that $\hat{h}$ lies in the P-dimensional subspace $\mathcal{S}_P$ spanned by $\{y_1^{\mathrm{p}}, \cdots, y_P^{\mathrm{p}}\}$. The LMMSE estimator minimizing the MSE in (3.12) can be expressed as

$$\hat{h}_{\mathrm{MMSE}} = \underset{\hat{h}\in\mathcal{S}_P}{\operatorname{argmin}}\, \mathbb{E}\{(h-\hat{h})^2\} = \underset{\hat{h}\in\mathcal{S}_P}{\operatorname{argmin}}\, \|h-\hat{h}\|^2 \tag{3.14}$$

The *projection theorem* [20] states that $\hat{h} \in \mathcal{S}_P$ which minimizes $\|h-\hat{h}\|^2$ is the *orthogonal projection* of h on $\mathcal{S}_P$. This means that $h-\hat{h}$ is orthogonal[1] to $\mathcal{S}_P$. Using (3.13) we have

$$\langle h - \mathbf{a}^T\mathbf{y}^{\mathrm{p}}, \mathbf{y}^{\mathrm{p}}\rangle \;=\; \mathbb{E}\{(h-\mathbf{a}^T\mathbf{y}^{\mathrm{p}})(\mathbf{y}^{\mathrm{p}})^H\} = \mathbf{0}. \tag{3.15}$$

We further calculate the expectation in (3.15) as

$$\begin{aligned}\mathbb{E}\{(h-\mathbf{a}^T\mathbf{y}^{\mathrm{p}})(\mathbf{y}^{\mathrm{p}})^H\} &= \mathbb{E}\{h(\mathbf{y}^{\mathrm{p}})^H\} - \mathbf{a}^T\mathbb{E}\{\mathbf{y}^{\mathrm{p}}(\mathbf{y}^{\mathrm{p}})^H\}\\ &= \mathbf{c}_{hy^{\mathrm{p}}} - \mathbf{a}^T\mathbf{C}_{y^{\mathrm{p}}y^{\mathrm{p}}}. \end{aligned}\tag{3.16}$$

Using (3.15) and (3.16) we get $\mathbf{a}^T = \mathbf{c}_{hy^{\mathrm{p}}}\mathbf{C}_{y^{\mathrm{p}}y^{\mathrm{p}}}^{-1}$. The LMMSE estimator can thus be given as

$$\hat{h}_{\mathrm{MMSE}} = \mathbf{c}_{hy^{\mathrm{p}}}\mathbf{C}_{y^{\mathrm{p}}y^{\mathrm{p}}}^{-1}\mathbf{y}^{\mathrm{p}}. \tag{3.17}$$

The corresponding variance of the estimation error $(h-\hat{h}_{\mathrm{MMSE}})$ can be calculated utilizing the orthogonality between $(h-\hat{h}_{\mathrm{MMSE}})$ and $\hat{h}_{\mathrm{MMSE}}$ as

$$\begin{aligned}\sigma_{\mathrm{MMSE}}^2 &= \mathbb{E}\{(h-\hat{h}_{\mathrm{MMSE}})(h-\hat{h}_{\mathrm{MMSE}})^H\}\\ &= \mathbb{E}\{(h-\hat{h}_{\mathrm{MMSE}})h^H\} - \mathbb{E}\{(h-\hat{h}_{\mathrm{MMSE}})\hat{h}_{\mathrm{MMSE}}^H\}\\ &= \mathbb{E}\{hh^H\} - \mathbb{E}\{\hat{h}_{\mathrm{MMSE}}h^H\}\\ &= \sigma_h^2 - \mathbf{c}_{hy^{\mathrm{p}}}\mathbf{C}_{y^{\mathrm{p}}y^{\mathrm{p}}}^{-1}\mathbf{c}_{y^{\mathrm{p}}h}. \end{aligned}\tag{3.18}$$

Equation (3.17) and (3.18) are actually the mean and variance of the *posterior* pdf $f(h|\mathbf{y}^{\mathrm{p}})$, respectively (cf. [20]). The variance of the pdf $f(h|\mathbf{y}^{\mathrm{p}})$ gives the accuracy of

[1]Two complex valued vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if their inner product is equals to zero, i.e., $\langle\mathbf{x},\mathbf{y}\rangle = \mathbb{E}\{\mathbf{x}^H\mathbf{y}\} = \mathbf{0}$

estimation. For a smaller variance the $f(h|\mathbf{y}^{\mathrm{p}})$ will be more concentrated about its mean.

To calculate the exact expressions of the $\hat{h}_{\mathrm{MMSE}}$ and σ^2_{MMSE} we need to evaluate the following covariances using the system model in (3.4) as

$$\begin{aligned}\mathbf{C}_{y^{\mathrm{p}}y^{\mathrm{p}}} &= \mathbb{E}\{\mathbf{y}^{\mathrm{p}}(\mathbf{y}^{\mathrm{p}})^H\} \\ &= \mathbb{E}\{(h\mathbf{x}^{\mathrm{p}}+\mathbf{w}^{\mathrm{p}})(h\mathbf{x}^{\mathrm{p}}+\mathbf{w}^{\mathrm{p}})^H\} \\ &= \mathbf{x}^{\mathrm{p}}\sigma_h^2(\mathbf{x}^{\mathrm{p}})^H+\sigma_w^2\mathbf{I}, \end{aligned} \tag{3.19}$$

$$\begin{aligned}\mathbf{c}_{hy^{\mathrm{p}}} &= \mathbb{E}\{h(\mathbf{y}^{\mathrm{p}})^H\} \\ &= \mathbb{E}\{h(h\mathbf{x}^{\mathrm{p}}+\mathbf{w}^{\mathrm{p}})^H\} \\ &= \sigma_h^2(\mathbf{x}^{\mathrm{p}})^H. \end{aligned} \tag{3.20}$$

By inserting (3.19) and (3.20) in (3.17) and (3.18) we get the final expressions

$$\hat{h}_{\mathrm{MMSE}} = \sigma_h^2(\mathbf{x}^{\mathrm{p}})^H(\mathbf{x}^{\mathrm{p}}\sigma_h^2(\mathbf{x}^{\mathrm{p}})^H+\sigma_w^2\mathbf{I})^{-1}\mathbf{y}^{\mathrm{p}}, \tag{3.21}$$

$$\sigma^2_{\mathrm{MMSE}} = \sigma_h^2-\sigma_h^2(\mathbf{x}^{\mathrm{p}})^H(\mathbf{x}^{\mathrm{p}}\sigma_h^2(\mathbf{x}^{\mathrm{p}})^H+\sigma_w^2\mathbf{I})^{-1}\mathbf{x}^{\mathrm{p}}\sigma_h^2. \tag{3.22}$$

The equations (3.21) and (3.22) can be further simplified. We begin by first calculating the following expression common in (3.21) and (3.22) as

$$\begin{aligned}\sigma_h^2(\mathbf{x}^{\mathrm{p}})^H(\mathbf{x}^{\mathrm{p}}\sigma_h^2(\mathbf{x}^{\mathrm{p}})^H+\sigma_w^2\mathbf{I})^{-1} &= \frac{1}{\sigma_w^2}\left(\mathbf{I}+\frac{\sigma_h^2}{\sigma_w^2}\mathbf{x}^{\mathrm{p}}(\mathbf{x}^{\mathrm{p}})^H\right)^{-1} \\ &= \frac{\sigma_h^2(\mathbf{x}^{\mathrm{p}})^H}{\sigma_w^2}\left(\mathbf{I}-\frac{\frac{\sigma_h^2}{\sigma_w^2}\mathbf{x}^{\mathrm{p}}(\mathbf{x}^{\mathrm{p}})^H}{1+\frac{\sigma_h^2}{\sigma_w^2}(\mathbf{x}^{\mathrm{p}})^H\mathbf{x}^{\mathrm{p}}}\right) \\ &= \frac{\sigma_h^2}{\sigma_w^2}\left((\mathbf{x}^{\mathrm{p}})^H-\frac{\frac{\sigma_h^2}{\sigma_w^2}PE_p(\mathbf{x}^{\mathrm{p}})^H}{1+\frac{\sigma_h^2}{\sigma_w^2}PE_p}\right) \\ &= \frac{\sigma_h^2}{\sigma_w^2+PE_p\sigma_h^2}\,(\mathbf{x}^{\mathrm{p}})^H \end{aligned} \tag{3.23}$$

where the inverse was calculated using the *Woodbury's identity* [20]. Using (3.23), we can compute a closed expression for $\hat{h}_{\text{MMSE}}$ and covariance σ^2_{MMSE}, as follows

$$\begin{aligned}\hat{h}_{\text{MMSE}} &= \sigma_h^2 \mathbf{I}(\mathbf{x}^{\text{p}})^H (\mathbf{x}^{\text{p}}\sigma_h^2\mathbf{I}(\mathbf{x}^{\text{p}})^H + \sigma_w^2\mathbf{I})^{-1}\mathbf{y}^{\text{p}} \\ &= \frac{\sigma_h^2}{\sigma_w^2 + PE_p\sigma_h^2}\,(\mathbf{x}^{\text{p}})^H\mathbf{y}^{\text{p}}.\end{aligned} \tag{3.24}$$

$$\begin{aligned}\sigma^2_{\text{MMSE}} &= \sigma_h^2 - \sigma_h^2(\mathbf{x}^{\text{p}})^H(\mathbf{x}^{\text{p}}\sigma_h^2(\mathbf{x}^{\text{p}})^H + \sigma_w^2\mathbf{I})^{-1}\mathbf{x}^{\text{p}}\sigma_h^2 \\ &= \sigma_h^2 - \frac{\sigma_h^2}{\sigma_w^2 + PE_p\sigma_h^2}\,(\mathbf{x}^{\text{p}})^H\mathbf{x}^{\text{p}}\sigma_h^2 \\ &= \sigma_h^2 - \frac{\sigma_h^2 PE_p\sigma_h^2}{\sigma_w^2 + PE_p\sigma_h^2} \\ &= \frac{\sigma_h^2\sigma_w^2}{\sigma_w^2 + \sigma_h^2 PE_p}.\end{aligned} \tag{3.25}$$

Now that we have calculated two different types of linear channel estimators, we can generalize linear channel estimation by analyzing the expressions of the LS channel estimate in (3.9) and the MMSE channel estimate in (3.24) as

$$\hat{h} = \alpha(\mathbf{x}^{\text{p}})^H\mathbf{y}^{\text{p}}, \tag{3.26}$$

where α is a constant scaling factor which depends on the type of channel estimator used. The constant α is characteristic for linear channel estimators. For the LS and MMSE channel estimators, α is given by

$$\alpha = \begin{cases} \dfrac{1}{PE_p} & \text{for LS estimation} \\ \dfrac{\sigma_h^2}{(\sigma_w^2 + \sigma_h^2 PE_p)} & \text{for MMSE estimation.} \end{cases} \tag{3.27}$$

By comparing the constant scaling factors for the two linear channel estimators given in (3.27), we notice that the two linear channel estimators can be put in relation to each other through the following equation

$$\hat{h}_{\text{MMSE}} = \frac{\sigma_h^2 PE_p}{(\sigma_w^2 + \sigma_h^2 PE_p)}\,\hat{h}_{\text{LS}}. \tag{3.28}$$

Thus, we have illustrated the similarity of the linear estimators. Linear estimators only differ to each other by a constant factor. For example if we know the LS channel estimate, the MMSE channel estimate can be calculated straight away by using an appropriate biasing factor. However, the channel statistics and the noise variance have to be known beforehand. Similar results have also been shown in [5].

It is useful to know this fact about linear estimators, when calculating the *posterior* pdf of the channel given its estimate, see Subsection 4.2. Linear estimators assure that this *posterior* pdf will remain unchanged in their distribution, because they only differ by a constant factor. In Subsection 4.2, we use the general form of the linear channel estimator in (3.26). We then prove the independence of this *posterior* pdf from the type of linear channel estimator used.

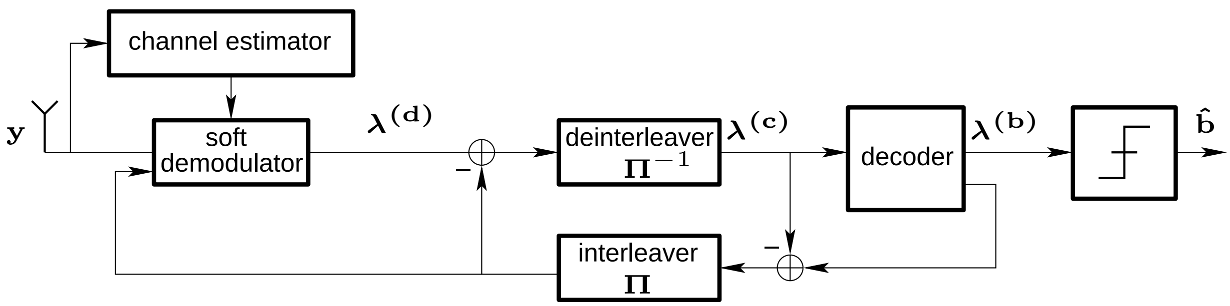


Figure 3.2: BICM-ID Receiver

3.3 BICM-ID Receiver

The BICM-ID receiver is composed of the following signal processing blocks also shown in Fig. 3.2. Detailed analytic descriptions can be found in the corresponding subsections. A short overview of the receiver is given here.

In the first stage at the receiver, the channel estimator utilizes the known pilots and the received pilot symbol sequence $\mathbf{y}^{\mathrm{p}}$ to estimate the channel h. The channel estimation can be done using linear estimation techniques like LS or MMSE estimation as explained in section 3.2. The soft demodulator signal processing block is of main interest in this work and will be investigated in more detail in Chapter 4. Here, the received symbol sequences are de-mapped to bit sequences. It delivers soft information about the individual bit probabilities in the form of log-likelihood ratios $\boldsymbol{\lambda}$ for each transmitted bit, see Subsection 3.3.2. The soft demodulator also receives feedback from the decoder carrying information about the coded bits and uses it as *a-priori* information for the de-mapping of the bits.

The exact inverse operation of the pseudo-random interleaver from Subsection 3.1 is done by the deinterleaver $\boldsymbol{\Pi}^{-1}$ at the receiver. The deinterleaver is necessary to put the log-likelihood ratios of the bit sequences in correct order, as expected by the decoder. The interleaver and the deinterleaver are critical for the performance of BICM-ID [9], because the interleaver introduces statistical independency between the encoded bits, which allows splitting up of the probabilities.

The decoder resolves the code, e.g., a convolutional code from the encoder and can be implemented with the BCJR-algorithm (cf. [21]). The BCJR algorithm computes the updated bit probabilities using the whole sequence and taking the code structure into account. As in Fig. 3.2, the decoder provides two outputs. The soft information for the decoded bits is sent to the slicer, where the received bit sequence $\hat{\mathbf{b}}$ is constructed. The second output delivers the *extrinsic* information for the de-mapped bits and is sent to the interleaver. This interleaved sequence serves as *a-priori* information for the soft-demodulator.

The exchange of information between the decoder and the soft demodulator happens every iteration or cycle. The decoded bit sequence for each iteration is calculated by the slicer. It applies decision on the incoming soft information from the decoder, based on the magnitude and sign of the log-likelihood ratios.

3.3.1 Maximum Likelihood Detection

To describe BICM-ID demodulation procedure, we first need to lay down some basic detection methods, which are based on the *Maximum likelihood* (ML) or the *Maximum a posteriori probability* (MAP) criterion [17]. In the case of uniformly distributed transmitted information bits or symbols, the two criterion are equivalent to each other. The MAP criterion maximizes the *a-posteriori* probability of detecting a bit sequence $\mathbf{b}$ given the received sequence $\mathbf{y}$ and is formulated by the following equation

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmax}}\, f(\mathbf{b}|\mathbf{y}). \tag{3.29}$$

The maximization process involves an exhaustive search over all possible combinations of bit sequences, which is exponentially complex in the length of the bit sequence. Due to this fact, it is practically impossible to implement the MAP criterion in (3.29). Thus, it is not possible to evaluate $f(\mathbf{b}|\mathbf{y})$ for each and every realization of the bit sequence $\mathbf{b}$. We will see in Subsection 3.3.3 how an alternative iterative algorithm is implemented for the demodulation and decoding [9], [17].

3.3.2 Log Likelihood Ratios

The iterative decoding procedure exchanges soft information between the demodulating and the decoding sub-blocks [19], [22]. In each iteration the demodulating and decoding sub-blocks utilize the *a-posteriori* information from each other as an additive information to get closer to the optimal criterion [17]. As the iterative decoding procedure for soft-in/soft-out demodulators and decoders has to deal with *a-posteriori* probabilities, it is convenient to define something called as the log-likelihood ratio (LLR). LLRs are the basic tools which prove to be very useful in describing iterative schemes [19]. Soft information is carried by the means of LLRs between the sub-blocks of the BICM receiver. If b be a binary random variable from the set $\{0, 1\}$, then the log-likelihood ratio of the random variable b is defined as

$$\lambda = \log \frac{p(b = 1)}{p(b = 0)},$$

where $p(b = 1)$ denotes the probability that the random variable b takes the value $b = 1$. It is possible to tell the value of the bit by looking at the sign of the LLR, i.e., a positive sign of the LLR would mean that the probability of the bit b being "1" is greater than the probability of b being a "0". The magnitude of the LLR shows the reliability of this decision. The larger the magnitude of LLR, the more certain is the decision. Probabilities $p(b = 0)$ or $p(b = 1)$ carry the all the soft information about the bit b. The soft information of the bit b is denoted by λ [22].

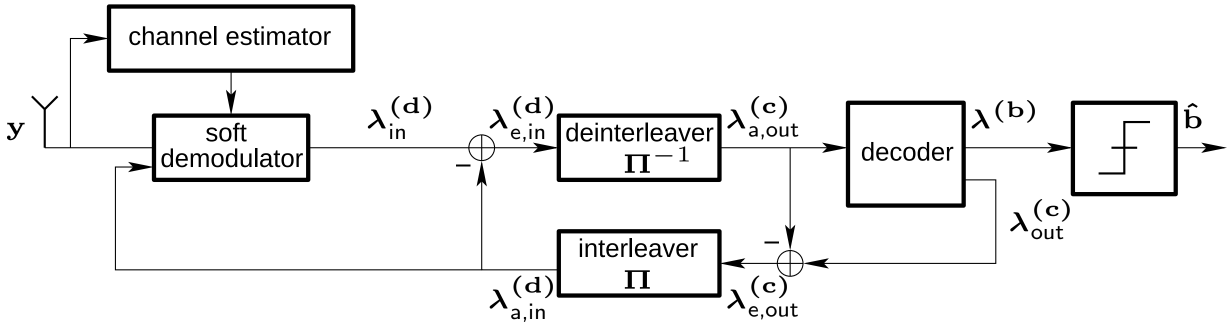


Figure 3.3: BICM Receiver

3.3.3 BICM-ID Demodulation & Decoding

In this subsection we describe how the iterative decoding procedure of a BICM-ID works. The fundamental task is to decode the received data symbol sequence $\mathbf{y}^{\mathrm{d}} = [y_1^{\mathrm{d}} \quad \cdots \quad y_N^{\mathrm{d}}]^T$ into the bit sequence $\hat{\mathbf{b}} = [\hat{b}_1 \quad \cdots \quad \hat{b}_K]^T$. We will use the MAP or ML criterion for detection in (3.29).

We know that there is a one to one correspondence between the bit sequences $\mathbf{b}$ and the encoded bit sequence $\mathbf{c}$. Therefore maximizing over the coded bits would be an equivalent way of maximizing the *a-posteriori* probability (APP). This has to be done only over the set of valid codewords $\mathbf{c} \in \mathcal{C}$

$$\hat{\mathbf{c}} = \arg\max_{\mathbf{c}\in\mathcal{C}} f(\mathbf{c}|\mathbf{y}^{\mathrm{d}}).$$

Assuming the interleaver to be linear and ideal the calculation of the APPs can also be done over the interleaved bit sequence $\mathbf{d}$ as

$$\hat{\mathbf{d}} = \arg\max_{\mathbf{d}\in\mathcal{D}}, f(\mathbf{d}|\mathbf{y}^{\mathrm{d}})$$

where the set $\mathcal{D}$ represents the set of codewords corresponding to the set of valid codewords $\mathcal{C}$ after interleaving. With the help of the Bayes's theorem [20], we can further simplify the *posterior* distribution into a more practical form as

$$f(\mathbf{d}|\mathbf{y}^{\mathrm{d}}) = \frac{f(\mathbf{y}^{\mathrm{d}}|\mathbf{d})f(\mathbf{d})}{f(\mathbf{y}^{\mathrm{d}})}.$$

We can further assume that the bits are uniformly distributed and independent. This would mean that we can omit $f(\mathbf{d})$ for the maximization process as it is constant and

$f(\mathbf{y}^{\mathrm{d}})$ can also be omitted as it is independent of $\mathbf{d}$. That means

$$f(\mathbf{d}|\mathbf{y}^{\mathrm{d}}) \propto f(\mathbf{y}^{\mathrm{d}}|\mathbf{d}) = \prod_{n=1}^{N} f(y_n^{\mathrm{d}}|\mathbf{d}_n),$$

where $\mathbf{d}_n = [d_n \quad d_{n+1} \quad \cdots \quad d_{n+m-1}]^T$, see section 3.1. The interleaved bit subsequence $\mathbf{d}_n$ corresponds to the symbol x_n^{d} and we can equivalently formulate $f(\mathbf{x}^{\mathrm{d}}|\mathbf{y}^{\mathrm{d}})$ as

$$f(\mathbf{x}^{\mathrm{d}}|\mathbf{y}^{\mathrm{d}}) \propto f(\mathbf{y}^{\mathrm{d}}|\mathbf{x}^{\mathrm{d}}) = \prod_{n=1}^{N} f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}).$$

The demodulator incorporates soft information (APPs of the mapped bits) from the decoder and the decoder also incorporates the soft information (APPs of the encoded bits) provided by the demodulator. Information is exchanged in an iterative way between the demodulator and the decoder until the desired performance is achieved [17], [19]. The information is exchanged with the help of LLRs, see Section 3.3.2.

Figure 3.3 shows the flow of soft information during the iterative demodulation and decoding process. The soft-demodulator and the decoder are the two main working sub-blocks of the receiver. Here, the vector $\boldsymbol{\lambda} = [\lambda_{1,1}\, \lambda_{1,2} \cdots \lambda_{1,m} \quad \lambda_{2,1}\, \lambda_{2,2} \cdots \lambda_{2,m} \quad \cdots \quad \lambda_{N,1}\, \lambda_{N,2} \cdots \lambda_{N,m}]$ contains the log-likelihood ratios of the whole bit sequence. The cardinality of the LLR vector $\boldsymbol{\lambda}$ can be calculated as $|\boldsymbol{\lambda}| = N\,m$. With $k \in \{1, \cdots, m\}$ and $n \in \{1, \cdots, N\}$, each element of the vector $\boldsymbol{\lambda}$ is represented by the likelihood ratio $\lambda_{n,k}$. This likelihood ratio belongs to the k-th bit of the n-th symbol. The superscripts in the Fig. 3.3 denote the corresponding type of soft information the vector $\boldsymbol{\lambda}$ is carrying. The soft information regarding the interleaved bit sequence is denote by $\boldsymbol{\lambda}^{(\mathbf{d})}$, for the coded bit sequence by $\boldsymbol{\lambda}^{(\mathbf{c})}$ and for the actual data bit sequence $\boldsymbol{\lambda}^{(\mathbf{b})}$.

The soft-demodulator has knowledge of the received symbol sequence $\mathbf{y}$ and takes the *a-priori* knowledge $\boldsymbol{\lambda}^{(\mathbf{d})}_{\mathrm{a,in}}$ of the mapped bits after the interleaver and calculates the so called *extrinsic* information $\boldsymbol{\lambda}^{(\mathbf{d})}_{\mathrm{e,in}}$ for all the encoded bits. This *extrinsic* information is then deinterleaved to produce $\boldsymbol{\lambda}^{(\mathbf{c})}_{\mathrm{a,out}}$ and sent to the soft-in/soft-out MAP decoder (uses the BCJR algorithm cf. [21]) as an *a-priori* input. The decoder in turn calculates the *extrinsic* information $\boldsymbol{\lambda}^{(\mathbf{c})}_{\mathrm{e,out}}$ on the coded bits. The LLR $\boldsymbol{\lambda}^{(\mathbf{c})}_{\mathrm{e,out}}$ is interleaved and produces $\boldsymbol{\lambda}^{(\mathbf{d})}_{\mathrm{a,in}}$, to serve as an *a-priori* input for the soft-demodulator. This completes an iteration. With every iteration the bit-error rate (BER) decreases, because with every iteration the *a-priori* information for the bits gets updated and thus becomes more reliable.

As we can see from the Fig. 3.3 that the decoder also has a second output, which delivers the LLR vector $\boldsymbol{\lambda}^{(\mathbf{b})}$ corresponding to the decoded bits. Finally, the Slicer makes decisions based on the sign and magnitude of each element in the LLR vector $\boldsymbol{\lambda}^{(\mathbf{b})}$ and produces the received bit sequence $\hat{\mathbf{b}}$.

We can now have a look at the log-likelihood algebra, with the help of which all operations of this iterative decoding procedure are done. Calculations with APPs using the log-likelihood ratios is a convenient method to describe iterative decoding algorithms. The separation of the *a-priori* information from the *extrinsic* information can be done by simple addition or subtraction of the LLRs.

First we will calculate the log-likelihood ratio $\lambda_{n,k}$. The use of conditioned LLRs is fully equivalent to the use of *posterior* probabilities. Each conditioned LLR can be separated into two entities, one contains the *a-priori* information and the second the *extrinsic* information [22]. We have

$$\begin{aligned}\lambda(b_k|\mathbf{y}^{\mathrm{d}},h) &= \log\frac{P(b_k=1|\mathbf{y}^{\mathrm{d}},h)}{P(b_k=0|\mathbf{y}^{\mathrm{d}},h)},\\ &= \underbrace{\log\frac{P(b_k=1)}{P(b_k=0)}}_{\textit{a-priori}\ \lambda_a^k}+\underbrace{\log\frac{f(\mathbf{y}^{\mathrm{d}}|b_k=1,h)}{f(\mathbf{y}^{\mathrm{d}}|b_k=0,h)}}_{\textit{extrinsic}\ \lambda_e^k}.\end{aligned}$$

For calculation purposes, we will denote the *a-priori* information as λ_a^k and the *extrinsic* information as λ_e^k. By splitting up the probabilities λ_e^k can further be expressed as

$$\lambda_{n,k} = \lambda_a^k+\log\frac{\sum\limits_{x_n^{\mathrm{d}}\in\mathbb{X}_k^{(1)}} f(y_n^{\mathrm{d}}|h,x_n^{\mathrm{d}})f(x_n^{\mathrm{d}})}{\sum\limits_{x_n^{\mathrm{d}}\in\mathbb{X}_k^{(0)}} f(y_n^{\mathrm{d}}|h,x_n^{\mathrm{d}})f(x_n^{\mathrm{d}})}, \tag{3.30}$$

where $\mathbb{X}_k^{(b)}$ is the set of all symbols whose bit labels at position k equals $b \in \{0,1\}$. Further, by exploiting the independence of the bits due to the assumption of an ideal interleaver, and we can express (3.30) as

$$\lambda_{n,k} \quad = \quad \lambda_a^k + \log \frac{\sum\limits_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(1)}} f(y_n^{\mathrm{d}}|h, x_n^{\mathrm{d}}) \prod\limits_{j \neq k} p(b_j)}{\sum\limits_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(0)}} f(y_n^{\mathrm{d}}|h, x_n^{\mathrm{d}}) \prod\limits_{j \neq k} p(b_j)}. \tag{3.31}$$

It is possible to express the individual bit probabilities as a function of a log-likelihood ratio. For the bit b_j we have the log-likelihood ratio $\lambda^j = \dfrac{p(b_j = 1)}{p(b_j = 0)}$. By applying $p(b_j = 0) = 1 - p(b_j = 1)$ we can express λ^j as

$$\lambda^j = \log \frac{p(b_j = 1)}{1 - p(b_j = 1)} \quad \text{or} \quad = \log \frac{1 - p(b_j = 0)}{p(b_j = 0)}. \tag{3.32}$$

It is now possible to express the individual bit probabilities $p(b_j)$ by using (3.32). We have

$$p(b_j = 0) = \frac{1}{1 + e^{\lambda^j}} \quad \text{and} \quad p(b_j = 1) = \frac{e^{\lambda^j}}{1 + e^{\lambda^j}}. \tag{3.33}$$

Moreover (3.33) can be generalized for any bit value b as

$$p(b_j = b) = \frac{e^{b\lambda^j}}{1 + e^{\lambda^j}}. \tag{3.34}$$

It is to be noted that the denominator in (3.34) is independent of the actual value of the bit b_j, and can thus be ignored when calculating the product term $\prod\limits_{j \neq k} p(b_j)$. We further on multiply the numerator and denominator in (3.31) with $\prod\limits_{j \neq k} \exp(-\frac{1}{2}\lambda^j)$ and get the expression

$$\lambda_{n,k} \quad = \quad \lambda_a^k + \log \frac{\sum\limits_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(1)}} f(y_n^{\mathrm{d}}|h, x_n^{\mathrm{d}}) \prod\limits_{j \neq k} \exp\left(\frac{1}{2}(2b_j - 1)\lambda^j\right)}{\sum\limits_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(0)}} f(y_n^{\mathrm{d}}|h, x_n^{\mathrm{d}}) \prod\limits_{j \neq k} \exp\left(\frac{1}{2}(2b_j - 1)\lambda^j\right)}, \tag{3.35}$$

where pdf of the likelihood function $f(y_n^{\mathrm{d}}|h, x_n^{\mathrm{d}})$ in (3.35) has a *complex Gaussian* distribution $\mathcal{CN}(hx_n^{\mathrm{d}}, \sigma_w^2)$, which follows from the system model in (3.3) and is given by

$$f(y_n^{\mathrm{d}}|h, x_n^{\mathrm{d}}) = \frac{1}{\pi\sigma_w^2} \exp\left(-\frac{|y_n^{\mathrm{d}} - hx_n^{\mathrm{d}}|^2}{\sigma_w^2}\right). \tag{3.36}$$

The likelihood function in (3.36) in can be easily incorporated in (3.35) to give the final expression for the LLR $\lambda_{n,k}$ as

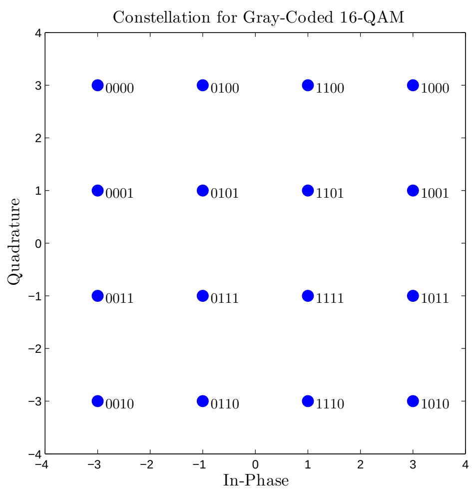
$$\lambda_{n,k} = \lambda_a^k + \log \frac{\sum_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(1)}} \exp\left\{-\frac{|y_n^{\mathrm{d}} - hx_n^{\mathrm{d}}|^2}{\sigma_w^2} + \frac{1}{2}\sum_{j \neq k}(2b_j - 1)\lambda^j\right\}}{\sum_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(0)}} \exp\left\{-\frac{|y_n^{\mathrm{d}} - hx_n^{\mathrm{d}}|^2}{\sigma_w^2} + \frac{1}{2}\sum_{j \neq k}(2b_j - 1)\lambda^j\right\}}. \tag{3.37}$$

The above expression is generally approximated via the *max-log* approximation [19], see Section 4.3. Equation (3.37) can be used for both types of LLRs, i.e., the ones carrying soft information on the mapped bits and the ones carrying soft information on the coded bits. For the mapped bits, λ_a^k will be an element of the *a-priori* LLR vector $\boldsymbol{\lambda}_{\mathrm{a,in}}^{(\mathbf{d})}$ and λ_e^k an element of the *extrinsic* LLR $\boldsymbol{\lambda}_{\mathrm{e,in}}^{(\mathbf{d})}$. Both being for the kth position. This also applies in the same sense to the coded bits.

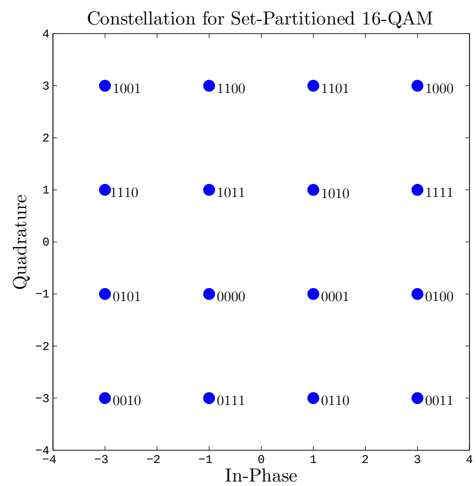
It is to be noted that during all calculations and system description, the channel h was considered as perfectly known. This is an assumption, which in reality would be incorrect. In practice, an estimate of the channel with the methods described in Subsection 3.2 is used instead. This will definitely cause performance degradation, because of channel estimation errors. The main concern of this thesis is to investigate methods, which are useful in improving the performance. And hereby reducing the performance loss due to channel estimation. We will see in the next chapters, how the performance can be improved even in the presence of channel estimation errors.

3.4 Symbol Constellations

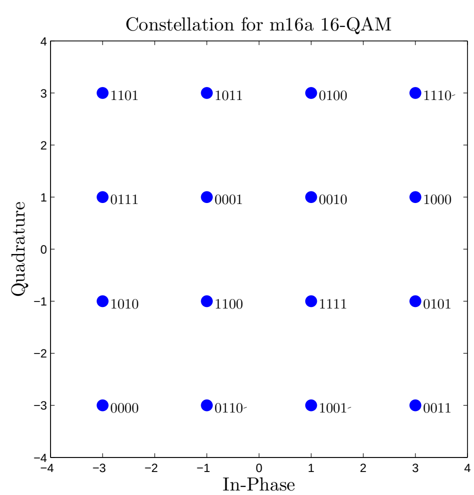
The BICM-ID system performance also depends upon the type of bit-to-symbol mapping used. Some constellations used with BICM-ID are given here for a 16QAM modulation.



(a) Gray-coded labeling



(b) Set-partitioned labeling



(c) m16a labeling

For an iterative system set-partitioning performs much better than Gray-coded constellation [8], [23]. The authors in [24] have described methods to derive an optimal bit labeling for BICM-ID with quadratic QAM constellations. (cf. [24], [23])

4

Modified Receiver for Iterative Systems

"[. . .] The enchanting charms of this sublime science reveal only to those who have the courage to go deeply into it." —— Carl Friedrich Gauss

IN this chapter we take a closer look at the *mismatched* and *modified* receiver designs for a BICM-ID system model. The use of these receivers is not limited to the BICM-ID system. The ML metrics of these two receivers are discussed here. In particular, the derivation of the *modified* ML metric is explained in detail. The *modified receiver* utilizes the statistics of the CEE and is thus able to adapt itself to the presence of CEE. This is a significant advantage over the *mismatched receiver*, which is unable to do so. The ML metrics of the *mismatched* and *modified* receivers are first discussed for a hard symbol decision decoding and are then adapted for soft decision bit-wise decoding in the BICM-ID system from Chapter 3. The simulation results at the end of the chapter show that the *modified receiver* outperforms the *mismatched receiver*.

4.1 Mismatched Receiver

In conventional design, the receiver expects the exact knowledge of the CSI for its ML based detector metric. However in practice, the receiver only has an imperfect channel estimate available. The channel estimate is imperfect due to the fact that only a limited number of pilots can be used to estimate the channel [2], (cf. Section 3.2). It is thus not possible to eliminate the CEE completely. Therefore, the *mismatched receiver* is suboptimal. The true channel is replaced by its noisy estimate in the receiver metric [2], which was originally designed for perfect CSI.

First we calculate the ML metric of the *ideal receiver*, under the assumption of a perfectly known channel h using the system model in (3.3). This is done by maximizing the pdf of the likelihood function $f(\mathbf{x}^{\mathrm{d}}|\mathbf{y}^{\mathrm{d}}, h)$. The likelihood function $f(\mathbf{x}^{\mathrm{d}}|\mathbf{y}^{\mathrm{d}}, h)$ can be approximated with the Bayes's theorem [20] as

$$\begin{aligned} f(\mathbf{x}^{\mathrm{d}}|\mathbf{y}^{\mathrm{d}}, h) &\propto f(\mathbf{y}^{\mathrm{d}}|\mathbf{x}^{\mathrm{d}}, h) \\ &= \prod_{n=1}^{N} f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, h). \end{aligned} \tag{4.1}$$

The pdf $f(\mathbf{y}^{\mathrm{d}}|\mathbf{x}^{\mathrm{d}}, h)$ in (4.1) for the ML detection can be factorized because the receiver does not take the channel correlation into account. This property allows symbol-wise ML detection at the receiver. With the pdf $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, h)$ distributed as $\mathcal{CN}(hx_n^{\mathrm{d}}, \sigma_w^2)$ given in (3.36), the ML decision rule for the *ideal receiver* is formulated as

$$\begin{aligned} \hat{x}_{n,\,\mathrm{ML}}^{\mathrm{d}} &= \operatorname*{argmax}_{x_n^{\mathrm{d}} \in \mathcal{A}} f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, h) \\ &= \operatorname*{argmax}_{x_n^{\mathrm{d}} \in \mathcal{A}} \ln f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, h) \\ &= \operatorname*{argmin}_{x_n^{\mathrm{d}} \in \mathcal{A}} \left(|y_n^{\mathrm{d}} - hx_n^{\mathrm{d}}|^2\right). \end{aligned} \tag{4.2}$$

By using the LS or MMSE channel estimate $\hat{h}$ from (3.10) or (3.24) in(4.2) we can write the ML decision rule for the *mismatched receiver* as

$$\begin{aligned} \hat{x}_{n,\,\mathrm{ML}}^{\mathrm{d}} &= \operatorname*{argmax}_{x_n^{\mathrm{d}} \in \mathcal{A}} f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, h)\,|_{h=\hat{h}} \\ &= \operatorname*{argmin}_{x_n^{\mathrm{d}} \in \mathcal{A}} \left(|y_n^{\mathrm{d}} - \hat{h}x_n^{\mathrm{d}}|^2\right). \end{aligned} \tag{4.3}$$

There are several possible *mismatched receivers* employing the metric in (4.3) depending on the type of channel estimation used. The use of the channel estimate in the ML metric of the *ideal receiver* causes a mismatch leading to performance degradation. It is not designed to deal with CEE. We will see in Section 4.2, how a *modified receiver* can be derived to overcome this problem.

4.1.1 BICM-ID with Mismatched Metric

The mismatched metric in (4.3) can easily be integrated in to the BICM-ID receiver in 3.3.3. We know that the mismatched metric is a result of replacing the true channel with its estimate in the metric of the *ideal receiver*. This allows us to formulate the LLR $\lambda_{n,k}^{mis}$ for the *mismatched receiver* employing iterative decoding by simply replacing h with $\hat{h}$ in the expression of the LLR for the *ideal receiver* in (3.37). The LLR for the mismatched receiver is thus given by

$$\lambda_{n,k}^{mis} = \lambda_a^k + \log \frac{\sum\limits_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(1)}} \exp\left\{ -\frac{|y_n^{\mathrm{d}} - \hat{h} x_n^{\mathrm{d}}|^2}{\sigma_w^2} + \frac{1}{2}\sum\limits_{j \neq k} (2b_j - 1)\lambda^j \right\}}{\sum\limits_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(0)}} \exp\left\{ -\frac{|y_n^{\mathrm{d}} - \hat{h} x_n^{\mathrm{d}}|^2}{\sigma_w^2} + \frac{1}{2}\sum\limits_{j \neq k} (2b_j - 1)\lambda^j \right\}}. \tag{4.4}$$

4.2 Modified Receiver

The ML metric can be adapted to the presence of CEE by utilizing its statistics. It further uses the *posterior* pdf of the channel conditioned on its estimate [2]. This method allows the receiver to have better knowledge of the channel statistics.

4.2.1 Posterior Channel Statistics

The *modified receiver* estimates the channel by using LS or MMSE estimators in Section 3.2. For illustrative purposes we use the general form of a linear channel estimator given in (3.26), i.e., $\hat{h} = \alpha(\mathbf{x}^{\mathrm{p}})^H \mathbf{y}^{\mathrm{p}}$. This will later on help us to understand the working of the receiver better.

Using the general form of the linear estimator the *posterior* pdf of the channel given its estimate, i.e., $f(h|\hat{h})$, is derived. This is calculated with the help of the Bayes's theorem [20] and is given by

$$f(h|\hat{h}) = \frac{f(\hat{h}|h)f(h)}{f(\hat{h})} = \frac{f(\hat{h}|h)f(h)}{\int\limits_h f(\hat{h}|h)f(h)} dh. \tag{4.5}$$

where we already know that $f(h)$ is distributed as $\mathcal{CN}(0,\sigma_h^2)$ and the pdf $f(\hat{h}|h)$ can be derived from the linear estimator in (3.26). Using the system model for the received pilot symbols in (3.4), the channel estimate in the general form can be expressed as

$$\hat{h} \;=\; \alpha(\mathbf{x}^{\mathrm{p}})^H(h\mathbf{x}^{\mathrm{p}} + \mathbf{w}^{\mathrm{p}}). \tag{4.6}$$

For the system model in (3.4) we know that $\mathbf{w}^{\mathrm{p}}$ has a Gaussian distribution. Thus, $f(\hat{h}|h)$ will also be Gaussian distributed. The mean and variance of the pdf $f(\hat{h}|h)$ with the distribution $\mathcal{CN}(\mu_{\hat{h}|h}, \sigma^2_{\hat{h}|h})$ is calculated using (4.6) and is given by

$$\begin{aligned}
\mu_{\hat{h}|h} &= \mathbb{E}\{\hat{h}|h\} \\
&= \mathbb{E}\{\alpha(\mathbf{x}^{\mathrm{p}})^H(h\mathbf{x}^{\mathrm{p}} + \mathbf{w}^{\mathrm{p}})|h\} \\
&= \alpha P E_p h, \qquad (4.7)\\
\sigma^2_{\hat{h}|h} &= \mathbb{E}\{\hat{h}\hat{h}^H|h\} \\
&= \mathbb{E}\{(\alpha(\mathbf{x}^{\mathrm{p}})^H(h\mathbf{x}^{\mathrm{p}} + \mathbf{w}^{\mathrm{p}}))(\alpha(\mathbf{x}^{\mathrm{p}})^H(h\mathbf{x}^{\mathrm{p}} + \mathbf{w}^{\mathrm{p}}))^H|h\} \\
&= \mathbb{E}\{(\alpha(\mathbf{x}^{\mathrm{p}})^H\mathbf{x}^{\mathrm{p}}h + \alpha(\mathbf{x}^{\mathrm{p}})^H\mathbf{w}^{\mathrm{p}})(\alpha(\mathbf{x}^{\mathrm{p}})^H\mathbf{x}^{\mathrm{p}}h + \alpha(\mathbf{x}^{\mathrm{p}})^H\mathbf{w}^{\mathrm{p}})^H|h\} \\
&= \alpha^2 P E_p \sigma_w^2. \qquad (4.8)
\end{aligned}$$

These results are only valid under the assumption that the pilot symbols have equal energy, i.e., $|x_1^{\mathrm{p}}|^2 = |x_2^{\mathrm{p}}|^2 = \cdots = |x_P^{\mathrm{p}}|^2 = E_p$. With this assumption, the term $(\mathbf{x}^{\mathrm{p}})^H\mathbf{x}^{\mathrm{p}}$ reduces to PE_p. Here, P is the number of pilots and E_p is the energy of each pilot. For further calculations we define the signal-to-noise ratio (SNR) as

$$\gamma \triangleq \frac{\sigma_h^2 P E_p}{\sigma_w^2}. \tag{4.9}$$

The product $f(\hat{h}|h)f(h)$ is a simple multiplication of Gaussian distributions, which again yields a Gaussian distribution given by

$$f(\hat{h}|h)f(h) = \frac{1}{\pi^2\alpha^2\gamma\sigma_w^4} \exp\left(-\frac{\alpha^2 PE_p(1+\gamma)\,h^2 - (2\alpha\gamma\hat{h})\,h + \gamma\hat{h}^2/PE_p}{\alpha^2\gamma\sigma_w^2}\right) \tag{4.10}$$

The exponent in (4.10) is expressed as a quadratic function in h. The integral to obtain $f(\hat{h})$ is given by

$$f(\hat{h}) = \int_h f(\hat{h}|h)f(h), \tag{4.11}$$

and we will utilize the integrand calculated in (4.10). The integral in (4.11) can be evaluated by using the following integral identity

$$\int_h \exp(-(ah^2+bh+c))dh = \sqrt{\frac{\pi}{a}}\exp\left(\frac{b^2-4ac}{4a}\right), \tag{4.12}$$

where a, b and c are arbitrary constants. Finally, using (4.10), (4.11) and (4.12) we get the expression for the pdf $f(\hat{h})$ as

$$f(\hat{h}) = \frac{1}{\pi\alpha^2 PE_p\sigma_w^2(1+\gamma)} \exp\left(-\frac{\hat{h}^2}{\alpha^2 PE_p\sigma_w^2(1+\gamma)}\right). \tag{4.13}$$

We calculate the *posterior* pdf using the results in (4.10), (4.13) and standard algebraic calculations. The final expression for $f(h|\hat{h})$ distributed as $\mathcal{CN}(\mu_{h|\hat{h}}, \sigma^2_{h|\hat{h}})$ is easier to analyze by inserting the general form of the linear estimator in the *posterior* pdf, i.e., $\hat{h} = \alpha(\mathbf{x}^{\mathrm{p}})^H\mathbf{y}^{\mathrm{p}}$ from (3.26). The *posterior* pdf of the channel, given its estimate $f(h|\hat{h})$ is given by the distribution

$$\mathcal{CN}\left(\frac{\gamma}{PE_p(1+\gamma)}\,(\mathbf{x}^{\mathrm{p}})^H\mathbf{y}^{\mathrm{p}},\ \frac{\sigma_h^2}{1+\gamma}\right), \tag{4.14}$$

with the mean and variance defined as

$$\mu_{h|\hat{h}} = \frac{\gamma}{PE_p(1+\gamma)}\,(\mathbf{x}^{\mathrm{p}})^H\mathbf{y}^{\mathrm{p}} \tag{4.15}$$

$$\sigma^2_{h|\hat{h}} = \frac{\sigma_h^2}{1+\gamma} \tag{4.16}$$

By observing the expression for the *posterior* pdf in (4.14), we conclude an important result. The pdf $f(h|\hat{h})$ is independent of the constant α and therefore independent of the type of linear channel estimator used. Similar results have also been shown in [5] but for a different system model. For high SNR, i.e., $\gamma \rightarrow \infty$ we make the following conclusions for the mean and variance in (4.15) and (4.16).

$$\begin{cases} \lim\limits_{\gamma\rightarrow\infty} \mu_{h|\hat{h}} = \lim\limits_{\gamma\rightarrow\infty} \dfrac{\gamma}{PE_p(1+\gamma)} \left(\mathbf{x}^{\mathrm{p}}\right)^H \mathbf{y}^{\mathrm{p}} = \dfrac{\left(\mathbf{x}^{\mathrm{p}}\right)^H \mathbf{y}^{\mathrm{p}}}{PE_p} \\ \\ \lim\limits_{\gamma\rightarrow\infty} \sigma^2_{h|\hat{h}} = \lim\limits_{\gamma\rightarrow\infty} \dfrac{\sigma_h^2}{1+\gamma} = 0. \end{cases} \tag{4.17}$$

It can be seen in (4.17) that the mean of the *posterior* pdf converges to the LS estimate in (3.10) and the variance converges down to zero. This means that the distribution of the *posterior* pdf $f(h|\hat{h})$ in (4.14) will become concentrated around the LS channel estimate for an increasing SNR value. Thus, we conclude that the mean of the *posterior* pdf is a scaled LS estimate. For higher values of SNR the channel estimation will become more accurate, because the error variance of the CEE as given in (3.11) tends to zero.

4.2.2 Modified ML Detection

The ML decision rule for the *modified receiver* can be formulated as

$$\begin{aligned} \hat{x}^{\mathrm{d}}_{n,\,\mathrm{ML}} &= \underset{x_n^{\mathrm{d}}\in\mathcal{A}}{\operatorname{argmax}}\, f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},\hat{h}) \\ &= \underset{x_n^{\mathrm{d}}\in\mathcal{A}}{\operatorname{argmax}} \ln f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},\hat{h}). \end{aligned} \tag{4.18}$$

It is to be noted that the likelihood function $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},\hat{h})$ is conditioned on the channel estimate. Thus, it is now possible to incorporate the channel statistics expressed as the *posterior* pdf of the channel in (4.14). To derive the *modified* ML metric, the likelihood function $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},\hat{h})$ needs to be calculated explicitly. The pdf $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},\hat{h})$ can be expressed as a marginalization of the joint likelihood function, with respect to the true channel and is given by

$$\begin{aligned} f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},\hat{h}) &= \int\limits_h f(y_n^{\mathrm{d}},h|x_n^{\mathrm{d}},\hat{h})dh \\ &= \int\limits_h f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},h)f(h|\hat{h})dh. \end{aligned} \tag{4.19}$$

The last step in (4.19) follows from the statistical independency of x^{d} from h and $\hat{h}$. Equation (4.19) can be interpreted as the following conditional expectation calculated over the *posterior* distribution of the true channel given its estimate [2] as

$$f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},\hat{h}) = \mathbb{E}_{h|\hat{h}}\{f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},h)|\hat{h}\}.$$

The product of the two known Gaussian distributions in the integrand of (4.19) remains Gaussian. The pdf $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},h)$ is given by the distribution $\mathcal{CN}(hx_n^{\mathrm{d}},\sigma_w^2)$ and $f(h|\hat{h})$ is to be taken from (4.14). After calculating the product of the Gaussian distributions, the exponent of the integrand in (4.19) can be formulated as a quadratic function in h and is given by

$$\begin{aligned} f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},h)f(h|\hat{h}) &= \\ \frac{1}{\pi^2\sigma_{h|\hat{h}}^2\sigma_w^2}\exp&\left(-\frac{(\sigma_{h|\hat{h}}^2|x_n^{\mathrm{d}}|^2+\sigma_w^2)h^2-2(y^{\mathrm{d}}x_n^{\mathrm{d}}\sigma_{h|\hat{h}}^2+\mu_{h|\hat{h}}\sigma_w^2)h+\sigma_{h|\hat{h}}^2y^{\mathrm{d}}+\mu_{h|\hat{h}}^2\sigma_w^2}{\sigma_{h|\hat{h}}^2\sigma_w^2}\right). \end{aligned} \tag{4.20}$$

It is now possible to apply the integral identity in (4.12) on (4.19) and after standard algebraic calculations, we get the likelihood function $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}},\hat{h})$ for the *modified receiver*, given by the distribution

$$\mathcal{CN}\left(\mu_{h|\hat{h}}x_n^{\mathrm{d}},\sigma_w^2+\sigma_{h|\hat{h}}^2|x_n^{\mathrm{d}}|^2\right) = \mathcal{CN}\left(\frac{\gamma\,(\mathbf{x}^{\mathrm{p}})^H\mathbf{y}^{\mathrm{p}}}{PE_p(1+\gamma)}\,x_n^{\mathrm{d}},\,\sigma_w^2+\frac{\sigma_h^2|x_n^{\mathrm{d}}|^2}{1+\gamma}\right), \tag{4.21}$$

where $\mu_{h|\hat{h}}$ and $\sigma_{h|\hat{h}}^2$ are given in (4.15) and (4.16), respectively. For the sake of simplicity we denote the mean and variance of the likelihood function $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, \hat{h})$ for the *modified receiver* as

$$\mu_{\mathrm{mod}} = \frac{\gamma\,(\mathbf{x}^{\mathrm{p}})^H \mathbf{y}^{\mathrm{p}}}{PE_p(1+\gamma)}\, x_n^{\mathrm{d}} \tag{4.22}$$

$$\sigma_{\mathrm{mod}}^2 = \sigma_w^2 + \frac{\sigma_h^2 |x_n^{\mathrm{d}}|^2}{1+\gamma} \tag{4.23}$$

It is to be noted that the mean and variance of the likelihood function are not constants, since they depend on the data symbol x_n^{d}. The final decision metric for the *modified receiver* is evaluated using (4.21) in the decision rule (4.18) and is given by

$$\hat{x}_{n,\,\mathrm{ML}}^{\mathrm{d}} = \underset{x_n^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmin}} \left(\ln\left(\sigma_w^2 + \frac{\sigma_h^2|x_n^{\mathrm{d}}|^2}{1+\gamma}\right) + \frac{\left| y^{\mathrm{d}} - \dfrac{\gamma\,(\mathbf{x}^{\mathrm{p}})^H \mathbf{y}^{\mathrm{p}}}{PE_p(1+\gamma)}\, x_n^{\mathrm{d}} \right|^2}{\left(\sigma_w^2 + \dfrac{\sigma_h^2|x_n^{\mathrm{d}}|^2}{1+\gamma}\right)} \right). \tag{4.24}$$

The *modified* ML metric can be analyzed for extreme cases of SNR. This results in the following two possibilities:

a) In the case of perfect CSI ($\hat{h} \longrightarrow h$) corresponding to $\gamma \longrightarrow \infty$, the *modified* metric reduces to the Euclidean distance metric of the *ideal receiver* in (4.2)

$$\hat{x}_{n,\,\mathrm{ML}}^{\mathrm{d}} = \underset{x_n^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmin}} \left(\left| y^{\mathrm{d}} - \frac{(\mathbf{x}^{\mathrm{p}})^H \mathbf{y}^{\mathrm{p}}}{PE_p}\, x_n^{\mathrm{d}} \right|^2 \right). \tag{4.25}$$

b) In the case of $\gamma \longrightarrow 0$, the metric reduces to the following expression

$$\hat{x}_{n,\,\mathrm{ML}}^{\mathrm{d}} = \underset{x_n^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmin}} \left(\ln\left(\sigma_w^2 + \sigma_h^2|x_n^{\mathrm{d}}|^2\right) + \frac{\left|y^{\mathrm{d}}\right|^2}{(\sigma_w^2 + \sigma_h^2|x_n^{\mathrm{d}}|^2)} \right). \tag{4.26}$$

It is to be noted that the metric in (4.26) does not contain the channel estimate. In b) it can be seen that the variance of the likelihood function $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, \hat{h})$ reduces to $\left(\sigma_w^2 + \sigma_h^2|x_n^{\mathrm{d}}|^2\right)$ for $\gamma \longrightarrow 0$. The presence of the channel variance σ_h^2 in (4.26) can be seen. Thus, we conclude that the *modified* metric takes the channel statistics into account even when the CSI is so bad that it is absent in the metric (4.26). Similar results have been shown in [5]. The *modified* metric in (4.24) is in a position to adapt itself to the CEE. Moreover, the *modified* metric proves to be independent of the linear channel estimator used because the constant α does not appear in (4.24).

The decision metric in (4.24) is fully equivalent to the one derived in [2]. The *modified* metric in (4.24) can be reformed by inserting the value of the SNR, i.e., $\gamma = \frac{\sigma_h^2 PE_p}{\sigma_w^2}$ and then using the constant $\rho = \frac{PE_p}{PE_p + \sigma_w^2}$. Under the assumption that the distribution of the true channel has a unit variance, i.e., $\sigma_h^2 = 1$, we get exactly the same expression for the decision metric as in [2].

4.2.3 Alternative Derivation of Modified ML Detector

An alternative method to detect the transmitted data is by maximizing the likelihood function $f(y_n^{\mathrm{d}}|x^{\mathrm{d}}, \mathbf{y}^{\mathrm{p}})$. This pdf can be calculated by the following marginalization

$$f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, \mathbf{y}^{\mathrm{p}}) = \int\limits_h f(y_n^{\mathrm{d}}|h, x_n^{\mathrm{d}}) f(h|\mathbf{y}^{\mathrm{p}}) dh, \tag{4.27}$$

where $f(h|\mathbf{y}^{\mathrm{p}})$ is to be calculated using Bayes's theorem [20] as

$$f(h|\mathbf{y}^{\mathrm{p}}) \quad = \quad \frac{f(h|\mathbf{y}^{\mathrm{p}}) f(h)}{f(\mathbf{y}^{\mathrm{p}})},$$

and is given by the distribution

$$\mathcal{CN}\left(\frac{\gamma}{PE_p(1+\gamma)} \left(\mathbf{x}^{\mathrm{p}}\right)^H \mathbf{y}^{\mathrm{p}}, \; \frac{\sigma_h^2}{1+\gamma}\right). \tag{4.28}$$

It can be seen that (4.28) is the same as the *posterior* pdf given in (4.14). This would mean that likelihood function in (4.21) is equivalent to the outcome of (4.27). This can easily be verified by evaluating the integral in (4.27) using the integral identity (4.12). The advantage of this method is that the *posterior* distribution of the true channel given its channel estimate, can directly be calculated with Bayes theorem. Moreover, due to the fact that the pdf $f(h|\hat{h})$ and $f(h|\mathbf{y}^{\mathrm{p}})$ are equivalent, we conclude that the calculation of the *posterior* pdf $f(h|\hat{h})$ with an LS estimate $\hat{h}$ yields the *posterior* pdf $f(h|\mathbf{y}^{\mathrm{p}})$ for an MMSE estimator. Derivation of $f(h|\hat{h})$ can be found in Subsection 4.2.1. In the following subsection we adapt the *modified receiver* for implementation in the BICM-ID system given in Chapter 3.

4.2.4 BICM-ID with Modified Metric

The *modified receiver* can be easily incorporated into the BICM-ID system model. As in Subsection 3.3.3, the likelihood function (3.36) for a perfectly known CSI is inserted in the expression (3.35) for the LLR $\lambda_{n,k}$. We only need to replace this likelihood function with our *modified* likelihood function given in (4.21). The LLR $\lambda_{n,k}^{mod}$ for the *modified receiver* can be formulated by using $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, \hat{h})$ in (3.35) and is given by

$$\lambda_{n,k}^{mod} = \lambda_a^k + \log \frac{\sum\limits_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(1)}} \exp\left\{ -\frac{|y_n - \mu_{\mathrm{mod}}|^2}{\sigma_{\mathrm{mod}}^2} + \frac{1}{2}\sum\limits_{j \neq k}(2b_j - 1)\lambda^j \right\}}{\sum\limits_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(0)}} \exp\left\{ -\frac{|y_n - \mu_{\mathrm{mod}}|^2}{\sigma_{\mathrm{mod}}^2} + \frac{1}{2}\sum\limits_{j \neq k}(2b_j - 1)\lambda^j \right\}}, \quad (4.29)$$

where $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, \hat{h})$ is distributed as $\mathcal{CN}(\mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}}^2)$. The mean μ_{mod} and variance σ_{mod}^2 of the *modified* likelihood function are given in (4.22) and (4.23), respectively. The iterative decoding procedure as described in Section 3.3.3 remains unchanged and we only use a different likelihood function for the soft decision decoding, i.e., the *modified* likelihood function.

4.3 Simulation Results

4.3.1 Implementation Issues

The BICM-ID receiver in Fig. 3.2 has been implemented in MATLAB©. The LLRs carrying soft information of the bit probabilities in this figure are expressed in general by (3.37). For the implementation of the iterative receiver, some approximations for the LLR are calculated as follows

$$\begin{aligned}
\lambda_{n,k} &= \lambda_a^k + \log \frac{\sum_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(1)}} \exp\left\{ -\frac{|y_n^{\mathrm{d}} - h x_n^{\mathrm{d}}|^2}{\sigma_w^2} + \frac{1}{2} \sum_{j \neq k} (2b_j - 1)\lambda^j \right\}}{\sum_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(0)}} \exp\left\{ -\frac{|y_n^{\mathrm{d}} - h x_n^{\mathrm{d}}|^2}{\sigma_w^2} + \frac{1}{2} \sum_{j \neq k} (2b_j - 1)\lambda^j \right\}} \\
&\approx \lambda_a^k + \max_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(1)}} \left\{ -\frac{|y_n^{\mathrm{d}} - h x_n^{\mathrm{d}}|^2}{\sigma_w^2} + \frac{1}{2} \sum_{j \neq k} (2b_j - 1)\lambda^j \right\} \\
&\quad - \max_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(0)}} \left\{ -\frac{|y_n^{\mathrm{d}} - h x_n^{\mathrm{d}}|^2}{\sigma_w^2} + \frac{1}{2} \sum_{j \neq k} (2b_j - 1)\lambda^j \right\} \\
&\approx \lambda_a^k + \min_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(0)}} \left\{ \frac{|y_n^{\mathrm{d}} - h x_n^{\mathrm{d}}|^2}{\sigma_w^2} - \frac{1}{2} \sum_{j \neq k} (2b_j - 1)\lambda^j \right\} \\
&\quad - \min_{x_n^{\mathrm{d}} \in \mathbb{X}_k^{(1)}} \left\{ \frac{|y_n^{\mathrm{d}} - h x_n^{\mathrm{d}}|^2}{\sigma_w^2} - \frac{1}{2} \sum_{j \neq k} (2b_j - 1)\lambda^j \right\}. \qquad (4.30)
\end{aligned}$$

The approximation was done by using the *max-log* approximation of the *Jacobian* logarithm, i.e., $\mathrm{jacln}(a, b) = \ln(e^a + e^a) + \ln(1 - e^{-|a-b|}) \approx \max(a, b)$ [19]. The *max-log* approximation facilitates efficient implementation of the iterative decoding system in MATLAB© with low performance loss.

4.3.2 Performance Evaluation of the BICM-ID Receivers

As discussed in Chapter 3 and Chapter 4, we have three BICM-ID receivers. The *ideal receiver* for perfect CSI, the *mismatched receiver* for imperfect CSI and the *modified receiver* for imperfect CSI with *posterior* channel statistics. We hereby evaluate the performance of these three receivers by analyzing the bit error rate (BER) performance for increasing SNR. The influence of the number of pilots and the pilot symbol energy on the BER is also discussed. These BICM-ID receivers use different metrics in the expression of their LLRs. To summarize, the LLR $\lambda_{n,k}$ used for the *ideal receiver* is in (3.37) and approximated in (4.30). The LLRs for the *mismatched receiver* and for the *modified receiver* are in (4.4) and (4.29), respectively and can be approximated in the same way as shown in (4.30).

4.3.3 Simulation Setup

A sequence of 512 information bits are encoded by a rate $R = 1/2$ (5,7) convolutional code with constraint length 3, where (5,7) represent the generator polynomials. The coded bits are sent through a pseudo-random interleaver. A 16QAM modulation ($M_a = 16$) is used to map the interleaved bits to symbols using a set-partitioning mapping (cf. Section 3.4). We use a i.i.d Rayleigh block fading channel of block length 16 symbols. For each transmitted block we use a different realization of the channel and it remains constant for the whole block. The pilots are inserted in the beginning of each block. A BCJR [21] decoder was used for iterative decoding. We will show the simulation results of a various number of pilots P and pilot energy E_p. The channel estimation is done with an LS estimator.

The SNR in the following figures is defined as

$$\gamma = 10 \log \frac{E_b}{\sigma_w^2} = 10 \log \frac{E_d}{\sigma_w^2} \frac{1}{R\, \mathrm{ld}(M_a)}, \tag{4.31}$$

where σ_w^2 is the noise variance. The respective mean bit and symbol energies are E_b, E_d. For a 16QAM modulation $M_a = 16$.

In Fig. 4.1 we see the BER performance of the receivers for $P = 1$ pilot with energy $E_p = 0.5$ for 10 iterations. The SNR required to achieve a BER $= 10^{-3}$ is about 1.0 dB lower for the *modified receiver* than for the *mismatched receiver*.

Fig. 4.2 shows the performance improvement for the *modified receiver* for an increasing number of iterations for $P = 1$ pilot with mean energy $E_p = 0.5$. The performance improvement is significant for the first few iterations and decreases for larger number of iterations. The type of bit labeling plays an important role in improving the performance over the iterations [9], see Section 3.4. The *ideal* and *mismatched* receivers also

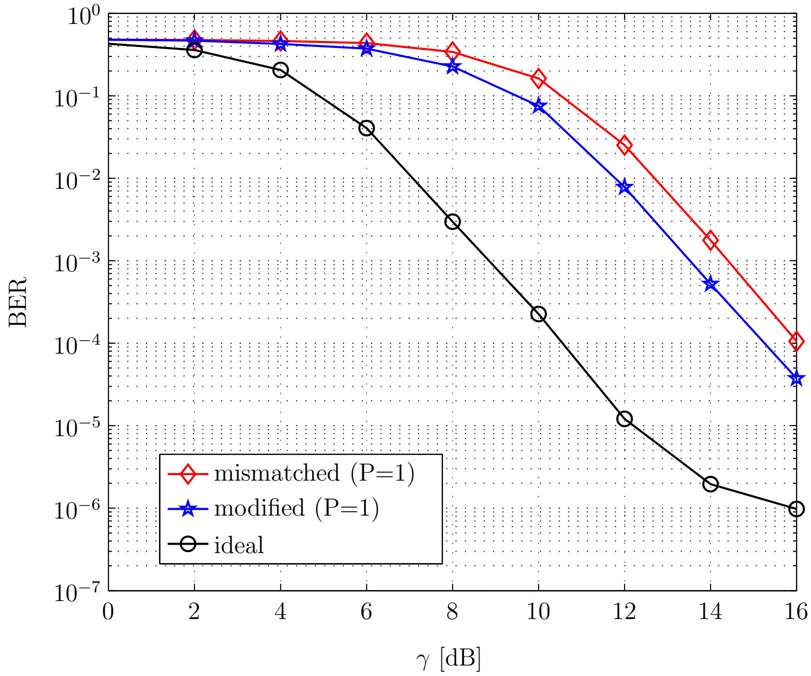


Figure 4.1: BER performance for 10 decoding iterations, with $P = 1$ pilot, $E_p = 0.5$ using set-partitioning labeling.

benefit from iterative decoding in the same way as the *modified receiver* does, see Fig. 4.3 for the *ideal receiver*.

Fig. 4.4 and Fig. 4.5 show the influence of the number of pilots and their symbol energy on the performance of the *modified* and *mismatched* receivers. Fig. 4.4 shows the improvement of the BER performance with $P = 1$, $P = 2$ and $P = 3$ pilots using constant mean pilot energy $E_p = 0.5$. The performance of the *modified receiver* for $P = 2$ pilots comes very close to that of a *mismatched receiver* for $P = 3$ pilots. For an increasing number of pilots the performance loss of the mismatched receiver becomes less. In Fig. 4.4 it can be seen that for $P = 3$ pilots the *mismatched receiver* comes quite close to the performance of the *modified receiver*.

By varying the pilot energy E_p and keeping the number of pilots constant $P = 1$, we get Fig. 4.5. We see that the *modified receiver* for pilot energies $E_p = 0.5$ and $E_p = 0.75$ achieves almost the same performance as the *mismatched receiver* for pilot

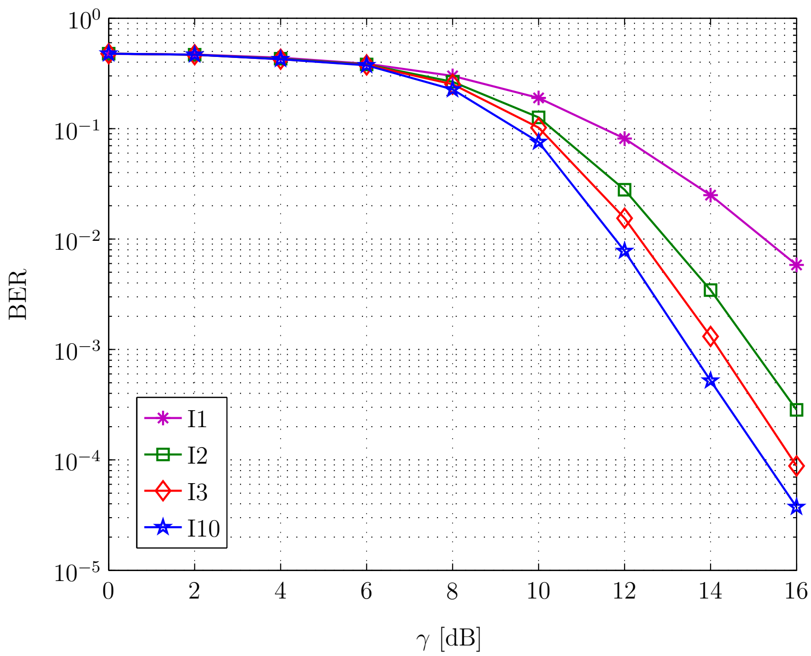


Figure 4.2: Performance of the *modified receiver* for various number of iterations, with $P = 1$ pilot, $E_p = 0.5$with $P = 1$ pilot, $E_p = 0.5$.

energies $E_p = 0.75$ and $E_p = 1$, respectively. By increasing the pilot energies the gap between the *mismatched receiver* and the *modified receiver* becomes insignificant. The channel estimation improves by increasing the number of pilots and their mean symbol energy. This consequently reduces the gap between the BER curves of the *modified* and *mismatched* receivers. The performance gain of the *modified receiver* over the *mismatched receiver* is maximum for $P = 1$ pilot with $E_p = 0.5$. The *modified receiver* outperforms the *mismatched receiver*, specially when fewer number of pilots are used for channel estimation. Similar results have also been shown in [2], [5] and [6].

Fig. 4.6 shows the BER performance of the receivers for $P = 1$ pilot with energy $E_p = 0.5$ for 10 iterations using a Gray bit labeling. The SNR required by the *ideal receiver* to achieve a BER $= 10^{-5}$ is increased by 4 dB if Gray labeling is used instead of set-partitioning as in Fig. 4.1. This is only due the fact that we used a different bit labeling because rest of the parameters are the same as for Fig. 4.1. Using a Gray labeling in an iterative system shows insignificant gain for more than one iteration.

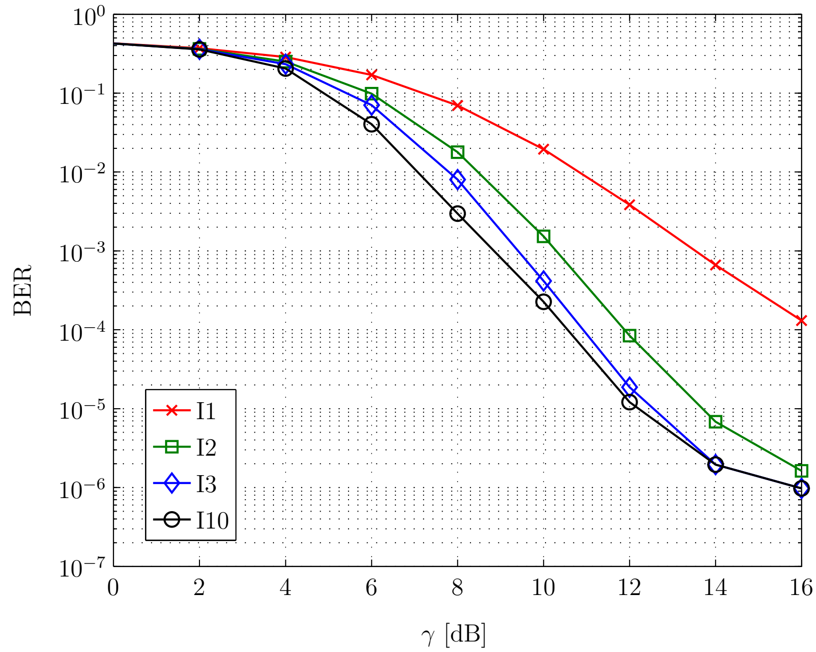


Figure 4.3: Performance of the *ideal receiver* for various number of iterations, with $P = 1$ pilot, $E_p = 0.5$.

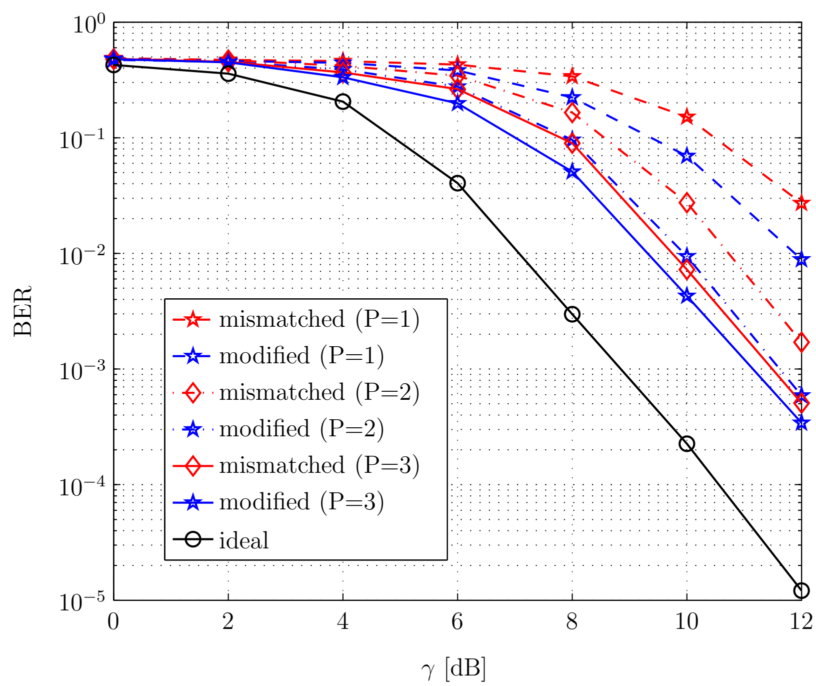


Figure 4.4: Performance of BICM-ID for varying number of pilots, $E_p = 0.5$.

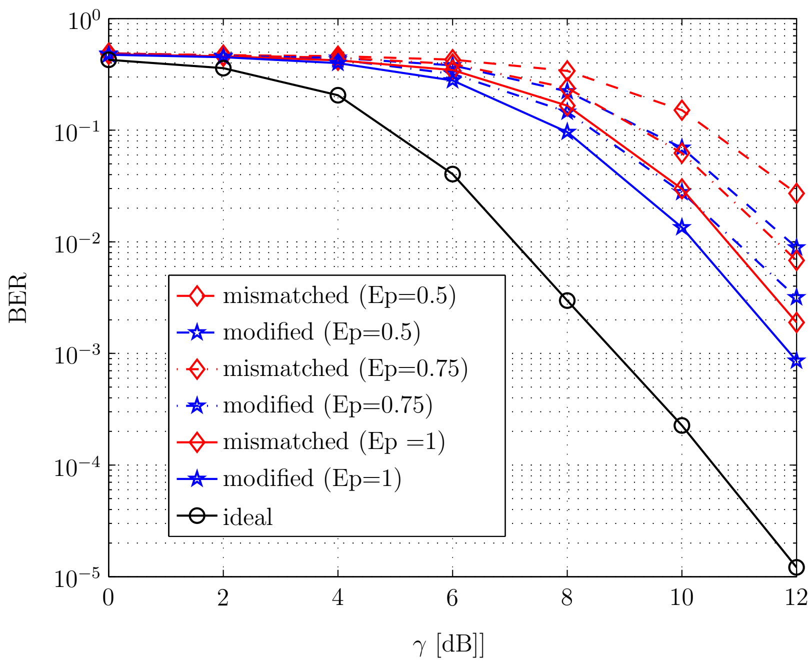


Figure 4.5: Performance of BICM-ID for varying pilot energy, $P = 1$.

This can be seen in Fig. 4.7, where the performance gain between 10 iterations is negligible. Gray labeling is not suitable for an iterative system as the maximum possible performance is achieved in the first iteration only (cf. [24], [23]).

Fig. 4.8 shows the BER performance of the receivers for $P = 1$ pilot with energy $E_p = 0.5$ for 10 iterations using the so-called m16a bit labeling (cf. Section 3.4). To achieve a BER $= 10^{-3}$ the *modified receiver* requires about 1.2 dB less SNR than the *mismatched receiver*. For a SNR of 12 dB the *ideal receiver* with m16a labeling achieves a BER $=10^{-6}$, whereas with set-partitioning labeling a BER $= 10^{-5}$ in Fig. 4.6 was possible. By looking at Fig. 4.8 and Fig. 4.6 we see that the m16a labeling achieves better performance for an SNR ≥ 10 for the *ideal receiver*. Moreover, it can be seen in Fig. 4.9 that the performance gain of the *ideal receiver* for the first two iterations is quite huge. For a larger number of iterations the decrease in the performance gain per iteration is much slower than in Fig. 4.1 for set-partitioning labeling. Thus, m16a labeling is able to achieve a better performance for a larger number of iterations than the set-partitioning labeling in the case of perfect CSI.

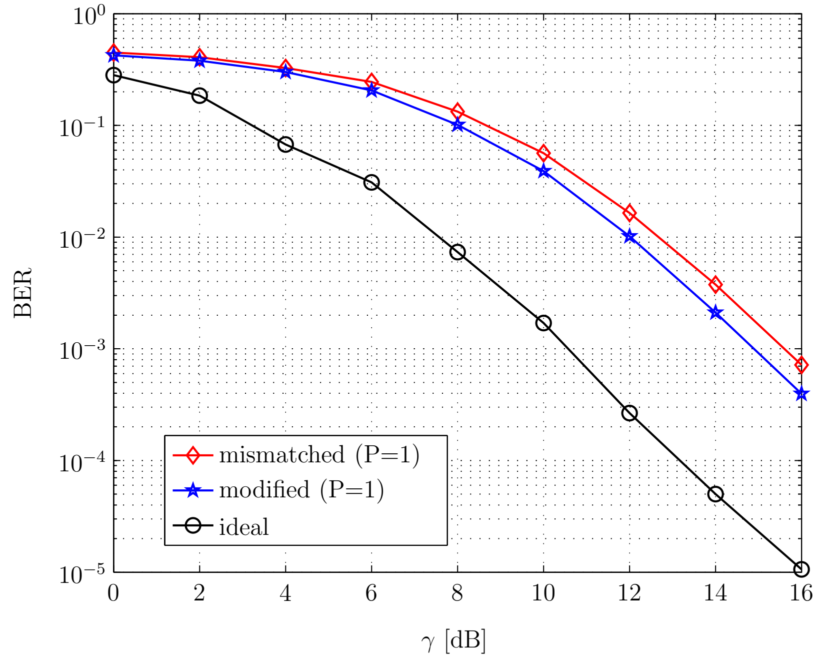


Figure 4.6: BER performance for 10 decoding iterations, with $P = 1$ pilot, $E_p = 0.5$ using Gray labeling.

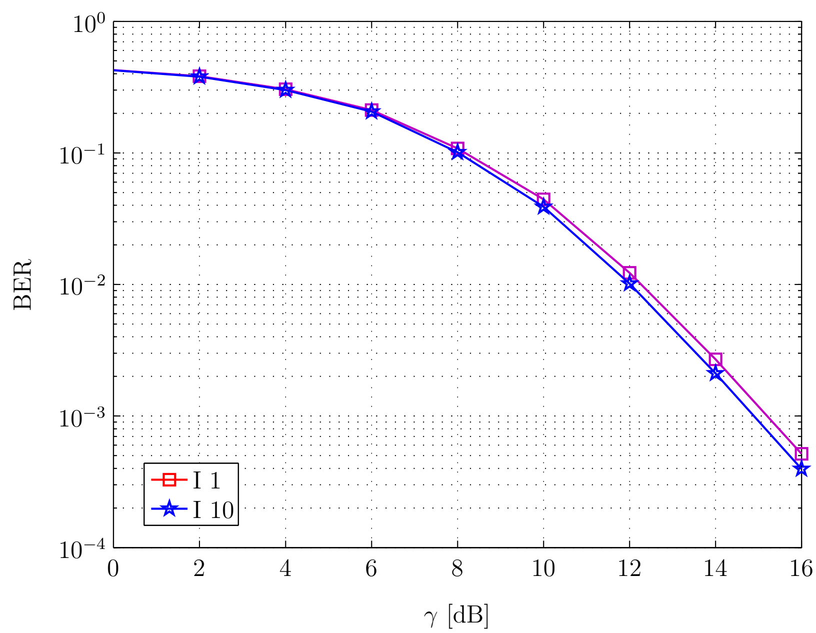


Figure 4.7: Performance of the *modified receiver* for various number of iterations, with $P = 1$ pilot, $E_p = 0.5$ using Gray labeling.

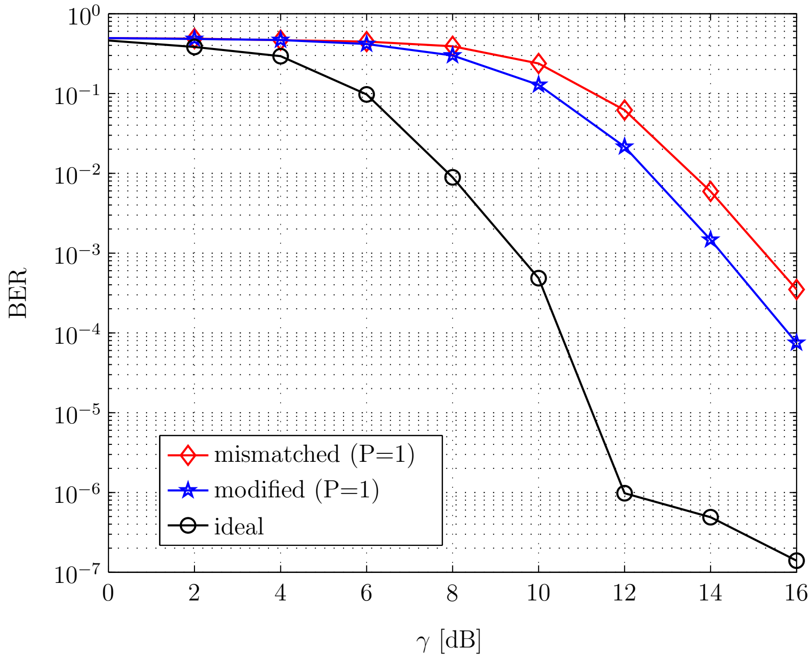


Figure 4.8: BER performance for 10 decoding iterations, with $P = 1$ pilot, $E_p = 0.5$ using m16a labeling.

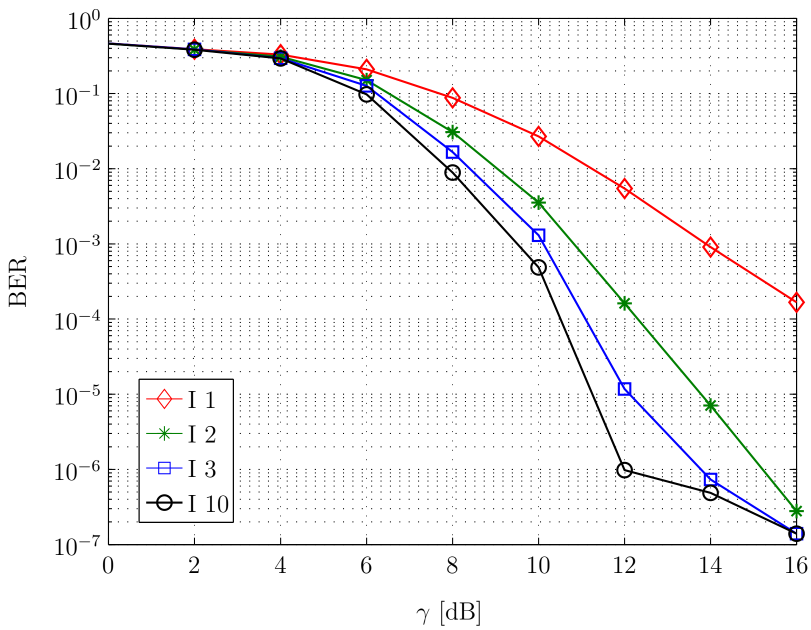


Figure 4.9: Performance of the *ideal receiver* for various number of iterations, with $P = 1$ pilot, $E_p = 0.5$ using m16a labeling.

5

Optimum Receiver for Non-iterative Systems

"[. . .] Engineering is the professional art of applying science to the optimum conversion of natural resources to the benefit of man." —— Ralph J. Smith

In this chapter we consider the derivation of the *optimum* metric for a pilot-assisted non-iterative system model. The received symbols are correlated due to an imperfect CSI. Symbol-wise detection for a correlated received symbol sequence leads to performance loss. In the derivation of the ML decision metric, the *optimum receiver* takes the channel correlation into account. This receiver delivers better performance through sequence detection, which is exponentially complex in the transmit sequence length. We have derived the *optimum receiver* for OFDM [11] and autoregressive (AR) [25] based channel models.

5.1 System Model

We define the $N \times 1$ transmit symbol sequence as a vector $\mathbf{x}$ containing both data symbols and pilot symbols given by

$$\mathbf{x} = \mathbf{x}^{\mathrm{d}} + \mathbf{x}^{\mathrm{p}}, \tag{5.1}$$

where $\mathbf{x}^{\mathrm{d}}$ and $\mathbf{x}^{\mathrm{p}}$ are the $N \times 1$ vectors carrying data symbols and pilots symbols, respectively. The pilot and data symbols are transmitted in an orthogonal manner. Depending on the channel model, the positions of the pilots in the transmitted symbol vector $\mathbf{x}$ correspond to the time or frequency slots in which they were sent. The positions of the pilots in the transmission is arbitrary. Let Ψ^{d} and Ψ^{p} denote the sets containing the positions of the symbols in the data and pilot vectors in a sequence, respectively. Then, Ψ^{d} and Ψ^{p} are disjoint sets, i.e., $\Psi^{\mathrm{d}} \cap \Psi^{\mathrm{p}} = \emptyset$ with cardinalities $|\Psi^{\mathrm{p}}| \triangleq P$ and $|\Psi^{\mathrm{d}}| \triangleq N - P$.

The data symbols are taken from the symbol alphabet $\mathcal{A}$ containing all possible symbols for the chosen modulation, i.e., with $n \in \Psi^{\mathrm{d}}$ we have $x_n^{\mathrm{d}} \in \mathcal{A}$. The cardinality of the symbol alphabet is given by $M_a = |\mathcal{A}| = 2^m$, where m is the number of bits assigned per symbol. The pilot symbols can be chosen from the symbol alphabet.

An example for the transmitted symbol vector $\mathbf{x}$ is given by sending pilot symbols periodically after every two data symbols, that yields the following sequence

$$\mathbf{x} = [x_1^{\mathrm{p}} \quad x_1^{\mathrm{d}} \quad x_2^{\mathrm{d}} \quad x_2^{\mathrm{p}} \quad x_3^{\mathrm{d}} \quad x_4^{\mathrm{d}} \quad x_3^{\mathrm{p}} \; \cdots\cdots]^T.$$

The system model is described by the following linear equation in its two equivalent forms given by

$$\mathbf{y} \triangleq \mathbf{H}\mathbf{x} + \mathbf{w} \triangleq \mathbf{X}\mathbf{h} + \mathbf{w}, \tag{5.2}$$

where the $N \times 1$ vector $\mathbf{y}$ denotes the received symbol sequence. The noise is represented by the $N \times 1$ vector $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I})$, where σ_w^2 denotes the noise variance and $\mathbf{I}$ is a $N \times N$ identity matrix. The transmitted $N \times N$ symbol matrix $\mathbf{X}$ is a diagonal matrix with the transmitted sequence $\mathbf{x}$ as its main diagonal, i.e., $\mathbf{X} = \mathrm{diag}(\mathbf{x})$. By definition of the symbol matrix we have $\mathbf{X} = \mathbf{X}^{\mathrm{d}} + \mathbf{X}^{\mathrm{p}}$, where $\mathbf{X}^{\mathrm{d}} = \mathrm{diag}(\mathbf{x}^{\mathrm{d}})$ and $\mathbf{X}^{\mathrm{p}} = \mathrm{diag}(\mathbf{x}^{\mathrm{p}})$ are diagonal matrices for the data symbols and pilot symbols, respectively. With the system model in (5.2) the received data symbol sequence $\mathbf{y}^{\mathrm{d}}$ and the received pilot symbol sequence $\mathbf{y}^{\mathrm{p}}$ are given by

$$\mathbf{y}^{\mathrm{d}} = \mathbf{H}\mathbf{x}^{\mathrm{d}} + \mathbf{w} = \mathbf{X}^{\mathrm{d}}\mathbf{h} + \mathbf{w}^{\mathrm{d}} \tag{5.3}$$

$$\mathbf{y}^{\mathrm{p}} = \mathbf{H}\mathbf{x}^{\mathrm{p}} + \mathbf{w} = \mathbf{X}^{\mathrm{p}}\mathbf{h} + \mathbf{w}^{\mathrm{p}}. \tag{5.4}$$

It is easy to verify with (5.1) and (5.2) that $\mathbf{y} = \mathbf{y}^{\mathrm{d}} + \mathbf{y}^{\mathrm{p}}$. The $N \times 1$ data and pilot noise vectors are distributed as $\mathbf{w}^{\mathrm{d}} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I})$ and $\mathbf{w}^{\mathrm{p}} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I})$, respectively

5.2 Channel Model

We consider a Rayleigh fast fading scenario with correlated channel coefficients. In general, the channel is modeled by a $N \times N$ diagonal channel matrix $\mathbf{H}$. The main diagonal of the channel matrix is given by the $N \times 1$ vector $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_h)$, such that $\mathbf{H} = \mathrm{diag}(\mathbf{h})$. The covariance matrix of the channel coefficients is defined as $\mathbf{C}_h \triangleq \mathbb{E}\{\mathbf{h}\mathbf{h}^H\}$, and is characteristic for each channel model, e.g., OFDM [11], AR [25] models. These two channel models are described in Subsections 5.2.1 and 5.2.2.

5.2.1 OFDM Based Channel Model

We consider an OFDM model [1] with the channel impulse response defined by the $N \times 1$ vector as

$$\tilde{\mathbf{h}} \triangleq [\tilde{h}_1 \quad \cdots \quad \tilde{h}_L \quad 0 \quad \cdots \quad 0]^T, \tag{5.5}$$

with L uncorrelated channel taps. The vector $\mathbf{h}$ contains the channel coefficients for each subcarrier and is defined as the discrete fourier transform of the vector $\tilde{\mathbf{h}}$ in (5.5) as

$$\mathbf{h} = \mathbf{F}\tilde{\mathbf{h}}, \tag{5.6}$$

where $\mathbf{F}$ denotes the $N \times N$ unitary discrete fourier transformation (DFT) matrix. An arbitrary element of the unitary DFT Matrix is defined by the following equation as

$$[\mathbf{F}]_{l,k} \triangleq \frac{1}{\sqrt{(N)}} W_N^{(l-1)(k-1)} \qquad l,k \in \{1,\cdots,N\}, \tag{5.7}$$

where W_N is the n-th root of unity is defined as $W_N \triangleq \exp\left(-\frac{2\pi i}{N}\right)$. Thus, the covariance matrix $\mathbf{C}_h$ for an OFDM based channel model is given by

$$\begin{aligned} \mathbf{C}_h &= \mathbb{E}\{\mathbf{h}\mathbf{h}^H\} \\ &= \mathbb{E}\{\mathbf{F}\,\tilde{\mathbf{h}}\,\tilde{\mathbf{h}}^H\,\mathbf{F}^H\} \\ &= \mathbf{F}\,\mathbf{C}_{\tilde{h}}\,\mathbf{F}^H, \end{aligned} \tag{5.8}$$

where $\mathbf{C}_{\tilde{h}} = \mathbb{E}\{\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\}$ is the diagonal covariance matrix of the channel taps vector $\tilde{\mathbf{h}}$ defined in (5.5).

5.2.2 AR(1) Based Channel Model

The channel coefficients in the vector $\mathbf{h}$ based on the autoregressive model are related by the following equation as

$$h_n = \alpha^2\, h_{n-1} + w_n \qquad n \in \{1,\cdots,N\}, \tag{5.9}$$

where α is the correlation coefficient and w_n is a *zero mean complex* Gaussian noise with the distribution $w_n \sim \mathcal{CN}(0,\sigma_w^2)$. The covariance matrix corresponding to (5.9) has a Toeplitz structure [26] with elements defined by the following equation as

$$[\mathbf{C}_h]_{l,k} \triangleq \mathbb{E}\{h_l h_k\} = \alpha^{2|l-k|} \qquad l,k \in \{1,\cdots,N\}. \tag{5.10}$$

The covariance matrix $\mathbf{C}_h$ can also be represented with the help of a singular value decomposition (SVD) [26] as

$$\mathbf{C}_h = \mathbf{U}\,\mathbf{V}\,\mathbf{U}^H, \tag{5.11}$$

where $\mathbf{U}$ is a unitary matrix and $\mathbf{V}$ contains the singular values of the covariance matrix in diagonal form. It is easy to see the similar form of the covariance matrices in (5.8) and (5.11). We will need the latter form of the covariance matrix in Subsection 5.5.2.

5.3 Pilot Assisted Channel Estimation

The pilot assisted channel estimation at the receiver is done with the help of the known pilot matrix $\mathbf{X}^{\mathrm{p}}$ defined in Section 5.1. Channel estimation is done by deriving the MMSE estimator for the channel using the system model in (5.4) for the received pilot symbol sequence $\mathbf{y}^{\mathrm{p}} = \mathbf{X}^{\mathrm{p}}\mathbf{h} + \mathbf{w}$. The MMSE channel estimation requires the *posterior* pdf of the channel given the received pilot symbol sequence $f(\mathbf{h}|\mathbf{y}^{\mathrm{p}})$ (cf. [20]).

We can utilize the MMSE estimator derived in Chapter 3 to calculate the mean and covariance of the multivariate Gaussian pdf $f(\mathbf{h}|\mathbf{y}^{\mathrm{p}})$. From the vector extension of (3.17) and (3.18) we get the expressions

$$\hat{\mathbf{h}}_{\mathrm{MMSE}} = \mathbb{E}\{\mathbf{h}|\mathbf{y}^{\mathrm{p}}\} = \mathbf{C}_{hy^{\mathrm{p}}}\,\mathbf{C}_{y^{\mathrm{p}}y^{\mathrm{p}}}^{-1}\,\mathbf{y}^{\mathrm{p}}, \tag{5.12}$$

$$\mathbf{C}_{h|y^{\mathrm{p}}} = \mathbf{C}_h - \mathbf{C}_{hy^{\mathrm{p}}}\,\mathbf{C}_{y^{\mathrm{p}}y^{\mathrm{p}}}^{-1}\,\mathbf{C}_{y^{\mathrm{p}}h}. \tag{5.13}$$

The pdf of the noise and channel vectors are defined in Sections 5.1 and 5.2, respectively. As in Section 3.2 we calculate the following covariances to evaluate 5.12 and 5.13 as

$$\begin{aligned}
\mathbf{C}_{y^{\mathrm{p}}y^{\mathrm{p}}} &= \mathbb{E}\{\mathbf{y}^{\mathrm{p}}\,(\mathbf{y}^{\mathrm{p}})^H\} \\
&= \mathbb{E}\{(\mathbf{X}^{\mathrm{p}}\,\mathbf{h} + \mathbf{w})(\mathbf{X}^{\mathrm{p}}\,\mathbf{h} + \mathbf{w})^H\} \\
&= \mathbf{X}^{\mathrm{p}}\,\mathbb{E}\{\mathbf{h}\mathbf{h}^H\}\,(\mathbf{X}^{\mathrm{p}})^H + \mathbb{E}\{\mathbf{w}\mathbf{w}^H\} \\
&= \mathbf{X}^{\mathrm{p}}\,\mathbf{C}_h\,(\mathbf{X}^{\mathrm{p}})^H + \sigma_w^2\mathbf{I},
\end{aligned} \tag{5.14}$$

$$\begin{aligned}
\mathbf{C}_{hy^{\mathrm{p}}} &= \mathbb{E}\{\mathbf{h}(\mathbf{y}^{\mathrm{p}})^H\} \\
&= \mathbb{E}\{\mathbf{h}(\mathbf{X}^{\mathrm{p}}\,\mathbf{h} + \mathbf{w})^H\} \\
&= \mathbb{E}\{\mathbf{h}\mathbf{h}^H\}\,(\mathbf{X}^{\mathrm{p}})^H \\
&= \mathbf{C}_h\,(\mathbf{X}^{\mathrm{p}})^H.
\end{aligned} \tag{5.15}$$

Using (5.14) and (5.15) the MMSE estimate for the channel $\mathbf{h}$ becomes

$$\hat{\mathbf{h}}_{\mathrm{MMSE}} = \mathbf{C}_h\,(\mathbf{X}^{\mathrm{p}})^H\,[\mathbf{X}^{\mathrm{p}}\,\mathbf{C}_h\,(\mathbf{X}^{\mathrm{p}})^H + \sigma_w^2\mathbf{I}]^{-1}\,\mathbf{y}^{\mathrm{p}}. \tag{5.16}$$

The covariance matrix of the estimation error is obtained as

$$\mathbf{C}_{h|y^{\mathrm{p}}} = \mathbf{C}_h - \mathbf{C}_h\,(\mathbf{X}^{\mathrm{p}})^H\,[\mathbf{X}^{\mathrm{p}}\,\mathbf{C}_h\,(\mathbf{X}^{\mathrm{p}})^H + \sigma_w^2\mathbf{I}]^{-1}\,\mathbf{X}^{\mathrm{p}}\,\mathbf{C}_h, \tag{5.17}$$

where $\mathbf{C}_h$ is the covariance matrix for the general channel model described in Section 5.2. The mean square error (MSE) [20] is defined as

$$\epsilon_{mse} = \frac{1}{N}\mathbb{E}\{\|\mathbf{h} - \hat{\mathbf{h}}_{\mathrm{MMSE}}\|^2\}, \tag{5.18}$$

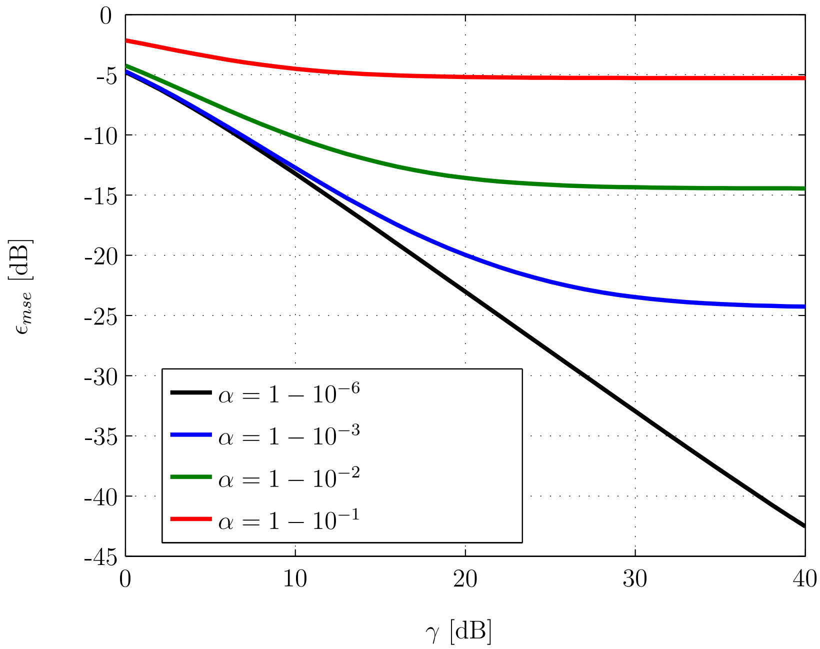
which is equal to the trace of the covariance matrix averaged over the length of the transmit symbol sequence[1]

$$\epsilon_{mse} = \frac{\mathrm{tr}\left(\mathbf{C}_{h|y^{\mathrm{p}}}\right)}{N}. \tag{5.19}$$

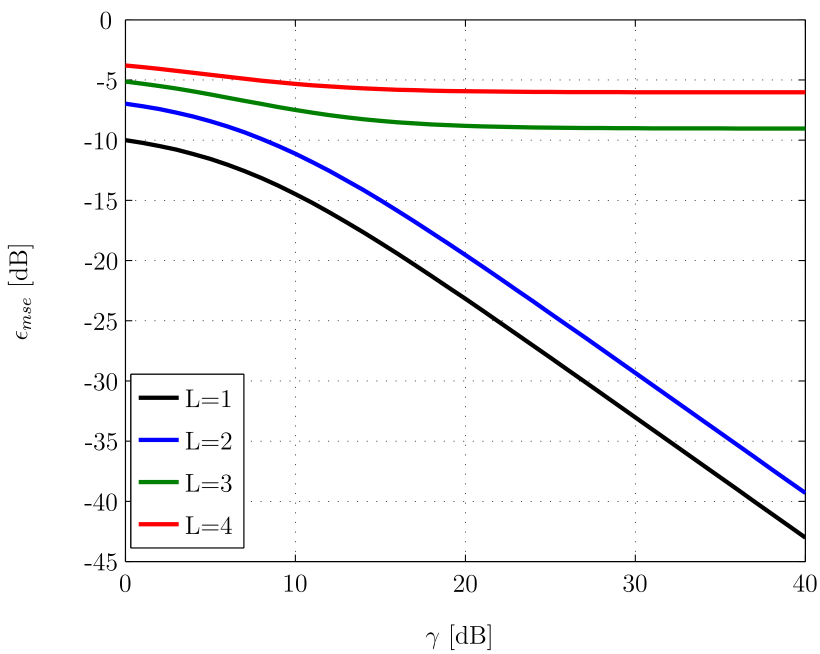
Fig. 5.1(b) and 5.1(a) show the theoretical (MSE) versus SNR $\gamma = 10\log(\sigma_w^2)^{-1}$ for the channel models in Subsection 5.2.1 and 5.2.2. The MSE was computed using $P = 2$ pilot symbols with an average pilot energy $E_p = 1$ and a transmit symbol sequence of length $N = 8$.

In Fig. 5.1(b) the MSE for an OFDM model is shown for a varying number of channel taps L. It can be seen that the channel estimation only works well for $L \leq P$, because the sampling theorem is not satisfied if $L > P$ [27], [28]. The MSE for the AR channel model in Fig. 5.1(a) is plotted for a different channel correlation coefficients α. The MSE has an error floor because the channel coefficients are not bandlimited [25]. Thus, it is not possible to reconstruct the channel properly with a limited number of pilots.

[1] The function tr(A) is equivalent to trace(A)



(a) MSE AR channel model.



(b) MSE OFDM channel model.

Figure 5.1: Mean square error with $P = 2$, $E_p = 1$ and $N = 8$ for (a) an AR channel model and (b) an OFDM channel model.

5.4 Mismatched Receiver

The ML decision rule for the *ideal receiver* based on the system model in (5.2) and the general channel model in Section 5.2, is formulated by maximizing the pdf of the likelihood function $f(\mathbf{X}^{\mathrm{d}}|\mathbf{y}^{\mathrm{d}}, \mathbf{h})$. Using the Bayes's theorem [20] we can write

$$\begin{aligned} f(\mathbf{X}^{\mathrm{d}}|\mathbf{y}^{\mathrm{d}}, \mathbf{h}) &\propto f(\mathbf{y}^{\mathrm{d}}|\mathbf{X}^{\mathrm{d}}, \mathbf{h}) \\ &= \prod_{\Psi^{\mathrm{d}}} f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, h_n). \end{aligned} \tag{5.20}$$

The pdf $f(\mathbf{y}^{\mathrm{d}}|\mathbf{X}^{\mathrm{d}}, \mathbf{h})$ can be factorized because the receiver does not take the correlation of the received symbol sequence into account. Because of the factorization, a symbol-wise detection of the received symbol sequence is possible. The pdf $f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, h_n)$ can be derived from the system model in (5.3) and is given by the distribution $\mathcal{CN}(h_n x_n^{\mathrm{d}}, \sigma_w^2)$. The ML decision rule for the *ideal receiver* with perfect CSI is formulated as

$$\begin{aligned} \hat{x}_{n,\mathrm{ML}}^{\mathrm{d}} &= \underset{x_n^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmax}}\, f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, h_n) \\ &= \underset{x_n^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmax}}\, \ln f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, h_n) \\ &= \underset{x_n^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmin}} \left(|y_n^{\mathrm{d}} - h_n x_n^{\mathrm{d}}|^2\right). \end{aligned} \tag{5.21}$$

As mentioned in Section 4.1, the *mismatched receiver* replaces the true channel by its estimate in the metric of an *ideal receiver*. The ML decision rule for the *mismatched receiver* is derived by using the channel estimate in the metric for perfect CSI in (5.21) and therefore given by

$$\begin{aligned} \hat{x}_{n,\mathrm{ML}}^{\mathrm{d}} &= \underset{x_n^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmax}}\, f(y_n^{\mathrm{d}}|x_n^{\mathrm{d}}, h_n)\, |_{h_n = \hat{h}_n} \\ &= \underset{x_n^{\mathrm{d}} \in \mathcal{A}}{\operatorname{argmin}} \left(|y_n^{\mathrm{d}} - \hat{h}_n x_n^{\mathrm{d}}|^2\right). \end{aligned} \tag{5.22}$$

The knowledge of the channel statistics is only used for the MMSE channel estimation as given in Section 5.3. In the following section we derive the *optimum receiver*, which incorporates the knowledge of the channel statistics in its ML decision metric.

5.5 Optimum Receiver

The *optimum receiver* utilizes the channel statistics in the derivation of the ML metric. ML detection at the receiver is performed by maximizing the joint pdf $f(\mathbf{X}|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{X})$, therefore no prior channel estimation is required. The symbol matrix $\mathbf{X}$ and received symbol vector $\mathbf{y}$ are defined in the Section 5.1 in (5.2). The likelihood function $f(\mathbf{y}|\mathbf{X})$ used for the ML detection cannot be factorized because the receiver takes the channel correlation into account, i.e., the receiver knows that the received symbols will not be statistically independent of each other. Thus, the *optimum receiver* will have to perform sequence ML detection.

5.5.1 Optimum ML Detection

The pdf of the general channel model $\mathbf{h}$ and the additive noise $\mathbf{w}$ is defined in Section 5.2 and 5.1 respectively. The likelihood function can be calculated by the following marginalization

$$f(\mathbf{y}|\mathbf{X}) = \int_{\mathbf{h}} f(\mathbf{y}|\mathbf{X},\mathbf{h}) f(\mathbf{h}) d\mathbf{h}. \tag{5.23}$$

By inspecting (5.2) and (5.23) it can be seen that $f(\mathbf{y}|\mathbf{X})$ has a Gaussian distribution, therefore it is sufficient to calculate the mean and covariance matrix. With respect to the system model in (5.2) we calculate the mean and variance as

$$\begin{aligned} \boldsymbol{\mu}_{\text{opt}} &= \mathbb{E}\{\mathbf{y}|\mathbf{X}\} \\ &= \mathbb{E}\{(\mathbf{X}\mathbf{h}+\mathbf{w})|\mathbf{X}\} \\ &= \mathbf{X}\mathbb{E}\{\mathbf{h}\} + \mathbb{E}\{\mathbf{w}\} = 0, \end{aligned} \tag{5.24}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{\text{opt}} &= \mathbb{E}\{\mathbf{y}\mathbf{y}^H|\mathbf{X}\} \\ &= \mathbb{E}\{(\mathbf{X}\mathbf{h}+\mathbf{w})(\mathbf{X}\mathbf{h}+\mathbf{w})^H|\mathbf{X}\} \\ &= \mathbf{X}\,\mathbb{E}\{\mathbf{h}\mathbf{h}^H\}\,\mathbf{X}^H + \mathbb{E}\{\mathbf{w}\mathbf{w}^H\} \\ &= \mathbf{X}\,\mathbf{C}_h\,\mathbf{X}^H + \sigma_w^2\,\mathbf{I}. \end{aligned} \tag{5.25}$$

It is to be noted that the covariance matrix $\mathbf{\Sigma}_{\text{opt}}$ is a function of the symbol matrix $\mathbf{X}$ and therefore a function of the data sequence $\mathbf{x}^{\text{d}}$. We represent a realization of the transmit symbol matrix as $\tilde{\mathbf{X}} = \tilde{\mathbf{X}}^{\text{d}} + \tilde{\mathbf{X}}^{\text{p}}$, where the pilot matrix $\tilde{\mathbf{X}}^{\text{p}}$ remains the same throughout. We define the set containing all possible realizations of the transmit sequence $\tilde{\mathbf{x}}^{\text{d}}$ as $\mathcal{X}^{\text{d}}$. It is to be noted that the cardinality $|\mathcal{X}^{\text{d}}| \triangleq |\mathcal{A}|^{(N-P)} = 2^{m(N-P)}$ is exponentially complex in the length of the transmit sequence N. The ML decision rule for the *optimum receiver* can now be formulated as

$$
\begin{aligned}
\hat{\mathbf{x}}^{\text{d}}_{\text{ML}} &= \underset{\tilde{\mathbf{x}}^{\text{d}} \in \mathcal{X}^{\text{d}}}{\operatorname{argmax}} f(\mathbf{y}|\tilde{\mathbf{X}}) \\
&= \underset{\tilde{\mathbf{x}}^{\text{d}} \in \mathcal{X}^{\text{d}}}{\operatorname{argmax}} \ln f(\mathbf{y}|\tilde{\mathbf{X}}) \\
&= \underset{\tilde{\mathbf{x}}^{\text{d}} \in \mathcal{X}^{\text{d}}}{\operatorname{argmax}} \ \ln \left(\frac{\exp\left(- \left(\mathbf{y} - \boldsymbol{\mu}_{\text{opt}}\right)^H \mathbf{\Sigma}_{\text{opt}}^{-1} \left(\mathbf{y} - \boldsymbol{\mu}_{\text{opt}}\right)\right)}{\pi^N \det(\mathbf{\Sigma}_{\text{opt}})} \right) \\
&= \underset{\tilde{\mathbf{x}}^{\text{d}} \in \mathcal{X}^{\text{d}}}{\operatorname{argmin}} \left(\ln \det(\mathbf{\Sigma}_{\text{opt}}) + \mathbf{y}^H \mathbf{\Sigma}_{\text{opt}}^{-1} \mathbf{y} \right), \qquad (5.26)
\end{aligned}
$$

We note that the term $\mathbf{y}^H \mathbf{\Sigma}_{\text{opt}}^{-1} \mathbf{y}$ in (5.26) has a quadratic form in $\mathbf{\Sigma}_{\text{opt}}$, which has a computational complexity $\mathcal{O}(N^2)$. In the following subsection we analyze the metric for a low-complexity computation.

5.5.2 Evaluation of the Optimum Metric

To find the minimum of the optimum metric, we need to compute (5.26) for all realizations of the transmitted symbol matrix $\tilde{\mathbf{X}}$. This involves two major matrix operations for each $\tilde{\mathbf{X}}$, i.e., we need to calculate the determinant and the inverse of the joint covariance matrix $\mathbf{\Sigma}_{\text{opt}}$. By using standard matrix determinant and inversion lemmas [20], it is easier to take an in depth look at the computational procedure.

We will now take a look at the optimum metric for the OFDM based channel model. For further calculations we assume that the transmitted symbols are from a constant modulus symbol alphabet, i.e., $\tilde{\mathbf{X}}^H \tilde{\mathbf{X}} = |x|^2 \mathbf{I}$. First we calculate the determinant of the joint covariance matrix $\boldsymbol{\Sigma}_{\text{opt}}$ using the matrix determinant lemma as

$$\begin{aligned}
\det(\boldsymbol{\Sigma}_{\text{opt}}) &= \det(\tilde{\mathbf{X}}\,\mathbf{C}_h\,\tilde{\mathbf{X}}^H + \sigma_w^2\,\mathbf{I}) \\
&= \det(\mathbf{C}_h^{-1} + \tilde{\mathbf{X}}^H(\sigma_w^2\,\mathbf{I})^{-1}\tilde{\mathbf{X}})\det(\mathbf{C}_h)\det(\sigma_w^2\,\mathbf{I}) \\
&= \det\left(\mathbf{C}_h^{-1} + \frac{\tilde{\mathbf{X}}^H\,\tilde{\mathbf{X}}}{\sigma_w^2}\right)\det(\mathbf{C}_h)\det(\sigma_w^2\,\mathbf{I}) \\
&= \det\left(\mathbf{C}_h^{-1} + \frac{|x|^2}{\sigma_w^2}\,\mathbf{I}\right)\det(\mathbf{C}_h)\det(\sigma_w^2\,\mathbf{I}), \qquad (5.27)
\end{aligned}$$

where the last step results from the constant modulus property of the transmitted symbols. The determinant of the joint covariance matrix turns out to be constant and independent of the data symbols $\mathbf{x}^{\text{d}}$. Thus, it is not necessary to include the determinant of the joint covariance matrix in the minimization process of the optimum metric in (5.26).

We next insert the channel covariance matrix $\mathbf{C}_h$, as given in (5.8) and (5.25) to yield

$$\begin{aligned}
\boldsymbol{\Sigma}_{\text{opt}} &= \tilde{\mathbf{X}}\,(\mathbf{F}\,\mathbf{C}_{\tilde{h}}\,\mathbf{F}^H)\,\tilde{\mathbf{X}}^H + \sigma_w^2\,\mathbf{I} \\
&= \tilde{\mathbf{X}}\,\left(\mathbf{F}\,\mathbf{C}_{\tilde{h}}\,\mathbf{F}^H + \tilde{\mathbf{X}}^{-1}\,(\sigma_w^2\mathbf{I})\,\tilde{\mathbf{X}}^{-H}\right)\,\tilde{\mathbf{X}}^H \\
&= \tilde{\mathbf{X}}\,\mathbf{F}\,\left(\mathbf{C}_{\tilde{h}} + \sigma_w^2\,\mathbf{F}^{-1}\,\tilde{\mathbf{X}}^{-1}\,\tilde{\mathbf{X}}^{-H}\,\mathbf{F}^{-H}\right)\,\mathbf{F}^H\,\tilde{\mathbf{X}}^H. \qquad (5.28)
\end{aligned}$$

Using (5.28), the inverse of $\boldsymbol{\Sigma}_{\text{opt}}$ can now be calculated as

$$\begin{aligned}
\boldsymbol{\Sigma}_{\text{opt}}^{-1} &= \tilde{\mathbf{X}}^{-H}\,\mathbf{F}^{-H}\,\left(\mathbf{C}_{\tilde{h}} + \sigma_w^2\,\mathbf{F}^{-1}\,\tilde{\mathbf{X}}^{-1}\,\tilde{\mathbf{X}}^{-H}\,\mathbf{F}^{-H}\right)^{-1}\,\mathbf{F}^{-1}\,\tilde{\mathbf{X}}^{-1} \\
&= \tilde{\mathbf{X}}^{-H}\,\mathbf{F}\,\left(\mathbf{C}_{\tilde{h}} + \sigma_w^2\,\mathbf{F}^H\,\tilde{\mathbf{X}}^{-1}\,\tilde{\mathbf{X}}^{-H}\,\mathbf{F}\right)^{-1}\,\mathbf{F}^H\,\tilde{\mathbf{X}}^{-1}, \qquad (5.29)
\end{aligned}$$

where the last step results from the fact that $\mathbf{F}^{-1} = \mathbf{F}^H$. For the sake of simplicity we define $\mathbf{D} = \left(\mathbf{C}_{\tilde{h}} + \sigma_w^2\,\mathbf{F}^H\,\tilde{\mathbf{X}}^{-1}\,\tilde{\mathbf{X}}^{-H}\,\mathbf{F}\right)^{-1}$. With the assumption of constant modulus symbols, the matrix $\mathbf{D}$ becomes a diagonal matrix, i.e.,

$$\begin{aligned}
\mathbf{D} &= \left(\mathbf{C}_{\tilde{h}} + \sigma_w^2\,\mathbf{F}^H\,\tilde{\mathbf{X}}^{-1}\,\tilde{\mathbf{X}}^{-H}\,\mathbf{F}\right)^{-1} \\
&= \left(\mathbf{C}_{\tilde{h}} + \frac{|x|^2}{\sigma_w^2}\,\mathbf{I}\right)^{-1}, \qquad (5.30)
\end{aligned}$$

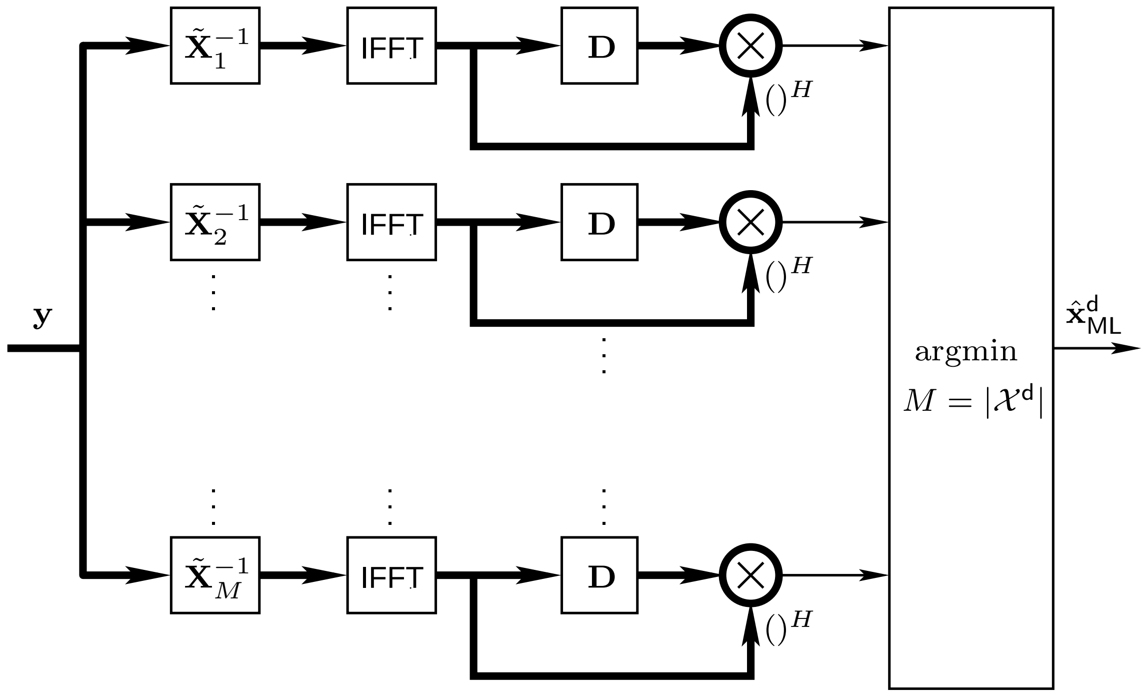


Figure 5.2: ML Detector for the *optimum receiver*

which is independent of $\mathbf{x}^{\mathrm{d}}$ and therefore the diagonal matrix $\mathbf{D}$ can be precalculated. With (5.27), (5.29) and (5.30) the optimum metric reduces to the following expression

$$\hat{\mathbf{x}}^{\mathrm{d}}_{\mathrm{ML}} = \underset{\tilde{\mathbf{x}}^{\mathrm{d}} \in \mathcal{X}^{\mathrm{d}}}{\operatorname{argmin}} \left(\mathbf{y}^H \, \tilde{\mathbf{X}}^{-H} \, \mathbf{F} \, \mathbf{D} \, \mathbf{F}^H \, \tilde{\mathbf{X}}^{-1} \, \mathbf{y} \right). \tag{5.31}$$

By further examining the optimum metric in (5.31), we see that the expression for the metric has a quadratic form in $\mathbf{D}$. With the vector $\tilde{\mathbf{y}} = \mathbf{F}^H \, \tilde{\mathbf{X}}^{-1} \, \mathbf{y}$, the optimum metric in (5.31) can now be expressed as

$$\hat{\mathbf{x}}^{\mathrm{d}}_{\mathrm{ML}} = \underset{\tilde{\mathbf{x}}^{\mathrm{d}} \in \mathcal{X}^{\mathrm{d}}}{\operatorname{argmin}} \left(\tilde{\mathbf{y}}^H \, \mathbf{D} \, \tilde{\mathbf{y}} \right). \tag{5.32}$$

We only need to calculate the vector $\tilde{\mathbf{y}} = \mathbf{F}^H \, \tilde{\mathbf{X}}^{-1} \, \mathbf{y}$ once for every realization of the transmitted symbol matrix $\tilde{\mathbf{X}}$ an we can Hermitian transpose it to get $\tilde{\mathbf{y}}^H$. The total number of realizations of the symbol matrix $\tilde{\mathbf{X}}$ is given by the cardinality of the set $\mathcal{X}^{\mathrm{d}}$ in Section 5.5.1. Since the symbol matrix $\tilde{\mathbf{X}}$ is diagonal, i.e., $\tilde{\mathbf{X}} = \mathrm{diag}(\tilde{\mathbf{x}})$, the term $\tilde{\mathbf{X}}^{-1} \, \mathbf{y}$ corresponds to a symbol-wise division of the received symbol sequence $\mathbf{y}$ and the symbol sequence $\tilde{\mathbf{x}}$.

The analysis of the optimum metric has shown that simplifications are possible and have also been implemented. The implementation procedure described in this section proves to be more efficient than using the original quadratic form of the optimum metric in (5.26). The complexity reduces from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$.

The evaluation of the optimum metric in this section is only valid for a constant modulus symbol alphabet and the channel model we used, i.e., the OFDM channel model from Subsection 5.2.1. The implementation of the ML detector for the optimum metric in (5.31) is illustrated in the block diagram in Fig. 5.2 and it shows the computation process for $M = 2^{m(N-P)}$ realizations of the transmit symbol matrix $\tilde{\mathbf{X}}$.

Even though the metric evaluation has been made efficient, the complexity of the ML detection is still a problem. The complexity increases exponentially in the length of the transmitted data symbol sequence. In [5] a sequence decoding algorithm like the Viterbi decoder has been used to overcome the exponential complexity (cf. 2.1.2.1). To apply this approach to our problem still remains an open issue.

The efficient implementation of the *optimum* metric remains the same for the AR(1) channel model and thus will not be repeated here. This is due to the fact that covariance matrices of the OFDM channel model (5.8) and the AR(1) channel model (5.11) have a similar form. However, the AR(1) model additionally requires a SVD for its channel covariance matrix to have the same form as in the OFDM channel model. The results for the AR(1) model can easily be derived by replacing the matrix $\mathbf{F}$ with $\mathbf{U}$ in the expression of the optimum metric.

5.6 Simulation Results

5.6.1 Performance Evaluation of the Optimum Receiver

In the following subsection, we will compare the BER performance of the *ideal receiver*, the *mismatched receiver* and the *optimum receiver*. The ML metrics of these receivers are given in (5.21), (5.22) and (5.26), respectively. The BER performance is evaluated for the OFDM and AR(1) channel models.

5.6.2 OFDM Channel Model

A total of $N = 8$ information bits are mapped to symbols using a BPSK modulation ($M_a = 2$). We consider a Rayleigh fast fading OFDM based channel model with N subcarriers and $L = 2$ channel taps. The channel estimation for the *mismatched receiver* is done with an MMSE estimator.

In the following figures the SNR for the OFDM based channel model is defined as

$$\gamma = 10 \log \frac{E_b}{\sigma_w^2} = 10 \log \frac{E_s}{\mathrm{ld}(M_a)} \frac{\|\mathbf{F}\,\tilde{\mathbf{h}}\|^2}{N}, \tag{5.33}$$

where σ_w^2 is the noise variance. Here, E_b and E_s denote the mean bit and symbol energies, respectively.

Fig. 5.3 shows the BER performance versus SNR of the three receivers using $P = 2$ and $P = 3$ pilots with mean pilot energy $E_p = 1/16$. It can be seen that the *optimum receiver* outperforms the *mismatched receiver*. For $P = 2$, the SNR required to achieve a BER $= 10^{-3}$ by the *optimum receiver* is about 10.3 dB less than that of a *mismatched receiver*. Moreover, for high SNR values the performance of the *optimum receiver* comes close to that of an *ideal receiver*. For a BER $= 10^{-3}$ and $P = 2$ pilots the *optimum receiver* has a 4.5 dB gap to the *ideal receiver* whereas the *mismatched receiver* has 14.8 dB gap.

In Fig. 5.4, we see the effect of varying the mean pilot symbol energy on the BER performance of the *mismatched* and *optimum* receivers. Fig. 5.4 shows the BER performance of the three receivers for $P = 2$ pilots with different mean symbol energies.

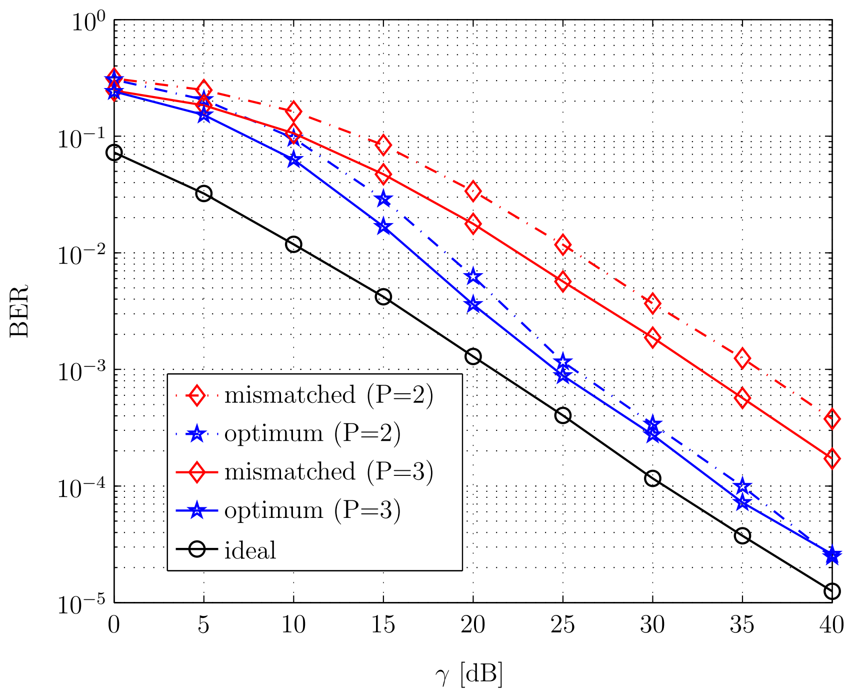


Figure 5.3: BER performance versus SNR of the three receivers for different number of pilots with mean pilot energy $E_p = 1/16$.

Increasing the pilot power helps to improve the BER performance of the *optimum receiver*, especially for low SNR values. This is visible in the resulting gap between the BER curves of the *optimum receiver* with $E_p = 1/4$ and $E_p = 1/16$ for SNR up to 25 dB. From Fig. 5.3 and Fig. 5.4 we observe that for high values of SNR the *optimum receiver* achieves almost the same BER performance irrespective of the number of pilots and their mean symbol energy. It can be seen in all figures that the receivers achieve the same diversity. Similar results have also been shown in [6], but for a different system model.

The MSE reduces by increasing the number of pilots and their mean symbol energy. This explains the improvement in BER performance of the *mismatched receiver* in Fig. 5.3 and 5.4. The *mismatched receiver* relies completely on the quality of the channel estimate, as its metric does not incorporate any methods to adapt to MSE. That is why the gap between the BER curves of the *mismatched receiver* and the *ideal receiver*

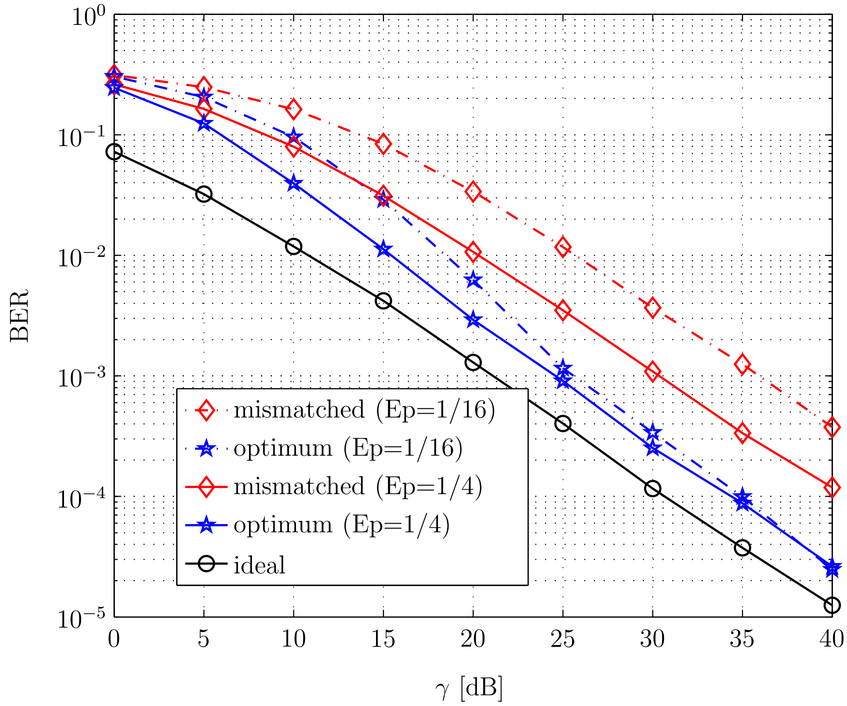


Figure 5.4: BER performance versus SNR of the three receivers, with $P = 2$ pilots and of mean energy $E_p = 1/16$ and $E_p = 1/4$.

remains constant for increasing SNR. From Fig. 5.4 it can be seen that the *mismatched receiver* for mean pilot symbol energy $E_p = 1/4$ needs about 6 dB less SNR than for $E_p = 1/16$ in order to achieve a BER $= 10^{-3}$.

5.6.3 AR(1) Channel Model

A total of $N = 8$ information bits are mapped to symbols using a BPSK ($M_a = 2$) modulation. We consider a AR(1) based channel model with the channel correlation coefficient α. The channel estimation for the *mismatched receiver* is done with an MMSE estimator.

In the following figures the SNR for the AR(1) based channel model is defined as

$$\gamma = 10 \log \frac{E_b}{\sigma_w^2} = 10 \log \frac{E_s}{\mathrm{ld}(M_a)}, \tag{5.34}$$

where σ_w^2 is the noise variance. Here, E_b and E_s are the respective mean bit and symbol

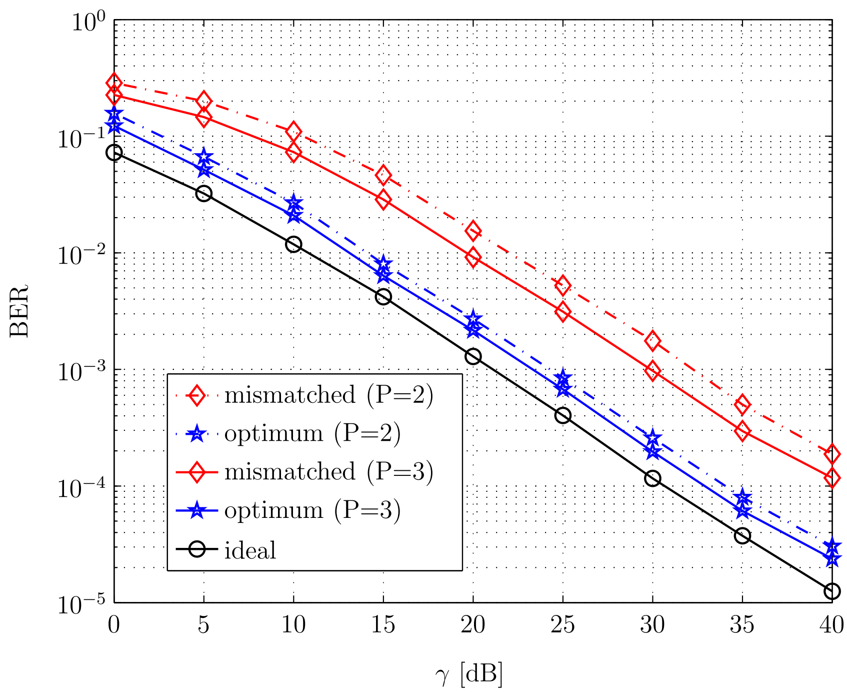


Figure 5.5: BER performance of the three receivers, with $\alpha = 1 - 10^{-6}$ and $E_p = 1/16$.

energies.

Fig. 5.5 shows the BER performance of the three receivers for $P = 2$ and $P = 3$ pilots with mean symbol energy $E_p = 1/16$ and a channel correlation coefficient $\alpha = 1 - 10^{-6}$. For $P = 2$, the SNR required to achieve a BER $= 10^{-3}$ is about 8 dB less if the *optimum receiver* is used instead of the *mismatched receiver*. There is no significant error floor in Fig. 5.5 because we have chosen a channel correlation coefficient very close to 1, see Fig. 5.1(a).

In Fig. 5.5 we can see the effect of decreasing the channel correlation coefficient. This figure shows the BER performance for channel correlation coefficients $\alpha = 1 - 10^{-3}$ and $\alpha = 1 - 10^{-6}$ for $P = 2$ pilots with mean symbol energy $E_p = 1/16$. For a BER $= 10^{-3}$ and $P = 2$ pilots the *optimum receiver* has a 3.1 dB gap to the *ideal receiver* whereas the *mismatched receiver* has 11.1 dB gap.

The error floor in the BER performance of the *mismatched* and *optimum* receivers can be clearly seen after a SNR of about 20 dB, which corresponds very well to the

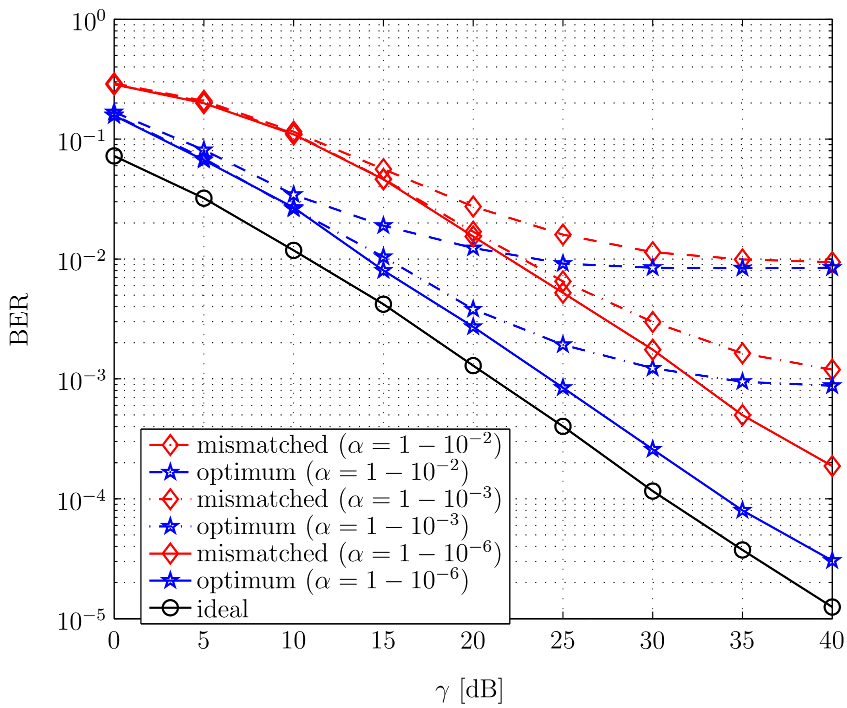


Figure 5.6: BER performance of the three receivers, with $\alpha = 1 - 10^{-2}$, $\alpha = 1 - 10^{-3}$ and $\alpha = 1 - 10^{-6}$ for $P = 2$ with $E_p = 1/16$.

theoretical error floor of the MSE for $\alpha = 1 - 10^{-3}$ shown in Fig. 5.1(a). By further decreasing the channel correlation coefficient the error floor can be at much lower SNR values. As shown in Fig. 5.1(a) the error floor for $\alpha = 1 - 10^{-2}$ starts at about 10 dB SNR and will be about the same for the BER performance of the *mismatched* and *optimum* receivers and can be seen in Fig. 5.6.

Thus, we have shown that the *optimum receiver* outperforms the *mismatched receiver* for both OFDM and AR(1) channel models.

6

Conclusions & Outlook

"[. . .] If I were again beginning my studies, I would follow the advice of Plato and start with mathematics." —— Galileo Galilei

In this thesis, we have investigated pilot-assisted communication systems for both iterative and non-iterative receiver architectures. The pilots are embedded in the transmitted sequence to facilitate channel estimation at the receiver, which was done using LS or MMSE estimators.

We have discussed the current research work on optimal receiver designs. These systems involve sophisticated technologies like OFDM, BICM or a combination of these two with multiple antennas.

The iterative receiver based on BICM-ID system has been described for a single antenna system. This receiver performs channel estimation prior to detection. A *modified receiver* was derived and adapted to a bit-wise soft decision decoding for this system. This was accomplished by incorporating the *posterior* distribution of the true channel conditioned on the channel estimate, into the likelihood function used for ML decoding. We made two important conclusions. Firstly, the ML metric of the *modified receiver* proved to be independent of the type of linear channel estimator used and secondly, it was able to suit itself to the quality of the available channel estimate.

The superior BER performance of the *optimal receiver* in comparison to a *mismatched receiver* using LS channel estimation was shown in the simulation results.

We have also derived an *optimal receiver* for a non-iterative system, which uses a ML metric derived by taking the channel correlation into account. The *optimal receiver* performs sequence detection without prior channel estimation, because the received pilot sequence is already incorporated in its ML metric. The ML metric of the *optimal receiver* was analyzed for low-complexity implementation. As a result, the computational time of the ML metric was significantly reduced. Simulation results have shown that the *modified receiver* outperforms the *mismatched receiver* in terms of BER.

There are a number of topics left open for further research on the *optimal receiver*, such as:

- The computational complexity of sequence detection increases exponentially in the length of the transmit sequence. This still poses a problem in the efficient implementation of the receiver. A low-complexity implementation of the *optimal receiver* can be done, e.g., using a sequence detector based on the Viterbi approach.
- Further, the BER performance of the *optimal receiver* could be researched for other channel models.
- The influence of the pilot placement on the BER performance could be examined.
- The *optimal receiver* could be extended to multiple antenna systems.

Bibliography

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications.* Cambridge University Press, 2005.

[2] S. Sadough, P. Piantanida, and P. Duhamel, "Achievable outage rates with improved decoding of bicm multiband ofdm under channel estimation errors," *Signals, Systems and Computers, 2006. ACSSC '06. Fortieth Asilomar Conference on*, pp. 1806–1810, Oct.-Nov. 2006.

[3] L. Tong, B. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *Signal Processing Magazine, IEEE*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[4] J. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh fading channels [mobile radio]," *Vehicular Technology, IEEE Transactions on*, vol. 40, no. 4, pp. 686–693, Nov 1991.

[5] G. Tarrico and E. Biglieri, "Space-time decoding with imperfect channel estimation," *Signal Processing and Information Technology, 2003. ISSPIT 2003. Proceedings of the 3rd IEEE International Symposium on*, pp. 302–305, 14-17 Dec. 2003.

[6] G. Taricco, "Optimum receiver design and performance analysis for fully correlated rician fading mimo channels with imperfect channel state information," *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pp. 1546–1550, 26-30 Nov. 2007.

[7] Y. Huang and J. Ritcey, "16-QAM BICM-ID in fading channels with imperfect channel state information," *Wireless Communications, IEEE Transactions on*, vol. 2, no. 5, pp. 1000–1007, Sept. 2003.

[8] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *Information Theory, IEEE Transactions on*, vol. 44, no. 3, pp. 927–946, May 1998.

[9] X. Li, A. Chindapol, and J. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8 psk signaling," *Communications, IEEE Transactions on*, vol. 50, no. 8, pp. 1250–1257, Aug 2002.

[10] A. Batra, J. Balakrishnan, G. Aiello, J. Foerster, and A. Dabak, "Design of a multiband ofdm system for realistic uwb channel environments," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 52, no. 9, pp. 2123–2138, Sept. 2004.

[11] H. Bölcskei, "MIMO-OFDM wireless systems: Basics, perspectives, and challenges," *IEEE Wireless Communications*, vol. 13, pp. 31–37, Aug. 2006. [Online]. Available: http://www.nari.ee.ethz.ch/commth/pubs/p/commag06

[12] P. Piantanida, G. Matz, and P. Duhamel, "Estimation-induced outage capacity of ricean channels," *Signal Processing Advances in Wireless Communications, 2006. SPAWC '06. IEEE 7th Workshop on*, pp. 1–5, 2-5 July 2006.

[13] S. Sadough, P. Piantanida, and P. Duhamel, "Mimo-ofdm optimal decoding and achievable information rates under imperfect channel estimation," *Signal Processing Advances in Wireless Communications, 2007. SPAWC 2007. IEEE 8th Workshop on*, pp. 1–5, June 2007.

[14] S. Alamouti, “A simple transmit diversity technique for wireless communications,” *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 8, pp. 1451–1458, Oct 1998.

[15] M. Rupp and C. Mecklenbrauker, “On extended alamouti schemes for space-time coding,” *Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on*, vol. 1, pp. 115–119 vol.1, Oct. 2002.

[16] E. Zehavi, “8-psk trellis codes for a rayleigh channel,” *Communications, IEEE Transactions on*, vol. 40, no. 5, pp. 873–884, May 1992.

[17] B. Muquet, “Novel receiver and decoding schemes for wireless ofdm systems with cyclic prefix or zero padding,” Ph.D. dissertation, ENST, Paris (France), June 2001.

[18] X. Li and J. Ritcey, “Bit-interleaved coded modulation with iterative decoding,” *Communications, 1999. ICC ’99. 1999 IEEE International Conference on*, vol. 2, pp. 858–863 vol.2, 1999.

[19] B. Hochwald and S. ten Brink, “Achieving near-capacity on a multiple-antenna channel,” *Communications, IEEE Transactions on*, vol. 51, no. 3, pp. 389–399, March 2003.

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Upper Saddle River (NJ), 1998.

[21] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate (corresp.),” *Information Theory, IEEE Transactions on*, vol. 20, no. 2, pp. 284–287, Mar 1974.

[22] J. Hagenauer, E. Offer, and L. Papke, “Iterative decoding of binary block and convolutional codes,” *Information Theory, IEEE Transactions on*, vol. 42, no. 2, pp. 429–445, Mar 1996.

[23] Y. Huang and J. Ritcey, "Exit chart analysis of bicm-id with imperfect channel state information," *Communications Letters, IEEE*, vol. 7, no. 9, pp. 434–436, Sept. 2003.

[24] A. Chindapol and J. Ritcey, "Design, analysis, and performance evaluation for bicm-id with square qam constellations in rayleigh fading channels," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 5, pp. 944–957, May 2001.

[25] K. Baddour and N. Beaulieu, "Autoregressive modeling for fading channel simulation," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 4, pp. 1650–1662, July 2005.

[26] R. Gray, "Toeplitz and circulant matrices: A review," 1977. [Online]. Available: citeseer.ist.psu.edu/gray01toeplitz.html

[27] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in ofdm systems," *Broadcasting, IEEE Transactions on*, vol. 48, no. 3, pp. 223–229, Sep 2002.

[28] R. Nilsson, O. Edfors, M. Sandell, and P. Borjesson, "An analysis of two-dimensional pilot-symbol assisted modulation for ofdm," *Personal Wireless Communications, 1997 IEEE International Conference on*, pp. 71–74, 17-19 Dec 1997.

[29] S. Le Goff, "Signal constellations for bit-interleaved coded modulation," *Information Theory, IEEE Transactions on*, vol. 49, no. 1, pp. 307–313, Jan 2003.

[30] E. Biglieri, G. Taricco, and A. Tulino, "Performance of space-time codes for a large number of antennas," *Information Theory, IEEE Transactions on*, vol. 48, no. 7, pp. 1794–1803, Jul 2002.

[31] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics.* Springer-Verlag, New York, 1999.

[32] H. Bölcskei and A. J. Paulraj, *Multiple-input multiple-output (MIMO) wireless systems*, 2002, pp. 90.1–90.14. [Online]. Available: http://www.nari.ee.ethz.ch/commth/pubs/p/mimo

[33] H. Bölcskei, *Principles of MIMO-OFDM wireless systems*, 2004. [Online]. Available: http://www.nari.ee.ethz.ch/commth/pubs/p/crc03

[34] J. Boutros, N. Gresset, and L. Brunel, "Turbo coding and decoding for multiple antenna channels," 2003. [Online]. Available: citeseer.comp.nus.edu.sg/721798.html

[35] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *Information Theory, IEEE Transactions on*, vol. 44, no. 2, pp. 744–765, Mar 1998.