DISSERTATION

WIRELESS OFDM SYSTEMS: CHANNEL PREDICTION AND SYSTEM CAPACITY

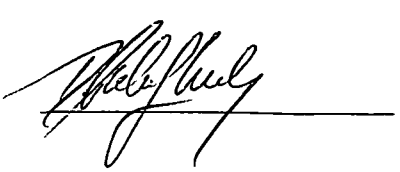
ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften

unter der Leitung von
Ao. Univ.-Prof. Dr. Franz Hlawatsch
Institut für Nachrichtentechnik und Hochfrequenztechnik (E389)

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von
Dipl.-Ing. Dieter Schafhuber
Kleistgasse 16/15
1030 Wien

Wien, im März 2004



Die Begutachtung dieser Arbeit erfolgte durch:

1. Ao. Univ.-Prof. Dr. Franz Hlawatsch

   Institut für Nachrichtentechnik und Hochfrequenztechnik
   Technische Universität Wien

2. Prof. Dr. Helmut Bölcskei

   Communication Technology Laboratory
   Department of Information Technology and Electrical Engineering
   Swiss Federal Institute of Technology (ETH) Zurich

Abstract

The general theme of this thesis is *orthogonal frequency division multiplexing* (OFDM) communications over time and frequency selective fading channels. We propose and study linear prediction techniques for acquiring *channel state information* (CSI) in OFDM receivers, and we perform an information-theoretic analysis of the performance of OFDM systems.

After a review of the generic discrete-time pulse-shaping OFDM system (which comprises conventional cyclic-prefix OFDM systems as a special case), we consider the transmission over a time and frequency selective fading channel. We arrive at an approximate multiplicative system input-output relation in which intersymbol and interchannel interference is neglected.

Based on this approximate input-output relation, we propose decision-directed channel predictors that are capable of yielding up-to-date CSI without regular transmission of pilot symbols. We derive the minimum mean-square error (MMSE) predictor and a reduced-complexity version that allows for an efficient DFT implementation. We also develop adaptive predictors that do not need statistical prior knowledge and can track nonstationary channels. Several applications of channel prediction are discussed, and the excellent performance of the proposed techniques is demonstrated by computer simulations.

The second major contribution of this thesis is an information-theoretic analysis of the performance of OFDM systems transmitting over time and frequency selective channels. We study the system capacity of wideband OFDM communications in the absence of CSI at the transmitter and the receiver. Using a codebook that is "peaky" in time and frequency, we show that OFDM can approach the infinite-bandwidth channel capacity. On the other hand, using a "nonpeaky" constant-modulus signaling scheme, we show that the information rate is reduced by a penalty term that is related to the predictability of the fading channel. We quantify the impact of the spread and shape of the scattering function on this penalty term. Finally, we formulate an upper and a lower bound on system capacity and demonstrate by simulations that both bounds are close to the AWGN channel capacity for large ranges of bandwidth and for practically relevant system parameters.

Kurzfassung

Diese Dissertation behandelt die Datenübertragung mittels orthogonaler Frequenzmultiplex-Technik (*orthogonal frequency division multiplexing*, OFDM) über zeit- und frequenzselektive Schwundkanäle. Wir entwickeln und untersuchen lineare Prädiktionsmethoden zur Erlangung von Kanalinformation im OFDM-Empfänger. Eine informationstheoretische Analyse von OFDM-Systemen liefert weiters Ergebnisse über deren Leistungsfähigkeit.

Nach der Beschreibung des OFDM-Systems mit Impulsformung (welches das OFDM-System mit zyklischem Präfix als Spezialfall enthält) behandeln wir die Übertragung über zeit- und frequenzselektive Schwundkanäle. Es ergibt sich näherungsweise eine Eingangs-Ausgangsbeziehung, die Intersymbol- und Interkanalinterferenz vernachlässigt.

Ausgehend von dieser Näherung entwickeln wir entscheidungsrückgekoppelte Kanalprädiktoren, die ohne Übertragung von Pilotsymbolen aktuelle Kanalinformation liefern können. Wir berechnen jenen Prädiktor, der den mittleren quadratischen Prädiktionsfehler minimiert und schlagen eine DFT-Implentierung geringerer Komplexität vor. Weiters entwickeln wir adaptive Prädiktoren, die kein statistisches Vorwissen benötigen und nichtstationären Kanälen folgen können. Verschiedene Anwendungen der Kanalprädiktion werden behandelt, und die ausgezeichnete Leistungsfähigkeit der vorgeschlagenen Methoden wird durch Simulationen gezeigt.

Der zweite wesentliche Beitrag dieser Dissertation ist eine informationstheoretische Analyse der OFDM-Übertragung über zeit- und frequenzselektive Schwundkanäle. Wir untersuchen die Systemkapazität ohne Kanalinformation am Sender und Empfänger. Mit Hilfe eines Codebuchs, das die Sendeleistung in Zeit und Frequenz konzentriert, zeigen wir, daß OFDM die Kanalkapazität für unendliche Bandbreite erreichen kann. Für Phasenmodulation-Codebücher zeigen wir hingegen, daß die Informationsrate durch einen Term reduziert wird, der mit der Prädizierbarkeit des Kanals zusammenhängt. Wir quantifizieren den Einfluß von Ausdehnung und Form der Streufunktion auf diesen Reduktionsterm. Abschließend formulieren wir eine obere und eine untere Schranke für die Systemkapazität und zeigen mittels Simulationen, daß beide Schranken innerhalb großer Bandbreitenbereiche und für praktisch relevante Systemparameter nahe der Kapazität des AWGN-Kanals sind.

Acknowledgment

I am indebted to my advisor Franz Hlawatsch for his support throughout the development of this thesis. His outstanding expertise and tireless advice greatly improved the technical content and the presentation of this thesis.

I am grateful to Helmut Bölcskei who invited me to visit the Communication Technology Laboratory at ETH Zürich. His encouragement and support were the basis for my information-theoretic research.

It is a pleasure for me to thank Gerald Matz who considerably contributed to this thesis by means of countless constructive advices and helpful discussions. He always found time to share his knowledge and the collaboration with him was a constant source of new ideas.

Regarding the ANTIUM project, sincere thanks for the fruitful collaboration are given to Rym Mhiri, Denis Masse and Philippe Loubaton.

I am grateful to all my colleagues at the Institute of Communications and Radio-Frequency Engineering for their support. In particular, I want to thank Harold Artés and Klaus Kopsa for many entertaining hours. Special thanks go to the people at the Communication Technology Laboratory for their warm welcome in Zürich.

Finally, I gratefully acknowledge the financial support by the Austrian Science Fund (Fonds zur Förderung der wissenschaftlichen Forschung). I also acknowledge the additional financial support by the European Commission in the course of the ANTIUM project.

Contents

1

Introduction

This thesis is concerned with *orthogonal frequency division multiplexing* (OFDM) communications over time and frequency selective Rayleigh fading channels. The investigation of OFDM systems is motivated by their increasing importance in applications. Moreover, time and frequency selective fading channels are relevant in wireless communications, which is one of the dominant applications of OFDM systems. Typically, wireless channels are small-scale fading channels and are thus inherently time and frequency selective (cf. e.g. [1–3]). An important problem in this context is the acquisition of *channel state information* (CSI) at the receiver of a wireless OFDM system. In this thesis, therefore, we propose and investigate the use of *channel prediction* for obtaining CSI. Another problem is that the information rate that can theoretically be achieved by wireless OFDM systems is unknown. We therefore present an information-theoretic analysis in which we study the system capacity and information rate of OFDM communication systems. This analysis will reveal a close relation between the achievable information rate and the predictability of the channel.

1.1 OFDM Communication Systems

OFDM is a modulation scheme that was first introduced in [4] where a general continuous-time pulse-shaping system was considered. An important development for OFDM was to recognize that a DFT can be used for modulation and demodulation [5]. However, for a long time OFDM was used only in military applications. The current success of OFDM is due to the invention of the so-called cyclic-prefix OFDM (CP-OFDM) system [6]. This system uses the DFT for modulation and demodulation, and thus CP-OFDM can be implemented with low complexity. Moreover, CP-OFDM also copes with delay-spread channels in a simple yet effective manner. Further classical work on OFDM is [7, 8], where the application of OFDM in communication systems is proposed. Since then, a continuously increasing number of publications have covered various aspects of OFDM communication systems such as synchronization, channel estimation, detection, implementation issues, etc.

Several of today's communication standards are based on OFDM. In particular, OFDM is used in commercial standards for wireless local area networks (WLAN), namely, IEEE 802.11a [9] and HIPERLAN/2 [10]; for terrestrial digital video broadcasting (DVB-T) [11]; for terrestrial digital audio broadcasting (DAB-T) [12]; and for asymmetric digital subscriber line (ADSL) systems. Furthermore, it is currently being standardized as an extension to the WLAN standard IEEE 802.11b under the name IEEE 802.11g and for the broadband wireless access system IEEE 802.15. Moreover, OFDM is a strong candidate for fourth-generation cellular communication systems, for future multi-input multi-output (MIMO) systems [13], and for ultra-wideband (UWB) systems [14].

OFDM is also known under the names *multicarrier modulation* and *discrete multitone* (DMT). Basically, these are simply different names for the same modulation scheme. However, in wired applications the designation DMT is widely accepted. A subtle difference is that for DMT real-valued transmit signals are desired and therefore only half of the available subcarriers are used for modulation; the other half is modulated by the conjugated complex symbols. In this thesis, we focus on wireless applications and therefore will use the name OFDM.

The basic idea of OFDM is to split the available transmission bandwidth into many parallel narrowband channels. In wireless systems, the channel introduces complicated impairments. It is here advantageous to deal with the low data-rate subcarriers individually. A similar approach is pursued by frequency division multiplexing. However, in OFDM the transmit/receive filters overlap in time and frequency. In this respect, OFDM is similar to code division multiplexing if the transmit/receive filters are considered as spreading sequences. However, in OFDM the transmit/receive filters have a certain time-frequency

modulation structure that aims at transmitting information at specific time-frequency locations.

There exist several extensions of OFDM systems that we briefly list but do not consider further in this thesis. In *OFDM offset quadrature amplitude modulated* (OFDM/OQAM) systems, the real part and the imaginary part of the data symbols are transmitted with a time offset of half the symbol duration [15–20]. OFDM/OQAM systems are related to Wilson bases [21, 22]. Furthermore, OFDM can be extended to systems operating with nonrectangular time-frequency lattices [23], and recently an extension of OFDM using multiple transmit/receive pulses was proposed [24, 25]. Finally, *precoded systems* [26–29] can also be regarded as an extension of OFDM.

There are certain aspects in OFDM systems that are important in practical implementations but are beyond the scope of this thesis. In particular, these topics include synchronization [27, 30–38] and the reduction of the high peak-to-average power ratio of OFDM [39–45].

1.2 Channel State Information: Relevance and Acquisition Techniques

In this section, we briefly discuss why the acquisition of *channel state information* (CSI) is important in communication systems. Furthermore, we give an overview of CSI acquisition techniques used in OFDM systems.

1.2.1 Relevance of Channel State Information

Most OFDM systems use coherent detection, which has approximately a 3 dB signal-to-noise ratio (SNR) gain over differential techniques [1] but requires CSI at the receiver. Moreover, CSI at the receiver and/or transmitter is also necessary for a number of advanced communication techniques. In particular, at the receiver CSI is required for antenna combining and space-time decoding. For example, in [46] it has been found that CSI is important to realize the full potential of MIMO communication systems. Furthermore, the transmitter needs CSI to apply link adaptation (bit and power loading), precoding, pre-equalization, and adaptive transmit antenna diversity [8, 47–50].

In wireless communication systems, it is much more difficult to obtain reliable CSI than in wired systems. This is because the estimation error for time and frequency selective channels contains, in addition to a component due to noise, a component that arises from

the time-variation of the channel. For a given (fixed) channel estimate, this additional error contribution increases gradually with time. Indeed, channel estimates are outdated after a time period equal to a fraction of the channel coherence time. Therefore, to obtain up-to-date CSI, time and frequency selective channels need to be tracked continuously.

For techniques that require CSI at the transmitter, outdated CSI is a severe problem. If CSI is obtained from the receiver via a feed-back link, a significant percentage of the data rate of the feed-back link may be required to transmit channel parameters. Here, CSI may be outdated due to transmission delays. On the other hand, in a time division duplex (TDD) communication scheme, if the channel is estimated by the transmitter while in receive mode, this CSI could be outdated as well when applied subsequently.

Depending on the application, accurate CSI is required to achieve performance gains similar to those that have been demonstrated with perfect channel knowledge. In [51], it is shown that, as a rule of thumb, CSI cannot be regarded as "perfect" if the mean square error (MSE) of channel estimation is larger than the reciprocal of the SNR. Hence, communications over time and frequency selective fading channels inherently suffers from channel uncertainty. The detrimental effects of channel uncertainty can be particularly pronounced for large bandwidths. In the wideband regime, the SNR typically is low and thus channel estimation errors tend to be large. Indeed, for spread-spectrum-like communication systems it has recently been reported that the information rate tends to zero for very large bandwidths, and the reason for this effect has been attributed to the large channel uncertainty [52–54].

1.2.2 Acquisition of Channel State Information

The approaches to channel estimation in OFDM systems can roughly be classified into four groups. These are pilot symbol assisted channel estimation, decision-directed channel estimation, blind channel estimation, and decision-directed channel prediction. Next, we briefly describe these approaches.

Pilot Symbol Assisted Channel Estimation

Channel estimation in time and frequency selective environments is usually performed in a pilot symbol assisted mode [55–72]. Here, known training symbols are regularly transmitted at certain subcarriers. For illustration, the pilot constellation in DVB-T is shown in Figure 1.1. The separation of the pilots in the time direction and in the frequency direction is four OFDM symbols and eight subcarriers, respectively; about 10 % of the transmitted symbols are pilots. In [73] it is shown that regular pilot locations are optimum. The performance

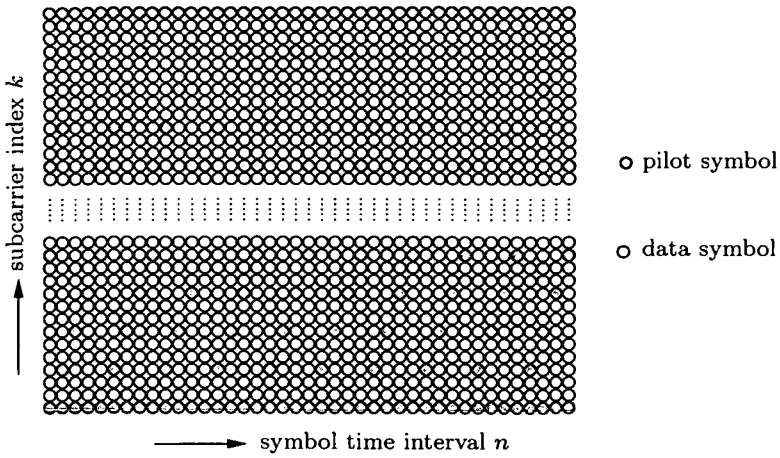


Figure 1.1: *Illustration of pilot symbol transmission in a DVB-T system.*

of the channel estimator increases with the number of pilots.

The channel coefficients at intermediate symbol time or subcarrier locations (i.e., between the pilot locations) are obtained by estimation (interpolation). A widely explored approach is linear minimum mean-square error (MMSE) channel estimation [55–57, 59, 60, 64, 65, 74], which requires (nominal or estimated) second-order channel statistics. Estimation of the channel statistics is considered in [64] and in a non-OFDM context in [75]. Explicit estimation of the channel statistics can be avoided by using adaptive channel estimators [66–70]. A drawback of pilot symbol assisted channel estimation is that it reduces the effective data rate and potentially introduces delays.

An alternative to the continuous transmission of pilots as illustrated in Figure 1.1 is to use a training data block at the beginning of each packet (this strategy is also employed in wireline communication systems). Here, the channel is estimated during the training block, possibly using MMSE channel estimation, and the resulting channel estimate is used for the duration of the respective packet. This strategy is used in IEEE 802.11a and HIPERLAN/2 where each frame starts with two known OFDM symbols (this training block is also used for synchronization). A drawback of using training data in blockwise form is that the channel cannot be tracked. Hence, IEEE 802.11a and HIPERLAN/2 cannot cope with fast time-varying channels but are intended for quasi-static transmission environments.

Decision-Directed Channel Estimation

An alternative to pilot symbol assisted channel estimation is decision-directed channel estimation as described in [76–79]. Here, previously detected symbols replace the pilot symbols for channel estimation. This has the advantage that once the symbols are detected,

all subcarriers can be used for channel estimation; this yields improved performance and avoids interpolation. On the other hand, error propagation can occur, which limits the performance of decision-directed channel estimation especially in the low SNR regime. Moreover, this technique performs poorly for fast time-varying channels because with the previously detected symbols only the past channel can be estimated. For fast time-variation of the channel, these estimates are outdated.

Blind Channel Estimation

Regular transmission of pilot symbols can be avoided by techniques for "blind" channel estimation [80] that exploit specific deterministic or stochastic structural properties of the multicarrier signal. Deterministic blind methods are attractive because they tend to feature better convergence properties and because no statistics need to be estimated at the receiver. So far, blind channel estimation has mostly been considered for time-invariant channels.

In particular, blind algorithms that are based on the cyclostationarity of the received signal have been presented in [81–84]. As observed in [83, 84], this cyclocstationarity is present even without any transmitter precoding. Alternative blind methods that are based on a deterministic signal structure have been presented in [27] (for precoded systems) and in [85] (exploiting the CP).

Decision-Directed Channel Prediction

As an alternative to conventional pilot symbol assisted or decision-directed estimation and tracking schemes, CSI can be acquired by decision-directed channel *prediction*. Here, exploiting the correlations of the fading channel, previously detected symbols are used to predict the channel into the future. While this approach is somewhat similar to decision-directed channel estimation, it is different in that is capable of yielding up-to-date CSI. Therefore, this technique has the advantage of allowing the tracking of fast time-varying channels without periodic transmission of training data. Only for initialization a short training block is required. Channel prediction also allows the application of advanced techniques such as link adaptation.

In a non-OFDM context, the prediction of fading channels and its applications were previously investigated in [50, 86–89]. In particular, the generic concept of the prediction of fading signals is described in [50], and in [89] the application of channel prediction to adaptive modulation is investigated. For OFDM systems, channel prediction and its application to equalization have been proposed in [90, 91]. However, [90] assumes that the channel can be perfectly observed (noiseless prediction), and [91] considers pilot symbol

based channel prediction.

In Chapter 3, we will consider the practically relevant case of noisy prediction, for which we will develop MMSE channel predictors as well as adaptive channel predictors [92–94]. Independently of the author's work, MMSE channel prediction in OFDM systems has recently been proposed in [65].

1.3 Information-Theoretic Aspects of Wireless Communications

Information-theoretic results known for time and frequency selective fading channels strongly depend on the CSI available at the transmitter and receiver [95]. Moreover, for the practically most important case that neither the transmitter nor the receiver has CSI, the channel capacity is known only for infinite bandwidth [96–99]; specifically, it is then identical to the capacity of an AWGN channel. This is a surprising result because it was expected that in the absence of CSI the capacity would decrease. To derive this result, frequency shift keying (FSK) signaling with an average power constraint but no peak power constraint was considered. Unfortunately, the analysis in [96–99] cannot be extended to the case of finite bandwidth, and thus it is unknown whether the capacity of time and frequency selective fading channels is identical to that of the AWGN channel for finite bandwidth.

In [100], the information rate of an "M-ary orthogonal communication system using stationary stochastic signals" and transmitting over an AWGN channel has been calculated. The system model in [100] pertains also to FSK signaling over (stationary) flat fading channels with both an average and a peak power constraint. For this system, the AWGN capacity is not achieved because the information rate is reduced by a "penalty" term due to the unknown channel.

Comparing the results of [96–99] and [100] indicates that the type of power constraint plays an important role in the analysis. In particular, it can be conjectured that "peaky" signaling schemes (i.e., no peak power constraint; sometimes the designation "flash signaling" is used for this case) are better suited for the transmission over time and frequency selective fading channels than "nonpeaky" ones.

Very recently, the interest in time and frequency selective fading channels increased again since modern wireless communication systems have to cope with mobile environments. Moreover, future communication systems will operate in the *wideband regime* with large bandwidth and low receive SNR (the SNR is low because the total transmit power is limited due to regulatory directives and/or technical reasons and the noise power is high due to the large bandwidth). It was long believed that the wideband regime is sufficiently

characterized by an asymptotic (i.e., infinite-bandwidth) analysis. However, in [101, 102] it is shown that this is not the case, and thus communication systems need to be analyzed specifically for this operation regime.

For wideband communications over time and frequency selective fading channels in the absence of CSI at the transmitter and the receiver, [52–54] report that the capacity asymptotically approaches zero if "nonpeaky" signaling schemes are used. This effect has been attributed to increasingly unreliable channel estimates when the transmit power is limited but the bandwidth is increased [52–54]. These results again indicate that one should use peaky signaling schemes for the transmission over time and frequency selective fading channels. However, it is difficult to assess whether the bounds in [52–54] have operational significance for current communication systems. This issue has recently been studied for spread-spectrum systems and pulse position modulation in [103, 104], with the conclusion that spread-spectrum-like signaling performs poorly for wideband communications. It is unclear if similar results also pertain to OFDM. For OFDM, it has been shown that the transmit and receive pulses (cf. Section 2.1) are approximate eigenfunctions of underspread time and frequency selective fading channels [105, 106]. Therefore, OFDM systems could be better suited for time and frequency selective environments than spread-spectrum systems. However, it can be conjectured that also for OFDM the type of power constraint (i.e., peaky or nonpeaky signaling) will be important for the analysis.

In Chapter 4, we will carry out an information-theoretic analysis of wireless OFDM communication systems. In particular, we will study the system capacity of wideband OFDM communications over time and frequency selective fading channels in the absence of CSI at the transmitter or receiver. To the author's knowledge, this information-theoretic analysis of OFDM systems is completely novel.

We will consider two specific signaling schemes that are peaky and nonpeaky, respectively. An advantage of the OFDM modulation format is that it allows to analyze both cases within the same framework. For infinite bandwidth, using peaky signaling, we obtain that OFDM can achieve the infinite-bandwidth capacity of time and frequency selective fading channels. With the nonpeaky signaling scheme, we demonstrate that, similar to the result in [100], the information rate decreases by a "penalty" term due to the unknown channel. In particular, also for OFDM, spreading the transmitted information over very large bandwidths results in vanishing information rate. However, in contrast to [52–54] (where bounds are derived that are only asymptotically tight), we can study the exact behavior of the information rate in the wideband regime. In this analysis, we will show that the information rate reduction is related to the predictability of the channel, which conforms to the results in [107]. Moreover, since in OFDM we know the exact behavior

of the information rate, the operating regime with vanishing information rate can easily be avoided by appropriately limiting the bandwidth actually used to transmit information. With this strategy, we obtain information rates for OFDM that do not approach zero but remain constant for very large bandwidths. For spread-spectrum systems, a similar strategy could be applied if the dependence of the information rate on bandwidth was known.

To assess the system capacity of OFDM in the finite-bandwidth case, we will derive upper and lower bounds since we cannot provide exact results. We will demonstrate that these bounds practically coincide for typical system and channel parameters over large and practically relevant ranges of bandwidth. Moreover, the upper bound is close to the capacity of the AWGN channel with the same transmit power and bandwidth.

From an information-theoretic point of view, our results indicate that OFDM is indeed well suited for wideband communications over time and frequency selective fading channels.

1.4 Overview of Contributions

We conclude this introductory chapter with an overview of the main contributions of this thesis.

Pulse-shaping OFDM system model (Chapter 2): Our analysis is based on a "pulse-shaping" discrete-time OFDM system that is closely related to the original OFDM system proposed in [4, 5]. We present an efficient digital implementation of pulse-shaping OFDM systems that is based on the discrete Fourier transform (DFT) and whose computational complexity is only slightly higher than that of the cyclic-prefix (CP) implementation of OFDM [6]. Furthermore, we introduce the concept of an *equivalent channel* that jointly characterizes the time and frequency selective fading channel and the pulse shaping at transmitter and receiver. Parts of this work have been previously reported in [108].

MMSE channel predictor for OFDM (Subsection 3.2.1): We derive a novel decision-directed MMSE channel predictor for OFDM that exploits the correlations of the time and frequency selective channel. We show that the MMSE predictor consists of two stages: a division by the data symbols and a time-varying MIMO filter. This work has been published in [69, 94].

Efficient implementation of channel predictor (Subsections 3.2.2 and 3.2.3): We show that the complexity of the full-blown MMSE channel predictor can be significantly reduced by replacing the second stage with a time-invariant MIMO filter,

and finding an efficient DFT implementation for the resulting reduced-complexity predictor. This work has been published in [69, 94].

Performance of infinite-length predictor (Subsections 3.2.4 and 3.2.5): We consider the MMSE channel predictor with infinite predictor memory. The performance (MMSE) of this predictor can be calculated in terms of the channel's scattering function and the noise variance. Moreover, for the special case of specular scattering for which the scattering function consists of discrete (Dirac impulse) components, we show that the channel can be predicted perfectly, i.e., the channel prediction MMSE is equal to zero. This result, which extends a similar result for noiseless prediction of nonregular processes [109], has been submitted for publication [110, 111].

Adaptive channel predictors (Section 3.3): Calculation of the MMSE channel predictor requires knowledge of the channel statistics. To avoid estimation of the channel statistics, we propose *adaptive* channel predictors that perform a continual update of the predictor coefficients using the normalized least-mean-square (NLMS) or recursive least-squares (RLS) algorithm. The adaptive channel predictors are capable of tracking nonstationary channels. This work has been published in [69, 93].

Application examples of channel prediction (Section 3.4): We consider the typical application of channel prediction, namely CSI acquisition for channel equalization at the receiver. We also propose an adaptive modulation strategy for wireless OFDM systems that is based on channel prediction and employs a novel protocol for feeding CSI back to the transmitter. Finally, we briefly consider channel prediction in OFDM systems that continuously transmit pilot symbols. Parts of this work have been published in [69, 93, 94].

System capacity of OFDM for infinite bandwidth (Section 4.3): We consider the system capacity of an OFDM system transmitting over a time and frequency selective Rayleigh fading channel and using an orthogonal codebook similar to [99, Section 8.6]. Analyzing the bit error probability of the maximum-likelihood detector in the asymptotic (i.e., infinite-bandwidth) regime, we show that error-free transmission is possible if the information rate is below a certain nonzero rate. This rate is equal to the capacity of time and frequency selective Rayleigh fading channels. This work has been submitted for publication [110, 111].

Information rate for constant-modulus signaling (Subsections 4.4.1–4.4.3): We study the practically important case of an OFDM system using constant-modulus signaling and derive its information rate. We show how the information rate is related

to the predictability of the channel (i.e., the prediction MMSE of the infinite-length MMSE channel predictor). In particular, we show that the information rate is close to the AWGN channel capacity if the channel prediction MMSE is close to zero. On the other hand, spreading the transmitted information over extremely large bandwidths results in zero information rate due to the large channel prediction MMSE. This work has been submitted for publication [110, 111].

Dependence of information rate on scattering function (Subsection 4.4.4): For finite bandwidth, we quantify how the scattering function influences the information rate of OFDM systems using constant-modulus signaling. In particular, we show that the information rate is reduced for a scattering function that is widely spread, and minimized by a flat scattering function. This work has been submitted for publication [110, 111].

Upper and lower bounds on system capacity (Section 4.5): To assess the OFDM system capacity for finite bandwidth, we develop an upper bound and we derive a lower bound from the information rate of OFDM using constant-modulus signaling (cf. Subsection 4.4.1). We demonstrate that for typical system and channel parameters, both bounds are close to the AWGN channel capacity. Furthermore, we show that the OFDM system capacity does not vanish for very large bandwidths. Parts of this work have been submitted for publication [110, 111].

Bounds on information rate for Gaussian signaling (Section 4.6): Finally, we present two lower bounds on the information rate for Gaussian signaling. One of these bounds is most useful in the narrowband regime, while the other is most useful in the wideband regime.

2

System Model

Orthogonal frequency division multiplexing (OFDM) communication systems were first introduced in the late '60s. In those days OFDM was hardly used and for a long time only little research was performed on this specific modulation format. It required the increased capabilities of digital signal processors and the desire for spectrally efficient wideband communications to again draw the interest to OFDM. Since the early '90s there has been a steadily increasing interest in OFDM and nowadays OFDM has been standardized for several communication services.

In this chapter, we describe the general system model for *pulse-shaping* OFDM communications over wireless channels. Pulse-shaping OFDM systems were initially introduced by [4, 5] and further investigated in, e.g., [106, 108, 112–116]. Based on this model, in subsequent chapters we will develop novel predictors for OFDM systems and carry out an information-theoretic analysis of wireless OFDM systems transmitting over time and frequency selective channels.

The outline of this chapter is as follows. In Section 2.1, we review the pulse-shaping OFDM modulator and demodulator and show how a pulse-shaping OFDM system can be implemented efficiently. As a special case, we explain the CP-OFDM system since it is the OFDM implementation used in practical applications. In Section 2.2, we then introduce the random time-varying channel model. We consider a time-varying channel since our interest is in wireless communications. In Section 2.3, the system input-output relation for OFDM communications over a time-varying channel is derived. Finally, we present a widely used approximation that neglects intersymbol and intercarrier interference and results in a very simple multiplicative input-output relation.

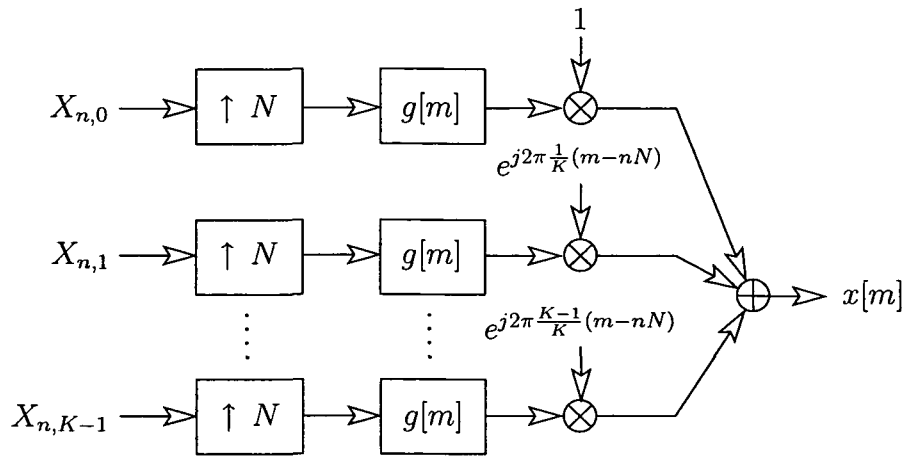


Figure 2.1: *Modulator of an OFDM system with K subcarriers and a symbol duration of N samples using pulse-shaping filters $g[m]$.*

2.1 OFDM Modulator and Demodulator

We next introduce the pulse-shaping OFDM modulator and demodulator [106, 108, 114, 115]. We discuss a low-complexity implementation and specialize for the case of CP-OFDM.

OFDM Modulator

Figure 2.1 shows the modulator for a pulse-shaping OFDM system with K subcarriers. The data symbols $X_{n,k}$ may belong to a single high-data-rate source (related by serial-to-parallel conversion) or to multiple sources/users. The subscripts n and k in $X_{n,k}$ denote the OFDM symbol and the OFDM subcarrier, respectively. The OFDM symbol duration is N signal samples. After upsampling by a factor of N, the transmit data is passed through the transmit filter using the transmit pulse $g[m]$ and modulated with the respective subcarrier center frequencies. The modulated discrete-time baseband transmit signal is given by

$$x[m] = \sum_{n=-\infty}^{\infty} \sum_{k=0}^{K-1} X_{n,k}\, g_{n,k}[m]\,, \tag{2.1}$$

with the time-frequency shifted transmit pulses

$$g_{n,k}[m] \triangleq g[m-nN]\, e^{j2\pi\frac{k}{K}(m-nN)}\,. \tag{2.2}$$

Note that $g[m]$ can in general have arbitrary shape and length. The number of subcarriers may vary in a wide range. For example, in IEEE 802.11a and in HIPERLAN/2 only $K = 64$

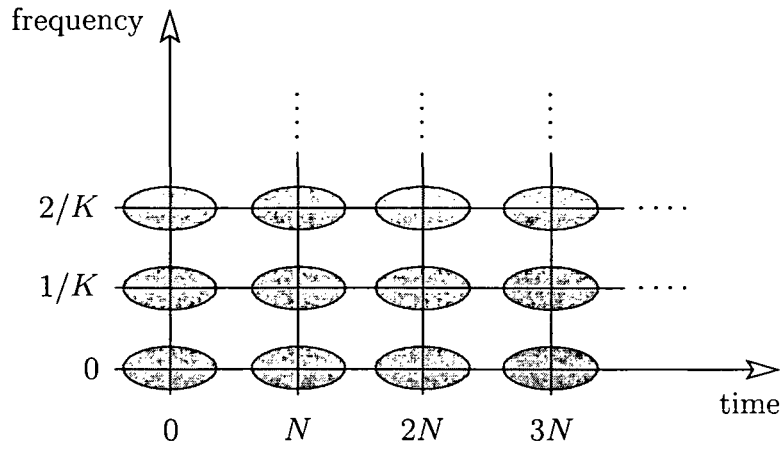


Figure 2.2: *OFDM modulation can intuitively be interpreted as sending the data symbols on top of time-frequency shifted pulses. These pulses are located on a rectangular time-frequency lattice with cell size $N \times 1/K$. Note that in contrast to this schematic picture, the pulses overlap with their neighbors.*

subcarriers are used [9, 10] but in DVB-T we have as many as $K = 8192$ subcarriers [11]. Moreover, the OFDM symbol duration is typically less than 25 % larger than K so that $1 < N/K \leq 1.25$. As will be discussed presently, the excess value of the ratio N/K above one corresponds to the amount of redundancy in the OFDM system, and thus the spectral efficiency is inversely proportional to N/K. Hence, a low value for N/K is desirable in practical OFDM systems.

It is intuitive to think of (2.1) as transmitting the data symbols $X_{n,k}$ on a rectangular time-frequency lattice of cell size $N \times 1/K$. This concept is illustrated in Figure 2.2. Around the grid points of the rectangular lattice, the time-frequency shifted pulses $g_{n,k}[m]$ are located. Since the $g_{n,k}[m]$ are generated from the single pulse $g[m]$ via time and frequency shifts, they all have the same shape. On top of these pulses, the data symbols are sent. However, (2.1) is more complicated than sketched in Figure 2.2 since a nonnegligible overlap between neighboring pulses does exist. Note that other lattice geometries than the rectangular are also possible. In particular, [23] proposes a hexagonal grid, which can have better interference robustness but is more difficult to implement and to analyze.

OFDM Demodulator

The demodulator of a pulse-shaping OFDM system is shown in Figure 2.3. During each OFDM symbol interval, the demodulator derives the K sequences $Y_{n,k}$ with $k = 0, \ldots, K-1$

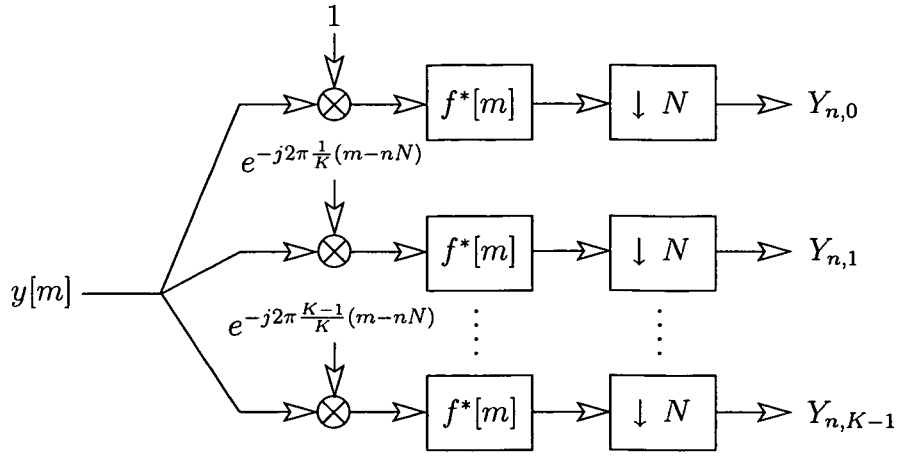


Figure 2.3: *Demodulator of a pulse-shaping OFDM system.*

by calculating the inner products

$$Y_{n,k} = \langle y, f_{n,k} \rangle = \sum_{m=-\infty}^{\infty} y[m]\, f_{n,k}^*[m]\,. \tag{2.3}$$

Here, the received signal is denoted by $y[m]$ and the time-frequency shifted receive pulses are defined as

$$f_{n,k}[m] \triangleq f[m - nN]\, e^{j2\pi\frac{k}{K}(m-nN)}\,, \tag{2.4}$$

with the receive pulse $f[m]$. Note that (2.2) and (2.4) use the same time-shift and frequency-modulation structure.

Biorthogonality Condition

In the absence of distortions and noise, the received signal is given by $y[m] = x[m]$. Inserting (2.1) into (2.3) then yields for the demodulated sequences

$$Y_{n,k} = \sum_{n'=-\infty}^{\infty} \sum_{k'=0}^{K-1} X_{n',k'} \langle g_{n',k'}, f_{n,k} \rangle\,.$$

Hence, perfect demodulation (i.e., $Y_{n,k} = X_{n,k}$) is obtained if and only if the transmit pulse $g[m]$ and receive pulse $f[m]$ satisfy the *biorthogonality condition*

$$\langle g_{n',k'}, f_{n,k} \rangle = \delta[n-n']\,\delta[k-k']\,. \tag{2.5}$$

In the special case where the transmit and receive pulses are identical, i.e., $g[m] = f[m]$, (2.5) is an orthogonality condition; this explains the name OFDM. However, throughout

this thesis we will use the term OFDM also for the case $g[m] \neq f[m]$. We note that the latter case is sometimes referred to as *biorthogonal* frequency division multiplexing.

With the transmit and receive pulses defined in (2.2) and (2.4), respectively, the left-hand side of (2.5) can be reformulated as

$$\langle g_{n',k'}, f_{n,k} \rangle = e^{j2\pi \frac{k'-k}{K} N} A^*_{f,g}\left((n'-n)N, \frac{k'-k}{K}\right),$$

with the *cross-ambiguity function* [21, 117]

$$A_{f,g}(l,\varphi) \triangleq \sum_{m=-\infty}^{\infty} f[m]\, g^*[m-l]\, e^{-j2\pi\varphi m}. \tag{2.6}$$

Hence, the biorthogonality condition (2.5) is equivalently given by

$$A_{f,g}\left(nN, \frac{k}{K}\right) = \delta[n]\,\delta[k]. \tag{2.7}$$

The biorthogonality condition (2.7) imposes a certain structure on the shape of the cross-ambiguity function: it requires that the cross-ambiguity function is equal to one at the origin and zero at all other grid points of the OFDM time-frequency lattice.

It can be shown that pulses that fulfill (2.5) or equivalently (2.7) exist if and only if $N/K \geq 1$, and furthermore the pulse sets $\{g_{n,k}\}$ and $\{f_{n,k}\}$ have to constitute a frame [21, 106, 115, 117]. For $N/K = 1$, a unique solution is obtained, i.e, there exists only one biorthogonal $f[m]$ to a prescribed $g[m]$ and vice versa. However, for $N/K > 1$ there exist several solution since (2.5) and (2.7) are underdetermined. This is a degree of freedom that can be exploited by the system designer to, e.g., minimize the interference when transmitting over a doubly dispersive channel [108]. If N/K is increased, the symbol rate of the OFDM system is decreased since less symbols can be transmitted within a certain time-frequency area. This reduces the spectral efficiency of the system. Therefore, the choice of N/K corresponds to a tradeoff between system design freedom and symbol rate (spectral efficiency). In practical systems, N/K is typically between 1.03 and 1.25.

2.1.1 Efficient Implementation

We next propose an efficient implementation of the modulator and demodulator of pulse-shaping OFDM systems [108]. The transmit and receive pulses are assumed to have finite length which will be denoted by L_g and L_f, respectively. The proposed implementation is shown for the modulator in Figure 2.4 and for the demodulator in Figure 2.5. It essentially consists of the usual length-K IDFT or DFT that is also a part of CP-OFDM systems

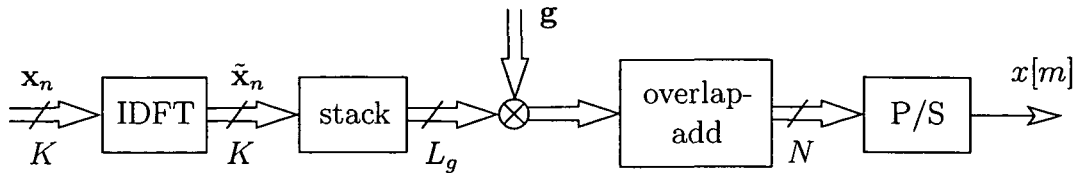


Figure 2.4: *Efficient implementation of a pulse-shaping OFDM modulator.*

(cf. Section 2.1.2), a pulse-shaping operation (elementwise multiplication by the vector $\mathbf{g} = (g[0]\, g[1] \cdots g[L_g-1])^T$ or $\mathbf{f} = (f[0]\, f[1] \cdots f[L_f-1])^T$), and an overlap-add or pre-aliasing operation. We note that in practically relevant scenarios where N/K is only slightly larger than one, polyphase implementations [118] are not possible since they would require N/K to be an integer.

Implementation of the Modulator

Within the nth OFDM symbol period, $x[m]$ in (2.1) can be written as

$$x[m] = \sum_{i=n-Q_g}^{n+Q_g} x_i^{(g)}[m-iN], \qquad m = nN, \ldots (n+1)N-1, \tag{2.8}$$

where $Q_g \triangleq \lceil L_g/(2N) \rceil$ and

$$x_n^{(g)}[m] \triangleq \tilde{x}_n[m]\, g[n], \qquad \text{with} \quad \tilde{x}_n[m] \triangleq \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} X_{n,k}\, e^{j2\pi \frac{mk}{K}}. \tag{2.9}$$

Equation (2.8) describes an overlap-add operation that involves $2Q_g + 1$ windowed IDFT signals $x_n^{(g)}[m]$ (see (2.9)). These can be computed as follows (see Figure 2.4). First, the vector $\tilde{\mathbf{x}}_n = (\tilde{x}_n[0]\ \tilde{x}_n[1] \cdots \tilde{x}_n[K-1])^T$ is periodically repeated (stacked) to form a length-L_g vector. Note that $\tilde{\mathbf{x}}_n$ contains the length-K normalized IDFT of $X_{n,k}$. Subsequently, this vector is multiplied elementwise by the transmit pulse vector $\mathbf{g}$.

Implementation of the Demodulator

Demodulation of the received signal $y[m]$ according to (2.3) can be efficiently implemented by means of the length-K normalized DFT

$$Y_{n,k} = \frac{1}{\sqrt{K}} \sum_{m=0}^{K-1} \tilde{y}_n^{(f)}[m]\, e^{-j2\pi \frac{km}{K}}.$$

Here, the length-K sequence $\tilde{y}_n^{(f)}[m]$ is obtained from the windowed received signal

$$y_n^{(f)}[m] \triangleq y[m + nN]\, f^*[m]$$

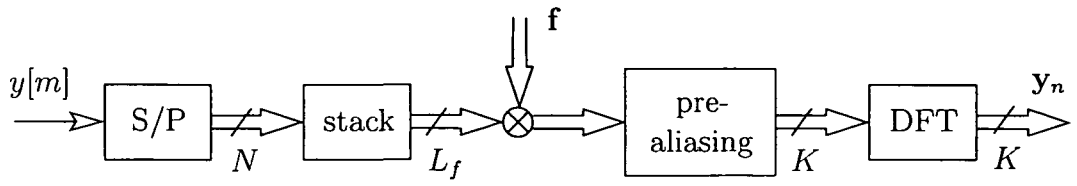


Figure 2.5: *Efficient implementation of a pulse-shaping OFDM demodulator.*

via the "pre-aliasing" operation

$$\tilde{y}_n^{(f)}[m] = \sum_{i=-Q_f}^{Q_f} y_n^{(f)}[m+iK],$$

with $Q_f \triangleq \lceil L_f/(2K) \rceil$.

Computational Complexity

The complexity of the modulator is determined by the IDFT and the pulse shaping, requiring a total of $\mathcal{O}(K \log_2 K + L_g)$ operations per symbol period. Similarly, the DFT and windowing at the receiver amount to $\mathcal{O}(K \log_2 K + L_f)$ operations per symbol period. Compared to a CP-OFDM system that requires only the IDFT/DFT (i.e., no pulse-shaping), this is an increase by $L_g + L_f$ operations per symbol period. As an example, for $K = 1024$ subcarriers, symbol length $N = 1280$, and pulse length $L_g = L_f = 2N$, the increase in computational complexity with respect to a CP-OFDM system is only 25%. We note that due to the overlap-add and pre-aliasing operations, pulse-shaping OFDM systems require additional memory and introduce a latency of several symbol periods.

2.1.2 Cyclic-Prefix OFDM System

The simplest and most widely used OFDM variant is the cyclic-prefix OFDM (CP-OFDM) system [5–7, 119]. It is used by all commercial communication systems based on OFDM.

CP-OFDM Pulses

In a CP-OFDM system, the transmit and receive pulse are both rectangular. They are given by

$$g[m] = \begin{cases} \frac{1}{\sqrt{K}}, & m = -L_{\text{cp}}, \ldots, K-1, \\ 0, & \text{else}, \end{cases} \quad \text{and} \quad f[m] = \begin{cases} \frac{1}{\sqrt{K}}, & m = 0, \ldots, K-1, \\ 0, & \text{else}. \end{cases}$$

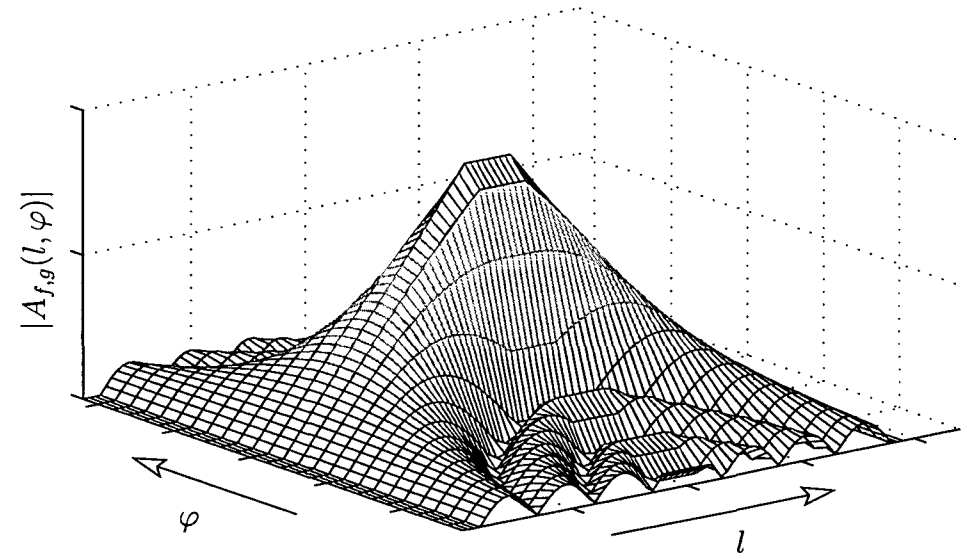


Figure 2.6: *Magnitude plot of the ambiguity function for a CP-OFDM system with* $N = 80$ *and* $K = 64$.

Note that the transmit pulse exceeds the receive pulse by the CP length L_{cp}. The symbol duration is $N = K + L_{\text{cp}}$. These pulses fulfill the biorthogonality condition (2.5) or (2.7). Moreover, the receive pulse set $\{f_{n,k}\}$ constitutes an orthonormal set by itself, i.e., $\langle f_{n,k}, f_{n',k'} \rangle = \delta[n-n']\delta[k-k']$. The magnitude of the cross-ambiguity function (2.6) for a CP-OFDM system with $N = 80$ and $K = 64$ is shown in Figure 2.6. In the delay direction at Doppler frequency $\varphi = 0$, this cross-ambiguity function is shaped like a trapezoid and in the Doppler frequency direction it is shaped like $\sin(x)/x$. Due to the CP, $A_{f,g}(l,0) = 1$ for $l = 0, 1, \ldots, L_{\text{cp}}$.

Modulator and Demodulator

For the rectangular transmit pulse, the transmit signal is given by

$$x[m] = \sum_{n=-\infty}^{\infty} x_n[m - nN],$$

where the nth CP-OFDM symbol is obtained as

$$x_n[m] = \begin{cases} \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} X_{n,k}\, e^{j2\pi mk/K}, & m = 0, \ldots, K-1, \\ x_n[m+K], & m = -L_{\text{cp}}, \ldots, -1, \\ 0, & \text{else}. \end{cases}$$

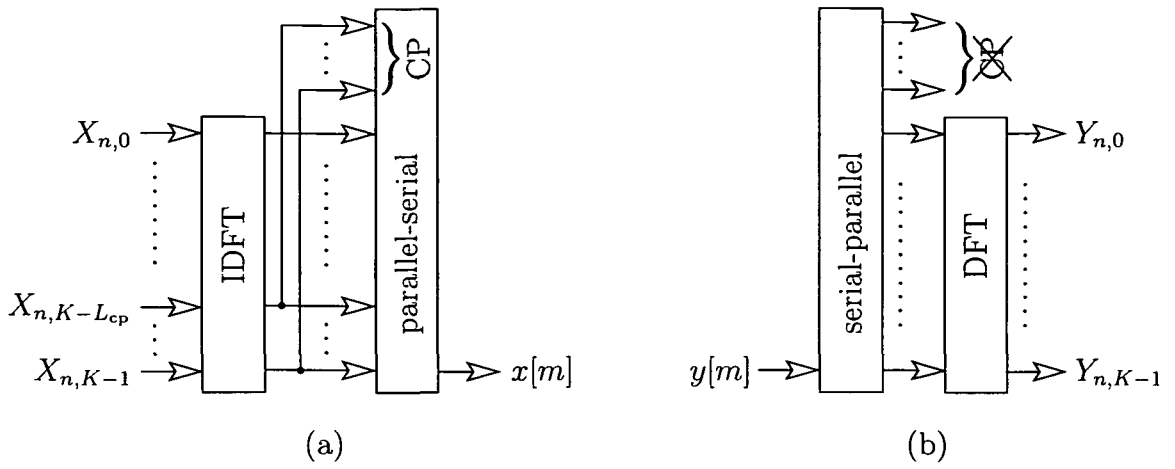


Figure 2.7: *Efficient DFT implementation of (a) the modulator and (b) the demodulator for a CP-OFDM system.*

Hence, the modulation merely involves a scaled K-point IDFT. The part of $x_n[m]$ for $m = -L_{\text{cp}}, \ldots, -1$ is the CP which is a copy of the last L_{cp} nonzero samples of $x_n[m]$. This modulator is shown in Figure 2.7(a).

Furthermore, the demodulation in (2.3) simplifies to a scaled K-point DFT,

$$Y_{n,k} = \frac{1}{\sqrt{K}} \sum_{m=0}^{K-1} y[nN + m]\, e^{-j2\pi km/K},$$

which is shown in Figure 2.7(b). Note that the CP is not used for the purpose of demodulation; it is discarded before applying the DFT.

A comparison of Figure 2.7 with the implementation of a pulse-shaping OFDM system shown in Figure 2.4 and Figure 2.5 reveals that we obtain some simplifications in the CP-OFDM case. In the modulator, only the IDFT, the stacking operation (prepending the CP), and the parallel-to-serial conversion are retained. The demodulator simplifies to a serial-to-parallel conversion and a DFT. No pulse-shaping is necessary for CP-OFDM because the transmit and receive pulses are rectangular. In practical systems, the (I)DFT is implemented by an (I)FFT and therefore the number of subcarriers is chosen as a power of two.

Purpose of the Cyclic Prefix

The CP can be regarded as a guard interval protecting consecutive OFDM symbols from intersymbol interference in the case of transmission over a multipath channel. However, a more detailed analysis shows that the CP may also prevent interference between the

subcarriers. In particular, for a time-invariant channel with impulse response shorter than $L_{\text{cp}} + 1$ samples and after discarding the CP at the receiver, the channel input-output relation is equivalent to a cyclic convolution. Therefore, after demodulation by means of a DFT, the system input-output relation is multiplicative, i.e., each subcarrier is affected by flat fading only. We will show in Section 2.3 that in practice such a multiplicative input-output relation approximately holds also for time-varying channels.

A further important advantage of using a CP is that the signal structure imposed by the CP can also be exploited for time synchronization [30, 34, 35, 120, 121].

There exist communication systems that are closely related to CP-OFDM. The classical one is single-carrier communications with frequency-domain equalization [119, 122, 123]. More recently, it has been suggested to use a CP also for this system [124–127]. The resulting system can easily be derived from Figure 2.7 by shifting the IDFT operation from the transmitter to the receiver. Another communication system that is related to CP-OFDM (indeed, it can be regarded as being dual to CP-OFDM) is known as *trailing-zero* OFDM [126, 128–131]. Instead of prepending a CP, it adds trailing zeros to the OFDM symbol at the transmitter. The receiver uses an overlap-add operation before DFT demodulation.

2.2 Wireless Fading Channels

In this thesis, we are interested in OFDM communications over wireless channels. The fundamental characteristic of a wireless channel is its time variation, which results from movements of the transmitter, the receiver, and/or the scatterers. Therefore, the rate of time variation strongly depends on the propagation scenario; e.g., in indoor scenarios the channel changes more slowly than in outdoor scenarios. The basic wave propagation mechanisms in wireless communications are free-space propagation, scattering, reflection, and diffraction [1, 3, 132]. We will not consider these physical propagation mechanisms in detail but use a statistical channel characterization known as the wide-sense stationary uncorrelated scattering (WSSUS) model [1, 3, 132, 133]. Moreover, we concentrate on the small-scale fading effects of the wireless channel. Large-scale fading effects like time variation of the path loss, the channel statistics, or the number of delay taps will not be modeled for reasons of analytic tractability. (We note, however, that a generalized channel model incorporating small-scale and large-scale fading effects has recently been introduced [134–136].)

This section is organized as follows. We first introduce the continuous-time channel

model in Subsection 2.2.1. We define several equivalent channel descriptions and explain the WSSUS model. Furthermore, some global channel parameters are defined. In Subsection 2.2.2, we consider the discrete-time channel that arises from the combination of analog-to-digital (A/D) conversion at the transmitter, transmission over the continuous-time channel, and digital-to-analog (D/A) conversion at the receiver. We again consider different channel descriptions and relate them with their continuous-time counterparts. Furthermore, we discuss a discrete WSSUS assumption and define some global channel parameters.

2.2.1 Continuous-Time Channel Model

Channel Input-Output Relation

We consider a wireless channel denoted by the operator symbol $\mathbb{H}_c$. Due to multipath propagation and Doppler shifts, $\mathbb{H}_c$ is time-varying. Its input-output relation is given by[1] [1, 3, 132, 133]

$$y(t) = (\mathbb{H}_c x)(t) + u(t) = \int_\tau h(t,\tau)\, x(t-\tau)\, d\tau + u(t)\,, \tag{2.10}$$

where $x(t)$ and $y(t)$ are the input and output, respectively, $h(t,\tau)$ is the channel's time-varying impulse response, and $u(t)$ is zero-mean white Gaussian noise with power spectral density N_0. For physical reasons, $\mathbb{H}_c$ has to be causal and can only introduce finite delays. Hence, $h(t,\tau)$ is zero for τ outside the interval $[0, \tau_{\max}]$. Moreover, the variation of $h(t,\tau)$ with respect to t cannot be arbitrarily fast; it is limited by the maximum Doppler frequency $\nu_{\max}$.

System Descriptions

Further system descriptions for $\mathbb{H}_c$ that are equivalent to the impulse response $h(t,\tau)$ can be obtained via Fourier transformation. For our purpose, it is sufficient to introduce the *time-dependent transfer function*

$$H_{\mathbb{H}_c}(t,f) \triangleq \int_\tau h(t,\tau)\, e^{-j2\pi f\tau}\, d\tau\,, \tag{2.11}$$

and the *spreading function*

$$S_{\mathbb{H}_c}(\tau,\nu) \triangleq \int_t h(t,\tau)\, e^{-j2\pi\nu t}\, dt\,. \tag{2.12}$$

Note that $H_{\mathbb{H}_c}(t,f)$ and $S_{\mathbb{H}_c}(\tau,\nu)$ are related by a symplectic 2-D Fourier transformation. Since wireless channels are causal and introduce finite delays and finite Doppler shifts,

[1] Integrals go from $-\infty$ to ∞ unless specified otherwise.

the nonzero support of $S_{\mathbb{H}_c}(\tau,\nu)$ is limited to $(\tau,\nu) \in [0,\tau_{\max}] \times [-\nu_{\max}/2, \nu_{\max}/2]$ where $\tau_{\max} = 10\,\text{ns} \cdots 10\,\mu\text{s}$ and $\nu_{\max} = 0 \cdots 100\,\text{Hz}$. Hence, wireless channels are typically highly underspread, i.e., $\tau_{\max}\nu_{\max} \ll 1$ [1, 106, 137–140].

Since the spreading function $S_{\mathbb{H}_c}(\tau,\nu)$ is of limited support and in 2-D Fourier relation with $H_{\mathbb{H}_c}(t,f)$, the time-dependent transfer function $H_{\mathbb{H}_c}(t,f)$ is a 2-D lowpass function. Its variation with respect to time and frequency is proportional to $1/\nu_{\max}$ and $1/\tau_{\max}$, respectively.

Channel Statistics

Throughout this thesis we will assume a random channel $\mathbb{H}_c$ that is wide-sense stationary with uncorrelated scatterers (WSSUS) and Rayleigh fading [1, 3, 132, 133]. Due to the WSSUS assumption, the correlation function of the impulse response is given by

$$\mathrm{E}\{h(t,\tau)h^*(t',\tau')\} = D_{\mathbb{H}_c}(t-t',\tau)\,\delta(\tau-\tau')\,,$$

with the *time-delay correlation function* $D_{\mathbb{H}_c}(\Delta t,\tau)$ [1, 133]. That is, the impulse response is stationary in time (WSS) and uncorrelated for different delays (US). Equivalently, the WSSUS assumption implies that the time-dependent transfer function is 2-D stationary:

$$\mathrm{E}\left\{H_{\mathbb{H}_c}(t,f)H^*_{\mathbb{H}_c}(t',f')\right\} = R_{\mathbb{H}_c}(t-t',f-f')\,,$$

with the *time-frequency correlation function* $R_{\mathbb{H}_c}(\Delta t,\Delta f)$ [1, 133]. Finally, the spreading function is 2-D white:

$$\mathrm{E}\left\{S_{\mathbb{H}_c}(\tau,\nu)S^*_{\mathbb{H}_c}(\tau',\nu')\right\} = C_{\mathbb{H}_c}(\tau,\nu)\,\delta(\tau-\tau')\,\delta(\nu-\nu')\,,$$

with the *scattering function* $C_{\mathbb{H}_c}(\tau,\nu)$ [1, 133]. Due to the limited support of $S_{\mathbb{H}_c}(\tau,\nu)$, the support of $C_{\mathbb{H}_c}(\tau,\nu)$ is also limited to $(\tau,\nu) \in [0,\tau_{\max}] \times [-\nu_{\max}/2, \nu_{\max}/2]$ with probability one. Moreover, $C_{\mathbb{H}_c}(\tau,\nu)$ is real-valued and nonnegative. It has the interpretation of a 2-D power density spectrum of the time-varying channel. Furthermore, the time-frequency correlation function $R_{\mathbb{H}_c}(\Delta t,\Delta f)$ is essentially supported within $(\Delta t,\Delta f) \in \left[-1/\nu_{\max}, 1/\nu_{\max}\right] \times \left[-1/\tau_{\max}, 1/\tau_{\max}\right]$, i.e., the transfer function of the channel decorrelates for time lags larger than $1/\nu_{\max}$ and frequency lags larger than $1/\tau_{\max}$. The correlation functions $D_{\mathbb{H}_c}(\Delta t,\tau)$ and $R_{\mathbb{H}_c}(\Delta t,\Delta f)$ and the scattering function $C_{\mathbb{H}_c}(\tau,\nu)$ are related by Fourier transforms similar to (2.11) and (2.12), i.e.,

$$R_{\mathbb{H}_c}(\Delta t,\Delta f) \triangleq \int_\tau D_{\mathbb{H}_c}(\Delta t,\tau)\,e^{-j2\pi f\tau}\,d\tau$$

and

$$C_{\mathbb{H}_c}(\tau,\nu) \triangleq \int_{\nu} D_{\mathbb{H}_c}(\Delta t,\tau)\, e^{-j2\pi f \Delta t}\, d\Delta t\,. \tag{2.13}$$

It should be noted that the statistics of a WSSUS channel do not depend on time even though the channel realizations do.

Channel Parameters

The *path loss* of a WSSUS channel $\mathbb{H}_c$ is defined as the integral of the scattering function over all delays and Doppler frequencies, i.e,

$$\sigma^2_{\mathbb{H}_c} \triangleq \int_{\tau}\int_{\nu} C_{\mathbb{H}_c}(\tau,\nu)\, d\tau\, d\nu\,.$$

The *delay spread* $\alpha_{\mathbb{H}_c}$ and *Doppler spread* $\beta_{\mathbb{H}_c}$ are defined as the normalized second moments of the scattering function, i.e.,

$$\alpha^2_{\mathbb{H}_c} \triangleq \frac{1}{\sigma^2_{\mathbb{H}_c}} \int_{\tau}\int_{\nu} (\tau-\tau_0)^2 C_{\mathbb{H}_c}(\tau,\nu)\, d\tau\, d\nu$$

and

$$\beta^2_{\mathbb{H}_c} \triangleq \frac{1}{\sigma^2_{\mathbb{H}_c}} \int_{\tau}\int_{\nu} (\nu-\nu_0)^2 C_{\mathbb{H}_c}(\tau,\nu)\, d\tau\, d\nu\,.$$

Here, (τ_0,ν_0) denotes the center of gravity of $C_{\mathbb{H}_c}(\tau,\nu)$. Roughly, $\alpha_{\mathbb{H}_c}$ and $\beta_{\mathbb{H}_c}$ are in the same range as $\tau_{\max}$ and $\nu_{\max}$.

We next define the *coherence time* $T_{\mathbb{H}_c}$ and the *coherence bandwidth* $B_{\mathbb{H}_c}$ of $\mathbb{H}_c$ as

$$T_{\mathbb{H}_c} \triangleq \frac{1}{\beta_{\mathbb{H}_c}} \qquad \text{and} \qquad B_{\mathbb{H}_c} \triangleq \frac{1}{\alpha_{\mathbb{H}_c}}\,.$$

$T_{\mathbb{H}_c}$ and $B_{\mathbb{H}_c}$ are approximately in the same range as $1/\nu_{\max}$ and $1/\tau_{\max}$, respectively. Therefore, coherence time and coherence bandwidth also specify the effective support area of $R_{\mathbb{H}_c}(\Delta t,\Delta f)$, i.e., the transfer function $H_{\mathbb{H}_c}(t,f)$ approximately decorrelates for time lags Δt larger than $T_{\mathbb{H}_c}$ and frequency lags Δf larger than $B_{\mathbb{H}_c}$.

2.2.2 Discrete-Time Channel Model

Discretization of the continuous-time channel $\mathbb{H}_c$ in (2.10) according to the model sketched in Figure 2.8 yields a discrete-time channel $\mathbb{H}_d$. At the transmitter, digital-to-analog (D/A) conversion is used to convert the discrete-time signal $x[m]$ into a continuous-time signal

$$x(t) = \sum_{m=-\infty}^{\infty} x[m]\, \gamma(t-mT_s)\,, \tag{2.14}$$

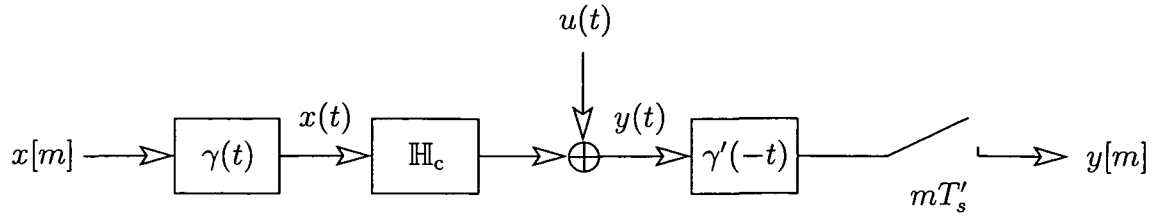


Figure 2.8: *Discretization of the continuous-time channel is performed by idealized D/A and A/D conversion at the transmitter and receiver, respectively.*

where $\gamma(t)$ is a transmit (interpolation) filter applied to each sample, $T_s = 1/B$ is the duration of one sample, and B is the transmit bandwidth. For simplicity we assume an ideal filter, i.e., $\gamma(t) = \sqrt{B}\,\mathrm{sinc}(\pi B t)$ with $\mathrm{sinc}(x) \triangleq \frac{\sin(x)}{x}$. This filter is normalized as $\|\gamma\|^2 = \int_t |\gamma(t)|^2 dt = 1$. The continuous-time transmit signal $x(t)$ propagates over the time-varying channel $\mathbb{H}_c$ and additive noise $u(t)$ is added.

At the receiver, analog-to-digital (A/D) conversion of the received signal $y(t)$ is performed according to

$$y[m] = \int_t y(t)\,\gamma'(t - mT'_s)\,dt\,, \tag{2.15}$$

where $\gamma'(t)$ is a receive (anti-alias and noise suppression) filter, $T'_s = 1/B'$ is the duration of one sample, and $B' = B + \nu_{\max}$ is the receive bandwidth. We assume an ideal filter with $\gamma'(t) = \sqrt{B'}\,\mathrm{sinc}(\pi B' t)$. The purpose of $\gamma'(t)$ is to limit the noise.

Channel Input-Output Relation

Since the bandwidth of $\gamma'(t)$ is $B' = B + \nu_{\max}$, the desired signal component in $y(t)$ (the information-bearing component $\int_\tau h(t,\tau)s(t-\tau)d\tau$) is not distorted by (2.15) but only sampled and scaled by $1/\sqrt{B'}$. Therefore, inserting (2.10) into (2.15) yields

$$y[m] = \frac{1}{\sqrt{B'}} \int_\tau h(mT'_s, \tau)\, x(mT'_s - \tau)\, d\tau + u[m]\,, \tag{2.16}$$

where

$$u[m] = \sqrt{B'} \int_t u(t)\,\mathrm{sinc}\left(\pi B'(t - mT'_s)\right) dt \tag{2.17}$$

is zero-mean white Gaussian discrete-time noise with variance $\sigma_u^2 = \mathrm{E}\{|u[m]|^2\} = N_0$. Furthermore, inserting (2.14) into (2.16) yields

$$y[m] = \sqrt{\frac{B}{B'}} \sum_{l=-\infty}^{\infty} x[m-l] \int_\tau h\big(mT'_s, \tau\big)\ \mathrm{sinc}\left(\pi B\big[\tau - m(T'_s - T_s) - lT_s\big]\right) d\tau + u[m]\,.$$

In this thesis, we consider wideband communication systems where the channel bandwidth is at least several MHz (e.g., in IEEE 802.11a the bandwidth is 20MHz). On the other hand, the Doppler frequency is maximally in the range of several tens of Hz. Hence, compared with B, $\nu_{\max}$ is negligibly small and therefore the sampling frequencies at transmitter and receiver are practically identical. We will consequently assume in the following that $B' = B$ and thus also $T_s' = T_s$. With this simplifying assumption, the input-output relation of the equivalent discrete-time channel $\mathbb{H}_\mathrm{d}$ is obtained as (cf. (2.10))

$$y[m] = (\mathbb{H}_\mathrm{d}x)[m] + u[m] = \sum_{l=-\infty}^{\infty} h[m,l]\,x[m-l] + u[m]\,, \tag{2.18}$$

with the time-varying discrete-time impulse response given by

$$h[m,l] \triangleq \int_\tau h(mT_s,\tau)\ \mathrm{sinc}\left(\pi B(\tau - lT_s)\right) d\tau\,. \tag{2.19}$$

For large channel bandwidths, (2.19) is well approximated as

$$h[m,l] \approx \frac{1}{B}\,h(mT_s, lT_s)\,. \tag{2.20}$$

Generally, $h[m,l]$ in (2.19) is nonzero for $l \in \mathbb{Z}$. However, for large B, it follows from (2.20) and the fact that $h(t,\tau) = 0$ for $\tau < 0$ and $\tau > \tau_{\max}$ that $h[m,l] \approx 0$ for $l < 0$ and $l > L$, where

$$L = \left\lceil \frac{\tau_{\max}}{T_s} \right\rceil\,. \tag{2.21}$$

Therefore, we will assume that $h[m,l]$ has nonzero support only for $l \in [0, L]$. Thus, (2.18) becomes

$$y[m] = (\mathbb{H}_\mathrm{d}x)[m] + u[m] = \sum_{l=0}^{L} h[m,l]\,x[m-l] + u[m]\,. \tag{2.22}$$

Note that the number of resolved channel delay taps is $L+1$.

System Description

For the discrete-time channel $\mathbb{H}_\mathrm{d}$, the *spreading function* is defined by [1, 133]

$$S_{\mathbb{H}_\mathrm{d}}(l,\xi) \triangleq \sum_{m=-\infty}^{\infty} h[m,l]\,e^{-j2\pi\xi m}\,. \tag{2.23}$$

Insertion of (2.19) and using the Fourier relation (2.12) yields

$$\begin{aligned} S_{\mathbb{H}_\mathrm{d}}(l,\xi) &= \int_\tau \int_\nu S_{\mathbb{H}_\mathrm{c}}(\tau,\nu)\ \mathrm{sinc}\left(\pi B(\tau - lT_s)\right) \sum_{m=-\infty}^{\infty} e^{j2\pi T_s\left(\nu - \frac{\xi}{T_s}\right)m}\, d\tau\, d\nu \\ &= B \sum_{m=-\infty}^{\infty} \int_\tau S_{\mathbb{H}_\mathrm{c}}\left(\tau, (\xi+m)B\right)\ \mathrm{sinc}\left(\pi B(\tau - lT_s)\right) d\tau\,, \end{aligned}$$

where we applied Poisson's sum formula. Note that $S_{\mathbb{H}_\mathrm{d}}(l,\xi)$ is periodic in the normalized Doppler variable ξ due to the discretization of the impulse response in the time direction. However, there does not occur aliasing in $S_{\mathbb{H}_\mathrm{d}}(l,\xi)$ because $S_{\mathbb{H}_\mathrm{c}}(\tau,\nu)$ is limited to Doppler frequencies $|\nu| \leq \nu_{\max}/2$ and $\nu_{\max} \ll B/2$. Hence, in the fundamental Doppler interval we obtain

$$S_{\mathbb{H}_\mathrm{d}}(l,\xi) = B \int_\tau S_{\mathbb{H}_\mathrm{c}}(\tau,\xi B)\, \mathrm{sinc}(\pi B(\tau - lT_s))\, d\tau\,, \qquad \xi \in [-1/2, 1/2)\,.$$

For large bandwidth B, this simplifies to

$$S_{\mathbb{H}_\mathrm{d}}(l,\xi) \approx S_{\mathbb{H}_\mathrm{c}}(lT_s, \xi B)\,. \tag{2.24}$$

In the fundamental Doppler interval, the support region for $S_{\mathbb{H}_\mathrm{d}}(l,\xi)$ is $(l,\xi) \in [0, L] \times [-\xi_{\max}/2, \xi_{\max}/2]$, where L is given in (2.21) and the maximum normalized Doppler frequency is

$$\xi_{\max} = T_s\, \nu_{\max} = \frac{\nu_{\max}}{B}\,. \tag{2.25}$$

The discrete-time channel $\mathbb{H}_\mathrm{d}$ is underspread because

$$\xi_{\max} L = \frac{\nu_{\max}}{B} \left\lceil \frac{\tau_{\max}}{T_s} \right\rceil \approx \nu_{\max}\, \tau_{\max} < 1\,.$$

Channel Statistics

The correlation function of the discrete-time impulse response $h[m,l]$ defined in (2.19) is given by

$$\mathrm{E}\left\{h[m,l]h^*[m',l']\right\} = \int_\tau D_{\mathbb{H}_\mathrm{c}}((m-m')T_s, \tau)\, \mathrm{sinc}(\pi B(\tau - lT_s))\, \mathrm{sinc}(\pi B(\tau - l'T_s))\, d\tau\,. \tag{2.26}$$

While the WSS property is preserved in the discrete-time channel $\mathbb{H}_\mathrm{d}$, the different taps of $\mathbb{H}_\mathrm{d}$ are in general correlated. However, these correlations are negligible for large bandwidth. Throughout this thesis, we therefore adopt the *discrete WSSUS* (DWSSUS) assumption defined by

$$\mathrm{E}\left\{h[m,l]h^*[m',l']\right\} = D_{\mathbb{H}_\mathrm{d}}[m-m', l]\, \delta[l-l']\,, \tag{2.27}$$

which corresponds to assuming that the integral in (2.26) vanishes for $l \neq l'$. The *time-delay correlation function* of $\mathbb{H}_\mathrm{d}$ is obtained from (2.26) for $l = l'$:

$$D_{\mathbb{H}_\mathrm{d}}[m,l] = \int_\tau D_{\mathbb{H}_\mathrm{c}}(mT_s, \tau) \left[\mathrm{sinc}(\pi B(\tau - lT_s))\right]^2 d\tau\,. \tag{2.28}$$

For large bandwidth, (2.28) simplifies to

$$D_{\mathbb{H}_d}[m,l] \approx \frac{1}{B^2} D_{\mathbb{H}_c}(mT_s, lT_s)\,.$$

In the spreading domain (delay-Doppler domain), the DWSSUS assumption can be expressed on the fundamental Doppler interval $\xi \in [-1/2, 1/2)$ as

$$\mathrm{E}\left\{S_{\mathbb{H}_d}(l,\xi) S^*_{\mathbb{H}_d}(l',\xi')\right\} = C_{\mathbb{H}_d}(l,\xi)\,\delta[l-l']\,\delta(\xi-\xi')\,, \qquad \xi,\xi' \in [-1/2, 1/2)\,. \tag{2.29}$$

Here, $C_{\mathbb{H}_d}(l,\xi)$ is the *scattering function* of $\mathbb{H}_d$, which is in Fourier relation with the time-delay correlation function:

$$C_{\mathbb{H}_d}(l,\xi) = \sum_{m=-\infty}^{\infty} D_{\mathbb{H}_d}[m,l]\, e^{-j2\pi\xi m}\,.$$

Inserting (2.28) and (2.13) yields

$$C_{\mathbb{H}_d}(l,\xi) = B^2 \int_\tau C_{\mathbb{H}_c}(\tau, \xi B)\, [\mathrm{sinc}(\pi B(\tau - lT_s))]^2\, d\tau\,,$$

and for large bandwidth we obtain

$$C_{\mathbb{H}_d}(l,\xi) \approx C_{\mathbb{H}_c}(lT_s, \xi B)\,, \qquad \xi \in [-1/2, 1/2)\,. \tag{2.30}$$

Hence, the scattering function of $\mathbb{H}_d$ is obtained by sampling $C_{\mathbb{H}_c}(\tau,\nu)$ in the delay (τ) direction and by scaling (normalizing) the Doppler frequency variable ν. The support region of $C_{\mathbb{H}_d}(l,\xi)$ is $(l,\xi) \in [0, L] \times [-\xi_{\max}, \xi_{\max}]$ where L and $\xi_{\max}$ are defined in (2.21) and (2.25), respectively.

Channel Parameters

The *path loss* of the discrete-time channel $\mathbb{H}_d$ is

$$\begin{aligned}\sigma^2_{\mathbb{H}_d} &\triangleq \sum_{l=0}^{L} \int_{-1/2}^{1/2} C_{\mathbb{H}_d}(l,\xi)\, d\xi \approx \sum_{l=0}^{L} \int_{-1/2}^{1/2} C_{\mathbb{H}_c}(lT_s, \xi B)\, d\xi \\ &\approx \frac{1}{T_s} \int_\tau \int_{-1/2}^{1/2} C_{\mathbb{H}_c}(\tau, \xi B)\, d\tau d\xi = \int_\tau \int_{-B/2}^{B/2} C_{\mathbb{H}_c}(\tau,\nu)\, d\tau d\nu = \sigma^2_{\mathbb{H}_c}\,,\end{aligned} \tag{2.31}$$

where we used (2.30) in the first approximation and assumed for the second approximation that T_s is small such that the summation over the delay taps can be replaced by integration over all delays. Hence, the path loss of the discrete-time channel is approximately equal to the path loss of the continuous-time channel.

The delay spread $\alpha_{\mathbb{H}_d}$ and the Doppler spread $\beta_{\mathbb{H}_d}$ of $\mathbb{H}_d$ are defined by

$$\alpha_{\mathbb{H}_d}^2 \triangleq \frac{1}{\sigma_{\mathbb{H}_d}^2} \sum_{l=0}^{L} \int_{-\xi_{max}}^{\xi_{max}} (l-l_0)^2 C_{\mathbb{H}_d}(l,\xi)\, d\xi \tag{2.32}$$

and

$$\beta_{\mathbb{H}_d}^2 \triangleq \frac{1}{\sigma_{\mathbb{H}_d}^2} \sum_{l=0}^{L} \int_{-\xi_{max}}^{\xi_{max}} (\xi-\xi_0)^2 C_{\mathbb{H}_d}(l,\xi)\, d\xi\,, \tag{2.33}$$

where (l_0, ξ_0) denotes the center of gravity of $C_{\mathbb{H}_d}(l,\xi)$. Inserting (2.30) and (2.31) into (2.32) yields

$$\begin{aligned}\alpha_{\mathbb{H}_d}^2 &\approx \frac{1}{\sigma_{\mathbb{H}_c}^2} \sum_{l=0}^{L} \int_{-\xi_{max}}^{\xi_{max}} (l-l_0)^2 C_{\mathbb{H}_c}(lT_s, \xi B)\, d\xi \\ &\approx \frac{1}{T_s^2 \sigma_{\mathbb{H}_c}^2} \int_\tau \int_\nu (\tau-\tau_0)^2 C_{\mathbb{H}_c}(\tau,\nu)\, d\tau d\nu = \frac{\alpha_{\mathbb{H}_c}^2}{T_s^2}\,.\end{aligned}$$

Hence, the delay spread of the discrete-time channel $\mathbb{H}_d$ is related to the delay spread of the continuous-time channel $\mathbb{H}_c$ as

$$\alpha_{\mathbb{H}_d} \approx \frac{\alpha_{\mathbb{H}_c}}{T_s} = B\alpha_{\mathbb{H}_c}\,.$$

A similar derivation for the Doppler spread (2.33) yields

$$\beta_{\mathbb{H}_d} \approx T_s \beta_{\mathbb{H}_c} = \frac{\beta_{\mathbb{H}_c}}{B}\,.$$

The delay spread and Doppler spread are inversely proportional to the coherence bandwidth and coherence time, respectively:

$$T_{\mathbb{H}_d} \triangleq \frac{1}{\beta_{\mathbb{H}_d}} \quad \text{and} \quad B_{\mathbb{H}_d} \triangleq \frac{1}{\alpha_{\mathbb{H}_d}}\,,$$

and thus it follows with the above approximations that

$$T_{\mathbb{H}_d} \approx \frac{T_{\mathbb{H}_c}}{T_s} = BT_{\mathbb{H}_c} \quad \text{and} \quad B_{\mathbb{H}_d} \approx T_s B_{\mathbb{H}_c} = \frac{B_{\mathbb{H}_c}}{B}\,.$$

2.3 Input-Output Relation of the OFDM System

For the communication system consisting of OFDM modulation (2.1), transmission over a time-varying channel (2.22), and OFDM demodulation (2.3), the system input-output relation is obtained as [106]

$$Y_{n,k} = \sum_{n'=-\infty}^{\infty} \sum_{k'=0}^{K-1} \langle \mathbb{H}_d\, g_{n',k'}, f_{n,k} \rangle\, X_{n',k'} + Z_{n,k}\,, \tag{2.34}$$

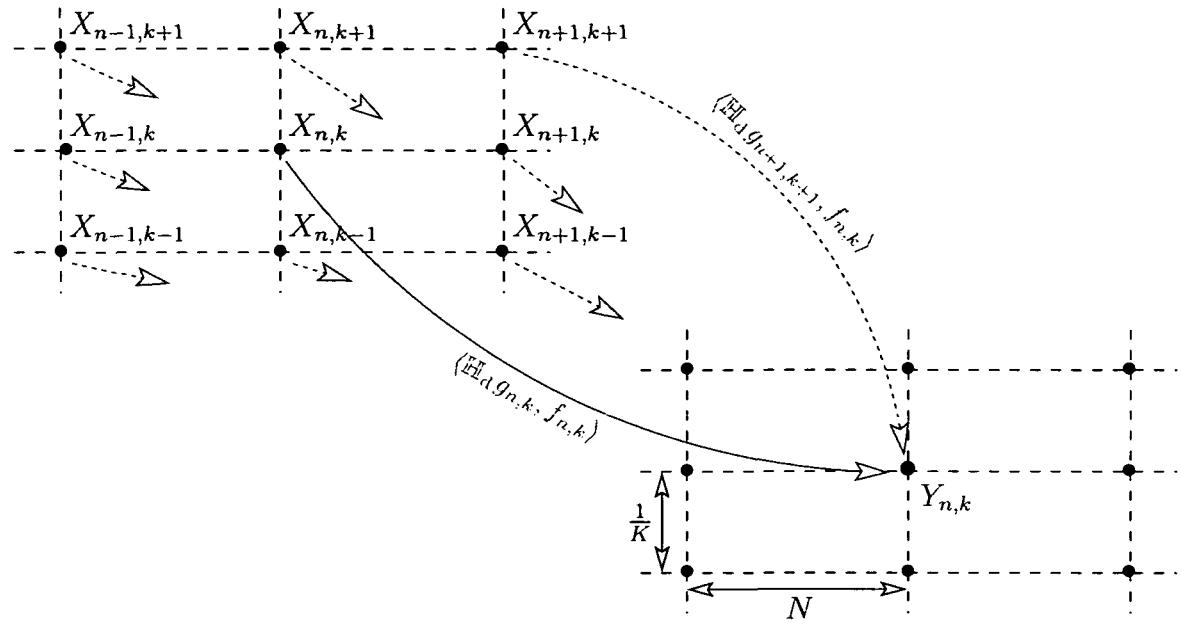


Figure 2.9: *With OFDM, one desires a multiplicative input-output relation where only* $\langle \mathbb{H}_d g_{n,k}, f_{n,k} \rangle X_{n,k}$ *contributes to* $Y_{n,k}$. *However, in general there exists interference from other transmitted data symbols.*

with

$$Z_{n,k} \triangleq \langle u, f_{n,k} \rangle . \tag{2.35}$$

Note that in (2.34), every transmit symbol $X_{n',k'}$ for $(n', k') \in \mathbb{Z} \times [0, K-1]$ contributes to a given received value $Y_{n,k}$. For this general input-output relation, decoding the information stream at the receiver would be highly complicated.

The sum in (2.34) can be split into a desired term and an interference term, which are respectively defined as

$$\langle \mathbb{H}_\mathrm{d}\, g_{n,k}, f_{n,k} \rangle X_{n,k} \qquad \text{and} \qquad \sum_{\substack{n'=-\infty \\ n' \neq n}}^{\infty} \sum_{\substack{k'=0 \\ k' \neq k}}^{K-1} \langle \mathbb{H}_\mathrm{d}\, g_{n',k'}, f_{n,k} \rangle X_{n',k'} .$$

This decomposition is illustrated in Figure 2.9. Frequently, the interference is further decomposed into *intersymbol interference* (ISI) and *intercarrier interference* (ICI) which are respectively defined by

$$\sum_{\substack{n'=-\infty \\ n' \neq n}}^{\infty} \sum_{k'=0}^{K-1} \langle \mathbb{H}_\mathrm{d}\, g_{n',k'}, f_{n,k} \rangle X_{n',k'} \qquad \text{and} \qquad \sum_{\substack{k'=0 \\ k' \neq k}}^{K-1} \langle \mathbb{H}_\mathrm{d}\, g_{n,k}, f_{n,k} \rangle X_{n,k} .$$

Thus, by ISI we denote all interference from past or future OFDM symbols, no matter at which subcarriers they are located. However, in most cases ISI is only experienced from the

neighboring OFDM symbols, and even this ISI is typically avoided by introducing a guard interval or CP between consecutive OFDM symbols. By ICI we denote the interference from different subcarriers within the same OFDM symbol. In most cases, substantial ICI contributions arise only from neighboring subcarriers. In contrast to ISI, ICI can never be completely avoided for time-varying channels. This is because the transmit pulses are distorted by the channel and the biorthogonality condition (2.5) is then not fulfilled anymore. However, for well-designed transmit and receive pulses and practical channels, the ICI terms are negligible compared to typical noise levels.

2.3.1 Approximate Input-Output Relation

The splitup of the system relation (2.34) into desired component and interference reflects the desire to obtain a system input-output relation that is approximately multiplicative and will be the basis for the rest of this thesis. The error introduced by neglecting the interference in (2.34) has been thoroughly analyzed in [24, 106, 108, 112, 113, 141–143]; related results can furthermore be found in [21, 105, 137, 138, 144]. From these investigations, we can draw the conclusion that for underspread channels and properly chosen pulses $g[m]$ and $f[m]$, the interference terms in (2.34) are approximately zero [105, 106, 108, 144]:

$$\langle \mathbb{H}_{\mathrm{d}}\, g_{n',k'}, f_{n,k} \rangle \approx 0 \qquad \text{for} \qquad (n',k') \neq (n,k)\,.$$

The system relation (2.34) then simplifies to the multiplicative input-output relation

$$Y_{n,k} \;=\; H_{n,k}\, X_{n,k} + Z_{n,k}\,. \tag{2.36}$$

where the *channel coefficients* are defined by

$$H_{n,k} \triangleq \langle \mathbb{H}_{\mathrm{d}}\, g_{n,k}, f_{n,k} \rangle\,.$$

Note that the $H_{n,k}$ depend on the channel and the transmit and receive pulses. In Subsection 2.3.2, we will discuss the properties of $H_{n,k}$ in terms of an equivalent channel $\mathbb{H}$.

Hence, because practical wireless channels are underspread, OFDM indeed allows to characterize the transmission over a wireless channel by pointwise multiplications of scalar complex-valued coefficients. This is the main advantage of this communication scheme over other wideband communication schemes. The practical benefit of this multiplicative system structure is the possibility of low-complexity equalization at the receiver.

An interpretation of the approximate system input-output relation (2.36) is that OFDM *diagonalizes* the time-varying channel [106, 145], i.e., the transmit pulses $g_{n,k}[m]$ in (2.2) and

the receive pulses $f_{n,k}[m]$ in and (2.4) are (approximate) *singular functions* of underspread wireless channels in the sense of the singular value decomposition [146]. This has the practical consequence that in OFDM systems, zero-forcing equalization of the received signal can be simply achieved through scalar multiplication of the $Y_{n,k}$ by the reciprocal of the $H_{n,k}$ [1].

According to (2.35), the additive Gaussian noise in (2.36) is given by $Z_{n,k} = \langle u, f_{n,k} \rangle$. Its correlation function is given by

$$R_Z[\Delta n, k, k'] = \mathrm{E}\left\{ Z_{n,k} Z^*_{n-\Delta n,k'} \right\} = N_0 \, e^{j2\pi \frac{N}{K} \Delta n k'} \langle f, f_{\Delta n, k-k'} \rangle \,, \tag{2.37}$$

since $u[m]$ in (2.17) is zero-mean white Gaussian noise with variance $\sigma_u^2 = N_0$. Hence, $Z_{n,k}$ is correlated if the receive pulse system $\{f_{n,k}\}$ is nonorthogonal. Moreover, $Z_{n,k}$ is stationary with respect to the OFDM symbol index n since its correlation function depends only on the difference Δn. However, for nonorthognal $\{f_{n,k}\}$, $Z_{n,k}$ is nonstationary with respect to the OFDM subcarrier index k. Uncorrelated and stationary noise is obtained if and only if the receive pulse set $\{f_{n,k}\}$ is orthogonal. In particular, if $\{f_{n,k}\}$ is orthonormal, the correlation function in (2.37) simplifies to $R_Z[\Delta n, k, k'] = N_0 \delta[\Delta n] \delta[k - k']$. In OFDM systems, which use identical transmit and receive pulses, the latter is always fulfilled. Furthermore, in CP-OFDM the receive pulse set is orthogonal and, hence, the noise is again white and stationary. Alternatively, $Z_{n,k}$ is stationary in k but correlated if $N/K \in \mathbb{N}$. However, such a choice of the OFDM lattice cell size is impractical since for $N/K = 1$ pulses of infinite length are necessary [21, 117] and for $N/K = 2, 3, \ldots$ the spectral efficiency is too small.

2.3.2 Equivalent Channel

We next introduce the concept of an *equivalent channel* to characterize the channel coefficients $H_{n,k} = \langle \mathbb{H}_\mathrm{d} \, g_{n,k}, f_{n,k} \rangle$ in (2.36). We can expect that the channel coefficients are related with the time-dependent transfer function $H_{\mathbb{H}_\mathrm{c}}(t, f)$ of the underlying continuous-time channel $\mathbb{H}_\mathrm{c}$ (see (2.11)) since the input-output relation in (2.36) characterizes the OFDM system in the time-frequency domain. Therefore, we express $H_{n,k}$ as the Fourier transform of a discrete-time time-varying impulse response $h_{n,l}$ in a manner analogous to (2.11),

$$H_{n,k} \triangleq \sum_{l=0}^{L} h_{n,l} \, e^{-j2\pi kl/K} \,. \tag{2.38}$$

The discrete-time system $\mathbb{H}$ whose impulse response is $h_{n,l}$ will be termed the *equivalent channel*. Thus, $H_{n,k}$ has the desired interpretation of the time-dependent transfer function

of $\mathbb{H}$. According to its definition, $h_{n,l}$ is given by

$$h_{n,l} \triangleq \sum_{m=-\infty}^{\infty} h[m,l]\, g[m-l-nN]\, f^*[m-nN]\,. \tag{2.39}$$

It is hence seen that the equivalent channel $\mathbb{H}$ incorporates the effects of the discrete-time channel $\mathbb{H}_\mathrm{d}$, of the transmit pulse $g[n]$, and of the receive pulse $f[n]$.

For a CP-OFDM system, (2.39) simplifies to

$$h_{n,l} = \frac{1}{K} \sum_{m=0}^{K-1} h[nN+m,l] \approx h[nN,l]\,, \tag{2.40}$$

where the approximation is accurate if the variation of $h[m,l]$ within an OFDM symbol interval is negligible. Hence, in a CP-OFDM system the impulse response of the equivalent channel $\mathbb{H}$ is obtained by subsampling the impulse response of the discrete-time channel $\mathbb{H}_\mathrm{d}$. Moreover, with (2.20) we obtain that $h_{n,l} \approx \frac{1}{B}\, h(nNT_s, lT_s)$. Hence, for CP-OFDM systems the impulse response of $\mathbb{H}$ is essentially obtained by sampling the impulse response of $\mathbb{H}_\mathrm{c}$. Even if (2.39) is more complicated than the special case (2.40), the same effect is basically observed there.

System Descriptions

The time-dependent transfer function $H_{n,k}$, expressed in terms of $h_{n,l}$ according to (2.38), was our starting point for the definition of the equivalent channel $\mathbb{H}$. Using (2.19), (2.38), (2.39), and the Fourier relation (2.11), it can be shown that $H_{n,k}$ is approximately given by

$$H_{n,k} \approx H_{\mathbb{H}_\mathrm{c}}(nT, kF)\,,$$

where we defined $T \triangleq NT_s$ and $F \triangleq B/K$. Therefore, $H_{n,k}$ is (approximately) a discretized version of the time-dependent transfer function of $\mathbb{H}_\mathrm{c}$ sampled on the rectangular time-frequency lattice of cell size $T \times F$. Recall that the time-frequency lattice of the discrete-time OFDM system had cell size $N \times 1/K$. Hence, we can conclude that the continuous-time OFDM symbol duration is $T = NT_s$ and the subcarrier frequency spacing is $F = B/K$.

We furthermore define the spreading function of $\mathbb{H}$ in a manner analogous to (2.23),

$$S_\mathbb{H}(l,\varphi) = \sum_{n=-\infty}^{\infty} h_{n,l}\, e^{-j2\pi\varphi n}\,. \tag{2.41}$$

Inserting (2.39) and (2.23) into (2.41) shows that $S_\mathbb{H}(l,\varphi)$ is related with the spreading

function of $\mathbb{H}_\mathrm{d}$ as

$$S_{\mathbb{H}}(l,\varphi) = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2}\int_{-1/2}^{1/2} S_{\mathbb{H}_\mathrm{d}}(l,\xi)\, G(\xi')\, F^*(\xi'')\, e^{-j2\pi l\xi'} \sum_{m=-\infty}^{\infty} e^{j2\pi(\xi+\xi'-\xi'')m} \cdot \sum_{n=-\infty}^{\infty} e^{-j2\pi N(\xi'-\xi''+\varphi/N)n}\, d\xi d\xi' d\xi''$$

$$= \frac{1}{N}\sum_{n=0}^{N-1} S_{\mathbb{H}_\mathrm{d}}\left(l, \frac{\varphi+n}{N}\right) A^*_{f,g}\left(l, \frac{\varphi+n}{N}\right). \tag{2.42}$$

Here, we used Poisson's sum formula and introduced the Fourier transforms of the transmit and receive pulses as $G(\xi) = \sum_{m=-\infty}^{\infty} g[m]\, e^{-j2\pi\xi m}$ and $F(\xi) = \sum_{m=-\infty}^{\infty} f[m]\, e^{-j2\pi\xi m}$, respectively. We also used the fact that the cross-ambiguity function defined in (2.6) can be expressed by

$$A_{f,g}(l,\xi) = \int_{-1/2}^{1/2} F(\xi')\, G^*(\xi'-\xi)\, e^{j2\pi(\xi'-\xi)l} d\xi',$$

and has the symmetry property $A_{f,g}(l,\xi) = e^{j2\pi l\xi} A^*_{g,f}(-l,-\xi)$. Furthermore, we note that $S_{\mathbb{H}_\mathrm{d}}(l,\xi)$ and $A_{f,g}(l,\xi)$ in (2.42) are periodic in the normalized Doppler frequency ξ with period one.

In the fundamental Doppler interval, we obtain from (2.42) that if aliasing is avoided, the spreading function of $\mathbb{H}$ is given by

$$S_{\mathbb{H}}(l,\varphi) = \frac{1}{N} S_{\mathbb{H}_\mathrm{d}}\left(l, \frac{\varphi}{N}\right) A^*_{f,g}\left(l, \frac{\varphi}{N}\right), \qquad \varphi \in [-1/2, 1/2). \tag{2.43}$$

Aliasing is avoided if (cf. (2.25))

$$\varphi_{\max} = N\xi_{\max} = N\frac{\nu_{\max}}{B} \le \frac{1}{2}. \tag{2.44}$$

In practical systems, this condition is always fulfilled because $\nu_{\max}$ is maximally in the range of several tens of Hz and B is minimally in the range of several MHz. Therefore, aliasing may occur only if N is larger than 10^5, which is not realistic (recall that in current WLAN systems we have $N = 80$ and even in DVB-T we maximally have $N = 10240$).

Equation (2.43) shows that the influence of the pulse-shaping OFDM system (in the spreading domain) is a weighting of the spreading function of $\mathbb{H}_\mathrm{d}$ with the cross-ambiguity function of the receive and transmit pulses and a normalization of the Doppler variable by the OFDM symbol duration N. For large bandwidths, we can use (2.24), with which (2.43) simplifies to

$$S_{\mathbb{H}}(l,\varphi) \approx \frac{1}{N} S_{\mathbb{H}_\mathrm{c}}\left(lT_s, \frac{\varphi}{NT_s}\right) A^*_{f,g}\left(l, \frac{\varphi}{N}\right), \qquad \varphi \in [-1/2, 1/2).$$

Hence, the spreading function of the continuous-time channel $\mathbb{H}_c$ is sampled in the delay direction with sampling period T_s and the Doppler frequency is normalized by NT_s.

The influence of the pulse shapes $g[m]$ and $f[m]$ on $\mathbb{H}$ is given by the weighting of $S_{\mathbb{H}_d}(l,\varphi)$ by $A_{f,g}(l,\varphi)$. Generally, the support region of $S_{\mathbb{H}_d}(l,\varphi)$ is small so that only the shape of $A_{f,g}(l,\varphi)$ about the origin comes into play. Within this support region, the magnitude of $A_{f,g}(l,\varphi)$ is close to one (see Figure 2.6). (In particular, recall that $A_{f,g}(0,0) = 1$ due to the biorthogonality of $g[m]$ and $f[m]$.) Hence, the influence on $\mathbb{H}$ of the particular pulses used by the OFDM system is small.

Channel Statistics

If $\mathbb{H}_d$ is DWSSUS (see (2.27)), then the equivalent channel $\mathbb{H}$ is also DWSSUS because the correlation function of the impulse response of $\mathbb{H}$ is given by

$$\mathrm{E}\{h_{n,l}\, h^*_{n',l'}\} = D_{\mathbb{H}}[n-n',l]\,\delta[l-l']\,, \tag{2.45}$$

where we used (2.27) and (2.39). Here, the time-delay correlation function is defined by

$$D_{\mathbb{H}}[n,l] \triangleq \sum_{m=-\infty}^{\infty}\sum_{m'=-\infty}^{\infty} D_{\mathbb{H}_d}[nN+m-m',l]\, g[m-l]\, g^*[m'-l]\, f[m']\, f^*[m]\,. \tag{2.46}$$

The channel coefficients $H_{n,k}$ in (2.38) constitute a 2-D stationary process with time-frequency correlation function

$$R_{\mathbb{H}}[n,k] \triangleq \mathrm{E}\{H_{n',k'}\, H^*_{n'-n,k'-k}\} = \sum_{l=0}^{L} D_{\mathbb{H}}[n,l]\, e^{-j2\pi kl/K}. \tag{2.47}$$

Moreover, since we assumed that $\mathbb{H}_c$ is Rayleigh fading, $H_{n,k}$ is circularly symmetric complex Gaussian (i.e., real part and imaginary part are independent and both are Gaussian). The support region of $R_{\mathbb{H}}[n,k]$ is essentially concentrated within $|n| \le 1/(N\beta_{\mathbb{H}_d})$ and $|k| \le K/\alpha_{\mathbb{H}_d}$. Hence, the channel coefficients $H_{n,k}$ are effectively correlated only for a limited range of time and frequency lags.

In the spreading domain, we obtain that $S_{\mathbb{H}}(l,\varphi)$ in (2.43) is 2-D white (cf. (2.29)) with the scattering function given by

$$C_{\mathbb{H}}(l,\varphi) = \sum_{m=-\infty}^{\infty} D_{\mathbb{H}}[m,l]\, e^{j2\pi\varphi m} = \frac{1}{N} C_{\mathbb{H}_d}\Big(l,\frac{\varphi}{N}\Big)\Big|A_{f,g}\Big(l,\frac{\varphi}{N}\Big)\Big|^2\,, \quad \varphi \in [-1/2,1/2)\,. \tag{2.48}$$

The support region of the scattering function $C_{\mathbb{H}}(l,\varphi)$ is $(l,\varphi) \in [0,L]\times[-\varphi_{\max},\varphi_{\max}]$ where L is defined in (2.21) and $\varphi_{\max}$ is given by (2.44). The support region of $|A_{f,g}(l,\varphi)|^2$

is much larger than that of $C_{\mathbb{H}_\mathrm{d}}(l,\varphi)$. Moreover, $|A_{f,g}(l,\varphi)|^2$ is close to one in the support region of $C_{\mathbb{H}_\mathrm{c}}(\tau,\nu)$, so that (2.48) can be well approximated by

$$C_{\mathbb{H}}(l,\varphi) \approx \frac{1}{N} C_{\mathbb{H}_\mathrm{d}}\left(l, \frac{\varphi}{N}\right) \approx \frac{1}{N} C_{\mathbb{H}_\mathrm{c}}\left(lT_s, \frac{\varphi}{NT_s}\right), \qquad \varphi \in [-1/2, 1/2). \tag{2.49}$$

Channel Parameter

The path loss of the equivalent channel is given by

$$\sigma_{\mathbb{H}}^2 \triangleq \sum_{l=0}^{L} \int_{-1/2}^{1/2} C_{\mathbb{H}}(l,\varphi)\, d\varphi = \sum_{l=0}^{L} \int_{-1/(2N)}^{1/(2N)} C_{\mathbb{H}_\mathrm{d}}(l,\xi)\, |A_{f,g}(l,\xi)|^2\, d\xi, \tag{2.50}$$

where the equality is due to (2.48) and (2.44) was assumed to be fulfilled. If we furthermore use the approximation (2.49), we obtain that the path losses of the channels $\mathbb{H}$, $\mathbb{H}_\mathrm{d}$, and $\mathbb{H}_\mathrm{c}$ are approximately equal:

$$\sigma_{\mathbb{H}}^2 \approx \sigma_{\mathbb{H}_\mathrm{d}}^2 \approx \sigma_{\mathbb{H}_\mathrm{c}}^2. \tag{2.51}$$

We furthermore note that using (2.47) and (2.48), the path loss of the equivalent channel $\mathbb{H}$ is obtained as

$$\sigma_{\mathbb{H}}^2 = R_{\mathbb{H}}[0,0] = \mathrm{E}\left\{|H_{n,k}|^2\right\}. \tag{2.52}$$

3

Channel Prediction in OFDM Systems

In most communication systems, *channel state information* (CSI) is required at the receiver. Therefore, training data is classically sent which enables the receiver to estimate the channel. If the channel is time-invariant, the training data is typically sent in one block preceding the data packet. However, in wireless communications the channel is subject to time and frequency selective fading. This necessitates continuous tracking of the channel, which is usually based on a periodically repeated transmission of training data. The amount of training data necessary to estimate and track the channel increases with the fading rate of the channel and can occupy a significant percentage of the transmission time.

The problems and drawbacks of channel estimation (cf. Subsection 1.2.2) can be overcome if channel prediction is used to obtain CSI. In this chapter, we propose novel schemes for decision-directed MMSE and adaptive channel prediction in wireless OFDM systems. Our channel predictors feature several advantages. They yield accurate, up-to-date CSI without additional latencies, even for fast time-varying channels or large prediction horizons. Due to their decision-directed mode of operation, no continual transmission of training data is required. Typically, a single known OFDM symbol suffices for the initialization

of the predictor. Thanks to efficient FFT implementations, the computational complexity of the proposed channel predictors is moderate. Our predictors enable key techniques for advanced wireless communication schemes such as antenna combining, space-time decoding, adaptive modulation, adaptive power control, and adaptive transmit diversity. Thus, they can help improve the system capacity and/or link reliability of wireless OFDM systems.

For the design of our MMSE channel predictors, knowledge of the second-order channel statistics is required; this is similar to the case of MMSE channel estimators. These channel statistics are unknown in practical applications, and thus they would have to be estimated prior to the design of the MMSE channel predictor. In the context of pilot symbol assisted MMSE channel estimation, this is considered in [64]. However, this approach is problematic since practical channels are stationary only for a certain time [135, 136]. Fortunately, the estimation of the channel statistics can be completely avoided by the application of *adaptive* channel predictors. The adaptive versions of our channel predictors do not require any statistical prior knowledge and are able to track nonstationary channel and noise statistics. Moreover, we demonstrate by simulations that the adaptive predictors perform close to the MMSE predictor.

Most parts of this chapter have been previously published by the author in [92–94]. Very recently, MMSE channel prediction for OFDM has independently been proposed in [65]. Further independent work on channel prediction for OFDM and its applications can be found in [90, 91]; we will briefly summarize the main differences of this work from our work. In [90], decision-directed channel prediction for equalization is proposed. *Noiseless* MMSE prediction is considered; this assumes that the channel impulse response can be observed without any error, which is however not a realistic assumption in practice. In contrast to [90], we propose a generic receiver structure employing channel prediction that can also be used for several purposes beyond equalization, such as adaptive modulation etc. Furthermore, we derive the MMSE predictor and simplified implementations for practical *noisy* systems. Additionally, we consider the performance of the MMSE predictor for infinite predictor memory and in specular scattering environments. Finally, we also propose adaptive predictors that do not require estimation of the channel statistics. In [91], pilot symbol assisted MMSE channel prediction is proposed and applied to adaptive modulation for HIPERLAN/2. This differs fundamentally from our approach since we avoid transmission of pilot symbols by using a decision-directed mode of operation.

In a non-OFDM context, the prediction of fading channels and its applications were previously investigated in [50, 86–89]. In particular, the generic concept of the prediction of fading signals is described in [50], and in [89] the application of channel prediction to adaptive modulation is thoroughly investigated.

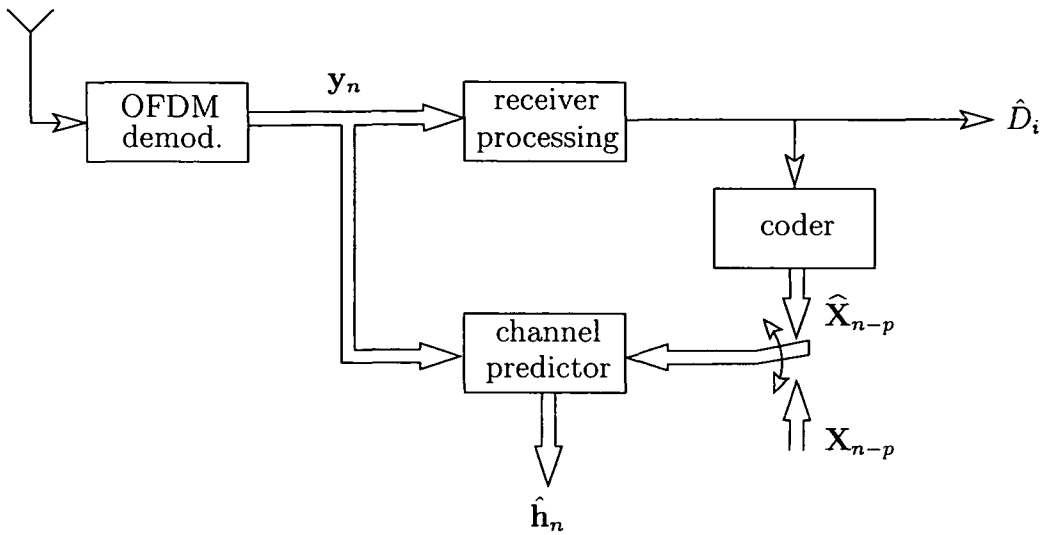


Figure 3.1: *OFDM receiver using channel prediction. The upper branch is a conventional OFDM receiver that outputs detected bits* $\widehat{D}_i$*. The lower part is required for channel prediction. During training mode (initialization), the switch is down.*

This chapter is organized as follows. In Section 3.1, we present the generic structure of a wireless OFDM receiver employing channel prediction. In Section 3.2, we derive the MMSE channel predictor. It will be seen that the computational complexity of the "full-blown" MMSE predictor is impractical. We therefore introduce a reduced-complexity MMSE predictor that allows an efficient DFT-based implementation. This DFT structure is also used to derive an analytic expression for the MMSE of the infinite-length one-step predictor and to show that error-free channel prediction is possible in point scattering environments. In Section 3.3, we develop adaptive channel predictors that do not require knowledge of the channel statistics. Specifically, we use the normalized least-mean-square (NLMS) algorithm and the recursive least-squares (RLS) algorithm for the adaptation of the channel predictor coefficients. In Section 3.4, applications of the proposed channel predictors are considered. We discuss predictive equalization, present an adaptive modulation scheme, and briefly consider pilot symbol augmented channel prediction. In Section 3.5, we finally assess the performance of our predictors by means of computer simulations.

3.1 OFDM Receiver Applying Channel Prediction

Figure 3.1 shows the generic receiver structure that we propose for channel prediction in wireless OFDM systems. This scheme yields CSI that is required for coherent receiver

processing (including, e.g., equalization, detection, and decoding) and can also be used for advanced techniques such as adaptive modulation. As a difference from [90], we consider the realistic case of noisy prediction. In contrast to the channel estimators in [55, 57, 59, 60, 64, 70, 74, 147] and to the predictive adaptive loading method in [91], our scheme operates in a decision-directed mode. Furthermore, different from the decision-directed channel estimators in [76–79], our channel predictor is able to yield up-to-date CSI.

We will now explain the structure shown in Figure 3.1. We recall from (2.36) that the OFDM system input-output relation is $Y_{n,k} = H_{n,k}X_{n,k} + Z_{n,k}$. In vector notation, this input-output relation can be written as

$$\begin{aligned} \mathbf{y}_n &= \mathbf{H}_n\mathbf{x}_n + \mathbf{z}_n \\ &= \mathbf{X}_n\mathbf{h}_n + \mathbf{z}_n\,, \end{aligned} \tag{3.1}$$

with the $K \times 1$ vectors $\mathbf{y}_n \triangleq [Y_{n,0}\, Y_{n,1} \cdots Y_{n,K-1}]^T$, $\mathbf{x}_n \triangleq [X_{n,0}\, X_{n,1} \cdots X_{n,K-1}]^T$, $\mathbf{h}_n \triangleq [H_{n,0}\, H_{n,1} \cdots H_{n,K-1}]^T$, $\mathbf{z}_n \triangleq [Z_{n,0}\, Z_{n,1} \cdots Z_{n,K-1}]^T$ and the diagonal $K \times K$ matrices $\mathbf{X}_n \triangleq \mathrm{diag}\{X_{n,0},\, X_{n,1},\, \cdots,\, X_{n,K-1}\}$ and $\mathbf{H}_n \triangleq \mathrm{diag}\{H_{n,0},\, H_{n,1},\, \cdots,\, H_{n,K-1}\}$.

Our decision-directed channel predictors process the demodulated receive vector $\mathbf{y}_n$ in order to yield an estimate $\hat{\mathbf{h}}_n \triangleq [\widehat{H}_{n,0}\, \widehat{H}_{n,1} \cdots \widehat{H}_{n,K-1}]^T$ of the current channel coefficient vector $\mathbf{h}_n$. To this end, past detected symbol matrices $\widehat{\mathbf{X}}_{n-p}$, $\widehat{\mathbf{X}}_{n-p-1}$, ...obtained by re-encoding previously detected bits $\widehat{D}_i$ are used. Here, $\widehat{\mathbf{X}}_n \triangleq \mathrm{diag}\{\widehat{X}_{n,0},\, \widehat{X}_{n,1},\, \cdots,\, \widehat{X}_{n,K-1}\}$ is a diagonal $K \times K$ matrix containing the detected data symbols. Furthermore, p denotes the *prediction horizon*, i.e., the number of OFDM symbols the channel is predicted ahead; its choice determined e.g. by the latency introduced by the receiver processing and the re-encoder (cf. Figure 3.1). The detected bits $\widehat{D}_i$ are obtained by OFDM demodulation and conventional receiver processing such as equalization and detection. Because the $\widehat{D}_i$ may be incorrect, re-encoding them yields (partly) incorrect $\widehat{\mathbf{X}}_n$. If too many $\widehat{\mathbf{X}}_n$ are incorrect, the accuracy of the predicted channel $\hat{\mathbf{h}}_n$ suffers. In particular, if $\hat{\mathbf{h}}_n$ is used e.g. for receiver processing in a feedback loop, error propagation may result. Simulation results in Section 3.5.5 show that error propagation is avoided if the SNR is above a certain threshold.

For initialization of the predictor, training data are required. During the training phase, a few known transmit symbols $\mathbf{X}_n$ serve as training data (numerical experiments in Section 3.5.1 suggest that already one OFDM symbol suffices for initialization). In packet-oriented communication systems such as IEEE 802.11a, each data packet is preceded by a known preamble which suffices for initialization of the predictor.

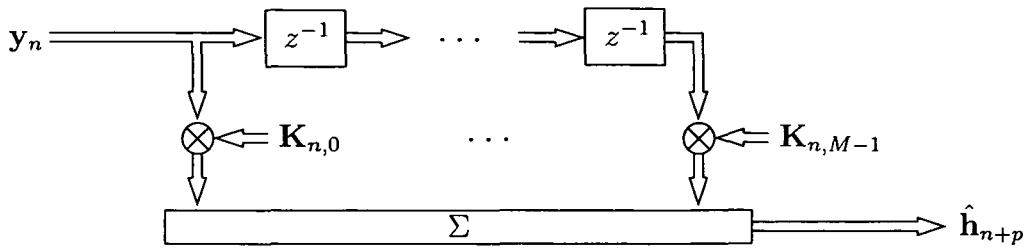


Figure 3.2: *The full-complexity MMSE channel predictor for OFDM is a linear time-varying MIMO filter of length M.*

3.2 MMSE Channel Predictors

We will now develop the MMSE predictor of the future channel coefficients[1] $H_{n+p,k}$, with prediction horizon $p \geq 1$. We have seen in Section 3.1 that the channel predictor can be operated in training mode or in decision-directed mode. For our derivation, we will assume training mode operation for mathematical tractability. Hence, the current and past data symbols $X_{n,k}$ are assumed known; the effects of detection errors are disregarded. Furthermore, we neglect ISI and ICI and thus assume that (2.36) or equivalently (3.1) holds[2], i.e., $Y_{n,k} = H_{n,k} X_{n,k} + Z_{n,k}$ or $\mathbf{y}_n = \mathbf{X}_n \mathbf{h}_n + \mathbf{z}_n$.

3.2.1 Full-Complexity MMSE Predictor

The MMSE predictor calculates a predicted channel coefficient vector $\hat{\mathbf{h}}_{n+p} = [\widehat{H}_{n+p,0}\ \widehat{H}_{n+p,1} \cdots \widehat{H}_{n+p,K-1}]^T$ from the current and past received vectors $\mathbf{y}_n, \mathbf{y}_{n-1}, \ldots, \mathbf{y}_{n-M+1}$ by means of a linear multi-input multi-output (MIMO) predictor filter of length M, i.e.,

$$\hat{\mathbf{h}}_{n+p} = \sum_{m=0}^{M-1} \mathbf{K}_{n,m}\, \mathbf{y}_{n-m}\,. \tag{3.2}$$

Figure 3.2 shows the structure of this predictor filter. A MIMO filter is used to exploit the correlations between different subcarriers that exist due to the channel's frequency correlation (nonzero coherence bandwidth). The memory of the predictor (corresponding to the filter length M) is used to exploit the time correlation between consecutive OFDM symbols that is introduced by the channel's time correlation (nonzero coherence time). The predictor coefficient matrices $\mathbf{K}_{n,m}$ of size $K \times K$ are time-varying to account for the nonstationarity of the received vector process $\mathbf{y}_n$. This nonstationarity is due to the

[1] For convenience, we shift the problem statement by p OFDM symbols.

[2] An analysis of the resulting errors can be found in e.g. [106, 108, 143].

multiplication of the stationary vector process $\mathbf{h}_n$ by the deterministic, time-varying symbol matrix $\mathbf{X}_n$ in (3.1) (recall that the transmit symbols are assumed known for the derivation of the MMSE predictor). Note that a linear predictor suffices since $\mathbf{h}_n$ and $\mathbf{z}_n$ in (3.1) are both circularly symmetric complex Gaussian [109, 148, 149].

Optimum Predictor Coefficients

The MMSE-optimum predictor coefficients $\mathbf{K}_{n,m}$ minimize the normalized MSE that is defined as [109, 148, 149]

$$\epsilon_{n+p} \triangleq \frac{1}{K} \mathrm{E}\{\|\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{n+p}\|^2\}.$$

According to the orthogonality principle [148], the optimum $\mathbf{K}_{n,m}$ are such that

$$\mathrm{E}\{(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{n+p})\, \mathbf{y}_{n-m}^H\} = \mathbf{0}, \qquad \text{for} \qquad m = 0, \ldots, M-1, \tag{3.3}$$

i.e., the prediction error vector $\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{n+p}$ is uncorrelated with all available observations $\mathbf{y}_n$. To streamline the calculation of the optimum $\mathbf{K}_{n,m}$, we rewrite the input-output relation of the MIMO predictor (3.2) as

$$\hat{\mathbf{h}}_{n+p} = \boldsymbol{\mathcal{K}}_n \boldsymbol{\mathcal{Y}}_n, \tag{3.4}$$

with the $KM \times KM$ predictor coefficient block matrix $\boldsymbol{\mathcal{K}}_n \triangleq [\mathbf{K}_{n,0} \cdots \mathbf{K}_{n,M-1}]$ and the $KM \times 1$ stacked vector $\boldsymbol{\mathcal{Y}}_n \triangleq [\mathbf{y}_n^T \cdots \mathbf{y}_{n-M+1}^T]^T$. Upon insertion of (3.1) and (3.4) into the orthogonality relation (3.3), the normal equations (Wiener-Hopf equations) are obtained as

$$\boldsymbol{\mathcal{K}}_n \boldsymbol{\mathcal{X}}_n \left(\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \boldsymbol{\mathcal{X}}_n^{-1} \boldsymbol{\mathcal{R}}_Z \boldsymbol{\mathcal{X}}_n^{-H}\right) = \boldsymbol{\mathcal{V}}_{\mathbb{H}}. \tag{3.5}$$

Here, $\boldsymbol{\mathcal{X}}_n \triangleq \mathrm{diag}\{\mathbf{X}_n, \ldots, \mathbf{X}_{n-M+1}\}$ is a diagonal matrix with the data symbols $X_{n-M+1,k}, \ldots, X_{n,k}$ as diagonal elements, the stacked channel correlation matrix $\boldsymbol{\mathcal{V}}_{\mathbb{H}}$ is defined as $\boldsymbol{\mathcal{V}}_{\mathbb{H}} \triangleq [\mathbf{R}_{\mathbb{H}}[p] \cdots \mathbf{R}_{\mathbb{H}}[p+M-1]]$, and the correlation matrices $\boldsymbol{\mathcal{R}}_{\mathbb{H}}$ and $\boldsymbol{\mathcal{R}}_Z$ are block-Toeplitz with first rows $[\mathbf{R}_{\mathbb{H}}[0] \cdots \mathbf{R}_{\mathbb{H}}[M-1]]$ and $[\mathbf{R}_Z[0] \cdots \mathbf{R}_Z[M-1]]$, respectively. Furthermore, the channel correlation matrix and the noise correlation matrix are given by

$$\mathbf{R}_{\mathbb{H}}[n] \triangleq \mathrm{E}\{\mathbf{h}_m \mathbf{h}_{m-n}^H\} = \begin{bmatrix} R_{\mathbb{H}}[n,0] & \ldots & R_{\mathbb{H}}[n,-K+1] \\ \vdots & & \vdots \\ R_{\mathbb{H}}[n,K-1] & \ldots & R_{\mathbb{H}}[n,0] \end{bmatrix} \tag{3.6}$$

and

$$\mathbf{R}_Z[n] \triangleq \mathrm{E}\{\mathbf{z}_m \mathbf{z}_{m-n}^H\} = \begin{bmatrix} R_Z[n,0,0] & \ldots & R_Z[n,0,K-1] \\ \vdots & & \vdots \\ R_Z[n,K-1,0] & \ldots & R_Z[n,K-1,K-1] \end{bmatrix},$$

respectively, where $R_{\mathbb{H}}[n,k]$ is defined in (2.47) and $R_Z[n,k,k']$ is given by (2.37). With (3.5), the optimum predictor coefficients are obtained as

$$\boldsymbol{\mathcal{K}}_{\text{opt},n} = \boldsymbol{\mathcal{W}}_n \boldsymbol{\mathcal{X}}_n^{-1}, \qquad \text{with } \boldsymbol{\mathcal{W}}_n \triangleq \boldsymbol{\mathcal{V}}_{\mathbb{H}} \left(\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \boldsymbol{\mathcal{X}}_n^{-1} \boldsymbol{\mathcal{R}}_Z \boldsymbol{\mathcal{X}}_n^{-H} \right)^{-1}. \tag{3.7}$$

We deliberately decompose $\boldsymbol{\mathcal{K}}_{\text{opt},n}$ as in (3.7) to partition the MMSE predictor into two operations/stages. We will come back to this point presently when considering the implementation of this predictor.

Minimum Mean Square Error

Again using the orthogonality principle [148], the error covariance matrix of the MMSE channel predictor is given by

$$\begin{aligned} \mathbf{B}_n &= \mathrm{E}\left\{ \left(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p} \right) \left(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p} \right)^H \right\} = \mathrm{E}\left\{ (\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p}) \mathbf{h}_{n+p}^H \right\} \\ &= \mathbf{R}_{\mathbb{H}}[0] - \boldsymbol{\mathcal{W}}_n \boldsymbol{\mathcal{V}}_{\mathbb{H}}^H = \mathbf{R}_{\mathbb{H}}[0] - \boldsymbol{\mathcal{V}}_{\mathbb{H}} \left(\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \boldsymbol{\mathcal{X}}_n^{-1} \boldsymbol{\mathcal{R}}_Z \boldsymbol{\mathcal{X}}_n^{-H} \right)^{-1} \boldsymbol{\mathcal{V}}_{\mathbb{H}}^H, \end{aligned} \tag{3.8}$$

where $\hat{\mathbf{h}}_{\text{opt},n+p} = \boldsymbol{\mathcal{K}}_{\text{opt},n} \boldsymbol{\mathcal{Y}}_n$. Note that the error covariance matrix depends on the actual OFDM symbol via the transmit symbols $X_{n,k}$ that are contained in the diagonal matrix $\boldsymbol{\mathcal{X}}_n$. The normalized MMSE achieved by $\boldsymbol{\mathcal{K}}_{\text{opt},n}$ is given by

$$\begin{aligned} \epsilon_{\text{min},n} &= \frac{1}{K} \mathrm{E}\left\{ \| \mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p} \|^2 \right\} = \frac{1}{K} \mathrm{tr}\{\mathbf{B}_n\} \\ &= R_{\mathbb{H}}[0,0] - \frac{1}{K} \mathrm{tr}\left\{ \boldsymbol{\mathcal{V}}_{\mathbb{H}} \left(\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \boldsymbol{\mathcal{X}}_n^{-1} \boldsymbol{\mathcal{R}}_Z \boldsymbol{\mathcal{X}}_n^{-H} \right)^{-1} \boldsymbol{\mathcal{V}}_{\mathbb{H}}^H \right\}. \end{aligned} \tag{3.9}$$

This depends on the data symbols, and thus is time-varying.

Implementation

Inserting (3.7) into (3.4) yields for the predicted channel vector

$$\hat{\mathbf{h}}_{\text{opt},n+p} = \boldsymbol{\mathcal{W}}_n \boldsymbol{\mathcal{X}}_n^{-1} \boldsymbol{\mathcal{Y}}_n .$$

This shows that the MMSE predictor implements two operations as depicted in Figure 3.3.

The first operation is a division of the $Y_{n,k}$ by the data symbols $X_{n,k}$:

$$\widetilde{\boldsymbol{\mathcal{H}}}_n = \boldsymbol{\mathcal{X}}_n^{-1} \boldsymbol{\mathcal{Y}}_n \qquad \text{or, equivalently,} \qquad \widetilde{H}_{n,k} = \frac{Y_{n,k}}{X_{n,k}}, \tag{3.10}$$

since the symbol matrix $\boldsymbol{\mathcal{X}}_n$ is diagonal. Recalling that $Y_{n,k} = H_{n,k} X_{n,k} + Z_{n,k}$, this division yields

$$\widetilde{H}_{n,k} = H_{n,k} + \widetilde{Z}_{n,k}, \qquad \text{with } \widetilde{Z}_{n,k} \triangleq \frac{Z_{n,k}}{X_{n,k}}. \tag{3.11}$$

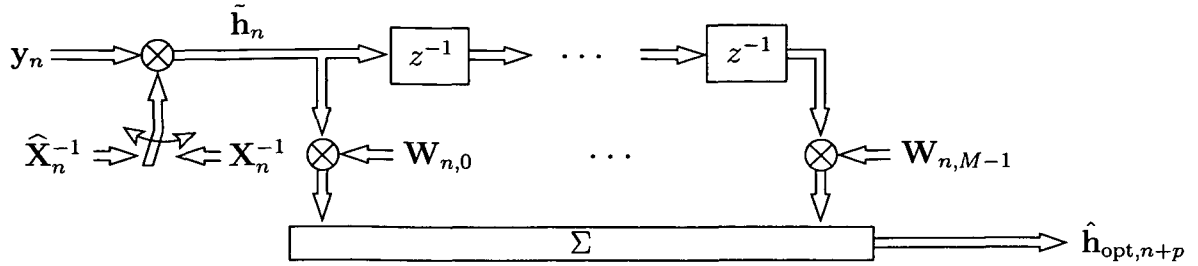


Figure 3.3: *The MMSE channel predictor consists of two stages: multiplication by the inverse data symbol matrix* $\mathbf{X}_n^{-1}$ *(or* $\widehat{\mathbf{X}}_n^{-1}$ *in decision-directed mode) and filtering by means of a time-varying MIMO system of length* M *with coefficients* $\mathbf{W}_{n,m}$.

Using $\tilde{\mathbf{h}}_n \triangleq [\widetilde{H}_{n,0}\, \widetilde{H}_{n,1} \cdots \widetilde{H}_{n,K-1}]^T$ and $\tilde{\mathbf{z}}_n \triangleq \left[\widetilde{Z}_{n,0}\, \widetilde{Z}_{n,1} \cdots \widetilde{Z}_{n,K-1}\right]^T$, this can be rewritten as

$$\tilde{\mathbf{h}}_n = \mathbf{X}_n^{-1}\mathbf{y}_n = \mathbf{h}_n + \tilde{\mathbf{z}}_n\,, \qquad \text{with} \qquad \tilde{\mathbf{z}}_n = \mathbf{X}_n^{-1}\mathbf{z}_n\,. \tag{3.12}$$

The second operation corresponds to a filtering by means of a time-varying MIMO system of length M, using (3.10) as input:

$$\hat{\mathbf{h}}_{\text{opt},n+p} = \boldsymbol{\mathcal{W}}_n\widetilde{\boldsymbol{\mathcal{H}}}_n = \boldsymbol{\mathcal{V}}_{\mathbb{H}}\left(\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \boldsymbol{\mathcal{X}}_n^{-1}\boldsymbol{\mathcal{R}}_Z\boldsymbol{\mathcal{X}}_n^{-H}\right)^{-1}\widetilde{\boldsymbol{\mathcal{H}}}_n\,. \tag{3.13}$$

This system is given by

$$\boldsymbol{\mathcal{W}}_n = [\mathbf{W}_{n,0}\ \ldots\ \mathbf{W}_{n,M-1}] = \boldsymbol{\mathcal{V}}_{\mathbb{H}}\left(\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \boldsymbol{\mathcal{X}}_n^{-1}\boldsymbol{\mathcal{R}}_Z\boldsymbol{\mathcal{X}}_n^{-H}\right)^{-1}\,. \tag{3.14}$$

In general it is time-varying since $\tilde{\mathbf{z}}_n = \mathbf{X}_n^{-1}\mathbf{z}_n$ is nonstationary, and it depends on the transmit symbols via $\boldsymbol{\mathcal{X}}_n$. The implementation of (3.13) would require on-line inversion of a $KM \times KM$ matrix in each symbol interval, which is not practical.

Comparison of the generic linear predictor in Figure 3.2 and the MMSE predictor in Figure 3.3 shows that we obtained an interesting two-stage structure. The first stage (the division stage) is often used in OFDM channel estimators, but it is typically introduced in an ad hoc manner (see e.g. [57, 65, 74, 90]). Our derivation, on the other hand, has shown that this stage is indeed part of the optimum (MMSE) channel estimator. The second stage (filtering stage) poses a problem for practical implementation because its computational complexity is excessive. Therefore, we will next develop a simplified MMSE predictor with significantly reduced complexity.

3.2.2 Reduced-Complexity Linear MMSE Predictor

The reduced-complexity MMSE channel predictor is constructed as follows. We retain the division stage (3.11), but to derive the subsequent processing we model the data $X_{n,k}$

as random. This will lead to a time-invariant MIMO filter for the second stage. More specifically, we assume that the $X_{n,k}$ are zero-mean and i.i.d. and that $1/X_{n,k}$ exists. Then, the noise term $\widetilde{Z}_{n,k} = Z_{n,k}/X_{n,k}$ in (3.11) has mean $\mathrm{E}\{\widetilde{Z}_{n,k}\} = \mathrm{E}\{Z_{n,k}\}\,\mathrm{E}\{1/X_{n,k}\} = 0$ and the correlation function is given by

$$\begin{aligned}\mathrm{E}\{\widetilde{Z}_{n,k}\widetilde{Z}^*_{n',k'}\} &= \mathrm{E}\{Z_{n,k}Z^*_{n',k'}\}\,\mathrm{E}\left\{\frac{1}{X_{n,k}X^*_{n',k'}}\right\} \\ &= R_Z[n-n',k,k']\,\mathrm{E}\left\{\frac{1}{|X_{n,k}|^2}\right\}\delta[n-n']\delta[k-k'] = \gamma^2\,\delta[n-n']\,\delta[k-k'],\end{aligned}$$

where $R_Z[n,k,k']$ is given by (2.37) and the "equivalent noise variance" is defined as

$$\gamma^2 \triangleq N_0\,\|f\|^2\,\mathrm{E}\left\{\frac{1}{|X_{n,k}|^2}\right\}. \tag{3.15}$$

Here, $\|f\|^2 = \sum_{m=-\infty}^{\infty}|f[m]|^2$. Hence, $\widetilde{Z}_{n,k}$ is zero-mean, stationary and white whereas $Z_{n,k}$ was nonstationary and correlated (cf. (2.37)). For PSK symbol alphabets, $\widetilde{Z}_{n,k}$ is Gaussian with variance $\gamma^2 = N_0\,\|f\|^2/\sigma_x^2$ with $\sigma_x^2 = |X_{n,k}|^2$. For non-PSK symbol alphabets, $\widetilde{Z}_{n,k}$ has a Gaussian mixture distribution. Thus, in general $\widetilde{Z}_{n,k}$ is non-Gaussian. It follows that the vector $\tilde{\mathbf{h}}_n = \mathbf{h}_n + \tilde{\mathbf{z}}_n$ is stationary and generally non-Gaussian. Due to the stationarity of $\tilde{\mathbf{h}}_n$, the linear MMSE predictor is given by a *time-invariant* MIMO predictor filter,

$$\hat{\mathbf{h}}_{n+p} = \sum_{m=0}^{M-1}\mathbf{W}_m\tilde{\mathbf{h}}_{n-m}\,. \tag{3.16}$$

In Subsection 3.2.3, we will show that this predictor filter can be efficiently implemented using the DFT.

Optimum Predictor Coefficients

The coefficient matrices $\mathbf{W}_m$ of the predictor filter are chosen such that the normalized MSE $\epsilon \triangleq \frac{1}{K}\mathrm{E}\{\|\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{n+p}\|^2\}$ is minimized. According to the orthogonality principle [148], the optimum $\mathbf{W}_m$ must be such that

$$\mathrm{E}\{(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{n+p})\,\tilde{\mathbf{h}}^H_{n-m}\} = \mathbf{0}\,, \qquad m = 0,\ldots,M-1\,, \tag{3.17}$$

i.e., the cross-correlations between the prediction error and the predictor input must vanish. Insertion of (3.12) and (3.16) into the orthogonality relation (3.17) yields the Wiener-Hopf equations

$$\sum_{m'=0}^{M-1}\mathbf{W}_{m'}\left(\mathbf{R}_{\mathbb{H}}[m-m'] + \gamma^2\delta[m-m']\,\mathbf{I}\right) = \mathbf{R}_{\mathbb{H}}[m+p]\,, \qquad m = 0,\ldots,M-1\,, \tag{3.18}$$

with the $K \times K$ Toeplitz correlation matrices $\mathbf{R}_{\mathbb{H}}[m]$ defined in (3.6). We will next stack the M equations in (3.18) into a single block matrix equation. To this end, we introduce the $K \times KM$ matrix $\boldsymbol{\mathcal{W}} \triangleq [\mathbf{W}_0 \cdots \mathbf{W}_{M-1}]$, the $K \times KM$ matrix $\boldsymbol{\mathcal{V}}_{\mathbb{H}} \triangleq [\mathbf{R}_{\mathbb{H}}[p] \cdots \mathbf{R}_{\mathbb{H}}[-M-1+p]]$, and the $KM \times KM$ block Toeplitz matrix $\boldsymbol{\mathcal{R}}_{\mathbb{H}}$ with first block row $[\mathbf{R}_{\mathbb{H}}[0] \cdots \mathbf{R}_{\mathbb{H}}[M-1]]$. Then, (3.18) can be compactly written as

$$\boldsymbol{\mathcal{W}} (\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \gamma^2 \mathbf{I}) = \boldsymbol{\mathcal{V}}_{\mathbb{H}} .$$

Thus, the MMSE-optimum predictor coefficient matrices $\mathbf{W}_{\text{opt},m}$ are given by

$$\boldsymbol{\mathcal{W}}_{\text{opt}} = \boldsymbol{\mathcal{V}}_{\mathbb{H}} (\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \gamma^2 \mathbf{I})^{-1}. \tag{3.19}$$

The computation of $\boldsymbol{\mathcal{W}}_{\text{opt}}$ in (3.19) is numerically stable (due to the term $\gamma^2\mathbf{I}$) and can be performed efficiently using the Wax-Kailath algorithm [150] because $\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \gamma^2\mathbf{I}$ is Hermitian Toeplitz/block-Toeplitz. Alternatively, the Levinson-Wiggins-Durbin algorithm can be used [151–153].

The main difference between the full-complexity predictor and the reduced-complexity predictor is that the filter $\boldsymbol{\mathcal{W}}_n$ in (3.14) is replaced by (3.19), and hence the term $\boldsymbol{\mathcal{X}}_n^{-1} \boldsymbol{\mathcal{R}}_Z \boldsymbol{\mathcal{X}}_n^{-H}$ is replaced by $\gamma^2\mathbf{I}$. Therefore, (3.19) is independent of the data symbols and the second (filtering) stage of the channel predictor becomes time-invariant. A special but important case where (3.14) and (3.19) coincide will be discussed presently.

Minimum Mean Square Error

Again using the orthogonality principle [148], the error covariance matrix of the reduced-complexity linear MMSE channel predictor is given by (cf. (3.8))

$$\begin{aligned} \mathbf{B} &= \mathrm{E}\{(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p})(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p})^H\} = \mathrm{E}\{(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p})\mathbf{h}_{n+p}^H\} \\ &= \mathbf{R}_{\mathbb{H}}[0] - \sum_{m=0}^{M-1} \mathbf{W}_{\text{opt},m} \mathbf{R}_{\mathbb{H}}[-m-p] = \mathbf{R}_{\mathbb{H}}[0] - \boldsymbol{\mathcal{V}}_{\mathbb{H}} (\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \gamma^2\mathbf{I})^{-1} \boldsymbol{\mathcal{V}}_{\mathbb{H}}^H , \end{aligned} \tag{3.20}$$

where $\hat{\mathbf{h}}_{\text{opt},n+p} = \sum_{m=0}^{M-1} \mathbf{W}_{\text{opt},m} \tilde{\mathbf{h}}_{n-m}$. Note that (3.20), in contrast to (3.8), does not depend on the transmit symbols $X_{n,k}$. The normalized MMSE achieved by $\mathbf{W}_{\text{opt},m}$ is (cf. (3.9))

$$\begin{aligned} \epsilon_{\min} &= \frac{1}{K} \mathrm{E}\left\{ \left\| \mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p} \right\|^2 \right\} = \frac{1}{K} \mathrm{tr}\{\mathbf{B}\} \\ &= R_{\mathbb{H}}[0,0] - \frac{1}{K} \mathrm{tr}\left\{ \boldsymbol{\mathcal{V}}_{\mathbb{H}} (\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \gamma^2\mathbf{I})^{-1} \boldsymbol{\mathcal{V}}_{\mathbb{H}}^H \right\} . \end{aligned} \tag{3.21}$$

This depends on the channel correlation matrix $\mathbf{R}_{\mathbb{H}}[m]$ and the equivalent noise variance γ^2. As will be verified experimentally in Subsection 3.5.2, $\epsilon_{\min}$ tends to decrease for channel coefficients $H_{n,k}$ that are more strongly correlated, i.e., for channels with a larger coherence time and/or a larger coherence bandwidth. Equivalently, $\epsilon_{\min}$ will be smaller for channels that are more underspread [1, 106, 138] in the sense of a smaller Doppler spread $\beta_{\mathbb{H}_d}$ (slower channel time-variations) and/or a smaller delay spread $\alpha_{\mathbb{H}_d}$. The prediction accuracy also improves for lower noise variance γ^2. The correlations of the $H_{n,k}$ can be fully exploited by the predictor (3.16) only if its memory exceeds the channel's coherence time, i.e., if $M > 1/(\beta_{\mathbb{H}_d} N)$. Hence, the choice of M is a trade-off between good prediction accuracy and low computational complexity. Note that the channel's coherence bandwidth is always completely covered because the predictor uses all subcarriers.

Optimality of the Reduced-Complexity MMSE Predictor

In general, the reduced-complexity MMSE predictor given by (3.19) is suboptimum because it does not fully exploit the knowledge of the data symbols $X_{n,k}$. However, if the data symbols $X_{n,k}$ are drawn from a PSK symbol alphabet and it the noise $Z_{n,k}$ is white, then the reduced-complexity predictor coincides with the MMSE predictor from Subsection 3.2.1 and hence is optimum. White noise in turn requires that the receive pulse is orthogonal (cf. (2.37)). Note that for a PSK symbol alphabet we obtain $\sigma_x^2 = \mathrm{E}\{|X_{n,k}|^2\} = |X_{n,k}|^2$ and furthermore $\mathrm{E}\{1/|X_{n,k}|^2\} = 1/\sigma_x^2$. For PSK symbol alphabet and white noise, the full-complexity MMSE predictor in (3.14) then simplifies to

$$\boldsymbol{\mathcal{W}}_n = \boldsymbol{\mathcal{V}}_{\mathbb{H}} \left(\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \frac{N_0}{\sigma_x^2} \mathbf{I} \right)^{-1}.$$

This does not depend on n and is identical to (3.19), with the equivalent noise variance given by $\gamma^2 = N_0/\sigma_x^2$ (cf. (3.15)).

We can conclude that when transmitting training data, the symbols should be chosen from a PSK alphabet.

3.2.3 Efficient DFT Implementation

The reduced-complexity MMSE predictor can be efficiently implemented as sketched in Figure 3.4. As will be seen presently, this implementation corresponds to a transformation of (3.16) into the eigenspace of the channel correlation matrices $\mathbf{R}_{\mathbb{H}}[m]$. Similar structures have been proposed for OFDM channel estimation and prediction e.g. in [56, 57, 65, 74, 147].

The first operation in Figure 3.4 is again the division operation $\tilde{\mathbf{h}}_n = \mathbf{X}_n^{-1}\mathbf{y}_n$, i.e., calculation of $\widetilde{H}_{n,k} = Y_{n,k}/X_{n,k}$ (cf. (3.7), (3.12)). Afterwards a K-point IDFT is applied

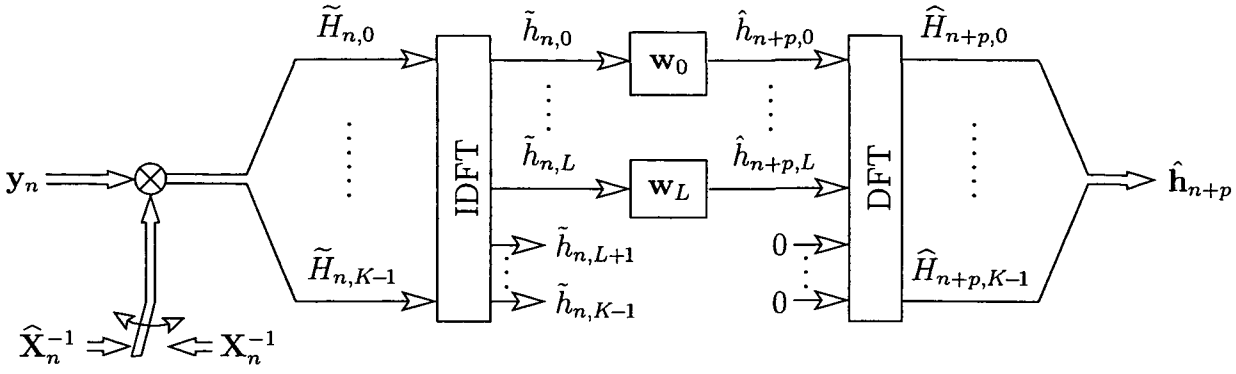


Figure 3.4: *Efficient DFT implementation of the reduced-complexity MMSE channel predictor. Note that in typical OFDM systems, the number of delay taps L is much smaller than the number of subcarriers K.*

to the $\widetilde{H}_{n,k}$, which yields (cf. (2.38))

$$\tilde{h}_{n,l} = \frac{1}{K}\sum_{k=0}^{K-1} \widetilde{H}_{n,k}\, e^{j2\pi lk/K} = \begin{cases} h_{n,l} + \tilde{z}_{n,l}\,, & l = 0,\ldots,L\,, \\ \tilde{z}_{n,l}\,, & l = L+1,\ldots,K-1\,. \end{cases} \tag{3.22}$$

Here, $h_{n,l}$ is the (subsampled) channel impulse response and $\tilde{z}_{n,l} = \frac{1}{K}\sum_{k=0}^{K-1} \widetilde{Z}_{n,k}\, e^{j2\pi lk/K}$ is white Gaussian noise with variance γ^2/K. Subsequently, we only need to process the first $L+1$ signals $\tilde{h}_{n,l}$ since they convey all information about $h_{n,l}$. A length-M MMSE predictor is used to predict the channel impulse response $h_{n+p,l}$ for each delay $l = 0,\ldots,L$. Due to the channel's DWSSUS property (2.45), $h_{n,l}$ and $h_{n,l'}$ are uncorrelated for $l \neq l'$. Therefore, the MMSE predictor decomposes into $L+1$ parallel single-input single-output (SISO) predictors given by

$$\hat{h}_{n+p,l} = \sum_{m=0}^{M-1} w_{m,l}^{*}\, \tilde{h}_{n-m,l} = \mathbf{w}_l^H \tilde{\mathbf{h}}_{n,l}\,, \qquad l = 0,\ldots,L\,, \tag{3.23}$$

with predictor coefficients $\mathbf{w}_l \triangleq [w_{0,l}\, w_{1,l} \cdots w_{M-1,l}]^T$ (specified below) and $\tilde{\mathbf{h}}_{n,l} \triangleq \left[\tilde{h}_{n,l}\, \tilde{h}_{n-1,l} \cdots \tilde{h}_{n-M+1,l}\right]^T$. Finally, the predicted channel coefficients $\widehat{H}_{n+p,k}$ are obtained from the predicted impulse response samples $\hat{h}_{n+p,l}$ via a K-point DFT,

$$\widehat{H}_{n+p,k} = \sum_{l=0}^{L} \hat{h}_{n+p,l}\, e^{-j2\pi kl/K}\,. \tag{3.24}$$

This implementation has significantly reduced complexity since $L+1$ SISO predictors are used instead of a MIMO predictor and because typically $L \ll K$. The computational complexity of the predictors is assessed in Subsection 3.2.6.

Optimum Predictor Coefficients

The Wiener-Hopf equations for the MMSE predictor coefficients $\mathbf{w}_l$ can be obtained by diagonalization of (3.18), corresponding to a transformation into the eigenspace of the correlation matrices $\mathbf{R}_{\mathbb{H}}[m]$. This diagonalizing transformation is a K-point DFT since the $K \times K$ matrices $\mathbf{R}_{\mathbb{H}}[m]$ are circulant. This can be seen by recalling that $\mathbf{R}_{\mathbb{H}}[m] = \mathrm{E}\left\{\mathbf{h}_n \mathbf{h}_{n-m}^H\right\}$ (see (3.6)). Furthermore, due to (2.38) the channel coefficients $H_{n,k}$ and the impulse response $h_{n,l}$ of the equivalent channel $\mathbb{H}$ are related by a DFT. In vector notation, we obtain the relation

$$\mathbf{h}_n = \begin{bmatrix} H_{n,0} \\ \vdots \\ H_{n,K-1} \end{bmatrix} = \mathbf{F} \begin{bmatrix} h_{n,0} \\ \vdots \\ h_{n,L} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{3.25}$$

where $\mathbf{F}$ is the orthogonal DFT matrix of dimension $K \times K$ with $\mathbf{F}\mathbf{F}^H = \mathbf{F}^H\mathbf{F} = K\mathbf{I}$. Hence, the correlation matrix is given by

$$\mathbf{R}_{\mathbb{H}}[m] = \mathrm{E}\left\{\mathbf{h}_n \mathbf{h}_{n-m}^H\right\} = \mathbf{F}\,\mathrm{diag}\left\{D_{\mathbb{H}}[m,0], \ldots, D_{\mathbb{H}}[m,L], 0, \ldots, 0\right\}\mathbf{F}^H, \tag{3.26}$$

where $D_{\mathbb{H}}[m,l]$ is the time-delay correlation function defined in (2.46). Note that the central matrix on the right-hand side of (3.26) is diagonal since $h_{n,l}$ and $h_{n,l'}$ are uncorrelated for $l \neq l'$. Hence, (3.26) is the eigenvalue decomposition of a circulant matrix with $L+1$ eigenvalues $KD_{\mathbb{H}}[m,l]$ (the factor K is due to $\mathbf{F}\mathbf{F}^H = K\mathbf{I}$) [146].

Diagonalizing the Wiener-Hopf equations in (3.18) via left multiplication by $\mathbf{F}$ and right multiplication by $\mathbf{F}^H$ yields equivalent Wiener-Hopf equations in the time-delay domain:

$$\left(\mathbf{D}_{\mathbb{H}}[l] + \frac{\gamma^2}{K}\mathbf{I}\right)\mathbf{w}_l = \mathbf{d}_{\mathbb{H}}[l], \qquad l = 0, 1, \ldots, L. \tag{3.27}$$

Here, the $M \times M$ Hermitian Toeplitz correlation matrices $\mathbf{D}_{\mathbb{H}}[l]$ and the cross-correlation vectors $\mathbf{d}_{\mathbb{H}}[l]$ are given by

$$\mathbf{D}_{\mathbb{H}}[l] \triangleq \mathrm{E}\left\{\mathbf{h}_{n,l}\mathbf{h}_{n,l}^H\right\} = \begin{bmatrix} D_{\mathbb{H}}[0,l] & \ldots & D_{\mathbb{H}}[M-1,l] \\ \vdots & & \vdots \\ D_{\mathbb{H}}[-M+1,l] & \ldots & D_{\mathbb{H}}[0,l] \end{bmatrix} \tag{3.28}$$

and

$$\mathbf{d}_{\mathbb{H}}[l] \triangleq \mathrm{E}\left\{\mathbf{h}_{n,l}h_{n+p,l}^*\right\} = \begin{bmatrix} D_{\mathbb{H}}[-p,l] \\ \vdots \\ D_{\mathbb{H}}[-p-M+1,l] \end{bmatrix}, \tag{3.29}$$

where $\mathbf{h}_{n,l} \triangleq \left[h_{n,l} \cdots h_{n-M+1,l}\right]^T$. Thus, the MMSE predictor coefficients are given by

$$\mathbf{w}_{\text{opt},l} = \left(\mathbf{D}_{\mathbb{H}}[l] + \frac{\gamma^2}{K}\mathbf{I}\right)^{-1} \mathbf{d}_{\mathbb{H}}[l], \qquad l = 0, 1, \ldots, L. \tag{3.30}$$

The computation of the $\mathbf{w}_{\text{opt},l}$ is numerically stable due to the term $\frac{\gamma^2}{K}\mathbf{I}$, and it can be done efficiently using the Levinson algorithm [152].

Moreover, from the derivation of (3.27) it follows that the coefficients of the reduced-complexity MMSE predictor in the time-frequency domain ($\mathbf{W}_m$ in (3.16)) and in the time-delay domain ($\mathbf{w}_l$ in (3.23)) can also be related by an eigenvalue decomposition or diagonalization:

$$\mathbf{W}_m = \mathbf{F}\mathbf{\Lambda}_m\mathbf{F}^H, \quad \text{where} \quad \mathbf{\Lambda}_m \triangleq \operatorname{diag}\{w_{m,0}, \ldots, w_{m,L}, 0, \ldots, 0\}.$$

This relation leads to the implementation shown in Figure 3.4. Note that the $w_{m,l}$ are the nonzero eigenvalues of the circulant matrix $\mathbf{W}_m$.

Minimum Mean-Square Error

Using the orthogonality principle [148] and the DFT relation (3.25), the error covariance matrix is given by (cf. (3.20))

$$\begin{aligned}\mathbf{B} &= \mathrm{E}\left\{\left(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p}\right)\left(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p}\right)^H\right\} \\ &= \mathbf{F}\operatorname{diag}\{\epsilon_{\min,0}, \epsilon_{\min,1}, \ldots, \epsilon_{\min,L}, 0, \ldots, 0\}\mathbf{F}^H, \end{aligned} \tag{3.31}$$

where the MMSE of the predictor for the lth delay tap is given by

$$\epsilon_{\min,l} \triangleq \mathrm{E}\left\{\left|h_{n+p,l} - \hat{h}_{\text{opt},n+p,l}\right|^2\right\} = D_{\mathbb{H}}[0,l] - \mathbf{w}_{\text{opt},l}^H \mathbf{d}_{\mathbb{H}}[l] \tag{3.32}$$

$$= D_{\mathbb{H}}[0,l] - \mathbf{d}_{\mathbb{H}}^H[l]\left(\mathbf{D}_{\mathbb{H}}[l] + \frac{\gamma^2}{K}\mathbf{I}\right)^{-1}\mathbf{d}_{\mathbb{H}}[l], \tag{3.33}$$

with $\hat{h}_{\text{opt},n+p,l} = \mathbf{w}_{\text{opt},l}^H \tilde{\mathbf{h}}_{n,l}$. Note that (3.31) is an eigenvalue decomposition of the error covariance matrix $\mathbf{B}$ into $L+1$ nonzero eigenvalues $K\epsilon_{\min,l}$ (the factor K is due to $\mathbf{F}\mathbf{F}^H = K\mathbf{I}$). Using (3.31) and (3.33), the normalized MMSE obtained with the optimum coefficients $\mathbf{w}_{\text{opt},l}$ is (cf. (3.21))

$$\begin{aligned}\epsilon_{\min} &= \frac{1}{K}\mathrm{E}\left\{\left\|\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{\text{opt},n+p}\right\|^2\right\} = \frac{1}{K}\operatorname{tr}\{\mathbf{B}\} \\ &= \frac{1}{K}\operatorname{tr}\left\{\mathbf{F}\operatorname{diag}\{\epsilon_{\min,0}, \epsilon_{\min,1}, \ldots, \epsilon_{\min,L}, 0, \ldots, 0\}\mathbf{F}^H\right\} = \sum_{l=0}^{L}\epsilon_{\min,l} \end{aligned} \tag{3.34}$$

$$= R_{\mathbb{H}}[0,0] - \sum_{l=0}^{L}\mathbf{d}_{\mathbb{H}}^H[l]\left(\mathbf{D}_{\mathbb{H}}[l] + \frac{\gamma^2}{K}\mathbf{I}\right)^{-1}\mathbf{d}_{\mathbb{H}}[l], \tag{3.35}$$

where we used that $\mathrm{tr}\{\mathbf{F}\mathbf{F}^H\} = K$ and $R_{\mathbb{H}}[0,0] = \sum_{l=0}^{L} D_{\mathbb{H}}[0,l]$ (cf. (2.47)). Note that $\epsilon_{\min}$ is equal to (3.21).

We note, however, that the DFT implementation does introduce a systematic error if some OFDM subcarriers (e.g. at the band edges) are not used for data transmission. This is frequently done in practical systems to reduce out-of-band emissions. An analysis of this error is provided in Appendix 3.A.

3.2.4 Infinite-Length MMSE Predictor

The best performance of the MMSE channel predictor is obtained if we exploit all existing channel correlations. This is always ensured for the frequency correlations of the channel. However, to exploit all time correlations, a prediction filter with infinite length is generally necessary. Whereas the infinite-length predictor and its MMSE are difficult to calculate for arbitrary prediction horizons p, an analytical result can be obtained for the case of one-step prediction ($p = 1$). This result is of interest for at least two reasons. First, we will use it for comparison with our finite-length predictors to determine the predictor length that is necessary for exploiting an essential part of the time correlations. Secondly, the infinite-length predictor will play an important role in Chapter 4 when we calculate the ergodic system capacity of OFDM systems.

For our calculations, we again use the predictor structure shown in Figure 3.4 since it is easier to analyze. Specifically, the predictor breaks up into the $L+1$ independent SISO predictors

$$\hat{h}_{n,l} = \sum_{m=1}^{\infty} w_{m,l}\, \tilde{h}_{m-n,l}\,, \qquad l = 0,1,\ldots,L\,. \tag{3.36}$$

Note that these input-output relations are equivalent to (3.23) for $p = 1$ and $M \to \infty$. (Note also that for notational convenience, we have formally replaced $w^*_{m,l}$ in (3.23) by $w_{m,l}$.)

Optimum Predictor Coefficients

In Subsection 3.2.3, we have seen that we can design the L SISO prediction filters (3.36) independently. Let us define the individual MSE of the lth predictor by $\epsilon_l \triangleq \mathrm{E}\{|h_{n,l}-\hat{h}_{n,l}|^2\}$. By the orthogonality principle [109, 148], the MSE ϵ_l is minimized by coefficients $w_{m,l}$ that satisfy the Wiener-Hopf equations (cf. (3.27))

$$D_{\mathbb{H}}[m,l] - \sum_{m'=1}^{\infty} w_{m',l}\left[D_{\mathbb{H}}[m-m',l] + \frac{\gamma^2}{K}\delta[m-m']\right] = 0\,, \qquad m \geq 0\,, \quad l = 0,1,\ldots,L\,.$$

We can equivalently formulate these equations as

$$D_{\mathbb{H}}[m,l] - \sum_{m'=1}^{\infty} w_{m',l}\left[D_{\mathbb{H}}[m-m',l] + \frac{\gamma^2}{K}\delta[m-m']\right] = g_{m,l}\,, \tag{3.37}$$

where the sequences $g_{m,l}$ have to be anticausal, i.e., $g_{m,l} = 0$ for $m \geq 0$, but are arbitrary otherwise. This follows because when considering the predictor coefficients $w_{m,l}$ for fixed l as sequences in m, the $w_{m,l}$ are strictly causal, i.e., $w_{m,l} = 0$ for $m \leq 0$. Next, we apply a Fourier transform with respect to the variable m to (3.37), thereby transforming the Wiener-Hopf equations from the time-delay domain to the scattering domain:

$$C_{\mathbb{H}}(l,\varphi) - W(l,\varphi)\left(C_{\mathbb{H}}(l,\varphi) + \frac{\gamma^2}{K}\right) = G(l,\varphi)\,. \tag{3.38}$$

Here, $C_{\mathbb{H}}(l,\varphi) = \sum_{m=-\infty}^{\infty} D_{\mathbb{H}}[m,l]e^{-j2\pi\varphi m}$ is the scattering function as given by (2.48), the predictor coefficients are transformed according to $W(l,\varphi) = \sum_{m=1}^{\infty} w_{m,l}\, e^{-j2\pi\varphi m}$, and $G(l,\varphi) \triangleq \sum_{-\infty}^{-1} g_{m,l}\, e^{-j2\pi\varphi m}$.

We now use following spectral decomposition [109, 154] of the scattering function and additive noise:

$$C_{\mathbb{H}}(l,\varphi) + \frac{\gamma^2}{K} = \rho_l \Phi(l,\varphi)\Phi^*(l,\varphi)\,, \tag{3.39}$$

with

$$\rho_l = \exp\left\{\int_{-1/2}^{1/2} \log\left(C_{\mathbb{H}}(l,\varphi) + \frac{\gamma^2}{K}\right) d\varphi\right\}, \tag{3.40}$$

where the sequence $\phi_{m,l} = \int_{-1/2}^{1/2} \Phi(l,\varphi)e^{j2\pi m\varphi}d\varphi$ is causal, minimum phase, and monic. Note that the sequence $\tilde{\phi}_{n,l} = \int_{-1/2}^{1/2} \frac{1}{\Phi(l,\varphi)}e^{j2\pi n\varphi}d\varphi$ exists and is again causal, minimum phase, and monic. The spectral decomposition (3.39) is guaranteed to exist since $\rho_l > -\infty$ due to the noise component γ^2/K, i.e., the Paley-Wiener condition is satisfied [109, 155]. Substituting (3.39) into (3.38) yields

$$\rho_l\left[\Phi(l,\varphi) - \Phi(l,\varphi)\, W(l,\varphi)\right] = \frac{G(l,\varphi)}{\Phi^*(l,\varphi)} + \frac{\gamma^2/K}{\Phi^*(l,\varphi)}\,. \tag{3.41}$$

Recall that the $w_{m,l}$ are strictly causal. Therefore, to solve (3.41) for the optimum $W(l,\varphi)$, we have to identify the part of (3.41) that yields a strictly causal sequence when inverse Fourier transformed with respect to the variable φ. Let us first consider the terms on the right-hand side of (3.41). Here, the time sequence corresponding to $\frac{G(l,\varphi)}{\Phi^*(l,\varphi)}$ is anti-causal because it is the convolution of two anticausal sequences. Hence, this term does not

contribute to the solution of the optimum $W(l,\varphi)$. Moreover, the sequence corresponding to $\frac{\gamma^2/K}{\Phi^*(l,\varphi)}$ is also anticausal. Hence, to obtain the optimum $W(l,\varphi)$, we have to equate the right-hand side of (3.41) to zero. We next have to identify those parts of the left-hand side of (3.41) that correspond to a strictly causal sequence. Here, the first term is $\Phi(l,\varphi)$; its strictly causal part corresponds to $\Phi(l,\varphi) - 1$ because the sequence $\phi_{m,l} - \delta[m]$ is strictly causal. Furthermore, the term $\Phi(l,\varphi)\,W(l,\varphi)$ corresponds to the convolution of a causal sequence and a strictly causal sequence and thus is strictly causal. Thus, the corresponding strictly causal part of the equation (3.41) is given by

$$\rho_l \left[\Phi(l,\varphi) - 1 - \Phi(l,\varphi)\,W(l,\varphi)\right] = 0\,.$$

Hence, the optimum infinite-length channel predictor coefficients are obtained from the inverse Fourier transform of

$$W_{\text{opt}}(l,\varphi) = 1 - \frac{1}{\Phi(l,\varphi)}\,. \tag{3.42}$$

Minimum Mean-Square Error

From (3.32), it follows that the MMSE of the lth infinite-length predictor filter is given by

$$\epsilon_{\min,l} = D_{\mathbb{H}}[0,l] - \sum_{m=1}^{\infty} w_{\text{opt},m,l}\, D_{\mathbb{H}}[-m,l] = \int_{-1/2}^{1/2} \left[C_{\mathbb{H}}(l,\varphi) - W_{\text{opt}}(l,\varphi) C_{\mathbb{H}}(l,\varphi)\right] d\varphi\,.$$

Inserting (3.39) and (3.42) yields

$$\epsilon_{\min,l} = \rho_l \int_{-1/2}^{1/2} \Phi^*(l,\varphi)\,d\varphi - \frac{\gamma^2}{K} \int_{-1/2}^{1/2} \frac{1}{\Phi(l,\varphi)}\,d\varphi = \rho_l - \frac{\gamma^2}{K}\,. \tag{3.43}$$

The final expression follows because both integrals evaluate to one since both $\Phi(l,\varphi)$ and $\frac{1}{\Phi(l,\varphi)}$ are monic. Inserting (3.40) into (3.43) yields further

$$\epsilon_{\min,l} = \frac{\gamma^2}{K} \left[\exp\left\{\int_{-1/2}^{1/2} \log\left(1 + \frac{K}{\gamma^2} C_{\mathbb{H}}(l,\varphi)\right) d\varphi\right\} - 1\right]. \tag{3.44}$$

For large bandwidth, we can furthermore insert (2.49), which yields

$$\epsilon_{\min,l} \approx \frac{\gamma^2}{K} \left[\exp\left\{NT_s \int_{\nu} \log\left(1 + \frac{K}{N\gamma^2} C_{\mathbb{H}_c}(lT_s,\nu)\right) d\nu\right\} - 1\right].$$

From (3.34) it follows that the total MMSE is given by summation of the $L+1$ individual MMSEs:

$$\epsilon_{\min} = \sum_{l=0}^{L} \epsilon_{\min,l} \approx \frac{\gamma^2}{K} \sum_{l=0}^{L} \left[\exp\left\{NT_s \int_{\nu} \log\left(1 + \frac{K}{N\gamma^2} C_{\mathbb{H}_c}(lT_s,\nu)\right) d\nu\right\} - 1\right]. \tag{3.45}$$

Hence, the minimum prediction error depends on the channel statistics via the scattering function $C_{\mathbb{H}_c}(\tau,\nu)$, on the equivalent noise variance γ^2, and on the OFDM system parameters N, K, and T_s. Note that the transmit power effectively enters inversely into γ^2 (cf. (3.15)). The behavior of $\epsilon_{\min}$ for different scattering functions $C_{\mathbb{H}_c}(\tau,\nu)$ and noise variances γ^2 can be directly assessed using (3.45).

3.2.5 Channel Prediction in Specular Scattering

We now consider OFDM channel prediction for a special type of "specular scattering" channels whose scattering function consists of discrete (specular) components:

$$C_{\mathbb{H}}(l,\varphi) = \sum_{i=0}^{I-1} \rho_i\, \delta[l-l_i]\, \delta(\varphi-\varphi_i)\,, \qquad \varphi \in [-1/2, 1/2)\,. \tag{3.46}$$

This model of a wireless fading channel is widely used in the literature for its simplicity and analytic tractability.

The parametric representation of $C_{\mathbb{H}}(l,\varphi)$ in (3.46) comprises I parameter triplets $\{\rho_i, l_i, \varphi_i\}$, i.e., each of the I scatterers is characterized by its gain $\sqrt{\rho_i}$, delay l_i, and Doppler shift φ_i. It is important to note that knowledge of the channel statistics (scattering function) entails exact knowledge of the delays and Dopplers of the individual scatterers. We can expect that this detailed knowledge about the structure of the time and frequency selective fading channel (which implies a reduced randomness of the channel) can be exploited for channel prediction. Indeed, our analysis will show that the minimum prediction error is equal to zero in this special case. This result extends a similar result for noiseless prediction of nonregular processes [109]. For our calculations, we again use the DFT-based implementation shown in Figure 3.4 since it allows a simplified analysis. Moreover, because $\epsilon_{\min} = \sum_{l=0}^{L} \epsilon_{\min,l}$ (cf. (3.35)), the global MMSE $\epsilon_{\min}$ vanishes if each of the $L+1$ individual MMSEs $\epsilon_{\min,l}$ is zero. Therefore, we can consider one (arbitrary but fixed) delay tap and drop the delay index l for notational simplicity.

Problem Statement

Let us consider the length-M linear predictor given by (cf. (3.23))

$$\hat{h}_n = \sum_{m=1}^{M} w_m^*\, \tilde{h}_{n-m}\,, \tag{3.47}$$

for the stationary process (cf. (3.22))

$$\tilde{h}_n = h_n + \tilde{z}_n\,. \tag{3.48}$$

Without loss of generality, we assume one-step prediction ($p = 1$); note that the prediction MSE will vanish for any $p > 1$ if it vanishes for $p = 1$ because several one-step predictors could be used one after the other to obtain any desired prediction horizon. For the power spectral density of h_n, we assume specular scattering defined by (cf. (3.46))

$$C_{\mathbb{H}}(\varphi) \triangleq \sum_{i=0}^{I-1} \rho_i \, \delta(\varphi - \varphi_i) \,, \qquad \varphi_i \in [-1/2, 1/2) \,, \quad \rho_i > 0 \,. \tag{3.49}$$

Furthermore, we require that $\varphi_i \neq \varphi_{i'}$ for $i \neq i'$ and assume that $\tilde{z}_n$ is zero-mean circularly complex white Gaussian noise with variance γ^2/K.

Optimum Predictor Coefficients

According to the orthogonality principle [109, 148], the MSE-optimum predictor coefficients are such that

$$\mathrm{E}\left\{(h_n - \hat{h}_n)\tilde{h}^*_{n-m}\right\} = 0 \,, \quad m = 1, \ldots, M \,.$$

Inserting (3.47) and (3.48), the Wiener-Hopf equations are obtained as (cf. (3.27))

$$\left(\mathbf{D}_{\mathbb{H}} + \frac{\gamma^2}{K}\mathbf{I}\right) \mathbf{w} = \mathbf{d}_{\mathbb{H}} \,, \tag{3.50}$$

where $\mathbf{w} = [w_1 \ \ldots \ w_M]^T$, $\mathbf{D}_{\mathbb{H}}$ is an $M \times M$ Toeplitz correlation matrix with first row $[D_{\mathbb{H}}[0], \ldots, D_{\mathbb{H}}[M-1]]$, and $\mathbf{d}_{\mathbb{H}} \triangleq [D_{\mathbb{H}}[-1], \ldots, D_{\mathbb{H}}[-M]]^T$ with $D_{\mathbb{H}}[m] \triangleq \mathrm{E}\{h_n h^*_{n-m}\}$.

The correlation function corresponding to the spectral density $C_{\mathbb{H}}(\varphi)$ in (3.49) is obtained as

$$D_{\mathbb{H}}[m] = \int_{-1/2}^{1/2} C_{\mathbb{H}}(\varphi) \, e^{j2\pi m\varphi} \, d\varphi = \sum_{i=0}^{I-1} \rho_i \, e^{j2\pi m\varphi_i} \,.$$

Therefore, the correlation matrix and cross-correlation vector in (3.50) are, respectively, given by

$$\mathbf{D}_{\mathbb{H}} = \sum_{i=0}^{I-1} \rho_i \, \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^H \qquad \text{and} \qquad \mathbf{d}_{\mathbb{H}} = \sum_{i=0}^{I-1} \rho_i \, e^{-j2\pi\varphi_i} \, \boldsymbol{\varphi}_i \,, \tag{3.51}$$

with the $M \times 1$ vectors

$$\boldsymbol{\varphi}_i \triangleq [1 \ e^{-j2\pi\varphi_i} \ \ldots \ e^{-j2\pi(M-1)\varphi_i}]^T \,.$$

The correlation matrix $\mathbf{D}_{\mathbb{H}}$ has rank I. For growing M (note that M can be chosen arbitrarily large), $\mathbf{D}_{\mathbb{H}}$ and $\mathbf{d}_{\mathbb{H}}$ can be arbitrarily well approximated as

$$\mathbf{D}_{\mathbb{H}} \approx \sum_{l \in \mathcal{L}} \tilde{\rho}_l \, \tilde{\boldsymbol{\varphi}}_l \tilde{\boldsymbol{\varphi}}_l^H \qquad \text{and} \qquad \mathbf{d}_{\mathbb{H}} \approx \sum_{l \in \mathcal{L}} \tilde{\rho}_l \, e^{-j2\pi \frac{l}{M}} \, \tilde{\boldsymbol{\varphi}}_l \,, \tag{3.52}$$

with the orthogonal DFT vectors

$$\tilde{\boldsymbol{\varphi}}_l \triangleq [1\, e^{-j2\pi l/M} \ \ldots\ e^{-j2\pi l(M-1)/M}]^T$$

and the index set $\mathcal{L}$ consisting of the I integers l corresponding to the frequencies $\frac{l}{M}$ that are closest to the actual Dopplers $\varphi_i \in [-1/2, 1/2)$. Note that $\tilde{\boldsymbol{\varphi}}_l^H \tilde{\boldsymbol{\varphi}}_k = M\delta[l-k]$. The cardinality of the set $\mathcal{L}$ is I and thus is identical to the number of scatterers of the parametric model. The $\tilde{\rho}_l$ in (3.52) are reordered but otherwise equal to the ρ_i.

Inserting the parametric model (3.52), the Wiener-Hopf equations (3.50) become

$$\left(\tilde{\rho}_l + \frac{\gamma^2}{KM}\right) \tilde{\boldsymbol{\varphi}}_l^H \mathbf{w}\, e^{j2\pi l/M} \approx \tilde{\rho}_l\,, \qquad l \in \mathcal{L}\,,$$

or equivalently

$$\left(\tilde{\rho}_l + \frac{\gamma^2}{KM}\right) W_l \approx \tilde{\rho}_l\,, \qquad l \in \mathcal{L}\,,$$

where

$$W_l \triangleq \tilde{\boldsymbol{\varphi}}_l^H \mathbf{w}\, e^{j2\pi l/M} = \sum_{m=1}^{M} w_m\, e^{j2\pi lm/M}$$

is the M-bin IDFT of the predictor coefficients. Hence, in the transformed domain the MMSE predictor coefficients are given by

$$W_{\mathrm{opt},l} \approx \frac{\tilde{\rho}_l}{\tilde{\rho}_l + \frac{\gamma^2}{KM}}\,, \qquad l \in \mathcal{L}\,.$$

The optimum predictor coefficients are finally obtained via a DFT,

$$w_{\mathrm{opt},m} = \frac{1}{M} \sum_{l \in \mathcal{L}} W_{\mathrm{opt},l}\, e^{-j2\pi ml/M}\,.$$

Minimum Mean-Square Error

Using (3.32), the prediction MMSE is given by (dropping the delay index l) $\epsilon_{\min}^{(M)} \triangleq D_{\mathbb{H}}[0] - \mathbf{d}_{\mathbb{H}}^H \mathbf{w}_{\mathrm{opt}}$. With our parametric scattering model,

$$D_{\mathbb{H}}[0] = \sum_{l \in \mathcal{L}} \tilde{\rho}_l = \sum_{i=0}^{I-1} \rho_i \qquad \text{and} \qquad \mathbf{d}_{\mathbb{H}}^H \mathbf{w}_{\mathrm{opt}} = \sum_{l \in \mathcal{L}} \frac{\tilde{\rho}_l^2}{\tilde{\rho}_l + \frac{\gamma^2}{KM}} = \sum_{i=0}^{I-1} \frac{\rho_i^2}{\rho_i + \frac{\gamma^2}{KM}}\,.$$

Therefore, the MMSE is finally obtained as

$$\epsilon_{\min}^{(M)} = \sum_{i=0}^{I-1} \left(\rho_i - \frac{\rho_i^2}{\rho_i + \frac{\gamma^2}{KM}}\right) = \frac{\gamma^2}{KM} \sum_{i=0}^{I-1} \frac{\rho_i}{\rho_i + \frac{\gamma^2}{KM}}\,.$$

Predictor type	Design	Prediction	Update
MMSE predictor	—	$\mathcal{O}(K^3M^3)$	—
Reduced-complexity MMSE predictor	$\mathcal{O}(K^2M^2)$	$\mathcal{O}(K^2M)$	—
DFT implementation	$\mathcal{O}(M^2L)$	$\mathcal{O}(K\log_2 K + ML)$	—
Adaptive predictor using NLMS algorithm	—	$\mathcal{O}(K\log_2 K + ML)$	$\mathcal{O}(ML)$
Adaptive predictor using RLS algorithm	—	$\mathcal{O}(K\log_2 K + ML)$	$\mathcal{O}(M^2L)$

Table 3.1: *Approximate computational complexity of the finite-length MMSE and adaptive channel predictors.*

This shows, in particular, that the MMSE vanishes for infinite predictor length, i.e.,

$$\epsilon_{\min}^{(\infty)} = \lim_{M\to\infty} \epsilon_{\min}^{(M)} = 0\,. \tag{3.53}$$

We can thus conclude that for unconstrained predictor filter length, the prediction error can be made arbitrarily small. A similar result exists for noiseless prediction of singular processes [109]. In our case, we assume perfect knowledge of the Doppler locations of the specular scattering model and, therefore, the only "randomness" in h_n enters through the I gains $\sqrt{\rho_i}$. If we are able to estimate the ρ_i without error, we can predict h_n arbitrarily well. But perfect estimation of ρ_i is indeed possible since h_n is a stationary process and infinite time is available to average out the noise. Hence, the asymptotic prediction error is zero.

3.2.6 Computational Complexity of MMSE Predictors

Thus far, we have discussed the "full-complexity" MMSE predictor (Subsection 3.2.1), the reduced-complexity linear MMSE predictor (Subsection 3.2.2), and the DFT implementation of the reduced-complexity linear MMSE predictor (Subsection 3.2.3). The computational complexity of these finite-length MMSE predictors is compared in Table 3.1. We specify the complexity both for the design (performed in advance only once) and for the actual channel prediction (performed in each symbol interval). For the full-blown MMSE predictor, the design has to be performed anew for each symbol and thus it is considered as part of the prediction complexity. It is assumed that the design is based on the Wax-Kailath algorithm [150] in the case of the reduced-complexity MMSE predictor and on the Levinson algorithm [152] in the case of the DFT implementation.

It is seen that for typical values of the parameters K, M, and L, the DFT implementation is significantly less complex than the original implementation of the reduced-complexity

MMSE predictor, which in turn is significantly less complex than the full-complexity MMSE predictor.

3.3 Adaptive OFDM Channel Predictors

In practice, calculation of the MMSE channel predictor in (3.30) presupposes that the channel correlation $\mathbf{R}_{\mathbb{H}}[m]$ and the noise variance γ^2 have been estimated from the received signal. Furthermore, the statistics of real-world channels are constant only over a certain time [133, 135, 136], thus necessitating reestimation of $\mathbf{R}_{\mathbb{H}}[m]$ and recalculation of the channel predictor once in a while. To avoid these problems, we propose *adaptive* channel predictors that perform a continual update of the predictor coefficients, do not require knowledge of the channel and noise statistics, and are capable of tracking nonstationary statistics.

The adaptive predictors will be based on the same structure as the DFT implementation of Figure 3.4 because it is computationally efficient and involves only $M(L+1)$ scalar coefficients, instead of the MK^2 scalar coefficients required by the original implementation of the reduced-complexity MMSE predictor. Note that adaptive algorithms generally perform better when fewer coefficients have to be adapted [156]. Assuming that the adaptation starts at $n = 0$, the predicted channel taps are (cf. (3.23))

$$\hat{h}_{n+p,l} = \mathbf{w}_l^H[n]\,\tilde{\mathbf{h}}_{n,l}\,, \qquad n \geq 0\,,\ \ l = 0,\ldots,L\,.$$

The time-varying predictor filters $\mathbf{w}_l[n]$ will be adaptively adjusted by means of the normalized least-mean-square (NLMS) algorithm or the recursive least-squares (RLS) algorithm [156].

3.3.1 NLMS Algorithm

The NLMS algorithm belongs to the family of stochastic gradient algorithms that iteratively estimate the MMSE predictor filters. We use the NLMS algorithm rather than the LMS algorithm because the selection of the adaptation constant is simpler. The predictor filters $\mathbf{w}_l[n]$ are updated according to [156]

$$\mathbf{w}_l[n] = \mathbf{w}_l[n-1] + \frac{\mu}{\|\tilde{\mathbf{h}}_{n-p,l}\|^2}\, e_{n,l}^*\, \tilde{\mathbf{h}}_{n-p,l}\,, \qquad n \geq p\,,$$

where μ is the adaptation constant, $\|\tilde{\mathbf{h}}_{n-p,l}\|^2 = \sum_{i=0}^{M-1} |\tilde{h}_{n-p-i,l}|^2$, and $e_{n,l}$ is the prediction error that would ideally be given by $h_{n,l} - \mathbf{w}_l^H[n-1]\,\tilde{\mathbf{h}}_{n-p,l}$. However, since we perform

noisy prediction, $h_{n,l}$ is unavailable. Thus we approximate $h_{n,l}$ by $y_{n,l} = h_{n,l} + \tilde{z}_{n,l}$ in (3.22), i.e.,

$$e_{n,l} \triangleq \tilde{h}_{n,l} - \mathbf{w}_l^H[n-1]\,\tilde{\mathbf{h}}_{n-p,l}\,, \qquad n \geq p\,. \tag{3.54}$$

The error introduced by this approximation will be small for most practical signal-to-noise ratios. For $n = 0, \ldots, p-1$, the prediction filters are initialized as

$$\mathbf{w}_l[n] = [1\ 0\ \ldots\ 0]^T, \qquad n = 0, \ldots, p-1\,. \tag{3.55}$$

Thus, $\hat{h}_{n+p,l} = \tilde{h}_{n,l}$ for $n = 0, \ldots, p-1$.

Stable operation of the NMLS algorithm requires $0 < \mu < 2$ [156]. The selection of μ is a trade-off between fast convergence and small excess MSE. We obtained good results with $\mu \approx 0.5$.

3.3.2 RLS Algorithm

With the RLS algorithm, the lth predictor filter $\mathbf{w}_l[n]$ is calculated such that it minimizes the error [156]

$$\epsilon_{\mathrm{RLS},l}[n] \triangleq \sum_{m=p}^{n} \lambda^{n-m} \left| h_{m,l} - \mathbf{w}_l^H[n]\,\tilde{\mathbf{h}}_{m-p,l} \right|^2.$$

Here, λ with $0 < \lambda \leq 1$ is a forgetting factor that accounts for possible nonstationarity of the input $\tilde{\mathbf{h}}_{n,l}$ (we obtained good results for $\lambda = 0.99$). The resulting update equation for $\mathbf{w}_l[n]$ is [156]

$$\mathbf{w}_l[n] = \mathbf{w}_l[n-1] + \mathbf{k}_{n-p,l}\, e_{n,l}^*\,, \qquad n \geq p\,,$$

where $e_{n,l}$ is as in (3.54) and $\mathbf{k}_{n,l}$ is the RLS gain vector given by

$$\mathbf{k}_{n,l} = \frac{\mathbf{P}_{n-1,l}\,\tilde{\mathbf{h}}_{n,l}}{\lambda + \tilde{\mathbf{h}}_{n,l}^H\,\mathbf{P}_{n-1,l}\,\tilde{\mathbf{h}}_{n,l}}\,, \qquad n \geq 1\,.$$

Here, the matrix $\mathbf{P}_{n,l}$ is the inverse of the $M \times M$ sample covariance matrix $\sum_{m=0}^{n} \lambda^{n-m}\,\tilde{\mathbf{h}}_{m,l}\,\tilde{\mathbf{h}}_{m,l}^H$; it can be calculated recursively as

$$\mathbf{P}_{n,l} = \frac{1}{\lambda}\left(\mathbf{I} - \mathbf{k}_{n,l}\,\tilde{\mathbf{h}}_{n,l}^H\right)\mathbf{P}_{n-1,l}\,, \qquad n \geq 1\,.$$

The RLS recursion is initialized as in (3.55). Furthermore, we set

$$\mathbf{P}_{0,l} = \left(\tilde{\mathbf{h}}_{0,l}\,\tilde{\mathbf{h}}_{0,l}^H + \delta\mathbf{I}\right)^{-1} = \frac{1}{\delta}\left[\mathbf{I} - \frac{\tilde{\mathbf{h}}_{0,l}\,\tilde{\mathbf{h}}_{0,l}^H}{\left\|\tilde{\mathbf{h}}_{0,l}\right\|^2 + \delta}\right]$$

and

$$\mathbf{k}_{0,l} = \mathbf{P}_{0,l}\,\tilde{\mathbf{h}}_{0,l} = \frac{1}{\left\|\tilde{\mathbf{h}}_{0,l}\right\|^2 + \delta}\,\tilde{\mathbf{h}}_{0,l}\,,$$

where the stabilization factor δ is in the range $0 < \delta \ll 1$ (we chose $\delta = 0.1$).

Compared to the NLMS algorithm, the RLS algorithm converges faster and has a smaller excess MSE. Its convergence rate is independent of the eigenvalue spread of the input process [156]. In our case, this is an important advantage over the NLMS algorithm since our channels are highly correlated (large coherence time).

3.3.3 Computational Complexity of Adaptive Predictors

The last two rows of Table 3.1 show the computational complexity of the channel prediction (filter operation) and coefficient update for the two adaptive predictors. Both operations have to be performed in each symbol interval. While the complexity of the channel prediction is equal for the NLMS and RLS predictors (it is the same as the complexity of the nonadaptive DFT implementation), the coefficient update is significantly more costly for the RLS predictor than for the NLMS predictor, especially for large filter lengths M.

3.4 Applications of OFDM Channel Prediction

Channel prediction is potentially useful for a variety of communication techniques. Out of these we consider predictive equalization in Subsection 3.4.1 and adaptive modulation in Subsection 3.4.2. Furthermore, in Subsection 3.4.3 we treat the practically relevant situation of an OFDM system that periodically transmits pilot symbols. Here, the decision-directed operation of the predictor can be augmented by the pilot symbols.

3.4.1 Predictive Equalization

As a first application example, we consider delay-free predictive equalization. The OFDM transmitter is sketched in Figure 3.5(a). We assume that a coder maps a bit stream D_i to symbol vectors $\mathbf{x}_n = [X_{n,0}\, X_{n,1} \cdots X_{n,K-1}]^T$ and an OFDM modulator generates the transmit signal $x[m]$. In the receiver shown in Figure 3.5(b), the channel predictor provides up-to-date channel state information without requiring regular transmission of pilot symbols. Similar receivers have been proposed e.g. in [65, 90, 94]. The vector signal

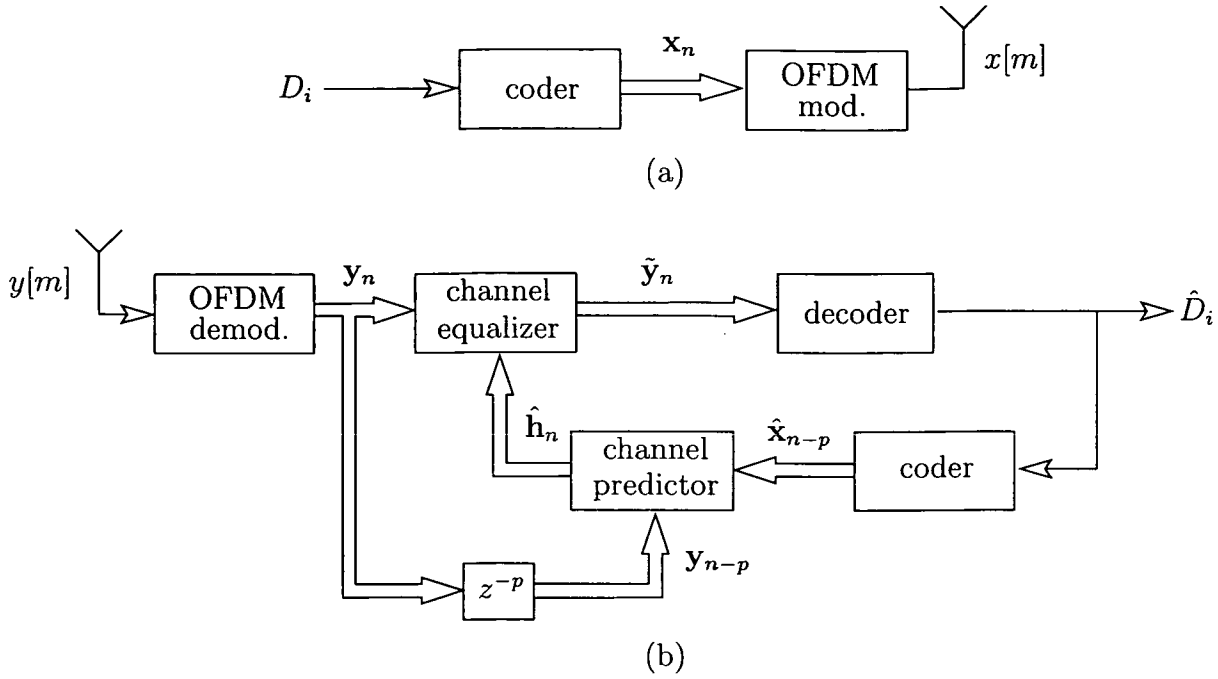


Figure 3.5: *Coded OFDM system using predictive channel equalization: (a) transmitter and (b) receiver.*

$\mathbf{y}_n = [Y_{n,0}\, Y_{n,1} \cdots Y_{n,K-1}]^T$ is obtained via the OFDM demodulator. A zero-forcing equalizer then calculates $\widetilde{Y}_{n,k} \triangleq Y_{n,k}/\widehat{H}_{n,k}$ (cf. (2.36)), where the estimated channel coefficients $\hat{\mathbf{h}}_n = \left[\widehat{H}_{n,0}\, \widehat{H}_{n,1} \cdots \widehat{H}_{n,K-1}\right]^T$ are provided by the channel predictor. Any of the finite length-predictors considered so far may be used depending on the required prediction accuracy and the available *a priori* knowledge. However, due to its efficient implementation, the DFT-based predictor structure is most attractive from a practical perspective. The equalized sequence $\tilde{\mathbf{y}}_n = \left[\widetilde{Y}_{n,0}\, \widetilde{Y}_{n,1} \cdots \widetilde{Y}_{n,K-1}\right]^T$ is decoded to obtain the detected bits $\hat{D}_i$.

The channel estimates $\widehat{H}_{n,k}$ are generated by the lower receiver branch in Figure 3.5(b). The channel predictor is operated in decision-directed mode, i.e., the true transmit symbols $\mathbf{x}_n = [X_{n,0}\, X_{n,0} \cdots X_{n,K-1}]^T$ required in (3.10) are replaced by symbols $\hat{\mathbf{x}}_n = \left[\widehat{X}_{n,0}\, \widehat{X}_{n,1} \cdots \widehat{X}_{n,K-1}\right]^T$ that are obtained by re-encoding the detected bits $\hat{D}_i$. Note that $\hat{\mathbf{x}}_n = \mathbf{x}_n$ only if all bits $\hat{D}_i$ were correct; otherwise, error propagation may result. Together, decoding and re-encoding introduce a delay of $p \geq 1$ OFDM symbols; this delay is compensated by channel prediction. We have $p = 1$ if coding is performed within only one OFDM symbol over the subcarriers or when no coding is used. Except for initialization of the predictor, no training data need to be transmitted.

For initialization of a channel predictor of length M, the symbol vectors $\mathbf{x}_n$ should ideally be known for M symbol periods. If less training data $\mathbf{x}_n$ are available, we set the corresponding unknown $\tilde{\mathbf{h}}_n$ to zero. With HIPERLAN/2 and IEEE 802.11a, two known

pilot OFDM symbols are transmitted at the beginning of each frame. Our simulations in Subsection 3.5.1 show that this suffices for initialization; even a single known OFDM symbol is sufficient. Predictive equalization thus enables the tracking of channel variations in systems like HIPERLAN/2 and IEEE 802.11a, which is difficult otherwise because the pilot symbols are not dense enough to interpolate the channel between them.

3.4.2 Adaptive Modulation

By adaptive modulation, we mean the adaptation of transmission parameters to the current channel realization. Such transmission parameters may be the symbol alphabet, power allocation, channel code, coding rate, etc. For adaptive modulation, channel state information (CSI) is required both at the transmitter and the receiver. Moreover, both the transmitter and the receiver have to be aware of the currently used set of transmission parameters. Implementations of adaptive modulation typically perform estimation of the channel at the receiver, calculation of the relevant transmission parameters according to some optimality criterion, and feed-back of these parameters to the transmitter [8, 49, 157]. The optimality criteria frequently encountered in the literature can be ordered into two groups. The first criterion aims at maximizing the data rate under the side constraint of achieving a certain bit error probability with a given transmit power and leads to techniques that are related to waterfilling [8, 158]. The other criterion tries to minimize the bit error probability under the constraints of a given transmit power and a fixed (or minimum) data rate [8, 159].

The idea of adaptive modulation is the key to obtaining high data rates in wired communications such as digital subscriber line (xDSL) transmission systems [158]. Here, the implementation of adaptive modulation is straightforward because the channel is practically time-invariant over the complete transmission period. Therefore, the channel can be estimated and the relevant information can be fed back to the transmitter before actually starting the transmission of data.

The success of adaptive modulation in wired communication systems has motivated research on the application of these techniques to wireless systems [47–50, 157, 160–165]. However, in wireless systems the channel generally changes during the transmission, and thus the implementation of methods proven useful in wired systems is not straightforward. Because the channel is time-varying, a packet-oriented strategy is necessary using packet lengths significantly shorter than the coherence time of the channel [49]. The adaptive modulation methods known from wired systems can then be used in wireless systems on a packet-by-packet basis. However, the overhead increases rapidly with the fading rate of the channel and possibly absorbs all advantages [50, 164, 165]. Moreover, things can be made

even worse by choosing unfavorable transmission parameters due to outdated CSI. This is where channel prediction comes into play as a possible remedy for the implementation problems plaguing adaptive modulation in wireless communication systems [50, 165]. By means of channel prediction, we are able to increase the maximum packet length by choosing transmission parameters that match the future channel state. Moreover, with prediction we can obtain up-to-date CSI even for fast time-varying channels or large prediction horizons. Because of these advantages, channel prediction could be the key to successful application of adaptive modulation in wireless systems.

System Description

Figure 3.6 shows a point-to-point, OFDM-based wireless communication system that implements an adaptive modulation strategy. We will propose a novel protocol for signaling the transmission parameters from the receiver back to the transmitter, and suggest the use of channel prediction for obtaining CSI at the receiver.

The transmitter sends either training data or real user data. In training data mode, the transmission parameters are fixed and preselected. For example, the available transmit power is evenly distributed over all subcarriers and a BPSK symbol alphabet is used. Training data is sent in blocks of length M'.

The receiver is based on the system previously described in Subsection 3.4.1 in the context of predictive equalization. However, the channel predictor is now used twice. First, it is used for predictive equalization during data transmission periods, where the prediction horizon p is chosen to compensate for decoding delays. Secondly, it is used to generate CSI for the calculation of a transmission parameter set that is optimum according to a suitable criterion. For the latter purpose, the predictor is operated with various prediction horizons p' to generate CSI for the complete subsequent data packet (see below).

Obtaining CSI

We suggest to use the efficient DFT implementation of the MMSE channel predictor as described in Subsection 3.2.3. The estimation of the required channel statistics and noise variance is discussed in [64].

Inserting the MSE-optimum predictor coefficients (3.30) into (3.23) yields

$$\hat{h}_{n+p',l} = \mathbf{d}_{\mathbb{H}}^{H}[l]\check{\mathbf{h}}_{n,l}\,, \qquad l = 0,\ldots,L\,, \quad p' \geq 1\,, \tag{3.56}$$

with

$$\check{\mathbf{h}}_{n,l} \triangleq \left(\mathbf{D}_{\mathbb{H}}[l] + \frac{\gamma^2}{K}\mathbf{I}\right)^{-1}\tilde{\mathbf{h}}_{n,l}\,, \tag{3.57}$$

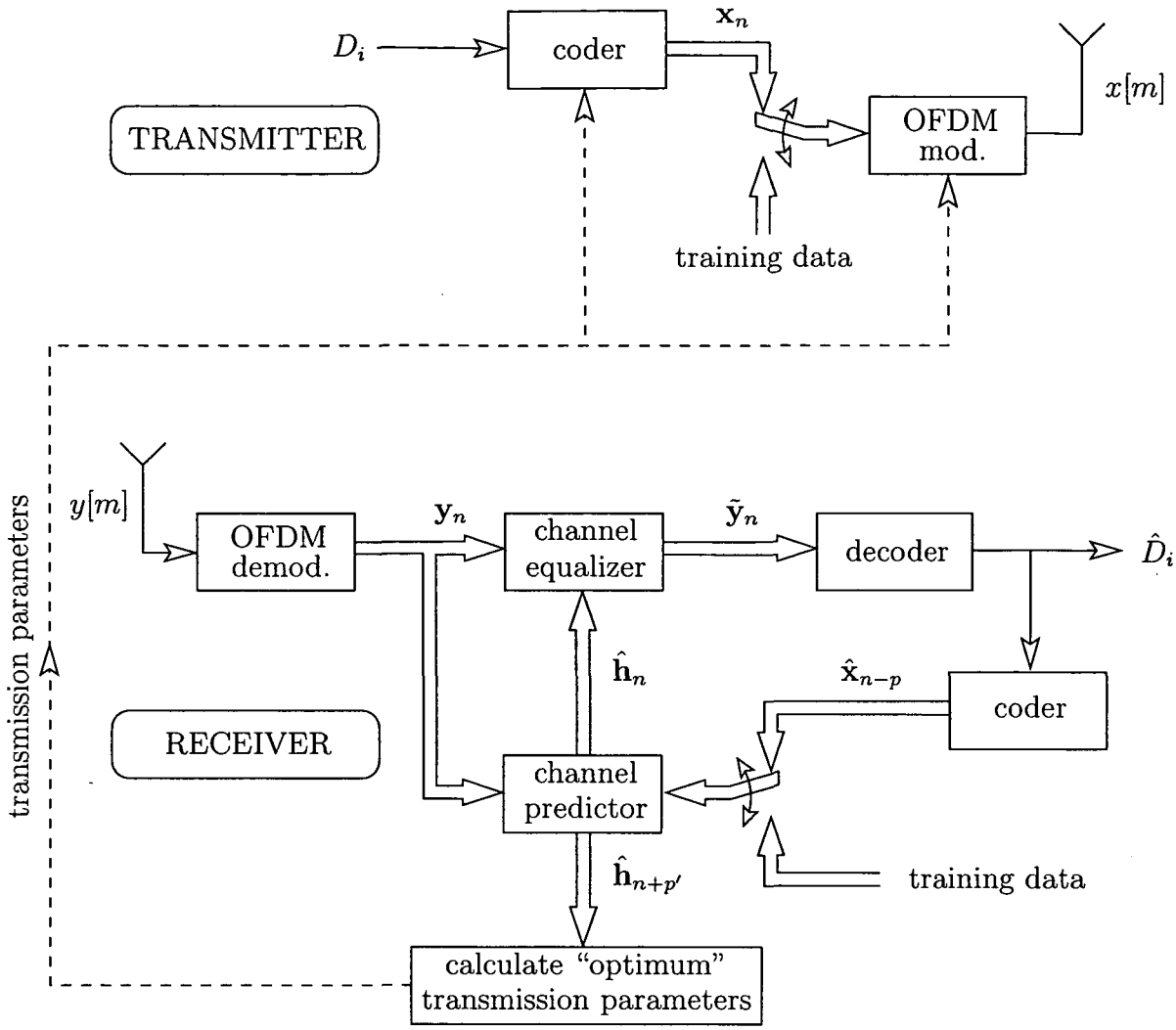


Figure 3.6: *Wireless OFDM communication system for implementation of adaptive modulation strategies. The transmitter is shown in the top part and the receiver using channel prediction in the lower part. The dashed line indicates the feed-back channel used to signal transmission parameters from the receiver to the transmitter.*

where $\mathbf{D}_{\mathbb{H}}[l]$ and $\mathbf{d}_{\mathbb{H}}[l]$ are given by (3.28) and (3.29), respectively. Note that while $\mathbf{d}_{\mathbb{H}}[l]$ in (3.56) depends on the prediction horizon p', (3.57) is independent of p' and therefore has to be calculated only once even if we wish to predict the channel for various p'.

Let us assume that the user data packet has length Q and starts with an offset O. To obtain CSI for the complete interval of the user data packet, we have to predict the channel for the Q prediction horizons $p' = O, O+1, \ldots, O+Q-1$ by calculating (3.57) once and subsequently evaluating (3.56) Q times for this range of p'.

Protocol

Finally, we propose a strategy for obtaining CSI and feeding the transmission parameters back to the transmitter. Consider the packet structure shown in Figure 3.7. The transmission starts with training data being sent from the transmitter to the receiver. The duration of the training data is $M' \geq M$, where M is the length of the channel predictor used in the receiver. Choosing M' larger than M may be required for reliable estimation of the channel statistics and the noise variance (see e.g. [64]). After the training data block, there follows a short information field by means of which the transmitter informs the receiver about the intended length of the data packet and possibly about constraints that need to be taken into account when selecting the future transmission parameters. This information field is sent using preselected transmission parameters (power level, symbol alphabet, channel code). The transmitter then remains silent during a fixed silence period before it starts transmitting the user data packet.

During the silent period, the receiver estimates the channel statistics, the receive power, and the noise variance using e.g. the algorithm in [64]. Based on the received signal during the initial training period, it then predicts the channel for the period of the information field. After decoding the information field, the receiver is informed about the intended length of the data packet and further transmitter-related parameters. Next, the receiver predicts the channel for the period of the data packet or up to a certain percentage of the channel coherence period (e.g., the prediction horizon can be limited to 10% of the coherence time $T_{\mathbb{H}_c}$). If p' is too large compared to $T_{\mathbb{H}_c}$, the prediction accuracy will be poor. If the user data packet is larger than this maximum prediction horizon, the receiver tells the transmitter to split the data packet into smaller snippets. Note that the receiver can calculate $T_{\mathbb{H}_c}$ after it has estimated the channel statistics.

Next, based on the predicted channel, the receiver calculates possible transmission parameters for the user data packet. For this calculation, it tiles the time-frequency rectangle of the data packet into smaller rectangles that correspond to regions of identical transmission parameters (e.g., identical power allocation and identical symbol alphabet). We propose to use a tiling as sketched in Figure 3.7, so that the information about the start point of one rectangle using a pointer and its size $A \times B$ suffices for the receiver to subsequently generate the same tiling. The parameters A and B have to be much smaller than the channel's coherence time and coherence bandwidth, respectively; they should be matched to regions of approximately constant channel gains. The information about the tiling of the data packet (starting pointer and parameters A and B) and the transmission parameters within each rectangle (power allocation, symbol alphabet, etc.) are signaled from the

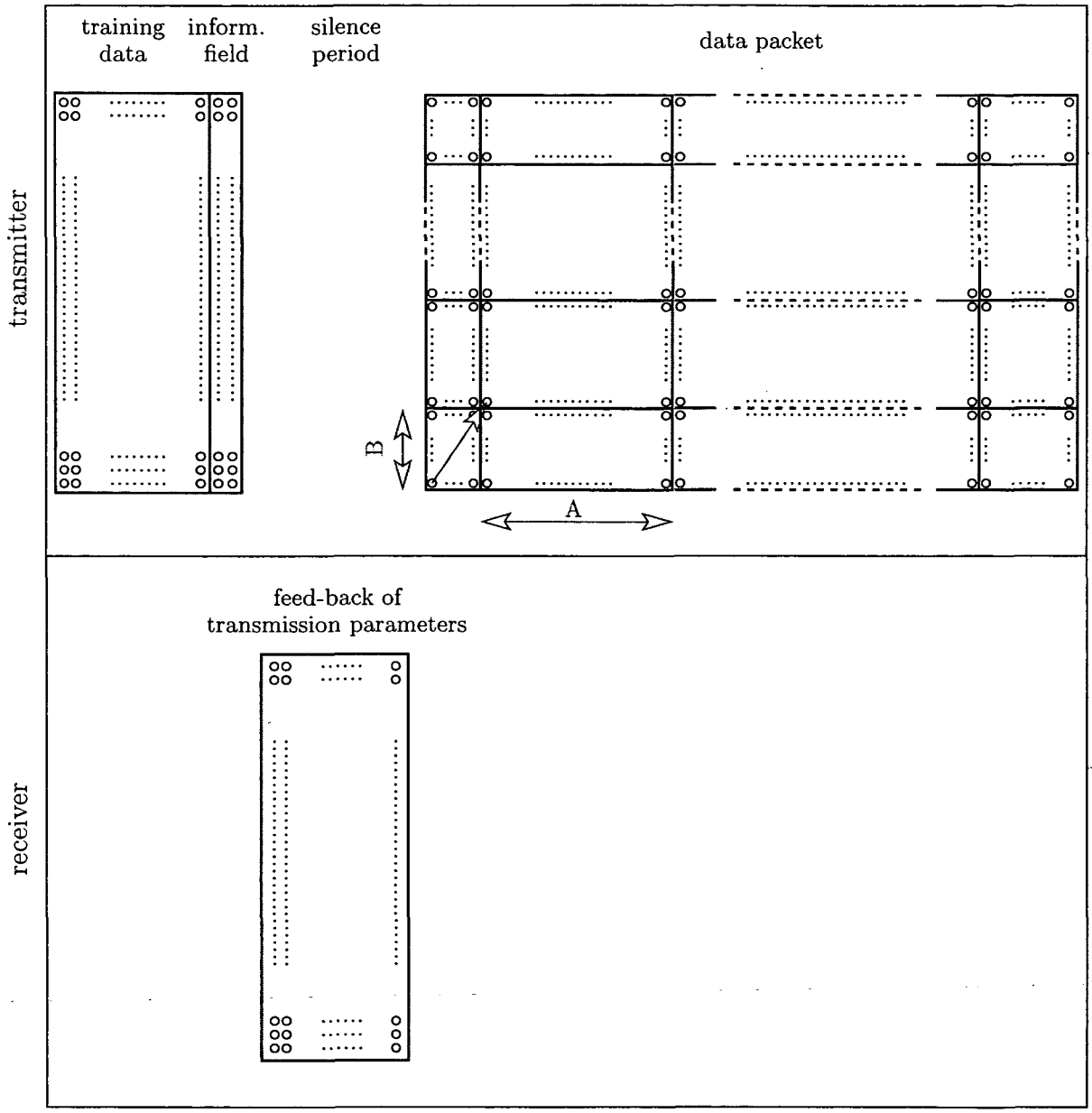


Figure 3.7: *Schematic illustration of the packet structure used in the proposed adaptive modulation scheme. Vertical and horizontal directions correspond to subcarriers and OFDM symbols, respectively.*

receiver to the transmitter during the transmitter's silence period. This transmission also uses a preselected format (power level, symbol alphabet, channel code). Additionally, the receiver informs the transmitter about the allowable size of the data packet. The feed-back channel carrying the transmitter parameter signaling has to be highly secure.

After receiving the transmission parameters, the transmitter adjusts its modulation and

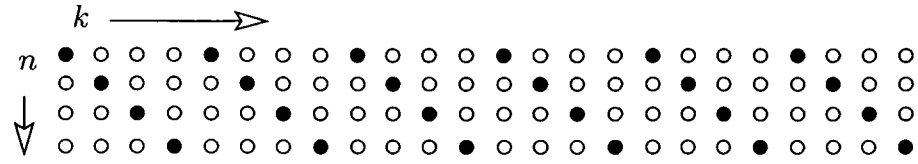


Figure 3.8: *Pilot symbols (•) are regularly distributed over the time-frequency plane and multiplexed with data symbols (○). This example shows scattered pilot locations for a system with $K = 24$ subcarriers and $P = 6$ pilots per OFDM symbol. The "subcarrier distance" between pilots is $S = 4$.*

possibly coding accordingly and starts to transmit the user data packet. This data packet is decoded at the receiver using predictive equalization. The accuracy of the already predicted channel can be increased by again running the channel predictor during the transmission of the data packet, this time in decision-directed mode.

The overhead of this protocol is determined by the sum of the lengths of the initial training data field, of the information field, and of the silence period compared to the size of the user data packet. For example, to obtain 10% overhead one could use 1000 OFDM symbols per packet, of which 100 are used for the overhead. For IEEE 802.11a, we would then obtain a total packet length of $1000 \cdot 4\mu\text{s} = 4\text{ms}$. Assuming indoor communications and a coherence time of 100 ms, we would have to predict up to $4\text{ms}/100\text{ms} = 4\%$ of the coherence time, for which channel prediction is very accurate (see Section 3.5).

3.4.3 Pilot Symbol Augmented Channel Prediction

Some wireless communication systems based on OFDM continually multiplex training data into the transmitted data stream. An example is DVB-T that transmits approximately 10% training data in the form of so-called scattered pilot symbols [166]. Here, pilot symbol based channel prediction can be used. Alternatively, one can use a channel predictor with decision-directed operation that is *augmented* by using the known pilot symbols instead of the corresponding symbol estimates $\widehat{X}_{n,k}$. These channel prediction schemes have the advantage of yielding improved CSI while enabling delay-free equalization and decoding.

Pilot Symbols

The pilots are regularly spread over the time-frequency plane as shown in Figure 3.8. The set of pilot symbol locations can be defined as [73]

$$\mathcal{P} \triangleq \left\{(n,k) \,\middle|\, n \in \mathbb{Z},\ k = iS + (n \bmod S),\ i \in [0, P-1]\right\},$$

where P is the number of pilots per OFDM symbol and $S = K/P$ is the "subcarrier distance" between two neighboring pilots in the frequency direction. We assume that P divides K so that S is an integer.

For reasons of simplicity, the pilots are assumed to be a BPSK-modulated sequence, i.e.,

$$\Omega_{n,k} = \begin{cases} 1 \text{ or } -1, & (n,k) \in \mathcal{P}, \\ 0, & \text{else}. \end{cases} \tag{3.58}$$

In [166], the pilot sequence is derived from a pseudo-noise sequence and transmitted with a "boosted" power level using amplification factor 4/3.

Pilot-based channel estimation or channel prediction basically means that we subsample the time-dependent transfer function of $\mathbb{H}$ and then apply some filtering and interpolation. The subsampling as well as the filtering/interpolation are two-dimensional, i.e., in both the time and frequency directions. Therefore, the pilots need to be dense enough to avoid aliasing. The four pilots per OFDM symbol used in the IEEE 802.11a standard are not dense enough for channel estimation; they are solely meant for frequency synchronization.

Pilot Symbol Based Channel Prediction

In the case where channel prediction is based exclusively on the pilot symbols, only the received sequences $Y_{n,k}$ for $(n,k) \in \mathcal{P}$ are used for channel prediction. (In contrast, we previously based channel prediction on all $Y_{n,k}$ for $n \in \mathbb{Z}$ and $k \in [0, K-1]$, see e.g. Figure 3.1.) The efficient DFT implementation discussed in Subsection 3.2.3 can be used for pilot symbol based channel prediction with small changes. Let us define the $K \times K$ diagonal pilot symbol matrix

$$\mathbf{\Omega}_n = \operatorname{diag}\{\Omega_{n,0}, \Omega_{n,1}, \cdots, \Omega_{n,K-1}\}.$$

We simply replace the multiplication by $\mathbf{X}_n^{-1}$ in Figure 3.4 with multiplication by $\mathbf{\Omega}_n$ and obtain the pilot symbol based channel predictor that is shown in Figure 3.9. The predictor coefficients $\mathbf{w}_l$ can be chosen according to the MMSE solution (see Subsection 3.2.3) or according to the adaptive algorithms of Section 3.3.

Pilot symbol based channel prediction has advantages as well as drawbacks. Our initial motivation for channel prediction was to reduce or altogether avoid the regular transmission of pilots. However, the simulations in Subsection 3.5.5 will reveal that a system purely based on decision-directed operation and predictive equalization exhibits an SNR threshold that has to be exceeded for satisfactory performance. The reason for this SNR threshold is error propagation. This problem is clearly avoided by pilot symbol based channel prediction. Furthermore, since the pilots are a BPSK-modulated sequence, the reduced-complexity

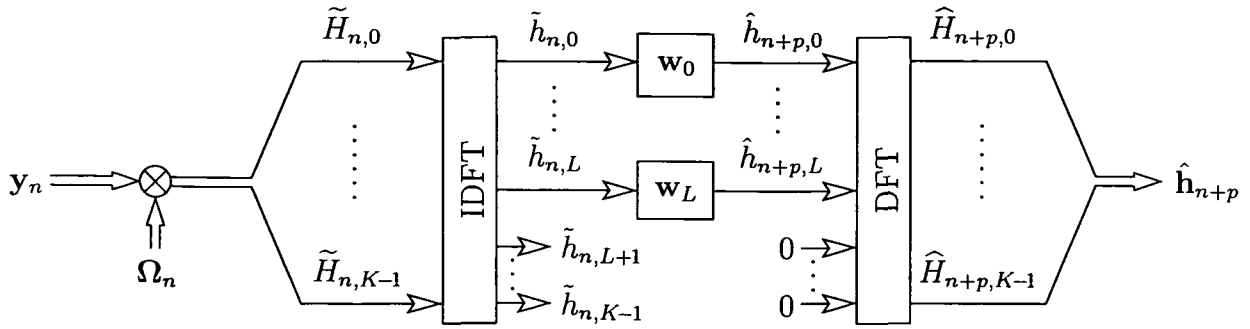


Figure 3.9: *Channel predictor for pilot symbol based channel prediction.*

linear predictor coincides with the full-complexity MMSE predictor (see Subsection 3.2.2 on the optimality of the reduced-complexity MMSE predictor for PSK symbols). On the other hand, for purely pilot-based prediction, the "reference" reduces from K observations per OFDM symbol to P observations per OFDM symbol, which leads to a larger MMSE.

We therefore suggest to employ channel prediction in an OFDM system where pilot symbols are available in the following two modes: below the SNR threshold, purely pilot symbol based channel prediction should be used; whereas above the SNR threshold, a combined pilot-based and decision-directed operation (see below) should be used.

Combined Decision-Directed and Pilot-Based Channel Prediction

In a nutshell, this combined operation mode uses pilot symbols where available and decision-directed operation at all other subcarriers. We thus replace (3.58) by

$$\widetilde{\Omega}_{n,k} = \begin{cases} 1 \text{ or } -1\,, & (n,k) \in \mathcal{P}\,, \\ Y_{n,k}/\widehat{X}_{n,k}\,, & \text{else}\,, \end{cases}$$

and define the diagonal matrix $\widetilde{\mathbf{\Omega}}_n = \operatorname{diag}\{\widetilde{\Omega}_{n,0}, \widetilde{\Omega}_{n,1}, \cdots, \widetilde{\Omega}_{n,K-1}\}$ accordingly. Replacing $\mathbf{\Omega}_n$ in Figure 3.9 by $\widetilde{\mathbf{\Omega}}_n$ leads to the combined operation mode. This mode is especially useful when the system operates above the SNR threshold (see Subsection 3.5.5); otherwise channel prediction accuracy may suffer from error propagation.

3.5 Simulation Results

We simulated a CP-OFDM system with $K = 120$ subcarriers and cyclic prefix length $L_{\text{cp}} = 20$. The OFDM symbol length was $N = K + L_{\text{cp}} = 140$. We used a coded

system with 16-QAM symbol alphabet and $\sigma_x^2 = \mathrm{E}\{|X_{n,k}|^2\} = 1$. The coding rate was approximately 1/2. A $(15, 7)$ Reed-Solomon code over $\mathrm{GF}(16)$ was used, with each code symbol consisting of one 16-QAM symbol. The code symbols were interleaved in frequency. No temporal coding was used. The coded system can exploit the frequency diversity offered by the multipath channel; however, time diversity is not exploited. In Subsections 3.5.5 and 3.5.6, we additionally simulated an uncoded system with 4-QAM symbol alphabet and $\sigma_x^2 = |X_{n,k}|^2 = 1$. Note that the coded and uncoded systems have approximately the same net data rate; however, the uncoded system cannot exploit diversity.

The channel was simulated by means of the technique described in [167]. We prescribed a scattering function with exponentially decaying delay profile and Jakes Doppler profile [132], i.e.,

$$C_{\mathbb{H}_\mathrm{d}}(l,\xi) = \begin{cases} \dfrac{\exp(-l/\lambda_0)}{\sqrt{\xi_{\max}^2 - \xi^2}}, & \text{for } l = 0,\ldots,L,\ |\xi| < \xi_{\max}, \\ 0, & \text{elsewhere.} \end{cases} \tag{3.59}$$

Unless stated otherwise, we chose $\lambda_0 = L/\log_e(2L)$ and $L = 19$. We considered both a "slow" (slowly time-varying) channel with $\xi_{\max}K = 0.001$ and a "fast" (rapidly time-varying) channel with $\xi_{\max}K = 0.01$. The quantity $\xi_{\max}K$ can be interpreted as the amount of channel variation within one OFDM symbol. The additive noise was zero-mean white and Gaussian with variance σ_u^2.

For the channel predictor, the DFT implementation of Figure 3.4 was used. The predictor consisted of $L+1 = 20$ SISO prediction filters, each of length $M = 10$. The prediction horizon was $p = 1$ unless indicated otherwise. The parameters of the NLMS and RLS algorithms were chosen as $\mu = 0.5$ and $\lambda = 0.99$, respectively. The SNR, defined as $(\sigma_{\mathbb{H}_\mathrm{d}}^2\sigma_x^2)/\sigma_u^2$, was 25 dB unless indicated otherwise.

3.5.1 Convergence of the Adaptive Predictors

Figure 3.10 shows the convergence of the adaptive channel predictors in the coded system for the slow and fast channels. The (normalized) prediction MSE was estimated from 100 realizations. For comparison, the estimated prediction MSE of the reduced-complexity MMSE predictor from Section 3.2.2 and the theoretical MMSE according to (3.21) are also plotted. We assume packet transmission with a packet length of 1000 OFDM symbols and a packet preamble of only a single known OFDM symbol. At the beginning of each packet, the channel predictors were initialized in training mode with the single known OFDM symbol to predict the channel for the next OFDM symbol. Afterwards, the predictors

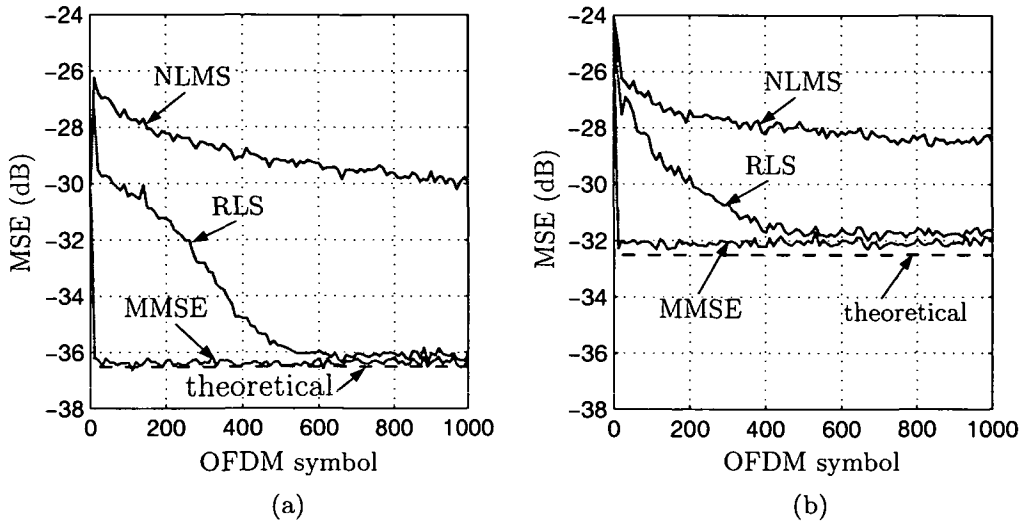


Figure 3.10: *Convergence behavior of the adaptive predictors in decision-directed mode for (a) the slow channel, and (b) the fast channel.*

were operated in decision-directed mode. With the adaptive channel predictors, coefficient adaptation started after $M = 10$ OFDM symbols had been received. With the MMSE channel predictor, the appropriate predictor length (between 1 and M) was used during the initialization phase, depending on the number of OFDM symbols received.

It is seen that the MSE achieved is below $1/\mathrm{SNR} = -25\,\mathrm{dB}$ for all predictors. The RLS-based predictor converges faster than the NLMS-based predictor and has almost no excess MSE relative to the reduced-complexity MMSE channel predictor. The excess MSE of the NLMS-based predictor is 4–6 dB. The prediction MSE obtained after convergence strongly depends on the channel's maximum Doppler frequency; it is about 2–4 dB higher for the fast channel than for the slow channel. We note that in this simulation, the predictors always converged. Thus, a single known OFDM symbol here suffices for startup.

3.5.2 Dependence of Prediction MSE on Maximum Delay and Doppler

Figure 3.11 shows how the normalized prediction MSE (after convergence in the case of adaptive predictors) depends on the channel's maximum delay L and maximum Doppler frequency $\xi_{\max}$ (cf. (3.59)). This simulation was carried out using the coded system. Note that a larger L ($\xi_{\max}$) implies a smaller coherence bandwidth (coherence time). Throughout this simulation, the predictors were operated in training mode to avoid error propagation effects. The prediction MSE was estimated by averaging over 100 realizations with 10^4 symbols each.

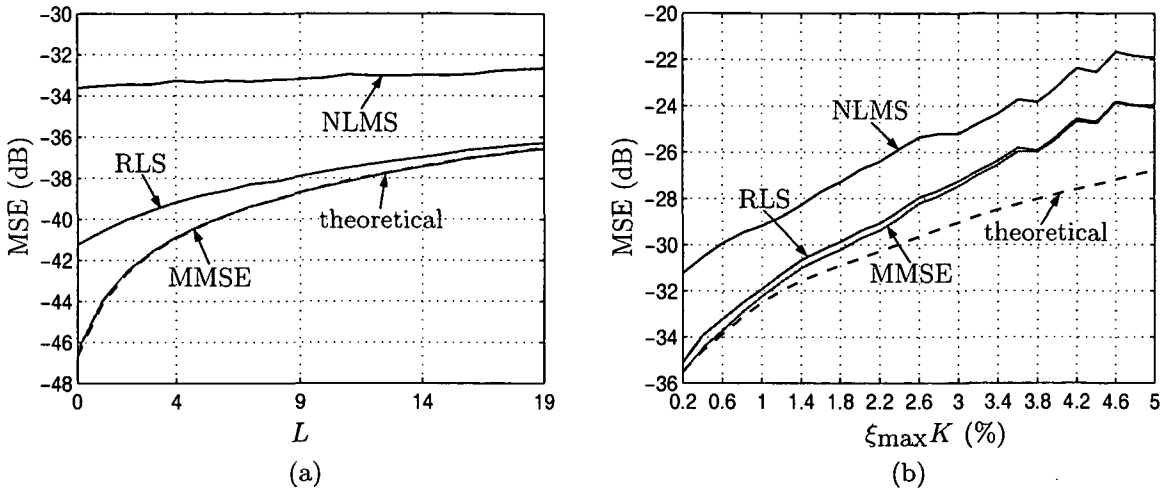


Figure 3.11: *Dependence of the prediction MSE on the channel's maximum delay L and maximum Doppler frequency $\xi_{\max}$: (a) varying L at fixed $\xi_{\max}K = 0.001$, (b) varying $\xi_{\max}$ at fixed $L = 19$.*

In Figure 3.11(a), L is varied between 0 and 19 while $\xi_{\max}K$ is fixed at 0.001 (slow channel). In Figure 3.11(b), $\xi_{\max}K$ is varied between 0.002 and 0.05 at a fixed L of 19. It can be seen that the prediction MSE increases with increasing L and increasing $\xi_{\max}$. This is an expected behavior since prediction is more difficult for weaker correlations, i.e., for a smaller coherence bandwidth and/or coherence time. We also see that the RLS-based predictor tends to perform nearly as well as the MMSE predictor. According to Figure 3.11(b), for high $\xi_{\max}$ all predictors deviate from the theoretical MMSE. This is because the approximate system relation (2.36) becomes less accurate with increasing $\xi_{\max}$ [108], due to increasing ICI.

3.5.3 Dependence of the Prediction MSE on the Prediction Horizon

Figure 3.12 shows the dependence of the normalized prediction MSE on the prediction horizon p for the coded system. The predictors were again operated in training mode. The prediction MSE was estimated by averaging over 10 realizations with 10^4 symbols each. For the slow channel considered in Figure 3.12(a), channel prediction is accurate even for large horizons. The excess MSE of the RLS-based predictor is negligible; that of the NLMS-based predictor is between 1 dB and 4 dB. For the fast channel in Figure 3.12(b), prediction over large horizons performs much less well. Indeed, for large p the prediction MSE of all methods is significantly higher than for the slow channel. Again, the RLS-based predictor performs practically as well as the MMSE predictor. However, the performance

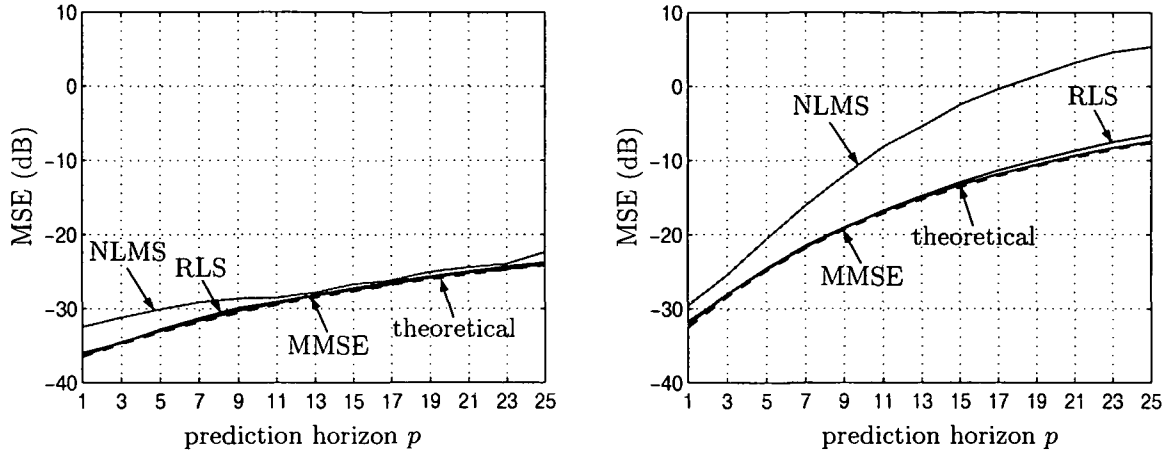


Figure 3.12: *Dependence of the prediction MSE on the prediction horizon p for (a) the slow channel and (b) the fast channel.*

of the NLMS-based predictor is substantially worse, with the MSE exceeding $-10\,\mathrm{dB}$ for $p \geq 10$.

3.5.4 Tracking of Nonstationary Channel Statistics

We next study the ability of the adaptive channel predictors to track nonstationary channel statistics. The nonstationary channel was derived from two different WSSUS channels. Channel 1 had a flat scattering function (i.e., rectangular delay and Doppler profiles) with maximum delay $L = 14$ and maximum Doppler $\xi_{\max} K = 0.004$, and the SNR was 25 dB. Channel 2 had the Jakes-exponential scattering function (3.59) with $L = 19$ and $\xi_{\max} K = 0.01$, and the SNR was 35 dB. Channel 1 was in force during the first 1000 OFDM symbols (phase A), and channel 2 was in force during the last 1000 OFDM symbols (phase C). During the intermediate transition phase (phase B), the channel impulse response and SNR at each time instant were obtained by linear interpolation of the impulse responses and SNR values of channel 1 and channel 2.

Figure 3.13 shows the normalized prediction MSE vs. the OFDM symbol index n for the adaptive channel predictors operated in decision-directed mode (initial convergence is not shown). The coded OFDM system was used. The prediction MSE was estimated from 100 realizations. It is seen that both adaptive algorithms succeed in tracking the nonstationary statistics. During phases A and C, the RLS-based predictor performs practically as well as the respective MMSE predictor. The NLMS-based predictor has an excess MSE of about 3 dB. During phase B, the adaptive predictors track the variation of the channel statistics

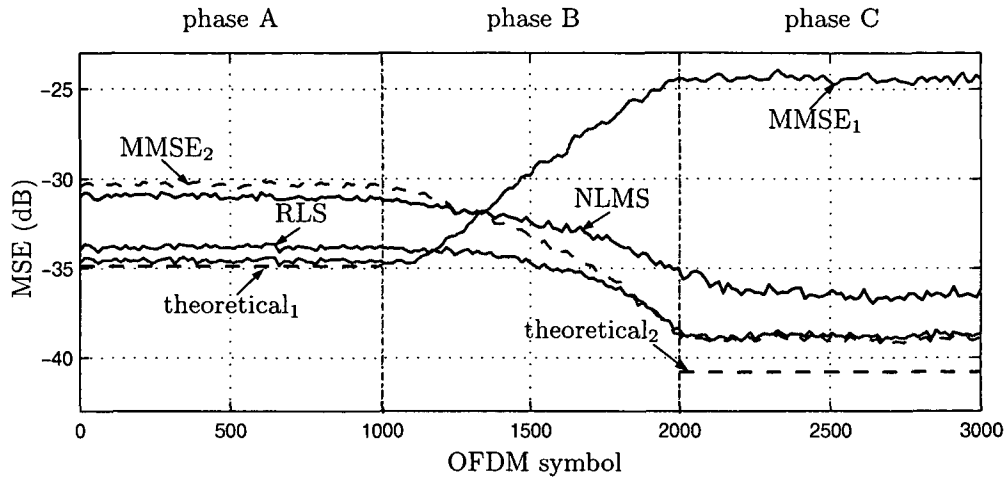


Figure 3.13: *Tracking of a channel with nonstationary statistics. The curves labeled* MMSE_1 *and* MMSE_2 *show the (estimated) MSE of the MMSE predictor designed for channel 1 and 2, respectively. The curves labeled* $\mathrm{theoretical}_1$ *and* $\mathrm{theoretical}_2$ *show the theoretical MMSE for channel 1 and channel 2, respectively.*

without problems. As expected, the MMSE channel predictors perform rather poorly if they are not matched to the current channel conditions. This is especially seen during phase C where MMSE_1 is about 13 dB above MMSE_2, which in turn is about 2 dB above the theoretical MMSE. This can be attributed to strong intercarrier interference that is caused by the fast channel and violates the system relation (2.36).

3.5.5 Performance of Predictive Equalization

We simulated the uncoded and coded OFDM systems applying predictive equalization that we proposed in Subsection 3.4.1 (see Figure 3.5). The results were averaged over 25 realizations with 10^4 OFDM symbols each. All channel predictors were operated in decision-directed mode except during the initial convergence for which training mode operation was used. For comparison, we also simulated an OFDM system using pilot symbol assisted (PSA) channel estimation [59]. For PSA channel estimation, approximately 10% of the transmitted symbols were used as pilots. The PSA channel estimator was designed according to the robust technique described in [59], using knowledge of the maximum delay L, maximum Doppler shift $\xi_{\max}$, and SNR, but not of the exact shape of the scattering function. In the MSE plots, we also show the theoretical performance of the MMSE channel predictor calculated according to (3.21). Furthermore, for the bit error rate (BER) results, the performance of an ideal receiver with perfect CSI is also plotted.

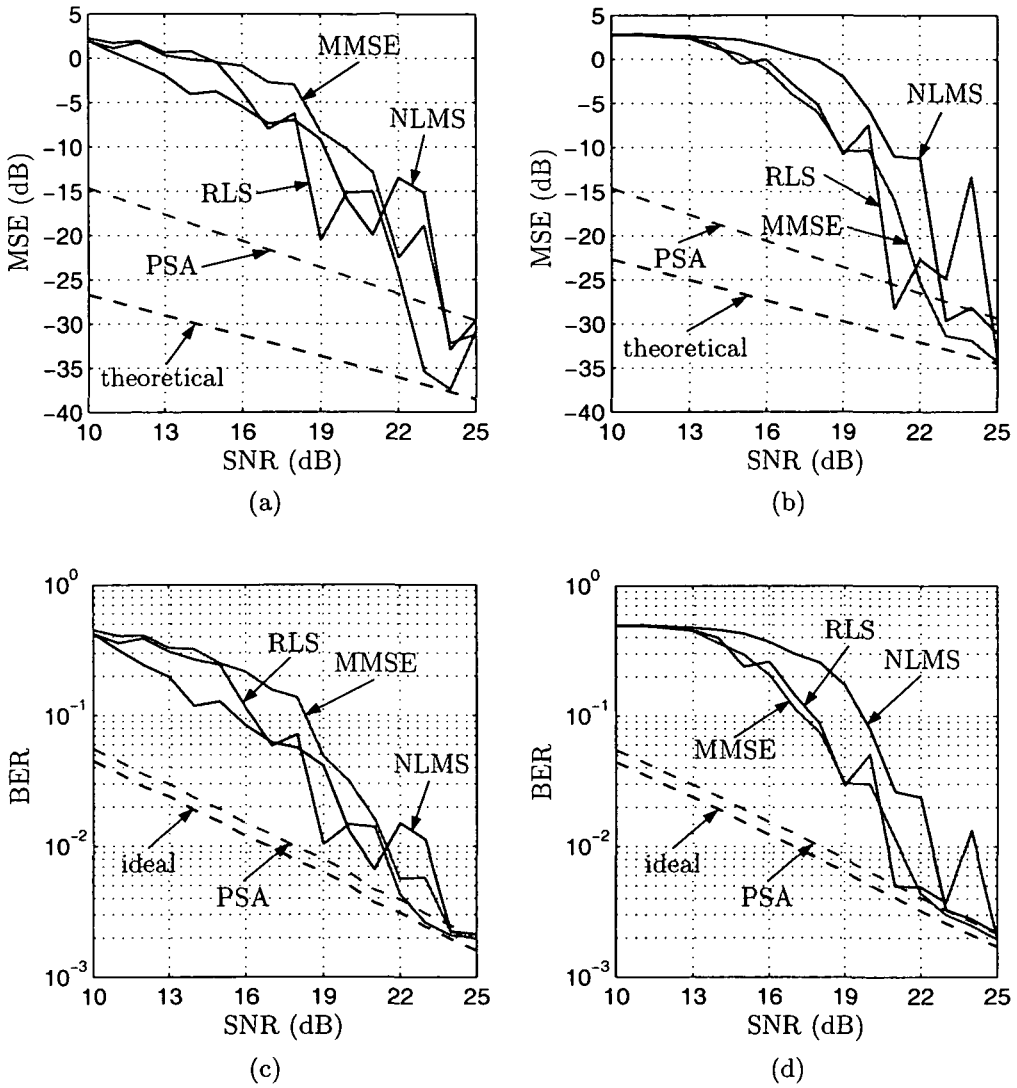


Figure 3.14: *MSE and BER performance of predictive equalization and PSA equalization within the uncoded system: (a) MSE for the slow channel; (b) MSE for the fast channel; (c) BER for the slow channel; (d) BER for the fast channel.*

Uncoded System

The uncoded system is considered in Figure 3.14. In Figure 3.14(a),(b), we show the normalized MSE obtained with the various receivers vs. the SNR for the slow channel and the fast channel, respectively. It is seen that the MSE for the predictive receivers is unsatisfactory up to a threshold as large as 23 ... 25 dB. This behavior is due to error propagation that results in poor channel tracking performance. The PSA system does not suffer from error propagation and thus exhibits no SNR threshold. However, it does not achieve the theoretical prediction MMSE since only 10 % of the subcarriers are used to

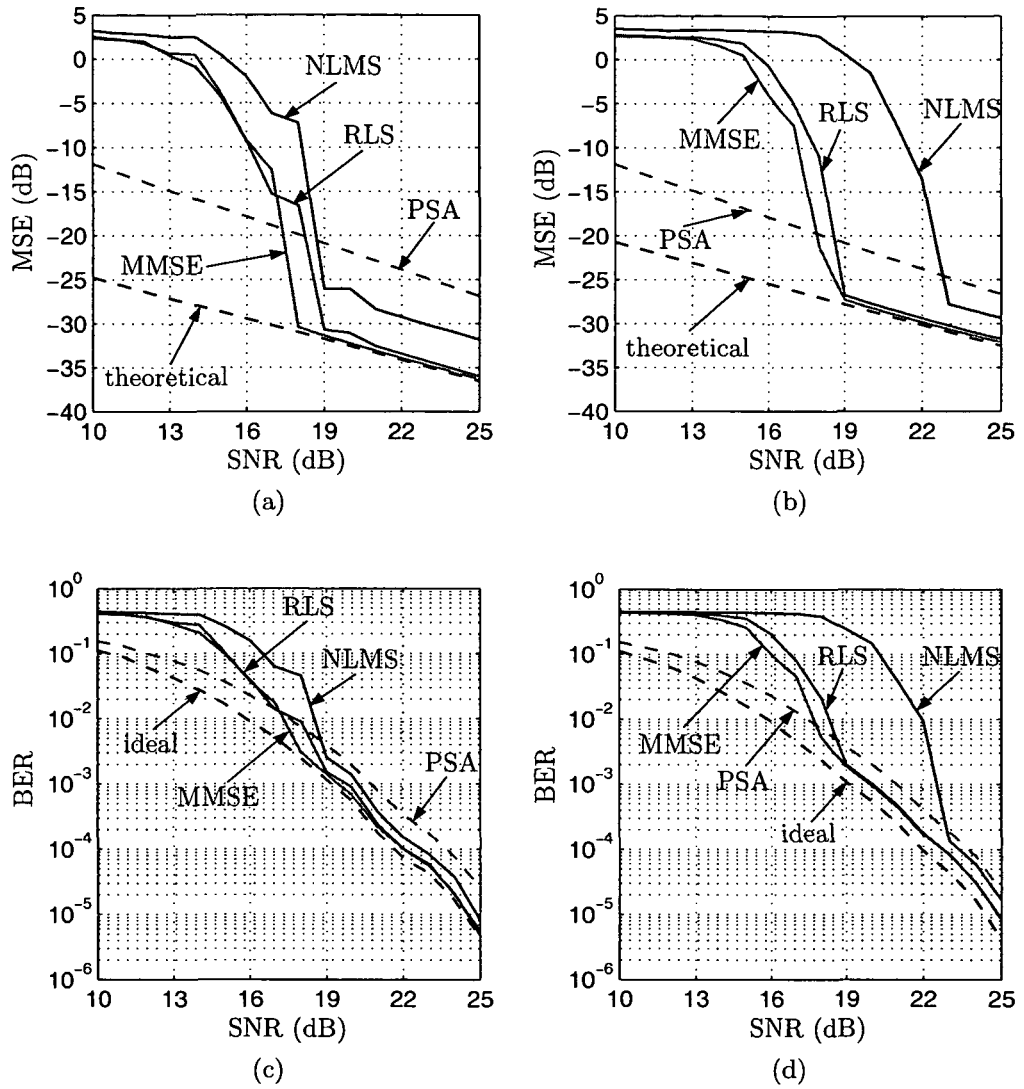


Figure 3.15: *MSE and BER performance of predictive equalization and PSA equalization within the coded system: (a) MSE for the slow channel; (b) MSE for the fast channel; (c) BER for the slow channel; (d) BER for the fast channel.*

transmit pilots.

In Figure 3.14(c),(d), the corresponding BERs are shown. For the predictive receivers, below the SNR threshold, the poor MSE results in high BER. Stable channel tracking is only possible above the SNR threshold of about 23 ... 25 dB.

Coded System

We next consider the coded system. This system has approximately the same net data rate as the uncoded system since a 16-QAM symbol alphabet and coding rate 1/2 are used. In Figure 3.15(a),(b), the normalized MSE is shown vs. the SNR for the slow channel and

the fast channel, respectively. Due to the decision-directed operation, good performance of the predictors again requires the SNR to be above a certain threshold. However, this SNR threshold is seen to be much lower compared to the uncoded system (cf. Figure 3.14(a),(b)). It is approximately equal for the RLS-based predictor and the MMSE predictor, and about 2–4 dB higher for the NLMS-based predictor. For SNRs higher than this threshold, the predictive receivers are close to the theoretical performance and clearly outperform the PSA receiver.

In Figure 3.15(c),(d), we show the BERs after channel decoding obtained with the various receivers vs. the SNR for the slow and fast channel, respectively. For SNR values above the threshold, the BER of the receivers with predictive channel equalization is nearly equal to the BER of the ideal receiver. The receivers using the MMSE-based and RLS-based channel predictors have practically equal performance. The receiver using the NLMS-based channel predictor has slightly poorer performance. Moreover, our decision-directed predictive receivers outperform the PSA receiver for SNRs above the threshold. This is because our channel predictors use symbol decisions from all K subcarriers (note that this causes their computational complexity to be larger than that of the PSA estimator).

The above results show that for the coded system, the SNR threshold is significantly lower than for the uncoded system. It may be expected that a more elaborate channel code will result in an ever lower threshold.

3.5.6 SNR Threshold

The SNR threshold also depends on the packet length used in the OFDM system. To appreciate this dependence, consider the realizations of the prediction errors in Figure 3.16 that were obtained with the uncoded OFDM system and the various predictive equalizers for the slow channel at an SNR of 18 dB. According to Figure 3.14(a), this SNR value is below the threshold for all channel predictors. For comparison, we also show the MSE of PSA channel estimation and the theoretical prediction MMSE. It is seen that the predictive receivers are able to track the time-varying channel for some time until catastrophic error propagation sets in and the predictors fail to track the channel. In Figure 3.16, this happens for the MMSE predictor and the RLS-based predictor after about 1800 OFDM symbols and for the NLMS-based predictor after about 3500 OFDM symbols. Note that the reverse effect may also occur: the predictors randomly converge without transmission of a training packet. In Figure 3.16, this happens for the NLMS predictor after about 10^4 OFDM symbols.

If we now calculate (estimate) MSEs by averaging the realizations of the prediction

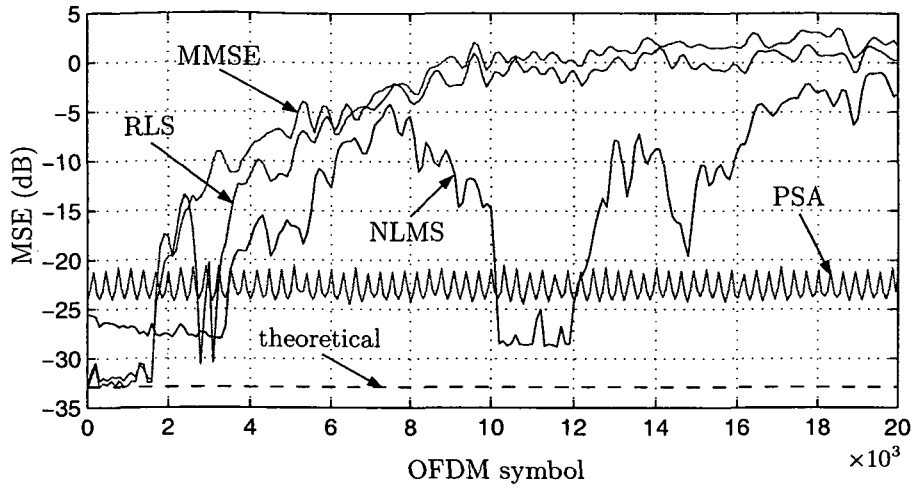


Figure 3.16: *Realization of the prediction error for the uncoded system at an SNR of* 18 *dB. The receivers applying predictive equalization are able to track the channel for about* 1800 *OFDM symbols (RLS and MMSE) or* 3500 *OFDM symbols (NLMS) before loosing track.*

errors in Figure 3.16 over time interfvals starting at the first OFDM symbol, we obtain different values depending on how many samples of the realizations are included in the calculation. For example, the MSE obtained by averaging over the first 10^3 OFDM symbols will be lower than the MSE obtained by averaging over the first 10^4 OFDM symbols.

To explore this dependence of the SNR threshold on the packet length, we show in Figure 3.17 the BERs of the various receivers averaged over 25 realizations vs. the SNR.

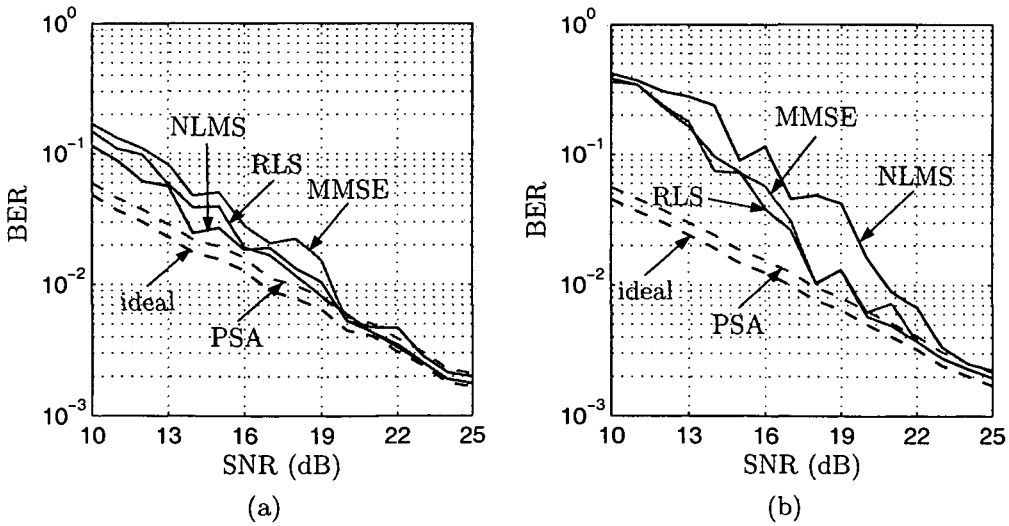


Figure 3.17: *BER performance of predictive equalization and PSA equalization: (a) BER for the slow channel; (b) BER for the fast channel.*

The performance of an ideal receiver with perfect CSI is also shown. A packet length of 10^3 OFDM symbols was used. Figure 3.17 should be compared with Figure 3.14(c),(d) which shows analogous results for a packet length of 10^4 OFDM symbols. For the shorter packet length, the SNR threshold is significantly reduced. In fact, for the slow channel in Figure 3.17(a), the threshold nearly disappears. Here, the predictive receivers perform close to optimum within the entire SNR range. For the fast channel in Figure 3.17(b), the SNR threshold is reduced to about 18 dB for the MMSE-based and RLS-based predictors and to about 22 dB for the NLMS-based predictor.

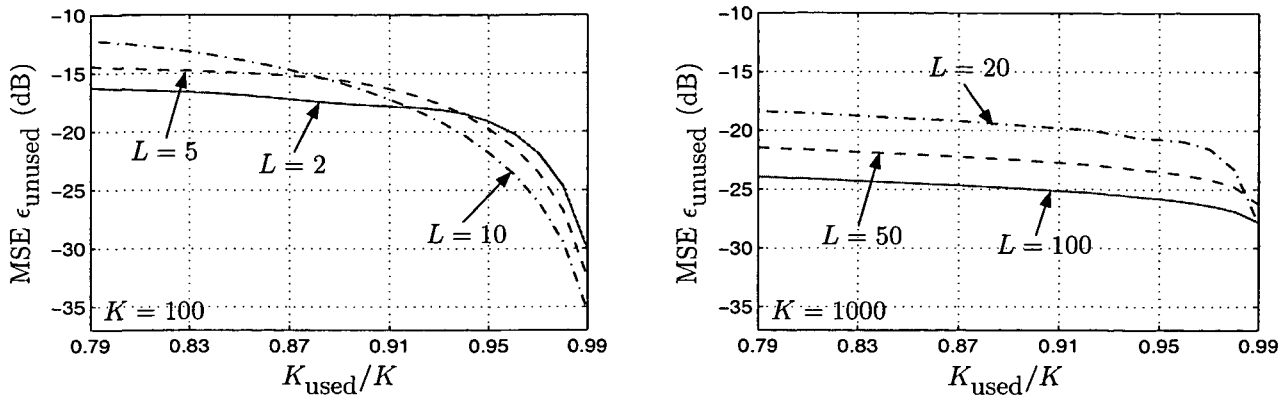


Figure 3.18: *Systematic error (MSE)* ϵ_{unused} *introduced by the DFT implementation when only* K_{used} *of the* K *subcarriers are used: (a) for* $K = 100$*, (b) for* $K = 1000$*.*

3.A Systematic Error Caused by Unused Subcarriers

We analyze the systematic error obtained with the DFT implementation from Subsection 3.2.3 (see Figure 3.4) when only the subcarriers $k \in \mathcal{K} \triangleq \{K_{\min}, K_{\min}+1, \ldots, K_{\max}\}$ are used for data transmission and the symbols of the remaining subcarriers are set to zero. Thus, only $K_{\text{used}} \triangleq K_{\max} - K_{\min} + 1$ subcarriers transmit data symbols. For simplicity, we assume zero noise ($Z_{n,k} = 0$) and a time-invariant channel ($H_{n,k} = H_k$ and $h_{n,l} = h_l$). Furthermore, we assume that (2.36) is satisfied exactly, so that $Y_{n,k} = H_k X_{n,k}$.

According to Figure 3.4, the first step of the DFT implementation is a division of the received sequences by the data symbols, which yields $\widetilde{H}_{n,k} = Y_{n,k}/X_{n,k} = H_k = \sum_{l=0}^{L} h_l\, e^{-j2\pi kl/K}$ for $k \in \mathcal{K}$. The last equation can be written as $\boldsymbol{\mathcal{H}} = \mathbf{U}\mathbf{h}$, where $\boldsymbol{\mathcal{H}} \triangleq \left[H_{K_{\min}} \cdots H_{K_{\max}}\right]^T$, $\mathbf{h} \triangleq \left[h_0 \cdots h_L\right]^T$, and the columns of the $K_{\text{used}} \times (L+1)$ matrix $\mathbf{U}$ are $\left[e^{-j2\pi K_{\min} l/K} \ldots e^{-j2\pi K_{\max} l/K}\right]^T$, $l = 0, \ldots, L$. Note that typically $K_{\text{used}} \gg L$. Applying to $\boldsymbol{\mathcal{H}} = \mathbf{U}\mathbf{h}$ an IDFT as in (3.22) followed by a DFT as in (3.24) yields

$$\widehat{\boldsymbol{\mathcal{H}}} = \frac{1}{K}\mathbf{U}\mathbf{U}^H\boldsymbol{\mathcal{H}} = \frac{1}{K}\mathbf{U}\mathbf{U}^H\mathbf{U}\mathbf{h}.$$

Because $\mathbf{U}\mathbf{U}^H \neq K\mathbf{I}$ for $K_{\text{used}} < K$, we have $\widehat{\boldsymbol{\mathcal{H}}} \neq \boldsymbol{\mathcal{H}}$. Using results from [168], the systematic MSE $\epsilon_{\text{unused}} \triangleq \frac{1}{K_{\text{used}}} \mathrm{E}\{\|\boldsymbol{\mathcal{H}} - \widehat{\boldsymbol{\mathcal{H}}}\|_F^2\}$ ($\|\cdot\|_F$ denotes the Froebenius norm) can be shown to be

$$\epsilon_{\text{unused}} = \frac{1}{K_{\text{used}}} \mathrm{tr}\left\{\mathbf{U}\Big(\mathbf{I} - \frac{1}{K}\mathbf{U}^H\mathbf{U}\Big)\mathbf{D}_{\mathbb{H}}\Big(\mathbf{I} - \frac{1}{K}\mathbf{U}^H\mathbf{U}\Big)\mathbf{U}^H\right\}, \tag{3.60}$$

where $\mathbf{D}_{\mathbb{H}} \triangleq \mathrm{E}\{\mathbf{h}\mathbf{h}^H\}$ and $\boldsymbol{\gamma}_l$ is the lth column of the matrix $\boldsymbol{\Gamma} \triangleq \mathbf{U}\big(\mathbf{I} - \frac{1}{K}\mathbf{U}^H\mathbf{U}\big)$. Because $\mathbf{D}_{\mathbb{H}}$ is diagonal due the WSSUS assumption (2.45), i.e., $\mathbf{D}_{\mathbb{H}} = \mathrm{diag}\{D_{\mathbb{H}}[0], \ldots, D_{\mathbb{H}}[L]\}$, we

obtain for (3.60)

$$\epsilon_{\text{unused}} = \frac{1}{K_{\text{used}}} \sum_{l=0}^{L} D_{\mathbb{H}}[l] \, \|\gamma_l\|^2 \, . \tag{3.61}$$

Figure 3.18 shows ϵ_{unused} (calculated from (3.61)) vs. the percentage of used subcarriers K_{used}/K for OFDM systems with $K = 100$ and $K = 1000$ subcarriers and three different values of the maximum channel delay L. We assumed $\mathbf{D}_{\mathbb{H}} = \frac{1}{L+1}\mathbf{I}$, i.e., all $L+1$ paths are equally strong and the path loss is normalized to $\text{tr}\{\mathbf{D}_{\mathbb{H}}\} = 1$. It is seen that, as expected, the error decreases as K_{used} increases. For $K_{\text{used}}/K = 0.9$, ϵ_{unused} is about $-16\,\text{dB}$ for $K = 100$ subcarriers and between $-20\,\text{dB}$ and $-25\,\text{dB}$ for $K = 1000$ subcarriers.

4

System Capacity of Wireless OFDM Systems

It has been reported in [52–54] that for spread-spectrum-like signaling over time and frequency selective fading channels without channel knowledge at the transmitter and receiver, the information rate approaches zero for very large bandwidths. In this chapter, we will study the information rate and system capacity of OFDM for this case. More specifically, we will carry out an information-theoretic analysis with the aim of calculating the system capacity of OFDM transmission over time and frequency selective Rayleigh fading channels (underspread WSSUS channels, cf. Subsection 2.2.1) in the wideband regime.

The outline of this chapter is as follows. We start with an overview of known results in Section 4.2. We concentrate on the asymptotic result of [96–99] and the bounds derived in [52–54].

Next, we consider two specific codebooks and calculate the resulting information rate. In Section 4.3, we use an orthogonal codebook similar to that in [99, Section 8.6]. By a similar derivation, we are able to calculate the asymptotic information rate of OFDM. This rate is equal to the asymptotic capacity in [99, Section 8.6] and the asymptotic AWGN

capacity, and hence it is the asymptotic system capacity of OFDM; moreover, OFDM achieves capacity for infinite bandwidth.

A more practical codebook is considered in Section 4.4 where we use i.i.d. data symbols drawn from a constant-modulus symbol alphabet. This codebook (among others) is used in IEEE 802.11a, Hiperlan/2, and DVB-T. For constant-modulus signaling and given channel statistics, we can calculate an exact expression for the information rate of wideband OFDM. Our result is based on the well-known relation between channel uncertainty and prediction MMSE [53, 107], and the new results on MMSE channel prediction in OFDM systems presented in Chapter 3. We show that for fixed transmit power, the information rate tends to zero in the large-bandwidth limit irrespectively of the channel statistics. Moreover, for finite bandwidths, we quantify the impact of the spread and shape of the scattering function on channel predictability and the associated reduction in information rate. In particular, we demonstrate that channel uncertainty (and hence reduction of information rate) is maximized if the scattering function is flat over its support region. We provide guidelines on how to choose system parameters for given channel statistics so as to ensure that the system operates below the *critical bandwidth* where the system capacity is close to the AWGN capacity. Finally, we show that information rate is not reduced in purely specular scattering environments since in this special case the channel can be predicted perfectly.

In Section 4.5, we consider the system capacity of OFDM. We derive an upper and a lower bound since we cannot provide an exact result. The lower bound follows from the calculation of the information rate for constant-modulus signaling in Section 4.4 for bandwidths below the critical bandwidth and is close to the AWGN capacity. Above the critical bandwidth, however, it remains at a fixed rate. Therefore, the system capacity of OFDM is close to the AWGN capacity up to a critical bandwidth. Moreover, it does not approach zero in the large-bandwidth limit.

In Section 4.6, we calculate two lower bounds on the information rate of OFDM for Gaussian signaling. However, no closed-form expressions for these lower bounds are found.

In Section 4.7, we demonstrate by numerical examples that the critical bandwidth is extremely high for typical system and channel parameters. Furthermore, up to the critical bandwidth our upper and lower bounds on the OFDM system capacity practically coincide and both are close to the AWGN channel capacity. We also demonstrate that current OFDM communication systems operate far below the critical bandwidth and thus can achieve system capacities close to the AWGN capacity.

We note that parts of this chapter have been submitted for publication [110, 111].

4.1 Definitions and Notation

In this chapter, it will be important to distinguish between random sequences/vectors and their realizations. Therefore, we will denote random quantities by capital letters and their realizations by lower-case letters. Thus, a random vector will be denoted by a boldface capital letter and its realization by a boldface lower-case letter. Some care is necessary with correlation matrices that will also be denoted by boldface capital letters even though they are nonrandom.

We define the *ergodic information rate* of the OFDM system by

$$R \triangleq \lim_{M\to\infty} \frac{1}{MT} I(\mathbf{Y};\mathbf{X}). \tag{4.1}$$

Here, $I(\cdot,\cdot)$ is the mutual information of two vectors, T is the duration of one OFDM symbol, and M is the number of OFDM symbols considered (hence, MT is the total transmission time); furthermore, the channel input and output vectors of size $MK \times 1$ are respectively defined by

$$\mathbf{X} \triangleq \left[\mathbf{X}_0^T\, \mathbf{X}_1^T \ldots \mathbf{X}_{M-1}^T\right]^T, \qquad \text{with} \quad \mathbf{X}_n \triangleq [X_{n,0}\, X_{n,1} \ldots X_{n,K-1}]^T,$$
$$\mathbf{Y} \triangleq \left[\mathbf{Y}_0^T\, \mathbf{Y}_1^T \ldots \mathbf{Y}_{M-1}^T\right]^T, \qquad \text{with} \quad \mathbf{Y}_n \triangleq [Y_{n,0}\, Y_{n,1} \ldots Y_{n,K-1}]^T,$$

where $X_{n,k}$ and $Y_{n,k}$ are the channel input and output, respectively.

Maximization of the information rate (4.1) over all possible codebooks or distributions of $\mathbf{X}$ yields the *OFDM system capacity*

$$S \triangleq \max R, \tag{4.2}$$

where we have to impose a power constraint on $\mathbf{X}$. This power constraint can be formulated as a peak power constraint or as an average power constraint.

We furthermore recall that, neglecting intersymbol and intercarrier interference, the system input-output relation is given by $Y_{n,k} = H_{n,k}X_{n,k} + Z_{n,k}$ (cf. (2.36)). This input-output relation can be expressed by

$$\mathbf{Y} = \operatorname{diag}\{\mathbf{H}\}\,\mathbf{X} + \mathbf{Z} = \operatorname{diag}\{\mathbf{X}\}\,\mathbf{H} + \mathbf{Z}, \tag{4.3}$$

where the $MK \times 1$ channel vector is defined as

$$\mathbf{H} = \left[\mathbf{H}_0^T\, \mathbf{H}_1^T \ldots \mathbf{H}_{M-1}^T\right]^T, \qquad \text{with} \quad \mathbf{H}_n = [H_{n,0}\, H_{n,1} \ldots H_{n,K-1}]^T,$$

and the $\operatorname{diag}\{\cdot\}$ operation generates a diagonal matrix whose diagonal elements are given by the argument vector. We will assume that $\mathbf{Z}$ is circularly symmetric complex Gaussian

white noise with covariance matrix $\mathbf{R}_Z = \sigma_z^2\mathbf{I}$, i.e., $\mathbf{Z} \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2\mathbf{I})$. Furthermore, note that $\sigma_z^2 = N_0$ is independent of the bandwidth B. Hence, our analysis pertains to OFDM with identical orthogonal pulses at the transmitter and the receiver (cf. Section 2.1).

For subsequent use, we note that the differential entropy of a circularly complex Gaussian vector $\mathbf{A}$ of dimension $M \times 1$ with covariance matrix $\mathbf{R}_A = \mathrm{E}\{\mathbf{A}\mathbf{A}^H\}$ is given by [169]

$$h(\mathbf{A}) = \log\left((\pi e)^M \det[\mathbf{R}_A]\right).$$

All calculations in this chapter are carried out in natural units (nats). One nat is equal to $1/\log(2) \approx 1.4427$ bit. We will use the unit bit only for the numerical results. Throughout our development, we consider time and frequency selective WSSUS Rayleigh fading channels and assume that neither the transmitter nor the receiver has CSI. However, it is assumed that the receiver has perfect knowledge of the channel statistics.

4.2 Overview of Known Results

We will here summarize the most important known information-theoretic results that are closely related to our analysis of OFDM system capacity. For fading channels, the results very much depend on the considered channel model and the state of information about the channel that is available at the transmitter and the receiver. We will only consider time and frequency selective WSSUS Rayleigh fading channels, among which flat Rayleigh fading channels are a special case. For completeness, we briefly consider the case that CSI is available at the receiver. However, the more interesting case is where the receiver has no CSI. An overview of results for the case that channel knowledge is available at the receiver and/or transmitter is given in [95].

4.2.1 CSI Available at Receiver

If knowledge of the channel realization $\mathbf{H} = \mathbf{h}$ is available at the receiver, we can apply the results of [170, 171] since (4.3) is simply the input-output relation of a multi-input multi-output (MIMO) channel. Hence, the instantaneous information rate of a block of M OFDM symbols is given by

$$\begin{aligned} R_{\mathrm{CSI,inst}} &= \frac{1}{MT}\log\left(\det\left[\mathbf{I} + \frac{\sigma_x^2}{N_0}\mathrm{diag}\{\mathbf{h}\}\,\mathrm{diag}\{\mathbf{h}^H\}\right]\right) \\ &= \frac{1}{MT}\sum_{n=1}^{M}\sum_{k=0}^{K-1}\log\left(1 + \frac{\sigma_x^2|h_{n,k}|^2}{N_0}\right), \end{aligned} \tag{4.4}$$

where we assumed i.i.d. data symbols $X_{n,k}$ with distribution $\mathcal{CN}(0, \sigma_x^2)$. Averaging over the realizations of the channel coefficients $h_{n,k}$ in (4.4) yields the (ergodic) system capacity

$$S_{\text{CSI}} = \frac{1}{MT} \sum_{n=1}^{M} \sum_{k=0}^{K-1} \mathrm{E}\left\{ \log\left(1 + \frac{\sigma_x^2 |H_{n,k}|^2}{N_0}\right)\right\} \leq \frac{K}{T} \log\left(1 + \frac{\sigma_x^2 \sigma_{\mathbb{H}}^2}{N_0}\right), \tag{4.5}$$

where Jensen's inequality [169] has been used; the deviation of the bound from the true value is known as Jensen's penalty. Moreover, we recall that the channel's path loss is $\sigma_{\mathbb{H}}^2 = \mathrm{E}\{|H_{n,k}|^2\}$ (cf. (2.52).

4.2.2 CSI Unavailable at Receiver

The situation where CSI is unavailable at the receiver is much more complicated than the previous one and no results for the channel capacity of OFDM are known. However, many information-theoretic results have been obtained for fading channels. We briefly recall some of these results here.

- The channel capacity of the time and frequency selective WSSUS channel at infinite bandwidth (see [96–98] and equation (8.6.37) in [99, Section 8.6]) is given by (in units of nats per second)

$$C = \frac{P\sigma_{\mathbb{H}_c}^2}{N_0}, \tag{4.6}$$

  where P is the transmit power, $\sigma_{\mathbb{H}_c}^2$ is the path loss, and N_0 is the power spectral density of the noise. Hence, for infinite bandwidth the capacity is identical to the asymptotic capacity of the AWGN channel operating at an SNR of $P\sigma_{\mathbb{H}_c}^2/N_0$. It is important to note that this result (4.6) is obtained without assuming a peak power constraint at the transmitter. More details will be presented in Section 4.3.

- The performance of an M-ary orthogonal communication system using stationary signals was investigated in [100]. This analysis pertains also to FSK signaling since the considered system model is

$$Y(t) = H(t)\, e^{j\omega_m t} + Z(t), \quad m = 1, 2, \ldots, M, \tag{4.7}$$

  using M frequencies to signal over a flat Rayleigh fading channel $H(t)$. Note that (4.7) implicitly imposes a peak and average power constraint at the transmitter since the "data symbols" have variance $|e^{j\omega_m t}|^2 = 1$. Hence, the result of [100] specifies the system capacity of FSK signaling over flat fading channels,

$$S_{\text{FSK}} = \frac{\sigma_{\mathbb{H}_c}^2}{N_0} - \int_{\nu} \log\left(1 + \frac{S_{\mathbb{H}_c}(\nu)}{N_0}\right) d\nu, \tag{4.8}$$

where $S_{\mathbb{H}_c}(\nu)$ is the Doppler spectrum of the channel. We see that the first component of (4.8) is identical to the infinite-bandwidth channel capacity (4.6) with transmit power $P = 1$. However, the second component in (4.8) is always negative since $S_{\mathbb{H}_c}(\nu) \geq 0$ and therefore it reduces S_{FSK}. This component can be interpreted as a penalty to system capacity due to the channel uncertainty at the receiver. For low SNR, we can use $\log(1+x) \approx x$ which yields $S_{\text{FSK}} = 0$. We recall that low SNR is obtained in the wideband regime.

An important difference between the assumptions underlying (4.6) and (4.8) is that for the first result an average power constraint is assumed whereas for the second result additionally a peak power constraint is imposed.

- The capacity of the discrete-time flat uncorrelated Rayleigh fading channel has recently been studied in [172]. The considered system model is given by

$$Y_n = H_n X_n + Z_n \,,$$

where X_n is the channel input, Y_n is the output, and H_n and Z_n are mutually independent complex Gaussian random variables. This corresponds to a flat Doppler profile, i.e., the channel coherence time is zero. The input X_n is furthermore subject to an average power constraint $\mathrm{E}\{|X_n|\} \leq P$. The main result of [172] is that the capacity-achieving distribution of X_i is discrete with a finite number of constellation points that increases monotonically with the SNR. For low SNR, it is reported that on-off keying (OOK) is optimal.

- Recent results on time and frequency selective fading channels that are closely related to our analysis are reported in [52–54]. These papers consider spread-spectrum-like signaling over time and frequency selective WSSUS channels. The common result is that under a peak power constraint the system capacity approaches zero in the large-bandwidth limit. Moreover, it is reported that the signaling scheme must be *peaky* in order to obtain nonzero rates in the asymptotic limit.

  - In [52], it is shown that for spread-spectrum-like signaling, the information rate is lower and upper bounded as

$$C_{\text{AWGN}} \left[1 - \frac{\tilde{L}}{\tilde{L}_{\text{crit}}} \log\left(1 + \frac{\tilde{L}_{\text{crit}}}{\tilde{L}} \right) \right] \leq R \leq C_{\text{AWGN}} \frac{\tilde{L}_{\text{crit}}}{\tilde{L}} \,,$$

    where $C_{\text{AWGN}} = P/N_0$ is the capacity of the infinite-bandwidth AWGN channel, $\tilde{L}$ is the number of resolved delay taps (which depends on the sampling rate),

and $\tilde{L}_{\text{crit}} \triangleq PT_c/N_0$ is a critical number of delay taps (here, T_c is the coherence time of the channel; note that in [52] the definition of T_c is different from our definition of $T_{\mathbb{H}_c}$ in Subsection 2.2.1). Furthermore, the path loss of the channel is normalized to one and a flat scattering function is assumed. Note that $\tilde{L}_{\text{crit}}$ increases with the SNR and the channel coherence time. If $\tilde{L} \ll \tilde{L}_{\text{crit}}$, then the information rate achievable with spread-spectrum signals is close to the capacity of the infinite-bandwidth AWGN channel, i.e., $R \approx C_{\text{AWGN}}$. However, if $\tilde{L} \gg \tilde{L}_{\text{crit}}$, the upper bound approaches zero and thus $R \approx 0$. Furthermore, note that the structure of the lower bound resembles (4.8).

- In [53], an upper bound on the mutual information for spread-spectrum-like transmission over time and frequency selective channels is developed. Under a fourth-moment constraint on the channel input signal, it is shown that in the large-bandwidth limit the mutual information approaches zero. The upper bound increases with increasing coherence time of the channel.

- Finally, in [54] an upper bound on the mutual information is derived using the concept of *capacity per unit cost* [173]. This bound depends on a fourth-order cost function called *forthegy* that is defined as

$$J_C(x) \triangleq \int_\tau \int_\nu |\chi(\tau,\nu)|^2 \, \Psi_{\mathbb{H}_c}(\tau,\nu)\, d\tau d\nu \,,$$

where

$$\Psi_{\mathbb{H}_c}(\tau,\nu) \triangleq \int_{\tau'} \int_{\nu'} C_{\mathbb{H}_c}(\tau',\nu')\, C_{\mathbb{H}_c}(\tau+\tau',\nu+\nu')\, d\tau' d\nu'$$

is the convolution of the scattering function with itself and

$$\chi(\tau,\nu) \triangleq \int_t x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{j2\pi\nu t} dt$$

is the ambiguity function of the channel input signal $x(t)$. Evaluating this bound for spread-spectrum signaling again shows that the information rate tends to zero in the large-bandwidth limit.

4.3 OFDM System Capacity for Infinite Bandwidth

We next calculate the system capacity of OFDM for infinite bandwidth. Our derivation follows the classical analysis of the capacity of time and frequency selective channels for

infinite bandwidth (see [99, Section 8.6] and [96–98]). We construct an orthogonal codebook that is peaky in time and in frequency, i.e., it concentrates the transmit energy in time and frequency instead of spreading it out. We then apply a maximum likelihood (ML) detector and calculate its error probability. The result is that we can obtain arbitrarily low error probability if the information rate is below a certain value—the infinite-bandwidth system capacity of OFDM transmitting over a time and frequency selective fading channel.

Surprisingly, this system capacity does not depend on OFDM system parameters such as the symbol duration T and the subcarrier separation F but only on the SNR at the receiver. Moreover, the OFDM system capacity is equal to the capacity of the time and frequency selective channel, which itself is equal to the infinite-bandwidth AWGN capacity. Hence, for infinite bandwidth OFDM is able to approach capacity.

Codebook

We use the orthogonal signaling scheme with K codewords that is sketched in Figure 4.1. The kth codeword is defined by setting

$$X_{n,k'} = \begin{cases} \alpha, & k' = k, \quad n = 0, M, 2M, \dots, \\ 0, & \text{else}, \end{cases}$$

i.e., every M OFDM symbols we only use the kth subcarrier to transmit one symbol $\alpha > 0$. All other subcarriers are not used. We note that this signaling scheme is similar to FSK signaling. The transmit bandwidth is given by $B = KF$ where the subcarrier frequency separation F is assumed fixed. Thus, to let B approach infinity we will let K grow. Furthermore, the time duration between two successive codewords is chosen as

$$T' = \frac{T}{\Theta}, \qquad 0 < \Theta \leq 1,$$

where T is the time duration of one OFDM symbol and Θ is the duty cycle (see Figure 4.1). The information rate of this signaling scheme in units of nats/sec is

$$R = \frac{\log K}{T'} = \Theta \frac{\log K}{T}. \tag{4.9}$$

To obtain constant transmit power P per codeword of duration T', we define

$$\alpha^2 = \frac{PT}{\Theta}. \tag{4.10}$$

The parameters we may vary are Θ and K. In particular, if we let Θ approach zero, fewer codewords are transmitted but α^2 increases. To obtain nonzero rates when Θ approaches

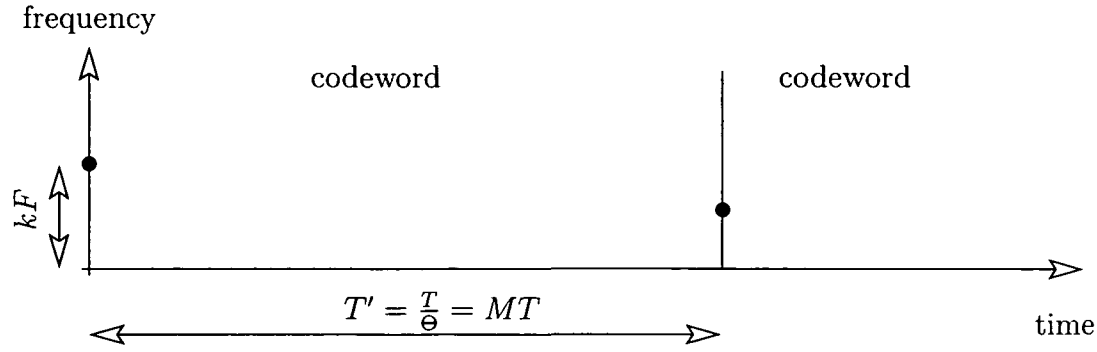


Figure 4.1: *Illustration of the orthogonal-signaling codebook used to calculate the infinite-bandwidth capacity of OFDM systems. The effective length of a codeword is T and the total length is MT. During the effective codeword length, a single OFDM symbol with only one active subcarrier is transmitted.*

zero, K must simultaneously grow exponentially fast. Since K is directly linked with the bandwidth via the relation $B = KF$, this amounts to increasing the transmit bandwidth exponentially fast. Hence, this codebook is rather inefficient with respect to spectral efficiency, which is a typical drawback of orthogonal signaling schemes [102].

ML Detector

Without loss of generality, we assume that the first codeword (i.e., one OFDM symbol at $n = 0$ and subcarrier k) is transmitted. With the OFDM input-output relation in (2.36), we have

$$Y_{0,k'} = \begin{cases} \alpha H_{0,k} + Z_{0,k}\,, & k' = k\,, \\ Z_{0,k'}\,, & \text{else}\,. \end{cases} \tag{4.11}$$

The distribution of $Y_{0,k'}$ is given by

$$p(y_{0,k'}) = \begin{cases} p_1(y_{0,k'}) = \mathcal{CN}\left(0, \alpha^2\sigma_{\mathbb{H}}^2 + N_0\right)\,, & k' = k\,, \\ p_0(y_{0,k'}) = \mathcal{CN}\left(0, N_0\right)\,, & \text{else}\,, \end{cases} \tag{4.12}$$

where $\sigma_{\mathbb{H}}^2 = \mathrm{E}\{|H_{0,k}|^2\}$ is the path loss (see (2.52)). Furthermore, $\mathrm{E}\left\{Y_{0,l}Y_{0,l'}^*\right\} = 0$ for $l \neq l'$, i.e., the $Y_{0,k'}$ are statistically independent.

The receiver is supposed to detect the index k of the transmitted codeword, i.e., the subcarrier used at the transmitter. To this end, the ML detector maximizes the conditional joint probability density function of the K received values $Y_{0,k'}, k' = 0, 1, \ldots, K-1$ (cf.

[99, Section 8.6]):

$$p(y_{0,0}, y_{0,1}, \ldots, y_{0,K-1}|k) = p_1(y_{0,k}) \prod_{\substack{k'=0 \\ k' \neq k}}^{K-1} p_0(y_{0,k'}) = \frac{p_1(y_{0,k})}{p_0(y_{0,k})} \prod_{k'=0}^{K-1} p_0(y_{0,k'}) \sim \frac{p_1(y_{0,k})}{p_0(y_{0,k})},$$

where in the last step we used the fact that $\prod_{k'=0}^{K-1} p_0(y_{0,k'})$ does not depend on k. Hence, the ML detector decides in favor of

$$\hat{k} = \arg \max_{k \in [0,K-1]} \frac{p_1(y_{0,k})}{p_0(y_{0,k})}.$$

Bound on Error Probability

Following [99, Section 8.6], we will now bound the average error probability. Assume that the transmitter used the kth subcarrier. Let us define the error event $\mathcal{E}_{k'}$ as the event that the ML detector erroneously decides in favor of subcarrier $k' \neq k$. This is the case if and only if $y_{0,k}$ is such that

$$\frac{p_1(y_{0,k'})}{p_0(y_{0,k'})} \geq \frac{p_1(y_{0,k})}{p_0(y_{0,k})}.$$

The probability of $\mathcal{E}_{k'}$ (given that subcarrier k was used by the transmitter and $y_{0,k}$ was received) is

$$\begin{aligned}
\Pr[\mathcal{E}_{k'}|k, y_{0,k}] &= \int_{y_{0,k'}: \frac{p_1(y_{0,k'})}{p_0(y_{0,k'})} \geq \frac{p_1(y_{0,k})}{p_0(y_{0,k})}} p_0(y_{0,k'})\, dy_{0,k'} \\
&\leq \int_{y_{0,k'}: \frac{p_1(y_{0,k'})}{p_0(y_{0,k'})} \geq \frac{p_1(y_{0,k})}{p_0(y_{0,k})}} p_0(y_{0,k'}) \left[\frac{p_1(y_{0,k'})p_0(y_{0,k})}{p_0(y_{0,k'})p_1(y_{0,k})}\right]^{\frac{1}{1+\rho}} dy_{0,k'} \\
&\leq \int_{y_{0,k'}} p_0(y_{0,k'}) \left[\frac{p_1(y_{0,k'})p_0(y_{0,k})}{p_0(y_{0,k'})p_1(y_{0,k})}\right]^{\frac{1}{1+\rho}} dy_{0,k'}, \qquad \rho \geq 0.
\end{aligned}$$

Therefore, the error probability, given the kth subcarrier was used at the transmitter, can be upper bounded by

$$\begin{aligned}
\Pr[\text{error}\,|k, y_{0,k}] &= \Pr\left[\bigvee_{k' \neq k} \mathcal{E}_{k'} \middle| k, y_{0,k}\right] \leq \sum_{k' \neq k} \Pr[\mathcal{E}_{k'}|k, y_{0,k}] \leq \left(\sum_{k' \neq k} \Pr[\mathcal{E}_{k'}|k, y_{0,k}]\right)^{\rho} \\
&\leq \left(\sum_{k' \neq k} \int_{y_{0,k'}} p_0(y_{0,k'}) \left[\frac{p_1(y_{0,k'})p_0(y_{0,k})}{p_0(y_{0,k'})p_1(y_{0,k})}\right]^{\frac{1}{1+\rho}} dy_{0,k'}\right)^{\rho} \\
&= (K-1)^{\rho} \left(\int_y p_0(y) \left[\frac{p_1(y)p_0(y_{0,k})}{p_0(y)p_1(y_{0,k})}\right]^{\frac{1}{1+\rho}} dy\right)^{\rho}, \quad 0 \leq \rho \leq 1,
\end{aligned}$$

where we used the union bound. To obtain the average error probability for the kth codeword, we now have to average over $Y_{0,k}$,

$$\begin{aligned}
P_{e,k} &\triangleq \mathrm{E}\left\{\Pr\left[\text{error}\,\middle|k, y_{0,k}\right]\right\} = \int_{y_{0,k}} \Pr\left[\text{error}\,\middle|k, y_{0,k}\right] p_1(y_{0,k})\, dy_{0,k} \\
&\leq \int_{y_{0,k}} (K-1)^\rho \left(\int_y p_0(y) \left[\frac{p_1(y)p_0(y_{0,k})}{p_0(y)p_1(y_{0,k})}\right]^{\frac{1}{1+\rho}} dy\right)^\rho p_1(y_{0,k})\, dy_{0,k} \\
&= \int_{y_{0,k}} (K-1)^\rho \left(\frac{p_0(y_{0,k})}{p_1(y_{0,k})}\right)^{\frac{\rho}{1+\rho}} \left(\int_y p_0(y) \left[\frac{p_1(y)}{p_0(y)}\right]^{\frac{1}{1+\rho}} dy\right)^\rho p_1(y_{0,k})\, dy_{0,k} \\
&= (K-1)^\rho \left(\int_{y_{0,k}} p_1^{\frac{1}{1+\rho}}(y_{0,k})\, p_0^{\frac{\rho}{1+\rho}}(y_{0,k})\, dy_{0,k}\right) \left(\int_y p_1^{\frac{1}{1+\rho}}(y)\, p_0^{\frac{\rho}{1+\rho}}(y)\, dy\right)^\rho \\
&= (K-1)^\rho \left(\int_y p_1^{\frac{1}{1+\rho}}(y)\, p_0^{\frac{\rho}{1+\rho}}(y)\, dy\right)^{1+\rho},
\end{aligned}$$

which is independent of k. Hence, the bound is valid for any codeword and we can omit the subscript k. Inserting (4.12) for $p_1(y)$ and $p_0(y)$, we obtain

$$\begin{aligned}
P_e &\leq (K-1)^\rho \left(\int_y \left[\frac{1}{\pi(\alpha^2\sigma_{\mathbb{H}}^2 + N_0)} e^{-\frac{|y|^2}{\alpha^2\sigma_{\mathbb{H}}^2+N_0}}\right]^{\frac{1}{1+\rho}} \left[\frac{1}{\pi N_0} e^{-\frac{|y|^2}{N_0}}\right]^{\frac{\rho}{1+\rho}} dy\right)^{1+\rho} \\
&= (K-1)^\rho \frac{1}{\pi^{1+\rho}(\alpha^2\sigma_{\mathbb{H}}^2 + N_0)N_0^\rho} \left(\int_y e^{-|y|^2 \frac{(1+\rho)N_0 + \rho\alpha^2\sigma_{\mathbb{H}}^2}{(1+\rho)(\alpha^2\sigma_{\mathbb{H}}^2+N_0)N_0}} dy\right)^{1+\rho} \\
&= (K-1)^\rho \frac{1}{(\alpha^2\sigma_{\mathbb{H}}^2 + N_0)N_0^\rho} \left(\frac{(1+\rho)(\alpha^2\sigma_{\mathbb{H}}^2 + N_0)N_0}{(1+\rho)N_0 + \rho\alpha^2\sigma_{\mathbb{H}}^2}\right)^{1+\rho} \\
&= (K-1)^\rho \frac{\left(1 + \frac{\alpha^2\sigma_{\mathbb{H}}^2}{N_0}\right)^\rho}{\left(1 + \frac{\rho}{1+\rho}\frac{\alpha^2\sigma_{\mathbb{H}}^2}{N_0}\right)^{1+\rho}} \\
&= (K-1)^\rho\, e^{\rho\log\left(1+\frac{\alpha^2\sigma_{\mathbb{H}}^2}{N_0}\right) - (1+\rho)\log\left(1+\frac{\rho}{1+\rho}\frac{\alpha^2\sigma_{\mathbb{H}}^2}{N_0}\right)}.
\end{aligned} \tag{4.13}$$

Inserting $K = e^{T'R}$ (cf. (4.9)) into (4.13) yields the following bound on the average error probability:

$$P_e \leq e^{T'(\rho R - E_0(\rho))}, \tag{4.14}$$

where we defined

$$E_0(\rho) \triangleq \frac{\Theta}{T}\left[(1+\rho)\log\left(1 + \frac{\rho}{1+\rho}\frac{\alpha^2\sigma_{\mathbb{H}}^2}{N_0}\right) - \rho\log\left(1 + \frac{\alpha^2\sigma_{\mathbb{H}}^2}{N_0}\right)\right].$$

System Capacity

We see from (4.14) that the error probability can be made arbitrarily small by increasing T', i.e.,

$$\lim_{T' \to \infty} P_e = 0 ,$$

iff

$$\rho R - E_0(\rho) < 0 . \tag{4.15}$$

Increasing T' can be accomplished by decreasing the duty cycle Θ. Furthermore, we have $E_0(0) = 0$,

$$\frac{\partial E_0}{\partial \rho} = \frac{\Theta}{T} \left[\frac{\frac{\alpha^2 \sigma_{\mathbb{H}}^2}{N_0}}{1 + \rho + \rho \frac{\alpha^2 \sigma_{\mathbb{H}}^2}{N_0}} - \log \left(1 + \frac{\frac{\alpha^2 \sigma_{\mathbb{H}}^2}{N_0}}{1 + \rho + \rho \frac{\alpha^2 \sigma_{\mathbb{H}}^2}{N_0}} \right) \right] > 0 , \qquad 0 \leq \rho \leq 1 ,$$

and

$$\frac{\partial^2 E_0}{\partial \rho^2} = -\frac{\Theta}{T} \frac{\frac{\alpha^4 \sigma_{\mathbb{H}}^4}{N_0^2}}{(1+\rho) \left(1 + \rho + \frac{\alpha^2 \sigma_{\mathbb{H}}^2}{N_0} \right)^2} < 0 , \qquad 0 \leq \rho \leq 1 .$$

Hence, to fulfill (4.15) the rate has to satisfy (cf. [99, Section 8.6])

$$R \leq \frac{\partial E_0}{\partial \rho}\Big|_{\rho=0} = \frac{\Theta}{T} \left[\frac{\alpha^2 \sigma_{\mathbb{H}}^2}{N_0} - \log \left(1 + \frac{\alpha^2 \sigma_{\mathbb{H}}^2}{N_0} \right) \right] .$$

Upon insertion of (4.10), we obtain the upper bound

$$R \leq \frac{\Theta}{T} \left[\frac{PT}{\Theta} \frac{\sigma_{\mathbb{H}}^2}{N_0} - \log \left(1 + \frac{PT}{\Theta} \frac{\sigma_{\mathbb{H}}^2}{N_0} \right) \right] . \tag{4.16}$$

We finally maximize the bound (4.16) with respect to the duty cycle Θ, which amounts to taking the limit for $\Theta \to 0$. For nonzero rates, $\Theta \to 0$ implies that the bandwidth approaches infinity. Thus, the information rate of OFDM using this orthogonal signaling scheme is bounded by

$$R \leq \frac{P \sigma_{\mathbb{H}}^2}{N_0} \approx \frac{P \sigma_{\mathbb{H}_c}^2}{N_0} , \tag{4.17}$$

where we used (2.51). We note that the right-hand side of (4.17) is equal to the infinite-bandwidth capacity of time and frequency selective WSSUS channels in (4.6) which is the maximally achievable information rate. Therefore, the infinite-bandwidth system capacity of OFDM is given by

$$S \approx \frac{P \sigma_{\mathbb{H}_c}^2}{N_0} . \tag{4.18}$$

The result in (4.17) shows that OFDM communications for infinite bandwidth can obtain arbitrarily low error probability if the information rate is bounded as $R \leq S$. Hence, we have shown that OFDM can achieve capacity for infinite bandwidth and we have described a capacity-achieving codebook.

4.4 Information Rate for Constant-Modulus Signaling

The infinite-bandwidth analysis in Section 4.3 has shown that when using a specific orthogonal codebook, the system capacity for OFDM is equal to the infinite-bandwidth capacity of time and frequency selective channels. While this result may be of theoretical interest, it suffers from the fact that the specific codebook considered and the infinite-bandwidth assumption are not relevant to practical systems. Therefore, we will now consider an OFDM system with a constant-modulus (i.e., PSK) symbol alphabet and finite transmission bandwidth.

In practice, the transmission bandwidth may be on the order of several hundreds of MHz but it will never be infinite. While until recently the asymptotic infinite-bandwidth performance has been regarded as a valid characterization of the wideband regime, in [101, 102] it is shown that this is not the case. The performance limits for infinite bandwidth thus do not allow us to draw conclusions about the wideband capabilities of communication systems. In particular, orthogonal signaling schemes tend to perform very well in the infinite-bandwidth case and often approach capacity. However, in the wideband regime their spectral efficiency is poor.

To obtain insight into the performance of practical OFDM systems, we would like to know how parameters such as bandwidth, SNR, channel coherence time and coherence bandwidth, etc. influence the information rate for practical codebooks such as those corresponding to constant-modulus signaling. This type of analysis seems to be completely novel; to the author's knowledge no results about the information rate of OFDM for time and frequency selective fading channels in the wideband regime exist. The problem is that the time and frequency selective channel is rather difficult to analyze and the only results obtained up to now are the infinite-bandwidth capacity [98, 99] and certain upper and lower bounds [52–54]. These bounds show that the capacity of signaling schemes that are not peaky in time and/or frequency approach zero for infinite bandwidth. However, it is not known for which bandwidth this disastrous "overspreading" effect occurs and how it is influenced by system and/or channel parameters.

Using OFDM, we are able to answer some of these open questions. Specifically, the

information rate of OFDM can be calculated exactly if we assume constant-modulus signaling, i.e., a PSK symbol alphabet. Since we are not able to maximize this information rate with respect to the input distribution (cf. (4.2)), we cannot calculate the system capacity of OFDM for finite bandwidth. However, our result on the information rate will allow us to develop a lower bound on the system capacity of OFDM (see Section 4.5).

4.4.1 Derivation of the Information Rate

We assume that the data symbols $X_{n,k}$ are zero-mean and i.i.d. They are drawn from a constant-modulus signal alphabet with $|X_{n,k}|^2 = \sigma_x^2$. For constant-modulus signaling, we are able to calculate the information rate (4.1) exactly because we know the distribution of the channel output vector and we can exploit that $\mathrm{E}\{|X_{n,k}|^2\} = |X_{n,k}|^2 = \sigma_x^2$. We start from

$$I(\mathbf{Y};\mathbf{X}) = h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}), \tag{4.19}$$

and separately calculate $h(\mathbf{Y})$ and $h(\mathbf{Y}|\mathbf{X})$.

Calculation of $h(\mathbf{Y})$

According to (4.3),

$$\mathbf{Y} = \mathbf{S} + \mathbf{Z}, \qquad \text{with} \quad \mathbf{S} \triangleq \operatorname{diag}\{\mathbf{H}\}\,\mathbf{X} \qquad \text{and} \quad \mathbf{Z} \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I}). \tag{4.20}$$

We next look at the distribution of $\mathbf{S}$. The individual components of $\mathbf{S}$ are given by

$$\begin{aligned} S_{n,k} &= H_{n,k}\,X_{n,k} = |H_{n,k}||X_{n,k}|\,e^{j(\arg\{H_{n,k}\}+\arg\{X_{n,k}\})} \\ &= \sigma_x|H_{n,k}|\,e^{j(\arg\{H_{n,k}\}+\arg\{X_{n,k}\})}. \end{aligned}$$

Here, $|H_{n,k}|$ is Rayleigh-distributed and $\arg\{H_{n,k}\}$ is uniformly distributed. Furthermore, since $X_{n,k}$ is drawn from a constant-modulus symbol alphabet, $|X_{n,k}| = \sigma_x$ is deterministic and $\arg\{X_{n,k}\}$ has a uniform discrete distribution. Therefore, $|S_{n,k}| = |H_{n,k}||X_{n,k}| = \sigma_x|H_{n,k}|$ is also Rayleigh-distributed. The probability density function (pdf) of $\arg\{S_{n,k}\} = \arg\{H_{n,k}\} + \arg\{X_{n,k}\}$ is given by the convolution of the pdf of $\arg\{H_{n,k}\}$ with the probability mass function of $\arg\{X_{n,k}\}$ since $H_{n,k}$ and $X_{n,k}$ are independent. Hence, the distribution of $\arg\{S_{n,k}\}$ is again uniform. Combining these two results, we find that $S_{n,k} \sim \mathcal{CN}(0, \sigma_x^2\sigma_{\mathbb{H}}^2)$. Note that $S_{n,k}$ is zero-mean since $H_{n,k}$ and $X_{n,k}$ are zero-mean. The correlation function of $S_{n,k}$ is

$$\mathrm{E}\{S_{n,k}S_{n',k'}^*\} = \mathrm{E}\{H_{n,k}H_{n',k'}^*\}\,\mathrm{E}\{X_{n,k}X_{n',k'}^*\} = \sigma_{\mathbb{H}}^2\sigma_x^2\,\delta_{n-n'}\,\delta_{k-k'}, \tag{4.21}$$

since the data symbols $X_{n,k}$ are uncorrelated. Note that in $S_{n,k} = H_{n,k}X_{n,k}$, the uncorrelated data symbols "destroy" the correlations existing between the channel coefficients $H_{n,k}$. Intuitively, this "randomizes" $S_{n,k}$ and thus maximizes entropy. For the joint distribution of $\mathbf{S}$ in (4.20), we thus find that it is circularly symmetric complex Gaussian according to $\mathcal{CN}(\mathbf{0}, \sigma_{\mathbb{H}}^2\sigma_x^2\mathbf{I})$ since the individual components of $\mathbf{S}$ are i.i.d., zero-mean, and circularly symmetric complex Gaussian.

Since $\mathbf{S}$ and $\mathbf{Z}$ are independent, the distribution of $\mathbf{Y}$ is also circularly symmetric complex Gaussian according to $\mathcal{CN}(\mathbf{0}, (\sigma_{\mathbb{H}}^2\sigma_x^2 + N_0)\mathbf{I})$. Hence, the entropy of $\mathbf{Y}$ is given by

$$\begin{aligned} h(\mathbf{Y}) &= \log\left((\pi e)^{MK} \det\left[(\sigma_{\mathbb{H}}^2\sigma_x^2 + N_0)\mathbf{I}\right]\right) \\ &= MK \log\left(1 + \frac{\sigma_{\mathbb{H}}^2\sigma_x^2}{N_0}\right) + MK\log(\pi e N_0)\,. \end{aligned} \tag{4.22}$$

In view of our derivation above, we can say that, for fixed variance, $h(\mathbf{Y})$ is maximized by constant-modulus signaling with i.i.d. data symbols since this results in white complex Gaussian channel output vectors. If $\mathbf{X}$ was distributed differently (e.g., Gaussian input with variance σ_x^2), then $\mathbf{S}$ would not be Gaussian. In that case, (4.22) would be an upper bound on $h(\mathbf{Y})$ (cf. Subsection 4.5.1). Note that (4.22) is also an upper bound on $h(\mathbf{Y})$ if the data symbols are drawn from a constant-modulus symbol alphabet with magnitude σ_x but are correlated (in that case, $\mathbf{Y}$ will generally be non-Gaussian).

Calculation of $h(\mathbf{Y}|\mathbf{X})$

For the calculation of $h(\mathbf{Y}|\mathbf{X})$, we follow [53, 107] and first use the chain rule of differential entropy [169]

$$h(\mathbf{Y}|\mathbf{X}) = h(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_M|\mathbf{X}) = \sum_{n=1}^{M} h(\mathbf{Y}_n|\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{n-1})\,. \tag{4.23}$$

Let us consider the nth component of this sum, i.e., $h(\mathbf{Y}_n|\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{n-1})$. The input-output relation for the OFDM system can be written as $\mathbf{Y}_n = \operatorname{diag}\{\mathbf{X}_n\}\,\mathbf{H}_n + \mathbf{Z}_n$. For fixed input $\mathbf{X} = \mathbf{x}$, the output $\mathbf{Y}_n = \operatorname{diag}\{\mathbf{x}_n\}\,\mathbf{H}_n + \mathbf{Z}_n$ is circularly symmetric complex Gaussian. If we furthermore fix the previous channel outputs $\mathbf{Y}_{n'} = \mathbf{y}_{n'}$ for $n' = 1, \ldots, n-1$, then $\mathbf{Y}_n$ is Gaussian with conditional mean given by

$$\begin{aligned} \mathrm{E}\{\mathbf{Y}_n|\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}\} &= \mathrm{E}\{\operatorname{diag}\{\mathbf{x}_n\}\,\mathbf{H}_n + \mathbf{Z}_n|\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}\} \\ &= \operatorname{diag}\{\mathbf{x}_n\}\,\mathrm{E}\{\mathbf{H}_n|\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}\} \\ &= \operatorname{diag}\{\mathbf{x}_n\}\,\widehat{\mathbf{H}}_n\,, \end{aligned}$$

where we define

$$\widehat{\mathbf{H}}_n \triangleq \mathrm{E}\{\mathbf{H}_n|\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}\} .$$

Hence, $\widehat{\mathbf{H}}_n$ is the MMSE estimate [109, 148, 149] of the channel given the channel input and the past channel outputs, i.e., $\widehat{\mathbf{H}}_n$ is the *predicted* channel vector (cf. Chapter 3). Furthermore, the conditional covariance matrix of $\mathbf{Y}_n$ is given by

$$\begin{aligned}
&\mathrm{cov}\{\mathbf{Y}_n|\mathbf{x},\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}\} \\
&\quad = \mathrm{E}\left\{(\mathbf{Y}_n - \mathrm{diag}\{\mathbf{x}_n\}\,\widehat{\mathbf{H}}_n)(\mathbf{Y}_n - \mathrm{diag}\{\mathbf{x}_n\}\,\widehat{\mathbf{H}}_n)^H|\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}\right\} \\
&\quad = \mathrm{diag}\{\mathbf{x}_n\}\,\mathrm{E}\left\{(\mathbf{H}_n - \widehat{\mathbf{H}}_n)(\mathbf{H}_n - \widehat{\mathbf{H}}_n)^H\right\}\mathrm{diag}\{\mathbf{x}_n^*\} + \mathrm{E}\{\mathbf{Z}_n\mathbf{Z}_n^H\} \\
&\quad = \mathrm{diag}\{\mathbf{x}_n\}\,\mathbf{B}_n\,\mathrm{diag}\{\mathbf{x}_n^*\} + N_0\mathbf{I}\,, \qquad (4.24)
\end{aligned}$$

where $\mathbf{B}_n \triangleq \mathrm{E}\{(\mathbf{H}_n - \widehat{\mathbf{H}}_n)(\mathbf{H}_n - \widehat{\mathbf{H}}_n)^H\}$ is the error covariance matrix of the one-step MMSE channel predictor. An expression for $\mathbf{B}_n$ has been provided in Subsection 3.2.1 (see (3.8) with $p = 1$). More specifically, for constant-modulus data symbols and white noise, (3.8) yields

$$\mathbf{B}_n = \mathbf{R}_{\mathbb{H}}[0] - \boldsymbol{\mathcal{V}}_{\mathbb{H}}\left(\boldsymbol{\mathcal{R}}_{\mathbb{H}} + \frac{N_0}{\sigma_x^2}\mathbf{I}\right)^{-1}\boldsymbol{\mathcal{V}}_{\mathbb{H}}^H\,, \qquad (4.25)$$

which does not depend on the actual realizations $\mathbf{x}$ and $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}$. However, $\mathbf{B}_n$ depends on the memory length of the predictor via the dimensions of the matrices $\boldsymbol{\mathcal{V}}_{\mathbb{H}}$ and $\boldsymbol{\mathcal{R}}_{\mathbb{H}}$. For constant-modulus data symbols and white noise, we know from Subsection 3.2.2 that the MMSE predictor coincides with the reduced-complexity MMSE predictor; this equivalence will be exploited presently.

With (4.24), we obtain that

$$\begin{aligned}
h(\mathbf{Y}_n|\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}) &= \log\left((\pi e)^K \det\left[\mathrm{cov}\{\mathbf{Y}_n|\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}\}\right]\right) \\
&= \log\left((\pi e)^K \det\left[\mathrm{diag}\{\mathbf{x}_n\}\,\mathbf{B}_n\,\mathrm{diag}\{\mathbf{x}_n^*\} + N_0\mathbf{I}\right]\right) \qquad (4.26) \\
&= \log\left((\pi e N_0)^K \det\left[\mathbf{I} + \frac{1}{N_0}\mathrm{diag}\{\mathbf{x}_n^*\}\,\mathrm{diag}\{\mathbf{x}_n\}\,\mathbf{B}_n\right]\right) \\
&= \log\left((\pi e N_0)^K \det\left[\mathbf{I} + \frac{\sigma_x^2}{N_0}\mathbf{B}_n\right]\right)\,, \qquad (4.27)
\end{aligned}$$

where we used that $\det[\mathbf{I} + \mathbf{AB}] = \det[\mathbf{I} + \mathbf{BA}]$ for matrices $\mathbf{A}$ and $\mathbf{B}$ of proper sizes [174] and $\mathrm{diag}\{\mathbf{x}_n^*\}\,\mathrm{diag}\{\mathbf{x}_n\} = \sigma_x^2\mathbf{I}$. The expression (4.27) shows that $h(\mathbf{Y}_n|\mathbf{x}; \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1})$ does not depend on the actual realizations $\mathbf{x}$ and $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}$, and by averaging we obtain

$$h(\mathbf{Y}_n|\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{n-1}) = \log\left((\pi e N_0)^K \det\left[\mathbf{I} + \frac{\sigma_x^2}{N_0}\mathbf{B}_n\right]\right) .$$

Inserting this result into (4.23) yields

$$h(\mathbf{Y}|\mathbf{X}) = \sum_{n=1}^{M} \log\left((\pi e N_0)^K \det\left[\mathbf{I} + \frac{\sigma_x^2}{N_0}\mathbf{B}_n\right]\right) . \tag{4.28}$$

In Subsection 3.2.3, we have shown that $\mathbf{B}_n$ can be decomposed as[1] (see (3.31) and note that we introduce the subscript/superscript n to make the dependence of the error covariance matrix $\mathbf{B}_n$ and the tap prediction error $\epsilon_{\text{min},l}^{(n)}$ on the predictor length explicit)

$$\mathbf{B}_n = \mathbf{F}\,\text{diag}\left\{\epsilon_{\text{min},0}^{(n)}, \epsilon_{\text{min},1}^{(n)}, \ldots, \epsilon_{\text{min},L}^{(n)}, 0, \ldots, 0\right\}\mathbf{F}^H ,$$

where $\epsilon_{\text{min},l}^{(n)}$ is given by (3.33) with $\gamma = \sigma_x^2/N_0$. Therefore, we obtain for (4.28)

$$\begin{aligned} h(\mathbf{Y}|\mathbf{X}) &= \sum_{n=1}^{M} \log\left((\pi e N_0)^K \prod_{l=0}^{L}\left(1 + K\frac{\sigma_x^2\,\epsilon_{\text{min},l}^{(n)}}{N_0}\right)\right) \\ &= \sum_{n=1}^{M}\sum_{l=0}^{L} \log\left(1 + K\frac{\sigma_x^2\,\epsilon_{\text{min},l}^{(n)}}{N_0}\right) + MK\log(\pi e N_0) . \end{aligned} \tag{4.29}$$

Information Rate for Constant-Modulus Signaling

Inserting (4.22) and (4.29) into (4.19) results in the following expression for the mutual information of channel input and output:

$$I(\mathbf{Y};\mathbf{X}) = MK\log\left(1 + \frac{\sigma_{\mathbb{H}}^2\sigma_x^2}{N_0}\right) - \sum_{n=1}^{M}\sum_{l=0}^{L}\log\left(1 + K\frac{\sigma_x^2\,\epsilon_{\text{min},l}^{(n)}}{N_0}\right) .$$

With (4.1), we therefore obtain for the ergodic information rate of OFDM systems with constant-modulus signaling

$$\begin{aligned} R_{\text{CM}} &= \lim_{M\to\infty}\frac{1}{MT} I(\mathbf{Y};\mathbf{X}) \\ &= \frac{K}{T}\log\left(1 + \frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\right) - \lim_{M\to\infty}\frac{1}{MT}\sum_{n=1}^{M}\sum_{l=0}^{L}\log\left(1 + K\frac{\sigma_x^2\,\epsilon_{\text{min},l}^{(n)}}{N_0}\right) \\ &= \frac{K}{T}\log\left(1 + \frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\right) - \frac{1}{T}\sum_{l=0}^{L}\log\left(1 + K\frac{\sigma_x^2\,\epsilon_{\text{min},l}^{(\infty)}}{N_0}\right) , \end{aligned} \tag{4.30}$$

[1] This is the eigendecomposition of the circulant matrix $\mathbf{B}_n$. The $L+1$ nonzero eigenvalues are $K\epsilon_{\text{min},l}^{(n)}$ where the factor K results from the fact that the DFT matrix $\mathbf{F}$ satisfies $\mathbf{F}^H\mathbf{F} = K\mathbf{I}$.

where $\epsilon_{\mathrm{min},l}^{(\infty)}$ denotes the one-step prediction MMSE for the lth delay tap using an infinite-length predictor. This specific predictor was analyzed in Subsection 3.2.4, where it was shown that $\epsilon_{\mathrm{min},l}^{(\infty)}$ is given by (3.44) with $\gamma^2 = N_0/\sigma_x^2$. Inserting (3.44) into (4.30) yields

$$R_{\mathrm{CM}} = \frac{K}{T} \log\left(1 + \frac{\sigma_x^2 \sigma_{\mathbb{H}}^2}{N_0}\right) - \frac{1}{T} \sum_{l=0}^{L} \int_{-1/2}^{1/2} \log\left(1 + K \frac{\sigma_x^2}{N_0} C_{\mathbb{H}}(l, \varphi)\right) d\varphi \,. \tag{4.31}$$

In the wideband regime, we can approximate $C_{\mathbb{H}}(l, \varphi)$ by (2.49). Furthermore, assuming constant subcarrier spacing F and constant transmit power P, we have $K = B/F$ and $\sigma_x^2 = PT/K$. Inserting this into (4.31) together with $\sigma_{\mathbb{H}}^2 \approx \sigma_{\mathbb{H}_{\mathrm{c}}}^2$ (cf. (2.51)) yields

$$R_{\mathrm{CM}} \approx \frac{B}{TF} \log\left(1 + \frac{TF}{B} \frac{P \sigma_{\mathbb{H}_{\mathrm{c}}}^2}{N_0}\right) - \sum_{l=0}^{L} \int_{\nu} \log\left(1 + \frac{P}{BN_0} C_{\mathbb{H}_{\mathrm{c}}}(lT_s, \nu)\right) d\nu \,,$$

where we also used that $T = NT_s = N/B$ and $\nu = \varphi/T$. Finally, noting that for large B the summation over the channel taps can be approximated by an integration over all delays, we finally obtain for the ergodic information rate of OFDM using constant-modulus signaling

$$R_{\mathrm{CM}} \approx \frac{B}{TF} \log\left(1 + \frac{TF}{B} \frac{P \sigma_{\mathbb{H}_{\mathrm{c}}}^2}{N_0}\right) - B \int_{\tau} \int_{\nu} \log\left(1 + \frac{P}{BN_0} C_{\mathbb{H}_{\mathrm{c}}}(\tau, \nu)\right) d\tau \, d\nu \,. \tag{4.32}$$

The first term on the right-hand side of (4.32) is equal to the AWGN channel capacity for effective bandwidth $B/(TF)$. The second term is related to the channel prediction error, and it is always negative since $C_{\mathbb{H}_{\mathrm{c}}}(\tau, \nu) \geq 0$. It can hence be interpreted as a loss in information rate due to the limited predictability of the channel. As long as the channel can be well predicted, the second term is small compared to the first one and thus the information rate loss is small. This is corroborated by our numerical results in Section 4.7.

Based on the derivation leading to (4.32), it can be conjectured that the information rate in (4.32) can be achieved by an OFDM communication system that uses constant-modulus signaling with independent data symbols. No pilot symbols are inserted in the transmit data stream, and at the receiver, CSI is obtained using a one-step, infinite-length, decision-directed MMSE channel predictor (cf. Subsection 3.2.4). This receiver corresponds to the receiver described in Subsection 3.4.1 and depicted in Figure 3.5(b), with prediction horizon $p = 1$. Let us recall some issues related to the practical application of such a system. For startup of the channel predictor, it is necessary to transmit some training data at the beginning of the transmission of a data packet. Furthermore, it will not be possible to implement an infinite-length predictor filter. However, for finite channel coherence times, the infinite-length MMSE channel predictor will reduce to the finite-length MMSE channel predictor which, for constant-modulus symbols, can be efficiently implemented as described in Subsection 3.2.3 and shown in Figure 3.4.

4.4.2 Alternative Derivation of the Information Rate

We now present an alternative derivation of (4.31). Applying the chain rule of mutual information [169] to $I(\mathbf{Y};\mathbf{X},\mathbf{H})$, we obtain the decomposition

$$I(\mathbf{Y};\mathbf{X}) = I(\mathbf{Y};\mathbf{X},\mathbf{H}) - I(\mathbf{Y};\mathbf{H}|\mathbf{X}). \tag{4.33}$$

Calculation of $I(\mathbf{Y};\mathbf{X},\mathbf{H})$

The first component of (4.33) is

$$I(\mathbf{Y};\mathbf{X},\mathbf{H}) = h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X},\mathbf{H}).$$

We use (4.22) for $h(\mathbf{Y})$; furthermore, due to $\mathbf{Y} = \operatorname{diag}\{\mathbf{H}\}\,\mathbf{X} + \mathbf{Z}$, we have

$$\begin{aligned} h(\mathbf{Y}|\mathbf{X},\mathbf{H}) &= \mathrm{E}_X\{\mathrm{E}_H\{h(\mathbf{Y}|\mathbf{x},\mathbf{h})\}\} = \mathrm{E}_X\{\mathrm{E}_H\{h(\operatorname{diag}\{\mathbf{h}\}\,\mathbf{x}+\mathbf{Z}|\mathbf{x},\mathbf{h})\}\} \\ &= \mathrm{E}_X\{\mathrm{E}_H\{h(\mathbf{Z}|\mathbf{x},\mathbf{h})\} = \mathrm{E}_X\{\mathrm{E}_H\{h(\mathbf{Z})\} = MK\log(\pi e N_0). \end{aligned} \tag{4.34}$$

We thus obtain

$$I(\mathbf{Y};\mathbf{X},\mathbf{H}) = MK\log\left(1+\frac{\sigma_{\mathbb{H}}^2\sigma_x^2}{N_0}\right). \tag{4.35}$$

Calculation of $I(\mathbf{Y};\mathbf{H}|\mathbf{X})$

For fixed channel input $\mathbf{X} = \mathbf{x}$, the channel output $\mathbf{Y} = \operatorname{diag}\{\mathbf{x}\}\,\mathbf{H} + \mathbf{Z}$ is circularly symmetric complex Gaussian with covariance matrix

$$\begin{aligned} \operatorname{cov}\{\mathbf{Y}|\mathbf{x}\} &= \mathrm{E}\left\{(\operatorname{diag}\{\mathbf{x}\}\,\mathbf{H}+\mathbf{Z})(\operatorname{diag}\{\mathbf{x}\}\,\mathbf{H}+\mathbf{Z})^H\right\} \\ &= \operatorname{diag}\{\mathbf{x}\}\,\boldsymbol{\mathcal{R}}_{\mathbb{H}}\operatorname{diag}\{\mathbf{x}^*\} + N_0\mathbf{I}. \end{aligned} \tag{4.36}$$

Here, $\boldsymbol{\mathcal{R}}_{\mathbb{H}}$ is an $MK \times MK$ block-Toeplitz correlation matrix with first row $[\mathbf{R}_{\mathbb{H}}[0]\ \cdots\ \mathbf{R}_{\mathbb{H}}[M-1]]$, with the $K \times K$ correlation matrix $\mathbf{R}_{\mathbb{H}}[m]$ defined in (3.6). Using (4.34) and (4.36), we therefore obtain

$$\begin{aligned} I(\mathbf{Y};\mathbf{H}|\mathbf{X}) &= h(\mathbf{Y}|\mathbf{X}) - h(\mathbf{Y}|\mathbf{H},\mathbf{X}) = \mathrm{E}_X\{h(\mathbf{Y}|\mathbf{x})\} - MK\log(\pi e N_0) \\ &= \mathrm{E}_X\left\{\log\left((\pi e)^{MK}\det\left[\operatorname{diag}\{\mathbf{x}\}\,\boldsymbol{\mathcal{R}}_{\mathbb{H}}\operatorname{diag}\{\mathbf{x}^*\} + N_0\mathbf{I}\right]\right)\right\} - MK\log(\pi e N_0) \\ &= \mathrm{E}_X\left\{\log\left(\det\left[\mathbf{I} + \frac{1}{N_0}\operatorname{diag}\{\mathbf{x}\}\,\boldsymbol{\mathcal{R}}_{\mathbb{H}}\operatorname{diag}\{\mathbf{x}^*\}\right]\right)\right\} \\ &= \mathrm{E}_X\left\{\log\left(\det\left[\mathbf{I} + \frac{1}{N_0}\operatorname{diag}\{\mathbf{x}^*\}\operatorname{diag}\{\mathbf{x}\}\,\boldsymbol{\mathcal{R}}_{\mathbb{H}}\right]\right)\right\} \end{aligned}$$

$$= \mathrm{E}_X\left\{\log\left(\det\left[\mathbf{I}+\frac{\sigma_x^2}{N_0}\boldsymbol{\mathcal{R}}_{\mathbb{H}}\right]\right)\right\} = \log\left(\det\left[\mathbf{I}+\frac{\sigma_x^2}{N_0}\boldsymbol{\mathcal{R}}_{\mathbb{H}}\right]\right),$$

where we exploited that the input is constant-modulus, i.e., $\mathrm{diag}\{\mathbf{x}^*\}\,\mathrm{diag}\{\mathbf{x}\} = \sigma_x^2\mathbf{I}$. Denoting the eigenvalues of $\boldsymbol{\mathcal{R}}_{\mathbb{H}}$ by $\lambda_i\{\boldsymbol{\mathcal{R}}_{\mathbb{H}}\}$, we further have

$$I(\mathbf{Y};\mathbf{H}|\mathbf{X}) = \sum_{i=0}^{MK-1}\log\left(1+\frac{\sigma_x^2}{N_0}\lambda_i\{\boldsymbol{\mathcal{R}}_{\mathbb{H}}\}\right). \tag{4.37}$$

Information Rate for Constant-Modulus Signaling

With (4.35) and (4.37), we obtain for (4.33)

$$I(\mathbf{Y};\mathbf{X}) = MK\log\left(1+\frac{\sigma_{\mathbb{H}}^2\sigma_x^2}{N_0}\right) - \sum_{i=0}^{MK-1}\log\left(1+\frac{\sigma_x^2}{N_0}\lambda_i\{\boldsymbol{\mathcal{R}}_{\mathbb{H}}\}\right).$$

Inserting this into (4.1), we furthermore obtain for the ergodic information rate

$$R_{\mathrm{CM}} = \frac{K}{T}\log\left(1+\frac{\sigma_{\mathbb{H}}^2\sigma_x^2}{N_0}\right) - \lim_{M\to\infty}\frac{1}{MT}\sum_{i=0}^{MK-1}\log\left(1+\frac{\sigma_x^2}{N_0}\lambda_i\{\boldsymbol{\mathcal{R}}_{\mathbb{H}}\}\right). \tag{4.38}$$

Since $\boldsymbol{\mathcal{R}}_{\mathbb{H}}$ in (4.38) is block-Toeplitz, we can apply Theorem 3 in [175], which is an extension of Szegö's theorem on the asymptotic eigenvalue distribution of Toeplitz matrices. This yields

$$R_{\mathrm{CM}} = \frac{K}{T}\log\left(1+\frac{\sigma_{\mathbb{H}}^2\sigma_x^2}{N_0}\right) - \frac{1}{T}\sum_{k=0}^{K-1}\int_{-1/2}^{1/2}\log\left(1+\frac{\sigma_x^2}{N_0}\lambda_k\{\mathbf{B}_{\mathbb{H}}(\varphi)\}\right)d\varphi, \tag{4.39}$$

where the $K\times K$ Toeplitz matrix $\mathbf{B}_{\mathbb{H}}(\varphi)$ is given by

$$\begin{aligned}\mathbf{B}_{\mathbb{H}}(\varphi) &= \sum_{n=-\infty}^{\infty}\mathbf{R}_{\mathbb{H}}[n]\,e^{-j2\pi n\varphi}\\ &= \sum_{n=-\infty}^{\infty}\mathbf{F}\,\mathrm{diag}\{D_{\mathbb{H}}[n,0],\ldots,D_{\mathbb{H}}[n,L],0,\ldots,0\}\,\mathbf{F}^H e^{-j2\pi n\varphi}\\ &= \mathbf{F}\,\mathrm{diag}\{C_{\mathbb{H}}(0,\varphi),\ldots,C_{\mathbb{H}}(L,\varphi),0,\ldots,0\}\,\mathbf{F}^H.\end{aligned} \tag{4.40}$$

Here, we used the eigenvalue decomposition of $\mathbf{R}_{\mathbb{H}}[n]$ in (3.26) and the Fourier relationship between the time-delay correlation function $D_{\mathbb{H}}[n,l]$ and the scattering function $C_{\mathbb{H}}(l,\varphi)$ in (2.48). Because the DFT matrix $\mathbf{F}$ is orthogonal up to a factor, i.e., $\mathbf{F}\mathbf{F}^H = \mathbf{F}^H\mathbf{F} = K\mathbf{I}$, (4.40) is recognized as the eigenvalue decomposition of $\mathbf{B}_{\mathbb{H}}(\varphi)$, the eigenvalues being

$$\lambda_l\{\mathbf{B}_{\mathbb{H}}(\varphi)\} = \begin{cases} KC_{\mathbb{H}}(l,\varphi), & l=0,1,\ldots,L,\\ 0, & \text{else}.\end{cases}$$

Hence, (4.39) becomes

$$R_{\mathrm{CM}} = \frac{K}{T}\log\left(1+\frac{\sigma_{\mathbb{H}}^2\sigma_x^2}{N_0}\right) - \frac{1}{T}\sum_{l=0}^{L}\int_{-1/2}^{1/2}\log\left(1+K\frac{\sigma_x^2}{N_0}C_{\mathbb{H}}(l,\varphi)\right)d\varphi\,,$$

which is identical to our previous result (4.31).

Hence, we have found an alternative derivation of the information rate of OFDM with constant-modulus signaling that is based on Szegö's Theorem. In contrast to our previous derivation in Subsection 4.4.1, this derivation does not directly reveal that the second term in (4.31) is related with the MMSE of the one-step infinite-length channel predictor.

4.4.3 Dependence of Information Rate on Bandwidth

With our expression for R_{CM} in (4.32), we can now assess how the information rate of OFDM with constant-modulus signaling depends on the bandwidth. We are specifically interested in the asymptotic limit and the wideband regime.

Asymptotic Limit

Using $\sigma_{\mathbb{H}_c}^2 = \int_\tau\int_\nu C_{\mathbb{H}_c}(\tau,\nu)\,d\tau\,d\nu$, it follows from (4.32) that in the limit of infinite bandwidth the OFDM information rate approaches zero:

$$\begin{aligned}\lim_{B\to\infty} R_{\mathrm{CM}} &\approx \lim_{B\to\infty}\left\{\frac{B}{TF}\log\left(1+\frac{TF}{B}\frac{P\sigma_{\mathbb{H}_c}^2}{N_0}\right) - B\int_\tau\int_\nu\log\left(1+\frac{P}{BN_0}C_{\mathbb{H}_c}(\tau,\nu)\right)d\tau\,d\nu\right\}\\ &= \frac{P\sigma_{\mathbb{H}_c}^2}{N_0} - \frac{P}{N_0}\int_\tau\int_\nu C_{\mathbb{H}_c}(\tau,\nu)\,d\tau d\nu = 0\,.\end{aligned}$$

Hence, as for CDMA [52–54], the wideband capacity of OFDM systems with constant-modulus signaling approaches zero. The reason for this asymptotic behavior is the fact that with constant-modulus signaling, the transmit power is uniformly spread over all time-frequency slots, resulting in signals that are not peaky in time or frequency. We emphasize, however, that in contrast to the results in [52–54], (4.32) allows to assess the behavior of the information rate in the wideband regime (see also the simulations in Section 4.7).

Wideband Regime

The information rate R_{CM} in (4.32) increases with B up to a maximum

$$R_{\max} \triangleq \max_B R_{\mathrm{CM}}\,,$$

that is obtained for the "critical bandwidth"

$$B_{\text{crit}} \triangleq \arg\max_{B} R_{\text{CM}} . \tag{4.41}$$

For $B < B_{\text{crit}}$, numerical results show that R_{CM} is close to the AWGN capacity (see Subsection 4.7.1). For $B > B_{\text{crit}}$, R_{CM} decreases to zero; we call this effect "overspreading." Overspreading means that the bandwidth is too large and the transmit power is too low to sufficiently "illuminate the channel." The expression for R_{CM} in (4.32) only allows a numerical calculation of $R_{\max}$ and B_{crit}.

4.4.4 Dependence of Information Rate on Scattering Function

Given a certain bandwidth, the information rate R_{CM} in (4.32) depends on the shape and spread of the scattering function $C_{\mathbb{H}_c}(\tau, \nu)$.

Impact of Spread of Scattering Function

For fixed $\sigma^2_{\mathbb{H}_c}$ and fixed shape of the scattering function, the information rate (4.32) decreases for increasing channel spread $\tau_{\max}\nu_{\max}$. This can easily be seen by replacing $C_{\mathbb{H}_c}(\tau, \nu)$ with $\frac{1}{ab}C_{\mathbb{H}_c}(\tau, \nu)\left(\frac{\tau}{a}, \frac{\nu}{b}\right)$ in (4.32) and noting that (4.32) decreases with increasing dilation factor product ab.

Impact of Shape of Scattering Function

To gain insight in how the information rate (4.32) depends on the shape of the scattering function, we now derive the worst-case scattering function $C^{\dagger}_{\mathbb{H}_c}(\tau, \nu)$ that minimizes (4.32) within the class of scattering functions with given path loss $\sigma^2_{\mathbb{H}_c}$ and given support region $[0, \tau_{\max}] \times [-\nu_{\max}/2, \nu_{\max}/2]$. To this end, we define a parametric representation of scattering functions as

$$C(\tau, \nu) = \sum_{p=0}^{2P} \sum_{q=-Q}^{Q} \alpha_{p,q} I(\tau - p\tau_0, \nu - q\nu_0) , \tag{4.42}$$

with the set of coefficients $\{\alpha_{p,q}\}$ and the indicator function

$$I(\tau, \nu) \triangleq \begin{cases} 1, & (\tau, \nu) \in [0, \tau_0] \times [-\nu_0/2, \nu_0/2] , \\ 0, & \text{else} , \end{cases}$$

where $\tau_0 = \tau_{\max}/(2P+1)$ and $\nu_0 = \nu_{\max}/(2Q+1)$. Note that in the limit $P \to \infty$ and $Q \to \infty$, this model is capable of approximating any finite-support scattering function with

arbitrary accuracy. With (4.42), the path loss is given by

$$\sigma_{\mathbb{H}}^2 = \int_\tau \int_\nu C(\tau,\nu)\, d\tau d\nu = \tau_0\nu_0 \sum_{p=0}^{2P} \sum_{q=-Q}^{Q} \alpha_{p,q}\,,$$

and the rate (4.32) is given by

$$\begin{aligned} R_{\text{CM}} &\approx \frac{B}{TF} \log\left(1 + \frac{TF}{B}\frac{P\sigma_{\mathbb{H}_\text{c}}^2}{N_0}\right) - B\int_\tau\int_\nu \log\left(1 + \frac{P}{BN_0}\, C(\tau,\nu)\right) d\tau\, d\nu \\ &= \frac{B}{TF} \log\left(1 + \frac{TF}{B}\frac{P\sigma_{\mathbb{H}_\text{c}}^2}{N_0}\right) - \tau_0\nu_0 B \sum_{p=0}^{2P} \sum_{q=-Q}^{Q} \log\left(1 + \frac{P}{BN_0}\,\alpha_{p,q}\right). \end{aligned} \tag{4.43}$$

Hence, the worst-case scattering function is obtained if $\sum_{p=0}^{2P}\sum_{q=-Q}^{Q} \log\left(1 + \frac{P}{BN_0}\,\alpha_{p,q}\right)$ is maximized for a given path loss $\sigma_{\mathbb{H}_\text{c}}^2$. To calculate the worst-case coefficients $\{\alpha_{p,q}^\dagger\}$, we define the cost function

$$J \triangleq \sum_{p=0}^{2P} \sum_{q=-Q}^{Q} \log\left(1 + \frac{P}{BN_0}\,\alpha_{p,q}\right) - \lambda\left(\tau_0\nu_0 \sum_{p=0}^{2P} \sum_{q=-Q}^{Q} \alpha_{p,q} - \sigma_{\mathbb{H}_\text{c}}^2\right), \tag{4.44}$$

where λ is a Lagrange multiplier. Solving $\partial J/\partial\alpha_{p,q} = 0$ yields

$$\alpha_{p,q}^\dagger = B\left(\frac{1}{\lambda} - \frac{N_0}{P}\right), \tag{4.45}$$

which does not depend on p or q. Inserting (4.45) into (4.44) and solving $\partial J/\partial\lambda = 0$ determines the Lagrange multiplier as

$$\frac{1}{\lambda} = \frac{N_0}{P} + \frac{\sigma_{\mathbb{H}_\text{c}}^2}{\tau_{\max}\nu_{\max}B}\,. \tag{4.46}$$

Inserting (4.46) into (4.45), we finally obtain the worst-case coefficients as

$$\alpha_{p,q}^\dagger = \frac{\sigma_{\mathbb{H}_\text{c}}^2}{\tau_{\max}\nu_{\max}}\,.$$

Letting $P \to \infty$ and $Q \to \infty$, it is seen that the scattering function $C^\dagger(\tau,\nu)$ minimizing (4.32) (maximizing the penalty term) among all scattering functions with path loss $\sigma_{\mathbb{H}_\text{c}}^2$ and support area $[0,\tau_{\max}] \times [-\nu_{\max}/2, \nu_{\max}/2]$ is the uniform scattering function given by

$$C^\dagger(\tau,\nu) = \begin{cases} \frac{\sigma_{\mathbb{H}_\text{c}}^2}{\tau_{\max}\nu_{\max}}, & (\tau,\nu) \in [0,\tau_{\max}] \times [-\nu_{\max}/2, \nu_{\max}/2]\,, \\ 0, & \text{else}\,. \end{cases} \tag{4.47}$$

This result is not surprising as it simply says that the channel prediction error is maximized when the random process describing the channel has a flat spectrum.

Insertion of (4.47) into (4.32) results in the information rate

$$R_{\mathrm{CM}}^{\dagger} \approx \frac{B}{TF} \log\left(1 + \frac{TF}{B}\frac{P\sigma_{\mathbb{H}_c}^2}{N_0}\right) - B\tau_{\max}\nu_{\max} \log\left(1 + \frac{P\sigma_{\mathbb{H}_c}^2}{B\tau_{\max}\nu_{\max}N_0}\right), \tag{4.48}$$

which is the minimum information rate for given path loss $\sigma_{\mathbb{H}_c}^2$ and given scattering function support region $[0, \tau_{\max}] \times [-\nu_{\max}/2, \nu_{\max}/2]$. Note that (4.48) depends on the channel spread, i.e., the product $\tau_{\max}\nu_{\max}$, but not on $\tau_{\max}$ and $\nu_{\max}$ individually.

Information Rate in Specular Scattering

For specular scattering, the scattering function consists of discrete components as defined by (3.46). In Subsection 3.2.5, we have shown that for this special case $\epsilon_{\min,l}^{(\infty)}$ is zero (cf. (3.53)). Inserting this result into (4.30) yields for the information rate

$$R_{\mathrm{CM}} = \frac{K}{T} \log\left(1 + \frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\right), \tag{4.49}$$

i.e., the "penalty term" due to channel uncertainty is zero. For constant subcarrier spacing F and constant transmit power P, we have $K = B/F$ and $\sigma_x^2 = PT/K$ and hence obtain for (4.49)

$$R_{\mathrm{CM}} = \frac{B}{TF} \log\left(1 + \frac{TF}{B}\frac{P\sigma_{\mathbb{H}_c}^2}{N_0}\right).$$

where we used (2.51). For $B \to \infty$, we thus obtain $R_{\mathrm{CM}} = P\sigma_{\mathbb{H}_c}^2/N_0$, i.e., the infinite-bandwidth information rate is nonzero. This is because the receiver perfectly knows the channel's delays and Doppler shifts.

4.4.5 Information Rate and Diversity

We have shown that the worst-case scattering function (maximizing the penalty on the information rate for an OFDM system transmitting over an unknown random time and frequency selective fading channel) is the flat scattering function in (4.47), and the maximum penalty is equal to the second term in (4.48)

$$B\tau_{\max}\nu_{\max} \log\left(1 + \frac{P\sigma_{\mathbb{H}_c}^2}{B\tau_{\max}\nu_{\max}N_0}\right).$$

We now recall from Subsection 2.2.1 that $\tau_{\max}$ and $\nu_{\max}$ are approximately equal to the reciprocal of the coherence bandwidth $B_{\mathbb{H}_c}$ and the coherence time $T_{\mathbb{H}_c}$, respectively. Hence,

the maximum penalty is approximately equal to

$$\frac{B}{T_{\mathbb{H}_c} B_{\mathbb{H}_c}} \log\left(1 + \frac{P\sigma^2_{\mathbb{H}_c}}{BN_0} T_{\mathbb{H}_c} B_{\mathbb{H}_c}\right).$$

From this expression, we see how diversity offered by the channel relates to obtainable information rate. From the diversity point of view, we would like to have a channel with small coherence time/bandwidth. However, in this case the penalty on the information rate will be large. Hence, high diversity and high information rate impose conflicting requirements on the channel parameters.

4.4.6 Impact of Information Spreading

(Pre)coding may be used to spread the transmitted information over time and/or frequency and thus introduce dependencies between the data symbols $X_{n,k}$. To obtain an understanding of how information spreading can affect the information rate R_{CM} in (4.32), we again look at the derivation in Subsection 4.4.1. For (4.22), it was essential that the received sequence $Y_{n,k}$ was i.i.d. with circularly symmetric complex Gaussian distribution. In our derivation, this was ensured by independent data symbols drawn from a constant-modulus alphabet. If information is spread over time and/or frequency at the transmitter, correlations between the $X_{n,k}$ are introduced. This leads to correlations in the received sequence $Y_{n,k}$ unless the time or frequency distance between correlated data symbols is larger than the channel coherence time or channel coherence bandwidth, respectively (cf. (4.21)). For correlated $Y_{n,k}$, (4.22) is an upper bound and thus the rate in (4.32) will generally be reduced. Hence, the channel has to decorrelate between two correlated data symbols or the information rate decreases. This requirement conforms with the aim of spreading to exploit diversity.

4.5 Bounds on System Capacity

While we cannot calculate the system capacity (4.2) exactly, we can derive an upper bound as well as a lower bound on system capacity. Our numerical results in Section 4.7 will demonstrate that the upper and lower bounds practically coincide for most bandwidths of practical interest. Only for bandwidths in the range of B_{crit} (see (4.41)) and above there is some deviation between the two bounds.

4.5.1 Upper Bound on System Capacity

To obtain an upper bound, we note that the mutual information for no CSI at the receiver will be upper bounded by the mutual information for perfect CSI at the receiver

$$\begin{aligned} I(\mathbf{Y};\mathbf{X}|\mathbf{H}) &= I(\mathbf{X};\mathbf{Y}|\mathbf{H}) = h(\mathbf{X}|\mathbf{H}) - h(\mathbf{X}|\mathbf{Y},\mathbf{H}) = h(\mathbf{X}) - h(\mathbf{X}|\mathbf{Y},\mathbf{H}) \\ &\geq h(\mathbf{X}) - h(\mathbf{X}|\mathbf{Y}) = I(\mathbf{X};\mathbf{Y}) = I(\mathbf{Y};\mathbf{X}), \end{aligned}$$

since $\mathbf{X}$ and $\mathbf{H}$ are independent and conditioning reduces entropy. Hence, $I(\mathbf{Y};\mathbf{X}) \leq I(\mathbf{Y};\mathbf{X}|\mathbf{H})$ and thus the system capacity without CSI at the receiver will be upper bounded by the mutual information for perfect CSI at the receiver, i.e., $S \leq S_{\text{CSI}}$. With (4.5), we obtain the desired upper bound on the system capacity of OFDM:

$$S \leq \frac{K}{T} \log\left(1 + \frac{\sigma_x^2 \sigma_{\mathbb{H}}^2}{N_0}\right). \tag{4.50}$$

Hence, the system capacity of OFDM is upper bounded by the capacity of K independent parallel AWGN channels, each operating at SNR $\sigma_x^2\sigma_{\mathbb{H}}^2/N_0$ and symbol rate T.

We next assume constant transmit power P and fixed subcarrier spacing F. We have $K = B/F$ and $\sigma_x^2 = PT/K = PTF/B$. Inserting this into (4.50) together with (2.51), we obtain

$$S \leq \frac{B}{TF} \log\left(1 + \frac{TF}{B}\frac{P\sigma_{\mathbb{H}_c}^2}{N_0}\right). \tag{4.51}$$

The only OFDM system parameter involved in (4.51) is the *product* of the OFDM symbol period T and the subcarrier separation F. For $TF = 1$, the bound (4.51) is the capacity of an AWGN channel with bandwidth B operating at an SNR equal to $P\sigma_{\mathbb{H}_c}^2/(BN_0)$. In OFDM systems, however, TF has to be larger than one; this amounts to introducing redundancy and design freedom into the system but obviously reduces the potential system capacity. The redundancy can e.g. be used for inserting a guard period or a cyclic prefix between consecutive OFDM symbols. However, TF is typically very close to one (e.g., in DVB-T a value of 1.03 is possible) so that the upper bound (4.51) is practically equal to the AWGN capacity. Furthermore, for any TF, the bound (4.51) can be interpreted as the capacity of an AWGN channel with "effective bandwidth" $B/(TF)$.

An interesting special case of the bound (4.51) is obtained for the infinite-bandwidth limit $B \to \infty$. Here, the bound is given by

$$\lim_{B\to\infty} \frac{B}{TF} \log\left(1 + \frac{TF}{B}\frac{P\sigma_{\mathbb{H}_c}^2}{N_0}\right) = \frac{P\sigma_{\mathbb{H}_c}^2}{N_0},$$

which is equal to the infinite-bandwidth capacity of time and frequency selective channels [99, Section 8.6]. Note that for this case, the bound does not depend on the OFDM system parameters T and F. In Section 4.3, we have shown that using a simple orthogonal codebook, this infinite-bandwidth system capacity can indeed be obtained by OFDM. Hence, for $B \to \infty$ the bound (4.51) is tight.

Furthermore, the bound (4.51) is identical to the first term of our expression (4.32) for the information rate for constant-modulus signaling, R_{CM}. R_{CM} deviates from the bound by the second term in (4.32) that is due to channel uncertainty at the receiver. For practical bandwidths, however, the second term is small compared to the first one, and thus R_{CM} is nearly equal to the bound (4.51). Only if the information is overspread, i.e. for $B > B_{\mathrm{crit}}$, R_{CM} deviates significantly from the bound.

4.5.2 Lower Bound on System Capacity

From the definition of the information rate in (4.1) and the system capacity in (4.2), we obtain the lower bound $S \geq R$ where R may be calculated using some specific codebook. Hence, we can use the information rate obtained with constant-modulus signaling R_{CM} in (4.32) to formulate a lower bound on the system capacity, i.e., $S \geq R_{\mathrm{CM}}$. However, for $B \geq B_{\mathrm{crit}}$, a tighter bound can be easily found by the information rate of an OFDM system with constant-modulus signaling that does not use all the available bandwidth but only bandwidth B_{crit}, i.e., only $K' = B_{\mathrm{crit}}/F$ adjacent subcarriers are used. This OFDM system would achieve $R_{\max} = \max_B R_{\mathrm{CM}} = R_{\mathrm{CM}}|_{B=B_{\mathrm{crit}}}$. Hence, we obtain the lower bound

$$S \geq \begin{cases} R_{\mathrm{CM}}\,, & B < B_{\mathrm{crit}}\,, \\ R_{\max}\,, & B \geq B_{\mathrm{crit}}\,. \end{cases}$$

Moreover, since in Subsection 4.4.4 we have shown that $R_{\mathrm{CM}} \geq R_{\mathrm{CM}}^{\dagger}$ with $R_{\mathrm{CM}}^{\dagger}$ given by (4.48), we also obtain the looser lower bound

$$S \geq \begin{cases} R_{\mathrm{CM}}^{\dagger}\,, & B < B_{\mathrm{crit}}\,, \\ R_{\max}^{\dagger}\,, & B \geq B_{\mathrm{crit}}\,. \end{cases} \tag{4.52}$$

From these lower bounds, it is clear that the OFDM system capacity does not approach zero at infinite bandwidth but is at least $R_{\max}^{\dagger}$.

4.5.3 Relation to Telatar and Tse's Result

As mentioned in Subsection 4.2.2, the following lower and upper bounds on the information rate for spread-spectrum-like signaling were given by Telatar and Tse [52]:

$$C_{\text{AWGN}}\left[1-\frac{\tilde{L}}{\tilde{L}_{\text{crit}}}\log\left(1+\frac{\tilde{L}_{\text{crit}}}{\tilde{L}}\right)\right] \leq R \leq C_{\text{AWGN}}\frac{\tilde{L}_{\text{crit}}}{\tilde{L}},$$

where $C_{\text{AWGN}} = P/N_0$ is the asymptotic AWGN capacity, $\tilde{L}_{\text{crit}} \triangleq PT_c/N_0$ is a critical number of delay taps, T_c is the coherence time of the channel (note that in [52] the coherence time is differently defined than $T_{\mathbb{H}_c}$ in Subsection 2.2.1), and $\tilde{L}$ is the number of resolved delay taps that increases linearly with bandwidth B. Furthermore, the path loss of the channel was normalized to one. The upper bound of [52] is meaningful if it is smaller than C_{AWGN} which is the case for $\tilde{L} \leq \tilde{L}_{\text{crit}}$.

With $\sigma^2_{\mathbb{H}_c} = 1$, our upper bound (4.51) is given by

$$S \leq \frac{B}{TF}\log\left(1+\frac{TF}{B}\frac{P}{N_0}\right) \leq \frac{P}{N_0} = C_{\text{AWGN}}.$$

Hence, our upper bound is different from that of [52].

For $B \leq B_{\text{crit}}$ and $\sigma^2_{\mathbb{H}} = 1$, our lower bound in (4.52) is equal to

$$S \geq \frac{B}{TF}\log\left(1+\frac{TF}{B}\frac{P}{N_0}\right) - B\tau_{\max}\nu_{\max}\log\left(1+\frac{P}{B\tau_{\max}\nu_{\max}N_0}\right),$$

where we used (4.48). Bounding the first term by means of $\log(1+x) \geq x - x^2/2$ where for small x (wideband regime) the quadratic term is negligible, we furthermore obtain the looser lower bound

$$\begin{aligned} S &\geq \frac{P}{N_0}\left[1-\frac{N_0 B\tau_{\max}\nu_{\max}}{P}\log\left(1+\frac{P}{B\tau_{\max}\nu_{\max}N_0}\right)\right] \\ &= \frac{P}{N_0}\left[1-\frac{N_0 B\tau_{\max}}{PT_c}\log\left(1+\frac{PT_c}{B\tau_{\max}N_0}\right)\right], \end{aligned}$$

where we defined the channel coherence time as $T_c \triangleq 1/\nu_{\max}$. Noting that the number of resolvable paths is approximately equal to $\tilde{L} = B\tau_{\max}$, defining $\tilde{L}_{\text{crit}} = PT_c/N_0$, and recalling that $C_{\text{AWGN}} = P/N_0$ as in [52], we finally realize that our looser lower bound is equal to that of [52]. Note, however, that for $B \geq B_{\text{crit}}$ we have a different lower bound, namely $S \geq R_{\max}$. Hence, the OFDM system capacity is nonzero also for $B \to \infty$.

4.6 Bounds on Information Rate for Gaussian Signaling

For Gaussian signaling, the inputs $X_{n,k}$ are i.i.d. with distribution $\mathcal{CN}(0, \sigma_x^2)$. For this case, the information rate cannot be derived exactly. However, we already derived the upper bound on the information rate (4.50) that is valid for any distribution of the transmit signal and hence also for Gaussian signaling. Here, we derive lower bounds for the Gaussian case. We again start with the decomposition of mutual information as in (4.33), i.e.,

$$I(\mathbf{Y};\mathbf{X}) = I(\mathbf{Y};\mathbf{X},\mathbf{H}) - I(\mathbf{Y};\mathbf{H}|\mathbf{X})\,. \tag{4.53}$$

Lower Bounds on $I(\mathbf{Y};\mathbf{X},\mathbf{H})$

A lower bound on $I(\mathbf{Y};\mathbf{X},\mathbf{H})$ can be calculated by using that

$$\begin{aligned} I(\mathbf{Y};\mathbf{X},\mathbf{H}) &= h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X},\mathbf{H}) \geq h(\mathbf{Y}|\mathbf{H}) - h(\mathbf{Y}|\mathbf{X},\mathbf{H}) = I(\mathbf{Y};\mathbf{X}|\mathbf{H}) \\ &= h(\mathbf{Y}|\mathbf{H}) - h(\mathbf{Y}|\mathbf{X},\mathbf{H}) = h(\mathbf{Y}|\mathbf{H}) - h(\mathbf{Z}) \\ &= h(\mathbf{Y}|\mathbf{H}) - MK\log(\pi e N_0)\,. \end{aligned} \tag{4.54}$$

For fixed channel realization $\mathbf{H} = \mathbf{h}$, the output $\mathbf{Y}$ is Gaussian with diagonal conditional covariance matrix. We thus have

$$\begin{aligned} h(\mathbf{Y}|\mathbf{H}) &= \mathrm{E}_H\{h(\mathbf{Y}|\mathbf{h})\} = \mathrm{E}_H\left\{\log\left((\pi e)^{MK}\det\left[\mathrm{E}\left\{\mathbf{Y}\mathbf{Y}^H|\mathbf{h}\right\}\right]\right)\right\} \\ &= \sum_{n=1}^{M}\sum_{k=0}^{K-1}\mathrm{E}_H\left\{\log\left(1+\frac{\sigma_x^2}{N_0}|H_{n,k}|^2\right)\right\} + MK\log(\pi e N_0) \\ &= MK\,\mathrm{E}_H\left\{\log\left(1+\frac{\sigma_x^2}{N_0}|H|^2\right)\right\} + MK\log(\pi e N_0)\,. \end{aligned}$$

Here, H is a random variable with distribution $\mathcal{CN}(0, \sigma_{\mathbb{H}}^2)$. Inserting in (4.54), we obtain

$$I(\mathbf{Y};\mathbf{X},\mathbf{H}) \geq MK\,\mathrm{E}_H\left\{\log\left(1+\frac{\sigma_x^2}{N_0}|H|^2\right)\right\}\,. \tag{4.55}$$

Based on (4.55), we can develop two different bounds that are useful for the wideband regime and the narrowband regime, respectively.

For the wideband regime, we lower bound (4.55) using $\log(1+x) \geq x - x^2/2$, which yields

$$I(\mathbf{Y};\mathbf{X},\mathbf{H}) \geq MK\frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\left(1-\frac{3}{4}\frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\right)\,. \tag{4.56}$$

Since the lower bound $\log(1+x) \geq x - x^2/2$ is only tight for small values of x, the bound (4.56) is most useful if $\frac{\sigma_x^2}{N_0}|H|^2$ is small compared to one, i.e., for small SNR. This is generally the case in the wideband regime.

For the narrowband regime, we lower bound (4.55) as

$$I(\mathbf{Y};\mathbf{X},\mathbf{H}) \geq MK\,\mathrm{E}_H\left\{\log\left(\frac{\sigma_x^2}{N_0}|H|^2\right)\right\} = MK\left[\log\left(\frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\right) + \mathrm{E}\left\{\log(|\chi|^2)\right\}\right],$$

where we used $\log(1+x) \geq \log(x)$ and introduced the random variable $\chi \triangleq H/\sigma_{\mathbb{H}}$ that is distributed as $\mathcal{CN}(0,1)$. It can be shown that [176]

$$\mathrm{E}\left\{\log(|\chi|^2)\right\} = -C_{\text{Euler}} \approx 0.5772\,,$$

where C_{Euler} is Euler's constant. Hence, we obtain the lower bound

$$I(\mathbf{Y};\mathbf{X},\mathbf{H}) \geq MK\left[\log\left(\frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\right) - C_{\text{Euler}}\right]. \tag{4.57}$$

This bound is tightest if $\frac{\sigma_x^2}{N_0}|H|^2$ is large compared to one, i.e., if the SNR is large. This is generally the case in the narrowband regime.

Upper Bound on $I(\mathbf{Y};\mathbf{H}|\mathbf{X})$

We start from

$$I(\mathbf{Y};\mathbf{H}|\mathbf{X}) = h(\mathbf{Y}|\mathbf{X}) - h(\mathbf{Y}|\mathbf{X},\mathbf{H}) = h(\mathbf{Y}|\mathbf{X}) - h(\mathbf{Z}) = h(\mathbf{Y}|\mathbf{X}) - MK\log(\pi e N_0)\,. \tag{4.58}$$

We can again decompose $h(\mathbf{Y}|\mathbf{X})$ as in (4.23). Fixing $\mathbf{X} = \mathbf{x}$ and $\mathbf{Y}_{n'} = \mathbf{y}_{n'}$ for $n' = 1,\ldots,n-1$, we find that (cf. (4.26))

$$h(\mathbf{Y}_n|\mathbf{x};\mathbf{y}_1,\ldots,\mathbf{y}_{n-1}) = \log\left(\det\left[\mathbf{I} + \frac{1}{N_0}\mathrm{diag}\left\{\mathbf{x}_n\right\}\mathbf{B}_n\mathrm{diag}\left\{\mathbf{x}_n^*\right\}\right]\right) + K\log(\pi e N_0)\,,$$

where $\mathbf{B}_n$ is given by (3.8). Since the matrix in the determinant is Hermitian symmetric, we can apply the Hadamard inequality [169] to obtain

$$\begin{aligned} h(\mathbf{Y}_n|\mathbf{x};\mathbf{y}_1,\ldots,\mathbf{y}_{n-1}) &\leq \log\left(\prod_{k=0}^{K-1}\left[\mathbf{I} + \frac{1}{N_0}\mathrm{diag}\left\{\mathbf{x}_n\right\}\mathbf{B}_n\mathrm{diag}\left\{\mathbf{x}_n^*\right\}\right]_{k,k}\right) + K\log(\pi e N_0) \\ &= \log\left(\prod_{k=0}^{K-1}\left(1 + \frac{|x_{n,k}|^2}{N_0}[\mathbf{B}_n]_{k,k}\right)\right) + K\log(\pi e N_0) \\ &= \sum_{k=0}^{K-1}\log\left(1 + \frac{|x_{n,k}|^2}{N_0}[\mathbf{B}_n]_{k,k}\right) + K\log(\pi e N_0)\,, \end{aligned} \tag{4.59}$$

which does not depend on $\mathbf{y}_i$, $i = 1, \ldots, n-1$. Unfortunately, this expression cannot easily be further simplified since we here deal with the full-complexity MMSE channel predictor of Subsection 3.2.1 and thus $\mathbf{B}_n$ depends on the data symbols (see (3.8)). Averaging (4.59) with respect to $\mathbf{X}$ and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{n-1}$ yields

$$\begin{aligned} h(\mathbf{Y}_n|\mathbf{X};\mathbf{Y}_1,\ldots,\mathbf{Y}_{n-1}) &= \mathrm{E}_{\mathbf{X},\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_{n-1}}\{h(\mathbf{Y}_n|\mathbf{x};\mathbf{y}_1,\ldots,\mathbf{y}_{n-1})\} \\ &\leq \sum_{k=0}^{K-1} \mathrm{E}_X\left\{\log\left(1+\frac{|X_{n,k}|^2}{N_0}[\mathbf{B}_n]_{k,k}\right)\right\} + K\log(\pi e N_0)\,, \end{aligned}$$

and inserting this into (4.23) results in

$$h(\mathbf{Y}|\mathbf{X}) \leq \sum_{n=1}^{M}\sum_{k=0}^{K-1} \mathrm{E}_X\left\{\log\left(1+\frac{|X_{n,k}|^2}{N_0}[\mathbf{B}_n]_{k,k}\right)\right\} + MK\log(\pi e N_0)\,.$$

Inserting this into (4.58) and using Jensens's inequality [169], we finally obtain the upper bound

$$I(\mathbf{Y};\mathbf{H}|\mathbf{X}) \leq \sum_{n=1}^{M}\sum_{k=0}^{K-1} \log\left(1+\frac{1}{N_0}\mathrm{E}_X\left\{|X_{n,k}|^2\,[\mathbf{B}_n]_{k,k}\right\}\right)\,. \tag{4.60}$$

The problem in finding closed-form expressions for these lower bounds is that the expectation $\mathrm{E}_X\{|X_{n,k}|^2[\mathbf{B}_n]_{k,k}\}$ with $\mathbf{B}_n$ given by (3.8) cannot be calculated since $\mathbf{B}_n$ depends on the data symbols in a complicated way.

Lower Bound on Information Rate

In the wideband regime, we obtain by inserting (4.56) and (4.60) into (4.53)

$$I(\mathbf{Y};\mathbf{X}) \geq MK\frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\left(1-\frac{3}{4}\frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\right) - \sum_{n=1}^{M}\sum_{k=0}^{K-1}\log\left(1+\frac{1}{N_0}\mathrm{E}_X\left\{|X_{n,k}|^2[\mathbf{B}_n]_{k,k}\right\}\right)\,.$$

With (4.1), a lower bound on the information rate for Gaussian signaling in the wideband regime is thus given by

$$R \geq \frac{K}{T}\frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\left(1-\frac{3}{4}\frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\right) - \lim_{M\to\infty}\frac{1}{MT}\sum_{n=1}^{M}\sum_{k=0}^{K-1}\log\left(1+\frac{1}{N_0}\mathrm{E}_X\left\{|X_{n,k}|^2[\mathbf{B}_n]_{k,k}\right\}\right)\,.$$

A similar derivation using (4.57) yields a lower bound on the information rate in the narrowband regime,

$$R \geq \frac{K}{T}\left(\log\left(\frac{\sigma_x^2\sigma_{\mathbb{H}}^2}{N_0}\right) - C_{\text{Euler}}\right) - \lim_{M\to\infty}\frac{1}{MT}\sum_{m=1}^{M}\sum_{k=0}^{K-1}\log\left(1+\frac{1}{N_0}\mathrm{E}_X\left\{|X_{n,k}|^2[\mathbf{B}_n]_{k,k}\right\}\right)\,.$$

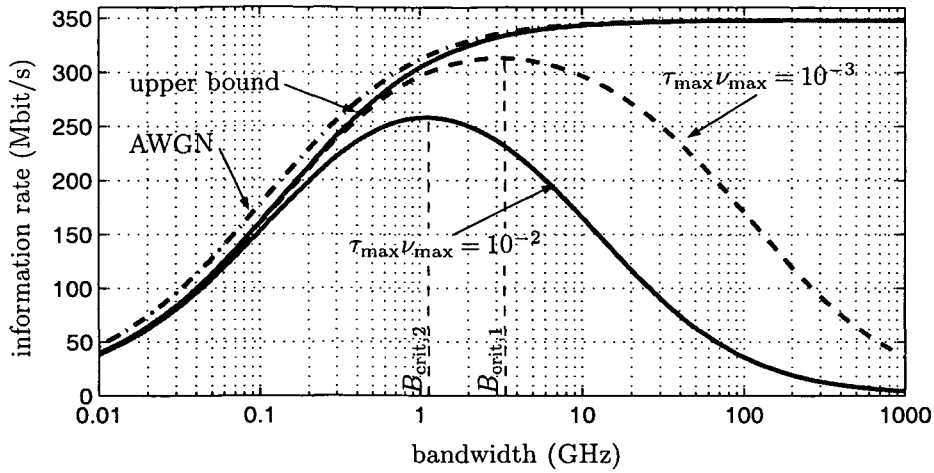


Figure 4.2: *OFDM information rate for constant-modulus signaling vs. bandwidth for two Rayleigh fading channels with different spread* $\tau_{\max}\nu_{\max}$. *For comparison, the upper bound (4.51) and the AWGN capacity are also shown.*

4.7 Simulation Results

In this section, we provide some numerical results aimed at assessing the performance of existing OFDM-based systems in the light of the results obtained in this chapter. Our simulation study will also yield system design guidelines ensuring that overspreading and, hence, small capacity are avoided. We use IEEE 802.11a related system parameters with subcarrier spacing $F = 312.5\text{kHz}$, $TF = N/K = 1.25$, and transmit power $P = 1\text{mW}$. We consider two channels with flat scattering functions and channel spreads $\tau_{\max}\nu_{\max} = 10^{-3}$ and $\tau_{\max}\nu_{\max} = 10^{-2}$. For both channels, the path loss is $\sigma^2_{\mathbb{H}_c} = 90\text{dB}$. Furthermore, $N_0 = k_0 \cdot 300\text{K} = 4.1421 \cdot 10^{-21}\,\text{W/Hz}$ where $k_0 = 1.3807 \cdot 10^{-23}\,\text{Ws/Hz}$ is the Boltzmann constant. For $B = 20\text{MHz}$, for example, these parameters yield a receive SNR of $P\sigma^2_{\mathbb{H}_c}/(BN_0) = 10.8\text{dB}$.

4.7.1 Dependence of Information Rate on Bandwidth

Figure 4.2 compares the AWGN capacity, the information rate for constant-modulus signaling R_{CM} in (4.32), and the upper bound in (4.51) as a function of bandwidth B. The asymptotic ($B \to \infty$) values for the AWGN capacity and the upper bound are both $P\sigma^2_{\mathbb{H}}/(N_0 \log(2)) = 348.3\text{Mbit/s}$. For $\tau_{\max}\nu_{\max} = 10^{-3}$, R_{CM} has a flat maximum at the critical bandwidth $B_{\text{crit},1} = 3.1\text{GHz}$. For $\tau_{\max}\nu_{\max} = 10^{-2}$, the critical bandwidth reduces to $B_{\text{crit},2} = 1.1\text{GHz}$. For bandwidths below B_{crit}, R_{CM} is very close to the upper bound; for

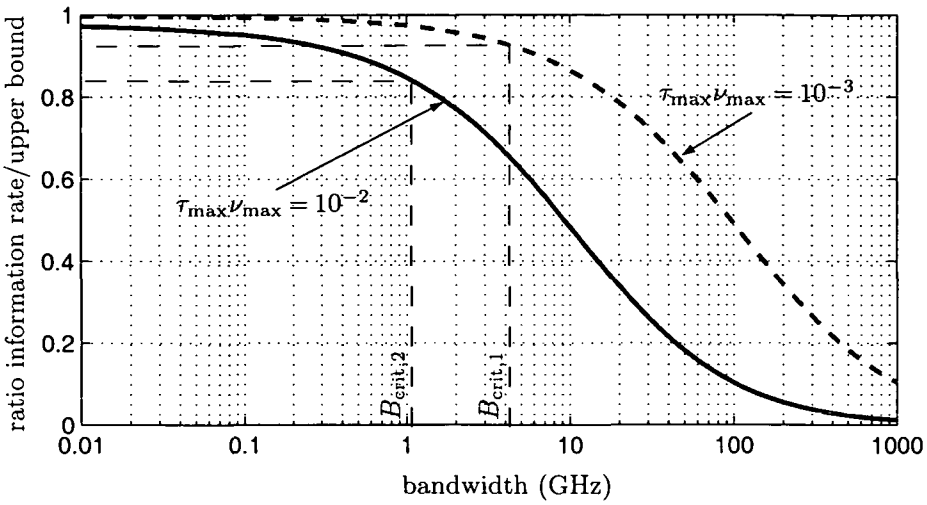


Figure 4.3: *Ratio of information rate for constant-modulus signaling and upper bound for two channels with different spread $\tau_{max}\nu_{max}$.*

bandwidths above B_{crit}, we enter the overspreading regime in which R_{CM} decreases with growing B and asymptotically approaches zero.

To better demonstrate that R_{CM} is indeed very close to the upper bound (4.51) for $B \leq B_{\text{crit}}$, Figure 4.3 shows the ratio of R_{CM} and the upper bound (4.51) vs. bandwidth for the two channels. For $\tau_{\max}\nu_{\max} = 10^{-3}$, R_{CM} decreases to 94% of the upper bound at $B = B_{\text{crit},1}$. For $\tau_{\max}\nu_{\max} = 10^{-2}$, the ratio is 84% at $B = B_{\text{crit},2}$. For $B > B_{\text{crit}}$, the ratios decrease more rapidly and constant-modulus signaling becomes rather inefficient. The system design has to ensure that this operating range is avoided.

The expression for R_{CM} in (4.32) allows to choose the OFDM system parameters (transmit power and bandwidth) such that the system operates close to the maximum information rate $R_{\max}$. This choice critically depends on the scattering function since B_{crit} decreases when the channel spread $\tau_{\max}\nu_{\max}$ increases. However, current OFDM-based wireless systems operate far below the overspreading regime. For example, in IEEE 802.11a $P > 1\text{mW}$ and $\tau_{\max}\nu_{\max} < 10^{-5}$. For $P = 1\text{mW}$ and $\tau_{\max}\nu_{\max} = 10^{-5}$, the critical bandwidth is $B_{\text{crit}} = 24\text{GHz}$ whereas the actual transmit bandwidth for IEEE 802.11a is $B = 20\text{MHz}$.

4.7.2 Dependence of Information Rate on Channel Spread

In Figure 4.4, we show the dependence of the information rate for constant-modulus signaling R_{CM} on the channel spread $\tau_{\max}\nu_{\max}$ for bandwidths of 100 MHz and 1 GHz. We again used a flat scattering function. Figure 4.4(a) shows the result for $B = 100$ MHz. Here, R_{CM} is practically equal to the upper bound (4.51) up to extremely large channel

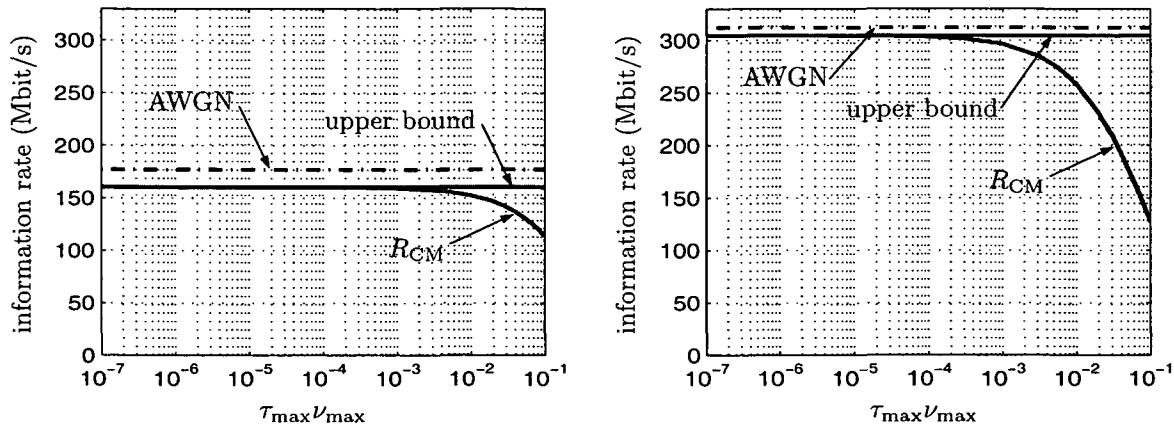


Figure 4.4: *Information rate for constant-modulus signaling vs. channel spread for a uniform scattering function for (a) bandwidth $B = 100$ MHz and (b) bandwidth $B = 1$ GHz.*

spreads. Note that for $\tau_{\max}\nu_{\max} \geq 10^{-2}$ and $TF = 1.25$, the approximative input-output relation (2.36) is not necessarily valid anymore, i.e., the intercarrier interference cannot be neglected for highly time-varying channels. We furthermore note that the AWGN capacity and the upper bound (4.51) do not depend on the channel spread but only on the SNR and the bandwidth.

Figure 4.4(b) shows the result for $B = 1$ GHz. Using ten times the bandwidth (and the same transmit power) nearly doubles the information rate. However, R_{CM} deviates from the upper bound (4.51) already for about $\tau_{\max}\nu_{\max} \geq 3 \cdot 10^{-4}$.

4.7.3 Spectral Efficiency

To further demonstrate the detrimental effect of overspreading, Figure 4.5 shows the spectral efficiency R/B as a function of E_b/N_0, where $E_b = P\sigma^2_{\mathbb{H}_c}/R$, for the AWGN channel, the upper bound (4.51), and the information rate for constant-modulus signaling (4.32). For the AWGN channel, reliable communications require $E_b/N_0 \geq -1.59$dB. For constant-modulus signaling (4.32) and $\tau_{\max}\nu_{\max} = 10^{-3}$ (10^{-2}), we numerically calculated the lowest (E_b/N_0) as $(E_b/N_0)_{\min} = -1.13$dB (-0.29dB), resulting in a spectral efficiency of 0.10bit/s/Hz (0.24bit/s/Hz). This point corresponds to the maximum of (4.32) at $B = B_{\mathrm{crit}}$ in Figure 4.2. Note that with increasing channel spread, we require a higher $(E_b/N_0)_{\min}$ for reliable communications. For $E_b/N_0 \geq (E_b/N_0)_{\min}$, we obtain two spectral efficiency values where the lower one is obtained when the OFDM system operates in the overspreading regime (i.e., $B \geq B_{\mathrm{crit}}$). While it is uncommon that the spectral efficiency vs. E_b/N_0 graph splits

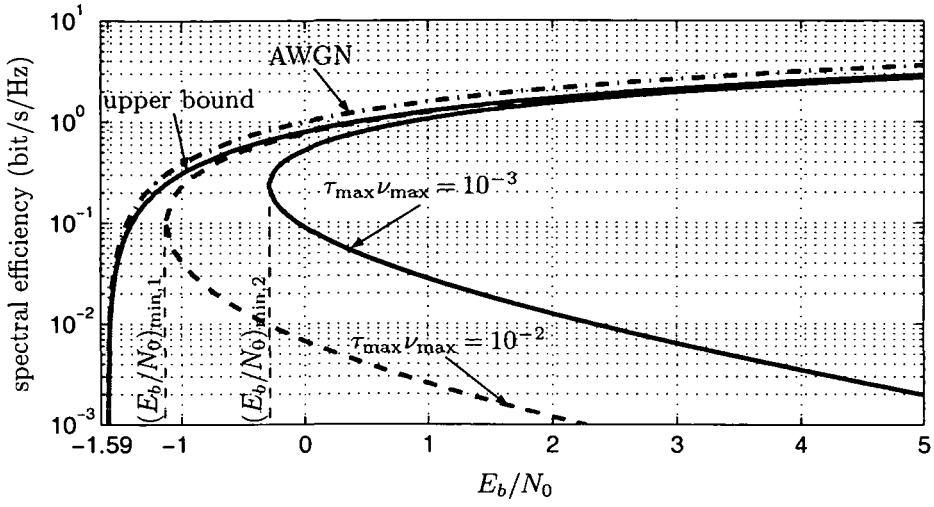


Figure 4.5: *Spectral efficiency R/B vs. E_b/N_0 for the AWGN channel, the upper bound, and the information rate for constant-modulus signaling.*

up into two branches, it should be noted that also in Figure 4.2 a given information rate can be achieved with two different bandwidths, one smaller and one larger than B_{crit}.

An interesting aspect of OFDM communications with constant-modulus signaling over time and frequency selective fading channels is that the minimum energy per bit for reliable communications, $(E_b/N_0)_{\text{min}}$, increases with the channel spread. This dependence is shown in Figure 4.6(a). For $\tau_{\text{max}}\nu_{\text{max}}$ below approximately 10^{-4}, $(E_b/N_0)_{\text{min}}$ is practi-

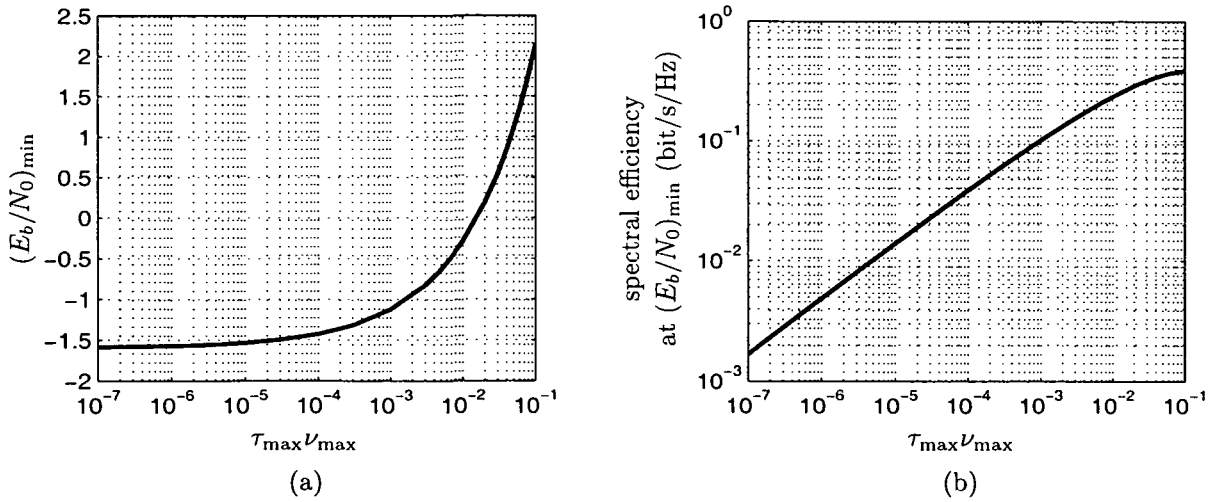


Figure 4.6: *Dependence on the channel spread of (a) the minimum E_b/N_0 required for reliable communications using constant-modulus, $(E_b/N_0)_{min}$, and (b) the spectral efficiency R/B at $(E_b/N_0)_{min}$.*

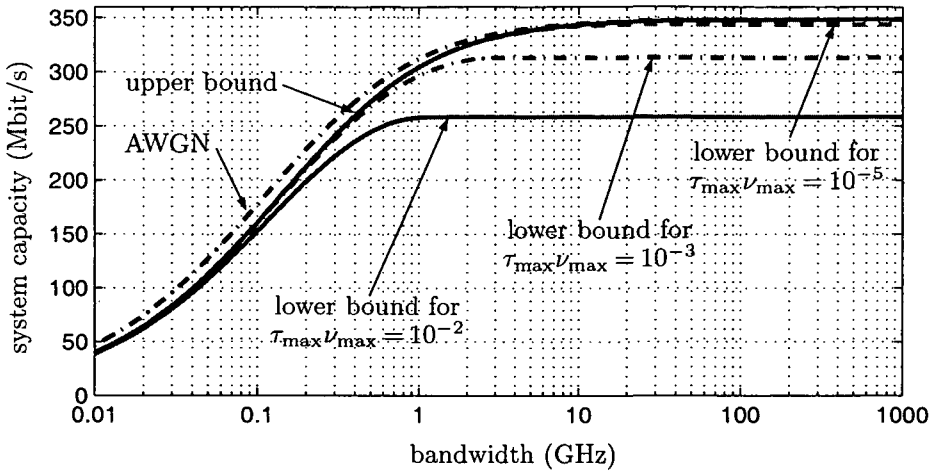


Figure 4.7: *OFDM system capacity bounds vs. bandwidth. For comparison the AWGN capacity is also shown.*

cally equal the $(E_b/N_0)_{\min}$ of the AWGN channel which is -1.59 dB. For larger channel spreads, $(E_b/N_0)_{\min}$ increases rapidly. In Figure 4.6(b), we show which spectral efficiencies are obtained at $(E_b/N_0)_{\min}$ for different channel spreads. Note that these values do not correspond to the lowest possible spectral efficiency but rather to the most efficient operating point with respect to E_b/N_0.

4.7.4 Bounds on System Capacity

In Figure 4.7, we finally compare the upper bound (4.51), the lower bound (4.52) for three different channel spreads, and the AWGN capacity as a function of bandwidth B. Note that for $B \leq B_{\text{crit}}$, the lower bounds in Figure 4.7 are identical to the OFDM information rate using constant-modulus signaling (4.32) for a flat scattering function (cf. Subsection 4.5.2). It is seen that the upper bound is only slightly lower than the AWGN capacity; indeed, for large bandwidths both coincide. The lower bound depends on the channel spread. However, for practical channel spreads, the lower bound is close to the upper bound. Specifically, for a channel spread of $\tau_{\max}\nu_{\max} = 10^{-5}$ that is typically encountered in slow-mobility environments such as indoor communications, the lower bound is almost identical to the upper bound. Even for $\tau_{\max}\nu_{\max} = 10^{-2}$, the lower bound is close to the upper bound for bandwidths up to about 200 MHz. Note that for current wireless communication systems, the bandwidth is typically smaller than 20 MHz (the bandwidth of IEEE 802.11a). However, OFDM is currently being considered for ultra-wideband (UWB) communications by the

standards committee IEEE 802.15 with a bandwidth of about 500 MHz [14]. Depending on the transmit power, path loss, and channel spread, such OFDM systems could be close to the critical bandwidth B_{crit}.

5

Conclusions

In this final chapter, we summarize the most important results of our work and suggest possible extensions of these results.

In Chapter 2, we reviewed a model for pulse-shaping OFDM systems together with a model for random time and frequency selective wireless channels. The input-output relation of the resulting system can be approximated by a simple pointwise multiplication of the transmit symbols by complex-valued channel coefficients, plus the addition of Gaussian noise. Hence, apart from the noise, each data symbol merely experiences flat fading. This simple approximate system input-output relation provided a basis for our development in subsequent chapters.

We presented an efficient DFT-based digital implementation of the pulse-shaping OFDM modulator and demodulator. The computational complexity of this implementation is only slightly larger than that of a conventional cyclic-prefix OFDM system. Furthermore, we characterized the channel coefficients by introducing the concept of an *equivalent channel*. It was seen that the channel coefficients can be obtained by sampling the time-dependent transfer function of the continuous-time channel on a rectangular time-frequency lattice that is induced by the OFDM modulation structure. The channel coefficients decorrelate for time lags larger than the channel's coherence time and for frequency lags larger than the channel's coherence bandwidth.

In Section 3, we presented decision-directed channel predictors for OFDM communications over time and frequency selective fading channels. Channel prediction is interesting because by compensating unavoidable delays (such as coding/decoding delays), it is capable

of yielding up-to-date channel state information (CSI). The proposed channel predictors can be operated in decision-directed mode and do not require regular transmission of pilot symbols. The successful application of channel prediction to delay-free channel equalization was demonstrated. Channel prediction also enables the use of advanced communication techniques such as adaptive modulation that hold the promise of improved system capacity and link reliability. For adaptive modulation, we proposed a novel protocol for signaling the transmission parameters from the receiver back to the transmitter, and we discussed the use of channel prediction for obtaining CSI at the receiver.

We derived the full-complexity MMSE channel predictor which, however, is unpractical because of its excessive computational complexity. We then developed a reduced-complexity MMSE predictor that allowed an efficient DFT-based implementation. We also proposed adaptive predictors using the normalized least-mean-square (NLMS) or recursive least-squares (RLS) algorithm. These adaptive predictors avoid an explicit predictor design, do not require any statistical prior knowledge, and are able to track nonstationary channel and noise statistics. Our simulation results demonstrated that in decision-directed mode, using only a single known OFDM symbol for initialization, adaptive prediction features excellent performance even in the case of fast time-varying channels. However, to avoid error propagation, a certain minimum SNR is required that depends on system and channel parameters. Above this SNR threshold, prediction of time-varying channels over large prediction horizons is feasible.

We furthermore analyzed the infinite-length one-step MMSE predictor and calculated its performance in terms of the channel's scattering function, noise variance, and OFDM system parameters. We also showed that for specular scattering, the prediction error can be made arbitrarily small.

In Chapter 4, we considered the system capacity of OFDM transmitting over time and frequency selective Rayleigh fading channels, under the assumption that transmitter and receiver have no CSI. We derived both an upper bound and a lower bound on the system capacity and demonstrated that these bounds practically coincide for typical system and channel parameters over large and practically relevant ranges of bandwidth, and that the upper bound is close to the AWGN channel capacity. Hence, the system capacity of wireless OFDM systems is close to the AWGN channel capacity. Moreover, even for very large bandwidths the OFDM system capacity does not necessarily vanish; this was demonstrated by the lower bound and by calculation of the infinite-bandwidth system capacity.

A main result of Chapter 4, which also led to the lower bound, was the derivation of the OFDM information rate obtained for constant-modulus signaling. This result is important because this signaling scheme is widely used in practical systems such as IEEE 802.11a

and DVB-T. From our derivation, it can be conjectured that this information rate can be achieved by a receiver that uses channel prediction for obtaining CSI.

It was shown that the information rate for constant-modulus signaling deviates from the upper bound on system capacity due to a "penalty term" which is related to the predictability of the channel, and which is small for channels that can be well predicted. Our numerical results demonstrated that this penalty term is small up to a critical bandwidth. For larger bandwidths, however, the penalty term is large and thus the information rate for constant-modulus signaling asymptotically approaches zero. Our expression for the information rate also allowed to study the impact of the shape and spread of the channel's scattering function on the information rate. In particular, we showed that the worst-case scattering function minimizing the information rate is the flat scattering function.

We finally provide some suggestions for future research concerning possible extensions of the work presented in this thesis.

- The idea of decision-directed channel prediction can be extended to wireless MIMO-OFDM systems which have been proposed for future communication systems. It has been demonstrated in [46] that CSI is required to exploit the full potential of MIMO systems. This CSI can be acquired by prediction. An important application of such a scheme could be space-time decoding. If the spatial correlations or structure of the MIMO channel can be exploited, the performance of channel prediction can be expected to improve. However, the results will critically depend on the channel model used.

- CSI at the *transmitter* is required for the application of techniques for link adaptation, pre-equalization, and precoding that hold the promise of improved system capacity and link reliability. Up-to-date CSI can be obtained through channel prediction. Assessing the sensitivity of the adaptive modulation algorithms with respect to channel uncertainty would be an interesting topic for future research. Furthermore, it would be interesting to implement strategies that only require the magnitude of the channel coefficients at the transmitter. For example, one could consider pre-equalization of the magnitude at the transmitter and phase equalization at the receiver.

- Our information-theoretic analysis of wireless OFDM systems could be extended to communication systems using orthogonal frequency division multiple access (OFDMA) for the uplink. Specifically, it would be interesting to consider the capacity regions for OFDMA communications over time and frequency selective fading channels.

- The information rate of an OFDM system applying pilot symbol assisted channel estimation could be derived and compared to the information rate obtained with channel prediction. How large is the loss in information rate due to pilot symbols and how does this system behave for very large bandwidths? Furthermore, one could attempt to derive the "optimal" pilot locations maximizing the information rate.

- The exact OFDM system input-output relation that includes intersymbol and inter-channel interference (cf. Section 2.3) could be the basis for an extended information-theoretic analysis of OFDM systems. Is it possible to calculate (bounds on) the information rate or system capacity? If so, how does the system behave when the channel spread approaches one?

- In practical systems, besides the ergodic information rate, the outage capacity is of great operational significance. To assess the outage capacity, the cumulative distribution function of the instantaneous information rate has to be calculated. If no analytic derivation is possible, one could resort to numerical methods. An interesting question here is how the outage capacity depends on the bandwidth as well as the spread and shape of the scattering function.

Bibliography

[1] J. G. Proakis, *Digital Communications.* New York: McGraw-Hill, 3rd ed., 1995.

[2] T. S. Rappaport, *Wireless Communications: Principles & Practice.* Upper Saddle River (NJ): Prentice Hall, 1996.

[3] J. D. Parsons, *The Mobile Radio Propagation Channel.* London: Pentech Press, 1992.

[4] R. W. Chang, "Synthesis of band-limited orthogonal signals for multi-channel data transmission," *Bell Syst. Tech. J.*, vol. 45, pp. 1775–1796, Dec. 1966.

[5] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency division multiplexing using the discrete Fourier transform," *IEEE Trans. Comm. Technol.*, vol. 19, pp. 628–634, Oct. 1971.

[6] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," in *Proc. IEEE ICASSP-80*, (Denver, CO), pp. 964–967, 1980.

[7] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Comm.*, vol. 33, pp. 665–675, July 1985.

[8] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Comm. Mag.*, vol. 28, pp. 5–14, May 1990.

[9] IEEE P802 LAN/MAN Committee, "The working group for wireless local area networks (WLANs)." http://grouper.ieee.org/groups/802/11/index.html.

[10] M. Radimirsch and V. Vollmer, "HIPERLAN type 2 standardisation — An overview," in *Proc. European Wireless '99, ITG-Fachtagung*, (Munich, Germany), pp. 139–144, Oct. 1999.

[11] ETSI, "Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television." Draft EN 300 744, V1.2.1, 1999. http://www.etsi.org.

[12] ETSI, "Digital audio broadcasting (DAB) to mobile, portable and fixed receivers." ETS 300 401, 1995. http://www.etsi.org.

[13] H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: Design, performance, and field trial results," *IEEE Comm. Mag.*, vol. 40, pp. 143–149, Sept. 2002.

[14] G. R. Aiello and G. D. Rogerson, "Ultra-wideband wireless systems," *IEEE Microwave Magazine*, vol. 4, pp. 36–47, June 2003.

[15] B. R. Saltzberg, "Performance of an efficient parallel data transmission system," *IEEE Trans. Comm. Technol.*, vol. 15, pp. 805–811, Dec. 1967.

[16] B. Hirosaki, "An orthogonally multiplexed QAM system using the discrete Fourier transform," *IEEE Trans. Comm.*, vol. 29, pp. 982–989, July 1981.

[17] B. LeFloch, M. Alard, and C. Berrou, "Coded orthogonal frequency division multiplex," *Proc. IEEE*, vol. 83, pp. 982–996, June 1995.

[18] H. Bölcskei, P. Duhamel, and R. Hleiss, "Design of pulse shaping OFDM/OQAM systems for high data-rate transmission over wireless channels," in *Proc. IEEE ICC-99*, (Vancouver, Canada), pp. 559–564, June 1999.

[19] C. Siclet and P. Siohan, "Design of OFDM/OQAM systems based on biorthogonal modulated filter banks," in *Proc. IEEE GLOBECOM-2000*, (San Francisco, CA), pp. 701–705, Dec. 2000.

[20] P. Siohan, C. Siclet, and N. Lacaille, "Analysis and design of OFDM/OQAM systems based on filterbank theory," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1170–1183, 2002.

[21] H. G. Feichtinger and T. Strohmer, eds., *Gabor Analysis and Algorithms: Theory and Applications*. Boston (MA): Birkhäuser, 1998.

[22] H. Bölcskei, H. G. Feichtinger, K. Gröchenig, and F. Hlawatsch, "Discrete-time Wilson expansions," in *Proc. IEEE-SP Int. Sympos. Time-Frequency Time-Scale Analysis*, (Paris, France), pp. 525–528, June 1996.

[23] T. Strohmer and S. Beaver, "Optimal OFDM design for time-frequency dispersive channels," *IEEE Trans. Comm.*, vol. 51, pp. 1111–1122, July 2003.

[24] M. M. Hartmann, G. Matz, and D. Schafhuber, "Multipulse multicarrier communications over time-varying fading channels: Performance analysis and system optimization." *IEEE ICASSP 2004*, accepted for publication.

[25] M. M. Hartmann, G. Matz, and D. Schafhuber, "Theory and design of multipulse multicarrier systems for wireless communications," in *Proc. 37th Asilomar Conf. Signals, Systems, Computers*, (Pacific Grove, CA), Nov. 2003.

[26] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers—Part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2006, July 1999.

[27] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers—Part II: Blind channel estimation, synchronization, and direct equalization," *IEEE Trans. Signal Processing*, vol. 47, pp. 2007–2022, July 1999.

[28] S. Barbarossa and A. Scaglione, "Optimal precoding for transmissions over linear time-varying channels," in *Proc. IEEE GLOBECOM-99*, (Rio de Janeiro, BR), pp. 2545–2549, Dec. 1999.

[29] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "Analysis of certain large isometric random precoded systems," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1293–1311, May 2003.

[30] R. Mhiri, D. Masse, and D. Schafhuber, "Synchronization for a DVB-T receiver in presence of co-channel interference," in *Proc. IEEE PIMRC-02*, (Lisbon, Portugal), pp. 2307–2311, Sept. 2002.

[31] H. Bölcskei, "Blind high-resolution uplink synchronization of OFDM-based multiple access schemes," in *Proc. Second IEEE Workshop on Signal Processing Applications in Wireless Communication*, (Annapolis, MD), pp. 166–169, May 1999.

[32] L. Wei and C. Schlegel, "Synchronization requirements for multi-user OFDM on satellite mobile and two-path Rayleigh fading channels," *IEEE Trans. Comm.*, vol. 43, pp. 887–895, Feb. 1993.

[33] W. D. Warner and C. Leung, "OFDM/FM frame synchronization for mobile radio data communication," *IEEE Trans. Veh. Technol.*, vol. 42, no. 3, pp. 302–313, 1993.

[34] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Comm.*, vol. 45, pp. 1613–1621, Dec. 1997.

[35] S. H. Müller-Weinfurtner, "On the optimality of metrics for coarse frame synchronization in OFDM: A comparison," in *IEEE PIMRC-98*, (Boston, USA), pp. 533–537, Sept. 1998.

[36] M. Speth, F. Classen, and H. Meyr, "Frame synchronization of OFDM systems in frequency selective fading channels," in *IEEE-VTC97*, pp. 1807–1811, 1997.

[37] A. Czylwik, "Synchronization for systems with antenna diversity," in *IEEE-VTC99-Fall*, pp. 728–732, 1999.

[38] F. Classen and H. Meyr, "Frequency synchronization algorithms for OFDM systems suitable for communication over frequency-selective fading channels," in *Proc. IEEE Vehic. Technol. Conf.*, (Stockholm, Sweden), pp. 1655–1659, 1994.

[39] V. Tarokh and H. Jafarkhani, "On the computation and reduction of the peak-to-average power ratio in multicarrier communications," *IEEE Trans. Comm.*, vol. 48, pp. 37–44, Jan. 2000.

[40] S. H. Müller and J. B. Huber, "A novel peak power reduction scheme for OFDM," in *IEEE PIMRC-97*, (Helsinki, Finland), pp. 1090–1094, Sept. 1997.

[41] S. H. Müller and J. B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *Electronics Letters*, vol. 33, pp. 368–369, Feb. 1997.

[42] T. A. Wilkinson and A. E. Jones, "Minimisation of the peak to mean envelope power ratio of multicarrier transmission schemes by block coding," in *Proc. IEEE VTC-95*, (Chicago, IL), pp. 825–829, 1995.

[43] A. Jones, T. Wilkinson, and S. Barton, "Block coding scheme for reduction of peak to mean envelope power ratio of multicarrier transmission schemes," *Electron. Letters*, vol. 30, pp. 2098–2099, Dec. 1994.

[44] P. V. Eetvelt, G. Wade, and M. Tomlinson, "Peak to average power reduction for OFDM schemes by selective scrambling," *Electronics Letters*, vol. 32, pp. 1963–64, Oct. 1996.

[45] S. Barbarossa, M. Pompili, and G. B. Giannakis, "Channel-independent synchronization of orthogonal frequency division multiple access systems," *IEEE J. Sel. Areas Comm.*, vol. 20, pp. 474–486, Feb. 2002.

[46] A. Paulraj, R. U. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge (UK): Cambridge Univ. Press, 2003.

[47] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Comm.*, vol. 17, pp. 1747–1758, Oct. 1999.

[48] A. Scaglione and S. Barbarossa, "Optimal power loading for OFDM transmissions over underspread Rayleigh time-varying channels," in *Proc. IEEE ICASSP-2000*, (Istanbul, Turkey), pp. 2969–2972, June 2000.

[49] S. Catreux, D. Gesbert, V. Erceg, and R. Heath, "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Comm. Mag.*, vol. 40, June 2002.

[50] Tung-Sheng Yang and A. Duel-Hallen, "Adaptive modulation using outdated samples of another fading channel," in *Proc. IEEE WCNC-2002*, (Orlando, FL), pp. 477–481, March 2002.

[51] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side information" be?," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1118–1134, May 2002.

[52] I. Telatar and D. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1384–1400, July 2000.

[53] M. Médard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Trans. Inf. Theory*, vol. 48, pp. 840–852, April 2002.

[54] V. G. Subramanian and B. Hajek, "Broad-band fading channels: Signal burstiness and capacity," *IEEE Trans. Inf. Theory*, vol. 48, pp. 809–827, April 2002.

[55] P. Hoeher, "TCM on frequency-selective land-mobile fading channels," in *Proc. 5-th Tirrenia International Workshop*, (Tirrenia, Italy), pp. 317–328, Sept. 1991.

[56] M. Sandell, *Design and Analysis of Estimators for Multicarrier Modulation and Ultrasonic Imaging.* PhD thesis, Lulea University of Technology, Lulea, Sweden, 1996.

[57] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Comm.*, vol. 46, pp. 931–939, July 1998.

[58] J. J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Börjesson, "On channel estimation in OFDM systems," in *Proc. IEEE VTC-95*, (Chicago, IL), pp. 815–819, July 1995.

[59] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 1207–1215, July 2000.

[60] P. Hoeher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," in *Proc. IEEE ICASSP-97*, (Munich, Germany), pp. 1845–1848, April 1997.

[61] M. J. Fernández-Getino Garciá, J. M. Páez-Borrallo, and S. Zazo, "DFT-based channel estimation in 2D-pilot-symbol-aided OFDM wireless systems," in *Proc. IEEE VTC-01 Spring*, pp. 810–814, 2001.

[62] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Processing*, vol. 49, pp. 3065–3073, Dec. 2001.

[63] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," in *Proc. IEEE-VTC97*, (Phoenix, USA), pp. 1639–43, May 1997.

[64] D. Schafhuber, G. Matz, F. Hlawatsch, and P. Loubaton, "MMSE estimation of time-varying channels for DVB-T systems with strong co-channel interference," in *Proc. EUSIPCO-02*, vol. III, (Toulouse, France), pp. 25–28, Sept. 2002.

[65] L. Hanzo, M. Münster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting.* IEEE Press and Wiley, 2003.

[66] D. Schafhuber, M. Rupp, G. Matz, and F. Hlawatsch, "Adaptive identification and tracking of doubly selective fading channels for wireless MIMO-OFDM systems," in *Proc. IEEE SPAWC-03*, (Rome, Italy), June 2003.

[67] D. Schafhuber, G. Matz, and F. Hlawatsch, "Adaptive Wiener filters for time-varying channel estimation in wireless OFDM systems," in *Proc. IEEE ICASSP-2003*, vol. 4, (Hong Kong), pp. 688–691, April 2003.

[68] D. Schafhuber, G. Matz, and F. Hlawatsch, "Kalman tracking of time-varying channels in wireless MIMO-OFDM systems," in *Proc. 37th Asilomar Conf. Signals, Systems, Computers*, (Pacific Grove, CA), Nov. 2003.

[69] D. Schafhuber, G. Matz, and F. Hlawatsch, "Adaptive tracking of time-frequency selective wireless channels in MIMO-OFDM communications," *IEEE Trans. Wireless Comm.*, in preparation.

[70] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Processing*, vol. 51, pp. 1615–1624, June 2003.

[71] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Comm.*, vol. 49, pp. 467–479, March 2001.

[72] P. Schniter, "Low-complexity estimation of doubly-selective channels," in *Proc. IEEE SPAWC-03*, (Rome, Italy), June 2003.

[73] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electron.*, vol. 44, pp. 1122–1128, Aug. 1998.

[74] Y. Li, L. Cimini, and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Comm.*, vol. 46, pp. 902–915, July 1998.

[75] H. Artés, G. Matz, and F. Hlawatsch, "Unbiased scattering function estimators for underspread channels and extension to data-driven operation," *IEEE Trans. Signal Processing*, 2003, to appear.

[76] V. Mignone and A. Morello, "CD3-OFDM: A novel demodulation scheme for fixed and mobile receivers," *IEEE Trans. Comm.*, vol. 44, pp. 1144–1151, Sept. 1996.

[77] P. Frenger and A. Svensson, "A decision directed coherent detector for OFDM," in *Proc. IEEE VTC-1996*, (Atlanta, GA), pp. 1584–1593, Apr./May 1996.

[78] A. Chini, Y. Wu, M. El-Tanany, and S. Mahmoud, "Filtered decision feedback channel estimation for OFDM-based DTV terrestrial broadcasting system," *IEEE Trans. Broadcasting*, vol. 44, pp. 2–11, March 1998.

[79] P. Frenger, N. Arne, and B. Svensson, "Decision-directed coherent detection in multicarrier systems on Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 490–498, March 1999.

[80] R. Liu and L. Tong (eds.), "Blind System Identification and Estimation," *Special Issue of Proc. IEEE*, vol. 86, Oct. 1998.

[81] R. W. Heath and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Processing*, vol. 47, pp. 848–856, March 1999.

[82] B. Muquet and M. de Courville, "Blind and semi-blind channel identification methods using second order statistics for OFDM systems," in *Proc. IEEE ICASSP-99*, (Phoenix, AZ), pp. 2745–2748, March 1999.

[83] H. Bölcskei, P. Duhamel, and R. Hleiss, "Blind channel identification in high-data-rate pulse shaping OFDM/OQAM systems," in *IEEE SP Workshop on Signal Processing Advances in Wireless Communications*, (Annapolis, MD), pp. 154–157, May 1999.

[84] H. Bölcskei, P. Duhamel, and R. Hleiss, "A subspace-based approach to blind channel estimation in pulse shaping OFDM systems," *IEEE Trans. Signal Processing*, vol. 49, April 2001.

[85] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Processing*, vol. 50, pp. 1699–1712, July 2002.

[86] F. Duel-Hallen, S. Hu, and H. Hallen, "Long-range prediction of fading signals," *IEEE Signal Processing Magazine*, vol. 17, pp. 62–75, May 2000.

[87] Y. Liu and S. D. Blostein, "Identification of frequency non-selective fading channels using decision feedback and adaptive linear prediction," *IEEE Trans. Comm.*, vol. 43, pp. 1484–1492, Feb.-March-April 1995.

[88] T. Ekman and G. Kubin, "Nonlinear prediction of mobile radio channels: Measurements and MARS model designs," in *Proc. IEEE ICASSP-99*, (Phoenix, AZ), pp. 2667–70, March 1999.

[89] T. Ekman, *Prediction of mobile radio channels.* PhD thesis, Uppsala University, Uppsala, Sweden, 2002.

[90] E. Al-Susa and R. F. Ormondroyd, "A predictor-based decision feedback channel estimation method for COFDM with high resilience to rapid time-variations," in *Proc. IEEE VTC-99 Fall*, (Amsterdam, The Netherlands), pp. 273–278, Sept. 1999.

[91] S. Thoen, L. Van der Perre, B. Gyselinckx, M. Engels, and H. De Man, "Predictive adaptive loading for HIPERLAN II," in *Proc. IEEE VTC-00 Fall*, (Boston, MA), pp. 2166–2172, Sept. 2000.

[92] D. Schafhuber and G. Matz, "MMSE and adaptive prediction of time-varying channels for OFDM systems," accepted for publication in *IEEE Trans. Wireless Comm.*, 2003.

[93] D. Schafhuber, G. Matz, and F. Hlawatsch, "Adaptive prediction of time-varying channels for coded OFDM systems," in *Proc. IEEE ICASSP-2002*, (Orlando, FL), pp. 2549–2552, May 2002.

[94] D. Schafhuber, G. Matz, and F. Hlawatsch, "Predictive equalization of time-varying channels for coded OFDM/BFDM systems," in *Proc. IEEE GLOBECOM-2000*, (San Franscisco, CA), pp. 721–725, Nov./Dec. 2000.

[95] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[96] I. Jacobs, "The asymptotic behavior of incoherent m-ary communication systems," *Proc. IEEE*, vol. 51, pp. 251–252, Jan. 1963.

[97] J. R. Pierce, "Ultimate performance of m-ary transmissions on fading channels," *IEEE Trans. Inf. Theory*, vol. 12, pp. 2–5, Jan. 1966.

[98] R. S. Kennedy, *Fading Dispersive Communication Channels.* New York: Wiley, 1969.

[99] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[100] A. J. Viterbi, "Performance of an m-ary orthogonal communication system using stationary stochastic signals," *IEEE Trans. Inf. Theory*, vol. 13, pp. 414–422, July 1967.

[101] S. Verdú, "Recent results on the capacity of wideband channels in the low-power regime," *IEEE Wireless Communications*, pp. 40–45, Aug. 2002.

[102] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1319–1343, June 2002.

[103] C. Zheng and M. Médard, "How far should we spread using DS-CDMA in time and frequency selective fading channels," in *Proc. IEEE GLOBECOM-2003*, (San Franscisco, CA), Dec. 2003.

[104] D. Porrat and D. Tse, "Bandwidth scaling in ultra wideband communication," in *Proc. 41st Allerton Conf. Commun., Contr., Comput.*, (Urbana, IL), Oct. 2003.

[105] W. Kozek, "On the transfer function calculus for underspread LTV channels," *IEEE Trans. Signal Processing*, vol. 45, pp. 219–223, Jan. 1997.

[106] W. Kozek and A. F. Molisch, "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE J. Sel. Areas Comm.*, vol. 16, pp. 1579–1589, Oct. 1998.

[107] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, pp. 933–946, May 2000.

[108] D. Schafhuber, G. Matz, and F. Hlawatsch, "Pulse-shaping OFDM/BFDM systems for time-varying channels: ISI/ICI analysis, optimal pulse design, and efficient implementation," in *Proc. IEEE PIMRC-02*, (Lisbon, Portugal), pp. 1012–1016, Sept. 2002.

[109] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing.* Englewood Cliffs (NJ): Prentice Hall, 1992.

[110] D. Schafhuber, H. Bölcskei, and G. Matz, "System capacity of wideband OFDM communications over fading channels without channel knowledge." *IEEE ISIT-04*, submitted.

[111] D. Schafhuber, H. Bölcskei, and G. Matz, "OFDM system capacity for wideband Rayleigh fading channels," *IEEE Trans. Inf. Theory*, in preparation.

[112] A. Vahlin and N. Holte, "Optimal finite duration pulses for OFDM," *IEEE Trans. Comm.*, vol. 4, pp. 10–14, Jan. 1996.

[113] R. Haas and J. C. Belfiore, "A time-frequency well-localized pulse for multiple carrier transmission," *Wireless Personal Comm.*, vol. 5, pp. 1–18, 1997.

[114] A. N. Akansu, P. Duhamel, X. Lin, and M. de Courville, "Orthogonal transmultiplexers in communication: A review," *IEEE Trans. Signal Processing*, vol. 46, pp. 979–995, April 1998.

[115] H. Bölcskei, "Efficient design of pulse shaping filters for OFDM systems," in *Proc. SPIE Wavelet Applications in Signal and Image Processing VII*, (Denver, CO), pp. 625–636, July 1999.

[116] P. Schniter, "On the design of non-(bi)orthogonal pulse-shaped FDM for doubly-dispersive channels." *IEEE ICASSP-04*, submitted.

[117] K. Gröchenig, *Foundations of Time-Frequency Analysis.* Boston: Birkhäuser, 2001.

[118] P. P. Vaidyanathan, *Multirate Systems and Filter Banks.* Englewood Cliffs (NJ): Prentice Hall, 1993.

[119] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Comm. Mag.*, vol. 33, pp. 100–109, Feb. 1995.

[120] J.-J. van de Beek, M. Sandell, and P.-O. Börjesson, "ML estimation of timing and frequency offset in OFDM systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 1800–1805, July 1997.

[121] D. Landström, S. K. Wilson, J.-J. van de Beek, P. Ödling, and P. O. Börjesson, "Symbol time offset estimation in coherent OFDM systems," in *Proc. IEEE ICC-99*, (Vancouver, Canada), pp. 500–505, June 1999.

[122] T. Walzman and M. Schwartz, "Automatic equalization using the discrete Fourier domain," *IEEE Trans. Inf. Theory*, vol. 19, pp. 59–68, Jan. 1973.

[123] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs (NJ): Prentice Hall, 3rd ed., 1996.

[124] G. Kadel, "Diversity and equalization in frequency domain—a robust and flexible receiver technology for broadband mobile communication systems," in *Proc. IEEE VTC-1997*, (Phoenix, AZ), pp. 894–898, May 1997.

[125] M. V. Clark, "Adaptive frequency-domain equalization and diversity combining for broadband wireless communications," *IEEE J. Sel. Areas Comm.*, vol. 16, pp. 1385–1395, Oct. 1998.

[126] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications," *IEEE Signal Processing Magazine*, vol. 17, pp. 29–48, May 2000.

[127] N. Al-Dhahir, "Single-carrier frequency-domain equalization for space-time block-coded transmission over frequency-selective fading channels," *IEEE Comm. Letters*, vol. 5, pp. 304–306, July 2001.

[128] Z. Wang, X. Ma, and G. B. Giannakis, "Optimality of single-carrier zero-padded block transmission," in *Proc. IEEE WCNC-2002*, (Orlando, FL), pp. 660–664, March 2002.

[129] B. Muquet, P. Magniez, P. Duhamel, M. de Courville, and G. B. Giannakis, "Turbo demodulation of zero-padded OFDM transmissions," in *Proc. 34th Asilomar Conf. Signals, Systems, Computers*, (Pacific Grove, CA), pp. 1815–1819, Oct./Nov. 2000.

[130] B. Muquet, M. de Courville, P. Duhamel, G. B. Giannakis, and P. Magniez, "Turbo demodulation of zero-padded OFDM transmissions," *IEEE Trans. Comm.*, vol. 50, pp. 1725–1728, Nov. 2002.

[131] B. Muquet, Z. Wang, G. B. Giannakis, M. de Courville, and P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions?," *IEEE Trans. Comm.*, vol. 50, pp. 2136–2148, Dec. 2002.

[132] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

[133] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Comm. Syst.*, vol. 11, pp. 360–393, 1963.

[134] G. Matz, "On doubly underspread wireless fading channels," *IEEE Trans. Wireless Comm.*, submitted.

[135] G. Matz, "Doubly underspread non-WSSUS channels: Analysis and estimation of channel statistics," in *Proc. IEEE SPAWC-03*, (Rome, Italy), June 2003.

[136] G. Matz, "Characterization of non-WSSUS fading dispersive channels," in *Proc. IEEE ICC-2003*, (Anchorage, AK), pp. 2480–2484, May 2003.

[137] G. Matz and F. Hlawatsch, "Time-frequency transfer function calculus (symbolic calculus) of linear time-varying systems (linear operators) based on a generalized underspread theory," *J. Math. Phys., Special Issue on Wavelet and Time-Frequency Analysis*, vol. 39, pp. 4041–4071, Aug. 1998.

[138] G. Matz and F. Hlawatsch, "Time-frequency characterization of random time-varying channels," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference* (B. Boashash, ed.), ch. 9.5, pp. 410–419, Oxford (UK): Elsevier, 2003.

[139] K. Kalliola, H. Laitinen, P. Vainikainen, M. Toeltsch, J. Laurila, and E. Bonek, "3-D double-directional radio channel characterisation for urban macrocellular applications," *IEEE Trans. Antennas and Propagation*, vol. 51, pp. 3122–3133, Nov. 2003.

[140] H. Hofstetter, M. Steinbauer, and C. F. Mecklenbräuker, "Double-directional radio channel estimation at 2GHz for high speed vehicular mobiles - experimantal results," in *Proc. Wireless Personal Multimedia Communications, WPMC'01*, (Aalborg, Denmark), Sept. 2001.

[141] T. Pollet, M. V. Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *IEEE Trans. Comm.*, vol. 43, pp. 191–193, Feb. 1995.

[142] P. K. Remvik and N. Holte, "Carrier frequency offset robustness for OFDM systems with different pulse shaping filters," in *Proc. IEEE GLOBECOM-97*, (Phoenix, AZ), pp. 11–15, 1997.

[143] Y. Li and L. Cimini, "Bounds on the interchannel interference of OFDM in time-varying impairments," *IEEE Trans. Comm.*, vol. 49, pp. 401–404, March 2001.

[144] G. Matz, *A Time-Frequency Calculus for Time-Varying Systems and Nonstationary Processes with Applications.* PhD thesis, Vienna Univ. Technology, Nov. 2000, (http://www.nt.tuwien.ac.at/dspgroup/tfgroup/doc/psfiles/GM-phd.ps.gz).

[145] W. Kozek and A. F. Molisch, "On the eigenstructure of underspread WSSUS channels," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, (Paris, France), pp. 325–328, April 1997.

[146] G. H. Golub and C. F. Van Loan, *Matrix Computations.* Baltimore: Johns Hopkins University Press, 3rd ed., 1996.

[147] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Sel. Areas Comm.*, vol. 17, pp. 461–471, March 1999.

[148] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs (NJ): Prentice Hall, 1993.

[149] L. L. Scharf, *Statistical Signal Processing.* Reading (MA): Addison Wesley, 1991.

[150] M. Wax and T. Kailath, "Efficient inversion of Toeplitz-block Toeplitz matrix," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1218–1221, Oct. 1983.

[151] R. A. Wiggins and E. A. Robinson, "Recursive solution to the multichannel filtering problem," *J. Geophys. Res.*, vol. 70, pp. 1885–1891, April 1965.

[152] S. M. Kay, *Modern Spectral Estimation.* Englewood Cliffs (NJ): Prentice Hall, 1988.

[153] H. Bölcskei, *Oversampled filter banks and predictive subband coders.* PhD thesis, Vienna University of Technology, Nov. 1997.

[154] E. A. Lee and D. G. Messerschmitt, *Digital Communication.* Boston (MA): Kluwer, 2nd ed., 1994.

[155] T. Kailath, *Linear Systems.* Englewood Cliffs (NJ): Prentice Hall, 1980.

[156] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs (NJ): Prentice Hall, 1991.

[157] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," *IEEE Trans. Comm.*, vol. 49, pp. 1561–1571, Sept. 2001.

[158] P. S. 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. Comm.*, vol. 43, no. 2–4, pp. 773–775, 1995.

[159] R. F. H. Fischer and J. B. Huber, "A new loading algorithm for discrete multitone transmission," in *Proc. IEEE GLOBECOM-96*, (London, UK), pp. 724–728, Nov. 1996.

[160] J. Cavers, "Variable-rate transmission for Rayleigh fading channels," *IEEE Trans. Comm.*, vol. 22, pp. 15–22, Feb. 1972.

[161] A. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Comm.*, vol. 45, pp. 1218–1230, Oct. 1997.

[162] W. T. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Trans. Comm.*, vol. 43, pp. 2223–2230, July 1995.

[163] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Comm.*, vol. 47, pp. 844–855, June 1999.

[164] A. Forenza and R. W. Heath Jr., "Link adaptation and channel prediction in wireless OFDM systems," in *Proc. 45th Midwest Symposium on Circuits and Systems*, vol. 3, pp. 211–214, Aug. 2002.

[165] M. R. Souryal and R. L. Pickholtz, "Adaptive modulation with imperfect channel information in OFDM," in *Proc. IEEE ICC-2001*, vol. 6, pp. 1861–1865, June 2001.

[166] ETSI, "Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television." EN 300 744, V1.4.1, 2001 (http://www.etsi.org).

[167] D. Schafhuber, G. Matz, and F. Hlawatsch, "Simulation of wideband mobile radio channels using subsampled ARMA models and multistage interpolation," in *Proc. 11th IEEE Workshop on Statistical Signal Processing*, (Singapore), pp. 571–574, Aug. 2001.

[168] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits and Systems*, vol. CAS-25, pp. 772–781, Sept. 1978.

[169] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[170] I. E. Telatar, "Capacity of multi-antenna gaussian channels," Tech. Rep. BL0112170-950615-07TM, AT&T Bell Laboratories, June 1995.

[171] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[172] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1290–1301, May 2001.

[173] S. Verdu, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.

[174] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge (UK): Cambridge Univ. Press, 1999.

[175] H. Gazzah, P. A. Regalia, and J.-P. Delmas, "Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1243–1251, March 2001.

[176] Ö. Oyman, R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Tight lower bounds on the ergodic capacity of Rayleigh fading MIMO channels," in *Proc. IEEE Globecom 2002*, (Taipei, Taiwan), pp. 1172–1176, Nov. 2002.