Automatic Modulation Classification in Mobile OFDM Systems with Adaptive Modulation

Von der Fakultät für Ingenieurwissenschaften

Abteilung Elektrotechnik und Informationstechnik

der Universität Duisburg-Essen

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften

genehmigte Dissertation

von

Yun Chen

aus

Jiangxi, China

1. Gutachter: Prof. Dr.-Ing. Andreas Czylwik

2. Gutachter: Prof. Dr.-Ing. Robert Fischer

Tag der mündlichen Prüfung: 21.03.2013

Abstract

Adaptive modulation is an efficient way to combat the effects of deep fades in broadband orthogonal frequency division multiplexing (OFDM) systems with time-varying multipath propagation channels. Modulation schemes are adapted to the current channel state. Bandwidth efficient modulation schemes are applied on subcarriers with high channel quality, while robust modulation schemes or even no modulation are preferred for subcarriers in deep fades. The resulting benefit in terms of required transmit power was demonstrated for a fixed data rate in [1], where a gain of $5 \cdots 15$ dB was recorded for a BER of 10^{-3} over the OFDM system with a fixed modulation. Alternatively, adaptive modulation can also be used to improve the bandwidth efficiency. The advantage in terms of throughput was shown for a predefined quality of service in [2]. In literature, several algorithms have been proposed for adaptive modulation with different emphasis on bandwidth efficiency and implementation complexity [3], [4], [5]. In the thesis, the algorithm proposed in [6] is used.

A main drawback of adaptive modulation is that it requires information about the adapted modulation scheme at the receiver to enable demodulation. Traditionally, this information can be provided in forms of explicit signalling, which reduces the bandwidth efficiency due to the signalling overhead. In the thesis, proposals are developed to reduce this undesirable overhead. These proposals exploit the correlation properties inherently existing in the transmission channel in both time and frequency domain. These correlations lead to memory effects in the signalling source. State-dependent Huffman coding schemes are then applied to reduce the redundancy resulting from these memory effects [7].

This signalling overhead can be totally eliminated via automatic modulation classification (AMC). In the past, AMC was mainly of interest in military fields like threat analysis and electronic surveillance [8], where no prior knowledge about the used modulation scheme is available. Under such circumstance, maximum likelihood (ML) based AMC provides the optimum solution in the sense that the classification error probability is minimized. Nowadays, AMC is drawing more and more research interest also in civilian applications like systems with adaptive modulation, where certain cooperations are organized as in the system considered in this thesis. These cooperations provide certain prior information, which can be utilized to improve the classification reliability. Consequently, the ML based approach does not deliver the minimum error probability any more [9], [10] [11]. Investigations have to be conducted to verify how much the performance can be improved by incorporating this prior information into the AMC algorithm. As one focus in this thesis, a maximum a posteriori (MAP) based AMC is developed, which is potentially

able to minimize the classification error probability again. Another focus is to reduce the implementation complexity to enable the application of AMC in systems with high time requirements, e.g. real-time systems.

In the last part of the thesis, comparisons are performed between these two approaches, namely explicit signalling and signalling-free AMC, in terms of end-to-end packet error probability. To ensure a fair comparison, the net data rate is maintained as a constant in both operation modes.

Acknowledgements

There are many people, who have directly or indirectly supported me in this thesis, whom I would like to thank. Here I am not able to mention all of them. First of all, I would like to express my sincere thanks to my supervisor and the head of the department Communication Systems (Nachrichtentechnische Systeme, NTS) of University of Duisburg-Essen, Prof. Dr.-Ing. Andreas Czylwik, who provided me with this opportunity to work in his department and gave me valuable advice during this work and most possible flexibility in the arrangement of this work. Thanks also go to other members of the department, among them Dr.-Ing. Lars Häring for interesting discussions and the joint publications, Qiong Liu and Bo Zhao for their organizational supports with printing and submitting my thesis during the time I left the university.

I would like to also express my special thanks to Prof. Dr.-Ing. Robert Fischer from the Institute of Communications Engineering of University of Ulm for taking over the role of my second supervisor and his valuable comments and correction suggestions for this thesis. Thanks also go to other members of the exam commission Prof. Dr. Roland Schmechel, Prof. Dr.-Ing. Uwe Maier and prof. Dr. Rainer Kokozinski.

I would like to take this chance to express my thanks to my parents, who enabled my study in Germany. Although, we live far away from each other, I can always feel them near me. This thesis is dedicated to them.

Contents

Notation

The used mathematical symbols are listed here. These symbols are alphabetically sorted.

SYMBOLS		PAGE
a_n	Transmit symbol on subcarrier n (symbol only used for illustrating OFDM principle)	8
$a_{i,l}$	The l-th signal point in A_i	5
$\breve{a}_n(t)$	Complex-valued baseband time-continous signal on the subcarrier n (symbol only used for illustrating OFDM principle)	8
$\tilde{a}_k$	The k-th sample of $\breve{a}_n(t)$ in the time-discrete domain (symbol only used for illustrating OFDM principle)	9
$\mathbf{a}$	Transmit symbol vector (symbol only used for illustrating OFDM principle)	9
$\tilde{\mathbf{a}}$	Transmit time-domain symbol vector (symbol only used for illustrating OFDM principle)	9
$\tilde{\mathbf{a}}_{\mathrm{cp}}$	CP-extended time-domain symbol vector	11
$\mathcal{A}$	The set of symbol alphabets A_i with $A_i \in \mathcal{A}$	5
A_i	Symbol alphabet of the i-th modulation candidate	5
α_{m}	Parameter required for building the decision interval in a moment-based interval detector	74
$\mathbf{b}$	The generated bit loading vector at the BS	6
b_i	Bandwidth efficiency (modulation level) of the i-th modulation candidate A_i with $b_i \in \mathcal{B}$	5
$b^{(n)}$	Number of loaded bits (modulation level) on the n-th subcarrier with $b^{(n)} \in \mathcal{B}$	15
$\hat{b}^{(n)}$	Classified modulation level on the n-th subcarrier with $\hat{b}^{(n)} \in \mathcal{B}$	6
$\hat{b}_{i,\mathrm{HLRT}}$	Classified modulation level based on the HLRT classifier	40
$\hat{b}_{i,\mathrm{1P}}$	Classified modulation level based on the 1-point approximation	43
$\hat{b}_{i,\mathrm{ED}}$	Classified modulation level based on the Euclidean distance approximation	44
$\hat{b}_{i,\mathrm{4-P}}$	Classified modulation level based on the 4-point approximation	48
$\hat{b}_{i,\mathrm{MAP}}^{(n)}$	Classifed modulation level on the n-th subcarrier based on the MAP classifier	56

viii

Abbreviations

The used abbreviations are listed here. These symbols are alphabetically sorted.

ABBREVIATIONS		PAGE
1-P	1 point (approximation)	42
1-D	1 dimensional (approximation of the a posteriori probability)	58
2-D	2 dimensional (approximation of the a posteriori probability)	59
3-D	3 dimensional (approximation of the a posteriori probability)	61
3GPP LTE	3rd generation partnership project long term evolution	8
4-P	4 point (approximation)	45
ADSL	Asymmetric digital subscriber line	8
ALRT	Average likelihood ratio test	38
AM	Adaptive modulation	3
AMC	Automatic modulation classification	2
ASK	Amplitude-shifted keying	39
BER	Bit error ratio	6
BLV	Bit loading vector	18
BS	Base station	4
CE	Channel estimation	18
CP	Cyclic prefix	11
CSI	Cubic spline interpolation	92
DAB	Digital audio broadcasting	8
DFT	Discrete Fourier transform	10
DL	Down link	4
DVB	Digital video broadcasting	8
ED	Euclidean distance (approximation)	44
EGPRS	Enhanced general packet radio service	83
FB	Feature-based	3
FB AMC	Feature-based automatic modulation classification	39
FSK	Frequency-shifted keying	39
FSDH	Frequency domain state-dependent Huffman coding	24
FFT	Fast Fourier transform	10
GLRT	Generalized likelihood ratio test	38
GPRS	General packet radio service	1

ABBREVIATIONS		PAGE
HLRT	Hybrid likelihood ratio test	38
HOS	High order statistic	3
HSDPA	High speed downlink packet access	1
ICI	Inter-carrier interference	9
ISI	Inter-symbol interference	8
IDFT	Inverse discrete Fourier transform	9
LB	Likelihood-based	3
LB AMC	Likelihood based automatic modulation classification	38
LI	Linear interpolation	92
LPI	Low-pass interpolation	92
LTE	Long term evolution	1
MAP	Maximum a posteriori	4
MCS	Modulation and coding scheme	83
ML	Maximum likelihood	48
MS	Mobile station	4
MIMO	Multiple input multiple output	12
NoTx	No transmission	5
OFDM	Orthogonal frequency division multiplexing	2
PAPR	Peak-to-average power ratio	12
PI	Polynomial interpolation	92
PDF	Probability density function	38
PSK	Phase-shifted keying	39
QAM	Quadrature amplitude modulation	1
QoS	Quality of service	2
QPSK	Quaternary phase shifted keying	1
SDH	State-dependent Huffman (coding)	29
TDD	Time division duplex	4
UL	Uplink	4
UMPT	Uniformly most powerful test	38
WLAN	Wireless local area networks	8

CHAPTER 1

Introduction

1.1 Current state in modulation classification

Adaptive techniques play an increasingly important role in the transceiver design of modern communication systems to provide high performance and high data rate services. General packet radio service (GPRS), the cellular system of the 2.5-th generation, has introduced as the first mobile system adaptive coding schemes, which increase the data rate from 9.05 kbps to maximal 21.4 kbps for the case that one time slot is assigned to the user [12]. In the following generations other parameters like modulation levels become also adaptive as in high-speed downlink packet access (HSDPA) and long term evolution (LTE) systems, which promise a peak speed of above 100 Mbps [13].

The adapted parameters have to be provided at the receiver to enable data detection. Conventionally, this can be accomplished via explicit signalling, e.g. the transmitter informs the receiver via an extra control channel about the applied transmission parameters. In most practical systems, a fixed number of bits is reserved for the purpose of explicit signalling, e.g., in GPRS 2 bits (after channel coding 8 binary symbols) are designated to signal 4 possible coding schemes. In HSDPA, 5 bits (after channel coding 15 binary symbols) are spent to distinguish 32 classes of adaptive parameters.

Usually, the signalling information is organized at the beginning of each transmission burst to minimize demodulation delay and memory requirements [14]. In general, robust modulation and channel coding schemes are applied to enable a well-protected transmission of the signalling information. In this thesis, the modulation level of the quadrature amplitude modulation (QAM) schemes is to be adapted. The task is to provide the receiver with information about the used modulation level to enable the demodulation process. In [15] and [16], quaternary phase shifted keying (QPSK) is used to modulate the signalling bits required to handle four different QAM candidates. Three identical QPSK symbols are transmitted consecutively and majority voting is performed to improve the transmission robustness. Otsuki et al. proposed a four symbol Walsh function to represent four modulation levels of QAM schemes, which are available for adaptation [17]. At the receiver, the modulation level is detected by correlating the received base

band signal with all Walsh functions and searching for the Walsh function that maximizes the real part of the correlation. Furthermore, it is confirmed from the standpoint of computer simulation that this signalling scheme is more robust than the majority voting technique for a four-level adaptive system. An unequal error protection signalling scheme is proposed in [18] for a five level adaptive modulation system proposed in [19]. The adaptation strategy optimizes the level-switching thresholds to maximize the instantaneous throughput for a predefined quality of service (QoS). Five QAM schemes with different modulation levels are available for adaptation. A 5-PSK is proposed to represent these five modulation levels. Since different modulation levels operate in different system SNR ranges, these five PSK signalling symbols are transmitted under different link conditions. These five phasors are optimized in accordance with the respective link quality to provide best possible signalling robustness, which leads to non-uniformly distributed phasors (unequal error protection).

The explicit signalling has the advantage of simple implementation. However, it causes an undesired signalling overhead, which has to be maintained as low as possible in order not to lose too much bandwidth efficiency. A number of proposals have been developed to provide bandwidth efficient signalling schemes at the cost of a limited adaptation function. In [14], a three-message-model is suggested for an adaptive system that can only switch up, down or remain at the same modulation level. The corresponding entropy of the signalling source resulting from this special adaptation strategy is also derived. Another even more simple adaptive system can only allow two states of modulation levels [20]. Therefore, only 1 bit is required to indicate whether the modulation level has been changed or not.

These signalling schemes described above do not exploit any information-theoretical features of the signalling source, e.g. occurrence probabilities of individual modulation levels and especially the existing memory effects due to time- and frequency-domain correlation of the transmission medium. Consequently, they are simple to implement. However, it may bear a large amount of redundancy as will be confirmed in this thesis. We will show that with source coding schemes, which take the existing memory effect into account, the signalling overhead can be reduced significantly.

Although explicit signalling is easy to implement, it wastes the scarce spectrum resource. In case of orthogonal frequency-division multiplexing (OFDM) based systems, each subcarrier is a potential resource unit for individual link adaptation and consequently requires its own signalling information. To eliminate this signalling overhead and improve the spectral efficiency, the adapted modulation level can be detected blindly via automatic modulation classification (AMC).

Originally, AMC was mainly of interests in military fields like electronic surveillance, secured transmission and threat analysis [21], [22]. Nowadays, it is drawing increasing research interest also in civil applications, e.g. in intelligent systems, spectrum management, network traffic administration and especially systems with adaptive modulation (AM) [23], [24] and [25]. Two general frameworks have been established to solve modulation classification problems [8]. One approach extracts modulation specific features (signatures) to distinguish between selectable modulation candidates and is called feature-based (FB) framework. The other approach exploits probabilistic relationship between the observed signal and the modulation candidates and evaluates the likelihood function thereof, i.e. likelihood-based (LB) framework.

In the FB framework, the discriminating features are normally chosen as information-bearing parameters like the carrier frequency, the instantaneous phase or amplitude information [26], [27], [28]. The main advantage of this method lies in efficient implementation and robustness against mismatch between theoretical model assumptions and real system conditions. Although an FB method may not be optimal, under real conditions it may provide a near-optimal classification performance if designed properly [29].

The LB framework formulates the likelihood function for each modulation candidate and decides on the hypothesis which provides the maximum value among all. So this method is viewed as a multiple hypothesis testing problem. In Bayesian sense, this approach represents the optimal classifier under ideal conditions and absence of knowledge about the modulation selection strategy. Consequently, it attracts much research interest. Especially for systems with additive noise only, if noise statistics are well defined, a solid theoretical framework has been established [30], [31], [32], [33]. Wei was able to derive a generic formula to asymptotically quantify the classification error probability [34]. Therefore, in later works concerning modulation classification in mobile radio environments, the LB approach was adopted to systems with fading channels [35], [36], [37], [38].

The emergence of the transmission technology OFDM opens also new opportunities, where the adaptive modulation can be performed on per-subcarrier basis. The required signalling overhead increases linearly with the number of subcarriers. Blind modulation detection is of great importance to minimize this signalling overhead and increase the bandwidth efficiency. Several types of adaptation strategies have been developed for different applications. One strategy targets at a predefined bit error ratio (BER) and tries to maximize the throughput by optimizing the switching levels [2]. Here several AMC algorithms have been proposed, i.e. minimum Euclidean distance based and multi-mode trellis decoder based AMC in [2] or modulation classification based on high order statistic (HOS) in [39], [40] and [43]. A more straightforward way is to estimate the

received SNR level and based on the switching threshold to blindly detect the adapted modulation scheme [41], [98].

Another strategy fixes the data rate and aims at minimizing the BER for a given transmit power. Since here the switching levels are normally not fixed, the LB framework is preferred to perform the modulation classification [44], [42], [9], [11], [48], [45], [46], [47].

Adaptive modulation on per-subcarrier basis requires modulation classification for each subcarrier, consequently the classification complexity increases also linearly with the number of subcarriers. For the computationally intensive LB framework, it may quickly exceed the computational capacity of the receiving terminal necessitating the search for simplification of likelihood functions. In this thesis, several methods will be proposed to simplify the underlying Bayesian likelihood function and consequently reduce the computational complexity. Furthermore, in an adaptive OFDM system with certain cooperation, where the receiver is at least informed about the applied adaptation strategy, certain prior information is available to assist the modulation classification. Consequently, the ML approach, which exploits only the received signal as the single information for modulation classification, is unable to minimize the classification error probability any more [10], [48], [9]. Investigations have to be conducted to show how much the performance can be improved by incorporating this prior knowledge into the AMC algorithm. In this thesis, maximum a posteriori (MAP) based algorithms will be developed, which can potentially minimize the classification error probability in the considered adaptive system again. Approximations will be developed to enable a practically feasible implementation.

1.2   Problem formulation

In multi-carrier transmission systems, the total transmit bandwidth is subdivided into multiple subchannels. Due to strongly time-dispersive nature of typical wireless channels, the link quality of individual subchannels distinguish from each other significantly. With a fix modulation scheme, subchannels with poor quality require that low order modulation formats have to be used to ensure certain QoS leading to high capacity waste. Adaptive modulation is a promising technique to efficiently utilize available resources by adapting modulation schemes to individual radio links. However, the adapted modulation schemes have to be provided at the receiver to enable demodulation.

In this dissertation, we consider a time-division duplex (TDD) based OFDM system with N data subcarriers. In the downlink (DL) transmission, the base station (BS) performs subcarrier-based AM according to the last uplink (UL) channel state vector $\mathbf{H}^{\mathrm{U}} = (H_0^{\mathrm{U}}, H_1^{\mathrm{U}}, ..., H_{N-1}^{\mathrm{U}})^{\mathrm{T}}$, e.g. perceived via preamble-based channel estimation. The mobile station (MS) has to classify the adapted modulation schemes to demodulate the re-

ceived DL signals with the aid of the DL channel state vector $\mathbf{H}^{\mathrm{D}} = (H_0^{\mathrm{D}}, H_1^{\mathrm{D}}, ..., H_{N-1}^{\mathrm{D}})^{\mathrm{T}}$ as shown in Figure 1.1. Assuming perfect time and frequency synchronization and that the use of a cyclic prefix eliminates inter-symbol interference, the received complex baseband signal is given by [58]:

$$r_{k,n} = H_{k,n}^{\mathrm{D}} s_{k,n} + w_{k,n}, \quad 0 \le k \le K - 1; \quad 0 \le n \le N - 1. \tag{1.1}$$

- k: OFDM symbol index, n: data subcarrier index, N: number of data subcarriers, K: number of OFDM symbols in a observed packet.

- $H_{k,n}^{\mathrm{D}}$: DL channel transfer function with $\mathrm{E}_{k,n}[|H_{k,n}^{\mathrm{D}}|^2] = 1$, the expectation is evaluated over the index k and n.

- $w_{k,n}$: the zero-mean complex-valued white Gaussian noise with a double-sided power spectral density $\frac{N_0}{2}$.

- $s_{k,n}$: the transmit symbol drawn randomly and independently from the symbol alphabet of a certain modulation scheme. In the sequel, we consider a finite-size candidate set of I quadrature amplitude modulation (QAM) schemes with the set of symbol alphabets $\mathcal{A} = \{A_1, A_2, ..., A_I\}$ and the set of corresponding bandwidth efficiencies (modulation levels) $\mathcal{B} = \{b_1, b_2, ..., b_I\}$. The i-th symbol alphabet A_i has the symbol set $\{a_{i,1}, a_{i,2}, ..., a_{i,M_i}\}$ with $\mathrm{E}_l[|a_{i,l}|^2] = 1$ and $M_i = 2^{b_i}$ denoting the number of constellation points correspondingly. The bits are mapped to signal points according to Gray-mapping [87].
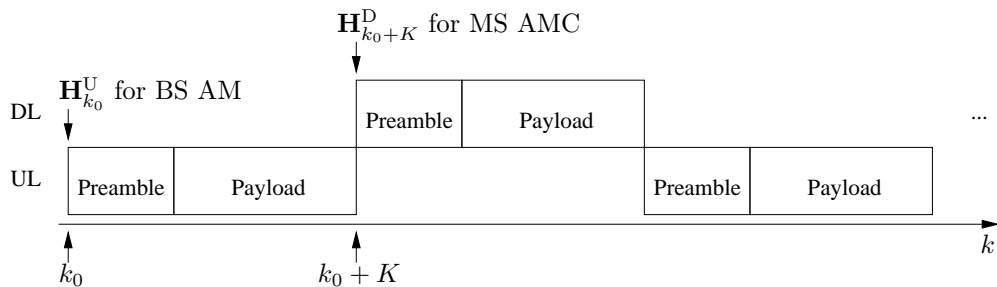
Figure 1.1: TDD-based AM and AMC data structure from the transmitter view

Throughout this dissertation, the following modulation schemes are considered: BPSK, 4-QAM, 16-QAM, 64-QAM which are widely used in modern radio systems like IEEE 802.11 a. In addition, we also consider a special case: no transmission (NoTx), which is used to handle extremely poor channel conditions. The signal constellations are schematically represented in Figure 1.2, corresponding to the modulation level set $\mathcal{B} = \{b_i\}_{i=1}^5 =$

$\{0, 1, 2, 4, 6\}$. So the adaptive modulation here is identical to an adaptive bit loading problem. The BS generates a bit loading vector $\mathbf{b} = (b^{(0)}, b^{(1)}, ..., b^{(N-1)})^{\mathrm{T}}$ according to certain criteria given in Section 2.2.1, where $b^{(n)}$ is the number of loaded bits on the n-th subcarrier with $b^{(n)} \in \mathcal{B}$. There are in general two kinds of bit loading. In one type, The BS may adaptively load a certain number of bits on each subcarrier based on $\mathbf{H}^{\mathrm{U}}$ and a given data rate R with the aim to minimize the bit error ratio (BER). In the other type, the transmitter may maintain certain QoS and maximize the throughput. We will consider the first type of bit loading, which keeps the data rate constant, since in many applications a constant data rate is desired. The details of this kind of bit loading will be provided in Section 2.2.
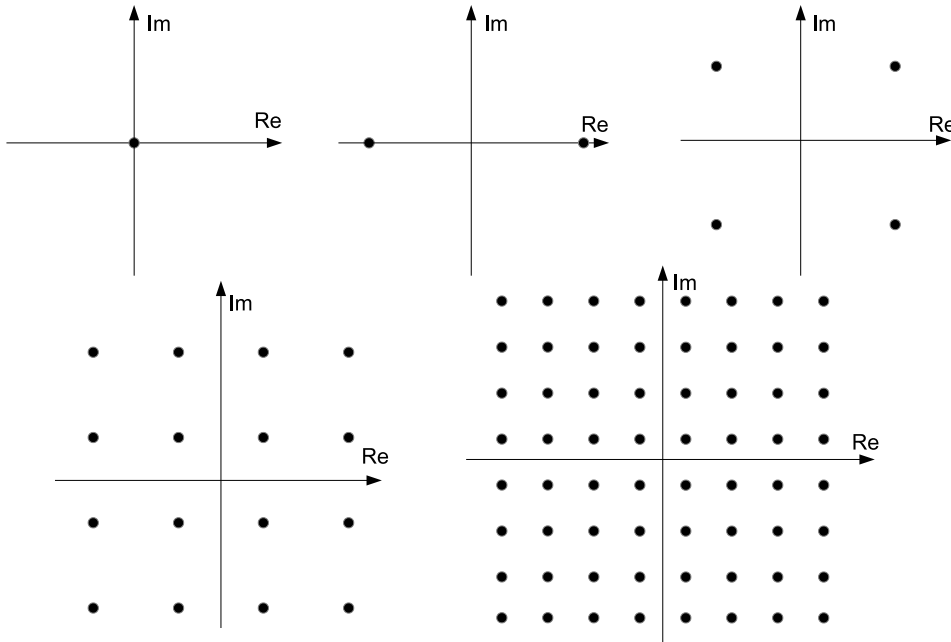
Figure 1.2:   Signal constellations of considered modulation candidates

The object of this dissertation is to detect $\mathbf{b}$, which is equivalent to classify b_i on each subcarrier as $\mathbf{b} = (b^{(0)}, b^{(1)}, ..., b^{(N-1)})^{\mathrm{T}}$ with $b^{(n)} \in \mathcal{B}$. The detected bit loading vector is denoted as $\hat{\mathbf{b}} = (\hat{b}^{(0)}, \hat{b}^{(1)}, ..., \hat{b}^{(N-1)})^{\mathrm{T}}$ with $\hat{b}^{(n)} \in \mathcal{B}$ too.

1.3   Structure of dissertation

This thesis is divided into six chapters. After introduction, some general points of the OFDM technology and adaptive OFDM concepts are given in Chapter 2. In Chapter 3, signalling is applied to communicate the adapted modulation schemes from the BS to the MS. Emphasis will be put on the possibility of reduction of required signalling

overhead. Automatic modulation classification will be investigated in depth in Chapter 4 as a signalling-free alternative. This method will be addressed from two different concepts, namely the likelihood-based framework and the feature-based framework. Also system-specific properties will be utilized to enhance the classification reliability. In the Chapter 5, signalling and automatic modulation classification will be compared with each other in terms of packet error ratio under the condition that the net data rate is constant. Then a brief conclusion will be given in the Chapter 6. In Appendices, some important mathematical derivations and simulation results are given to assist the descriptioins in the mentioned chapters.

CHAPTER 2

Adaptive OFDM

2.1 Basics of OFDM

Due to its high flexibility in resource allocation and ability to transform a inter-symbol interference (ISI) channel into multiple frequency-flat subchannels, OFDM has found widely used applications in wired as well as wireless communication systems. Digital audio broadcasting (DAB) and video broadcasting (DVB) systems are first consumer-oriented wireless OFDM systems followed by asymmetric digital subscriber line (ADSL) as the first OFDM-based wired system. Also in wireless local area networks (WLAN) systems, OFDM is a popular transmission technology. Recently, OFDM has been selected for the next generation cellular systems: the 3-rd generation partnership project long term evolution (3GPP LTE). An informative overview of the development of the OFDM technology can be found in [49], [50].

The basic idea is to split a high-speed serial data stream into a large number of N parallel low-speed substreams, so that the symbol duration is N times longer than the original one as shown in Figure 2.1 for $N = 8$. Consequently, OFDM is more robust against time-dispersive channels. Assume that the duration of an OFDM symbol is T and these N substreams are modulated on N harmonic sinusoids (subcarriers) with respecitve carrier frequency at

$$f_n = n\frac{1}{T} \quad \text{for} \quad n = 0, 1, \cdots, N - 1, \tag{2.1}$$

then the complex-valued base-band signal on subcarrier n is given by

$$\breve{a}_n(t) = a_n g_{\mathrm{p}}(t) \mathrm{e}^{\mathrm{j}2\pi f_n t}, \tag{2.2}$$

where a_n is the data symbol to be transmitted on subcarrier n and $g_{\mathrm{p}}(t)$ is the pulse shaping filter with

$$g_{\mathrm{p}}(t) = \begin{cases} 1 & : \text{if } 0 \leq t \leq T \\ 0 & : \text{otherwise.} \end{cases} \tag{2.3}$$

Normally, one OFDM symbol is followed by another. For illustration purpose, we consider here only one symbol.
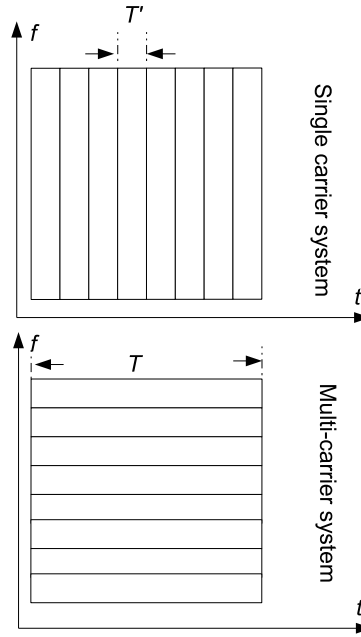
Figure 2.1: Visulization over time and frequency for single carrier and multi-carrier system [51]

It can be shown that the subcarrier signals $\breve{a}_n(t)$ are orthogonal to each other. There exists no inter-carrier interference (ICI). This property allows the simultaneous transmission of N data streams. The above idea would need N modulators/demodulators, which require high implementation complexity. The practical implementation of frequency multiplexing is performed via the inverse discrete Fourier transform (IDFT) .

2.1.1 Generation of OFDM signals via IDFT

Performing an equi-distant sampling of $\breve{a}_n(t)$ with a sampling rate of $f_\mathrm{s} = \frac{N}{T}$, we denote the k-th sample as

$$\tilde{a}_k = \breve{a}_n(\frac{kT}{N}), \tag{2.4}$$

then multiplexed signal in the time-discrete domain is given by

$$\tilde{a}_k = \sum_{n=0}^{N-1} a_n \exp(\mathrm{j}2\pi k \frac{n}{N}), \tag{2.5}$$

where k is the sample index. In fact, (2.5) is the IDFT of the parallel data symbols $\{a_n\}_{n=0}^{N-1}$, with the value taken at k-th sample [52]. Equivalently, (2.5) can be expressed in a vector matrix representation

$$\tilde{\mathbf{a}} = \mathbf{N}_\mathrm{IDFT}\, \mathbf{a}, \tag{2.6}$$

with $\mathbf{a} = (a_0, a_1, \cdots, a_{N-1})^{\mathrm{T}}$, $\tilde{\mathbf{a}} = (\tilde{a}_0, \tilde{a}_1, \cdots, \tilde{a}_{N-1})^{\mathrm{T}}$ and $\mathbf{N}_{\mathrm{IDFT}}$ the $N \times N$ IDFT matrix given by

$$\mathbf{N}_{\mathrm{IDFT}} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{(\mathrm{j}2\pi 1 \frac{1}{N})} & \cdots & e^{(\mathrm{j}2\pi 1 \frac{N-1}{N})} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{(\mathrm{j}2\pi(N-1)\frac{1}{N})} & \cdots & e^{(\mathrm{j}2\pi(N-1)\frac{N-1}{N})} \end{pmatrix}, \tag{2.7}$$

where the factor $\frac{1}{\sqrt{N}}$ is introduced to normalize the transmit power.

At the receiver side the discrete Fourier transform (DFT) is performed to recover the transmitted signal

$$\mathbf{N}_{\mathrm{DFT}} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{(-\mathrm{j}2\pi 1 \frac{1}{N})} & \cdots & e^{(-\mathrm{j}2\pi 1 \frac{N-1}{N})} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{(-\mathrm{j}2\pi(N-1)\frac{1}{N})} & \cdots & e^{(-\mathrm{j}2\pi(N-1)\frac{N-1}{N})} \end{pmatrix}. \tag{2.8}$$

The direct implementation of these two operations would require high computational efforts, especially for large DFT window size N as normally the case in practical systems. Practically feasible OFDM-based applications are enabled by the efficient implementation via fast Fourier transform (FFT) [53].

2.1.2   Cyclic prefix

In practice, we encounter a time-variant multipath channel described by the time-variant weighting function

$$\tilde{h}_{\mathrm{real}}(k, k') = \sum_{p_{\mathrm{k}}=0}^{P_{k'}-1} \tilde{h}_{p_{\mathrm{k}}}(k') \delta(k - k_{p_{\mathrm{k}}}(k')), \tag{2.9}$$

where k, k' and $k_{p_{\mathrm{k}}}$ are delay, observation, and encountered path delay index; $P_{k'}$ and $\tilde{h}_{p_{\mathrm{k}}}(k')$ are the time-variant number of path delays and time-variant complex-valued path gain of the $k_{p_{\mathrm{k}}}$-th path. For illustration purpose, we ignore the time variance for this moment and assume an equidistantly spaced path delay. The channel impulse response is then given by

$$\tilde{h}(k) = \sum_{p_{\mathrm{k}}=0}^{P_{\mathrm{k}}-1} \tilde{h}_{p_{\mathrm{k}}} \delta(k - k_{p_{\mathrm{k}}}), \tag{2.10}$$

The k-th received sample is given by the linear convolution of the transmit samples with the impulse response (for simple illustration purpose the additve noise is ignored)

$$\tilde{r}_k = \sum_{p_{\mathrm{k}}=0}^{P_{\mathrm{k}}-1} \tilde{h}_{p_{\mathrm{k}}} \tilde{a}_{k-p_{\mathrm{k}}}, \tag{2.11}$$

which as a compact form, can be given in the following matrix representation

$$\tilde{\mathbf{r}} = \begin{pmatrix} \tilde{h}_0 & 0 & \cdots & \cdots & \cdots & 0 \\ \tilde{h}_1 & \tilde{h}_0 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \tilde{h}_{P_k-1} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \tilde{h}_{P_k-1} & \cdots & \tilde{h}_1 & \tilde{h}_0 \end{pmatrix}_{N \times N} \begin{pmatrix} \tilde{a}_0 \\ \tilde{a}_1 \\ \vdots \\ \tilde{a}_{N-1} \end{pmatrix}. \tag{2.12}$$

It can be shwon by applying the DFT to (2.12) that the orthogonality is destroyed for the received signal $\tilde{\mathbf{r}}$ leading to ICI in (2.12). A method to preserve the orthogonality is to extend $\tilde{\mathbf{a}}$ with a cyclic prefix (CP) of length K_{cp} with $K_{\text{cp}} \geq P_k$, where the CP is a duplicate of the last K_{cp} samples of $\tilde{\mathbf{a}}$ [54], [55], [57]. The extended transmit signal is given by

$$\tilde{\mathbf{a}}_{\text{cp}} = (\underbrace{\tilde{a}_{N-K_{\text{cp}}+1}, \cdots, \tilde{a}_{N-1}}_{\text{cyclic prefix}}, \tilde{a}_0, \tilde{a}_1, \cdots, \tilde{a}_{N-1})^{\text{T}}. \tag{2.13}$$

To show how the orthogonality is preserved, we assume $K_{\text{cp}} = P_k$ for the purpose of simple illustration. The received signal with cyclic prefix is given by

$$\tilde{\mathbf{r}}_{\text{cp}} = \begin{pmatrix} \tilde{h}_0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \tilde{h}_1 & \tilde{h}_0 & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \tilde{h}_{P_k-1} & \ddots & \ddots & \tilde{h}_0 & 0 & \ddots & \ddots & \vdots \\ 0 & \tilde{h}_{P_k-1} & \ddots & \tilde{h}_1 & \tilde{h}_0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \tilde{h}_1 & \tilde{h}_0 & 0 & \vdots \\ \vdots & \ddots & \ddots & \tilde{h}_{P_k-1} & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \tilde{h}_{P_k-1} & \cdots & \tilde{h}_1 & \tilde{h}_0 \end{pmatrix}_{(N+P_k)\times(N+P_k)} \begin{pmatrix} \tilde{a}_{P_k-1} \\ \vdots \\ \tilde{a}_{N-1} \\ \tilde{a}_0 \\ \tilde{a}_1 \\ \vdots \\ \tilde{a}_{N-1} \end{pmatrix}. \tag{2.14}$$

After removal of the cyclic prefix, the received signal is given by

$$\tilde{\mathbf{r}} = \begin{pmatrix} \tilde{h}_0 & 0 & \cdots & 0 & \tilde{h}_{P_k-1} & \tilde{h}_{P_k-2} & \cdots & \tilde{h}_1 \\ \tilde{h}_1 & \tilde{h}_0 & 0 & \ddots & 0 & \tilde{h}_{P_k-1} & \ddots & \tilde{h}_2 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \tilde{h}_{P_k-2} & \ddots & \ddots & \tilde{h}_0 & 0 & \ddots & \ddots & \tilde{h}_{P_k-1} \\ \tilde{h}_{P_k-1} & \tilde{h}_{P_k-2} & \ddots & \ddots & \tilde{h}_0 & 0 & \ddots & 0 \\ 0 & \tilde{h}_{P_k-1} & \ddots & \ddots & \ddots & \tilde{h}_0 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \tilde{h}_{P_k-1} & \tilde{h}_{P_k-2} & \cdots & \tilde{h}_1 & \tilde{h}_0 \end{pmatrix}_{N \times N} \begin{pmatrix} \tilde{a}_0 \\ \tilde{a}_1 \\ \vdots \\ \tilde{a}_{N-1} \end{pmatrix}. \tag{2.15}$$

By adding cyclic prefix, (2.12) is transformed into (2.15), which corresponds to a transformation of the linear convolution into a linear cyclic convolution (elementwise view of $\mathbf{r}$). Now, a DFT operation can seperate the transmitted signal $\mathbf{a}$ without ICI as illustrated in detail in [55]

$$\mathbf{r} = \begin{pmatrix} H_0 & 0 & \cdots & 0 \\ 0 & H_1 & \cdots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & H_{N-1} \end{pmatrix}_{N \times N} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{pmatrix}. \tag{2.16}$$

With the CP extension at the transmitter and CP removal at the receiver, the received signal is distorted by a fading factor H_n (elementwise multiplication in frequency domain), which can be equalized by a simple multiplication of its reciprocal value $\frac{1}{H_n}$ enabling a simple receiver structure. On the other side, the cyclic extension has also drawbacks. It reduces bandwidth efficiency and wastes transmit power as this extension does not carry any non-redundant information.

2.1.3   Overview of advantages and disadvantages

Some typical advantages and disadvantages of the OFDM technology are listed here. More details can be found in the literature like [58], [59], [60] and [61]. Three important advantages are

- Suitable for applications in severe frequency selective channels requiring no complex time-domain equalization.

- High spectral efficiency in combination with multiple input multiple output (MIMO) technology.

- High flexibility in terms of resource allocation, e.g. application of adaptive techniques as will be discussed in the next section.

Some disadvantages are listed in the following:

- High peak-to-average power ratio (PAPR) due to multiplexing N signal branches leading to high challenges for power amplifiers.

- High sensitivity to synchronization impairments, e.g. due to imperfect Doppler-shift compensation.

- Reduction of power and spectral effiency due to CP and high amount of pilot symbols inserted for channel estimation purposes.

2.2   Adaptive techniques in OFDM

Data transmission via time-varying fading channels is subject to burst-wise error pattern, where the channel enters a deep fade for a relatively long duration. Such situation degrades the overall system performance. An efficient way to deal with this problem is to adapt the transmission scheme to the prevailing channel state.

Adaptive techniques have been investigated for a long time. At the beginning, adaptive power control was studied, which raises the power level if the channel quality falls under a predefined level, so that a quasi-constant SNR is maintained [62]. This technique may cause two problems: 1, the transmit power may exceed the given power constraint for very poor link conditions; 2, the increased power may cause severe cochannel interference. Another way is to change the pulse/symbol duration (or data/symbol rate) in response to the channel fading rate, where e.g. the modulator of such a system creates a low-pass signal $m_k^{(LP)}(t)$ with $k = 0$ or 1 to modulate the carrier signal [63], [64]. These two low-pass signals $m_k^{(LP)}(t)$ with $k = 0$ or 1 are orthogonal to each other, e.g. orthogonal sinusoids, of duration $\frac{1}{R^{(LP)}(t)}$, where $R^{(LP)}(t)$ is a adjustable symbol rate. The adaptive symbol rate $R^{(LP)}(t)$ is changed continuously with the received signal strength variation with the help of a feedback channel. Such a system is in deed a frequency shifted keying, where the freqeuncy spacing is a funtion of the channel fading variation. Obviously, due to the strong fluctuation nature of wireless channels, the variation in symbol duration could cause an expansion in bandwidth requirements, which is highly undesirable in view of the already very scarce radio spectrum resource. A more attractive method is to vary the coding rate or modulation level in accordance to the channel gain [65], [15]. There are also proposals, which jointly vary several parameters to achieve a higher adaptation granularity [66], [67], [68]. Note that adaptive techniques can be applied on duplex systems only, where the transmitter is able to possess the prevailling channel state information, e.g. via receiver feedback.

With the emergence of OFDM technology, the research interest in adaptive techniques has reached a new quality. In OFDM systems, each subcarrier can be viewed as a frequency flat subchannel and apply an individual transmission scheme. Consequently, compared to the single-carrier counterparty multicarrier systems are more flexible in terms of parameter adaptation as the transmitter can vary its transmission scheme in both time and frequency direction.

In the history, adaptive modulation has been shown as a powerful adaptive technique to improve the overall system performance [1], [2]. It can be used to minimize the error probability with maintaining a constant throughput and transmit power or to maximize the throughput with maintaining a constant power level and link quality. In [1], it was

shown for a fixed throughput that a power gain of $5 \cdots 15$ dB can be achieved for a BER of 10^{-3} if the receiver was perfectly informed about the adapted modulation schemes. The advantage of throughput improvement was documented for a predefined QoS in [2], where M-PSK symbols were used to perform the signalling.

2.2.1   Bit loading algorithm

Adaptive modulation is indeed adaptive bit loading, where more bits are loaded to sub-carriers with better link quality and less or no bit is loaded to deeply faded subcarriers. Several bit loading algorithms have been proposed with different emphasis on utilization of channel capacity and implemental complexity [6] [4] [3] [5]. In this paper, the bit loading is based on the algorithm proposed by Chow in [6]. Some modifications are made to reduce the implementation complexity [56]. In [1], it was shown that an additional adaptive power allocation provided only a marginal gain, so that with respect to the associated additional implemental complexity, the adaptive power control part of the algorithm is not applied in this thesis.

Here the used algorithm is briefly described. It has to fulfil two conditions

$$\sum_{n=0}^{N-1} b^{(n)} = B, \tag{2.17}$$

$$p_{\mathrm{AM}} = \min_{\mathbf{b}} p_{\mathrm{bit}}(\mathbf{b}), \quad \mathbf{b} = (b^{(0)}, ..., b^{(N-1)})^{\mathrm{T}}, \quad b^{(n)} \in \mathcal{B}$$

where $b^{(n)}$ is the number of bits loaded on the n-th subcarrier, B the total target number of bits, $\mathbf{b}$ the bit loading vector, $p_{\mathrm{bit}}(\mathbf{b})$ the resulting overall bit error probability if $\mathbf{b}$ is applied, $\mathcal{B}$ is the set of bit levels used for adaptive modulation and p_{AM} is the overall bit error probability after the adaptive bit loading is completed. The above equations say that in an OFDM symbol a certain target number of bits B has to be loaded meanwhile the overall bit error probability p_{bit} has to be minimized. In packet-based transmission systems, a transmission packet is comprised of several OFDM symbols. Optimum performance is achieved if the bit loading is carried out for each OFDM symbol individually. However this would cause enormous complexity. For short packet length and slowly time-varying channels, a reasonable solution is to perform packet-based adaptive modulation.

The algorithm carries out following steps based on [6] and [56]:

1. Measure the channel state $|H_n|$ for each subcarrier n at the transmitter. Initialize parameters: set the iteration counter $l_{\mathrm{it}} = 0$, $\bar{B}(l_{\mathrm{it}} = 0) = 0$ and $\lambda_{\mathrm{margin}}(l_{\mathrm{it}} = 0) = 1$ (or 0 dB if given in dB). Here $\bar{B}(l_{\mathrm{it}})$ denotes the already loaded number of bits in the l_{it}-th iteration and $\lambda_{\mathrm{margin}}(l_{\mathrm{it}})$ is called the system performance margin

indicating how much additional noise can be tolerated. If $\bar{B}(l_{\mathrm{it}} = 0) \neq B$, $\lambda_{\mathrm{margin}}(l_{\mathrm{it}})$ has to be adjusted correspondingly to satisfy the first equation of (2.17). Each increment/reduction of $\lambda_{\mathrm{margin}}$ means a degradation/improvement of the bit error probability.

2. While $\bar{B}(l_{\mathrm{it}}) \neq B$, do:

- Determine the number of bits which theoretically can be loaded on subcarrier n according to

$$\widetilde{b}^{(n)}(l_{\mathrm{it}}) = \log_2(1 + \frac{\rho_{n,\mathrm{r}}^2}{\Gamma \lambda_{\mathrm{margin}}(l_{\mathrm{it}})}), \quad n = 0, ..., N-1$$

$$\widetilde{B}(l_{\mathrm{it}}) = \sum_{n=0}^{N-1} \widetilde{b}^{(n)}(l_{\mathrm{it}}) \tag{2.18}$$

where $\rho_{n,\mathrm{r}}^2 = \frac{e_n|\hat{H}_n|^2}{N_0}$ is the received SNR in linear scale, e_n is the transmit power on subcarrier n which is constant over all n (no adaptive power loading) and Γ is the so called SNR gap also in linear scale, which characterizes the considered transmission system [69].

- Round down $\widetilde{b}^{(n)}(l_{\mathrm{it}})$ to the next smaller integer number belonging to $\mathcal{B}$

$$\mathrm{Round}(\widetilde{b}^{(n)}(l_{\mathrm{it}})) = \bar{b}^{(n)}(l_{\mathrm{it}}), \quad \text{with } \bar{b}^{(n)}(l_{\mathrm{it}}) \in \mathcal{B}$$

$$\Delta\widetilde{b}^{(n)}(l_{\mathrm{it}}) = \widetilde{b}^{(n)}(l_{\mathrm{it}}) - \bar{b}^{(n)}(l_{\mathrm{it}})$$

$$\bar{B}(l_{\mathrm{it}}) = \sum_{n=0}^{N-1} \bar{b}^{(n)}(l_{\mathrm{it}}). \tag{2.19}$$

- If $\bar{B}(l_{\mathrm{it}}) < B$

    Add bits to these subcarriers, which result in minimum increase in p_{bit}. Note that p_{bit} is not necessarily to be determined for AM. These subcarriers are determined by maximizing $\Delta\widetilde{b}^{(n)}(l_{\mathrm{it}})$ in (2.19).

    else

    Remove bits from these subcarriers, which result in maximum decrease in p_{bit}. These subcarriers are determined by minimizing $\Delta\widetilde{b}^{(n)}(l_{\mathrm{it}})$.

- Update $\lambda_{\mathrm{margin}}(l_{\mathrm{it}})$ given by

$$\lambda_{\mathrm{margin}}(l_{\mathrm{it}} + 1) = 2^{\frac{\bar{B}(l_{\mathrm{it}})-B}{N}} \lambda_{\mathrm{margin}}(l_{\mathrm{it}}) \tag{2.20}$$

- $l_{\mathrm{it}} = l_{\mathrm{it}} + 1$

3. if $\bar{B}(l_{\mathrm{it}}) = B$, the bit loading process is completed; set $l_{\mathrm{over}} = l_{\mathrm{it}}$ and $b^{(n)} = \bar{b}^{(n)}(l_{\mathrm{over}})$. Otherwise, go back to step 2.

The result of the algorithm is the bit loading vector denoted by $\mathbf{b} = (b^{(0)}, b^{(1)}, ..., b^{(N-1)})^{\mathrm{T}}$. In most cases $\bar{B}(1) \neq B$, so an interative adjustments of $\lambda_{\mathrm{margin}}$ is required. In each interation, additional bits are loaded/removed on/from these subcarriers which result in minimum/maximum increase/decrease in p_{bit}. To fulfil the conditions in (2.17), a scan process across the subcarriers is required which results in cross-subcarrier optimazation of the bit loading. So $b^{(n)}$ is not determined by H_n only, rather it is jointly determined by H_n and $\lambda_{\mathrm{margin}}(l_{\mathrm{over}})$, which in turn depends on B and $\mathbf{H} = (H_0, ..., H_n, ..., H_{N-1})^{\mathrm{T}}$. This knowledge will be exploited in Chapter 4 to enhance the reliability of modulation classification.
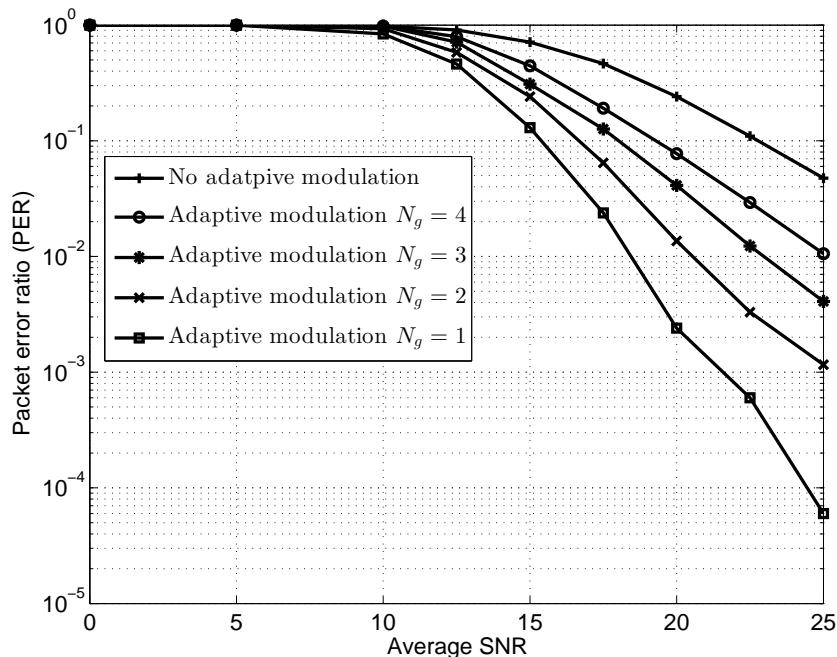
Figure 2.2:   Adaptive subband bit loading vs. fixed modulation

2.2.2   Performance improvement in PER

The above bit loading algorithm is applied on the considered system, whose parameters, e.g. data rate, channel coding, used radio channel model and channel estimation scheme are described in detail in Appendix A. The resulting performance improvement is plotted in Figure 2.2 in terms of end-to-end packet error ratio (PER) versus average SNR (transmit-side SNR) in dB, where average SNR is defined by SNR $= \frac{\mathrm{E}_l[|a_{i,l}|^2]}{N_0}$ and a packet is a transmitted burst with K OFDM data symbols as defined in Figure 1.1. The following scenarios are simulated:

1. The system uses a fixed modulation scheme of 16-QAM on all subcarriers (no adaptive bit loading);

2. The system uses adaptive modulation on per-subcarrier basis ($N_\mathrm{g} = 1$);

3. The system uses adaptive modulation on pairs of neighbouring subcarriers ($N_\mathrm{g} = 2$);

4. The system uses adaptive modulation on triples of neighbouring subcarriers ($N_\mathrm{g} = 3$);

5. The system uses adaptive modulation on four neighbouring subcarriers ($N_\mathrm{g} = 4$).

Considering a typical PER requirement of 10^{-1}, a significant gain in transmit power can be achieved by adaptive bit loading. A gain of around 7.5 dB can be achieved for per-subcarrier adaptive bit loading. With 1.5 dB degradation follows the subcarrier-pair-based adaptive bit loading. Even a subband adaptive bit loading with $N_\mathrm{g} = 4$ still outperforms the fixed modulation scheme by a gain of around 3 dB.

The results above are based on the assumption that the receiver is perfectly informed on the bit loading vector $\mathbf{b}$ calculated by the transmitter according to the bit loading algorithm above. In real world, this information has to be signalled to the receiver via explicit signalling, which will be examined in detail in Chapter 3. Alternatively, the receiver can blindly detect the applied $\mathbf{b}$ based on the received signal and certain boundary conditions provided by the system operation environment, which will be treated in depth in Chapter 4. This non-ideal knowledge of $\mathbf{b}$ will obviously degrade the observed gains. In the following parts of this thesis, we will try to show how the adaptive OFDM system performs under realistic conditions. Especially, we are interested in to which extend the gain can be maintained if $\mathbf{b}$ has to be signalled or blindly detected.

<div align="center">

CHAPTER 3

Explicit signalling

</div>

3.1 In-band signalling strategy

Explicit signalling is one method to communicate the adapted modulation schemes between the transmitter and the receiver. The modulation candidates are encoded in binary code words by some source coding scheme. The transmitter sends corresponding code words (signalling bits) to explicitly inform the receiver on which modulation scheme was adapted on which subcarrier.
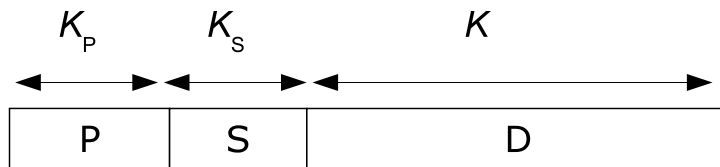
Figure 3.1: Data structure with in-band signalling

We propose an in-band signalling strategy for IEEE 802.11a based adaptive systems, where the signalling information is embedded into the existing data structure as shown in Figure 3.1. The signalling field denoted by S with K_S OFDM symbols contains the signalling symbols followed by the data field (payload) denoted by D with K OFDM symbols. The field P with K_P OFDM symbols carries preambles designated for channel estimation and synchronization issues (time and frequency synchronization). The total number of OFDM symbols in a burst (packet) is given by

$$K_F = K_P + K_S + K. \tag{3.1}$$

For a TDD based adaptive OFDM system, the in-band signalling for the adaptive DL transmission is schematically represented in Figure 3.2. The BS performs channel estimation (CE) using the preambles extracted from the latest UL transmission. Based on the channel estimate, the modulation schemes are determined on subcarrier-by-subcarrier basis for the next DL transmission by calculating a BLV $\mathbf{b} = (b^{(0)}, b^{(1)}, \cdots, b^{(N-1)})^T$, with N the number of data subcarriers and $b^{(n)}$ the number of loaded bits on the n-th subcarrier. The elements $b^{(n)}$ in $\mathbf{b}$ originate exclusively from a beforehand fixed set $\mathcal{B} = \{b_i\}_{i=1}^I$,

which represents the modulation levels (orders) of the used QAM schemes. The set $\mathcal{B}$ is at the same time the symbol alphabet and $\mathbf{b}$ is the signalling source, which has to be signalled to the MS via signalling bits embedded in the DL transmission. For this purpose, the signalling source $\mathbf{b}$ has to be source coded. Assuming that by a certain source coding scheme, the modulation level b_i is encoded into the code word c_i , the vector

$$\mathbf{b} = (b^{(0)}, b^{(1)}, \cdots, b^{(N-1)})^{\mathrm{T}} \tag{3.2}$$

is then signalled by the bit sequence vector

$$\mathbf{c} = (c^{(0)}, c^{(1)}, \cdots, c^{(N-1)})^{\mathrm{T}} \quad \text{with} \quad c^{(n)} \in \mathcal{C} = \{c_i\}_{i=1}^{I}, \tag{3.3}$$

where $\mathcal{C}$ is the code book of the source coding scheme.
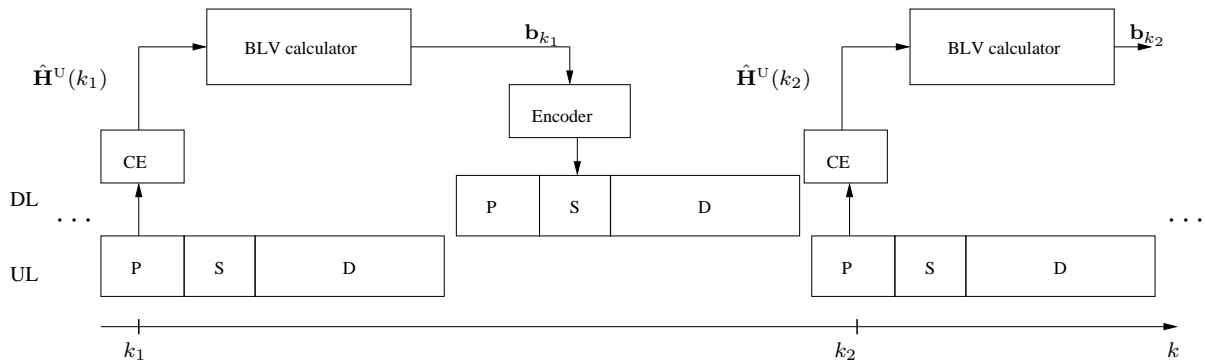
Figure 3.2:   Schematical representation of in-band signalling strategy

The signalling source $\mathbf{b}$ is characterized on one side by its elements $b^{(n)}$, on the other side its information-theoretical features like symbol probabilities and possible memory effects. Due to the bit loading algorithm described in Section 2.2.1, the symbols b_i in the symbol alphabet $\mathcal{B}$ are not uniformly distributed (the occurrence probabilites are not equal). Furthermore, memory effects exist inside a symbol sequence $\mathbf{b}_{k_1}$ as well as over consecutive symbol sequences $\mathbf{b}_{k_1}$ and $\mathbf{b}_{k_2}$, with $k_2 - k_1 = K_{\mathrm{F}}$ (see Figure 3.2). In the following, we distinguish between the following four source models:

1. Ignore all features and assume that the symbols are uniformly distributed. This model has the advantage that no prior knowledge about the source is necessary. We denote this kind of model as $\mathcal{B}_{\mathrm{u}, N_{\mathrm{g}}}$, where N_{g} is the number of symbols grouped into a joint symbol.

2. Ignore all memory effects and consider only the probabilities of the symbols (or joint symbols). This model is denoted as $\mathcal{B}_{\mathrm{H}, N_{\mathrm{g}}}$.

3. Ignore the memory effects over consecutive symbol sequences $\mathbf{b}_{k_1}$ and $\mathbf{b}_{k_2}$ and consider the probabilities and the memory effects inside a symbol sequence $\mathbf{b}_{k_1}$. This model is denoted as $\mathcal{B}_{\mathrm{FD},N_{\mathrm{g}}}$.

4. Ignore the memory effects inside a symbol sequence $\mathbf{b}_{k_1}$ and consider the probabilities and the memory effects over consecutive symbol sequences $\mathbf{b}_{k_1}$ and $\mathbf{b}_{k_2}$. This model is denoted as $\mathcal{B}_{\mathrm{TD},N_{\mathrm{g}}}$.

3.2 Per-burst coding scheme

In this class of coding scheme, the elements in $\mathbf{b}_{k_1}$ are assumed to be statistically independent on the elements in $\mathbf{b}_{k_2}$. So, no knowledge about the previous source sequences is required. The source models 1 to 3 can be used for per-burst coding scheme.

3.2.1 Trivial coding schemes

Without knowing the occurrence probabilities P_{b_i} of the symbols b_i, it is reasonable to assume uniformly distributed symbols with probabilities $P_{b_i} = \frac{1}{I}$. This kind of source model is denoted as $\mathcal{B}_{\mathrm{U},1}$. From information theory, the symbols b_i can be encoded into code words of the same length

$$W_1 = \lceil \log_2(I) \rceil, \tag{3.4}$$

where $\lceil \cdot \rceil$ is the round-up function. The bit loading vector $\mathbf{b}$ is comprised of N elements. The signalling overhead per transmission packet measured in bits is given by

$$L_{\mathrm{U},1} = N \lceil \log_2(I) \rceil. \tag{3.5}$$

Signalling can be organized more efficiently by jointly encoding multiple elements in $\mathbf{b}$. By grouping neighbouring two elements $b^{(n)}$ and $b^{(n+1)}$ into a symbol pair $(b^{(n)}, b^{(n+1)})$, the source model is now denoted by $\mathcal{B}_{\mathrm{U},2}$. The required signalling overhead is given by

$$L_{\mathrm{U},2} = \frac{N}{2} \lceil 2\log_2(I) \rceil. \tag{3.6}$$

In general, by grouping N_{g} neighbouring subcarriers with N_{g} as a divider of N, a joint symbol $(b^{(n)}, b^{(n+1)}, \cdots, b^{(n+N_{\mathrm{g}}-1)})$ is encoded by a bit sequence of the length

$$W_{N_{\mathrm{g}}} = \lceil N_{\mathrm{g}}\log_2(I) \rceil. \tag{3.7}$$

The corresponding source model is denoted as $\mathcal{B}_{\mathrm{U},N_{\mathrm{g}}}$. The corresponding signalling overhead is then given by

$$L_{\mathrm{U},N_{\mathrm{g}}} = \frac{N}{N_{\mathrm{g}}} \lceil N_{\mathrm{g}}\log_2(I) \rceil. \tag{3.8}$$

The entropy of the source $\mathcal{B}_{\mathrm{U},N_{\mathrm{g}}}$ is given by

$$E_{\mathcal{B}_{\mathrm{U}},N_{\mathrm{g}}} = -\frac{\sum_{i_1}^{I}\sum_{i_2}^{I}\cdots\sum_{i_{N_{\mathrm{g}}}}^{I} P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_{\mathrm{g}}}}} \log_2(P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_{\mathrm{g}}}}})}{N_{\mathrm{g}}} = \log_2(I), \qquad (3.9)$$

which is normalized to and independent on N_{g}, where $P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_{\mathrm{g}}}}}$ is the probability of the joint symbol $(b_{i_1}, b_{i_2}, \cdots, b_{i_{N_{\mathrm{g}}}})$. Then, it holds [70]

$$E_{\mathcal{B}_{\mathrm{U}},N_{\mathrm{g}}} \leq \frac{L_{\mathrm{U},N_{\mathrm{g}}}}{N} \leq E_{\mathcal{B}_{\mathrm{U}},N_{\mathrm{g}}} + \frac{1}{N_{\mathrm{g}}}. \qquad (3.10)$$

Signalling overhead may be reduced by increasing N_{g} at the price of rapidly growing coding complexity, which is exponential in N_{g}.

If we denote the set $\mathcal{N}_{\mathrm{g}}$, whose elements N_{g} are dividers of N

$$N \mod N_{\mathrm{g}} = 0, \qquad (3.11)$$

from the viewpoint of saving signalling overhead, the optimal group size $N_{\mathrm{g,op}}$ is determined by minimizing the redundancy of a code

$$N_{\mathrm{g,op}} = \underset{N_{\mathrm{g}}}{\operatorname{argmin}} \left(\frac{\lceil N_{\mathrm{g}}\log_2(I)\rceil}{N_{\mathrm{g}}} - E_{\mathcal{B}_{\mathrm{U}},N_{\mathrm{g}}} \right), \quad \text{with } N_{\mathrm{g}} \in \mathcal{N}_{\mathrm{g}}. \qquad (3.12)$$

Here we consider an OFDM system with $N = 48$ subcarriers and a modulation set with $I = 5$ candidates as specified in Appendix A. The possible N_{g} numbers, the average code word length $\bar{W}_{N_{\mathrm{g}}} = \frac{W_{N_{\mathrm{g}}}}{N_{\mathrm{g}}}$ and the corresponding signalling overheads $L_{\mathrm{U},N_{\mathrm{g}}}$ are given in Table 3.1. In this case, the number $N_{\mathrm{g,op}} = 3$ gives already a minimal overhead measured in bits

$$L_{\mathrm{U},N_{\mathrm{g,op}}} = \frac{48}{3}[3\log_2(5)] = \frac{48}{3}7 = 112. \qquad (3.13)$$

Note, due to the round-up operation, there may exist several group sizes $N_{\mathrm{g,op}}$ with the

Table 3.1: Normalized average code word length $\bar{W}_{N_{\mathrm{g}}} = \frac{W_{N_{\mathrm{g}}}}{N_{\mathrm{g}}}$ and signalling overhead $L_{\mathrm{U},N_{\mathrm{g}}}$ for different group size N_{g}.

N_{g}	1	2	**3**	4	6	8	12	16	24	48
$\bar{W}_{N_{\mathrm{g}}}$	3	2	**2.33**	2.5	2.33	2.38	2.33	2.38	2.33	2.33
$L_{\mathrm{U},N_{\mathrm{g}}}$	144	120	**112**	120	112	114	112	114	112	112
$E_{\mathcal{B}}$					2.3219					

same minimal signalling overhead. In this case, it is advised to select the smallest possible number $N_{\mathrm{g,min}}$ giving the lowst implementation complexity. Here $N_{\mathrm{g,min}} = 3$ satisfies this condition.

3.2.2   Huffman coding schemes

Knowing the occurrence probabilities P_{b_i} of the symbols b_i, the source is modelled by $\mathcal{B}_{\mathrm{H},1}$ and a Huffman coding scheme can be applied which minimize the average code word length [71]. The values P_{b_i} can be obtained beforehand by computer simulations.

As mentioned in Section 2.2.1 , the second constraint imposed on the AM requires that a constant B bits have to be loaded. Also due to frequency selectivity of the transmission channel, the probabilities P_{b_i} are not equal any more. Table 3.2 lists P_{b_i} as well as the code words c_i by Huffman coding for system parameters specified in the Appendix A (this system holds for all subsequent simulations if nothing else is stated). Because of $\frac{B}{N} = \frac{192 \text{ bit}}{48 \text{ subcarrier}} = 4$ bits/subcarrier, 16-QAM is the most frequently used modulation level, which is encoded by 1 bit. NoTx and BPSK are least frequently used which both require 4 bits.

Table 3.2:   Huffman coding for $B = 192$ bits and $N = 48$ data subcarrers.

b_i	0	1	2	4	6
P_{b_i}	0.027	0.026	0.186	0.481	0.280
c_i	0110	0111	010	1	00

Denote the code word length for i-th modulation level b_i as $W_{\mathrm{H},i}$, the signalling overhead is then given by

$$L_{\mathrm{H},1} = \sum_{i=1}^{I} N_{\mathrm{Sc},i} W_{\mathrm{H},i} \quad \text{with} \quad \sum_{i=1}^{I} N_{\mathrm{Sc},i} = N, \tag{3.14}$$

where $N_{\mathrm{Sc},i}$ is the number of subcarriers using b_i, i.e., elements in $\mathbf{b} = (b^{(0)}, b^{(1)}, \cdots, b^{(N-1)})^{\mathrm{T}}$ with $b^{(n)} = b_i$ and $b^{(n)} \in \mathcal{B}$. The vector $\mathbf{b}$ is determined by B and the time-variant $\hat{\mathbf{H}}^{\mathrm{U}}$. Consequently $N_{\mathrm{Sc},i}$ varies also with time leading to a time-variant signalling overhead $L_{\mathrm{H},1}(k)$ with k as index in the discrete-time domain.

The average code word length is given by

$$\bar{W}_{\mathrm{H},1} = \sum_{i=1}^{I} P_{b_i} W_{\mathrm{H},i}. \tag{3.15}$$

The average signalling overhead is

$$\bar{L}_{\mathrm{H},1} = N \bar{W}_{\mathrm{H},1}. \tag{3.16}$$

For the above considered example, numerical evaluations give: $\bar{W}_{\mathrm{H},1} = 1.811 \frac{\text{bit}}{\text{subcarrier}}$ and $\bar{L}_{\mathrm{H},1} = 48 \times 1.811 = 89.93$ bits. Compared to the overhead $L_{\mathrm{U},1}$ in Table 3.1, more than 50 bits are reduced on average.

Huffman coding scheme can also be combined with joint coding schemes for grouped subcarriers. Consider a joint symbol $(b_{i_1}, b_{i_2}, \cdots, b_{i_{N_g}})$ containing the modulation levels for N_g subcarriers. This source $\mathcal{B}_{H,N_g}$ has the symbol alphabet $\{(b_{i_1}, b_{i_2}, \cdots, b_{i_{N_g}})\}_{i_1,i_2,\cdots,i_{N_g}}^{I,I,\cdots,I}$ with the corresponding probabilities $P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_g}}}$. The entropy is given by

$$E_{\mathcal{B}_{H,N_g}} = \frac{-\sum_{i_1=1}^{I}\sum_{i_2=1}^{I}\cdots\sum_{i_{N_g}=1}^{I} P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_g}}} \log_2(P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_g}}})}{N_g}. \qquad (3.17)$$

We denote the code word length resulting from Huffman coding as $W_{H,i_1,i_2,\cdots,i_{N_g}}$, the average signalling overhead is then given by

$$\bar{L}_{H,N_g} = \frac{N}{N_g} \underbrace{\sum_{i_1=1}^{I}\sum_{i_2=1}^{I}\cdots\sum_{i_{N_g}=1}^{I} P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_g}}} W_{H,i_1,i_2,\cdots,i_{N_g}}}_{\bar{W}_{H,N_g}}, \qquad (3.18)$$

where $\bar{W}_{H,N_g}$ is the average code word length.

The Huffman coding scheme always minimizes the redundancy of a code

$$R_{H,N_g} = \min_{\mathcal{C}} \left(\bar{W}_{\mathcal{C},N_g} - E_{\mathcal{B}_{H,N_g}} \right), \qquad (3.19)$$

with $\bar{W}_{\mathcal{C},N_g}$ the average code word length of the code $\mathcal{C}$, in the sense that the average code word length $\bar{W}_{H,N_g}$ approaches the entropy $E_{\mathcal{B}_{H,N_g}}$. Consequently, $\bar{W}_{H,N_g}$ can be reduced by using source models which exploit special information-theoretical features as introduced in Section 3.1. The knowledge of the probabilities $P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_g}}}$ is one feature among many. Since the assumption of uniform distributed symbols results in maximal entropy $E_{\mathcal{B}_U}$, any deviation from this distribution will decrease $E_{\mathcal{B}_{H,N_g}}$.

For $N_g = 1$, the probabilities P_{b_i} and the code word lengths depend on B only, as P_{b_i} are determined by averaging long term observations, although the time-variant signalling overhead $L_{H,1}(k)$ depends on the temporary channel estimates $\hat{\mathbf{H}}^U$. For $N_g > 1$, however, by grouping neighbouring subcarriers, the probabilities $P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_g}}}$ depend on B and the frequency-domain correlation of $\hat{\mathbf{H}}^U$, which in turn depends on the precision of channel estimation. So the system SNR impacts $P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_g}}}$ and the resulting code word length $W_{H,i_1,i_2,\cdots,i_{N_g}}$. Consequently, the signalling overhead depends also on the system SNR.

For $N_g = 2$, $P_{b_{i_1},b_{i_2}}$ and the corresponding code words are provided in Appendix B for some representative SNR values and the simulation environment given in Appendix A. We also investigated the joint Huffman coding scheme for $N_g = 3$. Due to large symbol size ($5^3 = 125$ symbols), the individual probabilities $P_{b_{i_1},b_{i_2},b_{i_3}}$ and the code word lengths W_{H,i_1,i_2,i_3} are not listed here. But the average code word length per symbol $\frac{\bar{W}_{H,3}}{3}$ and the average signalling overheads $\bar{L}_{H,3}$ are given in Table 3.3.

Table 3.3: Average signalling overheads $\bar{L}_{\mathrm{H},N_{\mathrm{g}}}$, average code word lengths per symbol $\frac{\bar{W}_{\mathrm{H},N_{\mathrm{g}}}}{N_{\mathrm{g}}}$ for different system SNRs and N_{g}.

	Perfect CE			20 dB			0 dB		
N_{g}	1	2	3	1	2	3	1	2	3
$\frac{W_{\mathrm{H},N_{\mathrm{g}}}}{N_{\mathrm{g}}}$ in bits	1.81	1.48	1.35	1.81	1.49	1.36	1.81	1.56	1.45
$\bar{L}_{\mathrm{H},N_{\mathrm{g}}}$ in bits	86.88	71.04	64.80	86.88	71.52	65.28	86.88	74.88	69.60

Table 3.4: Entropy $E_{\mathcal{B}_{\mathrm{H},N_{\mathrm{g}}}}$, average code word lengths per symbol $\frac{\bar{W}_{N_{\mathrm{g}}}}{N_{\mathrm{g}}}$ and average signalling overheads $\bar{L}_{N_{\mathrm{g}}}$ for memoryless encoding schemes.

	Trivial encoding schemes			Huffman coding at 20 dB		
N_{g}	1	2	3	1	2	3
$E_{\mathcal{B}_{\mathrm{H},N_{\mathrm{g}}}}$	2.3219			1.7510	1.4593	1.3425
$\frac{\bar{W}_{N_{\mathrm{g}}}}{N_{\mathrm{g}}}$ in bits	3	2.5	2.33	1.81	1.49	1.36
$\bar{L}_{N_{\mathrm{g}}}$ in bits	144	120	112	86.88	71.52	65.28

Investigations for group size $N_{\mathrm{g}} > 3$ were not performed due to two reasons. First, no significant reduction of signalling overhead is expected since the degree of correlation reduces fast with increasing N_{g}. Second, the encoding and decoding complexity grows at least exponentially with increasing N_{g}. An overview on the signalling overhead is provided in Table 3.4 for the memoryless encoding schemes investigated till now.

3.2.3 Frequency-domain correlation

In OFDM systems, typically, a broadband frequency-selective channel is split into a large number of subcarriers. Each subcarrier is frequency-flat and can be separately modulated by a certain modulation format. The subcarrier spacing B_{Sub} is much smaller than the coherence bandwidth B_{Co}. The channel amplitudes, to which the modulation formats are adapted, are strongly correlated under neighbouring subcarriers.

This correlation has been already partially utilized in joint Huffman coding schemes, where neighbouring N_{g} subcarriers are grouped into a joint symbol $(b_{i_1}, b_{i_2}, \cdots, b_{i_{N_{\mathrm{g}}}})$ and signalled jointly. If the bandwidth of two neighbouring subcarrier groups $2N_{\mathrm{g}}B_{\mathrm{Sub}}$ is still smaller than B_{Co}, correlation exists also under two neighbouring joint symbols leading to memory effects in the signalling source. These memory effects bear certain amount of redundancy, which can be reduced by frequency-domain state-dependent Huffman coding (FSDH) .

In general, a discrete source with memory can be described by a Markov chain which can
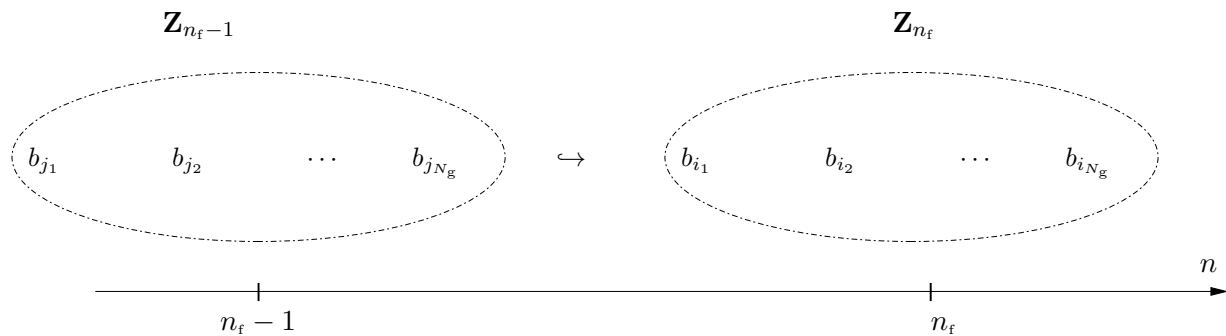
Figure 3.3: Schematical representation of symbol transition in the frequency domain

be encoded by state-dependent Huffman coding schemes based on transition probabilities. Transitions occur in frequency direction, which are schematically sketched in Figure 3.3. For a transmission packet, $\frac{N}{N_\mathrm{g}}$ joint symbols have to be encoded. We introduce the notation $Z_{n_\mathrm{f}-1}$ and Z_{n_f} to denote $(n_\mathrm{f}-1)$-th symbol $(b_{j_1}, b_{j_2}, \cdots, b_{j_{N_\mathrm{g}}})$ and n_f-th symbol $(b_{i_1}, b_{i_2}, \cdots, b_{i_{N_\mathrm{g}}})$ counted in the frequency domain respectively, independent of the size of the symbol N_g. Here, n_f is the symbol index. The first joint symbol Z_0 is called the initial symbol and encoded by Huffman coding introduced in Section 3.2.2. The remaining symbols Z_{n_f} with $n_\mathrm{f} = 1, 2, \cdots, \frac{N}{N_\mathrm{g}} - 1$, are encoded by state-dependent Huffman coding based on transition probabilities $P_{Z_{n_\mathrm{f}}|Z_{n_\mathrm{f}-1}}$. This type of source, which contains the inital symbol and the state-dependent symbols, is denoted as $\mathcal{B}_{\mathrm{F},N_\mathrm{g}}$.

Table 3.5: $P_{Z_{n_\mathrm{f}}|Z_{n_\mathrm{f}-1}}$, $W_{\mathrm{F},1}$ given in bold and $\bar{W}_{\mathrm{F},1}$ in bits per symbol for $N_\mathrm{g} = 1$ and SNR = 20 dB.

$Z_{n_\mathrm{f}-1}$ \ Z_{n_f}	0	1	2	4	6
0	0.2527 **2**	0.1679 **3**	0.4974 **1**	0.0800 **4**	0.0017 **4**
1	0.1919 **2**	0.1883 **3**	0.4938 **1**	0.1232 **4**	0.0026 **4**
2	0.0741 **3**	0.0707 **4**	0.5117 **1**	0.3388 **2**	0.0045 **4**
4	0.0043 **4**	0.0061 **4**	0.1300 **2**	0.7352 **1**	0.1242 **3**
6	6e-5 **4**	0.0002 **4**	0.0035 **3**	0.2114 **2**	0.7846 **1**
$\bar{W}_{\mathrm{F},1} = 1.4384$					

First we focus on the state-dependent symbols Z_{n_f} with $n_\mathrm{f} \geq 1$. We use a Markov chain to describe this source with the state-dependent transitions only, which is denoted as

$\mathcal{B}_{\text{FD},N_{\text{g}}}$. The entropy of $\mathcal{B}_{\text{FD},N_{\text{g}}}$ is given by

$$E_{\mathcal{B}_{\text{FD},N_{\text{g}}}} = -\frac{1}{N_{\text{g}}} \sum_{Z_{n_{\text{f}}-1}} \sum_{Z_{n_{\text{f}}}} P_{Z_{n_{\text{f}}-1}} P_{Z_{n_{\text{f}}}|Z_{n_{\text{f}}-1}} \log_2 \left(P_{Z_{n_{\text{f}}}|Z_{n_{\text{f}}-1}} \right). \tag{3.20}$$

By removing part of redundancy existing due to the frequency-domain memory effect [72], it holds

$$E_{\mathcal{B}_{\text{FD},N_{\text{g}}}} < E_{\mathcal{B}_{\text{H},N_{\text{g}}}}. \tag{3.21}$$

The entropy of the source model $\mathcal{B}_{\text{F},N_{\text{g}}}$ is compromised of two contributions and is given by

$$E_{\mathcal{B}_{\text{F},N_{\text{g}}}} = \frac{1}{N} \left((N - N_{\text{g}}) E_{\mathcal{B}_{\text{FD},N_{\text{g}}}} + N_{\text{g}} E_{\mathcal{B}_{\text{H},N_{\text{g}}}} \right) < E_{\mathcal{B}_{\text{H},N_{\text{g}}}}, \tag{3.22}$$

showing the potential of reducing average code word length by state-dependent Huffman coding. Due to the initial symbol Z_0, the relationship between these entropies are given by

$$E_{\mathcal{B}_{\text{FD},N_{\text{g}}}} < E_{\mathcal{B}_{\text{F},N_{\text{g}}}} < E_{\mathcal{B}_{\text{H},N_{\text{g}}}}. \tag{3.23}$$

Similar to the coding approaches given in Section 3.3.1, we consider $N_{\text{g}} = 1, 2, 3$, and different system SNR values. For $N_{\text{g}} = 1$, the transition probabilities $P_{Z_{n_{\text{f}}}|Z_{n_{\text{f}}-1}}$ and the corresponding code word lengths $W_{\text{F},1}$ are given in Table 3.5 for system SNR = 20 dB. Corresponding tables are given in Appendix C for perfect channel estimation and SNR = 0 dB.

The encoding of the initial symbol Z_0 has been already performed in Section 3.2.2. The resulting average code word length $\bar{W}_{\text{H},N_{\text{g}}}$ is defined in (3.18). In general, by considering correlation, it holds

$$\bar{W}_{\text{F},N_{\text{g}}} < \bar{W}_{\text{H},N_{\text{g}}}, \tag{3.24}$$

which can be derived from (3.23) as confirmed for $N_{\text{g}} = 1$ (compared to $\bar{W}_{\text{H},1}$ given in Table 3.3). The total average code word length including the initial symbol Z_0 is given by

$$\bar{W}_{\text{F},N_{\text{g}}} = \frac{N_{\text{g}}}{N} \left((N - N_{\text{g}}) \bar{W}_{\text{FD},N_{\text{g}}} + N_{\text{g}} \bar{W}_{\text{H},N_{\text{g}}} \right), \tag{3.25}$$

where $\bar{W}_{\text{FD},N_{\text{g}}}$ is the average code word length of $Z_{n_{\text{f}}}$ with $n_{\text{f}} \geq 1$. The resulting average signalling overhead is given by

$$\bar{L}_{\text{F},N_{\text{g}}} = (N - N_{\text{g}}) \bar{W}_{\text{FD},N_{\text{g}}} + N_{\text{g}} \bar{W}_{\text{H},N_{\text{g}}}. \tag{3.26}$$

Table 3.6 lists $\bar{L}_{\text{F},N_{\text{g}}}$ and $\frac{\bar{W}_{\text{F},N_{\text{g}}}}{N_{\text{g}}}$ for different N_{g} and system SNR.

Table 3.6: Average signalling overheads $\bar{L}_{\mathrm{F},N_{\mathrm{g}}}$, average code word lengths per subcarrier $\frac{\bar{W}_{\mathrm{F},N_{\mathrm{g}}}}{N_{\mathrm{g}}}$ for different system SNRs and N_{g} exploiting frequency-domain correlation.

	Perfect CE			20 dB			0 dB		
N_{g}	1	2	3	1	2	3	1	2	3
$\frac{W_{\mathrm{F},N_{\mathrm{g}}}}{N_{\mathrm{g}}}$ in bits	1.44	1.16	1.12	1.45	1.16	1.12	1.53	1.28	1.23
$\bar{L}_{\mathrm{F},N_{\mathrm{g}}}$ in bits	69.27	55.63	53.63	69.41	55.90	53.87	73.56	61.41	59.20

3.3   Over-burst coding scheme

The purpose of signalling is to inform the receiver on the current bit loading vector $\mathbf{b}$ (signalling source). In a packet-based OFDM transmission system with certain packet duration K_{F} (number of OFDM symbols), a common practice is to perform adaptive modulation on a packet-basis, i.e. $\mathbf{b}$ is updated in the period of K_{F}. For each packet, $\mathbf{b}$ has to be signalled only once. The next signalling is necessary if $\mathbf{b}$ is updated. We focus on typical indoor applications, the radio channel is assumed to be slowly time-variant leading to a strong correlation between successive BLVs $\mathbf{b}_{k_{\mathrm{s}}-1}$ and $\mathbf{b}_{k_{\mathrm{s}}}$ with k_{s} as packet index shown in Figure 3.4, provided if K_{F} is small. In fact, we expect a strong temporary memory effect in the signalling source for packet sizes typically used in practical applications. This memory effect bears also certain amount of redundancy.
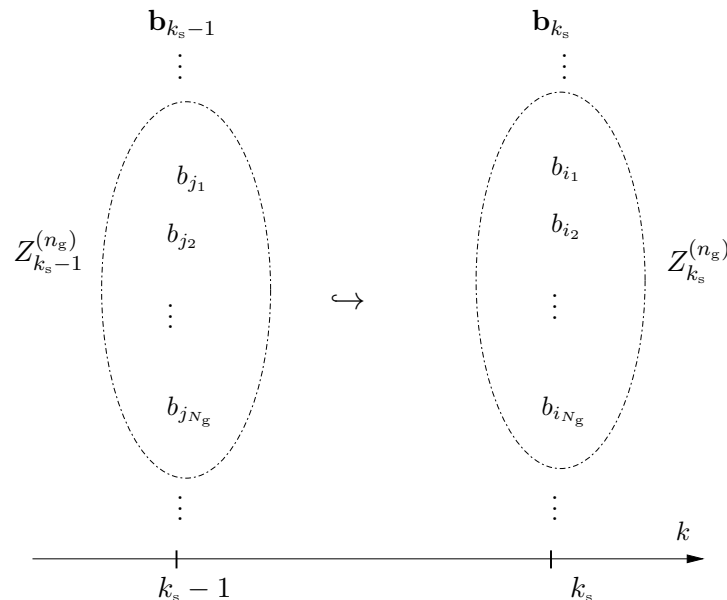
Figure 3.4:   Schematical representation of a symbol transition

3.3.1   Time-domain correlation

The redundancy can be reduced, if this memory effect is also taken into account. Beginning at the first packet $k_\mathrm{s} = 0$, we denote $\mathbf{b}_{k_\mathrm{s}}$ as the BLV for the k_s-th packet. Now, instead of $\mathbf{b}_{k_\mathrm{s}}$ itself, the transition $\overleftrightarrow{\mathbf{b}}_{k_\mathrm{s}}$ between $\mathbf{b}_{k_\mathrm{s}}$ and $\mathbf{b}_{k_\mathrm{s}-1}$, which is created by a transition indentifier as shown in Figure 3.5, is encoded and signalled. If we group N_g neighbouring subcarriers into a joint symbol $(b_{i_1}, b_{i_2}, \cdots, b_{i_{N_\mathrm{g}}})$ just as in the frequency-domain correlation, the transition of one symbol $(b_{j_1}, b_{j_2}, \cdots, b_{j_{N_\mathrm{g}}})$ of the packet $k_\mathrm{s} - 1$ to another symbol $(b_{i_1}, b_{i_2}, \cdots, b_{i_{N_\mathrm{g}}})$ of the next packet k_s is schematically represented in Figure 3.4. Here we also introduce $Z_{k_\mathrm{s}}^{(n_\mathrm{g})}$ to denote the n_g-th joint symbol counted in the frequency domain in the k_s-th packet, independent of N_g. Since $\overleftrightarrow{\mathbf{b}}_{k_\mathrm{s}}$ is completely described by $\mathbf{b}_{k_\mathrm{s}}$ and $\mathbf{b}_{k_\mathrm{s}-1}$, this source can be modelled by a Markov chain, too.
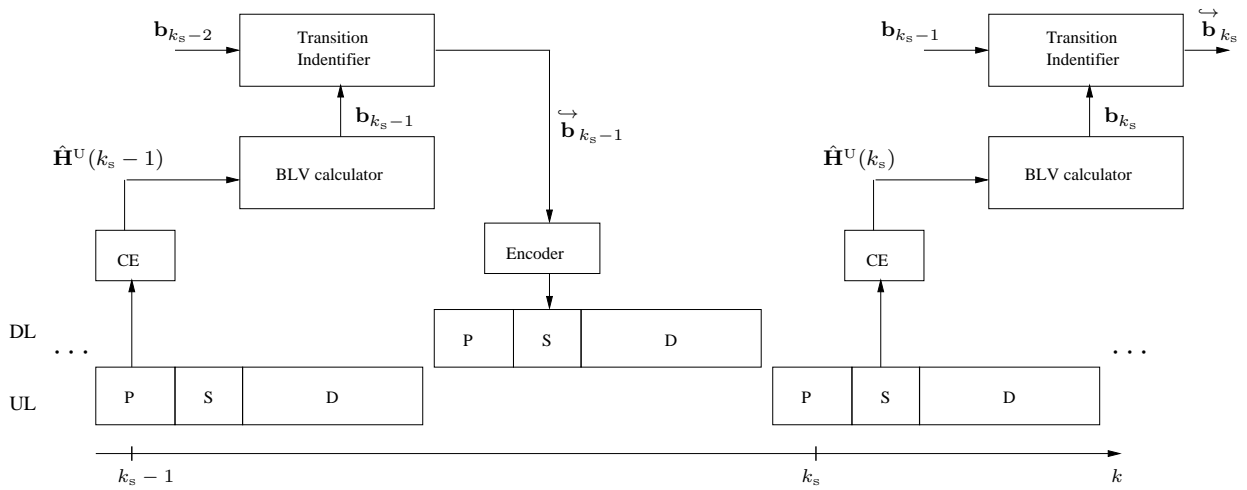
Figure 3.5:   Schematic representation of state-dependent in-band signalling

The initial bit loading vector $\mathbf{b}_0$ is encoded based on Huffman coding schemes introduced before. The entropy of this source $\mathcal{B}_{\mathrm{H},N_\mathrm{g}}$ is given in (3.17). For the subsequent bit loading vectors $\mathbf{b}_{k_\mathrm{s}}$ with $k_\mathrm{s} \geq 1$, only the transitions are encoded based on transition probabilities $P_{Z_{k_\mathrm{s}}^{(n_\mathrm{g})} | Z_{k_\mathrm{s}-1}^{(n_\mathrm{g})}}$, which have to be determined beforehand through computer simulations. We denote this source model with time-domain memory as $\mathcal{B}_{\mathrm{TD},N_\mathrm{g}}$, and assume that it is stationary, so that $P_{Z_{k_\mathrm{s}}^{(n_\mathrm{g})} | Z_{k_\mathrm{s}-1}^{(n_\mathrm{g})}}$ is independent on k_s and the symbol probability $P_{Z_{k_\mathrm{s}-1}^{(n_\mathrm{g})}}$. The entropy of $\mathcal{B}_{\mathrm{TD},N_\mathrm{g}}$ is given by

$$E_{\mathcal{B}_{\mathrm{TD},N_\mathrm{g}}} = -\frac{1}{N_\mathrm{g}} \sum_{Z_{k_\mathrm{s}-1}^{(n_\mathrm{g})}} \sum_{Z_{k_\mathrm{s}}^{(n_\mathrm{g})}} P_{Z_{k_\mathrm{s}-1}^{(n_\mathrm{g})}} P_{Z_{k_\mathrm{s}}^{(n_\mathrm{g})} | Z_{k_\mathrm{s}-1}^{(n_\mathrm{g})}} \log_2 \left(P_{Z_{k_\mathrm{s}}^{(n_\mathrm{g})} | Z_{k_\mathrm{s}-1}^{(n_\mathrm{g})}} \right) . \qquad (3.27)$$

This source model removes the part of redundancy existing due to the time-domain

memory, so it holds [72]

$$E_{\mathcal{B}_{\mathrm{TD},N_{\mathrm{g}}}} < E_{\mathcal{B}_{\mathrm{H},N_{\mathrm{g}}}}. \tag{3.28}$$

Consequently the average code word length can be also reduced by state-dependent Huffman coding as justified in Section 3.2.1.

The time-domain memory in the bit loading vector $\mathbf{b}_{k_{\mathrm{s}}}$ is provided by the time-domain correlation in the channel estimates. As mentioned, the precision of channel estimation depends on the system SNR. Consequently, the code word lengths depends also on the system SNR just as described in the frequency-domain joint coding schemes in Section 3.2.2.

Table 3.7: $P_{Z_{k_{\mathrm{s}}-1}^{(n_{\mathrm{g}})}|Z_{k_{\mathrm{s}}}^{(n_{\mathrm{g}})}}$, $W_{\mathrm{SDH},1}$ given in bold and $\bar{W}_{\mathrm{SDH},1}$ in bits per symbol for $N_{\mathrm{g}} = 1$ and SNR = 20 dB.

$Z_{k_{\mathrm{s}}-1}^{(n_{\mathrm{g}})}$ \ $Z_{k_{\mathrm{s}}}^{(n_{\mathrm{g}})}$	0	1	2	4	6
0	0.8120 **1**	0.1585 **2**	0.0293 **2**	0 **0**	0 **0**
1	0.1729 **2**	0.5587 **1**	0.2682 **2**	0 **0**	0 **0**
2	0.0042 **3**	0.0380 **3**	0.8885 **1**	0.0692 **2**	0 **0**
4	0 **0**	0 **0**	0.0265 **2**	0.9464 **1**	0.0269 **2**
6	0 **0**	0 **0**	0 **0**	0.0464 **1**	0.9535 **1**
$\bar{W}_{\mathrm{SDH},1} = 1.0708$					

For $N_{\mathrm{g}} = 1$, the transition probabilities $P_{Z_{k_{\mathrm{s}}}^{(n_{\mathrm{g}})}|Z_{k_{\mathrm{s}}-1}^{(n_{\mathrm{g}})}}$ and the corresponding code word lengths $W_{\mathrm{SDH},N_{\mathrm{g}}}$ resulting from the state-dependent Huffman (SDH) coding are given in Table 3.7 for system SNR = 20 dB. It is a common approach to represent $P_{Z_{k_{\mathrm{s}}}^{(n_{\mathrm{g}})}|Z_{k_{\mathrm{s}}-1}^{(n_{\mathrm{g}})}}$ in a transition matrix, where each entry is a transition probability for the transition from the symbol $Z_{k_{\mathrm{s}}-1}^{(n_{\mathrm{g}})}$ denoting the corresponding row to the symbol $Z_{k_{\mathrm{s}}}^{(n_{\mathrm{g}})}$ denoting the corresponding column. The average code word length is given by

$$\bar{W}_{\mathrm{SDH},N_{\mathrm{g}}} = \sum_{Z_{k_{\mathrm{s}}-1}^{(n_{\mathrm{g}})}} \sum_{Z_{k_{\mathrm{s}}}^{(n_{\mathrm{g}})}} P_{Z_{k_{\mathrm{s}}-1}^{(n_{\mathrm{g}})}} P_{Z_{k_{\mathrm{s}}}^{(n_{\mathrm{g}})}|Z_{k_{\mathrm{s}}-1}^{(n_{\mathrm{g}})}} W_{\mathrm{SDH},N_{\mathrm{g}}}, \tag{3.29}$$

describing the average number of bits required per symbol for the case that $P_{Z_{k_{\mathrm{s}}}^{(n_{\mathrm{g}})}|Z_{k_{\mathrm{s}}-1}^{(n_{\mathrm{g}})}}$ reaches the steady state (independent on k_{s}). For perfect channel estimation and system SNR = 0 dB, the corresponding coding is given in Appendix C.1.

For $N_\mathrm{g} = 2$, neighbouring 2 symbols are grouped into a joint symbol. The transition matrix possesses a dimension of 25×25. Due to large number of elements, we do not provide details about $P_{Z_{k_\mathrm{s}}^{(n_\mathrm{g})}|Z_{k_\mathrm{s}-1}^{(n_\mathrm{g})}}$ and $W_{\mathrm{SDH},N_\mathrm{g}}$ for individual transitions. But $\bar{W}_{\mathrm{SDH},N_\mathrm{g}}$ and the steady-state average signalling overheads $\bar{L}_{\mathrm{SDH},N_\mathrm{g}}$ are given in Table 3.8, as well as for $N_\mathrm{g} = 3$. Investigations for $N_\mathrm{g} > 3$ were not performed due to the reasons given in Section 3.2.2.

Table 3.8: Average signalling overheads $\bar{L}_{\mathrm{SDH},N_\mathrm{g}}$, average code word lengths per subcarrier $\frac{\bar{W}_{\mathrm{SDH},N_\mathrm{g}}}{N_\mathrm{g}}$ for different system SNR and N_g by state-dependent Huffman coding.

	Perfect CE			20 dB			0 dB		
N_g	1	2	3	1	2	3	1	2	3
$\frac{W_{\mathrm{SDH},N_\mathrm{g}}}{N_\mathrm{g}}$ in bits	1.04	0.57	0.42	1.07	0.64	0.50	1.58	1.30	1.21
$\bar{L}_{\mathrm{SDH},N_\mathrm{g}}$ in bits	49.92	27.56	20.14	51.36	30.89	24.13	75.84	62.32	58.11

Table 3.9: $\bar{L}_{N_\mathrm{g}}$, $\frac{\bar{W}_{\mathrm{X},N_\mathrm{g}}}{N_\mathrm{g}}$ and $E_{\mathcal{B}_{\mathrm{X},XN_\mathrm{g}}}$ given in bits for state-dependent Huffman coding schemes exploiting time-domain and frequency-domain memory at SNR $= 20$ dB.

	Time-domain correlation			Frequency-domain correlation		
N_g	1	2	3	1	2	3
$E_{\mathcal{B}_{\mathrm{X},N_\mathrm{g}}}$	0.4226	0.3991	0.2365	1.18	1.14	1.11
$\frac{W_{\mathrm{X},N_\mathrm{g}}}{N_\mathrm{g}}$	1.07	0.64	0.50	1.45	1.16	1.12
$\bar{L}_{\mathcal{B}_{\mathrm{X},XN_\mathrm{g}}}$	51.36	30.89	24.13	69.41	55.90	53.87

An overview on the entropies (representative notation $E_{\mathcal{B}_{\mathrm{X},N_\mathrm{g}}}$ with X standing for time-domain or frequency-domain based signalling scheme), normalized average code word lengths $\frac{\bar{W}_{\mathrm{X},N_\mathrm{g}}}{N_\mathrm{g}}$ and average signalling overheads $\bar{L}_{\mathcal{B}_{\mathrm{X},N_\mathrm{g}}}$ are provided for different N_g and system SNR in Table 3.9.

3.4   Performance comparison

In the following, the implementation issues and the performance of the introduced signalling schemes are discussed. The performance is compared in terms of required signalling overhead based on computer simulations. In the sequel, we will use the performance metric r_SD, which is the signalling bits to data bits ratio and defined as $r_\mathrm{SD} = \frac{L_{\mathcal{B}_{\mathrm{X},N_\mathrm{g}}}}{BK}$, to evaluate the required signalling overhead of different signalling schemes, where $L_{\mathcal{B}_{\mathrm{X},N_\mathrm{g}}}$ is the number of required signalling bits (overhead) based on the signalling scheme

X (representative notation for the above discussed signalling schemes), B is the number of loaded bits per OFDM symbol and K is the number of OFDM data symbols per packet.

3.4.1   Implementation aspects

The analyses performed until now aimed to reduce the signalling overhead. To produce practically feasible solutions, implementation issues have to be taken into account too. The coding approaches introduced in Section 3.2.1 do not require the a-priori knowledge of $P_{b_{i_1}, b_{i_2}, \cdots, b_{i_{N_g}}}$ which results in a maximum possible entropy $E_{\mathcal{B}_U}$ and consequently the highest amount of signalling overhead. By grouping multiple subcarriers into a coding unit (joint symbol), the overhead can be reduced as the average code word length per subcarrier $\frac{W_{N_g}}{N_g}$ approaches to $E_{\mathcal{B}_U}$. On the other side, the joint signalling scheme is more sensitive to decoding errors as each error may result in an incorrect signalling of N_g subcarriers. The coding complexity increases also exponentially with N_g. Based on this source model, $N_g = 3$ gives the lowest possible overhead as verified in Table 3.1. However, still an overhead of $\bar{L}_3 = 112$ bits is necessary giving $r_{SD} = \frac{\bar{L}_3}{BK} = \frac{112}{960} = 0.12$. The advantage of this coding scheme lies in easy implementation and universal application. No knowledge of the probability distribution is required. The number of signalling bits is always fixed so as the required transmission resource, independently on the current channel state $\hat{\mathbf{H}}^u$, packet duration K_F and the system SNR, leading to a simple resource allocation.

Once the probabilities $P_{b_{i_1}, b_{i_2}, \cdots, b_{i_{N_g}}}$ are available, more efficient coding schemes can be applied like the Huffman coding scheme which minimizes the redundancy of a code. For $N_g = 3$, the average signalling overhead is $\bar{L}_{H,3} = 65.28$ bits giving $r_{SD} = 0.068$, which reduces the overhead to the half in average. However, the number of required signalling bits is time-variant leading to a fluctuating resource requirement. A possible solution to this problem is provided by a mixed transmission strategy. First, a sufficiently large and fixed resource is reserved. The gap between the reserved resource and the currently occupied resource for signalling is filled by transmitting data bits. Obviously, the size of this gap has to be signalled too. A further fundamental problem of codes with variable code word length is catastrophic error propagation which leads to a total decoding mismatch after a single bit error. This phenomenon requires that the signalling bits have to be well protected via robust channel coding and modulation techniques.

The source to be signalled is the bit-loading vector $\mathbf{b}$, which is coupled to the channel estimate $\hat{\mathbf{H}}^u$ in a quantized form as described in Section 2.2.2. The correlation existing in the frequency and time domain for wireless channels corresponds to memory contained in this source, which can be utilized by state-dependent Huffman coding. For $N_g = 3$,

the exploitation of frequency-domain memory gives an average overhead of $\bar{L}_{\mathrm{F},3} = 53.87$ bits leading to $r_{\mathrm{SD}} = 0.056$, while by utilizing the temporary correlation the average overhead can be further reduced significantly to $\bar{L}_{\mathrm{T},3} = 24.13$ giving $r_{\mathrm{SD}} = 0.025$. Note that frequency-domain state-dependent Huffman coding cause an error propagation only in the frequency direction (inside a transmission packet), whereas, the time-domain state-dependent Huffman coding propagates errors in both frequency and time direction (the current packet and the subsequent packets) causing a severe error propagation problem. The above two state-dependent signalling schemes also require the knowledge of transitio probabilities, which is scenario-depedent. This fact limits the feasibility of these signalling schemes.

3.4.2   Simulation results

Computer simulations are conducted to evaluate the proposed signalling schemes in terms of r_{SD}. This ratio indicates the portion required for signalling bits in relation to the number of transmitted data bits. The underlying adaptive OFDM system is described in Appendix A.

Signalling overhead for a fixed packet size

We consider a constant data rate of $r_{\mathrm{D}} = 24$ Mpbs standardized as one possible transmission mode in IEEE 802.11a and a typical packet size of $K = 10$ OFDM data symbols, which gives a fixed number of $BK = 960$ data bits. In the standard, there is a 12-bits long $LENGTH$ field in physical layer convergence procedure (PLCP) header which indicates how many octets will be transmitted. The value K is determined based on the value of the $LENGTH$ field and the current data rate. In general, K depends on the used application, e.g. in real-time applications K is typically a small value.

For Huffman coding and the memory based signalling schemes, $L_{\mathcal{B}_{\mathrm{X},N_{\mathrm{g}}}}$ depends on the underlying channel state and the system SNR in addition. Consequently, r_{SD} is variable. In Figure 3.6, the cumulative distribution function of r_{SD} is plotted. The cumulative probability $P_{\mathrm{ss}}(r_{\mathrm{SD}}) = Pr(r_{\mathrm{SD}} \leq r_{\mathrm{SD},0})$ with $r_{\mathrm{SD},0}$ as a ratio of interest, is interpreted as the probability with which an overhead represented by $r_{\mathrm{SD},0}$ resulting from some signalling scheme X is sufficient for the signalling. We consider a typical SNR value of 20 dB and the channel model as well as the applied channel estimation algorithm given in Appendix A.

The signalling schemes in Section 3.2.1 are based on uniformly distributed symbols $P_{b_{i_1},b_{i_2},\cdots,b_{i_{N_{\mathrm{g}}}}}$. The signalling overhead $L_{\mathcal{B}_{\mathrm{U},N_{\mathrm{g}}}}$ is therefore constant. These signalling schemes result in a constant r_{SD}. Any value lower than the corresponding r_{SD} leads to a failed signalling trial. Values higher than r_{SD} ensures a successful signalling trial with
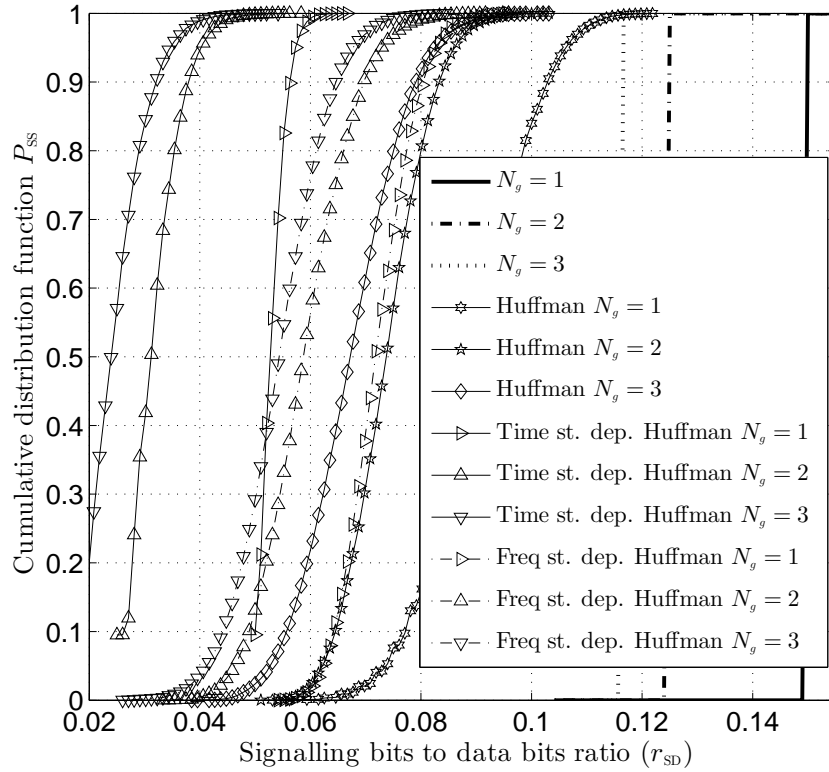
Figure 3.6:   Signalling overhead of proposed signalling schemes in terms of r_{SD} versus P_{SS}

$P_{\mathrm{SS}} = 1$. The relation between r_{SD} and P_{SS} is described by a step function with transition at exactly r_{SD}. For $N_{\mathrm{g}} = 1$, 2 and 3, the ratio is determined to $r_{\mathrm{SD}} = 0.15$, 0.125 and 0.1167 respectively.

Once the symbol probabilities $P_{b_{i_1}, b_{i_2}, \cdots, b_{i_{N_{\mathrm{g}}}}}$ are available, Huffman coding schemes can be applied. The signalling overheads are also simulated for $N_{\mathrm{g}} = 1$, 2 and 3. For $N_{\mathrm{g}} = 1$, a probability of $P_{\mathrm{SS}} = 0.9$ is achieved if $r_{\mathrm{SD}} = 0.1$ is ensured, while increasing N_{g} to 2 and 3, r_{SD} can be reduced to 0.083 and 0.08 respectively, for the same probability of $P_{\mathrm{SS}} = 0.9$. As shown in Figure 3.6, if $r_{\mathrm{SD}} = 0.1$ is allowed and $N_{\mathrm{g}} \geq 2$, a probability of $P_{\mathrm{SS}} \approx 1$ is possible compared to the signalling schemes above, a reduction of $\Delta r_{\mathrm{SD}} = 0.125 - 0.1 = 0.025$ is achieved, which corresponds to a net reduction of signalling bits of $\Delta \bar{L}_{\mathcal{B}_{\mathrm{H},2}} = \Delta r_{\mathrm{SD}} B K = 24$ bits.

The signalling schemes, which utilize the frequency-domain memory effect, are also simulated. The required signalling overhead $L_{\mathcal{B}_{\mathrm{F},N_{\mathrm{g}}}}$ is compromised of two contributions. The first contribution results from the initial symbol Z_0, which is encoded by Huffman coding scheme. The second contribution arises from the state-dependent symbols $Z_{n_{\mathrm{f}}}$ with $n_{\mathrm{f}} \geq 1$, which are encoded based on the state-dependent Huffman coding scheme as described in Section 3.2.3. For the symbol sizes of $N_{\mathrm{g}} = 1$, 2 and 3, a ratio of

$r_{\mathrm{SD}} = 0.08$, 0.07 and 0.065 are required respectively to reach a probability of $P_{\mathrm{SS}} = 0.9$. Comparing with the Huffman coding without utilizing memory, for $N_{\mathrm{g}} = 3$, a reduction in the overhead of $\Delta r_{\mathrm{SD}} = 0.08 - 0.065 = 0.015$ is possible giving a net reduction in signalling bits of $\Delta L_{\mathcal{B}_{\mathrm{F},3}} = \Delta r_{\mathrm{SD}} B K = 14.4$ bits. Furthermore, if a signalling overhead of $r_{\mathrm{SD}} = 0.08$ is provided, a probability of $P_{\mathrm{SS}} \approx 1$ can be achieved for both $N_{\mathrm{g}} = 2$ and 3 as confirmed in Figure 3.6. A comparison with the Huffman coding with $N_{\mathrm{g}} = 3$ gives $\Delta r_{\mathrm{SD}} = 0.1 - 0.08 = 0.02$ corresponding to $\Delta L_{\mathcal{B}_{\mathrm{F},3}} = 19.2$ bits. Compared to the coding schemes based on a uniformly distributed source, a reduction of $\Delta r_{\mathrm{SD}} = 0.1167 - 0.08 = 0.0367$ is achieved corresponding to $\Delta L'_{\mathcal{B}_{\mathrm{F},3}} = 35.23$ bits.

The signalling schemes based on the time-domain memory are simulated too. Just to show the potential of overhead reduction, only the steady-state signalling overhead is evaluated, which is slightly lower than the actually required overhead due to the overhead of the initial bit-loading vector $\mathbf{b}_0$. The more vectors $\mathbf{b}_{k_{\mathrm{s}}}$ are encoded, the smaller is the contribution of $\mathbf{b}_0$. In the practice, however, due to the problem of catastrophic error propagation, the number of packets using the time-domain state-dependent signalling scheme has to be limited.

For $N_{\mathrm{g}} = 1$, 2 and 3, the probability of $P_{\mathrm{SS}} = 0.9$ corresponds to an overhead of $r_{\mathrm{SD}} = 0.056$, 0.038 and 0.032 respectively. For $N_{\mathrm{g}} = 3$, a comparison of this signalling method with the methods based on the frequency-domain memory combined with state-dependent Huffman coding and the Huffman coding scheme results in a reduction of $\Delta r_{\mathrm{SD}} = 0.065 - 0.032 = 0.033$ and $0.08 - 0.032 = 0.048$ respectively, which corresponds to a reduction in net signalling bits of $\Delta L_{\mathcal{B}_{\mathrm{T},3}} = 31.68$ and 46.08 bits. For $N_{\mathrm{g}} = 1$, $r_{\mathrm{SD}} = 0.06$ gives approximately $P_{\mathrm{SS}} \approx 1$, while for $N_{\mathrm{g}} = 2$ and 3, r_{SD} can be reduced to around 0.045 for a probability of $P_{\mathrm{SS}} \approx 1$. This indicates that for each 100 data bits around 4.5 signalling bits are required in the steady state if the bit-loading vector $\mathbf{b}_{k_{\mathrm{s}}}$ is encoded by the time-domain state-dependent Huffman coding scheme, while the signalling method based on the uniformly distributed source model would require $0.1167 \times 100 = 11.67$ bits.

Signalling overhead for variable packet sizes

We have developed several signalling schemes, which utilized different information-theoretical properties of a common source. Some schemes result in an overhead $L_{X,N_{\mathrm{g}}}$ independent on K, e.g. schemes without considering time-domain memory effect. However, $L_{X,N_{\mathrm{g}}}$ based on the time-domain state-dependent Huffman coding schemes varies with K, as K impacts the time-domain correlation between consecutive bit-loading vectors. Intuitively, r_{SD} decreases with increasing K implying that large packet size causes low signalling overhead. From the standpoint of reducing signalling overhead, large K is desired. But in the practical system design K is limited due to time variance of the channel

and synchronization impairments.
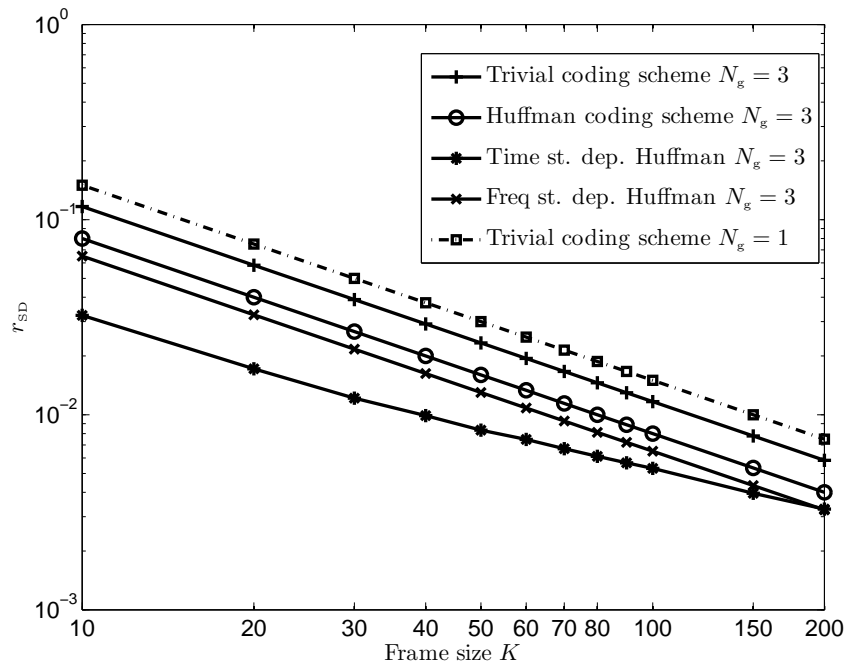
Figure 3.7: r_{SD} versus K for SNR $= 20$ dB and $P_{\mathrm{SS}} = 0.9$

Here, we investigate the dependence of r_{SD} on K for some signalling schemes, which is plotted in Figure 3.7. For the signalling schemes with variable signalling overhead, we consider a probability of $P_{\mathrm{SS}} = 0.9$ and a system SNR of 20 dB. The K-independent signalling schemes results in overheads, which decrease linearly in the log-log representation at the same slope. For small K values the Huffman coding based signalling scheme using the knowledge of the symbol probabilities $P_{b_1, \cdots, b_{N_{\mathrm{g}}}}$ achieves a remarkable reduction in signalling overhead, while for large K this reduction plays an unremarkable role in view of the error proprogation problem and implementation complexity. The further exploitation of frequency-domain memory effect results in a further reduction as confirmed in Figure 3.7. A high potential in reducing the overhead is achieved by the time-domain state-dependent Huffman coding, especially for small and middle packet size ($K \leq 50$). With increasing K, the time-domain correlation decrease more and more giving the fact that the decreasing slope becomes more flat and the overhead approaches that of other signalling schemes. If K is increased to certain value, it would require more signalling overhead than the frequency-domain correlation based signalling schemes. For sufficiently large K, this method will even degrade to Huffman coding based signalling scheme since the time-domain correlation will disappear completely.

PER of adaptive modulation with explicit signalling

The potential benefit of adaptive modulation in terms of packet error ratio (PER) improvement was demonstrated in Section 2.2.2. There, it was assumed that the adapted modulation schemes were perfectly synchronized between the transmitter and the receiver. We have to evaluate the PER under the realistic condition that the adapted modulation schemes have to be explicitly signalled. As mentioned, the PER performance will be certainly degraded to some extend. Simulations were conducted to quantify this degradation for the proposed signalling schemes. To avoid repetition, these simulation results will be shown in Chapter 5.

Compared to the system with fixed modulation, a system improvement is achieved if the system with adaptive modulation and explicit signalling shows a better PER performance. To avoid repetition, the performance evaluation in terms of PER will be provided in chapter 5, where all system scenarios are compared with each other: namely adaptive modulation with perfect knowledge about the adapted modulation at the receiver, explicit signalling, automatic modulation classification and fixed modulation. To ensure a fair comparison, the net data rate has to be constant for all scenarios.

3.4.3 Conclusion

Beginning with the most simple source model, which assumes uniformly distributed symbol probabilities $P_{b_{i_1}, b_{i_2}, \cdots, b_{i_{N_g}}}$, diverse source models have been proposed to describe the underlying signalling source. These models explore, from different aspects, the source-specific memory effects to reduce the required signalling overhead. The entropy was used as an information-theoretical criterion to quantify the potential of overhead reduction resulting from utilization of the existing memory effects.

CHAPTER 4

Automatic modulation classification

4.1 General framework

In the last chapter, we discussed modulation detection via explicit signalling, where the transmitter spends a certain number of OFDM symbols on transmitting information about the adapted modulation schemes. From the implementation point of view, this method requires few additional design and computational efforts as the signalling bits are just embedded into the existing data structure. At the receiver, before demodulation the signalling field has to be extracted and decoded.

Due to time-variance of wireless channels, this signalling information has to be updated periodically. Furthermore, since correct signalling is the necessary condition for a successful demodulation, this information has to be well-protected by applying robust modulation (e.g. BPSK) and powerful channel coding schemes (e.g. convolutional code with code rate of 1/2). So there exists a remarkable signalling overhead leading to a waste of data rate and transmit power. It was shown that even with sophisticated source coding methods, still a considerable portion of resources has to be reserved for the signalling purpose, especially for real-time applications where the data packet size is normally small. Moreover, methods based on the state-dependent Huffman coding scheme will also cause the well-known error propagation problem. Meanwhile, if no change on the existing data structure is desired, additional data rate is wasted by fixing the number of OFDM symbols used for signalling, since mostly the reserved signalling field is not completely occupied by the time-variant signalling overhead.

An approach to eliminate this signalling overhead is to perform automatic modulation classification (AMC) where the receiver is enabled to blindly recognize the adapted modulation formats solely relying on the received data symbols and some prior knowledge. AMC always includes two steps: signal conditioning and modulation classification as shown in Figure 4.1. Depending on the system under consideration, the signal conditioner may take steps like processing prior information and parameter estimation. In non-cooperative systems as normally the case in military applications, no prior knowledge is available about transmit parameters like signal power, noise power, carrier frequency,
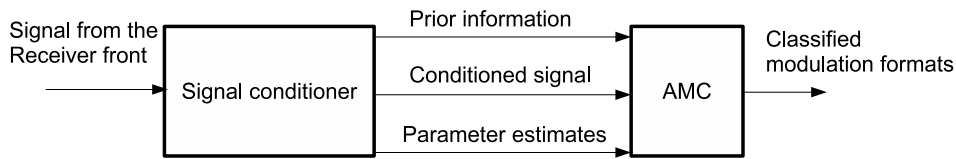
Figure 4.1: Schematic representation of modulation classification

transmit pulse, symbol duration and so on. Here the signal conditioner has to first estimate the above parameters. Certain results of signal conditioning are also necessary for the demodulation process afterwards. After these pre-processing tasks, the system is ready to carry out AMC. In the strategy, AMC can be categorized into two different frameworks: likelihood-based (LB) and feature-based (FB) [8]. In the LB framework, probabilistic properties of the received signal are used as the decision criterion. In the FB framework, certain modulation-specific features are extracted and compared against predefined thresholds to discriminate modulation candidates. Under the assumption that all information required for AMC is embedded in the observed signal, LB AMC is able to achieve the optimal classification performance in the sence that the misclassification probability is minimized, while FB AMC is in most cases only an suboptimal version but computationally more efficient than LB AMC.

Under the LB framework, AMC can be further classified into average likelihood ratio test (ALRT), generalized likelihood ratio test (GLRT) and hybrid likelihood ratio test (HLRT) depending on how the unknown parameters present in the received signal are treated. If all unknown parameters (the frequency and phase offset and the channel transfer function) are treated as random variables with known probability density functions (PDF), ALRT can be applied. In this approach, the unknown parameters are estimated by maximizing the PDFs with respect to the corresponding variables. This approach delivers the optimal solution in the sense that the misclassification probability is minimized. However, due to its high computational complexity and the unrealistic assumption that all PDFs are available, this method has only an academic interest as it provides the performance upper bound. In the GLRT, on the contrary, the unknowns are viewed as deterministic values, which have been already estimated by, e.g. training sequence based estimation methods. The best performance in this class is achieved by so-called uniformly most powerful test (UMPT) provided that this test exists [73]. The performance of GLRT is supposed to be poor since classification results depend sensitively on the precision of parameter estimation. In most practical systems, however, partial knowledge about the statistical properties of the unknowns is available. In this case HLRT can be used, which constitutes a compromise between ALRT and GLRT by estimating some unknowns

through probabilistic approaches and others through UMPT or training sequence based estimation.

Under the FB framework, certain discriminating features (signatures) are exacted from the received signal and compared against beforehand specified patterns. The modulation scheme with the pattern most closely matched to the received signal is selected as the classification output. The classification performance depends strongly on the choice of the discriminating features. Due to the heuristic nature of this framework, generally a long observation is required to achieve a reliable classification especially for modulation candidates from the same class, e.g., different QAM schemes. This requirement limits the application of the FB AMC in systems with small packet sizes like the IEEE 802.11 family. On the other side, AMC of this category can be constructed to be very computationally efficient and insensitive against mismatches in statistic models implying possible applications in non-cooperative environments like military applications.

Depending on the task of a classifier, whether to identify the modulation class, e.g. frequency-shifted keying (FSK), phase-shifted keying (PSK), amplitude-shifted keying (ASK) or QAM, or just to distinguish the modulation levels inside a class, different FB AMC algorithms have been proposed. In this work we concentrate on modulation classification inside a modulation class. The most primitive discriminating feature could be the information-bearing parameter as number of amplitude levels for ASK signals and phase levels for PSK signals. A second method uses a histogram of the information-bearing parameter, e.g. histogram of phase for PSK signals, where two peaks are expected for BPSK and four peaks for QPSK. Algorithms based on a histogram are grouped into the class called Haar wavelet transform [8]. Another group under the FB framework is based on high order statistics including moment-based AMC and cumulant-based AMC. Here a constellation diagram is mapped to another diagram through non-linear operations, e.g., 4-QAM is mapped to one single point by fourth order moment operation. Using the moment to cumulant conversion formula, cumulant-based AMC can be derived with two convenient properties: 1, the cumulant of the sum of two statistically independent random variables is equal to the sum of the cumulants of these two random variables; 2, the higher order cumulant (higher than second order) of an additve Gaussian noise is zero if the mean of the noise is zero. This implies that cumulant-based AMC is insensitive to zero-mean additive noise. There are large number of other AMC algorithms which also fall into FB category like fuzzy logic based AMC and neural network based AMC, which will not be discussed in this dissertation due to the limited scope.

In this chapter, both LB AMC and FB AMC will be applied to the considered system. System performance will be compared in terms of misclassification probability and

computational complexity.

4.2   Likelihood-based AMC

In this section, we apply LB AMC to our problem described in Section 1.2. The considered system is based on IEEE 801.11a. The received signal model is given in Section 1.2. Since a per-subcarrier bit-loading was carried out, we perform the modulation classification also on per-subcarrier basis. So, we drop the notation n for subcarrier index as long as a cross-subcarrier interaction is not necessary. In this signal model we have implicitly assumed:

- Synchronization parameters are ideally tracked and perfectly compensated.

- The unknown parameters are the channel transfer function H_k^{D}, the transmit symbol sequence $\{s_k\}_{k=0}^{K-1}$ and the adapted modulation scheme A_i represented by the modulation level b_i, which is to be classified.

In slowly time-variant channels $H_{k_1}^{\mathrm{D}} \approx H_{k_2}^{\mathrm{D}}$ for $0 \leq k_1, k_2 \leq K - 1$ and small packet size K, as typically the case in indoor applications, the channel can be assumed to be block-fading, so that the subscript k is dropped in H_k^{D}. In our considered system, H^{D} can be estimated by preamble-based channel estimation, which is specified in the standard. Consequently, H^{D} is viewed here as a determinstic value which is replaced by its estimate $\hat{H}^{\mathrm{D}}$. Instead, the unknown transmit symbol sequence $\{s_k\}_{k=0}^{K-1}$ is viewed as a random variable, which is randomly and independently drawn from the symbol alphabet A_i of the i-th modulation scheme. So the LB MAC applied here falls into the class of HLRT.

4.2.1   Maximum-likelihood-based AMC

The task of the HLRT classifier is to detect the used BLV **b** without any additional explicit signalling process, which is equivalent to classify b_i on each subcarriers. The detected number of loaded bits is denoted as $\hat{b}_i$ in the detected BLV $\hat{\mathbf{b}}$ with $\hat{b}_i \in \mathcal{B}$ too. Mathematically, the HLRT method can be formulated by

$$\hat{b}_{i,\mathrm{HLRT}} = \underset{b_i}{\mathrm{argmax}} \; p(b_i|\mathbf{r}), \tag{4.1}$$

where $\mathbf{r} = (r_0, r_1, ..., r_{K-1})^{\mathrm{T}}$ is the observed symbol vector. Note for simple notation, the subcarrier index n is omitted in the notation. Applying Bayes' law, it holds

$$p(b_i|\mathbf{r}) = \frac{p(\mathbf{r}|b_i)p(b_i)}{p(\mathbf{r})}. \tag{4.2}$$

Without any prior knowledge of $p(b_i)$ at the classifier, it is reasonable to assume uniformly distributed modulation levels:

$$p(b_i) = \tfrac{1}{I}. \tag{4.3}$$

Then, (4.1) can be rewritten as

$$\hat{b}_{i,\text{HLRT}} = \underset{b_i}{\text{argmax}} \, \frac{p(\mathbf{r}|b_i)}{I p(\mathbf{r})} = \underset{b_i}{\text{argmax}} \, p(\mathbf{r}|b_i), \tag{4.4}$$

as $\frac{1}{I}$ is a constant and $p(\mathbf{r})$ is irrelevant for building argmax. Since the channel state H^{D} is viewed as a known value and the elements r_k in $\mathbf{r}$ are uncorrelated

$$\text{E}[r_{k_1} r_{k_2}^*] = 0, \tag{4.5}$$

and Gaussian distributed, they are also statistically independent leading to

$$p(\mathbf{r}|b_i) = \prod_{k=0}^{K-1} p(r_k|b_i). \tag{4.6}$$

The received signal r_k is determined in (1.1), where H^{D} is replaced by the estimate $\hat{H}^{\text{D}}$. The noise w_k is assumed to be complex white Gaussian distributed. s_k is a discrete random variable conditioned on certain modulation scheme A_i. Consequently, it can take only a signal point from the symbol set $\{a_{i,1}, a_{i,2}, ..., a_{i,M_i}\}$. Assuming equiprobable signal points $p(a_{i,l}|b_i) = \frac{1}{M_i}$, it holds

$$p(r_k|b_i) = \sum_{l=1}^{M_i} p(r_k|a_{i,l}) p(a_{i,l}|b_i) = \frac{1}{M_i} \frac{1}{\pi N_0} \sum_{l=1}^{M_i} \exp(-\frac{|r_k - \hat{H}^{\text{D}} a_{i,l}|^2}{N_0}). \tag{4.7}$$

Then (4.4) is now equivalent to

$$\hat{b}_{i,\text{HLRT}} = \underset{b_i}{\text{argmax}} \, \prod_{k=0}^{K-1} \frac{1}{M_i} \frac{1}{\pi N_0} \sum_{l=1}^{M_i} \exp(-\frac{|\hat{H}^{\text{D}}|^2 |\hat{r}_k - a_{i,l}|^2}{N_0}), \tag{4.8}$$

with $\hat{r}_k = \frac{r_k}{\hat{H}^{\text{D}}}$ as the equalized received symbol. Cancelling the common term $(\frac{1}{\pi N_0})^K$, (4.4) is now equivalent to

$$\hat{b}_{i,\text{HLRT}} = \underset{b_i}{\text{argmax}} \, \left\{ \underbrace{(\frac{1}{M_i})^K \prod_{k=0}^{K-1} \sum_{l=1}^{M_i} \exp(-\rho^2 |\hat{r}_k - a_{i,l}|^2)}_{\kappa_1(\mathbf{r}|b_i)} \right\}, \tag{4.9}$$

with $\rho^2 = \frac{|\hat{H}^{\text{D}}|^2}{N_0}$. The right side of (4.9) is actually the HLRT-based likelihood function of the received symbol vector $\mathbf{r}$ conditioned on b_i and is denoted as $\kappa_1(\mathbf{r}|b_i)$. So HLRT AMC tests through all possible modulation candidates $\{b_i\}_{i=1}^I$ and selects $\hat{b}_{i,\text{HLRT}}$ out of this set which maximizes $\kappa_1(\mathbf{r}|m_i)$. In fact, HLRT AMC can be viewed as a multiple

hypothesis testing problem. We denote the likehood function of a received symbol r_k as $\kappa_1(r_k|b_i)$, which is given by

$$\kappa_1(r_k|b_i) = \left(\frac{1}{M_i}\right) \sum_{l=1}^{M_i} \exp(-\rho^2|\hat{r}_k - a_{i,l}|^2). \tag{4.10}$$

In the practical implementation, instead of the likelihood function itself, the logarithmic form is preferred. The reason will be explained in later sections. So (4.9) is equivalent to

$$\hat{b}_{i,\mathrm{HLRT}} = \underset{b_i}{\mathrm{argmax}} \ln \underbrace{\left\{(\frac{1}{M_i})^K \prod_{k=0}^{K-1} \sum_{l=1}^{M_i} \exp(-\rho^2|\hat{r}_k - a_{i,l}|^2)\right\}}_{\Lambda_1(\mathbf{r}|b_i)}, \tag{4.11}$$

with $\Lambda_1(\mathbf{r}|b_i)) = \ln\kappa_1(\mathbf{r}|b_i)$ and $\Lambda_1(r_k|b_i)) = \ln\kappa_1(r_k|b_i)$.

On a subcarrier basis, the evaluation of $\kappa_1(\mathbf{r}|b_i)$ requires M_iK exponential operations ($M_i = 2^{b_i}$). To make a decision, however, $\sum_{i=1}^{I} KM_i$ exponential functions have to be calculated, which shows an exponential increase of complexity with the modulation level b_i. In each exponential function itself, several multiplications and additions have to be computed. The likelihood function requires high computational efforts, which have to be reduced to enable practically feasible implementation.

4.2.2   1-point approximation

A straightforward simplification of $\kappa_1(\mathbf{r}|b_i)$ is based on the fast decaying property of the exponential function with decreasing arguments [46], [10], [74]. The principle is illustrated in Figure 4.2 for BPSK and 4-QAM. The extension to other QAM schemes is straightforward. Note that the result is exact for NoTx due to its one-point constellation. For BPSK, $\Lambda_1(r_k|b_i)$ of r_k is given by

$$\Lambda_1(r_k|b_i = 1) = \ln\frac{1}{2^1} + \ln[\exp(-\rho^2 D_{k,2,1}^2) + \exp(-\rho^2 D_{k,2,2}^2)], \tag{4.12}$$

with $D_{k,i,l}^2$ the squared Euclidean distance shown in Figure 4.2. Mathematically it is defined by

$$D_{k,i,l}^2 = |\hat{r}_k - a_{i,l}|^2 = \left(x_k - x^{(i,l)}\right)^2 + \left(y_k - y^{(i,l)}\right)^2, \tag{4.13}$$

where the complex baseband signals are represented in the coordinate system as $\hat{r}_k = x_k + \mathrm{j}y_k$ and $a_{i,l} = x^{(i,l)} + \mathrm{j}y^{(i,l)}$.

Due to $\exp(-\rho^2 D_{k,2,1}^2) \gg \exp(-\rho^2 D_{k,2,2}^2)$, for r_k with $\Re\{r_k\} < 0$, (4.12) is approximated by

$$\begin{aligned} \Lambda_1(r_k|b_i = 1) &\approx \ln\frac{1}{2^1} + \ln(\exp(-\rho^2 D_{k,2,1}^2)) \\ &\approx -\ln(2) - \rho^2 D_{k,2,1}^2. \end{aligned} \tag{4.14}$$
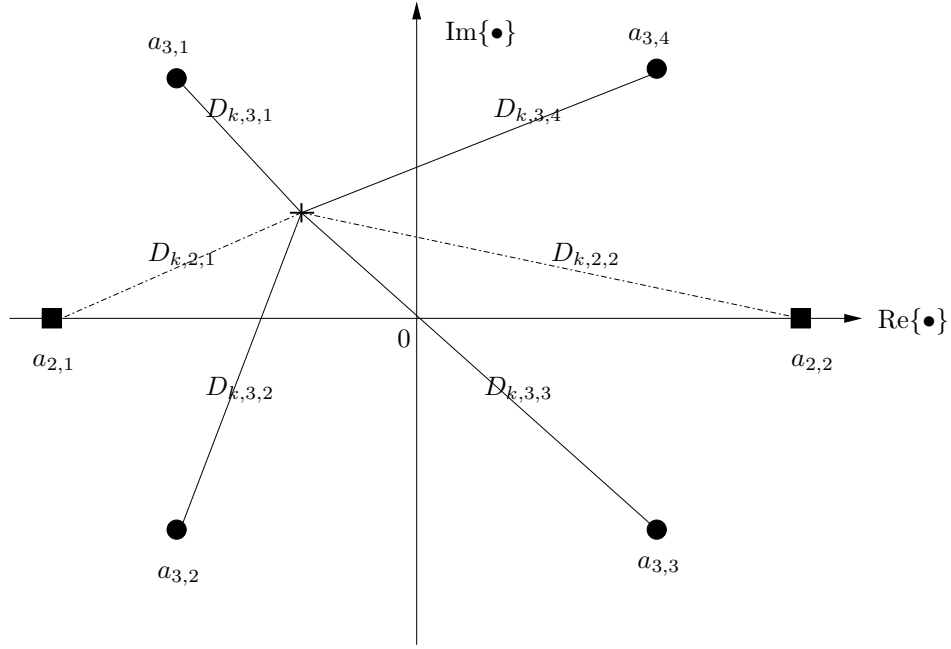
Figure 4.2: Principle of 1-point approximation for BPSK and 4-QAM

Similarly, the approximated $\Lambda_1(r_k|b_i = 2)$ is given by

$$
\begin{aligned}
\Lambda_1(r_k|b_i = 2) &\approx \ln\frac{1}{2^2} + \ln(\exp(-\rho^2 D_{k,3,1}^2)) \\
&\approx -2\ln(2) - \rho^2 D_{k,3,1}^2.
\end{aligned}
\tag{4.15}
$$

In general, if we denote $l_{\min,k,i}$ as the index of the signal point $a_{i,l_{\min,k,i}}$ of the i-th modulation format A_i which fulfils

$$
D_{k,i,l_{\min,k,i}}^2 \le D_{k,i,l}^2 \quad \text{with } l \ne l_{\min,k,i} \ \wedge \ 1 \le l_{\min,k,i}, l \le M_i,
\tag{4.16}
$$

the approximation is formulated in the general form

$$
\Lambda_1(r_k|b_i) \approx -b_i\ln(2) - \rho^2 \left\{ \left[x_k - x^{(i,l_{\min,k,i})}\right]^2 + \left[y_k - y^{(i,l_{\min,k,i})}\right]^2 \right\}.
\tag{4.17}
$$

Obviously, which signal point fulfils (4.16) depends on b_i and $\hat{r}_k$. The modulation classification is formulated by

$$
\hat{b}_{i,1P} = \operatorname*{argmax}_{b_i} \sum_{k=0}^{K-1} \left\{ \underbrace{-b_i\ln(2)}_{\text{prior probability}} - \right.
\tag{4.18}
$$

$$
\left. \underbrace{\rho^2 \left[-2x_k x^{(i,l_{\min,k,i})} - 2y_k y^{(i,l_{\min,k,i})} + \left(x^{(i,l_{\min,k,i})}\right)^2 + \left(y^{(i,l_{\min,k,i})}\right)^2 \right]}_{\text{Minimum distance based}} \right\},
$$

where $\rho^2(x_k^2 + y_k^2)$, giving the received signal power, is independent on b_i and can be omitted for AMC.

The approximated likelihood function in (4.18) is compromised of two contributions. One contribution is based on the minimum Euclidean distance between $\hat{r}_k$ and $a_{i,l_{\min,k,i}}$ stemming from A_i. The other contribution describes the prior probability of $a_{i,l_{\min,k,i}}$, conditioned on that $\hat{r}_k$ is nearest to $a_{i,l_{\min,k,i}}$ among all signal points from A_i, which is assumed to be equiprobable

$$p(a_{i,l_{\min,k,i}}|\hat{r}_k, b_i) = \tfrac{1}{2^{b_i}}, \qquad (4.19)$$

or in log-form: $\ln(p(a_{i,l_{\min,k,i}}|\hat{r}_k, b_i)) = b_i\ln(\tfrac{1}{2})$.

This assignment can be interpreted as a pessimistic probability assignment since the prior probability $p(a_{i,l_{\min,k,i}}|\hat{r}_k, b_i)$ of the point $a_{i,l_{\min,k,i}}$ is not assigned to be higher than other points $a_{i,l}$ out of A_i despite of the nearest distance.

In the following, we introduce another probability assignment: Keller proposed in [2] a probability assignment which can be viewd as an optimistic assignment. It states that if

$$D^2_{k,i,l_{\min,k,n,i}} < D^2_{k,i,l} \quad \text{for} \quad a_{i,k} \neq a_{i,l_{\min,k,i}} \qquad (4.20)$$

the prior probability must fulfil

$$p(a_{i,l}|\hat{r}_k, b_i) = \begin{cases} 1 & l = l_{\min,k,i} \\ 0 & l \neq l_{\min,k,i} \end{cases} \qquad (4.21)$$

It says that if $a_{i,l_{\min,k,i}}$ is nearest to $\hat{r}_{k,n}$, the transmitted signal point must be $a_{i,l_{\min,k,i}}$ so that it takes the assignment: $p(a_{i,l_{\min,k,i}}|\hat{r}_k, b_i) = 1$, and other points in A_i can not be the transmitted signal. We call the approximation, based on this optimistic assignment, as pure minimum Euclidean distance approximation, which will be described in Section 4.2.3.

4.2.3   Pure Euclidean distance approximation

Based on (4.21), (4.10) and (4.17), the log-likelihood function is approximated now by

$$\Lambda_1(r_k|b_i) \approx -b_i\ln(1) - \rho^2\left(\left(x_k - x^{(i,l_{\min,k,i})}\right)^2 + \left(y_k - y^{(i,l_{\min,k,i})}\right)^2\right) \qquad (4.22)$$

$$\approx \rho^2\left(\left(x_k - x^{(i,l_{\min,k,i})}\right)^2 + \left(y_k - y^{(i,l_{\min,k,i})}\right)^2\right).$$

This approximation leads to a modulation classification which is based purely on the Euclidean distance

$$\hat{b}_{i,\text{ED}} = \underset{b_i}{\operatorname{argmax}}\left\{\sum_{k=0}^{K-1} -\left[-2x_k x^{(i,l_{\min,k,i})} - 2y_k y^{(i,l_{\min,k,i})}\right.\right. \qquad (4.23)$$

$$\left.\left. + \left(x^{(i,l_{\min,k,i})}\right)^2 + \left(y^{(i,l_{\min,k,i})}\right)^2\right]\right\},$$

where as before $-\rho^2(x_k^2 + y_k^2)$ is omitted. Note that also ρ^2 is cancelled here which ignores the effect of the receiver-side SNR.

4.2.4 4-points approximation

Although the previous approximations have simplified the implementation complexity tremendously, these methods are based on one single signal point $a_{i,l_{\min,k,i}}$. The contribution of other signal points is completely ignored, which would cause a large simplification deviation especially for high level modulation schemes. On the other hand, we are treating signal constellations, which are strictly periodic and square-shaped. This convenient arrangement of the constellation points enables a more precise approximation, which will be demonstrated on the basis of 16-QAM. Figure 4.3 shows the complex signal plane with an arbitrary equalized symbol $\hat{r}_k = (x_k, y_k)$ and the signal points $a_{4,l} = (x^{(4,l)}, y^{(4,l)})$ with $l \in \mathcal{L}_{i=4} = \{1, 2, ..., 16\}$. The plane is segmented into 9 subplanes with boundaries given by dashed lines. The symbol (x_k, y_k) belongs to the p_k-th subplane with p_k as the subplane index and $p_k \in \mathcal{P}_{i=4} = \{1, 2, ..., 9\}$ if (x_k, y_k) lies in this subplane. The signal points $(x^{(4,l_k)}, y^{(4,l_k)})$ indicate these signal points which lie on the boundaries of or inside the subplane p_k and satisfy

$$D_{k,4,l_k}^2 < D_{k,4,\bar{l}_k}^2, \tag{4.24}$$

where both indices fulfil

$$l_k \in \mathcal{L}_{k,i=4}, \quad \bar{l}_k \in \bar{\mathcal{L}}_{k,i=4}, \text{ and } \bar{\mathcal{L}}_{k,i=4} \cap \mathcal{L}_{k,i=4} = \emptyset, \quad \bar{\mathcal{L}}_{k,i=4} \cup \mathcal{L}_{k,i=4} = \mathcal{L}_{i=4}, \tag{4.25}$$

where the subscripts k emphasizes the fact that $\mathcal{L}_{k,i=4}$ depends on $\hat{r}_k$. For the considered $\hat{r}_k$ in Figure 4.3, it holds

$$p_k = 1, \quad \mathcal{L}_{k,i=4} = \{1, 2, 3, 4\}. \tag{4.26}$$

Note that for simple notation purpose, an unconventional indexing ($\mathcal{L}_{k,i=4} = \{1, 2, 3, 4\}$) of signal points is used in Figure 4.3. Obviously, some points are shared by multiple subplanes, e.g. point $(x^{(4,3)}, y^{(4,3)})$ is shared by subplane 1, 2, 4 and 5. The points $(x^{(4,l_k)}, y^{(4,l_k)})$ build a square with $(x_0^{(4,p_k)}, y_0^{(4,p_k)})$ as the geometric central point, i.e.

$$x_0^{(4,p_k)} = \frac{1}{4} \sum_{l_k \in \mathcal{L}_{k,4}} x^{(4,l_k)} \text{and} \quad y_0^{(4,p_k)} = \frac{1}{4} \sum_{l_k \in \mathcal{L}_{k,4}} y^{(4,l_k)}. \tag{4.27}$$

The log-likelihood function in (4.10) for this received symbol $\hat{r}_k$ is given by

$$\Lambda_1(r_k|b_i = 4) = \ln \left\{ \frac{1}{16} \left[\sum_{l_k \in \mathcal{L}_{k,4}} \exp(-\rho^2 |\hat{r}_k - a_{4,l_k}|^2) \right. \right. \tag{4.28}$$
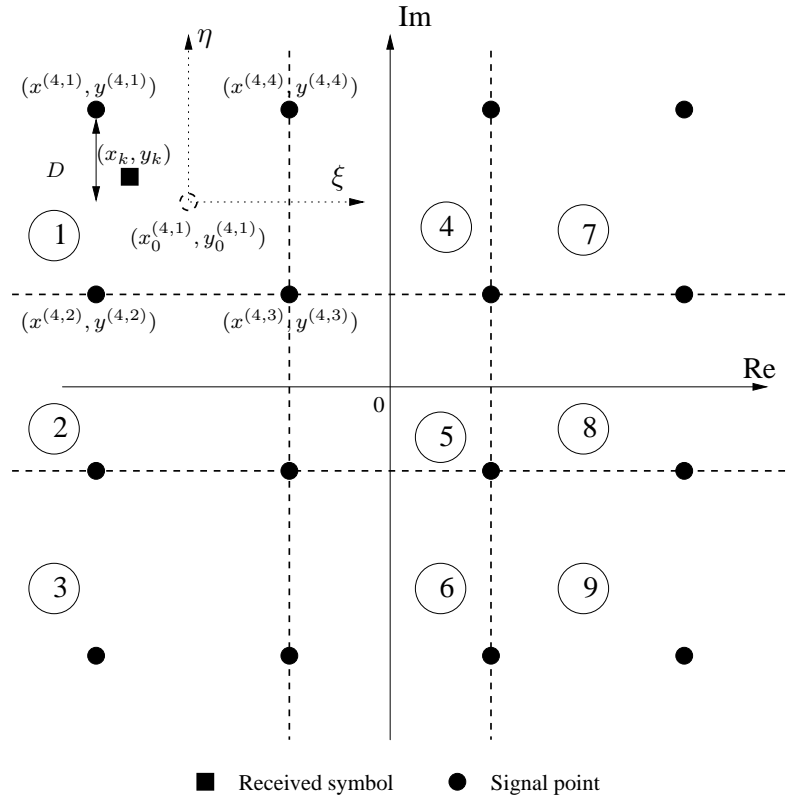
$$\left. \left. + \sum_{\bar{l}_k \in \bar{\mathcal{L}}_{k,4}} \exp(-\rho^2 |\hat{r}_k - a_{4,\bar{l}_k}|^2) \right] \right\}$$

Figure 4.3:   Principle of 4-points approximation for 16-QAM

Due to the fast decaying property of exponential functions, (4.28) is approximated by

$$\Lambda_1(r_k|b_i = 4) \approx \ln\left\{\frac{1}{16}\sum_{l_k \in \mathcal{L}_{k,4}} \exp(-\rho^2|\hat{r}_k - a_{4,l_k}|^2)\right\} \tag{4.29}$$

$$\approx \ln\left(\frac{1}{16}\sum_{l_k \in \mathcal{L}_{k,4}} \exp(-\rho^2\left[(x_k - x^{(4,l_k)})^2 + (y_k - y^{(4,l_k)})^2]\right)\right).$$

Perform the following coordinate transformation as shown in Figure 4.3

$$\xi = x - x_0^{(4,p_k)}, \quad \eta = y - y_0^{(4,p_k)}. \tag{4.30}$$

For the transformed coordinate system, the origin is now located at $(x_0^{(4,1)}, y_0^{(4,1)})$. Formulate (4.29) using the transformed coordinates

$$\Lambda(r_k|b_i = 4) \approx \ln\left(\frac{1}{16}\sum_{l_k \in \mathcal{L}_{k,4}} \exp\left(-\rho^2\left[\left(\xi_k - \xi^{(4,l_k)}\right)^2 + \left(\eta_k - \eta^{(4,l_k)}\right)^2\right]\right)\right). \tag{4.31}$$

Considering the constellation shown in Figure 4.3, due to the square and symmetric arrangement of QAM signals, it holds

$$\xi^{(4,1)} = \xi^{(4,2)} = -\xi^{(4,3)} = -\xi^{(4,4)}, \quad \eta^{(4,1)} = \eta^{(4,4)} = -\eta^{(4,2)} = -\eta^{(4,3)}. \tag{4.32}$$

Insert (4.32) into (4.31)

$$\Lambda_1(r_k|b_i = 4) \approx \ln\left\{\frac{1}{16}\left[\exp\left(-\rho^2\left[\left(\xi_k - \xi^{(4,1)}\right)^2 + \left(\eta_k - \eta^{(4,1)}\right)^2\right]\right) + \right.\right. \quad (4.33)$$

$$\exp\left(-\rho^2\left[\left(\xi_k - \xi^{(4,1)}\right)^2 + \left(\eta_k + \eta^{(4,1)}\right)^2\right]\right) +$$

$$\exp\left(-\rho^2\left[\left(\xi_k + \xi^{(4,1)}\right)^2 + \left(\eta_k + \eta^{(4,1)}\right)^2\right]\right) +$$

$$\left.\left.\exp\left(-\rho^2\left[\left(\xi_k + \xi^{(4,1)}\right)^2 + \left(\eta_k - \eta^{(4,1)}\right)^2\right]\right)\right]\right\},$$

which is rewritten by

$$\Lambda_1(r_k|b_i = 4) \approx \ln\left\{\frac{1}{4}\exp\left(-\rho^2\left(\xi_k^2 + \eta_k^2 + \left(\xi^{(4,1)}\right)^2 + \left(\eta^{(4,1)}\right)^2\right)\right) \times \right. \quad (4.34)$$

$$\left.\cosh\left(2\rho^2\xi_k\xi^{(4,1)}\right)\cosh\left(2\rho^2\eta_k\eta^{(4,1)}\right)\right\}$$

$$\approx -\rho^2\left(\xi_k^2 + \eta_k^2 + \left(\xi^{(4,1)}\right)^2 + \left(\eta^{(4,1)}\right)^2\right) +$$

$$\ln\left(\cosh\left(2\rho^2\xi_k\xi^{(4,1)}\right)\right) + \ln\left(\cosh\left(2\rho^2\eta_k\eta^{(4,1)}\right)\right).$$

It holds

$$\ln(\cosh(x)) = \ln\left(\frac{e^x + e^{-x}}{2}\right) \approx \begin{cases} \ln\left(\frac{e^x}{2}\right) = x - \ln(2), & \text{for } x \gg 1 \\ \ln\left(\frac{e^{-x}}{2}\right) = -x - \ln(2), & \text{for } x \ll -1 \end{cases} \quad (4.35)$$

$$\approx |x| - \ln(2) \quad \text{for}|x| \gg 1,$$

so that (4.33) is further approximated to

$$\Lambda_1(r_k|b_i = 4) \approx -\rho^2\left[\left(\xi_k^2 + \eta_k^2\right) + \left(\xi^{(4,1)}\right) + \left(\eta^{(4,1)}\right) + \right. \quad (4.36)$$

$$\left.2\left(\left|\xi_{k,n}\xi^{(4,1)}\right| + \left|\eta_{k,n}\eta^{(3,1)}\right|\right)\right] - \ln(2).$$

By back-transforming the coordinates, the 4-P approximated log-likelihood function is given by

$$\Lambda_{4-\text{P}}(r_k|b_i = 4) = -\rho^2\left[\left(x_k - x_0^{(4,1)}\right)^2 + \left(y_k - y_0^{(4,1)}\right)^2 + 2D_{0,i=4}^2\right] + \quad (4.37)$$

$$2D_{0,i=4}\rho^2\left[\left|x_k - x_0^{(4,1)}\right| + \left|y_k - y_0^{(4,1)}\right|\right] - 4\ln(2)$$

$$= -\rho^2\left[E_{r_k} + E_{4,1} - 2x_kx_0^{(4,p_k)} - 2y_ky_0^{(4,p_k)} + 2D_{0,i=4}^2\right] +$$

$$2D_{0,i=4}\rho^2\left[\left|x_k - x_0^{(4,1)}\right| + \left|y_k - y_0^{(4,1)}\right|\right] - 4\ln(2),$$

with $E_{r_k} = x_k^2 + y_k^2$, $E_{4,1} = \left(x_0^{(4,1)}\right)^2 + \left(y_0^{(4,1)}\right)^2$ and $D_{0,i=4} = |x(4,1) - x_0^{(4,1)}|$ as defined in Figure 4.3.

The 4-point approximation for 64-QAM can be similarly performed with corresponding meaning of the center point $(x_0^{(5,p_k)}, y_0^{(5,p_k)})$ and $D_{0,i=4} = x^{(5,p_k)} - x_0^{(5,p_k)}$. For NoTx, BPSK and 4-QAM, the principle is illustrated in Appendix E. In Table 4.1, the approximated $\frac{\Lambda(\mathbf{r}|b_i))}{\rho^2}$ is provided for considered modulation formats.

Table 4.1: Simplified log-likelihood function to obtain 4-point based classification $\hat{b}_{i,4-\mathrm{P}}$; $\gamma = \frac{\ln(2)}{\rho^2}$, $E_{i=2} = |a_{2,l}|^2$, $E_{i=3} = |a_{3,l}|^2$, $E_{4,p_k} = |x_0^{(4,p_k)}|^2 + |y_0^{(4,p_k)}|^2$ and $E_{5,p_k} = |x_0^{(5,p_k)}|^2 + |y_0^{(5,p_k)}|^2$.

b_i	4-point approximated log-Likelihood functions $\frac{\Lambda(\mathbf{r}|b_i))}{\rho^2}$
0	-0γ
1	$\sum_{k=0}^{K-1} 2|x_k| - KE_{i=2} - 1\gamma$
2	$\sum_{k=0}^{K-1} 2(|x_k x^{(3,1)}| + |y_k y^{(3,1)}|) - KE_{i=3} - 2\gamma$
4	$\sum_{k=0}^{K-1} \left[2(x_k x_0^{(4,p_k)} + y_k y_0^{(4,p_k)}) + 2D_{0,i=4}\left(\left|x_k - x_0^{(4,p_k)}\right| + \left|y_k - y_0^{(4,p_k)}\right|\right) \right] -$ $K(E_{4,p_k} + 2D_{0,i=4}^2) - 4\gamma$
6	$\sum_{k=0}^{K-1} \left[2(x_k x_0^{(5,p_k)} + y_k y_0^{(5,p_k)}) + 2D_{0,i=6}(|x_k - x_0^{(5,p_k)}| + |y_k - y_0^{(5,p_k)}||\right] -$ $K(E_{5,p_k} + 2D_{0,i=6}) - 6\gamma$

4.2.5   Bias effect and computational complexity

Under the assumption of equally likely distributed modulation formats $P_{b_i} = P_{b_j}$ for $b_i, b_j \in \mathcal{B}$, it is commonly recognized that maximum-likelihood (ML)-based classifier defined by $\kappa_1(\mathbf{r}|b_i)$ in (4.9) is optimal in the sense that it achieves the minimum misclassification probability according to Bayes' decision theory. We denote $P_{\mathrm{ML}}(\hat{b}_i|b_i)$ as the probability that the ML-based classifier decides on the modulation level $\hat{b}_i$

$$\hat{b}_i = \underset{b_i}{\mathrm{argmax}} \; \kappa_1(\mathbf{r}|b_i), \tag{4.38}$$

where b_i was adapted at the transmitter. The optimal ML-based classification rule according to (4.38) is considered as an unbiased decision criterion. The conditional misclassification probability is given by

$$P_{\mathrm{ML}}(mis|b_i) = 1 - P_{\mathrm{ML}}(\hat{b}_i|b_i) \quad \text{for } \hat{b}_i = b_i. \tag{4.39}$$

For I modulation candidates, the minimum misclassification probability based on the ML approach is given by

$$p_{\min} = \frac{1}{I} \sum_{i=1}^{I} P_{\mathrm{ML}}(mis|b_i). \tag{4.40}$$

Approximations of the likelihood function $\kappa_{1,\text{App}}(\mathbf{r}|b_i)$ will increase the misclassification probability

$$p_{\text{App}} = \frac{1}{I} \sum_{i=1}^{I} P_{\text{App}}(mis|b_i) > p_{\min}, \tag{4.41}$$

where it holds

$$P_{\text{App}}(mis|b_i) = 1 - P_{\text{App}}(\hat{b}_{i,\text{App}}|b_i) \quad \text{for } \hat{b}_{i,\text{App}} = b_i, \tag{4.42}$$

with

$$\hat{b}_{i,\text{App}} = \operatorname*{argmax}_{b_i} \kappa_{1,\text{App}}(\mathbf{r}|b_i). \tag{4.43}$$

Denoting $\Delta\kappa_{1,\text{App}}(\mathbf{r}|b_i)$ as the approximation deviation

$$\Delta\kappa_{1,\text{App}}(\mathbf{r}|b_i) = \kappa_1(\mathbf{r}|b_i) - \kappa_{1,\text{App}}(\mathbf{r}|b_i), \tag{4.44}$$

(4.43) is given by

$$\hat{b}_{i,\text{App}} = \operatorname*{argmax}_{b_i} \{\kappa_1(\mathbf{r}|b_i) - \Delta\kappa_{1,\text{App}}(\mathbf{r}|b_i)\}, \tag{4.45}$$

showing that the unbiased criterion given in (4.38) is now biased by $\Delta\kappa_{1,\text{App}}(\mathbf{r}|b_i)$. As $\Delta\kappa_{1,\text{App}}(\mathbf{r}|b_i)$ is obviously a function of b_i, the approximation given in (4.45) results in an b_i-dependent classification bias compared to the unbiased decision criterion.

In general, $\Delta\kappa_{1,\text{App}}(\mathbf{r}|b_i)$ depends also on the selected approximation. For 1-point approximation, it holds

$$\begin{aligned}
\Delta\kappa_{1,1-\text{P}}(\mathbf{r}|b_i) &= \kappa_1(\mathbf{r}|b_i) - \kappa_{1,1-\text{P}}(\mathbf{r}|b_i) \\[2mm]
&= \left(\frac{1}{2^{b_i}}\right)^K \left[\prod_{k=0}^{K-1} \sum_{l=1}^{2^{b_i}} \exp\left(-\rho^2 D_{k,i,l}^2\right) - \prod_{k=0}^{K-1} \exp\left(-\rho^2 D_{k,i,l_{\min,k,i}}^2\right) \right] \\[2mm]
&= \left(\frac{1}{2^{b_i}}\right)^K \left[\underbrace{\sum_{l_0}^{2^{b_i}} \cdots \sum_{l_{K-1}}^{2^{b_i}} \exp\left(-\rho^2 D_{1,i,l_0}^2\right) \cdots \exp\left(-\rho^2 D_{K,i,l_{K-1}}^2\right)}_{\left(2^{b_i}\right)^K \text{ addends}} - \right.
\end{aligned}$$

$$\left. \underbrace{\exp\left(-\rho^2 D_{1,i,l_{\min,k,i}}^2\right) \cdots \exp\left(-\rho^2 D_{K,i,l_{\min,K,i}}^2\right)}_{\text{maximum term}} \right], \tag{4.46}$$

with $D_{k,i,l_{\min,k,i}}^2$ and $l_{\min,k,i}$ defined in (4.13) and (4.16), respectively. The approximation deviation is given by the difference between $\kappa_1(\mathbf{r}|b_i)$, which is the sum of $\left(2^{b_i}\right)^K$ addends

given in multiplication form of exponential functions, and the maximum term under these addends (1-P approximation).

Generally, the relation of the likelihood functions depends on $\mathbf{r}$ and b_i leading to

$$\kappa_1(\mathbf{r}|b_i) \not> \kappa_1(\mathbf{r}|b_j) \quad \text{for } b_i > b_j, \tag{4.47}$$

which states that modulation level with more signal points will not be prefered in the decision than that with less signal points. This is actually the condition for an unbiased decision criterion. However, inside a modulation level b_i, due to the fast decaying property of exponential functions, the maximum addend dominates the other addends. Investigations have also shown that, after discarding this maximum addend, for any r_k the remaining addends take values from comparable ranges, quite independently on the underlying modulation level b_i

$$\frac{1}{2^{b_i}} \prod_{k=0}^{K-1} \exp\left(-\rho^2 D_{k,i,l_k}^2\right) \approx \frac{1}{2^{b_j}} \prod_{k=0}^{K-1} \exp\left(-\rho^2 D_{k,j,l_k}^2\right) \quad \text{for } b_i \neq b_j, \tag{4.48}$$

and

$$l_0 \neq l_{\min,0,i} \wedge l_1 \neq l_{\min,1,i} \wedge \cdots \wedge l_{K-1} \neq l_{\min,K-1,i}, \tag{4.49}$$

and

$$l_0 \neq l_{\min,0,j} \wedge l_1 \neq l_{\min,1,j} \wedge \cdots \wedge l_{K-1} \neq l_{\min,K-1,j}, \tag{4.50}$$

respectively. After substracting the respective dominating term (maximum term), it holds

$$\Delta\kappa_{1,1-\text{P}}(\mathbf{r}|b_i) > \Delta\kappa_{1,1-\text{P}}(\mathbf{r}|b_j) \quad \text{for } b_i > b_j, \tag{4.51}$$

as the number of the remaining terms fulfils the relation

$$\left(2^{b_i}\right)^K - 1 \gg \left(2^{b_j}\right)^K - 1 \quad \text{for } b_i > b_j. \tag{4.52}$$

Consequently, $\Delta\kappa_{1,1-\text{P}}(\mathbf{r}|b_i)$ takes larger values for larger b_i and is biased to higher modulation level. With negative sign prior to $\Delta\kappa_{1,1-\text{P}}(\mathbf{r}|b_i)$, decisions according to (4.45) are made in favour of smaller b_i and therefore biased to lower modulation levels.

Similarly, the 4-P approximation presents also a bias towards lower level modulations due to

$$\Delta\kappa_{1,4-\text{P}}(\mathbf{r}|b_i) > \Delta\kappa_{1,4-\text{P}}(\mathbf{r}|b_j) \quad \text{for } b_i > b_j, \tag{4.53}$$

Table 4.2: Computational efforts for a decision on the subcarrier basis; bias effects (- denotes bias towards lower level; + towards higher level); values in brackets give numbers of operations required for the considered system; † Required multiplication and addition operations are m_i-dependent: $\mathcal{O}_\times(0) = \mathcal{O}_+(0) = 0$; $\mathcal{O}_\times(1) = \mathcal{O}_\times(2) = K + 1$; $\mathcal{O}_\times(4) = \mathcal{O}_\times(6) = 4K + 2$; $\mathcal{O}_+(1) = K + 1$; $\mathcal{O}_+(2) = 2K + 1$; $\mathcal{O}_+(4) = \mathcal{O}_+(6) = 4K + 3$.

	exp func.	Multiplications	Additions	Bias
ML	$K \sum_{i=1}^{I} 2^{b_i}$ (870)	$2K \sum_{i=1}^{I} 2^{b_i}$ (1740)	$K \sum_{i=1}^{I} 2^{b_i}$ (870)	No
1-P	0	$I(1 + 5K)$ (255)	$4IK$ (200)	$--$
ED	0	$5IK$ (250)	$4I(K - 1)$ (200)	$+$
4-P	0	$\sum_{i=1}^{I} \mathcal{O}_\times(b_i)^\dagger \leq I(4K + 2)$ (≈ 255)	$\sum_{i=1}^{I} \mathcal{O}_+(b_i)^\dagger \leq I(4K + 3)$ (≈ 200)	-

as confirmed by the derivations in Appendix E.1. Here instead of only 1 maximum term, 4 largest terms are considered leading to

$$\Delta\kappa_{1,1-\mathrm{P}}(\mathbf{r}|b_i) > \Delta\kappa_{1,4-\mathrm{P}}(\mathbf{r}|b_i) \tag{4.54}$$

Consequently, the 4-P approximation produces a smaller bias effect and therefore, this approach is closer to the unbiased decision criterion.

In the contrast, the Euclidean distance based classifier shows a bias towards higher modulation levels as

$$\Delta\kappa_{1,\mathrm{ED}}(\mathbf{r}|b_i) < \Delta\kappa_{1,\mathrm{ED}}(\mathbf{r}|b_j) \quad \text{for } b_i > b_j. \tag{4.55}$$

The derivations thereof are also provided in the Appendix E.2. This phenomenon can be intuitively explained by the fact that, since in this approximation method only pure Euclidean distance is used as the decision criterion, higher level modulation schemes with higher density of signal points in the signal plane show potentially a higher probability to be able to provide a signal nearer to a received signal point (smaller Euclidean distance between a potential signal point and the received point).

The purpose of the introduced approximations is to reduce the implementation complexity. Table 4.2 gives an overview of the presented approximations with respect to computational complexity in a per-subcarrier-based decision and the associated bias effect. Note that these approximations also simplify the subsequent demodulation process.

Once the modulation format has been decided, the demodulated symbol is also determined for 1-point approximation and Euclidean distance or limited to maximal four signal points for 4-point approximation.

4.2.6   Simulation results

In the previous sections, we assumed uniformly distributed symbols b_j in the set $\mathcal{B}$. In other words, the modulation levels are randomly and equiprobably assigned to subcarriers. In systems with adaptive modulation, however, this assumption does not hold anymore. With the adaptive bit loading algorithm adopted in this thesis, modulation levels are assigned based on the channel estimate vector $|\hat{\mathbf{H}}|$ and the number of bits B to be loaded in one OFDM symbol. Consequently, lower level modulations are assigned to subcarriers with poor link quality and potentially less classifiable. Correspondingly, higher level formats are assigned to high quality subcarriers, where signal points are transmitted more reliably leading to a better classifiability.

Conditional misclassification probabilities $P_{\mathrm{ML}}(mis|b_i)$

The above analysis is confirmed in Figure 4.4, where the conditional misclassification probabilities $P_{\mathrm{ML}}(mis|b_i)$, which are based on the ML classifier, are plotted versus transmit-side SNR, which is defined as

$$\text{Average SNR} = \tfrac{E_{\mathrm{S}}}{E_{\mathrm{N}}} = \frac{\mathrm{E}_l[|a_{i,l}|^2]}{N_0}, \tag{4.56}$$

with E_{S} and E_{N} as the transmit signal power and noise power, respectively. Two experiments are conducted here: 1, random and equiprobable modulation (REM) assignment ($P'_{\mathrm{ML}}(mis|b_i)$ given in thin dashed line) and 2, adaptive modulation (AM) assignment ($P_{\mathrm{ML}}(mis|b_i)$ given in thick solid line).

In the first experiment, the conditional misclassification probabilities fulfil

$$P'_{\mathrm{ML}}(mis|b_i) > P'_{\mathrm{ML}}(mis|b_j) \quad \text{for} \quad b_i > b_j, \tag{4.57}$$

as higher level formats carry more information per unit bandwidth and power and are consequently more difficult to classify than lower level formats. Due to adaptive modulation, however, the classification performance is significantly shifted as shown in the second experiment

$$P_{\mathrm{ML}}(mis|b_i = 0) \gg P_{\mathrm{ML}}(mis|b_i = 1) \approx P_{\mathrm{ML}}(mis|b_i = 2) \approx \tag{4.58}$$
$$P_{\mathrm{ML}}(mis|b_i = 4) \gg P_{\mathrm{ML}}(mis|b_i = 6)$$

Here lower level formats are less classifiable than higher level formats. The similar performance observed for $b_i = 1, 2, 4$ results from the B-constraint imposed on the system
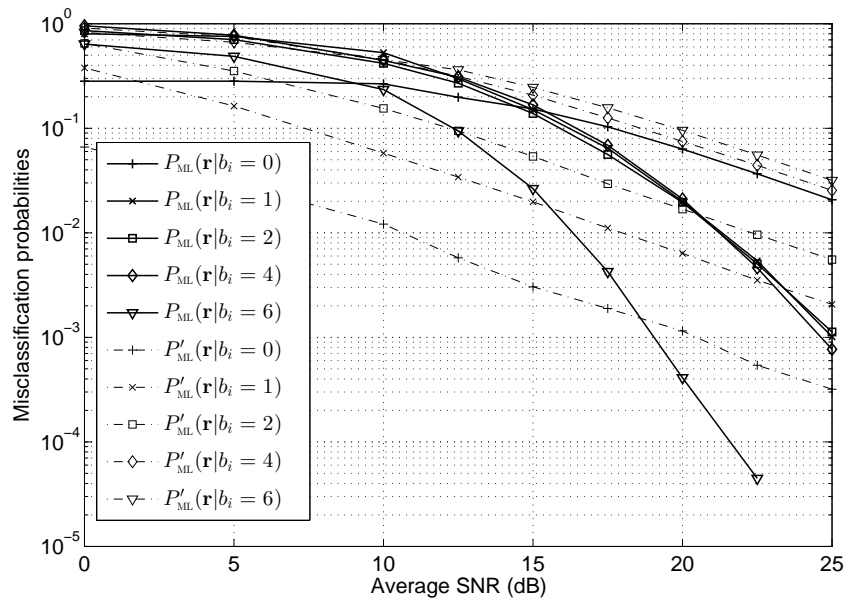
Figure 4.4: $P_{\mathrm{ML}}(mis|b_i)$ based on ML classifier in two scenarios 1) REM (dashed) and 2) AM (solid)

as formulated in (2.17). In this experiment, we fix the data rate $r_{\mathrm{D}} = 24$ Mpbs giving $B = 192$ bits, which corresponds to 4 bits per subcarrier in average. To satisfy this constraint, the modulation level $b_i = 4$ has to be frequently adapted. This leads to overloading problems, i.e. subcarriers are forced to use higher level formats than it would be allowed from the viewpoint of transmission quality. So on some subcarriers, modulation level $b_i = 2$ may be adjusted to $b_i = 4$. Similarly, the levels $b_i = 1$ and $b_i = 0$ are changed to their next higher level $b_i = 2$ and $b_i = 1$, respectively. This overloading problem leads to degradation in classification reliability. On the other side, the adaptive modulation always tries to ensure the best possible performance under a given circumstance, which requires that the transmission quality of these modulation schemes, which are involved in the overloading problem, must be balanced under each other. Consequently these modulation levels share similar classification performance.

Probability of subcarrier misclassification P_{S}

Now the ML classifier and the introduced approximated classifiers are applied on these two experiments to recognize the assigned modulation levels on each subcarrier. Classification performance is evaluated in Figure 4.5 in terms of probability of subcarrier misclassification P_{S} given by

$$P_{\mathrm{S}} = \frac{\mathrm{E}[N_{\mathrm{mis}}]}{N}, \tag{4.59}$$

where $\mathrm{E}[N_{\mathrm{mis}}]$ is the average number of misclassified subcarriers per packet. In the first experiment with random and equally likely modulation assignment, the ML classifier with unbiased decision criterion delivers the minimal misclassification probability $P'_{\mathrm{S,ML}}$ as stated in Section 4.2.1, which is followed by the biased 4-P approximation $P'_{\mathrm{S,4-P}}$ and 1-P approximation $P'_{\mathrm{S,1-P}}$. As 4-P approximation produces less bias effect, it outperforms the 1-P approximation. The Euclidean distance approximation gives the worst performance $P'_{\mathrm{S,ED}}$ as it is unsuitable to classify nested modulation formats, i.e. 4-QAM, 16-QAM and 64-QAM [37].
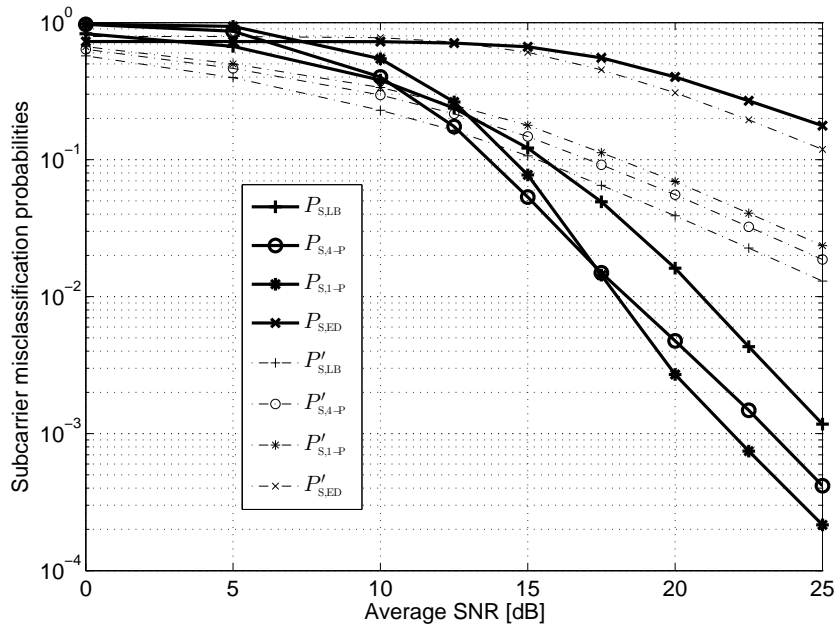
Figure 4.5: P_{S} based on different classifier in two scenarios 1) REM (dashed) and 2) AM (solid)

The same classifiers applied on the second experiment, however, produce totally different results. Here, the 1-P $P_{\mathrm{S,1-P}}$ and 4-P approximation $P_{\mathrm{S,4-P}}$ with bias towards lower level formats outperform the unbiased ML classifier $P_{\mathrm{S,ML}}$, which is supposed to be optimal. This phenomenon is caused by the link adaptation, which shifts the classifiability as given in (4.58). In the situation, where due to poor link quality the classification is not reliable, 1-P and 4-P approximation make decisions in favour of lower level formats, which are, in fact, most likely adapted to this link. This match of the bias effect to the nature of the adaptive modulation can be interpreted as an implicit utilization of certain prior knowledge about the adaptation strategy, which is now incidently incorporated in the approximation. This prior knowledge is not utilized in the unbiased ML classifier. Due to the additional prior knowledge the 1-P and 4-P approximation outperform the supposed optimal ML classifier.

Furthermore, compared to the results in the first experiment all the three classifiers improve their own performance, since adaptive modulation improves in general the transmission quality. An exception is given by the ED approximation, which continues to give the worst performance. Due to the bias towards higher level formats, which is mismatched to the adaptation strategy, the performance $P_{\text{s,ED}}$ is even degraded by the link adaptation.

4.3 MAP-based AMC

In the past, modulation classification was mainly of interest in military fields like thread analysis, secured transmission and electronic surveillance. In such non-cooperative environments, the classifier is not able to possess any prior knowledge about the signal parameters such as signal power, carrier frequency, symbol duration as well as the applied modulation scheme. The received (observed) signal is the only source to provide information for modulation classification. Under this circumstance, two general frameworks have been developed, which were introduced in the previous section, to solve blind modulation recognition problems. The LB framework is formulated as a multiple hypothesis testing problem and builds a probabilistic argument as the decision rule based on statistical properties of the received signal. The detected modulation scheme is the one who maximizes this probabilistic argument. The FB framework, which will be discussed in more detail in next section, instead, follows a hierarchical decision strategy and divides the modulation candidates into binary subsets, which are represented as nodes in a binary decision tree. Characteristic signatures of the candidates are extracted from the received signal to distinguish between these two subsets. The classification process is completed when these nodes, who contain only a single candidate, are reached in the constructed decision tree.

Nowadays, AMC is drawing more and more research interest also in civilian applications such as intelligent communication systems with adaptive modulation, where certain cooperation between the transmitter and receiver is organized. This cooperation enables the estimation of unknown parameters embedded in the signal through inserting training (preamble) symbols. The estimated parameters can be viewed as deterministic known values and used for subsequent modulation classification, which significantly simplifies the classification complexity.

Furthermore, the classifier can be informed about the applied AM strategy. Based on this a-priori information, the classifier can adjust its classification strategy to fit with the applied AM. The ML approach does not deliver the minimum misclassification probability any more as confirmed in Section 4.2.6, as the PDF of the observed waveform,

conditioned on the modulated signal, does not contain the full information for modulation classification [10], [46], [48]. It was shown that if the a-priori information is utilized efficiently, a significant improvement in the classification performance is possible [11], [9]. In this section we will formulate a maximum-a-posteriori (MAP) based classifier, which is potentially able to minimize the misclassification probability again.

4.3.1 Optimal MAP classifier

If we consider the DL transmission of a TDD based adaptive OFDM system as described in Section 1.2, where the BS adapts the modulation formats according to the latest incoming UL channel estimate $\hat{\mathbf{H}}^{\mathrm{U}}$ under the consideration of the constraints given in (2.17). Knowing the AM strategy, the MS can in turn incorporate $\hat{\mathbf{H}}^{\mathrm{U}}$ and B, which represents the total number of bits to be loaded, as a-priori information into the classification algorithm. In the standard of IEEE 802.11a, B is already provided at the receiver through in-band signalling [87]. If the transmission channel is reciprocal, $\hat{\mathbf{H}}^{\mathrm{U}}$ can be approximated by $\hat{\mathbf{H}}^{\mathrm{D}}$, which is estimated at the MS based on preambles received in the last incoming DL transmission packet as shown in Figure 1.1. Due to the iterative cross-subcarrier bit loading mentioned in Section 2.2.1, the number of bits $b_i^{(n)}$ loaded on the n-th subcarrier depends on the whole channel state vector $\hat{\mathbf{H}}^{\mathrm{U}}$ leading to a modulation classification with interactions between subcarriers. In the following, we classify the modulation level $b_i^{(n)}$ on the n-th subcarrier and denote the classifed modulation level as $\hat{b}_i^{(n)}$. Based on the observed signal and the knowledge about the used bit loading algorithm, the optimum modulation classification approach is formulated by maximizing the a posteriori probability

$$\hat{b}_{i,\mathrm{MAP}}^{(n)} = \underset{b_i}{\operatorname{argmax}}\, p(b_i|\mathbf{r}_n, \hat{\mathbf{H}}^{\mathrm{D}}, B), \tag{4.60}$$

where $\hat{b}_{i,\mathrm{MAP}}^{(n)}$ is the classified modulation level based on MAP approach, $\mathbf{r}_n = (r_{0,n}, r_{1,n}, \cdots, r_{K-1,n})^{\mathrm{T}}$ is the observed symbol vector and $\hat{\mathbf{H}}^{\mathrm{D}}$ is the channel estimate vector at the MS. We consider here the DL modulation classification. Obviously, the derived classification rules can be also applied for the UL.

Bayes′ law gives:

$$p(b_i|\mathbf{r}_n, \hat{\mathbf{H}}^{\mathrm{D}}, B) = \frac{p(\mathbf{r}_n, b_i, B, \hat{\mathbf{H}}^{\mathrm{D}})}{p(\mathbf{r}_n, B, \hat{\mathbf{H}}^{\mathrm{D}})}. \tag{4.61}$$

Omitting $p(\mathbf{r}_n, B, \hat{\mathbf{H}}^{\mathrm{D}})$ due to the independence on b_i, (4.60) is equavelent to

$$\hat{b}_{i,\mathrm{MAP}}^{(n)} = \underset{b_i}{\operatorname{argmax}}\, p(b_i, \mathbf{r}_n, \hat{\mathbf{H}}^{\mathrm{D}}, B). \tag{4.62}$$

Applying Bayes' law again, we obtain

$$p(b_i, \mathbf{r}_n, \hat{\mathbf{H}}^{\mathrm{D}}, B) = p(\mathbf{r}_n | \hat{\mathbf{H}}^{\mathrm{D}}, b_i, B) p(\hat{\mathbf{H}}^{\mathrm{D}}, b_i, B). \qquad (4.63)$$

We assume statistically independent symbols as before

$$p(\mathbf{r}_n | \hat{\mathbf{H}}^{\mathrm{D}}, b_i, B) = \prod_{k=0}^{K-1} p(r_{k,n} | \hat{\mathbf{H}}^{\mathrm{D}}, b_i, B) = \prod_{k=0}^{K-1} p(r_{k,n} | \hat{H}_n^{\mathrm{D}}, b_i), \qquad (4.64)$$

where, for a given b_i, $r_{k,n}$ is independent of B and $\hat{H}_{n'}^{\mathrm{D}}$ for $n' \neq n$ as shown in (1.1). Assuming equiprobable constellation points, $p(r_{k,n} | \hat{H}_n^{\mathrm{D}}, b_i)$ is given in (4.7), so that (4.64) is actually the likelihood function defined by $\kappa_1(\mathbf{r}_n | b_i)$ in (4.9).

On the other hand, the joint probability can be rewritten as

$$p(\hat{\mathbf{H}}^{\mathrm{D}}, b_i, B) = p(|\hat{\mathbf{H}}^{\mathrm{D}}|, \hat{\boldsymbol{\Phi}}, b_i, B) \qquad (4.65)$$

where $|\hat{\mathbf{H}}^{\mathrm{D}}| = (|\hat{H}_0^{\mathrm{D}}|, |\hat{H}_1^{\mathrm{D}}|, ..., |\hat{H}_{N-1}^{\mathrm{D}}|)^{\mathrm{T}}$ is the estimated channel magnitude vector and $\hat{\boldsymbol{\Phi}} = (\hat{\Phi}_0, \hat{\Phi}_1, ..., \hat{\Phi}_{N-1})^{\mathrm{T}}$ is the estimated phase vector. Since phase information is irrelevant to AM, it can be integrated out here for AMC. Combining (4.61), (4.62), (4.63) and (4.64) the sufficient statistic for modulation classification is then provided by

$$\hat{b}_{i,\mathrm{MAP}}^{(n)} = \underset{b_i}{\mathrm{argmax}} \left\{ p(|\hat{\mathbf{H}}^{\mathrm{D}}|, b_i, B) \prod_{k=0}^{K-1} p(r_{k,n} | \hat{H}_n^{\mathrm{D}}, b_i) \right\}. \qquad (4.66)$$

Futhermore, it holds

$$p(|\hat{\mathbf{H}}^{\mathrm{D}}|, b_i, B) = p(b_i | |\hat{\mathbf{H}}^{\mathrm{D}}|, B) p(|\hat{\mathbf{H}}^{\mathrm{D}}|, B), \qquad (4.67)$$

and $p(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)$ is independent of b_i, so the sufficient statistic becomes

$$\hat{b}_{i,\mathrm{MAP}}^{(n)} = \underset{b_i}{\mathrm{argmax}} \left\{ p(b_i | |\hat{\mathbf{H}}^{\mathrm{D}}|, B) \prod_{k=0}^{K-1} p(r_{k,n} | \hat{H}_n^{\mathrm{D}}, b_i) \right\}$$

$$= \underset{b_i}{\mathrm{argmax}} \left\{ \underbrace{p(b_i | |\hat{\mathbf{H}}^{\mathrm{D}}|, B)}_{\kappa_2(b_i | |\hat{\mathbf{H}}^{\mathrm{D}}|, B)} \underbrace{(\frac{1}{M_i})^K \prod_{k=0}^{K-1} \sum_{l=1}^{M_i} \exp(-\frac{|r_{k,n} - \hat{H}_n^{\mathrm{D}} a_{i,l}|^2}{N_0})}_{\kappa_1(\mathbf{r}_n | b_i)} \right\}, \qquad (4.68)$$

where $\kappa_1(\mathbf{r}_n | b_i)$ is the likelihood function as already defined in (4.9) and $\kappa_2(b_i | |\hat{\mathbf{H}}^{\mathrm{D}}|, B)$ is the a-priori probability of b_i based on the a priori information $|\hat{\mathbf{H}}^{\mathrm{D}}|$ and B. Note that for some $|\hat{\mathbf{H}}_0^{\mathrm{D}}|$ and B_0, the link adaptation described in Section 2.2.1 leads to $\kappa_2(b_i | |\hat{\mathbf{H}}_0^{\mathrm{D}}|, B_0) \neq \kappa_2(b_j | |\hat{\mathbf{H}}_0^{\mathrm{D}}|, B_0)$ for $i \neq j$. This reflects the nature of the used adaptive bit loading algorithm. In the literature, especially for non-cooperative environments,

however, the a-priori probability of different modulation schemes is usually assumed to be equal. This is true only if the classifier does not possess any prior information, e.g., in military applications. Here, the a-priori information available at the receiver is incorporated into the classification rule resulting in a maximum-a-posteriori (MAP) classifier. In the sequel, for an easier implementation we will consider (4.66).

The estimated channel magnitude vector $|\hat{\mathbf{H}}^{\mathrm{D}}|$ in $p(|\hat{\mathbf{H}}^{\mathrm{D}}|, b_i, B)$ has a dimension of N with N the number of data subcarriers in the considered OFDM systems. In the practice, N usually is a large number, e.g. in the IEEE 802.11a system $N = 48$ is used, so that the evaluation of $p(|\hat{\mathbf{H}}^{\mathrm{D}}|, b_i, B)$ requires a treatment of a high-dimensional distribution function. To enable practically feasible solutions, we propose several approximations to reduce the dimension of this function.

4.3.2   1-D approximated MAP classifier

Suppose that the modulation scheme on the n_0-th subcarrier is to be classified, which is equivalent to detect the number of loaded bits $b^{(n_0)}$. The AM algorithm requires that the total error probability p_{AM} has to be minimized

$$p_{\mathrm{AM}} = \min_{\mathbf{b}} p_{\mathrm{bit}}(\mathbf{b}), \quad \text{with } \mathbf{b} = (b^{(0)}, b^{(1)}, \cdots, b^{(N-1)})^{\mathrm{T}} \text{ and } b^{(n)} \in \mathcal{B}. \tag{4.69}$$

Due to the second constraint imposed on the adaptive modulation, which requires

$$\sum_{n=0}^{N-1} b^{(n)} = B, \tag{4.70}$$

an iterative adjustment of bit loading is necessary leading to a cross-subcarrier optimization as analyzed in Section 2.2.1. Consequently the entire elements of the magnitude vector $|\hat{\mathbf{H}}^{\mathrm{D}}|$ have to be considered as formulated above. This iterative adjustment begins with an initial value $\bar{b}^{(n_0)}(l_{\mathrm{it}} = 0)$ given by

$$\bar{b}^{(n_0)}(l_{\mathrm{it}} = 0) = \left\lfloor \log_2 \left(1 + \frac{\rho_{n,\mathrm{r}}^2}{\Gamma \lambda_{\mathrm{margin}}(l_{\mathrm{it}})} \right) \right\rfloor, \tag{4.71}$$

where $\lfloor \cdot \rfloor$ is a round-down function to ensure $\bar{b}^{(n_0)}(l_{\mathrm{it}} = 0) \in \mathcal{B}$. The parameters $\rho_{n,\mathrm{r}}^2$, Γ and $\lambda_{\mathrm{margin}}(l_{\mathrm{it}} = 0)$ are given by the system requirements as explained in Section 2.2.1. So $\bar{b}^{(n_0)}(l_{\mathrm{it}} = 0)$ is an unique function of $\hat{H}_{n_0}^{\mathrm{U}}$. The final value $b^{(n_0)}(l_{\mathrm{it}} = l_{\mathrm{over}})$, when the iterative bit loading is completed, however, may have to take another number $b^{(n_0)} \neq \bar{b}^{(n_0)}(l_{\mathrm{it}} = 0)$ to satisfy (4.70). But due to the error-constraint given in (4.69), $b^{(n_0)}$ will not deviate much from $\bar{b}^{(n_0)}(l_{\mathrm{it}} = 0)$ so that $|\hat{H}_{n_0}^{\mathrm{D}}|$ with $|\hat{H}_{n_0}^{\mathrm{D}}| \approx |\hat{H}_{n_0}^{\mathrm{U}}|$ still provides a-priori information of most significance. Channel magnitudes of other subcarriers $|\hat{H}_n^{\mathrm{D}}|$ for $n \neq n_0$ play here only a secondary role. For simplification purpose we ignore this

secondary effect at this moment and take only the most dominant channel coefficient $|\hat{H}_{n_0}^{\mathrm{D}}|$ into account [11]. So $|\hat{\mathbf{H}}^{\mathrm{D}}|$ is substituted by $|\hat{H}_{n_0}^{\mathrm{D}}|$ in (4.66). The dimension is now reduced from N to one. So we call it one-dimensional (1-D) approximation based MAP classification. The *1-D Classifier* for the n-th subcarrier is given by

$$\hat{b}_{i,1-\mathrm{D}}^{(n)} = \underset{b_i}{\operatorname{argmax}} \left\{ p(|\hat{H}_n^{\mathrm{D}}|, b_i, B)\kappa_1(\mathbf{r}_n|b_i) \right\}. \tag{4.72}$$

We use the relationship

$$p(|\hat{H}_n^{\mathrm{D}}|, b_i, B) = p(|\hat{H}_n^{\mathrm{D}}| \,|b_i, B)p(b_i, B) = p(|\hat{H}_n^{\mathrm{D}}| \,|b_i, B)p(b_i|B)p(B). \tag{4.73}$$

As $p(B)$ can be omitted from the maximization, it follows

$$\hat{b}_{i,1-\mathrm{D}} = \underset{b_i}{\operatorname{argmax}} \left\{ \underbrace{p(|\hat{H}_n^{\mathrm{D}}| \,|b_i, B)p(b_i|B)}_{\kappa_{2,1-\mathrm{D}}(b_i| \,|\hat{\mathbf{H}}^{\mathrm{D}}|,B)} \kappa_1(\mathbf{r}_n|b_i) \right\}, \tag{4.74}$$

where $\kappa_{2,1-\mathrm{D}}$ represents the a-priori probability under 1-D approximation. In fact, the 1-D approximation considers only the initial iteration in the iterative bit loading process.

4.3.3 2-D approximated MAP classifier

Iterative adjustments are necessary if the initial number of total loaded bits $\bar{B}(l_{\mathrm{it}} = 0) = \sum_{n=0}^{N-1} \bar{b}^{(n)}(l_{\mathrm{it}} = 0)$ is not equal to B. Based on the derivations of 1-D approximation, the less iterations are necessary, the more precise is this approximation. Less iterations imply $\bar{B}(l_{\mathrm{it}} = 0) \approx B$. As given in (2.18), $\bar{B}(l_{\mathrm{it}} = 0)$ is determined by the channel magnitude vector $|\hat{\mathbf{H}}^{\mathrm{U}}| = (|\hat{H}_0^{\mathrm{U}}|, |\hat{H}_1^{\mathrm{U}}|, \cdots, |\hat{H}_{N-1}^{\mathrm{U}}|)^{\mathrm{T}}$. Since wireless channels are subject to strong fluctuations, $\bar{B}(l_{\mathrm{it}} = 0)$ varies inside a relatively large range as well. But B is always a constant determined by the data-rate of the system. Therefore, iterative adjustments are unavoidable. This fact implies that the a-priori probability is more precisely specified if besides the own channel magnitude $|\hat{H}_n^{\mathrm{U}}|$, the magnitudes of other subcarriers are also incorporated.

As described in Section 2.2.1, in each iteration indexed by l_{it}, the AM algorithm assigns a integer number $\bar{b}^{(n)}(l_{\mathrm{it}})$ with $\bar{b}^{(n)}(l_{\mathrm{it}}) \in \mathcal{B}$ to the subcarrier n. So the integer number $\bar{b}^{(n)}(l_{\mathrm{it}})$ undergoes a process described by

$$\bar{b}^{(n)}(0) \rightarrow \bar{b}^{(n)}(1) \cdots \rightarrow \bar{b}^{(n)}(l_{\mathrm{it}}) \cdots \rightarrow \bar{b}^{(n)}(l_{\mathrm{over}}), \tag{4.75}$$

reflecting the iterative adjustment until $\bar{b}^{(n)}(l_{\mathrm{over}}) = b^{(n)}$ (l_{over} is the number of iterations required for AM) or in other words until (4.70) is fulfilled. The number of loaded bits $\bar{b}^{(n)}(l_{\mathrm{it}} + 1)$ in the next iteration is determined by $|\hat{H}_n^{\mathrm{U}}|$ and the difference $\bar{B}(l_{\mathrm{it}}) - B$ as

given in (2.20), where $\bar{B}(l_{\mathrm{it}})$ is the total number of loaded bits in the current iteration. In the computation of $\bar{B}(l_{\mathrm{it}})$, the entire vector $|\hat{\mathbf{H}}^{\mathrm{U}}|$ is taken into account. According to (4.71), each state passed in (4.75) is uniquely determined by the corresponding state in the following process

$$\{|\hat{H}_n^{\mathrm{U}}|, \lambda_{\mathrm{margin}}(0)\} \rightarrow \{|\hat{H}_n^{\mathrm{U}}|, \lambda_{\mathrm{margin}}(1)\} \cdots \rightarrow \cdots \rightarrow \{|\hat{H}_n^{\mathrm{U}}|, \lambda_{\mathrm{margin}}(l_{\mathrm{over}})\}, \quad (4.76)$$

which, based on (2.20), is in turn determined by

$$|\hat{H}_n^{\mathrm{U}}| \rightarrow \{|\hat{H}_n^{\mathrm{U}}|, \bar{B}(0)\} \cdots \rightarrow \{|\hat{H}_n^{\mathrm{U}}|, \bar{B}(l_{\mathrm{it}} - 1)\} \quad (4.77)$$
$$\cdots \rightarrow \{|\hat{H}_n^{\mathrm{U}}|, \bar{B}(l_{\mathrm{over}} - 1)\}.$$

This knowledge implies that the a-priori information, previously given by $|\hat{\mathbf{H}}^{\mathrm{U}}|$ with dimension N, can be now sufficiently provided by a more compact vector

$$Pr = (|\hat{H}_n^{\mathrm{U}}|, \bar{B}(0), \bar{B}(1), \cdots, \bar{B}(l_{\mathrm{over}} - 1)), \quad (4.78)$$

with dimension $l_{\mathrm{over}} + 1$, where l_{over} is the number of required iterations. In other words, for adaptive modulation it is necessary to know the entire vector $|\hat{\mathbf{H}}^{\mathrm{U}}|$, but for modulation classification the information provided in $|\hat{\mathbf{H}}^{\mathrm{U}}|$ is equivalently given by (4.78).

In practical implementations, l_{over} is usually much smaller than N leading to a dimension reduction from N to $l_{\mathrm{over}} + 1$. The approximated MAP classifier is formulated in

$$\hat{b}_{i,l_{\mathrm{over}}+1} = \operatorname*{argmax}_{b_i} \left\{ p(Pr, b_i, B) \prod_{k=0}^{K-1} p(r_{k,n}|\hat{H}_n^{\mathrm{D}}, b_i) \right\}. \quad (4.79)$$

As wireless channels are mostly time-variant so that l_{over} is also time-dependent resulting in a time-dependent dimensionality of the density function $p(Pr, b_i, B)$. On the other side, under some circumstance the number l_{over} can still be large enough to let the evaluation of $p(Pr, b_i, B)$ be very difficult. Investigations have shown that the largest adjustment occurs mostly in the first iteration

$$\bar{b}^{(n)}(l_{\mathrm{it}} = 0) \rightarrow \bar{b}^{(n)}(l_{\mathrm{it}} = 1), \quad (4.80)$$

which is determined by $\{|\hat{H}_n^{\mathrm{D}}|, \bar{B}(0)\}$. In fact, for most subcarriers $\bar{b}^{(n)}(l_{\mathrm{it}})$ does not undergo a large adjustment any more for $l_{\mathrm{it}} \geq 1$. An approach to simplify the implementation is to approximate Pr by

$$Pr_{2-D} = (|\hat{H}_n^{\mathrm{D}}|, \bar{B}(0)). \quad (4.81)$$

Since the dimension is now reduced from N to two, we call it two dimensional (2-D) approximation. Through $(|\hat{H}_n^{\mathrm{D}}|, \bar{B}(0))$, the initial number $\bar{b}^{(n)}(0)$ and the first adjusted

number $\bar{b}^{(n)}(1)$ are uniquely determined. So the *2-D approximation* is justified if it can be assumed that the adjustment carried out in the first iteration dominates the subsequent adjustments, or mathematically formulated

$$|\bar{b}^{(n)}(1) - \bar{b}^{(n)}(0)| \gg |\bar{b}^{(n)}(l_{\text{over}}) - \bar{b}^{(n)}(1)|. \tag{4.82}$$

We observe that $\bar{B}(0)$ is a quantized integer number in $p(|\hat{H}_n^{\mathrm{D}}|, \bar{B}(0), b_i, B)$. To avoid quantization errors and facilitate practical implementations, $\bar{B}(0)$ is in turn approximated by its corresponding continuous-valued $\tilde{B}(0)$ defined in (2.18).Then the *2-D classifier* is given by

$$\hat{b}_{i,2-\mathrm{D}} = \underset{b_i}{\operatorname{argmax}} \left\{ p(|\hat{H}_n^{\mathrm{D}}|, \tilde{B}(0), b_i, B) \kappa_1(\mathbf{r}_n|b_i) \right\}. \tag{4.83}$$

Because of

$$
\begin{aligned}
p(|\hat{H}_n^{\mathrm{D}}|, \tilde{B}(0), b_i, B) &= p(|\hat{H}_n^{\mathrm{D}}| \,|\tilde{B}(0), b_i, B) p(\tilde{B}(0), b_i, B) \\
&= p(|\hat{H}_n^{\mathrm{D}}| \,|\tilde{B}(0), b_i, B) p(b_i|\tilde{B}(0), B) p(\tilde{B}(0), B),
\end{aligned}
\tag{4.84}
$$

omitting the irrelevant term $p(\tilde{B}(0), B)$, (4.83) is equivalent to

$$\hat{b}_{i,2-\mathrm{D}} = \underset{b_i}{\operatorname{argmax}} \left\{ \underbrace{p(|\hat{H}_n^{\mathrm{D}}| \,|\tilde{B}(0), b_i, B) p(b_i|\tilde{B}(0), B)}_{\kappa_{2,2-\mathrm{D}}(b_i| |\hat{\mathbf{H}}^{\mathrm{D}}|, B)} \kappa_1(\mathbf{r}_n|b_i) \right\}, \tag{4.85}$$

where $\kappa_{2,2-\mathrm{D}}$ represents the a-priori probability under *2-D approximation*. In fact, the 2-D approximation considers only the first two iterations.

4.3.4   3-D approximated MAP classifier

As indicated in (2.18), the evaluation of $\tilde{B}(0)$ included in $\kappa_{2,2-\mathrm{D}}$ requires N times logarithm operations at the receiver. These operations are very computationally intensive which can significantly delay the classification process. Especially, in delay-sensitive applications such operations have to be avoided as much as possible. An approach to solve this problem is to expand $\tilde{B}(0)$ by Taylor's series and truncate the expansion after the second order. As shown in Appendix G, the second order approximation of $\tilde{B}(0)$ is uniquely determined by two parameters: mean squared channel amplitude

$$\mu_{\mathrm{H}}^2 = \frac{1}{N} \sum_{n=0}^{N-1} |\hat{H}_n^{\mathrm{D}}|^2, \tag{4.86}$$

giving the average power gain of the channel and mean squared channel variance

$$\sigma_{\mathrm{H}}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (|\hat{H}_n^{\mathrm{D}}|^2 - \mu_{\mathrm{H}}^2)^2, \tag{4.87}$$

characterizing the level of frequency selectivity of the channel. Then (4.83) is approximated by

$$\hat{b}_{i,3-\mathrm{D}}^{(n)} = \operatorname*{argmax}_{b_i} \left\{ p(|\hat{H}_n^{\mathrm{D}}|, \mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, b_i, B) \kappa_1(\mathbf{r}_n|b_i) \right\}. \tag{4.88}$$

In this approximation, the dimension is reduced from N to three, so we call this approximation *3-D approximation*. Due to

$$
\begin{aligned}
p(|\hat{H}_n^{\mathrm{D}}|, \mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, b_i, B) &= p(|\hat{H}_n^{\mathrm{D}}| \,| \mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, b_i, B) p(\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, b_i, B) \\
&= p(|\hat{H}_n^{\mathrm{D}}| \,| \mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, b_i, B) p(b_i|\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, B) p(\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, B), \quad (4.89)
\end{aligned}
$$

by omitting $p(\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, B)$, the *3-D classifier* is given by

$$\hat{b}_{i,3-\mathrm{D}}^{(n)} = \operatorname*{argmax}_{b_i} \left\{ \underbrace{p(|\hat{H}_n^{\mathrm{D}}| \,| \mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, b_i, B) p(b_i|\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, B)}_{\kappa_{2,3-\mathrm{D}}(b_i|\,|\hat{\mathbf{H}}^{\mathrm{D}}|, B)} \kappa_1(\mathbf{r}_n|b_i) \right\}. \tag{4.90}$$

where $\kappa_{2,3-\mathrm{D}}$ denotes the a-priori probability under *3-D* approximation. The computational complexity is reduced from N logarithm operations to N multiplications and $2N$ additions with the price of one additional dimension. However, $p(|\hat{H}_n^{\mathrm{D}}| \,| \mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, b_i, B)$ and $p(b_i|\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, B)$ can be measured or simulated beforehand and stored in a look-up table. As soon as the current $|\hat{H}_n^{\mathrm{D}}|$, μ_{H}^2 and σ_{H}^2 are available, $\kappa_{2,3-\mathrm{D}}(b_i|\,|\hat{\mathbf{H}}^{\mathrm{D}}|, B)$ can be determined based on the values called from the corresponding look-up tables.

4.3.5 Determination of the prior probabilities

The prior probabilities $\kappa_{2,1-\mathrm{D}}(b_i|\,|\hat{\mathbf{H}}^{\mathrm{D}}|, B)$, $\kappa_{2,2-\mathrm{D}}(b_i|\,|\hat{\mathbf{H}}^{\mathrm{D}}|, B)$ and $\kappa_{2,3-\mathrm{D}}(b_i|\,|\hat{\mathbf{H}}^{\mathrm{D}}|, B)$ have to be determined in practice through extensive field measurements or extensive computer simulations in an off-line way. These probabilities are then stored in a look-up table for real-time read-out.

Table 4.3: $B = 192$, 4 bits/subcarrier on average.

b_i	0	1	2	4	6
$p(b_i|B)$	0.027	0.026	0.186	0.481	0.280
$p(b_i|\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, B)$	0.023	0.022	0.176	0.522	0.257

In this thesis, computer simulations are performed to determine these probabilities. The Matlab built-in function *ksdensity* is used to graphically approximate the involved conditional PDFs based on a sufficiently large number of typical channel realizations. For the considered $B = 192$ bits (corresponding to a data-rate of $r_{\mathrm{D}} = 24$ Mpbs), $p(|\hat{H}_n^{\mathrm{D}}| \,| b_i, B)$
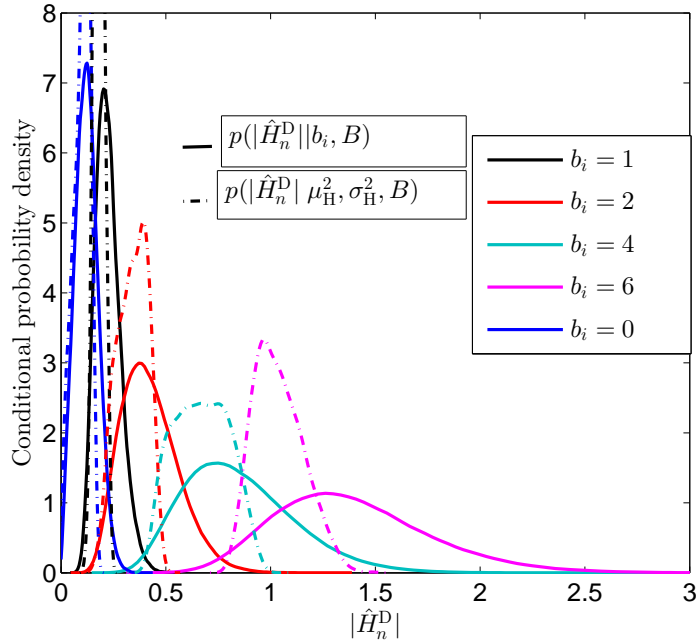
Figure 4.6: $p(|\hat{H}_n^{\mathrm{D}}||b_i, B)$ and $p(|\hat{H}_n^{\mathrm{D}}||\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, b_i, B)$ with $B = 192$

is plotted in Figure 4.6 for the set $\{b_i\}_{i=1}^{I=5}$. Note that the channel gains $|\hat{H}_n^{\mathrm{D}}|$ are normalized here with $\mathrm{E}\{|\hat{H}_n^{\mathrm{D}}|^2\} = 1$ just as the signal power stated before. A certain value $p(|\hat{H}_{n'}^{\mathrm{D}}||b_j, B)$, for certain $|\hat{H}_{n'}^{\mathrm{D}}|$ and b_j, is then estimated by inserting $|\hat{H}_{n'}^{\mathrm{D}}|$ into the graphic of $p(|\hat{H}_n^{\mathrm{D}}||b_j, B)$ belonging to b_j (Figure 4.6), which is obtained via the function *ksdensity* based on the whole set of channel realizations. Whereas, a certain value $p(|\hat{H}_{n'}^{\mathrm{D}}||\tilde{B}(0), b_j, B)$ is estimated by inserting $|\hat{H}_{n'}^{\mathrm{D}}|$ into the graphic of $p(|\hat{H}_n^{\mathrm{D}}||\tilde{B}(0), b_j, B)$ corresponding to b_j, which is obtained by considering only a subset of channel realizations. This subset includes those channel realizations which result in $\tilde{B}(0)$ in this quantized interval $[\tilde{B}(0) - \Delta\tilde{B}, \tilde{B}(0) + \Delta\tilde{B}]$ (quantization of continous-valued $\tilde{B}(0)$ for PDF). Here, $\Delta\tilde{B}$ is the quantization step (or resolution for determining the PDF with limited number of data samples). This parameter $\Delta\tilde{B}$ is designed so that a good trade-off between estimation precision and computational effort is achieved. Here the resolution is set to $\Delta\tilde{B} = 5$ $\frac{\mathrm{bit}}{\mathrm{s}\cdot f_{\mathrm{BW}}}$, where f_{BW} is the total bandwidth of the considered OFDM system. Compared to the 1-D approximation, in the 2-D approximation only the channel realizations in the beforehand fixed interval (subset) are used to generate the graphics $p(|\hat{H}_n^{\mathrm{D}}||\tilde{B}(0), b_i, B)$. By the same means $p(|\hat{H}_n^{\mathrm{D}}||\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, b_i, B)$ can be also approximated with the two channel resolution parameters $\Delta\mu_{\mathrm{H}}^2 = 0.05$ and $\Delta\sigma_{\mathrm{H}}^2 = 0.015$. The corresponding graphics are also depicted in Figure 4.6 for $0.5 \leq \mu_{\mathrm{H}}^2 \leq 0.55$ and $0.14 \leq \sigma_{\mathrm{H}}^2 \leq 0.155$. A clear shift can be observed in the corresponding profiles indicating the inaccuracy of the 1-D approximation.

Similarly, the conditional probabilities $p(b_i|B)$, $p(b_i|\tilde{B}(0), B)$ and $p(b_i|\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, B)$ can be estimated. Table 4.3 provides the estimates of $p(b_i|B)$ and $p(b_i|\mu_{\mathrm{H}}^2, \sigma_{\mathrm{H}}^2, B)$ for the same μ_{H}^2 and σ_{H}^2 intervals as above.

4.3.6   Heuristic MAP classifier

The previously proposed MAP classifiers utilize the DL channel estimates $\hat{\mathbf{H}}^{\mathrm{D}}$ and the applied data rate r_{D} as the a-priori information. The application of the same bit loading algorithm in the UL transmission is not necessary and mostly also not possible, since BS and MS have to meet different system requirements. Especially, it is not required that the mobile station has to perform the computationally intensive calculation of the bit loading vector $\mathbf{b}$, too. If, however, the system is designed to realize a symmetric adaptive modulation in both transmission directions and the mobile station is powerful enough to perform real-time calculations of $\mathbf{b}$, a heuristic MAP (HMAP) classifier can be used which has been proven to be able to significantly improve the classification performance compared to classifiers without using any prior information [9].
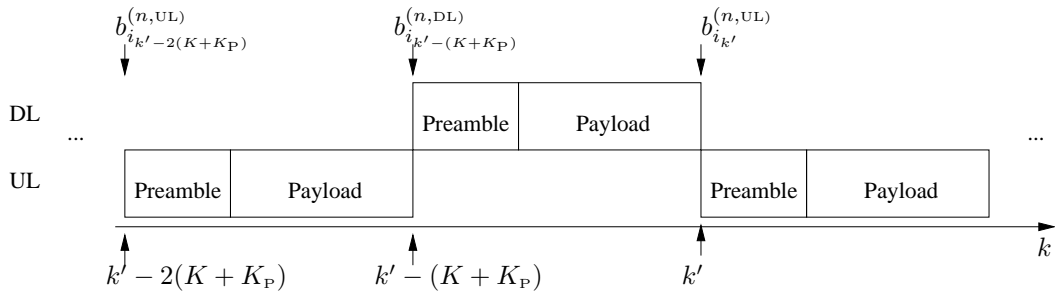
Figure 4.7:   The principle of heuristic MAP approach in a TDD system

In Figure 4.7, $b_{i_{k'}}^{(n,\mathrm{UL})}$ and $b_{i_{k'-(K+K_{\mathrm{P}})}}^{(n,\mathrm{DL})}$ denote the adapted modulation level on subcarrier n, for the current UL at time slot k' and the last DL at $k' - (K + K_{\mathrm{p}})$ with $K + K_{\mathrm{p}} = K_{\mathrm{F}}$ the packet duration, respectively. Here the duration of the preambles is denoted as K_{p}. Suppose that the MS is now at time slot k' and intends to demodulate the incoming DL packet of $k' - (K + K_{\mathrm{p}})$. It has to first detect the applied modulation level $b_{i_{k'-(K+K_{\mathrm{p}})}}^{(n,\mathrm{DL})}$ by maximizing the a posteriori probability

$$\hat{b}_{i_{k'-(K+K_{\mathrm{p}})}}^{(n,\mathrm{DL})} = \underset{b_i}{\operatorname{argmax}} \left\{ \kappa_2 \left(b_i | \, |\hat{\mathbf{H}}_{k'-(K+K_{\mathrm{p}})}^{\mathrm{D}}|, B \right) \kappa_1(\mathbf{r}|b_i) \right\}. \tag{4.91}$$

The idea now is to heuristically approximate the a-priori probability in (4.91) by a quantized conditional probability

$$\kappa_2 \left(b_i | \, |\hat{\mathbf{H}}_{k'-(K+K_{\mathrm{p}})}^{\mathrm{D}}|, B \right) \approx \kappa_2'(b_i | b_{i_{k'}}^{(n,\mathrm{UL})}). \tag{4.92}$$

In a TDD system, if the system operates in a slowly time-variant channel environment, the modulation level to be classified $b_{i_{k'-(K+K_{\mathrm{p}})}}^{(n,\mathrm{DL})}$ for the last DL is much more likely the one $b_{i_{k'}}^{(n,\mathrm{UL})}$ being used for the next UL. It is therefore reasonable to assign

$$\kappa_2'(b_i|b_{i_{k'}}^{(n,\mathrm{UL})}) > \kappa_2'(b_j|b_{i_{k'}}^{(n,\mathrm{UL})}) \ \text{ for } \ b_j \neq b_i = b_{i_{k'}}^{(n,\mathrm{UL})}, \tag{4.93}$$

with

$$\sum_{i=1}^{I} \kappa_2'(b_i|b_{i_{k'}}^{(n,\mathrm{UL})}) = 1. \tag{4.94}$$

In [9], the following assignment was proposed

$$\kappa_2'(b_i|b_{i_{k'}}^{(n,\mathrm{UL})}) = \begin{cases} P_{b_i|b_{i_{k'}}^{(n,\mathrm{UL})}} & : \text{if } b_i = b_{i_{k'}}^{(n,\mathrm{UL})} \\ \dfrac{1-P_{b_i|b_{i_{k'}}^{(n,\mathrm{UL})}}}{I-1} & : \text{otherwise}, \end{cases} \tag{4.95}$$

where $P_{b_i|b_{i_{k'}}^{(n,\mathrm{UL})}}$ is a design parameter. For slowly time-variant channels, it is reasonable to assign

$$P_{b_i|b_{i_{k'}}^{(n,\mathrm{UL})}} \gg \frac{1 - P_{b_i|b_{i_{k'}}^{(n,\mathrm{UL})}}}{I-1}, \tag{4.96}$$

or in other words

$$P_{b_i|b_{i_{k'}}^{(n,\mathrm{UL})}} \gg \frac{1}{I-1}, \tag{4.97}$$

with I the number of modulation candidates.

Obviously, the classification performance depends on the value of $P_{b_i|b_{i_{k'}}^{(n,\mathrm{UL})}}$. An analytical derivation of the optimal assignment is impossible. In the practical implementation, an extensive computer simulation is an option to find a meaningful assignment. Investigations have shown that it suffices to limit $P_{b_i|b_{i_{k'}}^{(n,\mathrm{UL})}}$ to a certain range

$$P_{\mathrm{lower}} < P_{b_i|b_{i_{k'}}^{(n,\mathrm{UL})}} < P_{\mathrm{upper}} \ \text{ with } \ \frac{1}{I} < P_{\mathrm{lower}} < P_{\mathrm{upper}} < 1. \tag{4.98}$$

Inside this range, the classification performance is insensitive to the exact value of $P_{b_i|b_{i_{k'}}^{(n,\mathrm{UL})}}$.

The approximation in (4.92) can be interpreted as a mapping of the continuous valued $\hat{\mathbf{H}}_{k'-(K+K_{\mathrm{p}})}^{\mathrm{D}}$ describing the a-priori information to a quantized bit level $b_{i_{k'}}^{(n,\mathrm{UL})}$ with $b_{i_{k'}}^{(n,\mathrm{UL})} \in \mathcal{B}$ under the B-constraint. Certainly, this quantization process results in irreversable

quantization errors. Furthermore, the assignment of identical value to $\kappa_2'(b_i|b_{i_{k'}}^{(n,\mathrm{UL})})$ for $b_i \neq b_{i_{k'}}^{(n,\mathrm{UL})}$ in (4.95) is a too rough simplification, as in reality it follows

$$\kappa_2'(b_i|b_{i_{k'}}^{(n,\mathrm{UL})}) \geq \kappa_2'(b_j|b_{i_{k'}}^{(n,\mathrm{UL})}) \quad \text{for} \quad |b_i - b_{i_{k'}}^{(n,\mathrm{UL})}| > |b_j - b_{i_{k'}}^{(n,\mathrm{UL})}|. \tag{4.99}$$

It means that the modulation levels from the neighbourhood of $b_{i_{k'}}^{(n,\mathrm{UL})}$ are more likely adapted than those who are far away from $b_{i_{k'}}^{(n,\mathrm{UL})}$. This is justified by slow time-variance of the radio channel and a short packet size. In [45], a continuous Gaussian distributed value $\kappa_2'(b_{i,\mathrm{con}}|b_{i_{k'}}^{(n,\mathrm{UL})})$ with the mean value at $b_{i_{k'}}^{(n,\mathrm{UL})}$ and a variance of σ_{G}^2 is proposed to interpolate the discrete-valued function $\kappa_2'(b_i|b_{i_{k'}}^{(n,\mathrm{UL})})$. This proposal assumes first a continuous modulation level $b_{i,\mathrm{con}}$. The desired discrete value $\kappa_2'(b_i|b_{i_{k'}}^{(n,\mathrm{UL})})$ is obtained by sampling $\kappa_2'(b_{i,\mathrm{con}}|b_{i_{k'}}^{(n,\mathrm{UL})})$ at $b_{i,\mathrm{con}} = b_i$. As $\kappa_2'(b_{i,\mathrm{con}}|b_{i_{k'}}^{(n,\mathrm{UL})})$ allows continuous value domain $b_{i,\mathrm{con}}$, the quantization effect is reduced. Moreover, the Gaussian distribution guarantees also the satisfaction of (4.99). But the approximation continues to have the problem of no analytical expressions for the design parameter σ_{G}^2. Therefore, we will not pursue this approach in more detail.

4.3.7   MAC based on symmetric and bi-directional bit loading

In Section 4.3.6, it is observed that, in a TDD transmission system with a reciprocal radio channel assumed, the classification performance can be improved if the MS applies the same bit loading algorithm on the estimated channel and the resulting bit loading vector is incorporated into the classification algorithm as a priori information. A meaningful use of this a priori information requires that the BS and MS apply the same bit loading algorithm for the DL and UL, respectively (symmetric and bidirectional bit loading).

If the channel is reciprocal and the same bit loading algorithm is used in both UL and DL direction, due to the slow time-variance of the channel as assumed, the question arises, what performance can be expected if the MS performs the classification solely based on the DL channel estimate. The MS uses a powerful channel estimator and applies the same bit loading algorithm on the estimated channel. The calculated bit loading vector for the upcoming UL is used as the classified bit loading vector for the last DL:

$$\hat{b}^{(n)} = \hat{b}_{\hat{\mathbf{H}}}^{(n)} \quad \text{for } n = 0, 1, \cdots, N-1, \tag{4.100}$$

where $\hat{b}^{(n)}$ is the classified modulation level and $\hat{b}_{\hat{\mathbf{H}}}^{(n)}$ is the calculated modulation level resulting from the application of the same bit loading algorithm on the channel estimate $\hat{\mathbf{H}}$.

Computer simulations are conducted to evaluate the classification performance of the above classifier for the adaptive OFDM system given in Appendix A. A short packet
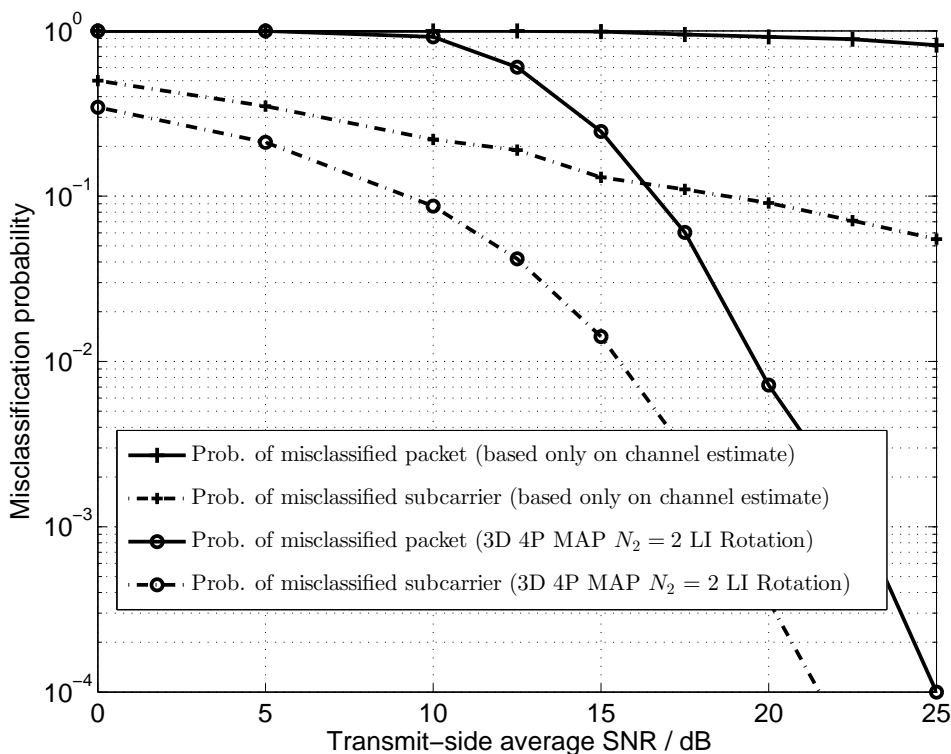
Figure 4.8: Performance of the classifier based only on the channel estimate

of 10 OFDM payload symbols is considered to ensure the low channel variance between successive packets. The channel estimation is based on a minimum mean squared error (MMSE) estimator. Additionally, a Wiener-filter based interpolation is used to further improve the estimation precision by utilizing the statistical properties of the used channel and noise model. As shown in Figure 4.8, compared to the classifier 3-D MAP 4-P $N_2 = 2$ LI Rotation, this classifier shows a significant degradation in both probability of misclassified packet and subcarrier.

Table 4.4: Transition probabilities $P_{Z_{n_s}|Z_{n_s-1}}$ based on the perfect channel estimation.

Z_{n_s-1} \ Z_{n_s}	0	1	2	4	6
0	0.8976	0.092	0.0103	0	0
1	0.0976	0.7516	0.150	0	0
2	0.0015	0.0224	0.9417	0.0342	0
4	0	0	0.0134	0.9724	0.0141
6	0	0	0	0.0235	0.9764

Actually, this poor performance could have been expected from the previous results given

in Section 3.3.1, where the transition probabilities of the modulation levels between successive packets were simulated. The transition probabilities are given in Table 4.4 for perfect channel estimator and a packet length of 10 OFDM payload symbols again. Even with the highest transition probability $P_{6\to 6} = 0.9764$ for the transition from modulation level 6 to 6, the probability, that a packet error occurs, is estimated to:

$$P_{\mathrm{F}} = 1 - P_{6\to 6}^{N} = 1 - 0.9764^{48} = 0.6822, \tag{4.101}$$

where N is the number of used data subcarriers and $P_{b_i \to b_i}^{N}$, with b_i as a modulation level, is the probability that the bit loading vector calculated for the upcoming transmission based on the perfect channel estimator is identical to that of the last received packet. This result indicates that the classifier based solely on the channel estimate is unsuitable for classification problem formulated in this thesis. Furthermore, classifiers of this kind are even practically unfeasible due to following reasons:

- This classifier utilizes only the channel estimate for the classification. The information from the received signal is totally ignored. This leads to a strong dependence of the classification performance on the channel estimation precision. A precise channel estimation requires a certain SNR value, which is not always given on any subcarrier and in any time slot.

- Powerful channel estimator usually utilizes statistical properties of the used channel model, which, in case of a model mismatch would cause a high estimation error.

- In most practical applications, due to different system requirements in the MS and BS, different bit loading algorithms are used for downlink and uplink. So, the assumption that both stations use the same algorithm is not always given.

Due to the poor performance and the infeasibility in practical systems, this classifier will not be further pursued in this thesis. However, if the used bit loading algorithm is designed such that the modulation levels switch not so sensitively to the time variation of the radio channel and the computational efforts of the bit loading algorithm can be reduced to be implemented in real-time applications, this classifier can draw again interest because the classifier requires only the channel estimates, which are available anyhow at the MS for the channel equalization. So, it would be a task of joint optimization of the bit loading algorithm at the BS and the modulation classification at the MS.

4.3.8 Simulation results

Computer simulations are conducted to evaluate the classification performance of the proposed MAP classifiers for the adaptive OFDM system given in Appendix A. Performance is given in terms of probability of subcarrier misclassification P_{s} as defined before.
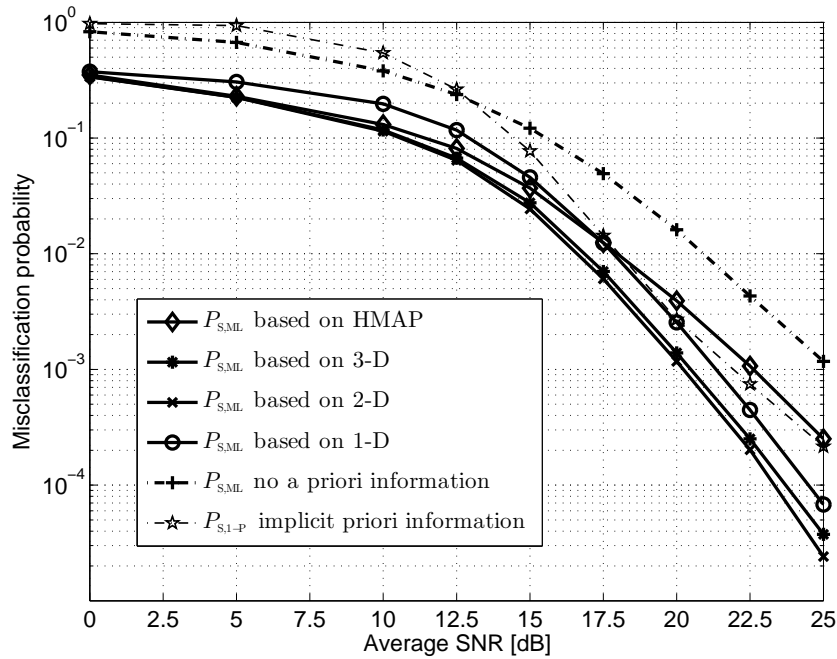
Figure 4.9: P_S of different MAP classifiers based on ML function and the 1-P approximated classifier

Figure 4.9 shows P_S using no approximated $\kappa_1(\mathbf{r}_n|b_i)$, namely the full ML function given in (4.9), while the a-priori probability $\kappa_2(b_i|\,|\hat{\mathbf{H}}^\mathrm{D}|, B)$ is approximated by the proposed approaches (1-D, 2-D, 3-D and HMAP). The combination of the ML classifier with the 1-D, 2-D, 3-D or HMAP based approximation of the a-priori probability builds an approximated MAP-based ML classifier and is denoted as 1-D, 2-D or 3-D MAP ML classifier or HMAP ML classifier, respectively. The 2-D MAP ML classifier gives the best performance, which is closely followed by the 3-D MAP ML classifier. It verifies an efficient utilization of the available prior knowledge by both classifiers. The 1-D MAP ML classifier shows 1 dB degradation for $P_\mathrm{S} = 10^{-3}$. Another $1 \ldots 1.5$ dB is required for the HMAP classifier to achieve the same performance. Finally, the pure ML classifier delivers by far the worst performance, which underlines the importance of incorporating the prior knowledge into the classification strategy. As a comparison, the performance of the 1-P approximation is also provided (dashed thin line) as it implicitly utilizes certain prior knowledge. The performance is at high SNR comparable to that of HMAP.

In the following, we introduce the abbreviations 2-D MAP ML, 2-D MAP 4-P, 2-D MAP 1-P and 2-D MAP ED for 2-D approximated a-priori probability $\kappa_2(b_i|\,|\mathbf{H}^\mathrm{D}|, B)$ in combination with ML, 4-P, 1-P and ED approximated likelihood function $\kappa_1(\mathbf{r}_n|b_i)$. Correspondingly 3-D based MAP classifiers are abbreviated by 3-D MAP ML, 3-D MAP 4-P, 3-D MAP 1-P and 3-D MAP ED. Figure 4.10 gives the classification performance for 2-D
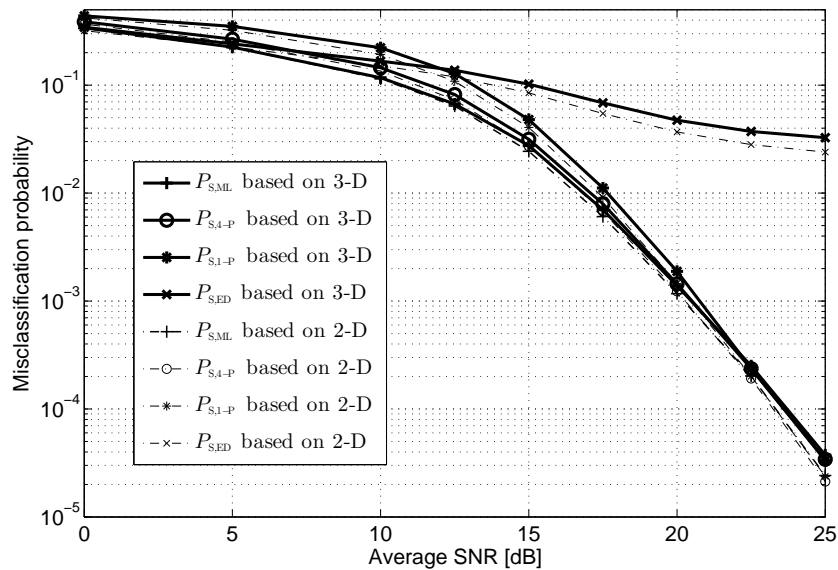
Figure 4.10: P_{S} of 2-D and 3-D based classifiers using different approximations of the ML function

(thin, dashed) and 3-D (thick, solid) based MAP classifiers with mentioned combinations. The 2-D based MAP classifiers provide in any case a slightly better performance than their 3-D correspondences. Due to the nested signal constellations, the ED approximation continues to show a very poor performance even under the use of the prior knowledge. For both 2-D and 3-D approximations, the 4-P based classifier provides almost the same classification robustness as the ML-based classifier, while the 1-P based classifiers show some degradation especially in the low SNR range.

Based on the observations in Figures 4.9 and 4.10, we conclude that 3-D MAP 4-P classifier provides quite robust classification performance. The performance is comparable to that of the 2-D MAP ML classifier which shows the best performance among all proposals. At the same time the 3-D MAP 4-P classifier eliminates the computation of N exponential functions required for the 2-D approximation and the $K \sum_{i=1}^{I} 2^{b_i}$ exponential functions required for the likelihood function. Therefore, it maintains a manageable implementation complexity. In the sequel, we will use this classifier to approximate the optimal classifier provided by (4.68).

4.4 Featured-based AMC

Although LB AMC is able to achieve the optimal classification performance in the case that no prior information is available in the classifier, it suffers from high implementation complexity in general. FB AMC draws increasing research interest recently as it can be constructed to be very computationally efficient and more resistant against mismatches

in statistical models. FB AMC extracts a feature set $\mathcal{F} = \{f_{l'}\}_{l'=1}^{F}$ from the received signal which is suitable to discriminate between modulation candidates with a manageable complexity, where $f_{l'}$ is the l'-th extracted feature and F is the number of used features. We concentrate on classification rules applicable for QAM signals. The problem is now to recognize modulation levels (orders) inside one modulation class. Investigations have shown that many FB AMC algorithms with suboptimal performasnce appear to be results of approximations of the LB AMC framework: q_p-type classifiers suitable for classification of QAM signals proposed in [75] and Euclidean distance approximation-based classifier [2], just to mention some.

Compared to LB AMC, where modulation classification is a multiple hypothesis testing problem, FB AMC is usually organized in a hierarchical structure also known as binary decision tree. The modulation level set $\{b_i\}_{i=1}^{I}$ is successively divided into binary subsets (nodes in the tree) $\{b_{i_1}\}_{i_1 \in \mathcal{I}_1}$ and $\{b_{i_2}\}_{i_2 \in \mathcal{I}_2}$ with

$$\mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset \quad \text{and} \quad \mathcal{I}_1 \cup \mathcal{I}_2 \in \{1, 2, \cdots, I\}, \tag{4.102}$$

till all nodes in the tree are occupied by a single modulation level, i.e. number of elements in subsets is one. For each branch, a feature set has to be established as a binary decision rule

$$\mathcal{F} \underset{\mathcal{I}_2}{\overset{\mathcal{I}_1}{\gtrless}} \mathcal{F}_{\text{th}}, \tag{4.103}$$

where $\mathcal{F}_{\text{th}}$ denotes the set of predefined thresholds. Ideally, for each decision layer $\mathcal{F}$ and $\mathcal{F}_{\text{th}}$ has to be optimized to distinguish between the involved two subsets. So, in worst case $I - 1$ decision rules are required.

4.4.1   Higher order moment-based AMC

The idea of higher order moment-based classifier is that, through nonlinear operations, the constellation of each modulation candidate is mapped to a characteristic image, which is then projected to a feature set $\mathcal{F}_{\text{mom}}$ by building statistical moments. Mostly, for the sake of practical implementation this feature set is a scalar (moment) $\mathcal{F}_{\text{mom}} = \{\mu\}$. This scalar can serve as the discriminating parameter (feature). For this purpose, we define

$$\mu_{n_{\text{m}}, n'_{\text{m}}}(\mathbf{x}) = \text{E}[\mathbf{x}^{n_{\text{m}} - n'_{\text{m}}}(\mathbf{x}^*)^{n'_{\text{m}}}], \quad \text{with } n'_{\text{m}} \leq n_{\text{m}} \tag{4.104}$$

as the n_{m}-th order mixed moment with the n'_{m}-th order conjugation of a complex-valued stationary random process $\mathbf{x}$ [76]. For any modulation format A_i, the theoretical moments (noise-free and asymptotic value) are given by

$$\mu_{n_{\text{m}}, n'_{\text{m}}}(A_i) = \text{E}[a_{i,l}^{n_{\text{m}} - n'_{\text{m}}}(a_{i,l}^*)^{n'_{\text{m}}}] = \frac{1}{M_i} \sum_{l=1}^{M_i} a_{i,l}^{n_{\text{m}} - n'_{\text{m}}}(a_{i,l}^*)^{n'_{\text{m}}}. \tag{4.105}$$
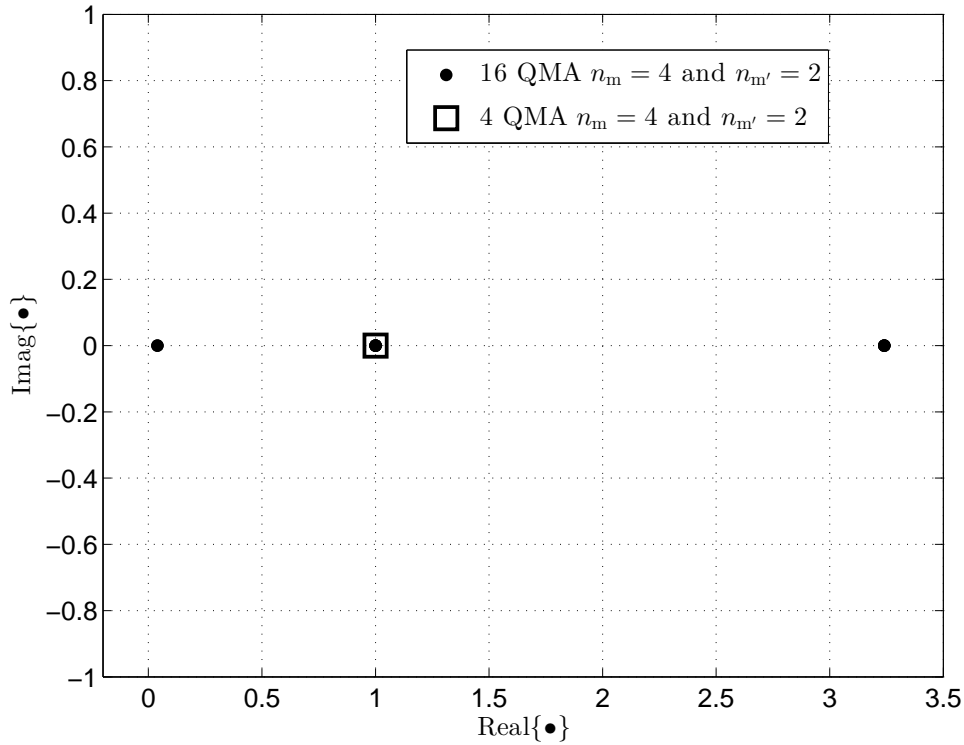
Figure 4.11:   Characteristic image of 4 and 16 QAM for $n_{\mathrm{m}} = 4$ and $n'_{\mathrm{m}} = 2$

For $n_{\mathrm{m}} = 4$ and $n'_{\mathrm{m}} = 2$, the nonlinear operation $\{|a_{i,l}|^4\}_{l=1}^{M_i}$ maps 4-QAM and 16-QAM to their specific images as in Figure 4.11, which shows the respective amplitude levels to the power 4. Then take the expectation value of the mapped constellation points resulting in

$$\mu_{4,2}(A_3) = \mathrm{E}[|a_{3,l}|^4] = 1, \tag{4.106}$$

for 4-QAM and

$$\mu_{4,2}(A_4) = \mathrm{E}[|a_{i,l}|^4] = 1.32, \tag{4.107}$$

for 16-QAM. Using (4.105), the theoretical mixed moments of order till $n_{\mathrm{m}} = 8$ is tabulated in Appdendix F for considered candidate set $\{A_i\}_{i=1}^{I}$. Based on this table, $\mu_{2,0}$ is able to discriminate between the two subsets

$$\mathcal{I}_1 = \{i_1 = 2\} \quad \text{and} \quad \mathcal{I}_2 = \{i_2 = 1, 3, 4, 5\}, \tag{4.108}$$

because of

$$\mu_{2,0}(A_{i_1}) = 1 \quad \text{and} \quad \mu_{2,0}(A_{i_2}) = 0. \tag{4.109}$$

A possible hierarchical classifier could be constructed as in Figure 4.12. At each layer $\tilde{l}$, a binary decision is made according to the decision rule

$$\hat{\mu}_{n_{\mathrm{m}}, n'_{\mathrm{m}}} \overset{\mathcal{I}_1}{\underset{\mathcal{I}_2}{\gtrless}} \mu_{\mathrm{th}, \tilde{l}}, \tag{4.110}$$

where $\mu_{n_{\mathrm{m}},n'_{\mathrm{m}}}$ is the selected discriminating feature, $\hat{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}$ is the feature estimate of the received data sequence $\mathbf{r}$ and $\mu_{\mathrm{th},\tilde{l}}$ is the threshold value, respectively. The feature is estimated by

$$\hat{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}} = \frac{1}{K} \sum_{k=0}^{K-1} \hat{r}_k^{\,n_{\mathrm{m}}-n'_{\mathrm{m}}} \left(\hat{r}_k^*\right)^{n'_{\mathrm{m}}}, \tag{4.111}$$

with $\hat{r}_k = \frac{H^{\mathrm{D}}}{\hat{H}^{\mathrm{D}}} s_k + \frac{n_k}{\hat{H}^{\mathrm{D}}}$ as the equalized symbol. In general, the optimal value of $\mu_{\mathrm{th},\tilde{l}}$ is
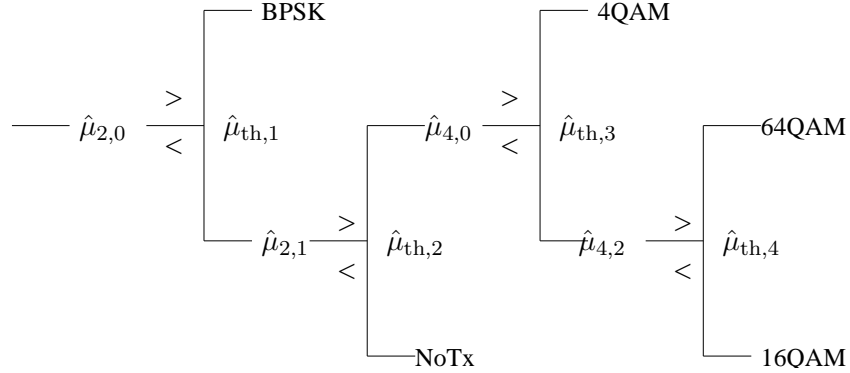
Figure 4.12: A possible binary classification tree based on moments

prohibitively difficult to determine as the PDF of $\hat{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}$ is not available. The following properties are needed to simplify this problem. The principle thereof is illustrated in case of a binary decision between subsets with single element $\mathcal{I}_{s,\tilde{l}} = \{i_{s,\tilde{l}}\}$ for $s = 1, 2$. The extension to subsets with multiple elements requires more sophisticated mathematical derivations and is partially provided in [78] and [79].

Property 1: Under the assumption of statistically independent r_k and sufficiently long observation length K, central limit theorem can be used to approximate the PDF of the feature [78]

$$\hat{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}} \sim \mathcal{N}\left(\tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)}, \tilde{\sigma}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)}\right). \tag{4.112}$$

In (4.112), $\tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)}$ and $\tilde{\sigma}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)}$ are the mean and variance of the Gaussian distributed random variable $\hat{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}$, respectively and the overscript (i) accounts for the dependence of these parameters on the underlying modulation formats A_i.

Property 2: If it holds $\tilde{\sigma}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i_1)} \approx \tilde{\sigma}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i_2)}$ and $\tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i_1)} < \tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i_2)}$, minimum misclassification probability is provided by a threshold detector under the assumption of equally likely subsets [80].

$$\hat{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}(\mathbf{r}_n) \underset{\mathcal{I}_2}{\overset{\mathcal{I}_1}{\gtrless}} \mu_{\mathrm{th}}, \tag{4.113}$$

where the optimal threshold is

$$\mu_{\text{th,opt}} = \frac{\tilde{\mu}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)} + \tilde{\mu}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)}}{2}. \tag{4.114}$$

Property 3: If instead $\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)} < \tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)}$, the optimal detector is given by an interval detector: [78]

$$\text{Decide the subset for } \mathcal{I}_{1,\tilde{l}} = \{i_{1,\tilde{l}}\}, \quad \text{if } \hat{\mu}_{n_{\text{m}},n'_{\text{m}}}(\mathbf{r}_n) \in [\alpha_{\text{m}} - \beta_{\text{m}} \; \alpha_{\text{m}} + \beta_{\text{m}}], \tag{4.115}$$

with

$$\alpha_{\text{m}} = \left(\frac{\tilde{\mu}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)}}{\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)}} - \frac{\tilde{\mu}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)}}{\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)}} \right) \frac{\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)} \tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)}}{\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)} - \tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)}}, \tag{4.116}$$

and

$$\beta_{\text{m}} = \frac{\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)} \tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)}}{\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)} - \tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)}} \left[\ln \left(\frac{\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)}}{\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)}} \right) + \frac{(\tilde{\mu}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)} - \tilde{\mu}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)})^2}{\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_2)} - \tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i_1)}} \right]. \tag{4.117}$$

So the problem of setting μ_{th} is reduced to determine $\tilde{\mu}_{n_{\text{m}},n'_{\text{m}}}^{(i)}$ and $\tilde{\sigma}_{n_{\text{m}},n'_{\text{m}}}^{(i)}$ with

$$\tilde{\mu}_{n_{\text{m}},n'_{\text{m}}}^{(i)} = \text{E}[\hat{\mu}_{n_{\text{m}},n'_{\text{m}}}] = \text{E}[\frac{1}{K} \sum_{k=0}^{K-1} \hat{r}_k^{n_{\text{m}}-n'_{\text{m}}} (\hat{r}_k^*)^{n'_{\text{m}}}] = \frac{1}{K} \sum_{k=0}^{K-1} \text{E}[\hat{r}_k^{n_{\text{m}}-n'_{\text{m}}} (\hat{r}_k^*)^{n'_{\text{m}}}] \tag{4.118}$$

$$= \frac{1}{K} \sum_{k=0}^{K-1} \int \hat{r}_k^{n_{\text{m}}-n'_{\text{m}}} (\hat{r}_k^*)^{n'_{\text{m}}} p(r_k|b_i) \mathrm{d}r_k,$$

where $p(r_k|b_i)$ is defined in (4.7) and the integration has to be evaluated numerically over the whole complex plane. For low orders n_{m}, an alternative can be used to avoid the two-dimensional integration. We assume perfect channel knowledge for this moment and illustrate the computation in case of $\tilde{\mu}_{4,2}^{(i)}$. So H^{D} is a deterministic value in $\hat{r}_k = s_k + \frac{n_k}{H^{\text{D}}}$, with

$$\hat{r}_k^2 (\hat{r}_k^*)^2 = |s_k|^4 + \left| \frac{n_k}{H^{\text{D}}} \right|^4 + 4|s_k|^2 \left| \frac{n_k}{H^{\text{D}}} \right|^2 \tag{4.119}$$

$$+ 4|s_k|^2 \Re \left\{ s_k \left(\frac{n_k}{H^{\text{D}}} \right)^* \right\} \quad + 4 \left| \frac{n_k}{H^{\text{D}}} \right|^2 \Re \left\{ s_k \left(\frac{n_k}{H^{\text{D}}} \right)^* \right\} + 2\Re \left\{ s_k^2 \left(\frac{n_k^*}{(H^{\text{D}})^*} \right)^2 \right\},$$

and

$$\text{E}\left[\hat{r}_k^2 (\hat{r}_k^*)^2 \right] = \text{E}\left[|s_k|^4 \right] + \frac{\text{E}[|n_k|^4]}{|H^{\text{D}}|^4} + 4\text{E}\left[|s_k|^2 \right] \frac{N_0}{|H^{\text{D}}|^2}, \tag{4.120}$$

where N_0 is the noise power and the last three terms in (4.120) disappear as the noise n_k is zero-mean circularly symmetric Gaussian distributed noise. It follows

$$\tilde{\mu}_{4,2}^{(i)} = \mu_{4,2}^{(i)} + 4\mu_{2,1}^{(i)} \frac{N_0}{|H^{\mathrm{D}}|^2} + \frac{\mathrm{E}[|n_k|^4]}{|H^{\mathrm{D}}|^4}, \tag{4.121}$$

and it shows $\tilde{\mu}_{4,2}^{(i)} \neq \mu_{4,2}^{(i)}$. Consequently, this feature parameter is biased leading to a mean value dependent on the noise power, an inconvenient property which can be eliminated by using higher order cumulants given in the next section. For low noise power ($\frac{N_0}{|H^{\mathrm{D}}|^2} \to 0$ and $\frac{\mathrm{E}[|n_k|^4]}{|H^{\mathrm{D}}|^4} \to 0$), however, it holds $\tilde{\mu}_{4,2}^{(i)} \approx \mu_{4,2}^{(i)}$.

Property 4: Assuming low noise power, it holds in general

$$\tilde{\mu}_{n_{\mathrm{m}},n_{\mathrm{m}}'}^{(i)} \approx \mu_{n_{\mathrm{m}},n_{\mathrm{m}}'}^{(i)}, \tag{4.122}$$

which can be viewed as an unbiased feature again.

Similarly, the variance can be determined by

$$\begin{aligned} \tilde{\sigma}_{n_{\mathrm{m}},n_{\mathrm{m}}'}^{(i)} &= \mathrm{var}[\hat{\mu}_{n_{\mathrm{m}},n_{\mathrm{m}}'}] = \mathrm{var}[\frac{1}{K} \sum_{k=0}^{K-1} \hat{r}_k^{n_{\mathrm{m}}-n_{\mathrm{m}}'} (\hat{r}_k^*)^{n_{\mathrm{m}}'}] \\ &= \frac{1}{K^2} \sum_{k=0}^{K-1} \mathrm{var}[\hat{r}_k^{n_{\mathrm{m}}-n_{\mathrm{m}}'} (\hat{r}_k^*)^{n_{\mathrm{m}}'}] \\ &= \frac{1}{K^2} \sum_{k=0}^{K-1} \int \left\{ \hat{r}_k^{n_{\mathrm{m}}-n_{\mathrm{m}}'} (\hat{r}_k^*)^{n_{\mathrm{m}}'} - \mathrm{E}\left[\hat{r}_k^{n_{\mathrm{m}}-n_{\mathrm{m}}'} (\hat{r}_k^*)^{n_{\mathrm{m}}'} \right] \right\} p(r_k|b_i) \mathrm{d}r_k, \end{aligned} \tag{4.123}$$

where again in general, a numerical computation of the two dimensional integration is necessary.

Property 5: Under assumption of low noise power, a similar approximation holds

$$\mathrm{var}[\hat{r}_k^{n_{\mathrm{m}}-n_{\mathrm{m}}'} (\hat{r}_k^*)^{n_{\mathrm{m}}'}] \approx \mu_{2n_{\mathrm{m}},n_{\mathrm{m}}}^{(i)} - \left| \mu_{n_{\mathrm{m}},n_{\mathrm{m}}'}^{(i)} \right|^2. \tag{4.124}$$

Using Property 4 and Property 5, the approximated mean and variance of the discriminating features selected for the decision tree in Figure 4.12 are provided in Table 4.5. For sufficiently large K, it holds $\tilde{\sigma}_{n_{\mathrm{m}},n_{\mathrm{m}}'}^{(i_1)} \approx \tilde{\sigma}_{n_{\mathrm{m}},n_{\mathrm{m}}'}^{(i_2)}$ with $i_1 \in \mathcal{I}_1$ and $i_2 \in \mathcal{I}_2$, which enables the application of Property 2. So, the classification rule is described by a threshold detector. If we further assume $\tilde{\mu}_{n_{\mathrm{m}},n_{\mathrm{m}}'}^{(i_1)} < \tilde{\mu}_{n_{\mathrm{m}},n_{\mathrm{m}}'}^{(i_2)}$, at each decision level $\tilde{l}$, the threshold value is given by

$$\mu_{\mathrm{th},\tilde{l}} = \frac{\tilde{\mu}_{\max}^{(i_1,\tilde{l})} + \tilde{\mu}_{\min}^{(i_2,\tilde{l})}}{2}, \tag{4.125}$$

Table 4.5: Approximated $\tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)}$ and $K\tilde{\sigma}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)}$ for moments selected in the decision tree.

	NoTx ($i=1$)	BPSK ($i=2$)	4QAM ($i=3$)	16QAM ($i=4$)	64QAM ($i=5$)
$\tilde{\mu}_{2,0}^{(i)}$	0	1	0	0	0
$K\tilde{\sigma}_{2,0}^{(i)}$	0	0	1	1.32	1.38
$\mu_{\mathrm{th},1}$	0.5				
$\tilde{\mu}_{2,1}^{(i)}$	0	1	1	1	1
$K\tilde{\sigma}_{2,1}^{(i)}$	0	0	0	0.32	0.38
$\mu_{\mathrm{th},2}$	0.5				
$\tilde{\mu}_{4,0}^{(i)}$	0	1	-1	-0.68	-0.619
$K\tilde{\sigma}_{4,0}^{(i)}$	0	0	0	2.66	3.58
$\mu_{\mathrm{th},3}$	-0.84				
$\tilde{\mu}_{4,2}^{(i)}$	0	1	1	1.32	1.38
$K\tilde{\sigma}_{4,2}^{(i)}$	0	0	0	1.38	2.05
$\mu_{\mathrm{th},4}$	1.35				

with

$$\tilde{\mu}_{\max}^{(i_1,\tilde{l})} = \max_{i_1} \tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i_1,\tilde{l})}, \qquad \tilde{\mu}_{\min}^{(i_2,\tilde{l})} = \min_{i_2} \tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i_2,\tilde{l})} \tag{4.126}$$

where $\mu_{n_{\mathrm{m}},n'_{\mathrm{m}}}$ is the selected discriminating feature at this decision level. Using (4.125) the corresponding threshold values are also given in Table 4.5.

Some important conclusions can be drawn with respect to the moment-based AMC. For illustration purpose, consider a binary decision problem $\{A_1, A_2\}$ using a one-dimensional feature space $\mathcal{F}_{\mathrm{mom}} = \{\mu_{n_{\mathrm{m}},n'_{\mathrm{m}}}\}$. Under the assumption of a low noise power, Property 4 and Property 5 imply that the parameter set $\{n_{\mathrm{m}}, n'_{\mathrm{m}}\}$ determines $\tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)}$ and $\tilde{\sigma}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)}$ and consequently the classification performance. Generally speaking, any $\{n_{\mathrm{m}}, n'_{\mathrm{m}}\}$, which ensures

$$\tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i_1)} \neq \tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i_2)} \quad \text{and} \quad \tilde{\sigma}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)} < \infty, \tag{4.127}$$

is qualified as a discriminating feature. For concrete practical problems, there exists usually an optimal set $\{n_{\mathrm{m,opt}}, n'_{\mathrm{m,opt}}\}$, which from mathematical viewpoint has to maximize $\Delta\tilde{\mu} = |\tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i_1)} - \tilde{\mu}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i_2)}|$ meanwhile minimize $\tilde{\sigma}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)}$. This, however, in most case is a contradictory requirement. For phase shifted keying (PSK) signals, it was shown that both $\Delta\tilde{\mu}$ and $\tilde{\sigma}_{n_{\mathrm{m}},n'_{\mathrm{m}}}^{(i)}$ are monotonically increasing functions with respect to n_{m} [81], [82]. This statement can be applied to the considered QAM schemes too, since QAM schemes can be viewed as multiple-level ring-formed PSK signals, e.g., 16 QAM can be treated

as three-level PSK signals with QPSK, 8-PSK and QPSK, respectively. In general it is impossbile to determine $\{n_{\mathrm{m,opt}}, n'_{\mathrm{m,opt}}\}$ analytically. Instead, depending on application environment extensive simulations have to be performed to search for optimal parameter set $\{n_{\mathrm{m,opt}}, n'_{\mathrm{m,opt}}\}$. This search process begins at small integer numbers and is very tedious especially for classification problems with multiple modulation candidates. In the practice, a performance requirement is set and stop the search process once the requirement is satisfied.

In literature it was also proposed to use several moments as a linearly combined discriminating feature $\mathcal{F} = \{f_{\mathrm{com}}\}$ [83]

$$f_{\mathrm{com}} = \sum_q \lambda_q \mu_{n_{\mathrm{m},q}, n'_{\mathrm{m},q}}, \tag{4.128}$$

where λ_q are optimization parameters, which mostly have to be determined by numerical methods. Notice that the goal of combining several moments is to raise the difference in mean value

$$\tilde{\mu}_{\mathrm{com}} = \sum_q \lambda_q \tilde{\mu}_{n_{\mathrm{m},q}, n'_{\mathrm{m},q}}, \tag{4.129}$$

but at the same time maintain the variance of the sum

$$\tilde{\sigma}_{\mathrm{com}} = \mathrm{var}[\sum_q \lambda_q \mu_{n_{\mathrm{m},q}, n'_{\mathrm{m},q}}], \tag{4.130}$$

inside a certain range, where $\mu_{n_{\mathrm{m},q}, n'_{\mathrm{m},q}}$ are statistically dependent. Here the key issue is to optimize the parameter set $\{\lambda_q\}_q$.

4.4.2   Higher order cumulant-based AMC

We mentioned that linear combinations of moments can create new discriminating features. Cumulant-based AMC is actually a special linear combination with some characteristic properties which have been shown to be very useful for classification problems. Cumulants can be obtained with the help of a so-called cumulant-generating function [84]

$$\Psi(\nu) = \ln\{\mathrm{E}\left[\exp(j\nu\mathbf{x})\right]\} = \sum_q \kappa_q \frac{\nu^q}{q!} \tag{4.131}$$

$$= \kappa_1 \nu + \kappa_2 \frac{\nu^2}{2!} + \cdots + \kappa_q \frac{\nu^q}{q!} + \cdots,$$

where $\Phi(\nu) = \mathrm{E}\left[\exp(j\nu\mathbf{x})\right]$ is the moment generating function and $\mathbf{x}$ is a random variable. The n_{c}-th order cumulant is the coefficient of $\frac{\nu^{n_{\mathrm{c}}}}{n_{\mathrm{c}}!}$ in the Taylor series expansion. Cumulants $\kappa_{n_{\mathrm{c}}}$ can be obtained also from moments $\mu_{n_{\mathrm{m}}}$ by coefficients comparison in

Taylor series of $\Psi(\nu)$ and $\Phi(\nu)$. For real random variables, the conversion from moments to cumulants and vice versa is provided in [85].

By analogy with moments, mixed cumulants can be denoted as $\kappa_{n_c, n'_c}(\mathbf{x})$ for the n_c-th order cumulant with n'_c-th order conjugation of a complex random variable $\mathbf{x}$. Here the relation between cumulants and moments are much more complicated and depends on the parameter pair (n_c, n'_c). A definition thereof can be found in [86]. As an example, $\kappa_{4,2}$ is given by

$$\kappa_{4,2}(\mathbf{x}) = \mathrm{E}[\mathbf{x}^2 (\mathbf{x}^*)^2] - \mathrm{E}[\mathbf{x}^2]\mathrm{E}\left[(\mathbf{x}^*)^2\right] - \mathrm{E}\left[\mathbf{x}\mathbf{x}^*\right]\mathrm{E}\left[\mathbf{x}\mathbf{x}^*\right] - \mathrm{E}[\mathbf{x}\mathbf{x}^*]\mathrm{E}\left[\mathbf{x}\mathbf{x}^*\right] \quad (4.132)$$

$$= \mu_{4,2} - \mu_{2,0}\mu_{2,2} - 2\mu_{2,1}^2 = \mu_{4,2} - |\mu_{2,0}|^2 - 2\mu_{2,1}^2.$$
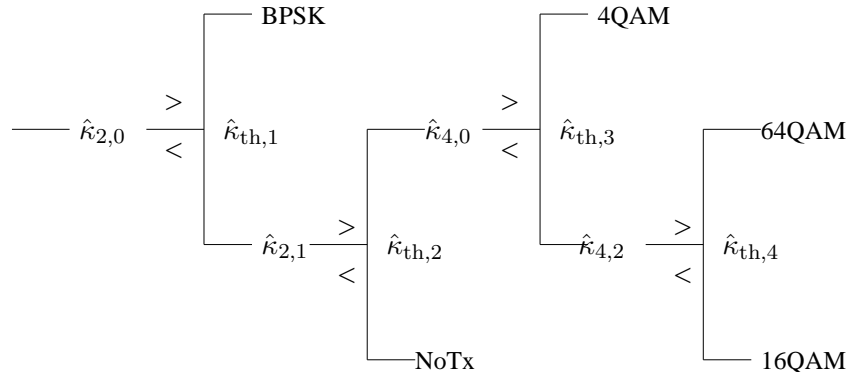
Figure 4.13: A possible binary classification tree based on cumulants

Relations between moments and cumulants are provided till $n_c = 8, n'_c = 4$ in the Appendix of [24]. A general relationship can be found in [76], [77]. We apply these relations to calculate the theoretical values $\kappa_{n_c, n'_c}(A_i)$ for the considered modulation formats $\{A_i\}_{i=1}^I$, which are given in Appendix F. Because of

$$\kappa_{2,0}(A_{i_1}) \neq \kappa_{2,0}(A_{i_2}) \quad \text{for } i_1 \in \{2\},\, i_2 \in \{1, 3, 4, 5\}, \quad (4.133)$$

$\kappa_{2,0}$ is able to discriminate between BPSK and the rest of the modulation formats. According to the same principle as stated in the moment-based AMC, a binary decision tree can be constructed (Figure 4.13). Similarly, $\hat{\kappa}_{n_c, n'_c}(\mathbf{r})$ is the estimate of the cumulant κ_{n_c, n'_c} obtained from involved moment estimates of the received symbol $\mathbf{r}$. For example, it holds

$$\hat{\kappa}_{4,2}(\mathbf{r}) = \hat{\mu}_{4,2}(\mathbf{r}) - |\hat{\mu}_{2,0}(\mathbf{r})|^2 - 2\hat{\mu}_{2,1}^2(\mathbf{r}) \quad (4.134)$$

$$= \frac{1}{K}\sum_{k=0}^{K-1} \hat{r}_k^2 (\hat{r}_k^*)^2 - \left|\frac{1}{K}\sum_{k=0}^{K-1} \hat{r}_k^2\right|^2 - 2\left(\frac{1}{K}\sum_{k=0}^{K-1} |\hat{r}_k|^2\right)^2,$$

with $\hat{r}_k = s_k + \frac{w_k}{H^{\mathrm{D}}}$.

The thresholds $\hat{\kappa}_{\mathrm{th},\tilde{l}}$ are determined with the help of some special properties of cumulants provided in the following without proof [76].

Property 6: For a random variable $\mathbf{x}$ and any constant $b, c \in \Re$, it holds

$$\kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(\mathbf{x} + b) = \kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(\mathbf{x}) \quad \text{for } n_{\mathrm{c}} \geq 2, \tag{4.135}$$

which is called semi-invariance. Furthermore, it holds

$$\kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(c\mathbf{x}) = c^{n_{\mathrm{c}}} \kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}, \tag{4.136}$$

which is called homogeneity.

Property 7: For two independent random variables $\mathbf{x}$ and $\mathbf{y}$, it holds

$$\kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(\mathbf{x} + \mathbf{y}) = \kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(\mathbf{x}) + \kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(\mathbf{y}), \tag{4.137}$$

which is called "cumulants accumulate" and is also the reason for the name.

Property 8: For a zero-mean Gaussian noise n_k, it holds

$$\kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(n_k) = 0 \quad \text{for } n_{\mathrm{c}} > 2, \tag{4.138}$$

Under the assumption of sufficiently long observation length K, Property 1 can be also applied for cumulant-based AMC [76]

$$\hat{\kappa}_{n_{\mathrm{c}},n_{\mathrm{c}}'} \sim \mathcal{N}\left(\tilde{\kappa}_{n_{\mathrm{c}},n_{\mathrm{c}}'}^{(i)}, \tilde{\varsigma}_{n_{\mathrm{c}},n_{\mathrm{c}}'}^{(i)}\right), \tag{4.139}$$

where $\tilde{\kappa}_{n_{\mathrm{c}},n_{\mathrm{c}}'}^{(i)}$ and $\tilde{\varsigma}_{n_{\mathrm{c}},n_{\mathrm{c}}'}^{(i)}$ are the mean and variance of the Gaussian distributed random variable $\hat{\kappa}_{n_{\mathrm{c}},n_{\mathrm{c}}'}$. The problem is now reduced to determine $\tilde{\kappa}_{n_{\mathrm{c}},n_{\mathrm{c}}'}^{(i)}$ and $\tilde{\varsigma}_{n_{\mathrm{c}},n_{\mathrm{c}}'}^{(i)}$ as for the moment-based AMC. In general, it is extremely difficult to analytically describe these two parameters. We assume here an infinitely long observation length, so that asymptotically, the mean value is given by

$$\tilde{\kappa}_{n_{\mathrm{c}},n_{\mathrm{c}}'}^{(i)} = \mathrm{E}\left[\hat{\kappa}_{n_{\mathrm{c}},n_{\mathrm{c}}'}(\mathbf{r})\right] \approx \mathrm{E}\left[\kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(r_k)\right] \quad \text{for } K \to \infty, \tag{4.140}$$

where K denotes the observation length. Using Property 6 and Property 7, it holds

$$\kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(r_k) = \kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(s_k) + \left(\frac{1}{\hat{H}^{\mathrm{D}}}\right)^{n_{\mathrm{c}}} \kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(w_k), \tag{4.141}$$

where $\hat{H}^{\mathrm{D}}$ is viewed as a deterministic known value. Using Property 8, it gives

$$\tilde{\kappa}^{(i)} \approx \mathrm{E}\left[\kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(r_k)\right] = \mathrm{E}\left[\kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}(s_k)\right] = \kappa_{n_{\mathrm{c}},n_{\mathrm{c}}'}^{(i)}, \tag{4.142}$$

with $n_{\mathrm{c}} \geq 4$, and where s_k is a signal point from the modulation format A_i. Note that under this condition the feature $\kappa_{n_c,n_c'}$ is unbiased and the low noise power assumption can be even relaxed. This shows the advantage of cumulants against moments. But this property holds only for $K \to \infty$. However, for $n_{\mathrm{c}} = 2$ and $n_{\mathrm{c}}' = 1$

$$\tilde{\kappa}_{n_c,n_c'}^{(i)} = \kappa_{n_c,n_c'}^{(i)} + \frac{N_0}{|H^{\mathrm{D}}|^2}, \tag{4.143}$$

the feature is biased, which becomes unbiased again only for low noise power. In Table 4.6, $\tilde{\kappa}^{(i)}$ are tabulated for the applied cumulants in the decision tree.

Generally, there is also no analytical formula to determine $\tilde{\varsigma}^{(i)}$. Under certain circumstance, approximations are possible. We illustrate the derivation in case of $\tilde{\varsigma}_{4,2}^{(i)}$. Other derivations can be found in [78].

Table 4.6: Approximated $\tilde{\kappa}_{n_c,n_c'}^{(i)}$ and $K\tilde{\varsigma}_{n_c,n_c'}^{(i)}$ for cumulants selected in the decision tree.

	NoTx ($i=1$)	BPSK ($i=2$)	4QAM ($i=3$)	16QAM ($i=4$)	64QAM ($i=5$)
$\tilde{\kappa}_{2,0}^{(i)}$	0	1	0	0	0
$K\tilde{\varsigma}_{2,0}^{(i)}$	0	0	1	1.32	1.38
$\kappa_{\mathrm{th},1}$			0.5		
$\tilde{\kappa}_{2,1}^{(i)}$	0	1	1	1	1
$K\tilde{\varsigma}_{2,1}^{(i)}$	0	0	0	0.32	0.38
$\kappa_{\mathrm{th},2}$			0.5		
$\tilde{\kappa}_{4,0}^{(i)}$	0	-2	-1	-0.68	-0.619
$K\tilde{\varsigma}_{4,0}^{(i)}$	0	0	0	2.66	3.58
$\kappa_{\mathrm{th},3}$			-0.84		
$\tilde{\kappa}_{4,2}^{(i)}$	0	-2	-1	-0.68	-0.619
$K\tilde{\varsigma}_{4,2}^{(i)}$	0	0	0	1.38	2.05
$\kappa_{\mathrm{th},4}$			-0.65		

Property 9: For normalized QAM signals, it holds

$$\kappa_{2,0} = 0; \quad \kappa_{2,1} = 1. \tag{4.144}$$

It is shown in [78]

$$\tilde{\varsigma}_{4,2} \approx \frac{1}{K} \left(\mu_{8,4} - |\mu_{4,2}|^2 \right). \tag{4.145}$$

Since $\mathbf{r}$ is a noisy process, in order to use Property 7, the moments in (4.145) have to be converted to cumulants by using

$$
\begin{aligned}
\mu_{8,4} &= \kappa_{8,4} + 16\kappa_{6,3}\kappa_{2,1} + |\kappa_{4,0}|^2 + 18\kappa_{4,2}^2 + 72\kappa_{4,2}\kappa_{2,1}^2 + 24\kappa_{2,1}^4, \tag{4.146} \\
\mu_{4,2} &= \kappa_{4,2} + 2\kappa_{2,1}^2.
\end{aligned}
$$

Inserting (4.146) into (4.145) gives:

$$\tilde{\varsigma}_{4,2} \approx \kappa_{8,4} + 16\kappa_{6,3}\kappa_{2,1} + |\kappa_{4,0}|^2 + 17\kappa_{4,2}^2 + 68\kappa_{4,2}\kappa_{2,1}^2 + 20\kappa_{2,1}^4. \tag{4.147}$$

In (4.147), based on Property 7, each cumulant $\kappa_{n_c,n_c'}$ has to be interpreted as $\kappa_{n_c,n_c'}^{(i)} + \kappa_{n_c,n_c'}(\frac{w_k}{\hat{H}^D})$, where $\hat{H}^D$ can be viewed as a deterministic value enabling the use of Property 6. Note that the approximation does not require the assumption of low niose power. Furthermore, based on Property 6 and Property 8, $\kappa_{n_c,n_c'}(\frac{w_k}{H^D})$ disappears except for $n_c = 2$ and $n_c' = 1$, which strongly limits the contribution of noise to the variance. This can be viewed as another advantage of cumulants against moments.

For $n_c \geq 2$ and $n_c' \geq 1$, $\kappa_{2,1}(\frac{w_k}{H^D})$ is usually present in $\tilde{\varsigma}_{n_c,n_c'}$ resulting in a noise power dependent variance. To assess the relationship between the variances, i.e., to choose the detector type as proposed in Property 2 and Property 3, we neglect the contribution from the noise as similar to the procedure in the moment-based AMC. The resulting variances are also provided in Table 4.6 for the selected cumulants in the decision tree in Figure 4.13. We observe that under the assumption of low noise power it holds $\tilde{\varsigma}_{n_c,n_c'} \approx \tilde{\sigma}_{n_c,n_c'}$, so that the detector type selected for moment-based AMC continue to hold and the corresponding thresholds are calculated according to (4.125) for each decision level $\tilde{l}$. We point out that, however, the low noise power assumption can be relaxed leading to a noise power independent variance, which may still not affect the selection of the detector type, if K is sufficiently large.
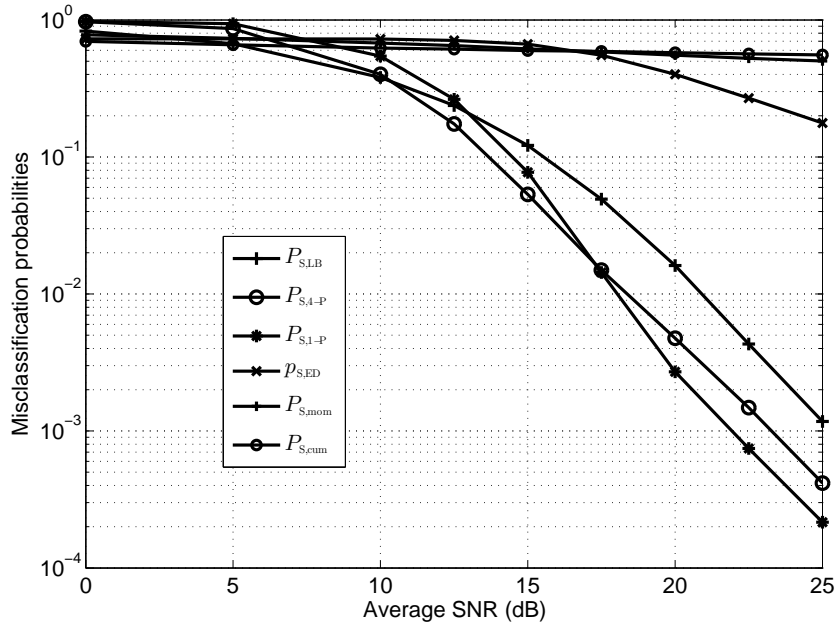
Figure 4.14: P_S of different classifiers based on ML function, 1-P, 4-P, ED approximated, moment-based and cumulat-based classifier

4.4.3   Simulation results

Simulations are conducted to evaluate the performance of the proposed moment- and cumulant-based classification algorithms in terms of probability of subcarrier misclassification P_s. The moment-based classifier uses the hierarchical structure shown in Figure 4.12. The cumulant-based classifier uses the structure illustrated in Figure 4.13. The results are plotted in Figure 4.14, which are comparable to the classification results of the Euclidean distance based classifier. The performance is not improved with increasing SNR showing little sensitivity to noise power level. The poor performance confirms the fact that feature-based classifiers may not be suitable for the considered modulation classification problem. This is explained by the following reasons:

1. Feature-based classifiers are generally not suitable to classify nested modulation candidates (modulation schemes from the same class);

2. Feature-based classifiers require a long observation (more than 10.000 symbols have to be observed to make a reliable decision [78]). In this work, we aim at designing AMC algorithms for real-time applications, so that the observation length is short;

3. Most feature-based classifiers assume an additive noise only. Deep fades encountered in wireless fading channels cause high disagreement between theoretical and estimated feature values;

4. Due to the short observation period, the density funtion is not Gaussian distibuted any more leading to model mismatch. For short observation lengths, a mathematical treatment of feature-based modulation classification is extremely difficult. Consequently, the assumptions made are not valid any more.

We observed that the cumulant-based AMC shows also no improvement over moment-based AMC, which otherwise would be expected from previous theoretical derivations. This fact is mainly caused by short observation length. Although, at the same decision layer the cumulants, which are given by the linear combination of moments, have higher difference in mean value, but the short observation length also leads to a much higher variance, which then compensates the benifit resulting from the increased difference in the mean values. Due to the above reasons, the feature-based approach will not pursued in this thesis any more.

4.5   System-specific measures for performance improvement

Practical systems often have to satisfy boundary conditions imposed from the operation environment. Certain boundary conditions are even included in the adaptive modulation algorithm. For instance, in the cellular system enhanced general packet radio service (EGPRS), where both adaptive modulation and coding are implemented in a cellular system for the first time, a set of boundary parameters is required as inputs for the so-called Downlink Modulation and Coding Scheme (MCS) Adviser and Downlink Intelligent Override, who decide which MCS will be used in the next DL. These parameters can be service class and QoS type requested by the MS, indication for retransmission or fresh block and remaining data volume to be transferred [88].

4.5.1   Boundary conditions
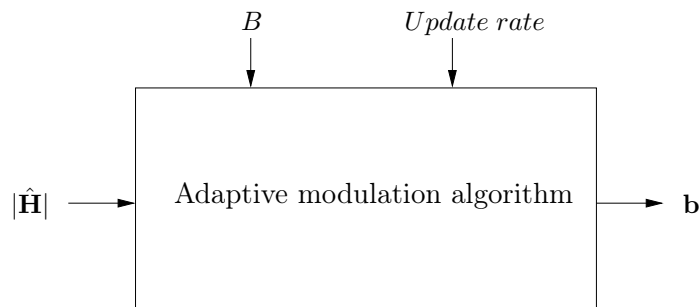
Figure 4.15:   Input parameters for adaptive modulation

The adaptive modulation algorithm considered in this thesis requires input parameters as shown in Figure 4.15. The parameter *Update rate* indicates how frequently the bit loading vector **b** has to be updated within one packet. An update on an OFDM symbol basis would require too high computational efforts. So, for slowly time-variant channels, no update is designed for practical systems due to high implementation complexity. The parameter B and **H** were explained in previous sections. The B-constraint requires that the number of bits received in an OFDM block must be

$$\sum_{n=0}^{N-1} \hat{b}^{(n)} = B, \tag{4.148}$$

otherwise, at least one subcarrier must have had a misclassified modulation scheme. Apparently, this boundary condition can be used to check whether any classification error occurred in the detected bit-loading vector $\{\hat{b}^{(n)}\}_{n=0}^{N-1}$. On the other side, it can

also serve as a measure to improve the overall classification performance. Obviously, only limited number of combinations of modulation levels can fulfil (4.148). We introduce the set $\mathcal{C}_{\hat{b}_i^{(n)}}$ including all combinations as

$$\mathcal{C}_{\hat{b}_i^{(n)}} = \left\{ \left\{ \hat{b}_{i_{n_{\mathrm{cb}}}}^{(n)} \right\}_{n=0}^{N-1} \right\}_{n_{\mathrm{cb}}=0}^{N_{\mathrm{CB}}-1} = \left\{ \left\{ \hat{b}_{i_0}^{(n)} \right\}_{n=0}^{N-1}, \left\{ \hat{b}_{i_1}^{(n)} \right\}_{n=0}^{N-1}, \cdots, \left\{ \hat{b}_{i_{N_{\mathrm{CB}}-1}}^{(n)} \right\}_{n=0}^{N-1} \right\}, \quad (4.149)$$

which fulfils the sum constraint given in (4.148) with N_{CB} as the number of possible combinations and n_{cb} as the combination index, respectively. The optimal classifier in the sense of minimizing the misclassification probability is given by maximizing the N-dimensional joint a posteriori probability

$$\left\{ \hat{b}^{(n)} \right\}_{n=0}^{N-1} = \underset{n_{\mathrm{cb}}}{\operatorname{argmax}} \left\{ p\left(\left\{ \hat{b}_{i_{n_{\mathrm{cb}}}}^{(n)} \right\}_{n=0}^{N-1} \mid \{\mathbf{r}_n\}_{n=0}^{N-1}, \hat{\mathbf{H}}^{\mathrm{D}}, B \right) \right\}, \quad (4.150)$$

which reflects the mentioned cross-subcarrier bit loading process of the used AM. Note that the maximization is now limited inside the set $\mathcal{C}_{\hat{b}_i^{(n)}}$ only, so that the number of possible solutions are reduced from I^N to N_{CB} with $N_{\mathrm{CB}} \ll I^N$ in general. But the evaluation of the N-dimensional joint probability requires extremely high computational efforts for large N. Furthermore, the number N_{CB} grows approximately exponentially with N. In practical systems, (4.150) is not implementable. In [9], a compromise is proposed which divides the N subcarriers into two groups according to a certain grouping criterion. The first group contains N_1 subcarriers $\{n_{n_1}\}_{n_1=0}^{N_1-1}$, whose classified modulation levels $\{\hat{b}^{(n_{n_1})}\}_{n_1=0}^{N_1-1}$ are viewed as reliable. The classification strategy for this group follows the decision rule formulated in the previous sections. The remaining $N_2 = N - N_1$ subcarriers $\{n_{n_2}\}_{n_2=0}^{N_2-1}$ are assigned into the second group, whose classified modulation levels $\{\hat{b}^{(n_{n_2})}\}_{n_2=0}^{N_2-1}$ is assessed as less reliable. The classification performance of the second group is then enhanced by the above boundary condition, which is now given by

$$B_{\mathrm{rem}} = \sum_{n_2=0}^{N_2-1} \hat{b}^{(n_{n_2})} = B - \sum_{n_1=0}^{N_1-1} \hat{b}^{(n_{n_1})}. \quad (4.151)$$

Now, only a very small number of combinations is possible, if N_2 is selected to be a small number, as summarized in the following set

$$\mathcal{C}_{\hat{b}_i^{(n_{n_2})}} = \left\{ \left\{ \hat{b}_{i_{n_{\mathrm{cb},2}}}^{(n_{n_2})} \right\}_{n_2=0}^{N_2-1} \right\}_{n_{\mathrm{cb},2}=0}^{N_{\mathrm{cb},2}-1}, \quad (4.152)$$

where, similarly, $N_{\mathrm{CB},2}$ is the number of possible modulation combinations for the subcarriers in group two and $n_{\mathrm{cb},2}$ is the combination index, respectively. Consequently, the most likely applied modulation combination at the transmitter is the one which maximizes the N_2-dimensional joint a-posteriori probability

$$\left\{ \hat{b}^{(n_{n_2})} \right\}_{n_2=0}^{N_2-1} = \underset{n_{\mathrm{cb},2}}{\operatorname{argmax}} \left\{ p\left(\left\{ \hat{b}_{i_{n_{\mathrm{cb},2}}}^{(n_{n_2})} \right\}_{n_2=0}^{N_2-1} \mid \{\mathbf{r}_{n_{n_2}}\}_{n_2=0}^{N_2-1}, \hat{\mathbf{H}}^{\mathrm{D}}, B \right) \right\}, \quad (4.153)$$

which, in case of statistical independence, can be approximated by

$$\left\{\hat{b}^{(n_2)}\right\}_{n_2=0}^{N_2-1} = \underset{n_{\mathrm{cb},2}}{\mathrm{argmax}} \left\{ \prod_{n_2=0}^{N_2-1} p(\hat{b}_{i_{n_{\mathrm{cb},2}}}^{(n_2)} | \mathbf{r}_{n_2}, \hat{\mathbf{H}}^{\mathrm{D}}, B) \right\}, \tag{4.154}$$

provided that the spectral spacing between the subcarriers $\{n_{n_2}\}_{n_2=0}^{N_2-1}$ is at least equally large as the coherence bandwidth of the channel B_{Co}, which is mostly fulfiled for small N_2. The decision rule in (4.154) can be also given in the following form

$$\left\{\hat{b}^{(n_2)}\right\}_{n_2=0}^{N_2-1} = \underset{n_{\mathrm{cb},2}}{\mathrm{argmax}} \left\{ \sum_{n_2=0}^{N_2-1} \ln\left[p(\hat{b}_{i_{n_{\mathrm{cb},2}}}^{(n_2)} | \mathbf{r}_{n_2}, \hat{\mathbf{H}}^{\mathrm{D}}, B) \right] \right\}. \tag{4.155}$$

The classification of the bit loading vector $\mathbf{b}$ is now organized in four steps:

1. According to a classification reliability index (will be given later), the N subcarriers are divided into two groups with N_1 subcarriers in group one having higher classification reliability and N_2 subcarriers in group two having lower classification reliability;

2. The first group with subcarriers $\{n_{n_1}\}_{n_1=0}^{N_1-1}$ is classified using MAP-based decision rules formulated before;

3. The elements of the set given in (4.152) are listed out for the remaining subcarriers $\{n_{n_2}\}_{n_2=0}^{N_2-1}$ in group two;

4. These subcarriers are classified using the decision rule in (4.155).

Until now the grouping criterion mentioned above is not yet formulated. The criterion indicating the classification reliability depends necessarily on the applied adaptive modulation and modulation classification algorithm. As reported before, link quality indicates, to certain extend, the classification reliability. The link quality is specified by the amplitude of the channel coefficients $\{|\hat{H}_n^{\mathrm{D}}|\}_{n=0}^{N-1}$. We denote $P(mis| |\hat{H}_n^{\mathrm{D}}|)$ as the conditional misclassification probability conditioned on that data is transmitted via the n-th subcarrier with a transfer factor H_n^{D} and the receiver obtains the channel estimate $\hat{H}_n^{\mathrm{D}}$. We expect

$$P(mis| |\hat{H}_{n_1'}^{\mathrm{D}}|) > P(mis| |\hat{H}_{n_2'}^{\mathrm{D}}|) \ \ \text{for} \ \ |\hat{H}_{n_1'}^{\mathrm{D}}| < |\hat{H}_{n_2'}^{\mathrm{D}}|. \tag{4.156}$$

Based on this criterion, subcarriers are divided into two groups $\{n_{n_1'}\}_{n_1'=0}^{N_1'-1}$ and $\{n_{n_2'}\}_{n_2'=0}^{N_2'-1}$ by comparing against a threshold

$$|\hat{H}_{n_{n_1'}}^{\mathrm{D}}| > |H_{\mathrm{th}}| \ \ \text{and} \ \ |\hat{H}_{n_{n_2'}}^{\mathrm{D}}| \le |H_{\mathrm{th}}|. \tag{4.157}$$

This criterion is disturbed by the overloading problem as illustrated in Subsection 4.2.6. Furthermore, the precision of channel estimation plays also an important role. In general, coefficients with small amplitudes $|H_n^{\mathrm{D}}|$ are potentially more difficult to estimate. So the application of this criterion is limited.

A more practical criterion is proved to work well [9]. It uses the difference between the largest and the second largest value of the decision metric $\kappa_{\mathrm{DM}}(b^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B))$, which can be any proposed modulation classification metric formulated before. The argument $Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)$ specifies the prior knowledge included in MAP-based classifiers. Depending on whether and how the prior knowledge shall be incorporated into the classification rule, $Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)$ is given by different expressions as illustrated before. In detail, the grouping method performs the following 3 steps:

1. Sort $\kappa_{\mathrm{DM}}(b^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B))$ across b_i with $i = 1, \cdots, I$

$$\kappa_{\mathrm{DM}}(b_{i_1}^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)) \;>\; \kappa_{\mathrm{DM}}(b_{i_2}^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)) \tag{4.158}$$
$$\cdots > \kappa_{\mathrm{DM}}(b_{i_I}^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)).$$

2. Build the difference between metrics based on b_{i_1} and b_{i_2}

$$\Delta\kappa_{\mathrm{DM}}(b_i^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)) \;=\; \kappa_{\mathrm{DM}}(b_{i_1}^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)) - \tag{4.159}$$
$$\kappa_{\mathrm{DM}}(b_{i_2}^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)).$$

3. Divide subcarriers into two groups $\{n_{n_1}\}_{n_1=0}^{N_1-1}$ and $\{n_{n_2}\}_{n_2=0}^{N_2-1}$ by comparing against a threshold

$$\{n_{n_1}\}_{n_1=0}^{N_1-1} \quad \text{with} \quad \Delta\kappa_{\mathrm{DM}}(b_i^{(n_{n_1})}|\mathbf{r}_{n_{n_1}}, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)) > \Delta\kappa_{\mathrm{DM,th}} \tag{4.160}$$
$$\{n_{n_2}\}_{n_2=0}^{N_2-1} \quad \text{with} \quad \Delta\kappa_{\mathrm{DM}}(b_i^{(n_{n_2})}|\mathbf{r}_{n_{n_2}}, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B)) \leq \Delta\kappa_{\mathrm{DM,th}}.$$

Boundary condition splitting technique

Incorporation of the above boundary condition into the classification algorithm can significantly improve the system performance as will be confirmed in later computer simulations [9]. If the system has sufficiently large number of subcarriers, the boundary condition can be even applied multiple times.

Suppose that the system has N subcarriers, via which B bits have to be transmitted as stated before. Now the N subcarriers are alternately split into n_{g} groups in the following

way

$$\text{These subcarriers belong to group 1:} \quad 0, n_{\mathrm{g}}, 2n_{\mathrm{g}}, \cdots, (\frac{N}{n_{\mathrm{g}}} - 1)n_{\mathrm{g}} \qquad (4.161)$$

$$\text{These subcarriers belong to group 2:} \quad 1, n_{\mathrm{g}} + 1, 2n_{\mathrm{g}} + 1, \cdots, (\frac{N}{n_{\mathrm{g}}} - 1)n_{\mathrm{g}} + 1$$

$$\vdots$$

$$\text{These subcarriers belong to last group } n_{\mathrm{g}}: \quad n_{\mathrm{g}} - 1, 2n_{\mathrm{g}} - 1, 3n_{\mathrm{g}} - 1, \cdots, N - 1.$$

n_{g} is the number of groups, which the N subcarriers are split into. Each group has $\frac{N}{n_{\mathrm{g}}}$ subcarriers, via which now $\frac{B}{n_{\mathrm{g}}}$ bits have to be transmitted. This splitting technique results in n_{g} times utilization of a single boundary condition. The classification rule given in (4.148), (4.149) and (4.154) is applied on each individual groups.

But, with fixed N, increasing n_{g} means reduction of frequency diversity leading to reduction of the efficiency of the adaptive modulation. In practical implementations, depending on system specifications, an optimal group number $n_{\mathrm{g,opt}}$ can be found

4.5.2   Channel interpolation

In a TDD-based adaptive OFDM system, the BS uses the out-dated UL channel estimate $\hat{\mathbf{H}}^{\mathrm{U}}$ to perform the adaptive modulation. The MS use the DL channel estimate $\hat{\mathbf{H}}^{\mathrm{D}}$ to equalize the received symbols $\{\mathbf{r}_n\}_{n=0}^{N-1}$, which is in turn required for the subsequent modulation classification. In the previously proposed MAP classifiers, $\hat{\mathbf{H}}^{\mathrm{D}}$ is incorporated into the classification algorithm as the prior knowledge.
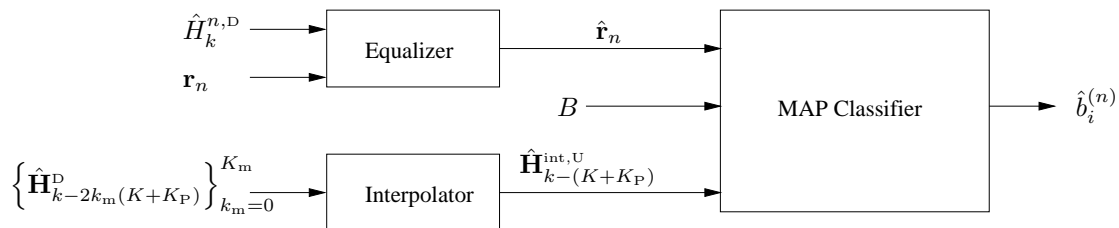
Figure 4.16:   MAP based classifier with channel interpolation

Obviously, the prior knowledge would be more precisely described by $\hat{\mathbf{H}}^{\mathrm{U}}$. But $\hat{\mathbf{H}}^{\mathrm{U}}$ is not available at the MS. The mismatch between $\hat{\mathbf{H}}^{\mathrm{U}}$ and $\hat{\mathbf{H}}^{\mathrm{D}}$ degrades the effectiveness of the MAP classifier. On the other side, $\hat{\mathbf{H}}^{\mathrm{U}}$ represents the channel measurement in the same frequency band as $\hat{\mathbf{H}}^{\mathrm{D}}$ but at the time slot exactly one packet duration earlier than $\hat{\mathbf{H}}^{\mathrm{D}}$ as shown in Figure 1.1. This property enables the application of time-domain interpolation to improve the effectiveness of the MAP classifier. We assume that the

MS is at the time slot k and intends to detect the bit loading vector $\mathbf{b}$ calculated on the basis of the channel measurement $\hat{\mathbf{H}}_{k-(K+K_{\mathrm{P}})}^{\mathrm{U}}$ at $k - (K + K_{\mathrm{P}})$. The current channel measurements $\hat{\mathbf{H}}_k^{\mathrm{D}}$ is available at the MS. In addition, the previous channel measurements $\hat{\mathbf{H}}_{k-2(K+K_{\mathrm{P}})}^{\mathrm{D}}$, $\hat{\mathbf{H}}_{k-4(K+K_{\mathrm{P}})}^{\mathrm{D}}$, $\cdots$, $\hat{\mathbf{H}}_{k-2K_{\mathrm{m}}(K+K_{\mathrm{P}})}^{\mathrm{D}}$ can be stored in the MS intern memory space, where K_{m} denotes the memory capability in the MS, with k_{m} as the memory index. Using these measurements, an interpolation can be applied to create an interpolated channel measurements $\hat{\mathbf{H}}_{k-(K+K_{\mathrm{P}})}^{\mathrm{int,U}}$ for the time slot $k - (K + K_{\mathrm{P}})$ to approximate the original channel measurement $\hat{\mathbf{H}}_{k-(K+K_{\mathrm{P}})}^{\mathrm{U}}$. The classification strategy is now graphically represented in Figure 4.16. The interpolation is performed on a per-subcarrier basis and illustrated in the following for some simple and well-known interpolation methods.

Linear interpolation

The most simple method is the linear interpolation

$$\hat{\mathbf{H}}_{k-(K+K_{\mathrm{P}})}^{\mathrm{int,U}} = \frac{1}{2} \left(\hat{\mathbf{H}}_k^{\mathrm{D}} + \hat{\mathbf{H}}_{k-2(K+K_{\mathrm{P}})}^{\mathrm{D}} \right). \tag{4.162}$$

In this method only one memory unit is required ($K_{\mathrm{m}} = 1$). Under the condition of low noise power and slowly time-variant channels, this method is recommended.

Polynomial interpolation
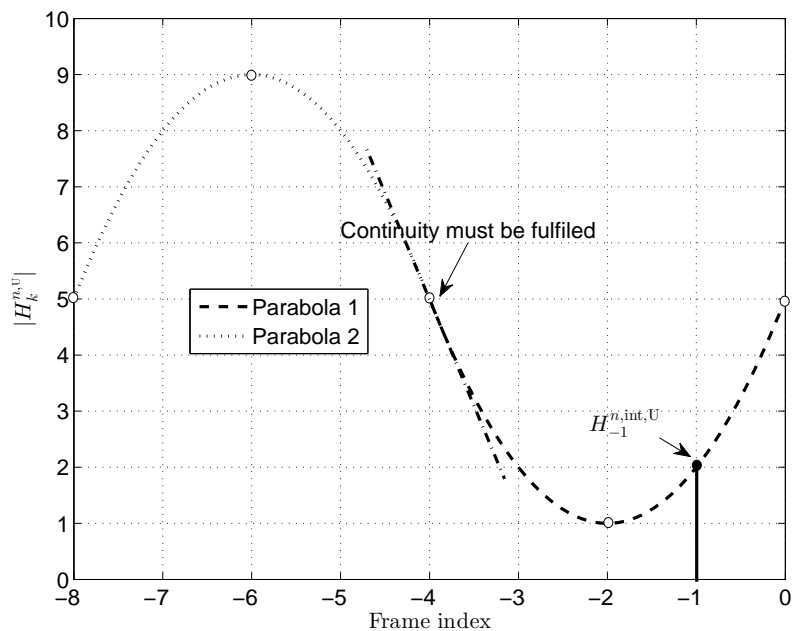
Figure 4.17: Parabolic interpolation

For p_{i} given points, there exists an unique polynomial of degree $p_{\mathrm{i}} - 1$ or less, which passes through these points [89]. For $K_{\mathrm{m}} \geq 2$, $K_{\mathrm{m}} + 1$ channel measurements are available.

Figure 4.17 shows the principle of constructing a parabolic interpolation. There are five channel measurements $\hat{H}_k^{n,\mathrm{D}}$ available at packet index $k = 0$, -2, -4, -6 and -8, respectively. Beginning at the data pairs $(0, \hat{H}_0^{n,\mathrm{D}})$ and $(-2, \hat{H}_{-2}^{n,\mathrm{D}})$, piecewise parabolas are construced for each neighbouring data pairs. A parable is given by

$$y_{\mathrm{p}} = C_{\mathrm{p},2}x_{\mathrm{p}}^2 + C_{\mathrm{p},1}x_{\mathrm{p}} + C_{\mathrm{p},0}, \tag{4.163}$$

including three unknown coefficients $C_{\mathrm{p},2}$, $C_{\mathrm{p},1}$ and $C_{\mathrm{p},0}$. The determination of these three coefficients requires three independent equations. Each data pair provides 2 independent equations. The third equation is created by requiring continuity condition of the first derivative of y_{p} to be fulfilled at the interpolating nodes, e.g. at $k = -2$ for the first data pair. The desired channel measurements $\hat{H}_{k'-(K+K_{\mathrm{P}})}^{n,\mathrm{U}}$ is approximated by $\hat{H}_{k'-(K+K_{\mathrm{P}})}^{n,\mathrm{int,U}}$, which is the sampled value of the constructed parable at $k = -1$ as shown in Figure 4.17. If more data points are available, higher order polynomial interpolation can be applied, e.g. cubic spline interpolation. More details for this topic can be found in [89].

4.5.3   Rotation of signal constellations

Classification of modulation candidates of the same class arises the problem of nested signal constellations. This leads to poor distinguishability between the candidates. An easy and effective solution to improve the distinguishability is to rotate certain signal constellations [42], [48]. For the QAM signals considered in this thesis, the 16 QAM signal constellation is rotated clockwise by 45° as shown in Figure 4.18. In the regular 16 QAM constellation, each signal point is surrounded by 4 signal points of 64 QAM, i.e., the 16 QAM is nested in the 64 QAM. By rotating the 16 QAM, this regular nesting arrangement is distorted. Especially, the 4 points at the four corners are rotated even outside of the surrounding of the 64 QAM.

The 4-P approximation proposed for simplifying the likelihood function $\kappa_1(\mathbf{r}|b_i)$ can be also applied on the rotated 16 QAM constellation. The detailed derivation is given in Appendix E. Theoretically, any distortion of this nested structure of the signal constellations may result in an improved distinguishability. In case of a complex signal constellation, the phase and the amplitude of the signal points can be adjusted for this purpose. It was shown through computer simulation, although varying amplitude can also improve the classification reliability, an meaningful variation in sense of improving PER performance depends on the system environment like the channel state and system SNR. It is difficult to find an analytic expression of this dependence. A variation in the amplitude causes also a variation in transmit power, which is mostly undesired in practical systems. Importantly, a rotation in phase by $\frac{\pi}{4}$, which does not change the transmit power, not only distorts the nested structure but also preserves the regular QAM arrangement which
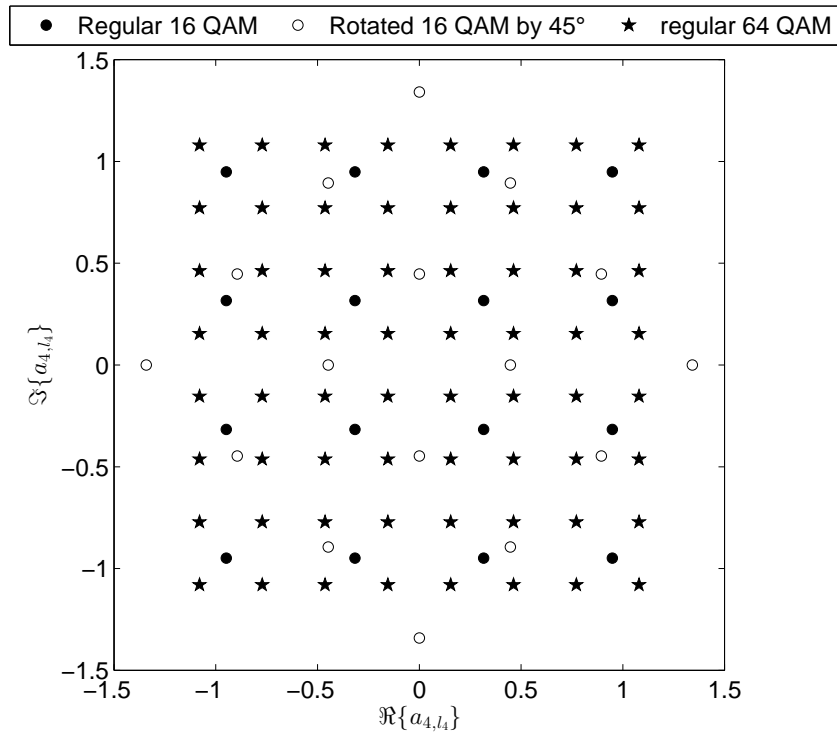
Figure 4.18: Rotation of 16 QAM by 45° to reduce the effect of nested signal constellation

allows an application of the 4-P approximation again.

4.5.4   Simulation results

Since the described system-specific measures above involve cross-subcarrier modulation classificaiton, like the B-constraint boundary condition, the performance improvement, achieved by incorporating these additional measures into the proposed 2-D MAP ML and 3-D MAP 4-P classifier, is now evaluated in terms of probability of packet misclassification P_{F} with

$$P_{\text{F}} = P(\mathbf{b} \neq \hat{\mathbf{b}}). \tag{4.164}$$

The performance improvement of the probability of subcarrier misclassification P_{S} is less significant in this case, but is provided also in Figure H.1 in Appendix H. The performance metric P_{F} is meaningful if all subcarriers are assigned to a single service as normally the case in IEEE 802.11a, since any misclassified subcarrier will lead to a packet detection error due to bit number mismatch. For reasonably high SNR, it holds $P_{\text{S}} \approx \frac{P_{\text{F}}}{N}$. With $N = 48$, $P_{\text{F}} = 0.1$ corresponds to $P_{\text{S}} = 0.002$. The underlying system is described in Appendix A. We used here a realistic channel estimation method, which is also illustrated in Appendix A. The division of subcarriers into two groups with different classification reliability follows the mentioned two criteria, namely the channel

amplitude $|\hat{H}_n^{\mathrm{D}}|$ and the metric difference $\Delta\kappa_{\mathrm{DM}}(b^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B))$. The number of subcarriers, whose classifiability is rated as less reliable, is denoted as N_2 for the criterion $\Delta\kappa_{\mathrm{DM}}(b^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B))$ and N_2' for the criterion of $|\hat{H}_n^{\mathrm{D}}|$, respectively.
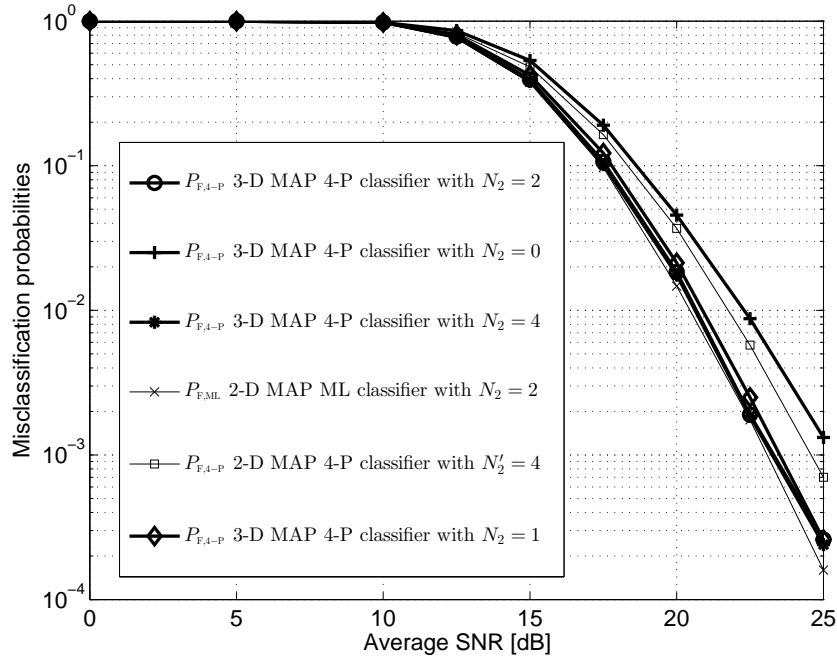
Figure 4.19:   Performance improvement by boundary condition in P_{F}

.

Data rate assisted classifiers

The performance is shown in Figure 4.19. Using the 3-D MAP 4-P classifier and the criterion $|\hat{H}_n^{\mathrm{D}}|$, only a small improvement in P_{F} is observed even if N_2' is set to $N_2' = 4$. This result indicates the inefficiency of this criterion. On the other side, using the same 3-D MAP 4-P classifier combined with the second criterion $\Delta\kappa_{\mathrm{DM}}(b^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B))$, a remarkable improvement is observed in P_{F} already for $N_2 = 1$. Around 1 dB performance gain is possible for a classification reliability of $P_{\mathrm{F}} = 0.1$. Increasing N_2 to 2, a further slight improvement is observable. However, the performance stagnates at $N_2 = 2$. Further increase of N_2 leads to no performance gain as confirmed by setting $N_2 = 4$. The performance stagnation implies that this criterion allows only a limited capability for evaluation of the classification reliability. Once N_2 exceeds this limit, this criterion is not suitable to distinguish the subcarriers with respect to their classification reliability any longer. Improvements are also registered in the performance of P_{S} (see Appendix H). However, the gain achieved there is less remarkable as the performance metric P_{S} is normalized to the total number of subcarriers N. This behaviour shows that P_{S} benefites less from the usage of this boundary condition.

For comparison, we also simulated the 2-D MAP ML classifier combined with the criterion of $\Delta\kappa_{\mathrm{DM}}(b^{(n)}|\mathbf{r}_n, Pr(|\hat{\mathbf{H}}^{\mathrm{D}}|, B))$ with $N_2 = 4$. The performance behaves very similarly as that of 3D MAP 4P with $N_2 = 2$, which confirms the classification feasibility of the computationally efficient 3-D MAP 4-P classifier also under the application of the boundary condition.

Channel interpolation assisted classifiers

The performance improvement, which is enabled by interpolating the channel estimates by the MS to approximate the channel measurement used for the adaptive modulation at the BS, is evaluated in terms of P_{F} in Figure 4.20 (the correspongding P_{S} performance is provided in Appendix H). In addition to the boundary condition given above, the 3-D MAP 4-P classifier is now equiped with the following three interpolation techniques: polynomial interpolation (PI), Matlab built-in low-pass interpolation (LPI) and the linear interpolation (LI). The polynomial interpolation constructs piece-wise polynomial of degree n_{pl} between the data points and requires furthermore continuity to be satisfied at the interpolating nodes for the polynomial itself, 1-st, 2-nd, $\cdots$, and $(n_{\mathrm{pl}} - 1)$-th order derivative thereof. Here n_{pl} is set to 3 which results in a cubic spline interpolation (CSI). The low-pass filter follows the three steps. First, the original data vector is expanded into the desired length by just inserting zeros between the data points. Then, a symmetric finite impulse response filter (FIR) of adjustable order is applied, which allows the original data points pass through unchanged but interpolates the data point in between such that the mean square errors between the interpolated points and their ideal values are minimized. Finally, the desired data point is obtained by selecting the sample at correct time instance [90].

Based on the 3-D MAP 4-P classifier with $N_2 = 2$, an additional equipment with CSI shows no effects on the performance in low SNR region. A slight improvement is observable in the moderate SNR region. The performance is improved more evidently as the SNR value further increases. This behaviour indicates that this interpolation technique is highly sensitive to the precision of the channel estimation. However, the LPI and LI assisted classifiers show a considerable improvement already in the moderate SNR region which indicates less sensitivity to channel estimation errors. For a performance of $P_{\mathrm{F}} = 0.1$, around 0.5 dB gain is registered by equipping the 3-D MAP 4-P classifier with a simple linear interpolator. In the sequel, we prefer the linear interpolation as it delivers similar performance as the low-pass interpolation but provokes much less additional computational efforts and memory requirements.
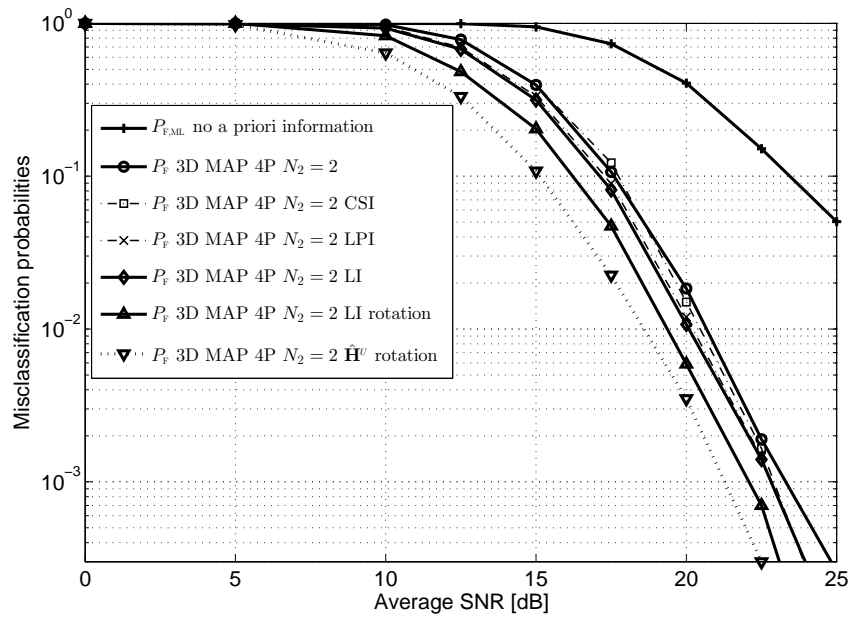
Figure 4.20: P_F for boundary condition and interpolation assisted classifier

Rotation of 16 QAM by 45°

Based on the data rate assisted and linear interpolation equipped 3D MAP 4P classifier, the classification performance is evaluated for the signal constellations with improved distinguishability. This improvement is achieved by simply rotating the signal points of 16 QAM clock-wise by 45°. This rotation does not scale the magnitude of the signal points and consequently causes no change in the transmit power but provides a reasonable performance gain in P_F as confirmed in Figure 4.20. Consider again a PER $P_F = 0.1$, about 1 dB can be saved by this rotation. To analyze the effect of the mismatch between the channel state $\hat{\mathbf{H}}^D$ used to describe the prior knowledge and the channel state $\hat{\mathbf{H}}^U$ used at the BS to perform the adaptive modulation, simulations are conducted based on the assumption that $\hat{\mathbf{H}}^U$ were available at the MS. The classification performance can be further improved, especially in the moderate SNR region. The corresponding classification performance of P_S is provided in the Appendix H.

End-to-end PER

In OFDM systems with adaptive modulation on per-subcarrier basis, AMC can be applied to blindly detect the modulation format applied on each subcarrier. A detection error on any subcarrier would lead to a severe bit mismatch in the channel decoding and consequently a packet error, so that the end-to-end packet error ratio (PER) P_T degrades by the classification error probability.

We simulate the impact of using AMC on P_T for the developed modulation detection

algorithms as given in Figure 4.21. For comparison purpose, we include the end-to-end PER performance of the reference system specified in the IEEE 802.11a standard given in Appendix A, which does not apply any adaptive modulation (Non AM) and naturally requires no modulation classification. This reference system outperforms its corresponding adaptive system with the pure ML-based modulation classifier implying poor reliability of this kind of classifier in adaptive systems. Besides, the pure ML-classifier causes also high implementation complexity which is undesirable in real-time applications. This fact leads to the 4-point approximation (Pure 4-P) based classifier, which on the one side reduces the complexity drastically and on the other side even improve the PER performance as certain prior information is implicitly utilized in this method. Interesting to note that the pure 4-P classifier also outperforms the reference system, around 1.5 dB gain can be achieved for $P_{\mathrm{T}} = 0.1$. This shows the potential benefits of the application of adaptive modulation in WLAN-based systems in combination with an implementally efficient 4-P approximation based modulation classifier.
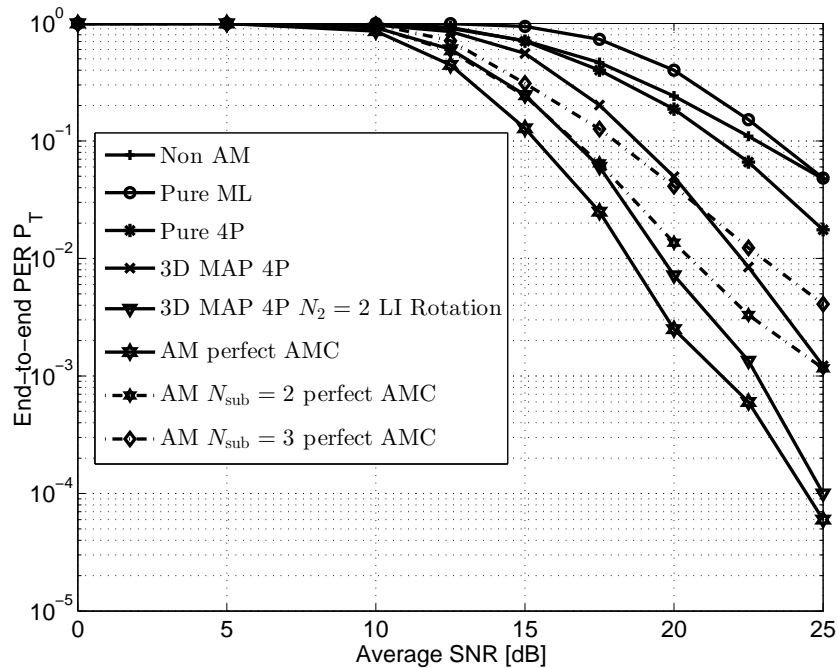
Figure 4.21: P_{T} based on proposed AMC and perfect AMC with subcarrier and subband adaptation

If more prior information is incorporated in the 4-P classifier in forms of the channel estimate of the considered subcarrier $|\hat{H}_n^{\mathrm{D}}|$, the average channel power gain $\mu_{\mathrm{H}}^2 = \frac{1}{N}\sum_{n=0}^{N-1}|\hat{H}_n^{\mathrm{D}}|^2$ and the mean squared channel variance $\sigma_{\mathrm{H}}^2 = \frac{1}{N}\sum_{n=0}^{N-1}(|\hat{H}_n^{\mathrm{D}}|^2 - \mu_{\mathrm{H}}^2)^2$ (3-D MAP 4-P), the PER performance can be further significantly improved. For $P_{\mathrm{T}} = 0.1$, round 2.5 ... 3 dB gain is recorded compared to the pure 4-P based classifier. By

considering other system-specific features, which can be utilized to improve the PER performance like the fixed data rate, the channel interpolation and the rotation of certain signal constellation (3-D MAP 4-P $N_2 = 2$, LI Rotation), the PER is further enhanced to a remarkable extend. A further 2 ... 2.5 dB gain is registered for $P_T = 0.1$.

As the upper bound of P_T, we provide the end-to-end PER $P_{c,1}$ for the subcarrier-based adaptive system with perfect information about the adapted modulation schemes at the receiver, or in other words packet errors are caused only by erronous data decoding. Our proposed classifier causes certain performance degradation, e.g., around 1 dB for $P_T = 0.1$. But this degradation shrinks with increasing SNR value.
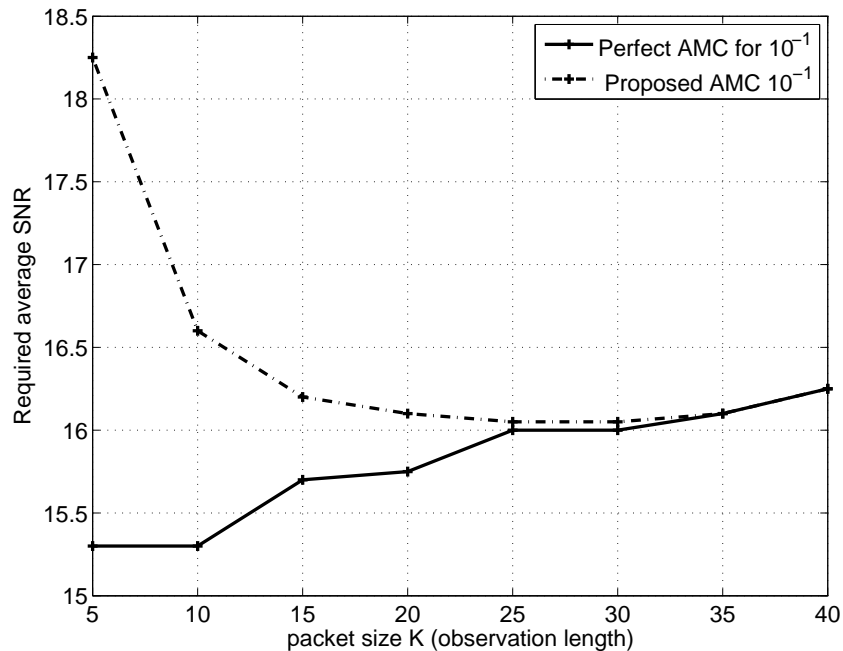
Figure 4.22:   Required system average SNR versus packet size K for different P_T

We also want to point out that the investigation is based on a short observation length ($K = 10$) aiming to study the worst case scenarios. With increasing K, the performance degradation caused by AMC will decrease as shown in Figure 4.22, where the above classifier denoted by 3-D MAP 4-P $N_2 = 2$ LI Rotation is used. We choose again a typical PER value of $P_T = 10^{-1}$. As expected, the required average SNR value (transmit-side SNR $\frac{E_l[|a_{i,l}|^2]}{N_0}$) falls with growing K as the classification reliability increases with K. From a certain value of $K = 25$ on, the degradation becomes negligible. Note that here synchronization impairments are not considered and pilot tones are not used to track the time variance of the channel. These facts will obviously impact the performance, but the general profile will remain. So it is a task of the system designer to find a proper K value

to meet the system requirements like real-time property, transmit power, computational complexity and memory requirements.

In Figure 4.21, we also provide $P_{c,N_{sub}}$ for the subband-based adaptive system with receiver-side perfect information about the adapted modulation schemes on subbands, where N_{sub} denotes the number of subcarriers in a subband using the same modulation format. In general, by grouping subcarriers into a subband, classification reliability can be improved as the number of observed symbols is increased from K to $N_{sub}K$. So in literature, most of classifiers are based on subband adaptation. However, subband adaptation degrades frequency diversity ultilized by the adaptive modulation, especially if the total number of subcarriers N is not large (or the subcarrier spacing $B_{sub} = \frac{B}{N}$ is large) as in our system. To minimize P_T, which may result either from classification error or the subsequent data decoding error, the subband size N_{sub} has to be optimized for the applied classifier. We observe that our proposed classifier outperforms the subband-based adaptive system with perfect AMC already for $N_{sub} = 2$ giving no reason for subband adaptation. Although classification performance is improved by increasing N_{sub}, the end-to-end PER P_T is not decreased, which indicates the high classification reliability of the proposed classification algorithm. More simulation results for some other mentioned classifiers can be found in Appendix H.

CHAPTER 5

Comparison of signalling with automatic modulation classification

In OFDM systems with adaptive modulation, the adapted modulation schemes have to be provided at the receiver via signalling as a traditional approach. Alternatively, automatic modulation classification can be applied to eliminate this additional signalling overhead as extensively illustrated in Chapter 4. We now compare these two approaches in terms of the end-to-end PER P_{T} for the considered OFDM system with adaptive modulation.

5.1 Parameter setting

To ensure a fair comparison, the following two conditions have to be fulfiled:

1. The total packet size is identical for both cases. So it requires:

$$K_{\mathrm{S}} + K_{\mathrm{DS}} = K, \tag{5.1}$$

   where K_{S} and K_{DS} are the number of OFDM symbols for signalling symbols and payload symbols under explicit signalling, respectively, and K is the number of OFDM symbols for data transmission under modulation classification as shown in Figure 5.1.

2. The net data rate must be equal:

$$\frac{B_{\mathrm{DS}} K_{\mathrm{DS}}}{K_{\mathrm{S}} + K_{\mathrm{DS}}} = B_{\mathrm{D}}, \tag{5.2}$$

   Where B_{DS} and B_{D} are the number of data bits carried per OFDM data symbol under signalling and modulation classification, respectively.

To be consistent with the previous parameter setting, B_{D} is set to 96 bits and the code rate of the convolutional channel coding scheme remains $\frac{1}{2}$. But K is modified to give a more reasonable data structure under signalling ($K_{\mathrm{DS}} = K - K_{\mathrm{S}} \gg K_{\mathrm{S}}$ shall be fulfilled in order to maintain an efficient utilization of bandwidth), which will be explained later in more detail.

5.1.1   Selection of signalling schemes

In Chapter 3, a dozen of signalling schemes have been proposed which result in different signalling overheads L_S. Using time-domain state-dependent Huffman coding based signalling schemes, L_S can be reduced to a very low level. However, the resulting L_S depends on time and the packet size $K_S + K_{DS}$. Furthermore, these signalling schemes produce a catastrophic error propagation in the time direction leading to a domino-effect-similar signalling failure. Consequently, they are normally not used in practical systems and also not further investigated in this thesis. Here we concentrate on the trivial signalling schemes with fixed L_S, the Huffman coding and the frequency-domain state-dependent Huffman coding based signalling schemes with variable L_S, which fortunately do not cause catastrophic error propagation in the time direction (no error propagation to subsequent packets). Concretely, the following three schemes are compared one after another with

K_P	K_S	K_{DS}

(a)

K_P	K

(b)

Figure 5.1:   Data structure with inband signalling (a); data structure without signalling (b)

the proposed automatic modulation classifier.

1. The trivial signalling scheme S_1 with joint coding with $N_g = 3$ giving a fixed signalling overhead of $L_S = 112$ bits.

2. The Huffman coding based signalling scheme S_2 with $N_g = 3$, too. The symbol probabilities $P_{b_{i_1}, b_{i_2}, b_{i_3}}$ are obtained through computer simulations as already explained in Section 3.2.2. This scheme results in a variable L_S. According to our computer simulation based on 10000 channel realizations, an overhead of around $L_S = 96$ bits are required to achieve a signalling probability of $P_S \approx 1$.

3. The signalling scheme S_3 based on frequency-domain state-dependent Huffman coding with $N_g = 3$ too. Here, the required symbol probabilities and transition prob-

abilities are determined also through computer simulations. This scheme reduces the overhead L_S to 87 bits by exploiting frequency-domain memory effect.

Note that all the three overheads are independent on the time variance of the channel (the mobility of the mobile terminal) and the packet size as no temporary memory effect is exploited here.

5.1.2   Determination of K_S

The length of the signalling field K_S is determined by L_S and the signalling rate r_S or in other words, the applied modulation scheme M_S and code rate R_SC for signalling bits. We restrict r_S, correspondingly M_S and R_SC to values which are specified in the IEEE 802.11a standard, as our investigations are based on this standard. Furthermore, we do not apply any adaptive techniques on the transmission of the signalling bits as this would, otherwise, itself require a signalling or classification process. To provide a well-protected transmission, we further restrict r_S to the low data rate transmission schemes as given in the Table 5.1. Note that in the standard each r_S corresponds to a fixed combination of M_S and R_SC.

Table 5.1: Possible signalling rate r_S based on IEEE 802.11a, the corresponding M_S, R_SC and the relationship between K_S, signalling resource unit U_S^* measured in time-frequency grids and signalling capability L_S' (maximum transmittable signalling bits) under each signalling rate.

r_S in Mbps	M_S	R_SC	$K_\mathrm{S} \rightarrow$	1	2	3	4	5
			$U_\mathrm{S}^* \rightarrow$	48	96	144	192	240
$r_\mathrm{S,1} = 6$	BPSK	1/2	$L_\mathrm{S}' \rightarrow$	24	48	72	96	120
$r_\mathrm{S,2} = 9$	BPSK	3/4	$L_\mathrm{S}' \rightarrow$	36	72	108	144	180
$r_\mathrm{S,3} = 12$	4 QAM	1/2	$L_\mathrm{S}' \rightarrow$	48	96	144	192	240
$r_\mathrm{S,4} = 18$	4 QAM	3/4	$L_\mathrm{S}' \rightarrow$	72	144	216	288	360

The signalling information has to be detected prior to detecting the data symbols implying that the signalling bits have to be separately channel encoded and decoded. As the standard specifies the convolutional codes with a memory length of $l_\mathrm{c} = 6$ as the channel coding scheme, 6 termination bits have to be added to the required signalling overhead $\tilde{L}_\mathrm{S} = L_\mathrm{S} + 6$, where $\tilde{L}_\mathrm{S}$ denotes the terminated signalling overhead. Obviously, to enable a successful signalling, the signalling capability L_S' of certain r_S and K_S has to fulfil

$$\tilde{L}_\mathrm{S} \leq L_\mathrm{S}'. \tag{5.3}$$

This relationship determines the required K_S for any combination of a selected signalling scheme and $r_{\mathrm{s},i}$, where $r_{\mathrm{s},i}$ with $i = 1, 2, 3, 4$ denotes the above mentioned 4 signalling rates, by searching for the corresponding minimum L'_S in Table 5.1. In other words, once the signalling scheme and r_s are selected, K_S is fixed. For an example, if S_1 and $r_{\mathrm{s},1}$ are used, from Table 5.1 it follows $K_\mathrm{S} = 5$, as 120 is the smallest number fulfilling (5.3) (120 $> (112 + 6) = 118$). For the considered signalling schemes and signalling rates, K_S are summerized in Table 5.2.

Table 5.2: Look-up table for K_S under certain signalling scheme and signalling rate.

Signalling scheme	L_S	$\tilde{L}_\mathrm{S}$	$r_{\mathrm{S},1}$	$r_{\mathrm{S},2}$	$r_{\mathrm{S},3}$	$r_{\mathrm{S},4}$
S_1	112	118	$K_\mathrm{S} = 5$	$K_\mathrm{S} = 4$	$K_\mathrm{S} = 3$	$K_\mathrm{S} = 2$
S_2	96	102	$K_\mathrm{S} = 5$	$K_\mathrm{S} = 3$	$K_\mathrm{S} = 3$	$K_\mathrm{S} = 2$
S_3	87	93	$K_\mathrm{S} = 4$	$K_\mathrm{S} = 3$	$K_\mathrm{S} = 2$	$K_\mathrm{S} = 2$

5.1.3 Determination of B_DS

In the previous investigations of applicability of automatic modulation classification, we simulated an almost worst case scenario with a small packet size of $K = 10$. This packet size would imply a portion of $K_\mathrm{DS} = K - K_\mathrm{S} \approx \frac{K}{2}$ just for signalling. In other words, almost half of the bandwidth efficiency is wasted by signalling bits. So, in this chapter we first simulate a packet size of $K = 20$. For the above example of $K_\mathrm{S} = 5$, it results in $K_\mathrm{DS} = K - K_\mathrm{S} = 15$ and consequently an increase in the number of carried data bits per OFDM symbol from 96 bits to

$$B_\mathrm{DS} = 96\frac{K}{K_\mathrm{DS}} = 96\frac{20}{15} = 128 \text{ bits.} \tag{5.4}$$

The parameter B_DS is uniquely determined by K and $K_\mathrm{DS} = K - K_\mathrm{S}$, while K_S is in turn determined by the selected signalling scheme and signalling rate r_s. The possible B_DS values are given in Table 5.3 for $K = 20$ and possible K_S values resulting from Table 5.2. Note that B_DS has to be a integer number. The data bits B_DS under signalling are also channel-coded with the same code rate $R_\mathrm{C} = 1/2$ as that under modulation classification so that the number of transmit bits is $B'_\mathrm{DS} = 2B_\mathrm{DS}$. The fact, that under signalling an OFDM symbol has to carry more data bits $B_\mathrm{DS} > B_\mathrm{D}$ or after channel coding $B'_\mathrm{DS} > B$, results from the fairness conditions mentioned before, which says that the net data rate has to be maintained the same for both approaches.

5.1.4 Overview of K_S, B_DS and r_s

As mentioned above, once K, the signalling scheme and r_s are selected, all other parameters like K_S and B_DS can be determined as shown in Table 5.4.

Table 5.3: Look-up table of B_{DS} for $K = 20$; the last column represents the signalling-free case (automatic modulation classification).

K_{S}	5	4	3	2	0
B_{DS}	128	120	113	107	96
B'_{DS}	256	240	226	214	192

Table 5.4: Look-up table of K_{S} and B_{DS} for $K = 20$ by fixing $r_{\mathrm{S},i}$ and S_j.

	$r_{\mathrm{S},1}$	$r_{\mathrm{S},2}$	$r_{\mathrm{S},3}$	$r_{\mathrm{S},4}$
S_1	$K_{\mathrm{S}} = 5$, $B_{\mathrm{DS}} = 128$	$K_{\mathrm{S}} = 4$, $B_{\mathrm{DS}} = 120$	$K_{\mathrm{S}} = 3$, $B_{\mathrm{DS}} = 113$	$K_{\mathrm{S}} = 2$, $B_{\mathrm{DS}} = 107$
S_2	$K_{\mathrm{S}} = 5$, $B_{\mathrm{DS}} = 128$	$K_{\mathrm{S}} = 3$, $B_{\mathrm{DS}} = 113$	$K_{\mathrm{S}} = 3$, $B_{\mathrm{DS}} = 113$	$K_{\mathrm{S}} = 2$, $B_{\mathrm{DS}} = 107$
S_3	$K_{\mathrm{S}} = 4$, $B_{\mathrm{DS}} = 120$	$K_{\mathrm{S}} = 3$, $B_{\mathrm{DS}} = 113$	$K_{\mathrm{S}} = 2$, $B_{\mathrm{DS}} = 107$	$K_{\mathrm{S}} = 2$, $B_{\mathrm{DS}} = 107$

5.1.5   Selection of the classification scheme

In Chapter 4, we proposed also a dozen of modulation classification schemes showing different classification reliability and implementation complexity. In general, the application of blind detection techniques in adaptive systems increases the bandwidth efficiency as signalling always wastes certain transmission resources. Furthermore, signalling is only possible in cooperative adaptive systems. In many situations like in military applications and intelligent systems, where the terminal of interest is not the desired receiver, the modulation schemes can only be detected blindly. So, automatic modulation classification shows a higher flexibility and therefore is a more universal method. On the other hand, blind modulation detection requires in general much more implementation complexity, especially the likelihood-based detection techniques. Besides, it is typically less reliable especially in real-time applications where the observation interval is short.

Based on the system-specific properties, we proposed an automatic modulation classifier, which approximates the computationally intensive likelihood function by an implementation-efficient 4-point approximation (4-P). To increase the classification reliability, the existing prior knowledge about the adaptation strategy is incorporated into the classifier (3-D MAP). Furthermore, some system boundary conditions like the knowledge about the data rate r_{D} ($N_2 = 2$), the application of a linear interpolation technique (LI) and the rotation of certain signal constellation (rotation) are taken into account to further enhance the classification reliability. This classifier is denoted as 3-D MAP 4-P $N_2 = 2$ LI Rotation and balances well between the two conflicting requirements, namely classification performance and implementation complexity as conformed in Chapter 4. This classifier will be used to compare with the previously described signalling schemes.

5.2 Computer simulations

Computer simulations are conducted to compare signalling with AMC in terms of the end-to-end PER P_T. Since the data rate is constant and the packet size is the same for both cases as formulated in Section 5.1, due to the additional bandwidth requirement for signalling bits, modulation schemes of higher bandwidth efficiency (higher modulation levels) have to be used in case of signalling compared to modulation classification. As comparison reference, we also provide P_T for fixed modulation (non AM), i.e., all sub-carriers use the same modulation scheme (16 QAM for the simulated data rate $r_\mathrm{D} = 24$ Mpbs) and the performance upper bound (AM perfect AMC), namely the end-to-end PER with adaptive modulation under perfect signalling or modulation classification in Figure 5.2. In reality, a signalling failure leads to a packet error as data symbols cannot be demodulated correctly. Consequently, any signalling scheme will degrades P_T to certain extend.
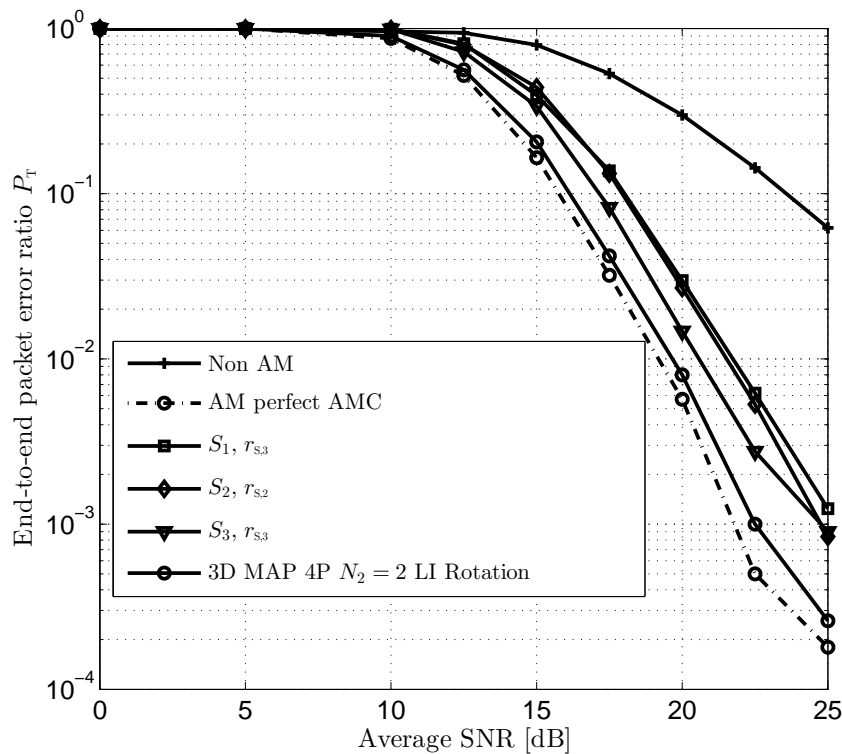
Figure 5.2: P_T comparison between signalling and automatic modulation classification

We restrict the signalling scheme to the above three methods S_j with $j = 1, 2, 3$, each of which can be combined with one of the 4 signalling rates $r_{\mathrm{s},i}$ with $i = 1, 2, 3, 4$. It results in $3 \times 4 = 12$ signalling strategies, which are uniquely identified by the parameter pair $(S_j, r_{\mathrm{s},i})$ (see Table 5.4). The scheme S_1 does not require any prior information about the signalling source like symbol probabilities and therefore is mostly applied in practical

systems. The resulting signalling bits can be transmitted with any $r_{\mathrm{S},i}$. In general, there exists a performance trade-off in selecting $r_{\mathrm{S},i}$. Lower $r_{\mathrm{S},i}$ provides more protection for the signalling bits but results in less protection for data bits as data symbols have to be modulated by higher modulation levels on average due to higher B_{DS} and vice versa. For each S_j, an optimum $r_{\mathrm{S,opt}}$ can be found which delivers the best PER performance. For S_1, the parameter pair $(S_1, r_{\mathrm{S},3})$ gives the best PER performance. The scheme S_2 reduces the overhead and potentially also the length of signalling fields K_{S} as confirmed in Table 5.2. Using S_2, B_{DS} can be reduced as shown in Table 5.4. But S_2 requires the knowledge about the symbol probabilities and is consequently less flexible. The pair $(S_2, r_{\mathrm{S},2})$ delivers the best PER and shows similar performance to that of $(S_1, r_{\mathrm{S},3})$. Finally, the scheme S_3 reduces the overhead further by additionally requiring transition probabilities. But the pair $(S_3, r_{\mathrm{S},3})$ shows a reasonable improvement in PER performance compared to $(S_1, r_{\mathrm{S},3})$ and $(S_2, r_{\mathrm{S},2})$. Around 0.8 dB gain is observed for $P_{\mathrm{F}} = 0.1$. The PER performance of other siganlling strategies are provided in Appendix I. All the signalling schemes applied on the adaptive system show a significant performance improvement compared to the fixed modulation but at the same time, a remarkable degradation with respect to the performance upper bound as shown in Figure 5.2.

AMC is applied to replace signalling and consequently reduce B_{DS} to its original value $B_{\mathrm{D}} = 96$ bits. Using the proposed classifier, around 1 dB gain is observable for $P_{\mathrm{F}} = 0.1$ compared to the performance of the best signalling strategy $(S_3, r_{\mathrm{S},3})$. This result demonstrates that from the viewpoint of PER, AMC is able to outperform the conventional signalling philosophy. However, the classifier shows also a slight degradation compared to the upper bound, which can be compensated by more sophisticated classification algorithms proposed in Chapter 4.

We want to point out here that the performance relationship or the observed gain may be shifted if another parameter constellation is simulated, e.g., another data rate r_{D} (B_{D}) or another packet size K. Generally, with increasing K the bit overload due to signalling $\Delta B_{\mathrm{DS}} = B_{\mathrm{DS}} - B_{\mathrm{D}}$ is reduced. The performance gap between the signalling-based performance and the upper bound will be decreased. This holds also for the modulation classification based performance as the classification reliability increases with increasing observation interval. In Appendix I, we provide some simulation results for a smaller packet size of $K = 10$, which was the parameter setting for the previous investigations.

CHAPTER 6

Conclusion

We have investigated adaptive QAM for a WLAN-based OFDM system. Under the ideal condition, that the adapted modulation level is perfectly synchronized between the transmitter and the receiver, a significant performance improvement in error probability was observed over the system with fixed modulation. Then, two realistic modulation detection strategies were addressed: explicit signalling and automatic modulation classification. The impact thereof on the system performance and implementation complexity was extensively analyzed. Finally, these two approaches are compared in terms of end-to-end PER.

Explicit signalling is a conventional method to provide the receiver with unknown transmission parameters. Due to the large number of subcarriers, signalling modulation scheme on per-subcarrier basis would occupy a large portion of transmission resources. It was shown in the considered OFDM system, 144 bits were required if trivial source coding scheme was used. We proposed a joint coding scheme to reduce the signalling overhead, which groups neighbouring subcarriers into a joint symbol. The overhead could be reduced to 112 bits by grouping 3 subcarriers. Higher overhead reduction could be achieved by exploiting information theoretical properties of the source. Specifically, we utilized the memory effects resulting from the correlation properties of the radio channel existing in both time and frequency domain. Computer simulations have shown that the utilization of the frequency-domain memory effect could reduce the overhead to the half (around 60 bits), while by exploiting the time-domain memory effect the overhead could be even reduced to one third (about 40 bits). We also applied the proposed signalling schemes to signal the adapted modulation schemes. Compared to the ideal condition, a high degradation was abserved, while compared to the fixed modulation scheme, around 5 dB gain was achieved for an end-to-end PER of $P_{\mathrm{T}} = 0.1$ and a packet size of $K = 20$. Even by a joint coding scheme with 3 subcarriers in a group and without any utilization of memory effect, still about 4 . . . 4.5 dB gain was observed compared to the system with fixed modulation. These results show potential advantages of using adaptive modulation combined with explicit signalling.

As a signalling-free alternative, AMC was investigated. This approach was addressed from two different concepts namely: a likelihood-based and a feature-based framework.

It was shown that the feature-based modulation classification was not suitable for the problem stated here mainly due to two reasons: 1, the modulation candidates stem from the same modulation class resulting in nested signal constellations; 2, the observation time is too short to build a reliable discriminating feature. The likelihood-based concept could provide significantly better classification performance. Here, emphasis was put on reducing the implementation complexity of the computationally intensive likelihood function. Based on properties of an exponential function, one-point approximation and Euclidean distance approximation were analyzed. Those two simplifications take only a single signal point from each modulation candidate into account and consequently present a too strong simplification especially for high level modulation schemes like 64 QAM. Next, based on the special arrangement of the signal constellations, a 4-point approximation method was developed. The 4-point approximation considers 4 signal points and therefore reduces the simplification error drastically. Furthermore, we observed that the one-point and 4-point approximations not only enormously reduced the complexity but also improved the classification reliability as they implicitly incorporated the prior information into the decision rule. Driven by this phenomenon, we formulated the general MAP-based modulation classification framework, which involved the N-dimensional channel magnitude vector. Next, proposals were developed to reduce the dimensionality to enable practically feasible implementations. Based on the strategy of the used adaptive modulation, one-dimensional (1-D), two-dimensional (2-D), and three-dimensional (3-D) approximations were derived to simplify the N-dimensional density function.

We exploited system-specific properties like a given constant data rate, channel reciprocity to further enhance the classification reliability. It was also shown that QAM constellations become more distinguishable if the nested signal structure is distorted by rotating certain constellation, which improved also the classification performance. A classifier, which utilized these system specific measures, was able to perform highly reliable modulation classification already for a small packet size of $K = 10$. The resulting performance degradation measured in the end-to-end PER P_T was shown to be almost negligible (for $P_\mathrm{T} = 0.1$): around 1 dB for $K = 10$, around $0.2 \ldots 0.5$ dB for $K = 20$ and negligible for $K > 25$, respectively.

We also compared the above two approaches: signalling and AMC under the condition that the net data rate is constant. We found out that both approaches outperformed the fixed modulation scheme and modulation classification in turn achieved a remarkable improvement compared to its signalling-based alternative: around $1 \ldots 1.5$ dB gain for $K = 20$ and 2 - 2.5 dB gain for $K = 10$ at $P_\mathrm{T} = 0.1$.

The results achieved here show potential benefits of using adaptive modulation in combi-

nation with AMC in OFDM systems. We want to point out that in this thesis we assumed perfect time and frequency synchronization. It would be interesting to investigate the effect of synchronization impairments on the proposed modulation classification algorithms. Further investigations may be conducted to test the algorithms in multiple-user scenarios, where interference would play a role. Also new technologies like multiple input and multiple output (MIMO), which has been introduced in WLAN standards recently, may inspire new design concepts. Some investigations have already been conducted in this area [91], [92].

APPENDIX A

Simulation system

The simulated OFDM system is mainly based on the IEEE 802.11a standard and is schematically represented in Figure A.1 [87]. The k-th block of data stream $\mathbf{d}_k = (d_{k,0}, d_{k,1}, \cdots, d_{k,N-1})^{\mathrm{T}}$ is mapped onto the QAM symbols $\mathbf{s}_k = (s_{k,0}, s_{k,1}, \cdots, s_{k,N-1})^{\mathrm{T}}$ based on a bit loading vector $\mathbf{b} = (b^{(0)}, b^{(1)}, \cdots, b^{(N-1)})^{\mathrm{T}}$, which hold for blocks $k = 0, 1, \cdots, K-1$ with the following notations:

- k: block index or OFDM symbol index.

- K: number of blocks in a transmit packet.

- n: subcarrier index.

- N: number of data subcarriers, N_{P}: number of pilot tones, N_{r}: number of subcarriers in guard band and $N_{\mathrm{F}} = N + N_{\mathrm{P}} + N_{\mathrm{r}}$: FFT/IFFT window size.
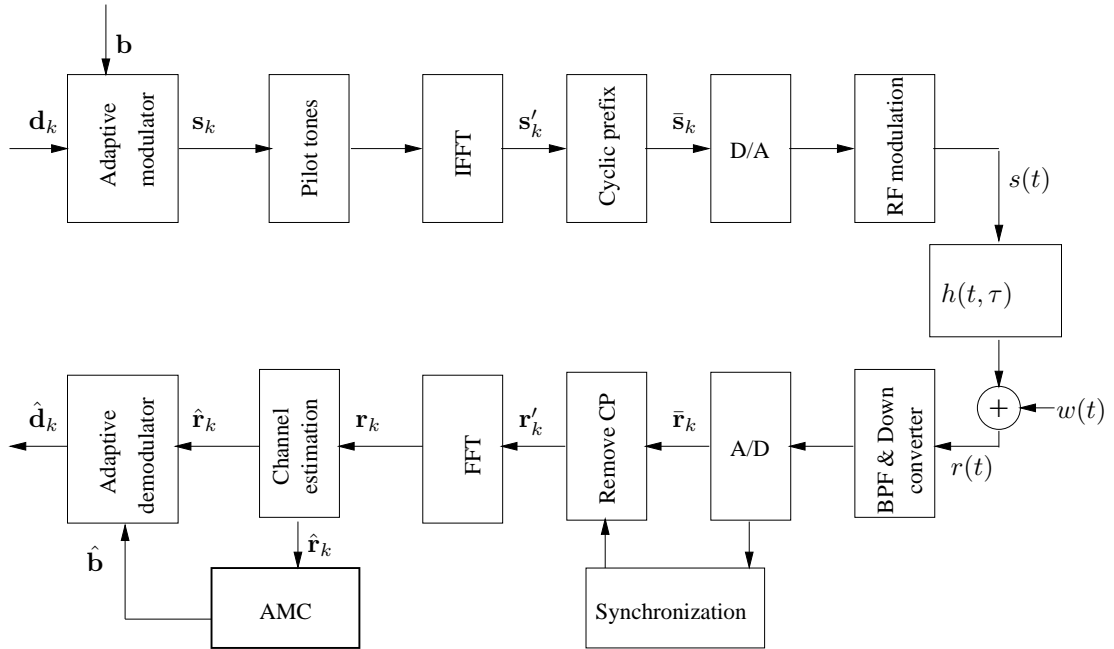
Figure A.1: System block diagrom

Based on the Gray-coding scheme specified in the standard [87], the binary bit sequence $(d_{k,n}, \cdots, d_{k,n+b^{(n)}-1})$ is mapped onto the symbol $s_{k,n}$ stemming from the signal point set

$$s_{k,n} \in \{a_{i,1}, a_{i,2}, \cdots, a_{i,M_i}\}, \tag{A.1}$$

of the modulation level b_i with $b^{(n)} = b_i$, where the number of bits carried in $s_{k,n}$ is equal to $b_i = \log_2(M_i)$. Then pilot symbols are inserted onto the four pilot tones for synchronization and channel tracking purposes, which are not exploited in this thesis. The symbols are modulated on orthogonal subcarriers via IFFT

$$s'_{k,n'} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} s_{k,n} e^{j2\pi \frac{nn'}{N}} \quad \text{for } n' = 0, 1, \cdots, N-1. \tag{A.2}$$

The modulated symbol block is extended with a cyclic prefix of length K_{cp} to avoid inter-symbol and inter-carrier interference

$$s'_{k,n'} = s'_{k,N_{\text{F}}-n'+1} \quad \text{for } n' = -K_{\text{cp}}, -(K_{\text{cp}} - 1), \cdots, -1, \tag{A.3}$$

with $\bar{\mathbf{s}}_k = (s'_{k,-K_{\text{cp}}}, \cdots, s'_{k,0}, s'_{k,1}, \cdots, s'_{k,N-1})$. Then, the digital symbols are converted to analog signals and up-converted to the transmission band around the center frequency f_{c}.

Here we consider a single-cell and single-user environment. Possible interference from other cells and users is not treated. Furthermore, the following assumptions hold until stated others:

- perfect time and frequency synchronization and,

- the noise power spectrum density N_0 has been estimated beforehand.

The base station performs the calculation of $\mathbf{b}$ (adaptive modulation) as described in Section 2.2.1 for the downlink transmission; the mobile station applies AMC algorithms to blindly detect the adapted modulation levels. The modulation candidates are given in Table A.1. Besides the modulation format no transmission (NoTx), which is adopted to handle the deeply faded subcarriers, the remaining four modulation formats have been already implemented in the standard.

Table A.1: Modulation candidates for adaptive modulation

Modulation candidates A_i	NoTx	BPSK	4-QAM (QPSK)	16-QAM	64-QAM
Modulation level b_i	0	1	2	4	6

The equivalent baseband propagation channel model of $h(\tau;t)$ is given by a time-varying multipath Rayleigh fading channel

$$\tilde{h}_{\mathrm{b}}(\tau;t) = \sum_{p_{\mathrm{k}}=0}^{P_{\mathrm{k}}-1} \tilde{h}_{p_{\mathrm{k}}}(t)\delta(\tau - \tau_{p_{\mathrm{k}}}(t)), \tag{A.4}$$

where $\tilde{h}_{p_{\mathrm{k}}}(t)$ is the p_{k}-th complex-valued path coefficients with statistically independent and Gaussian distributed real and imaginary part. The delays $\tau_{p_{\mathrm{k}}}(t)$ are assumed to be time-invariant and equi-distantly located at

$$\tau_{p_{\mathrm{k}}}(t) = \tau_{p_{\mathrm{k}}} = p_{\mathrm{k}}T_{\mathrm{s}}, \tag{A.5}$$

with T_{s} the system sampling period. We simulate a typical exponentially decaying power delay profile [93]

$$\mathrm{E}[|\tilde{h}_{p_{\mathrm{k}}}|^2] = \frac{1-\alpha_{\mathrm{k}}}{1-\alpha_{\mathrm{k}}^{P_{\mathrm{k}}}}\alpha_{\mathrm{k}}^{p_{\mathrm{k}}}, \quad 0 \le p_{\mathrm{k}} \le P_{\mathrm{k}} - 1 \tag{A.6}$$

with $0 < \alpha_{\mathrm{k}} < 1$ as the decaying parameter as specified for IEEE 802.11a indoor applications [87]. Furthermore, we observe that the power delay profile is normalized so that it holds

$$\sum_{p_{\mathrm{k}}=0}^{P_{\mathrm{k}}-1} \mathrm{E}[|\tilde{h}_{p_{\mathrm{k}}}|^2] = 1. \tag{A.7}$$

Depending on application scenarios, $\alpha_{p_{\mathrm{k}}}$ is assigned with different values. Here we consider a typical large office enviroment with

$$\alpha_{p_{\mathrm{k}}} = \mathrm{e}^{-\frac{1}{3}}, \tag{A.8}$$

corresponding to an average normalized root-mean-squared (rms) delay spread of (normalized on the sampling period) [94]

$$\sigma_\tau \approx \frac{\sqrt{\alpha_{p_{\mathrm{k}}}}}{1-\alpha_{p_{\mathrm{k}}}} \approx 3. \tag{A.9}$$

With $T_{\mathrm{s}} = 50$ ns, the rms delay spread $\tau_{\mathrm{rms}} = 150$ ns corresponds to the Model C specified in the standard [95].

The time-variance of the channel is described by a Doppler spectrum based on the Jakes' model with a maximum Doppler frequency $f_{\mathrm{D}} = 55.56$ Hz, which corresponds to a maximum mobility speed of $v_{\mathrm{max}} = 3.33$ m/s for a center frequency of $f_{\mathrm{c}} = 5$ GHz.

A preamble-based least square (LS) estimator is used to estimate the channel

$$\hat{\mathbf{h}} = \operatorname*{argmax}_{\mathbf{h}} (\mathbf{y} - \mathbf{X}\mathbf{N}_{\mathrm{DFT}}\mathbf{h})^{\mathrm{H}}(\mathbf{y} - \mathbf{X}\mathbf{N}_{\mathrm{DFT}}\mathbf{h}), \tag{A.10}$$

where $(\cdot)^{\mathrm{H}}$ is the Hermitian transpose operation; $\mathbf{X}$ is a diagonal matrix with the elements of $\mathbf{x}$ on its diagonal; $\mathbf{x}$ is the transmitted preamble vector whose values are specified in the standard; $\mathbf{y}$ is the received preamble vector; $\mathbf{N}_{\mathrm{DFT}}$ is the N point DFT matrix and $\mathbf{h}$ is the channel impulse response vector. The channel estimate is given by [96], [97]

$$\hat{\mathbf{h}}_{\mathrm{LS}} = \mathbf{N}_{\mathrm{DFT}} \left(\mathbf{N}_{\mathrm{DFT}}^{\mathrm{H}} \mathbf{X}^{\mathrm{H}} \mathbf{X} \mathbf{N}_{\mathrm{DFT}} \right)^{-1} \mathbf{N}_{\mathrm{DFT}}^{\mathrm{H}} \mathbf{X}^{\mathrm{H}} \mathbf{y}. \qquad (A.11)$$

The time variance of $\mathbf{h}(t)$ can be tracked by pilot-based channel estimation and interpolation techniques which is not considered in this thesis as we assume small packet size and low time-variance.

At the receiver side after band pass filtering (BPF), down-conversion and analog to digital conversion (A/D), synchronization is performed to correct the symbol timing offset, which is necessary for the removal of the cyclic prefix. In this thesis we assume perfect synchronization. The time-domain symbols $\tilde{\mathbf{r}}_k$ are transformed to the frequency domain symbols $\mathbf{r}_k$ via DFT and the channel distortion is equalized using the estimated channel state described above. The equalized symbols $\hat{\mathbf{r}}_k$ are input to AMC, which outputs a detected bit loading vector $\hat{\mathbf{b}}$, which is in turn used for demodulating $\hat{\mathbf{r}}_k$.

Further relevant system parameters are listed in Table A.2. Here $B = 192$ corresponds

Table A.2:   Baseband system parameters

bandwidth in MHz f_{BW}	20	FFT window N_{F}	64
Number of pilot N_{P}	4	Number of data subcarriers N	48
Frame size K	10	Number of bit per OFDM symbol B	192
Guard Interval length K_{cp}	16	Number of multipaths P_{k}	16
Max. Dopp. freq. in Hz f_{D}	55.56	Mobility speed in m/s v_{max}	3.33

to one standardized data rate $r_{\mathrm{D}} = 24$ Mbps. We consider a packet size of $K = 10$ OFDM data symbols and a subcarrier-by-subcarrier and packet-based adaptive modulation, which means the receiver is allowed to observe 10 symbols to make a decision. Obviously, the more symbols can be observed, the more reliable is the modulation classification. The packet size is not fixed in IEEE 802.11a. There is a 12-bits $LENGTH$ indicator in the PLCP header, which indicates how many bits are transmitted in a packet. Based on the $LENGTH$ value, the maximum packet size can be determined for standardized data rates as in Table A.3 [87]. In practical systems, K is limited due to synchronization impairments and time-variance of the channel, so that K is much smaller than the maximum allowable values. Furthermore, K depends also on the data service class, e.g., real-time applications allow in general smaller K than off-line applications. Here we intend to simulate real time applications and choose a small packet size of $K = 10$

to demonstrate the potential of applying AMC in WLAN-based systems with adaptive modulation.

Table A.3: Possible transmission modes and the corresponding parameters [87]

r_{D} in Mbps	Modulation	Code rate	Coded bits per OFDM symbol	Max. number of OFDM symbol per packet
6	BPSK	1/2	48	1366
9	BPSK	3/4	48	910
12	4-QAM	1/2	96	683
18	4-QAM	3/4	96	455
24	16-QAM	1/2	192	341
36	16-QAM	3/4	192	227
48	64-QAM	1/2	288	170
54	64-QAM	3/4	288	151

APPENDIX B

State-independent Huffman coding

The table is divided into two parts. The first part is given in Table B.1.

Table B.1: Group size $N_{\mathrm{g}} = 2$, symbol pairs (b_{i_1}, b_{i_2}), symbol probabilities $P_{b_{i_1}, b_{i_2}}$, code word length W_{H, i_1, i_2} and average code word length $\bar{W}_{\mathrm{H}, 2}$

	perfect CE		20 dB		0 dB	
(b_{i_1}, b_{i_2})	$P_{b_{i_1}, b_{i_2}}$	W_{H, i_1, i_2}	$P_{b_{i_1}, b_{i_2}}$	W_{H, i_1, i_2}	$P_{b_{i_1}, b_{i_2}}$	W_{H, i_1, i_2}
(0, 0)	**0.0068**	**7**	**0.0067**	**6**	**0.0050**	**7**
(0, 1)	0.0048	7	0.0049	7	0.0041	8
(0, 2)	0.0138	6	0.0137	6	0.0147	6
(0, 4)	0.0020	9	0.0021	8	0.0047	7
(0, 6)	4e-5	13	4e-5	12	0.0001	11
(1, 0)	0.0048	7	0.0048	7	0.0040	8
(1, 1)	**0.0051**	**7**	**0.0051**	**7**	**0.0042**	**7**
(1, 2)	0.0130	6	0.0131	6	0.0122	6
(1, 4)	0.0032	8	0.0032	8	0.0054	7
(1, 6)	0.0001	12	6e-5	12	0.0002	11
(2, 0)	0.0139	5	0.0141	6	0.0146	6
(2, 1)	0.0127	6	0.0129	6	0.0125	6

On the next page, the second part is given in Table B.2.

Table B.2: Group size $N_{\mathrm{g}} = 2$, symbol pairs (b_{i_1}, b_{i_2}), symbol probabilities $P_{b_{i_1}, b_{i_2}}$, code word length W_{H, i_1, i_2} and average code word length $\bar{W}_{\mathrm{H},2}$

	perfect CE		20 dB		0 dB	
(2, 2)	**0.0957**	**3**	**0.0949**	**3**	**0.0851**	**3**
(2, 4)	0.0626	4	0.0626	4	0.0719	4
(2, 6)	0.0008	11	0.0009	9	0.0028	9
(4, 0)	0.0021	8	0.0021	8	0.0043	7
(4, 1)	0.0031	8	0.0032	7	0.0056	7
(4, 2)	0.0624	4	0.0627	4	0.0716	4
(4, 4)	**0.3544**	**2**	**0.3538**	**2**	**0.3215**	**2**
(4, 6)	0.0590	4	0.0592	4	0.0712	4
(6, 0)	2e-5	14	3e-5	12	0.0001	11
(6, 1)	4e-5	14	4e-5	12	0.0001	11
(6, 2)	0.0009	10	0.0009	10	0.0028	8
(6, 4)	0.0590	4	0.0588	4	0.0714	4
(6, 6)	**0.2197**	**2**	**0.2199**	**2**	**0.2089**	**2**
$\bar{W}_{\mathrm{H},2}$	2.9695		2.9738		3.1134	

114

APPENDIX C

State-dependent Huffman coding

C.1   Time-domain correlation

Table C.1: $P_{Z_{k_\mathrm{s}}^{(n_\mathrm{g})}|Z_{k_\mathrm{s}-1}^{(n_\mathrm{g})}}$, $W_{\mathrm{SDH},1}$ given in bold and $\bar{W}_{\mathrm{SDH},1}$ in bits per symbol for $N_\mathrm{g} = 1$ and perfect channel estimation

$Z_{k_\mathrm{s}-1}^{(n_\mathrm{g})}$ \ $Z_{k_\mathrm{s}}^{(n_\mathrm{g})}$	0	1	2	4	6
0	0.8976 **1**	0.092 **2**	0.0103 **2**	0 **0**	0 **0**
1	0.0976 **2**	0.7516 **1**	0.150 **2**	0 **0**	0 **0**
2	0.0015 **3**	0.0224 **3**	0.9417 **1**	0.0342 **2**	0 **0**
4	0 **0**	0 **0**	0.0134 **2**	0.9724 **1**	0.0141 **2**
6	0 **0**	0 **0**	0 **0**	0.0235 **1**	0.9764 **1**
$\bar{W}_{\mathrm{SDH},1} = 1.0377$					

Table C.2: $P_{Z_{k_s}^{(n_g)}|Z_{k_s-1}^{(n_g)}}$, $W_{\text{SDH},1}$ given in bold and $\bar{W}_{\text{SDH},1}$ in bits per symbol for $N_g = 1$ and SNR = 0 dB

$Z_{k_s-1}^{(n_g)}$ \ $Z_{k_s}^{(n_g)}$	0	1	2	4	6
0	0.1495 **3**	0.1192 **4**	0.4872 **1**	0.2326 **2**	0.0113 **4**
1	0.1284 **3**	0.1094 **4**	0.4678 **1**	0.2813 **2**	0.0128 **4**
2	0.0740 **3**	0.0674 **4**	0.4179 **1**	0.4148 **2**	0.0257 **4**
4	0.0136 **4**	0.0154 **4**	0.1629 **3**	0.6325 **1**	0.1754 **2**
6	0.0008 **4**	0.0011 **4**	0.0176 **3**	0.2952 **2**	0.6851 **1**
$\bar{W}_{\text{SDH},1} = 1.5825$					

C.2   Frequency-domain correlation

Table C.3: $P_{Z_{n_f}|Z_{n_f-1}}$, $W_{\text{FSDH},1}$ given in bold and $\bar{W}_{\text{FSDH},1}$ in bits per symbol for $N_g = 1$ and perfect channel estimation

Z_{n_f-1} \ Z_{n_f}	0	1	2	4	6
0	0.2538 **2**	0.1641 **3**	0.5037 **1**	0.0768 **4**	0.0014 **4**
1	0.1869 **3**	0.2019 **2**	0.4871 **1**	0.1229 **4**	0.0009 **4**
2	0.0749 **3**	0.0680 **4**	0.5152 **1**	0.3370 **2**	0.0047 **4**
4	0.0040 **4**	0.0062 **4**	0.1296 **2**	0.7365 **1**	0.1235 **3**
6	6e-5 **4**	0.0001 **6**	0.0030 **3**	0.2128 **2**	0.7838 **1**
$\bar{W}_{\text{FSDH},1} = 1.4354$					

Table C.4: $P_{Z_{n_f}|Z_{n_f-1}}$, $W_{\text{FSDH},1}$ given in bold and $\bar{W}_{\text{FSDH},1}$ in bits per symbol for $N_g = 1$ and SNR = 0 dB

Z_{n_f-1} \ Z_{n_f}	0	1	2	4	6
0	0.1727 **2**	0.1459 **4**	0.5106 **1**	0.1655 **3**	0.0051 **4**
1	0.1513 **4**	0.1608 **3**	0.4622 **1**	0.2175 **2**	0.0080 **4**
2	0.0785 **3**	0.0673 **4**	0.4552 **1**	0.3864 **2**	0.0124 **4**
4	0.0092 **4**	0.0122 **4**	0.1524 **2**	0.6778 **1**	0.1481 **3**
6	0.0007 **4**	0.0008 **4**	0.0119 **3**	0.2634 **2**	0.7230 **1**
$\bar{W}_{\text{FSDH},1} = 1.5265$					

APPENDIX D

Derivation of 4-points approximation

D.1 Approximation of regular 4 QAM

The 4-point approximation for 4-QAM is derived here. Consider an arbitrary received symbol $\hat{r}_{k,n}$ (after equalization) with the Cartesian representation $(x_{k,n}, y_{k,n})$ as depicted in Figure D.1, where n denotes the subarrier index and k denotes the OFDM symbol index. The log-likelihood function is given by

$$
\begin{aligned}
\Lambda_1(r_{k,n}|b_i = 2) &= \ln\left\{\frac{1}{4}\left[\sum_{l=1}^{4}\exp(-\rho_n^2|\hat{r}_{k,n} - a_{3,l}|^2)\right]\right\} \qquad\text{(D.1)}\\
&= \ln\left\{\frac{1}{4}\left[\exp\left(-\rho_n^2\left[(x_{k,n} - x^{(3,1)})^2 + (y_{k,n} - y^{(3,1)})^2\right]\right) + ...\right.\right.\\
&\qquad\qquad \exp\left(-\rho_n^2\left[(x_{k,n} - x^{(3,2)})^2 + (y_{k,n} - y^{(3,2)})^2\right]\right) + ...\\
&\qquad\qquad \exp\left(-\rho_n^2\left[(x_{k,n} - x^{(3,3)})^2 + (y_{k,n} - y^{(3,3)})^2\right]\right) + ...\\
&\qquad\qquad \left.\left.\exp\left(-\rho_n^2\left[(x_{k,n} - x^{(3,4)})^2 + (y_{k,n} - y^{(3,4)})^2\right]\right)\right)\right\}.
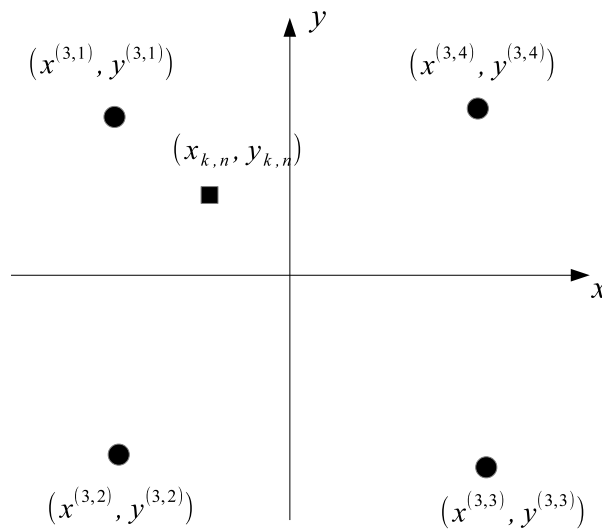\end{aligned}
$$

Figure D.1: Principle of 4-point approximation for 4-QAM

Using the relationship of coordinates shown in Figure D.1, (D.1) is equal to

$$
\Lambda_1(r_{k,n}|b_i = 2) = \ln\left\{\frac{1}{4}\left[\exp\left(-\rho_n^2\left[(x_{k,n}-x^{(3,1)})^2 + (y_{k,n}-y^{(3,1)})^2\right]\right) + ...\right.\right. \quad \text{(D.2)}
$$

$$
\exp\left(-\rho_n^2\left[(x_{k,n}-x^{(3,1)})^2 + (y_{k,n}+y^{(3,1)})^2\right]\right) + ...
$$

$$
\exp\left(-\rho_n^2\left[(x_{k,n}+x^{(3,1)})^2 + (y_{k,n}+y^{(3,1)})^2\right]\right) + ...
$$

$$
\left.\left.\exp\left(-\rho_n^2\left[(x_{k,n}+x^{(3,1)})^2 + (y_{k,n}-y^{(3,1)})^2\right]\right)\right]\right\}
$$

$$
= \ln\left\{\exp\left(-\rho_n^2\left(x_{k,n}^2 + y_{k,n}^2 + \left(x^{(3,1)}\right)^2 + \left(y^{(3,1)}\right)^2\right)\right) \times \cdots\right.
$$

$$
\left.\cosh\left(2\rho_n^2 x_{k,n}x^{(3,1)}\right)\cosh\left(2\rho_n^2 y_{k,n}y^{(3,1)}\right)\right\}
$$

$$
= -\rho_n^2\left(x_{k,n}^2 + y_{k,n}^2 + \left(x^{(3,1)}\right)^2 + \left(y^{(3,1)}\right)^2\right) + \cdots
$$

$$
\ln(\cosh\left[2\rho_n^2 x_{k,n}x^{(3,1)}\right]) + \ln(\cosh\left[2\rho_n^2 y_{k,n}y^{(3,1)}\right]).
$$

Due to $\ln(\cosh(x)) = \ln\left(\frac{e^x + e^{-x}}{2}\right) \approx |x| - \ln(2)$ for $|x| \gg 1$, (D.2) is approximated to

$$
\Lambda_1(r_{k,n}|b_i = 2) \approx -\rho_n^2(|\hat{r}_{k,n}|^2 + |a_{3,1}|^2) + 2\rho_n^2\left[|x_{k,n}x^{(3,1)}| + |y_{k,n}y^{(3,1)}|\right] - 2\ln(2). \quad \text{(D.3)}
$$

D.2   Approximation of rotated 4 QAM

In the thesis, a rotation of 16 QAM by 45° is proposed. Here, the principle of the 4-point approximation is derived for rotated 4 QAM for the purpose of simple illustration. A transfer to the rotated 16 QAM is achieved by a coordinate transformation as performed in Section 4.2.4. Consider an arbitrary received symbol $\hat{r}_{k,n}$ with the Cartesian representation $(x_{k,n}, y_{k,n})$ as depicted in Figure D.2. The log-likelihood function is given by

$$
\breve{\Lambda}_1(r_{k,n}|b_i = 2) = \ln\left\{\frac{1}{4}\left[\sum_{l=1}^{4}\exp(-\rho_n^2|\hat{r}_{k,n} - \breve{a}_{3,l}|^2)\right]\right\} \quad \text{(D.4)}
$$

$$
= \ln\left\{\frac{1}{4}\left[\exp\left(-\rho_n^2\left[(x_{k,n}-\breve{x}^{(3,1)})^2 + (y_{k,n}-\breve{y}^{(3,1)})^2\right]\right) + ...\right.\right.
$$

$$
\exp\left(-\rho_n^2\left[(x_{k,n}-\breve{x}^{(3,2)})^2 + (y_{k,n}-\breve{y}^{(3,2)})^2\right]\right) + ...
$$

$$
\exp\left(-\rho_n^2\left[(x_{k,n}-\breve{x}^{(3,3)})^2 + (y_{k,n}-\breve{y}^{(3,3)})^2\right]\right) + ...
$$

$$
\left.\left.\exp\left(-\rho_n^2\left[(x_{k,n}-\breve{x}^{(3,4)})^2 + (y_{k,n}-\breve{y}^{(3,4)})^2\right]\right)\right]\right\},
$$

where $\{\breve{a}_{3,l}\}_{l=1}^4$ and $\{(\breve{x}^{(3,1)}, \breve{y}^{(3,1)})\}_{l=1}^4$ are rotated signal points and the corresponding Cartesian coordinates, respectively. Due to the special arrangement of the rotated 4 QAM, it follows

$$\breve{x}^{(3,1)} = -\breve{y}^{(3,2)} = -\breve{x}^{(3,3)} = \breve{y}^{(3,4)} = |\breve{D}^{(3)}| \tag{D.5}$$
$$\breve{y}^{(3,1)} = -\breve{x}^{(3,2)} = -\breve{y}^{(3,3)} = \breve{x}^{(3,4)} = |\breve{D}^{(3)}|.$$
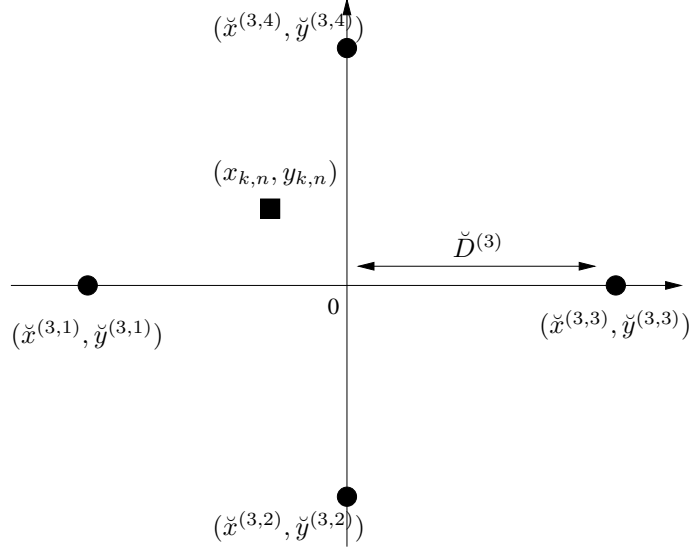

Figure D.2:   Principle of 4-points approximation for rotated 4-QAM

The log-likelihood function is now simplified to

$$
\begin{aligned}
\breve{\Lambda}_1(r_{k,n}|b_i = 2) &= \ln\left\{\frac{1}{4}\left[\sum_{l=1}^4 \exp(-\rho_n^2|\hat{r}_{k,n} - \breve{a}_{3,l}|^2)\right]\right\} \\
&= \ln\left\{\frac{1}{4}\left[\exp\left(-\rho_n^2\left[(x_{k,n} - \breve{x}^{(3,1)})^2 + y_{k,n}^2\right]\right) + ... \right.\right.\\
&\qquad\qquad \exp\left(-\rho_n^2\left[x_{k,n}^2 + (y_{k,n} + \breve{x}^{(3,1)})^2\right]\right) + ... \\
&\qquad\qquad \exp\left(-\rho_n^2\left[(x_{k,n} + \breve{x}^{(3,1)})^2 + y_{k,n}^2\right]\right) + ... \\
&\qquad\qquad \left.\left. \exp\left(-\rho_n^2\left[x_{k,n}^2 + (y_{k,n} - \breve{x}^{(3,1)})^2\right]\right)\right]\right\} \\
&= \ln\left\{\frac{1}{4}\exp\left(-\rho_n^2\left[x_{k,n}^2 + y_{k,n}^2 + \left(\breve{x}^{(3,1)}\right)^2\right]\right)\left[\exp\left(2\rho_n^2\breve{x}^{(3,1)}x_{k,n}\right) + ... \right.\right.\\
&\qquad \left.\left. \exp\left(-2\rho_n^2\breve{x}^{(3,1)}y_{k,n}\right) + \exp\left(-2\rho_n^2\breve{x}^{(3,1)}x_{k,n}\right) + \exp\left(2\rho_n^2\breve{x}^{(3,1)}y_{k,n}\right)\right]\right\},
\end{aligned} \tag{D.6}
$$

where the second factor is given by

$$\breve{factor} = \exp\left(2\rho_n^2 \breve{x}^{(3,1)} x_{k,n}\right) + \exp\left(-2\rho_n^2 \breve{x}^{(3,1)} y_{k,n}\right) + \ldots \tag{D.7}$$
$$\exp\left(-2\rho_n^2 \breve{x}^{(3,1)} x_{k,n}\right) + \exp\left(2\rho_n^2 \breve{x}^{(3,1)} y_{k,n}\right)$$
$$= \exp\left(\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} + y_{k,n} + x_{k,n} - y_{k,n}\right)\right) + \exp\left(\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} - y_{k,n} - x_{k,n} - y_{k,n}\right)\right) + \ldots$$
$$\exp\left(\rho_n^2 \breve{x}^{(3,1)} \left(-x_{k,n} + y_{k,n} - x_{k,n} - y_{k,n}\right)\right) + \exp\left(\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} + y_{k,n} - x_{k,n} + y_{k,n}\right)\right)$$
$$= \left[\exp\left(\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} + y_{k,n}\right)\right) + \exp\left(-\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} + y_{k,n}\right)\right)\right] \times$$
$$\left[\exp\left(\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} - y_{k,n}\right)\right) + \exp\left(-\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} - y_{k,n}\right)\right)\right]$$
$$= 4\cosh\left(\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} + y_{k,n}\right)\right) \cosh\left(\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} - y_{k,n}\right)\right).$$

It follows

$$\breve{\Lambda}_1(r_{k,n}|b_i = 2) = \ln\left\{\exp\left(-\rho_n^2 \left[x_{k,n}^2 + y_{k,n}^2 + \left(\breve{x}^{(3,1)}\right)^2\right]\right) \times \right. \tag{D.8}$$
$$\left. \cosh\left(\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} + y_{k,n}\right)\right) \cosh\left(\rho_n^2 \breve{x}^{(3,1)} \left(x_{k,n} - y_{k,n}\right)\right)\right\}.$$

Due to

$$\ln(\cosh(x)) = \ln\left(\frac{e^x + e^{-x}}{2}\right) \approx |x| - \ln(2) \quad \text{for } |x| \gg 1, \tag{D.9}$$

the 4-point approximation for rotated 4 QAM is expressed by

$$\breve{\Lambda}_1(r_{k,n}|b_i = 2) \approx -\rho_n^2 \left(x_{k,n}^2 + y_{k,n}^2 + \left(\breve{x}^{(3,1)}\right)^2\right) + \ldots \tag{D.10}$$
$$\rho_n^2 \breve{x}^{(3,1)} |x_{k,n} + y_{k,n}| - \ln(2) + \rho_n^2 \breve{x}^{(3,1)} |x_{k,n} - y_{k,n}| - \ln(2)$$
$$\approx -\rho_n^2 \left(x_{k,n}^2 + y_{k,n}^2 + \left(\breve{D}^{(3)}\right)^2\right) + \ldots$$
$$\rho_n^2 \breve{D}^{(3)} |x_{k,n} + y_{k,n}| - \ln(2) + \rho_n^2 \breve{D}^{(3)} |x_{k,n} - y_{k,n}| - \ln(2).$$

<div align="center">

APPENDIX E

Derivation of approximation deviations

</div>

The deviations due to the proposed approximations are provided in the following.

E.1 Deviations due to 4-P approximation

The 4-P approximation takes only these 4 signal points of a modulation level b_i into account, which belong to the subplane defined in Section 4.2.4. If $b_i \leq 2$, there exists only one subplane. All signal points belong to this plane. It follows:

$$\Delta\kappa_{1,4-\text{P}}(\mathbf{r}_n|b_i) = 0 \quad \kappa_1(\mathbf{r}_n|b_i) = \kappa_{1,4-\text{P}}(\mathbf{r}_n|b_i) \ \text{ for } \ b_i \leq 2 \text{ and } \rho_n^2 \gg 1. \tag{E.1}$$

Consequently, no bias effect occurs, since all signal points are considered. In practice, however, by using $\ln(\kappa_{1,4-\text{P}}(\mathbf{r}_n|b_i))$ further approximations are possible according to (4.36), which produce also a bias effect. But this bias effect is negligible for sufficient high system SNR.

For $b_i > 2$, considering a received symbol $r_{k,n}$ on the n-th subcarrier of the k-th OFDM symbol, the 4-P approximation causes an algorithm-specific bias effect, which is analyzed in the following. Similarly to 1-P approximation, the deviation is built

$$\Delta\kappa_{1,4-\text{P}}(\mathbf{r}_n|b_i) = \kappa_1(\mathbf{r}_n|b_i) - \kappa_{1,4-\text{P}}(\mathbf{r}_n|b_i) \tag{E.2}$$

$$= \left(\frac{1}{2^{b_i}}\right)^K \left[\prod_{k=0}^{K-1}\sum_{l=1}^{2^{b_i}} \exp\left(-\rho_n^2 D_{k,n,i,l}^2\right) - \prod_{k=0}^{K-1}\sum_{l_{k,n}\in\mathcal{L}_{k,n,i}} \exp\left(-\rho_n^2 D_{k,n,i,l_{k,n}}^2\right) \right]$$

$$= \left(\frac{1}{2^{b_i}}\right)^K \left[\underbrace{\sum_{l_0=1}^{2^{b_i}}\cdots\sum_{l_{K-1}=1}^{2^{b_i}} \exp\left(-\rho_n^2 D_{0,n,i,l_0}^2\right)\cdots\exp\left(-\rho_n^2 D_{K-1,n,i,l_{K-1}}^2\right)}_{\left(2^{b_i}\right)^K \text{addends}} - \right.$$

$$\left. \underbrace{\sum_{l_{0,n}\in\mathcal{L}_{0,n,i}}\cdots\sum_{l_{K-1,n}\in\mathcal{L}_{K-1,n,i}} \exp\left(-\rho_n^2 D_{0,n,i,l_{0,n}}^2\right)\cdots\exp\left(-\rho_n^2 D_{K-1,n,i,l_{K-1,n}}^2\right)}_{4^K \text{ maximum terms among } \left(2^{b_i}\right)^K \text{ addends}} \right].$$

According to the reasoning in Section 4.2.4, it holds

$$\Delta\kappa_{1,4-\mathrm{P}}(\mathbf{r}_n|b_i) > \Delta\kappa_{1,4-\mathrm{P}}(\mathbf{r}_n|b_j) \ \text{ for } \ b_i > b_j \tag{E.3}$$

Compared to $\Delta\kappa_{1,1-\mathrm{P}}(\mathbf{r}_n|b_i)$ given in (4.46), it follows

$$\Delta\kappa_{1,4-\mathrm{P}}(\mathbf{r}_n|b_i) < \Delta\kappa_{1,1-\mathrm{P}}(\mathbf{r}_n|b_i) \ \text{ for } \ \forall b_i \tag{E.4}$$

Consequently, the 4-P approximation produces also a bias effect towards lower b_i, but this effect is by far weaker than that of the 1-P approximation.

E.2 Deviations due to Euclidean distance approximation

Deviations due to the Euclidean distance (ED) approximation are similarly derived as for the 1-P approximation. Both algorithms retain the maximum exponential term, which dominates the remaining exponential terms. The only difference is the assignment of the probability $p(a_{i,l_{\min,k,n,i}}|r_{k,n}, b_i)$. For 1-P approximation, this value is assigned to

$$p_{1-\mathrm{P}}(a_{i,l_{\min,k,n,i}}|r_{k,n}, b_i) = \frac{1}{2^{b_i}}, \tag{E.5}$$

which is viewed as a pessimistic probability assignment, while for ED approximation it follows

$$p_{\mathrm{ED}}(a_{i,l_{\min,k,n,i}}|r_{k,n}, b_i) = 1, \tag{E.6}$$

which is an optimistic probability assignment. The ED approximation is given by

$$\kappa_{1,\mathrm{ED}}(\mathbf{r}_n|b_i) = \prod_{k=0}^{K-1} \exp\left(-\rho_n^2 D_{k,n,i,l_{\min,k,n,i}}^2\right). \tag{E.7}$$

The approximation deviation is given by

$$
\begin{aligned}
\Delta\kappa_{1,\mathrm{ED}}(\mathbf{r}_n|b_i) &= \kappa_1(\mathbf{r}_n|b_i) - \kappa_{1,\mathrm{ED}}(\mathbf{r}_n|b_i) \\
&= \left(\frac{1}{2^{b_i}}\right)^K \prod_{k=0}^{K-1}\sum_{l=1}^{2^{b_i}}\exp\left(-\rho_n^2 D_{k,n,i,l}^2\right) - \prod_{k=0}^{K-1}\exp\left(-\rho_n^2 D_{k,n,i,l_{\min,k,n,i}}^2\right) \\
&= \left(\frac{1}{2^{b_i}}\right)^K \Bigg[\underbrace{\sum_{l_0}^{2^{b_i}}\cdots\sum_{l_{K-1}}^{2^{b_i}}\exp\left(-\rho_n^2 D_{0,n,i,l_0}^2\right)\cdots\exp\left(-\rho_n^2 D_{K-1,n,i,l_{K-1}}^2\right)}_{\left(2^{b_i}\right)^K\text{addends}} - \\
&\quad \underbrace{\left(2^{b_i}\right)^K\exp\left(-\rho_n^2 D_{0,n,i,l_{\min,0,n,i}}^2\right)\cdots\exp\left(-\rho_n^2 D_{K-1,n,i,l_{\min,K-1,n,i}}^2\right)}_{\left(2^{b_i}\right)^K\times\text{maximum term}} \Bigg] \\
&= \left(\frac{1}{2^{b_i}}\right)^K \sum_{l_0}^{2^{b_i}}\cdots\sum_{l_{K-1}}^{2^{b_i}}\Bigg[\exp\left(-\rho_n^2 D_{0,n,i,l_0}^2\right)\cdots\exp\left(-\rho_n^2 D_{K-1,n,i,l_{K-1}}^2\right) - \\
&\quad \exp\left(-\rho_n^2 D_{0,n,i,l_{\min,0,n,i}}^2\right)\cdots\exp\left(-\rho_n^2 D_{K-1,n,i,l_{\min,K-1,n,i}}^2\right)\Bigg].
\end{aligned}
$$

(E.8)

Due to

$$
\exp\left(-\rho_n^2 D_{k,n,i,l_k}^2\right) < \exp\left(-\rho_n^2 D_{k,n,i,l_{\min,k,n,i}}^2\right)\quad\text{for } 0 \le k \le K-1,
$$

(E.9)

it follows

$$
\Delta\kappa_{1,\mathrm{ED}}(\mathbf{r}_n|b_i) < 0\quad\text{for }\forall b_i
$$

(E.10)

and

$$
|\Delta\kappa_{1,\mathrm{ED}}(\mathbf{r}_n|b_i)| > |\Delta\kappa_{1,\mathrm{ED}}(\mathbf{r}_n|b_i)|\quad\text{for } b_i > b_j,
$$

(E.11)

due to

$$
2^{b_i} > 2^{b_j}\quad\text{for } b_i > b_j.
$$

(E.12)

consequently, the Euclidean distance based approximation is biased towards high modulation levels in contrast to the 1-P approximation, although the principle of approximation is the same. This is caused by different assignments of the prior probability.

APPENDIX F

Higher-order moments and cumulants

F.1   Moment-based AMC

Based on (4.105), the theoretical mixed moments of order up to $n_{\mathrm{m}} = 8$ are given in Table F.1 for considered candidate set $\{A_i\}_{i=1}^{I}$. By creating Table F.1, the following facts are

Table F.1:   Theoretical values of mixed moments for the normalized constellations [8]

	NoTx ($i=1$)	BPSK ($i=2$)	4QAM ($i=3$)	16QAM ($i=4$)	64QAM ($i=5$)
$\mu_{2,0}$	0	1	0	0	0
$\mu_{2,1}$	0	1	1	1	1
$\mu_{4,0}$	0	1	1	-0.68	-0.619
$\mu_{4,1}$	0	1	0	0	0
$\mu_{4,2}$	0	1	1	1.32	1.38
$\mu_{6,0}$	0	1	0	0	0
$\mu_{6,1}$	0	1	1	-1.32	-1.298
$\mu_{6,2}$	0	1	1	0	0
$\mu_{6,3}$	0	1	1	1.96	2.22
$\mu_{8,0}$	0	1	1	2.2	1.91
$\mu_{8,1}$	0	1	0	0	0
$\mu_{8,2}$	0	1	1	-2.48	-2.75
$\mu_{8,3}$	0	1	0	0	0
$\mu_{8,4}$	0	1	1	3.12	3.96

taken into account:

1. For symmetrical constellations $\mu_{n_{\mathrm{m}},n_{\mathrm{m}}'} = 0$ if n_{m} is odd

2. $\mu_{n_{\mathrm{m}},n_{\mathrm{m}}'} = \left(\mu_{n_{\mathrm{m}},(n_{\mathrm{m}}-n_{\mathrm{m}}')}\right)^{*}$

F.2   Cumulant-based AMC

Based on the relations between cumulants and moments provided in [76], the theoretical (noise-free and infinitely long observation time) cumulants of order up to $n_{\mathrm{m}} = 8$ are tabulated in table F.2 for the considered candidate set $\{A_i\}_{i=1}^{I}$.

Table F.2:   Mixed cumulants for normalized constellations [8]

	NoTx ($i = 1$)	BPSK ($i = 2$)	4QAM ($i = 3$)	16QAM ($i = 4$)	64QAM ($i = 5$)
$\kappa_{2,0}$	0	1	0	0	0
$\kappa_{2,1}$	0	1	1	1	1
$\kappa_{4,0}$	0	-2	1	-0.68	-0.619
$\kappa_{4,1}$	0	-2	0	0	0
$\kappa_{4,2}$	0	-2	-1	-0.68	-0.619
$\kappa_{6,0}$	0	16	0	0	0
$\kappa_{6,1}$	0	16	-4	2.08	1.80
$\kappa_{6,2}$	0	16	0	0	0
$\kappa_{6,3}$	0	16	4	2.08	1.80
$\kappa_{8,0}$	0	-272	-34	-13.98	-11.50
$\kappa_{8,1}$	0	-272	0	0	0
$\kappa_{8,2}$	0	-272	34	-13.98	-11.50
$\kappa_{8,3}$	0	-272	0	0	0
$\kappa_{8,4}$	0	-272	-34	-13.98	-11.50

<div align="center">

APPENDIX G

3-D approximation

</div>

Define $\Gamma' = \Gamma\lambda_{\mathrm{margin}}(0)$ given in linear scale (not in dB) (usually $\lambda_{\mathrm{margin}}(0)$ is initialized to 1), so it holds

$$\widetilde{B}(0) = N\frac{1}{N}\sum_{n=0}^{N-1}\log_2\left(1 + \frac{e_n|H_n|^2}{N_0\Gamma'}\right). \tag{G.1}$$

Using Taylor series expansion, we obtain [99]

$$\widetilde{B}(0)=N\left\{\log_2\left(1 + \frac{e_n\mu_{\mathrm{H}}^2}{N_0\Gamma'}\right)-\log_2 e\sum_{l=1}^{N-1}\frac{(-1)^{l+1}(\frac{e_n\mu_{\mathrm{H}}^2}{N_0\Gamma'})^{l+1}}{(l+1)(1+\frac{e_n\mu_{\mathrm{H}}^2}{N_0\Gamma'})^{l+1}}\frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{|H_i|^2 - \mu_{\mathrm{H}}^2}{\mu_{\mathrm{H}}^2}\right)^l\right\} \tag{G.2}$$

with $\mu_{\mathrm{H}}^2 = \frac{1}{N}\sum_{n=0}^{N-1}|H_n|^2$. Truncating (G.2) after second order ($l \geq 2$) gives:

$$\widetilde{B}(0) \approx \widetilde{B}^{(\mathrm{2nd})}(0) = N\left\{\log_2\left(1 + \frac{e_n\mu_{\mathrm{H}}^2}{N_0\Gamma'}\right) - \log_2 e\frac{(\frac{e_n\mu_{\mathrm{H}}^2}{N_0\Gamma'})^2}{2(1+\frac{e_n\mu_{\mathrm{H}}^2}{N_0\Gamma'})^2}\frac{\sigma_{\mathrm{H}}^2}{\mu_{\mathrm{H}}^4}\right\}, \tag{G.3}$$

with e the Euler's constant and $\sigma_{\mathrm{H}}^2 = \frac{1}{N}\sum_{n=0}^{N-1}(|H_n|^2 - \mu_{\mathrm{H}}^2)^2$. So $\widetilde{B}^{(\mathrm{2nd})}(0)$ is completely characterized by the mean-squared channel gain μ_{H}^2, which gives the average power gain of the channel, and the mean-squared channel variance σ_{H}^2, which describes the level of frequency selectivity. This leads to the approximation of (4.88).

<div style="text-align: center">

APPENDIX H

Further simulation results

</div>

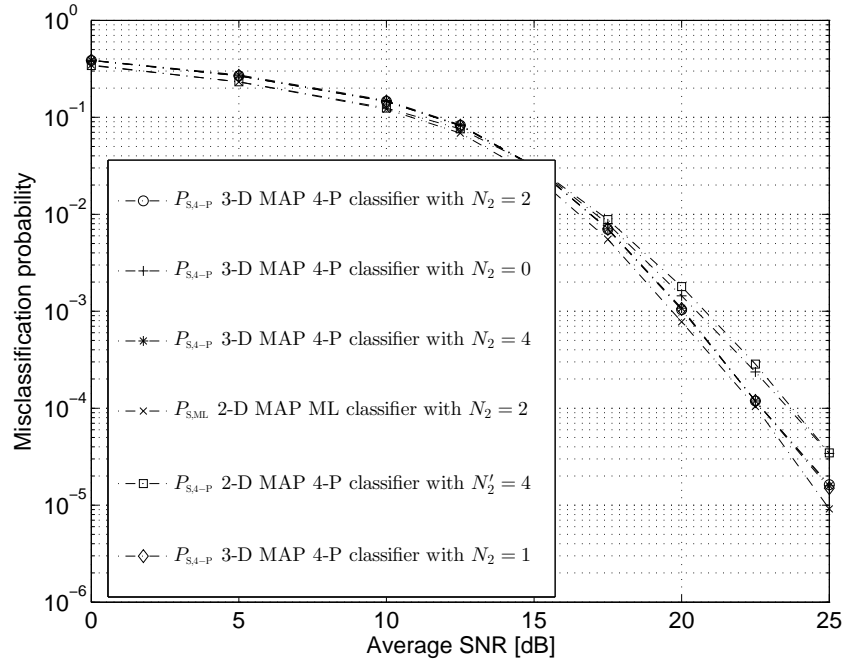In this appendix, some further simulation results are provided:

Figure H.1: The probability of misclassified subcarriers P_{S} based on the boundary condition

Figure H.1 shows that using 3D MAP 4P classifier and the criterion of $|\hat{H}_n^{\mathrm{D}}|$, P_{S} remains almost unchanged even if N_2' is set to $N_2' = 4$. On the other side, using 3D MAP 4P classifier and the criterion $\Delta\kappa_{\mathrm{DM}}(b^{(n)}|\mathbf{r}_n, Pr(|\hat{H}^{\mathrm{D}}|, B))$, certain improvements are registered in P_{S}. However, the gain here is less remarkable compared to P_{F} as the performance metric P_{S} is normalized to the total number of subcarriers N. This behaviour shows that P_{S} is less benefited of the usage of this boundary condition.
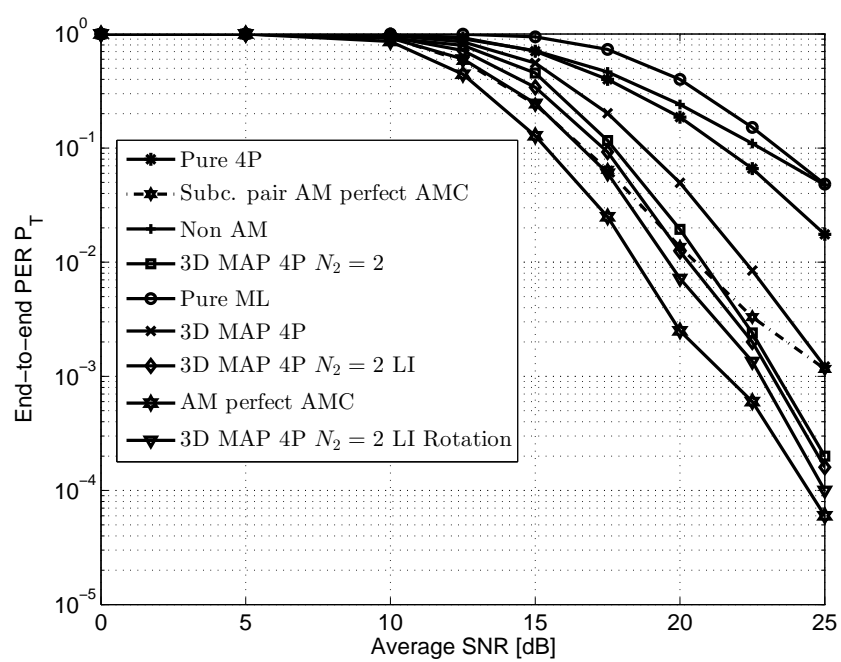
Figure H.2: P_{T} based on different classification algorithms for $N_{\mathrm{sub}} = 1$ and $N_{\mathrm{sub}} = 2$

Signalling versus automatic modulation classification

I.1 Further results for $K = 20$

Here the end-to-end PER performance P_{T} of other signalling strategies are provided. Referred to Table 5.4, for a fixed K the signalling rate $r_{\mathrm{s},i}$ determines the error probability of the signalling bits and B_{DS} determines the error probability of the data bits (for a fixed receiver setting and channel model). As B_{DS} is uniquely determined by K_{s}. In fact the parameter pair $(r_{\mathrm{S},i}, K_{\mathrm{s}})$ determines P_{T}. The 12 signalling strategies in Table 5.4 can be now reduced to the 7 strategies as shown in Figure I.1.

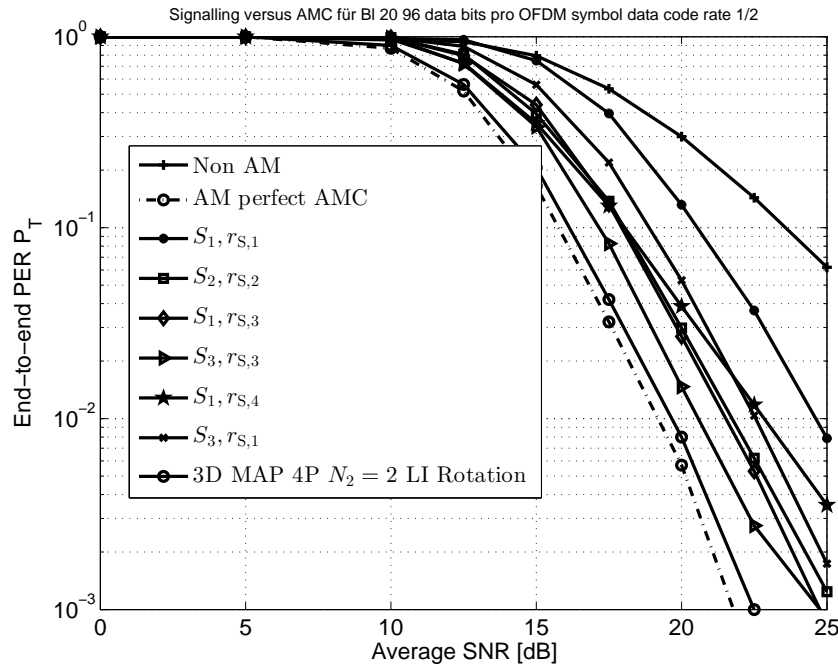**Figure I.1:** P_{T} comparison between signalling and automatic modulation classification

I.2 Simulation results for $K = 10$

The highest modulation level used in this system is $b_{i,\max} = 6$. It gives a maximum number of bits which can be transmitted in an OFDM symbol $Nb_{i,\max} = 48 * 6 = 288$

bits. Based on (5.2), for $K = 10$, K_{DS} has to fulfill

$$K_{\mathrm{DS}} = K - K_{\mathrm{S}} \geq \left\lceil \frac{B_{\mathrm{D}} K}{N b_{i,\max}} \right\rceil = \left\lceil \frac{1960}{288} \right\rceil = 7. \tag{I.1}$$

It follows

$$K_{\mathrm{S}} \leq 10 - 7 = 3. \tag{I.2}$$

Referring to the derived look-up Table 5.2, $r_{\mathrm{S},1}$ can not be used here. Only $r_{\mathrm{S},2}$, $r_{\mathrm{S},3}$ and $r_{\mathrm{S},4}$ are applicable of transmitting the signalling bits. Using these signalling rates, simulations are conducted to evaluate P_{T} as shown in Figure I.2. The best performance in terms of P_{T} is achieved by the signalling scheme S_3 signalled at the rate $r_{\mathrm{S},3}$, which shows approximately 4 dB gain over the fixed modulation scheme (Non AM) at $P_{\mathrm{T}} = 10^{-1}$. Also using S_1 signalled at $r_{\mathrm{S},4}$, a gain of around 3.5 dB was recorded. This implies the benefit of adaptive modulation combined with signalling. On the other side, the proposed classfier 3D MAP 4P $N_2 = 2$ LI Rotation outperforms all signalling schemes for at least 2 dB at $P_{\mathrm{T}} = 10^{-1}$, which confirms again the advantage of automatic modulation classification.
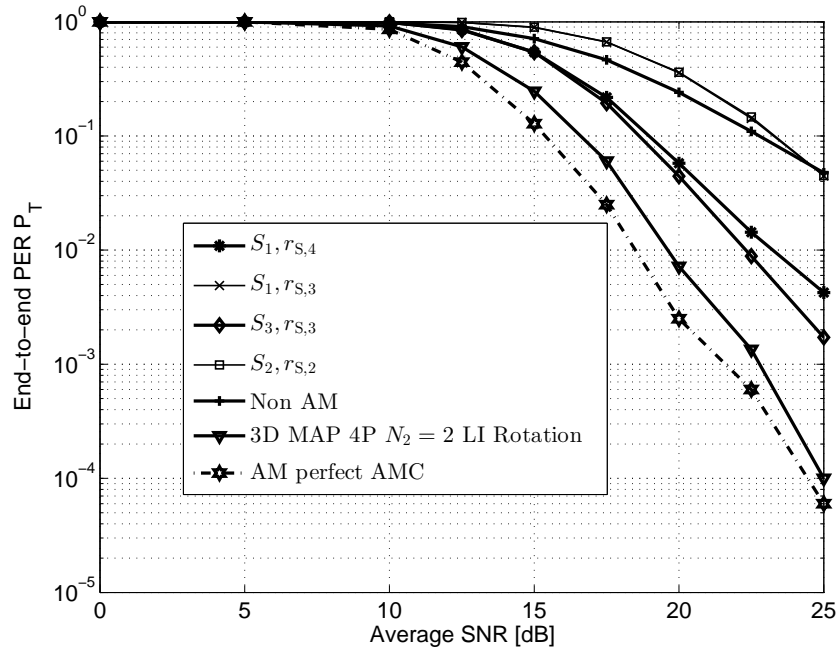

Figure I.2: P_{T} comparison between signalling and automatic modulation classification for $K = 10$

Bibliography

[1] A. Czylwik. Adaptive OFDM for wideband radio channels. In *In Proc. of the Global Telecommunications Conference*, 1996.

[2] T. Keller and L. Hanzo. Adaptive modulation techniques for duplex OFD transmission. *IEEE Trans. Veh. Technol.*, 49(5):1893–1906, 2000.

[3] R. F. H. Fischer and J. B. Huber. A new loading algorithm for discrete multitone transmission. In *In Proc. of the Global Telecommunications Conference*, 1996.

[4] D. Hughes-Hartogs. Ensemble modem structure for imperfect transmission media, US-patent, 1987.

[5] J. Campello. Optimal discrete bit loading for multicarrier modulation systems. *in Proc. GLOBECOM*, 1998.

[6] P. Chow, J. M. Cioffi, and J. A. C. Bingham. A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels. *IEEE Trans. Commun.*, 43:773–775, 1995.

[7] Y. Chen, L. Häring, and A. Czylwik. Reduction of AM-induced signaling overhead in WLAN-based OFDM systems. *in Proc. of the 14th International OFDM-Workshop (InOWo), Hamburg, Sep.*, 2009.

[8] A. Dobre, Y. Bar-Ness, W. Su, and O. A.Abdi. Survey of automatic modulation classification techniques: classical approaches and new trends. *IET Communications*, 1:137–156, 2007.

[9] L. Häring, Y. Chen, and A. Czylwik. Automatic modulation classification methods for wireless OFDM systems in TDD mode. *IEEE Trans. on Communications*, vol. 58, no. 9:pp. 2480–2485, 2010.

[10] T. Yücek and H. Arslan. A novel sub-optimum maximum likelihood modulation classification algorithm for adaptive OFDM systems. In *IEEE Wireless Commun. Networking Conf. (WCNC), vol. 2, pp. 739–744, Atlanta, GA*, 2004.

[11] Y. Chen, L. Häring, and A. Czylwik. MAP-based automatic modulation classification with reduced complexity for TDD-based adaptive OFDM systems. *in Proceedings of the 16-th International OFDM-Workshop (InOWo), Hamburg, Germany*, 2011.

[12] Digital cellular telecommunications system (phase 2+); channel coding (GSM 05.03 version 8.4.0 release 1999).

[13] 3rd generation partnership project; technical specification group radio access network; high speed downlink packet access: Physical layer aspects (release 5).

[14] J. F. Paris, M. Carmen Aguayo-Torres, and J. T. Entrambasaguas. Non-ideal adaptive modulation: bounded signaling information and imperfect adaptation. In *Globecom*, 2004.

[15] W. T. Webb and R. Steele. Variable rate QAM for mobile radio. *IEEE Trans. on Communications*, 43:2223–2230, 1995.

[16] R. Steele and W. T. Webb. Variable rate QAM for data transmission over Rayleigh fading channel. In *Wireless' 91 IEEE, Calgary, Alberta*, 1991.

[17] S. Sampei S. Otsuki, and N. Morinaga. Square QAM adaptive modulation TDMA/TDD systems using modulation level estimation with walsh function. *Electron. Lett.*, 32, (3):169–171, 1995.

[18] J. M. Torrance and L. Hanzo. Demodulation level selection in adaptive modulation. *Electronics Letters*, 32:1751–1752, 1996.

[19] J. M. Torrance and L. Hanzo. Optimization of switching levels for adaptive modulation in a slow Rayleigh fading channel.*Electron. Lett.*, 32:1167–1169, 1996.

[20] J. F. Paris, M. C. Aguayo-Torres, and J. T. Entrambasaguas. Optimum discrete-power adaptive QAM scheme for Rayleigh fading channels. *IEEE Communications Letters*, 5:281–283, 2001.

[21] J. E. Hipp. Modulation classification based on statistical moments. In *Proc. Milcom*, 1986.

[22] D. Nicholson. Issues in signal design to lower probability of classification and identification. In *Proc. Milcom*, 1987.

[23] G. Hagn, D. Jansky, and T. Dayharsh. Definition of a measurement capacity for spectrum managers. *IEEE Trans. Electromagnetic Compatibility*, 19:216–224, 1977.

[24] A. Swami, S. Barbarossa, and B. M. Sadler. Blind source separation and signal classificaiton. In *Signals, Systems and Computers, Pacific Grove, CA, USA*, 2000.

[25] K. E. Nolan, L. Doyle, D. O'Mahony, and P. Mackenzie. Signal space based adaptive modulation for software defined radio. In *IEEE WCNC, Orlando, Florida*, 2002.

[26] E. E. Azzouz and A. K. Nandi. *Automatic modulation recognition of communication signals*. Kluwer Academic, 1996.

[27] A. K. Nandi and E. E. Azzouz. Algorithms for automatic recognition of communication signals. *IEEE Trans. Commun.*, 46:431–436, 1998.

[28] H. Deng et al. Instantaneous feature based algorithm for HF digital modulation classification. *in Proc. CISS Conf.*, 2003.

[29] Y. Yang and C. H. Liu. An asymptotic optimal algorithm for modulation classificaiton. *IEEE Communications Letters*, 2:117–119, 1998.

[30] Y. Yang, C. H. Liu, and T. W. Soong. A log-likelihood function-based algorithm for QAM signal classification. *Signal Processing*, 70:61–71, 1998.

[31] C. Y. Hwang. Advanced methods for digital quadrature and offset modulation classification. In *IEEE MILCOM*, 1991.

[32] P. C. Sapiano and J. D. Martin. Maximum likelihood PSK classifier. In *IEEE MILCOM*, 1996.

[33] P. Panagiotou, A. Anastasopoulos, and A. Polydoros. Likelihood ratio tests for modulation classification. In *IEEE MILCOM*, 2000.

[34] W.Wei and J. Mendel. Maximum likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, vol. 48, no. 2, pp. 189–193,, 2000.

[35] Y. Yang, J. N. Chang, J. C. Liu, and C. H. Liu. Maximum log-likelihood function-based QAM signal classification over fading channel. *Wireless Personal Communications*, 28:77–94, 2004.

[36] A. E. El-Mahdy and N. M. Namazi. Classification of multiple m-ary frequency-shift keying signals over a Rayleigh fading channel. *IEEE Trans. Commun.*, 50:967–974, 2002.

[37] O. A. Dobre and F. Hameed. Likelihood-based algorithms for linear digital modulation classification in fading channels. *in Proc. CCECE*, 2006.

[38] S. Taira. Automatic classification of QAM signals in fading channel. In *IEEE VT, Tokyo, May*, 2000.

[39] Q. S. Huang, Q. C. Peng, and H. Z. Shao. Blind modulation classification algorithm for adaptive OFDM systems. *IEICE Trans. Commun.*, 2:296–301, 2007.

[40] S. Chaterjee and W. A. C. Fernando. Blind estimation of channel and modulation scheme in adaptive modulation schemes for OFDM-CDMA based 4G systems. *IEEE Trans. Consum. Electron.*, 50:1065–1075, 2004.

[41] S. B. Reddy, T. Yücek, and H. Arslan. An efficient blind modulation detection for adaptive OFDM systems. In *Vehicular Technology Conference 2003-Fall*, 2003.

[42] M. Gaida, S. Edinger, and N. J. Fliege. Classification of QAM signals for multicarrier systems. In *15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland*, 2007.

[43] J. Leinonen and M. Juntti. Modulation classification in adaptive OFDM systems. In *Vehicular Technology Conference 2004-Spring*, 2004.

[44] S. Lu and O.A. Dobre. Blind modulation recognition for adaptive orthogonal frequency division multiplexing systems. *in CD IEEE NECEC, St. John's, Canada.*, 2007.

[45] L. Häring, Y. Chen, and A. Czylwik. Utilizing side information in modulation classification for wireless OFDM systems with adaptive modulation. *in Proceedings of the IEEE Vehicular Technology Conference 2011 Fall, San Francisco, USA*, 2011.

[46] L. Häring, Y. Chen, and A. Czylwik. Efficient modulation classification for adaptive wireless OFDM systems in TDD mode. *in Proceedings of the Wireless Communications and Networking Conference, Sydney, Australia*, 2010.

[47] L. Häring, A. Czylwik, and Y. Chen. Automatic modulation classification in application to wireless OFDM systems with adaptive modulation in TDD mode. *in Proc. 13th International OFDM-Workshop (InOWo), Hamburg, Germany*, 2008.

[48] Y. Chen, L. Häring, and A. Czylwik. Likelihood function-based automatic modulation classification in mobile radio OFDM systems. *in Proceedings of the 15th International OFDM-Workshop (InOWo), Hamburg, Germany*, 2010.

[49] F.B. Frederiksen. An overview of OFDM and related techniques towards development of future wireless multimedia communications. In *IEEE Radio and Wireless Conference*, 2002.

[50] S. B. Weinstein. The history of orthogonal frequency division multiplexing. *IEEE Communications Magazine*, 47:26–35, 2009.

[51] R. F. H. Fischer. Communications engineering. *Lecture notes*, Ulm University, Germany, 2012. page: 499.

[52] S. B. Weinstein and P. W. Ebert. Data transmission for frequency-division multiplexing using the discrete Fourier transform. *IEEE Trans. Commun. Tech.*, 19:628–634, 1971.

[53] J. Cooley and J. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, 19:297–301, 1965.

[54] J. A. C. Bingham. Multicarrier modulation for data transmission: An idea whose time has come. *IEEE Communications Magazine*, pages 5–14, 1990.

[55] A. Peled and A. Ruiz. Frequency domain data transmission using reduced computational complexity algorithms. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP*, 1980.

[56] A. Camargo. Adaptive modulation, channel coding and MIMO schemes for practical OFDM systems. *Dissertation*, 2009.

[57] B. Hirosaki. An orthogonally multiplexed QAM system using the discrete Fourier transform. *IEEE Transactions on Communications*, 29:982–989, 1981.

[58] T. D. Chiueh and P. Y. Tsai. *OFDM Basedband Receiver Design for Wireless Communications*. John Wiley & Sons, Singapore, 2007.

[59] Richard D.J. van Nee. *OFDM for wireless multimedia communications*. Artech House Publishers, 1999.

[60] Ramjee Prasad. *OFDM for wireless communications systems*. Artech House Publishers, 2004.

[61] Gordon L. Stuber Ye Geoffrey Li. *Orthogonal frequency division multiplexing for wireless communications*. Springer, 2006.

[62] J. F. Hayes. Adaptive feedback communications. *IEEE Trans. Commun. Technol*, 16:29–34, 1968.

[63] G. J. Clowes. Variable rate data transmission for a Rayleigh fading channel. Technical report, Commun. I,ab., Defence Res. Telecommun. Establishment, Ottawa, Ont., Canada, 1969.

[64] J. K. Cavers. Variable-rate transmission for Rayleigh fading channels. *IEEE Trans. Commun.*, vol. COM-20:15–22, 1972.

[65] B. Vucetic. An adaptive coding scheme for time-varying channels. *IEEE Trans. Commun.*, 39:653–663, 1991.

[66] S. M. Alamouti and S. Kallel. Adaptive trellis-coded multiple-phase shift keying for Rayleigh fading channels. *IEEE Trans. Commun.*, 42:2305–2314, 1994.

[67] A. J. Goldsmith and S.-G. Chua. Variable-rate variable-power MQAM for fading channels. *IEEE Trans. Commun*, 45:1218–1230, 1997.

[68] N. Morinaga H. Matsuoka, S. Sampei, and Y. Kamio. Symbol rate and modulation level controlled adaptive modulation/TDMA/TDD for personal communication systems. *in Proc. IEEE VTC*, 1996.

[69] G. D. Forney Jr. and M. V. Eyuboglu. Combined equalization and coding using precoding. *IEEE Communications Magazine*, 29(12):25–34, 1991.

[70] J. H. van Lint. *Introduction to coding theory*. Springer, 1991.

[71] D. A. Huffman. A method for the construction of minimum-redundancy codes. In *Proceedings of the IRE*, 1952.

[72] Graham Wade. *Signal coding and processing*. Cambridge University Press, 1994.

[73] H. L. V. Trees. *Detection, estimation and modulation theory - Part I.* Wiley, New York, 2001.

[74] A. Puengnim, N. Thomas and J. Y. Tourneret. Digital modulation classification in flat-fading channels. In *14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy*, 2006.

[75] Chuang yu Hunag. *Advanced methods for digital modulation classification.* PhD thesis, University of Southern California, Los Angeles, 1991.

[76] D. R. Brillinger. *Time Series: Data analysis and theory.* New York: McGraw-Hill, 1981.

[77] C. L. Nikia and A. P. Petropulu *Higher-order spectra analysis: A nonlinear signal processing framework.* Prentice Hall, Inc., Englewood Fliffs, New York, 1993.

[78] A. Swami and B. M. Sadler. Hierarchical digital modulation classification using cumulants. *IEEE Transactions on Communication*, 48:416–429, 2000.

[79] M. D. Srinath and P.K. Rajasekaran. *Introduction to statistical signal processing with applications.* Prentice-Hall, NJ, 1996.

[80] S. M. Kay. *Fundamentals of statistical signal processing volume I and volume II.* Prentice Hall, 1993.

[81] Yawpo Yang and Samir S. Soliman. Statistical moments based classifier for MPSK signals. In *Globecom*, 1991.

[82] Samir S. Soliman and Shue-Zen Hsue. Signal classification using statistical moments. *IEEE Transactions on Communication*, 40. No 50:908–916, 1992.

[83] C. J. Le Martret. Modulation classification by means of different orders statistical moments. In *MILCOM*, 1997.

[84] M. B. Priestley. *Spectra analysis and time series.* Academic Press, London, UK, 1981.

[85] P. McCullagh. *Tensor methods in statistics.* Champman & Hall, 1987.

[86] Jerry M. Mendel. Tutorial on high-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications. In *Proceedings of the IEEE, Vol 79. No. 3*, volume 79, 1991.

[87] Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 5 GHz band.

[88] Lucent technology inc. Link adaptation in enhanced general packet radio service networks, 2006.

[89] de Boor. *A practical guide to splines*. Springer-Verlag, 1978.

[90] IEEE. Programs for digital signal processing. Technical report, IEEE Press, New York, 1979.

[91] Y. Liu, O. Simeone S. Wei J. Dabin, and A. M. Haimovich. Modulation classificaiton of MIMO-OFDM signals by independent component analysis and support vector machines. In *Signals, Systems and Computers (ASILOMAR)*, 2011.

[92] K. Hassan et. Blind modulation identification for MIMO systems. In *GLOBECOM*, 2010.

[93] J. D. Parsons. *The mobile radio propagation channel*. Wiley, New York, 2001.

[94] Dennis R. Morgan. Analysis and realization of an exponentially-decaying impulse response model for frequency-selective fading channels. *IEEE SIGNAL PROCESSING LETTERS*, VOL. 15:441–444, 2008.

[95] Channel models for HIPERLAN/2 in different indoor scenarios. *3ERI085B, HIPERLAN/2 ETSI/BRAN Contribution*, 30 March 1998

[96] A. L. T. Chinchilla. *Synchronization and channel estimation in OFDM: algorithms for efficient implementation of WLAN systems*. PhD thesis, University of Cottbus, Germany, 2004.

[97] J. J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Börjesson. On channel estimation in OFDM systems. *In Proceedings of Vehicular Technology Conference (VTC 95)*, 1995.

[98] M.G. Deng and X.D. Zhang. An improved blind modulation detection method for adaptive OFDM systems. In *In Proc. ICSP*, 2006.

[99] S. H. Mo Y. Y. Kang, and J. H. Cho. RMS delay spread vs. mean-square path gain for characterization of channel capacity. *IEEE Radio and Wireless Symposium, Orlando, FL, USA*, 2008.