Modeling and Estimation of Wireless Multipath Channels

-

An Application within Pilot-assisted Channel Estimation for Downlink OFDM

Morten Lomholt Jakobsen

MSc Thesis
September 1st 2008 - June 2nd 2009
Department of Mathematical Sciences
Aalborg University, Denmark

AALBORG UNIVERSITY

Department of Mathematical Sciences
Fredrik Bajers Vej 7 G
9220 Aalborg Ø
http://www.math.aau.dk

Title:

Modeling and Estimation of Wireless Multipath Channels
- An Application within Pilot-assisted Channel
Estimation for Downlink OFDM

Project period:

September 1st 2008 - June 2nd 2009

Semesters:

MAT5-MAT6

Author:

Morten Lomholt Jakobsen

Supervisors:

Kasper K. Berthelsen
Bernard H. Fleury

Number of copies: 10

Report - number of pages: 121

Deadline: June 2nd 2009

Morten Lomholt Jakobsen

Abstract

This thesis investigates estimation of time-vaying multipath propagation delay parameters in a channel estimation framework for communication systems using orthogonal frequency-division multiplexing. When using state-of-the-art channel estimation algorithms which are robust to lack of a priori channel knowledge, an expensive overhead of pilot symbol data must be introduced. In order to decrease this overhead, an approach is to use a channel estimation algorithm exploiting (or presupposing) a priori knowledge on time-varying multipath channel model parameters. That is, a *parametric model* of the wireless multipath channel is invoked. The key-parameters required/presupposed by the channel estimation algorithm are the *time-varying propagation delays* of the parametric channel model. Hence, the practical applicability of this algorithm approach depends critically on whether a low complexity and sufficiently accurate *delay estimation module* can be successfully included in the receiver. The objective of this thesis is to achieve insight on whether this particular algorithm approach is realistic or not.

In recent literature, a number of methods for multipath propagation delay estimation have been proposed. A particular method proposed in literature, known as the ESPRIT algorithm, is investigated in this thesis. It is subsequently demonstrated, that the particular performance of this method depends critically on the channel model assumed.
Specifically, a parametric channel model may or may not take into account the time-varying behavior of the multipath propagation delays. If not, the channel model reflects a static mobile receiver in an equally static environment. This is clearly not a very realistic scenario. Anyway, assuming a static channel model allows for satisfactory estimation accuracy of the propagation delays, and hence, satisfactory system performance. Assuming on the contrary, a channel model taking more reasonably into account the dynamic behavior of the multipath propagation delays, the conclusions are refashioned indeed. The method investigated, i.e the ESPRIT algorithm, completely fails to supply satisfactory delay estimation accuracy. Accordingly, we propose to use (or develop) algorithms which are applicable in the more realistic scenario, where the multipath propagation delays exhibit dynamic behavior.

Inspired by the above conclusion, propagation delay estimation methods other than the ESPRIT algorithm have been investigated in this thesis - namely in a survey alike algorithm study. Preliminary results indicate a promising applicability of a particular method, however, further investigation is needed since only incipient knowledge and superficial implementations have been attained during the project work leading to the present thesis. The method is referred to as the sequential beamforming algorithm (SBA) and indeed it shows promising results in terms of performance, robustness and computational complexity.

Dansk Resumé (kort version)

Den forhåndenværende rapport omhandler estimation af tidsvarierende signal forsinkelser indenfor kommunikations systemer, som kræver estimation af det aktuelle udbredelsesmedie, også kaldet den trådløse kanel. Algoritmer anbefalet i litteraturen, som er robuste overfor manglende viden om den trådløse kanal, kræver imidlertid overførsel af uhensigtsmæssigt store mængder redundant data - såkaldte pilotsymboler. En mulig måde at undgå dette problem, er ved at benytte algoritmer som antager forhåndsviden om visse parametre i en såkaldt kanalmodel. Netop signal forsinkelserne spiller en helt afgørende rolle. Den praktiske anvendelighed af en sådan algoritme afhænger direkte af om et tilpas simpelt og tilpas præcist estimations modul kan implementeres i de trådløse modtagere. Formålet med denne rapport er at opnå indsigt i om den påtænkte algoritmeløsning er realistisk eller ej.

Acknowledgements

This project work has been financially supported by Infineon Technologies Denmark A/S, and the project proposal was suggested by Christian Rom, PhD, Infineon Technologies Denmark A/S. I will like to thank Infineon Technologies Denmark A/S for their support and Christian Rom for all the inspiration and motivation and the practical aspects provided during the project period.
The entire project work has been jointly carried out in coorporation with Department of Electronic Systems, Section for Navigation and Communications Engineering, Aalborg University. During the project period I have been working together with a fellow student, Kim Laugesen, Department of Electronic Systems, Aalborg University. I will like to thank him for the last ten months of daily coorporation. Furthermore, I will like to express my gratitude for the weekly correspondence with Gunvor Elisabeth Kirkelund, PhD candidate, Department of Electronic Systems, Aalborg University.

In particular, I will like to express my sincere appreciation for the almost daily correspondence with Carles Navarro Manchon, PhD candidate, Department of Electronic Systems, Aalborg University. His guidance, comments, patience and extraordinary amount of time spent in favor of this project work have been indispensable.

Finally, I would like to thank Bernard H. Fleury, Dr. Sc. Techn., Department of Electronic Systems, Aalborg University, and Kasper K. Berthelsen, PhD, Department of Mathematical Sciences, Aalborg University. They have served as my main supervisors and I will like to thank them both for their thorough supervision, guidance and patience during the project period.

Morten Lomholt Jakobsen
Aalborg, June 2009

Contents

Notations and Symbols

This page provides the reader with a quick insight to the notational conventions and symbols used throughout the report.

Functions

A function f which maps from the set V to the set W is written commonly as $f : V \to W$. By $f(x_1, \ldots, x_n; c_1, \ldots, c_m)$ we refer to a function in n variables and with m parameters, e.g.

$$f\left(x; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

is the Gaussian probability density function in the single real variable x and with two parameters: the mean $\mu \in \mathbb{R}$ and the variance $\sigma^2 > 0$.

Vectors and Matrices

Vectors are recognized as boldface lowercase letters, e.g. $\boldsymbol{x}$, $\boldsymbol{w}$ or $\boldsymbol{s}$. Vectors are always considered as columns by default. Matrices are recognized as boldface capitalized letters, e.g. $\mathbf{X}$, $\mathbf{A}$ or $\mathbf{R}$.

List of symbols and abbreviations

cdf	Abbreviation for *cumulative distribution function*
pdf	Abbreviation for *probability density function*
i.i.d.	Abbreviation for *independent and identically distributed*
$\mathbb{N}$	The set of natural number $\{1, 2, 3, \ldots\}$
$\mathbb{Z}$	The set of integer number $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$
$\mathbb{R}$	The set of real numbers
$\mathbb{C}$	The set of complex numbers
$\mathbb{1}[\cdot]$	Indicator function
i	The imaginary unit, i.e. $i = \sqrt{-1}$
$:=$	Left hand side defined as right hand side, e.g. $f(t) := t^2 + 1$
$\equiv$	Constantly equal to, e.g. $f(t) \equiv 1$
$*$	Convolution operator
$\mathbb{E}[\cdot]$	Expected value operator
$\mathbb{V}\text{ar}[\cdot]$	Variance operator
$\mathbb{C}\text{ov}[\cdot, \cdot]$	Co-variance operator
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian distribution of dimension d
$\mathcal{U}(a, b)$	Uniform distribution with pdf $f(\cdot) = \mathbb{1}[a \leq \cdot \leq b](b-a)^{-1}$
$\text{Pois}(\mu)$	Poisson distribution with mean parameter μ
$\text{Expo}(\lambda)$	Exponential distribution with rate parameter λ
$\sim$	Distributed as, e.g. $X \sim \mathcal{N}(\mu, \sigma^2)$
$(\cdot)^\top$	Matrix or vector transpose
$(\cdot)^*$	Complex conjugation of all entries in matrix or vector
$(\cdot)^H$	Hermitian (conjugate transpose) of matrix or vector
$\text{tr}(\mathbf{A})$	Trace of square matrix $\mathbf{A}$
$\det(\mathbf{A})$	Determinant of square matrix $\mathbf{A}$

Chapter 1

Introduction

The need and desire for communication systems to support even more features, and still at even higher data rates, are truly notable today. Cell phones are capable of accessing the internet and thereby used to obtain or download specific web contents, e.g. music, video streams or documents. Cell phones also hold the ability to function as localization (GPS) devices and laptops can work as service providers for other laptops and cell phones through (wireless) local area networking.
The number of mobile users has increased rapidly during the last decade, and most people are carrying multiple devices, e.g. phones, laptops, and personal digital assistants (PDA's). In fact, *cellular systems* have experienced exponential growth over the last decade **[1]**, and the mobile devices involved are used in several distinct matters - including personal, business and pure entertainment. The principle of a cellular system is grounded on a centrally located non-mobile device (often referred to as the *base station*) providing services to a number of mobile users within its range of operation (the *cell*). In urban areas, cellular systems are densely deployed and during cell phone conversations, one or more cellular systems are utilized to exchange information. In general, exchange of information may be performed in terms of mobile unit to base station communication, mobile unit to mobile unit or perhaps base station to *all* mobile units within the cell. The latter version of communication is referred to as *broadcast* transmission. Reception of television and radio signals are typical examples of one-way communication schemes using base station broadcast.

With a cellular system in mind, consider more generally a situation where certain entities desire to exchange information. A schematic illustration is given in terms of a transmitter, a communication or transmission medium and a receiver, see **Figure 1.1**. The transmitter may for instance be a person talking in a cell phone, a laptop transferring data files or an antenna array broadcasting television signals. The communication medium may be a copper wire, fibre optics, a digital versatile disc (DVD) or the *wireless medium*[1]. The receiver may be a cell phone, a television, a

---

[1] The wireless medium is essentially the entire physical environment, including the atmosphere, buildings, trees, cars and everything that impact the propagation conditions and propagation pat-

laptop or any device intended to obtain the signals provided by the transmitter. In **Figure 1.1**, the input signal is denoted by x and the output signal is denoted by y. For any given system, the relationship between the received y and the transmitted x depends on the communication medium. From this point and on we will refer to any medium applicable for communication as a *channel* and in general we may think of a channel as a '*black box*' that we can only adapt to - not control. Different channels inherit different properties depending on materials, physical lengths and so on. Since we cannot control the channel, it is challenging to build and adapt systems such that two entities are able to communicate across a given channel. The challenge becomes even more demanding when certain requirements are imposed on the system, for instance in terms of high data rates and with a large number of users operating simultaneously. These challenges illustrate just a few parts of the very essentials of *communication systems engineering*, see for instance **[3, Chapter 1]** for further introduction.

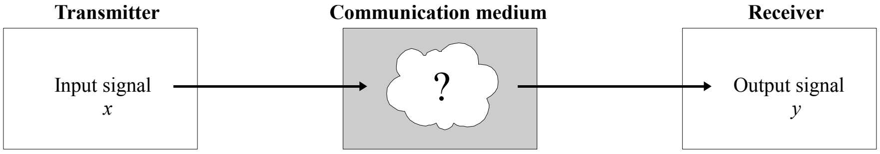


Figure 1.1: Communication system with a transmitter, a medium that information is sent through and a receiver. The communication medium is the physical tool that allows for information to be exchanged.

Now, consider a wireless channel connecting a transmitter with a mobile receiver. At time t_1, the transmitter launches a signal x into the wireless channel and at time $t_1 + \tau_1$, the receiver observes the signal y. How to describe the relationship between y and x? At time t_2 the transmitter launches the exact same signal x, and at time $t_2 + \tau_2$ the receiver observes a signal, however, now the signal is different from y. This phenomenon may occur because the mobile receiver has moved since the previous transmission and the properties of the wireless channel have changed. How can we describe such changes?

As mentioned in **[3, Section 1.4]**, it is convenient to construct *mathematical models* which reflect the most important characteristics of the channel. Some models are more appropriate than others, but all models are limited in their capabilities, and essentially, *all models are wrong*! It is important to realize that any model is constructed with a specific purpose. Thus, a model may describe one physical phenomenon quite well, while completely failing to even imitate another. However, if the model is not intended to describe the second phenomenon, then it should not be put into the context of this phenomenon either. Two different models intended to describe the same phenomenon are perfectly comparable, and which to make use of depends on individual simplicity, accuracy, computational complexity, etc.

---

terns of electromagnetic waves. Communication performed across the wireless medium is referred to as *unguided* transmission and when performed across wires or fibres, the transmission is called *guided* **[2, Chapter 2]**.

In **Chapter 3**, we investigate mathematical models of wireless channels, in particular, models of so-called *multipath channels*. Today, such channels and their modeling are of growing interest since even higher data rates are required in modern communication systems. In order to obtain the data rates demanded, consecutive information symbols are transmitted closely spaced in time and the symbols can therefore easily happen to interfere with one another. This interference emerges from delayed (scattered) signal components associated with different symbols arriving on top of each other. When the duration of each information symbol is very short even tiny delays introduced at earlier transmitted symbols can cause such symbol interference. Several techniques exist for mitigation of deteriorating effects such as *inter-symbol interference* (ISI), which is the common name for the phenomenon just briefly described. Primarily, ISI is introduced when communication is performed across multipath channels, because these channels will to some degree spread or smear the transmitted signals in time. At the receiver side this causes the incoming information symbols to superimpose in time and may introduce severe signal distortion. *Orthogonal frequency-division multiplexing* (OFDM) is a digital discrete-time implementation of *multicarrier modulation* (ISI circumventing technique treated in detail in **Chapter 4**) and has recently been applied in several applications and standards[2] **[1, 4]**. The basic ideas of this multicarrier modulation technique have been known for half a century, however, at the time of its invention it was too complex for actual system implementations.

1.1 Long Term Evolution

The *third generation partnership project* (3GPP) is a global standards-developing organization for mobile technologies. Today's most widely deployed cellular standards are specified by the 3GPP **[5, Chapter 1]**. One of the ongoing developments by 3GPP is an evolution of the current, third generation mobile communication system (3G). Partly, the development is concerned with evolved *radio access* - referred to as the *long-term evolution* (LTE). Evolved *packet access* is concerned in the development too, however, this subject is non-essential for the scope of this project. LTE is presumed to initiate deployment by 2009-2010, and the objective is to provide a long term technology, competitive for a decade and possibly beyond. Citing **[5]**: *'The 3GPP Long-Term Evolution is intended to be a mobile-communication system that can take the telecom industry into the 2020s'.*

To meet and accomplish this long term competitiveness the standardization is not restricted by previous work in the 3GPP, i.e. LTE need not be backward compatible. This is highly advantageous since radio interfaces can be designed from scratch without restrictions from designs and implementations of the late 1990s. However, several other requirements are present, including spectrum flexibility due to ascending saturation of the electromagnetic spectrum (frequency band allocation for new applications and increased number of operators). Rigorous demands on data rates are imposed too, especially at the *cell edges*, and in the ability to apply multi-antenna

---

[2] Including for instance, European digital audio broadcasting (DAB) and IEEE 802.11a WLAN.

technologies (*spatial multiplexing*). A communication system using multiple antennas at both transmitter and receiver side is referred to as a multiple-input multiple-output (MIMO) system. Much research attention is payed to MIMO systems today, since the use of multiple antennas can increase system coverage, received *signal-to-noise ratio* (SNR) and *channel capacity* **[5, Section 12.1, 14.4]**, **[1, Chapter 10]**.

1.1.1 LTE with Downlink OFDM

The OFDM multicarrier modulation technique holds a number of key features, including robustness against multipath signal distortion, high achievable peak data rates and flexibility in usage of *bandwidth*. For these reasons, OFDM has been chosen as the LTE *downlink* (base station to mobile terminal) transmission scheme. In order to recover the transmitted information at the receiver side, *channel estimation* is needed to be able to undo[3] the effects introduced by the wireless medium. Estimation of the channel (treated in **Chapter 5**) is performed by transmitting so-called *pilot symbols*, which are symbols known to both transmitter and receiver. The pilot symbols are distributed in both time and frequency according to certain patterns and this technique is commonly referred to as *pilot-assisted channel estimation* (PACE). In **[4]**, a variety of existing state-of-the-art channel estimation algorithms for OFDM are studied. Conclusions are that estimators of optimal performance exist, however, the computational complexity is relatively excessive and furthermore the correlation properties of the channel and the noise statistics need to be known. Suggestions are to put more investigation on channel and noise statistic estimators and investigate their robustness against estimation errors and the corresponding impact on overall system performance.

In **[6]**, several existing channel estimation algorithms are presented in a unified framework, including a number of performance comparisons from simulation studies. Conclusions from **[6]** partly constitute the motivation for the project work leading to the present thesis. In **[7]**, two particular channel estimation algorithms are proposed for LTE. The conclusions together with future work suggestions of **[7]** constitute the main motivation for this project. In the following, we introduce the background material needed to present the problem statement for this project.

1.2 OFDM in Brief

The basic idea of OFDM, and multicarrier modulation in general, is to modulate the information symbols intended for transmission onto a large set of subchannels (in frequency). On each subchannel, only a small amount of bandwidth is available, thus prolonging the symbol duration. Hence, in contrast of transmitting symbols consecutively in time across a single channel of large bandwidth (i.e. symbol duration is small), the concept is to transmit the symbols simultaneously in time across separate frequency channels of narrow bandwidth. Think of having ten 1Mbps internet connections instead of a single 10Mbps connection. Even though the transmission rate on each subchannel (1Mbps connection) is low, the overall transmission rate

---

[3] The process of undoing or removing adverse channel effects is called *channel equalization*.

(10Mbps) is maintained since there are ten connections working in parallel. Introducing this prolonged symbol duration on each subchannel causes enhanced robustness against multipath signal distortion, i.e. inter-symbol interference (ISI). In fact, ISI can be completely eliminated in OFDM through the use of a so-called *cyclic prefix* (explained in **Section 4.1.1**). The process of modulating the information symbols onto the different subchannels is swift and digitally performed by use of *discrete-time fast Fourier transform* (DFT) operations. Half a century ago when multicarrier modulation was invented there was no cheap and fast hardware implementations of the DFT. Hence, to build or realize a multicarrier system at that time required a separate modulator (transmitter side oscillator) and a separate demodulator (receiver side oscillator) for every single subchannel to be used. With N being the number of subchannels for the system, both transmitter and receiver would therefore have to physically comprise N independent oscillators - implying huge devices and massive amounts of power consumption.

1.2.1 Overlapping Subchannel Spectrum

In the frequency domain, one can choose to distribute the center frequencies of the N subchannels sufficiently distant, thereby ensuring the bandwidth (range of frequencies[4]) of any two adjacent subchannels to be practically non-overlapping. In this way no interference between the subchannels will occur (which is preferable of course), however, it will make the system occupy a relatively wide spectrum or large bandwidth. The spectral efficiency can be improved by allowing the spectrum of adjacent subchannels to overlap. By doing so, in order to demodulate the signal at the receiver side, the subchannels must be *orthogonal* in a certain sense (defined in **Section 4.1.2**). In **Figure 1.2** the principle of non-overlapping spectrum versus overlapping spectrum is illustrated.

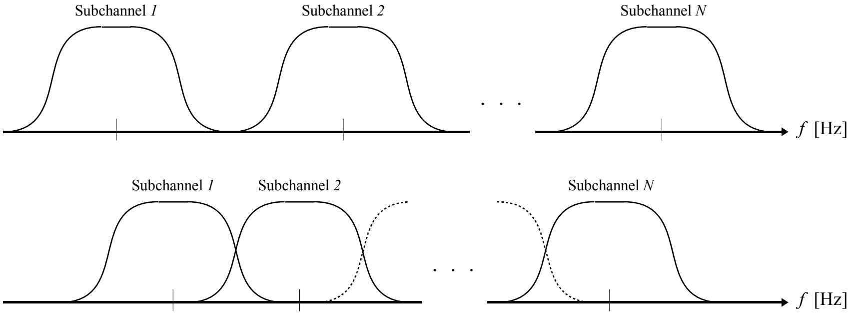


Figure 1.2: Frequency domain view of non-overlapping subchannels (top) and when the spectrum of adjacent subchannels are overlapping (bottom).

Since a multipath channel is dispersive in delay (i.e. it smears the transmitted signals in time), the channel is said to be *delay-dispersive*. A signal being spread in delay is

---

[4] See **[8]** for a splendid discussion on the connection between *reality* and *models*. Furthermore, a lucid definition of the concept of bandwidth is provided.

a time domain phenomenon. When translating or recasting delay dispersiveness in the frequency domain it corresponds to so-called *frequency-selectivity* of the channel. That is, the response of the channel to signal components on different frequencies is non-constant and the channel may strongly attenuate some frequency components while passing other frequency components practically unaffected. In **Figure 1.3**, a fictitious channel frequency response is jointly illustrated with a set of overlapping subchannels in an OFDM system. In the figure we have denoted the center frequencies of the N subchannels by $f_1, f_2, \ldots, f_N$, and these are referred to as *subcarrier frequencies* or simply as *subcarriers* for short. With a slight abuse of terminology the terms 'subchannel' and 'subcarrier' are often used interchangeably.

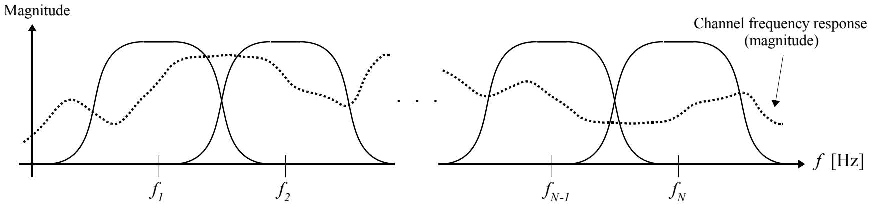


Figure 1.3: Frequency domain view of overlapping OFDM subchannels (solid lines) with the dotted line representing a fictitious frequency response of a multipath (delay-dispersive/frequency-selective) channel.

By choosing the number N of subcarriers sufficiently large, the frequency response of the multipath channel can be considered approximately constant (flat) on each individual subchannel. When transmitting information symbols across one of the N subchannels, the symbol distortion introduced on this particular subchannel therefore amounts to a single multiplication by a complex attenuation coefficient. This complex attenuation is the channel frequency response at the corresponding subcarrier frequency. Since there are N subchannels there are N of such channel attenuation coefficients and these need to be estimated to recover all the transmitted information symbols. This is the earlier mentioned channel estimation task that needs to be performed and since the number N is required to be large (to ensure a flat response on every subchannel) the estimation problem is high-dimensional in nature.

1.3 State-of-the-art Channel Estimation

In **[9, 10, 11]** several techniques for channel estimation in OFDM systems are suggested. All methods are based on pilot symbol transmissions with different one and two-dimensional (frequency and time domain) filtering techniques included. The basic approach is to impose a certain parametric structure on the multipath channel in order to lower the dimensionality of the estimation problem. Some existing channel estimation algorithms require knowledge of certain system parameters which are unknown in general, e.g. channel correlation properties or signal-to-noise ratio. In such cases, these parameters can be substituted by fixed nominal values ensuring robustness against correlation mismatch and ensuring optimized performance within situations of high signal-to-noise. Other algorithms assume exact knowledge of the time-varying parameters describing the channel model. With such knowledge excellent estimation performance can by achieved, however, these time-varying channel parameters are in general unknown as well. The channel model parameters needed can be estimated from pilot symbol obtained data but such estimations are typically computationally demanding since high estimation accuracy is needed. In **[6, 7]**, it is shown how even small estimation errors in the channel model parameters can notably decrease performance of the channel estimation algorithms.

1.4 Problem Statement

A relatively simple and intuitively appealing parametric channel model reflecting a multipath channel scenario may seek to imitate phenomenons well-known from acoustics. With such a model the different signal copies arriving at the receiver (each copy individually delayed and attenuated), are to a certain extent, modeled as peaks located on a time axis, where the time represents delay. **Figure 1.4** illustrates a delay axis with five echoes corresponding to a multipath channel with a five-fold peak response when exposed to an impulse launched by the transmitting entity. To illustrate this, think of a person shouting out loud in a large and empty hall. A certain number of dominant and distinguishable reflections of the shout will be captured by the human ear. This phenomena is directly inherited in the multipath channel model.

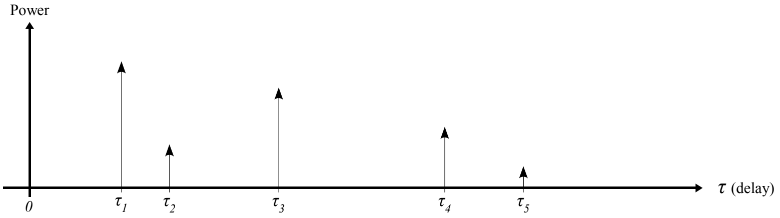


Figure 1.4: Schematic presentation of a five-fold response in a multipath channel. In acoustics, such a scheme is referred to as an *echogram*.

Each of the five arrows in **Figure 1.4** represents an individually delayed and attenuated signal copy, where the value of τ_l for $l = 1, 2, 3, 4, 5$, represents the delay at which the l'th copy arrives. The length of each arrow represents the power contents in each signal copy, i.e. the shorter the arrow the more attenuated the signal copy is. As mentioned in the beginning of this chapter, transmitting the signal x at two different points in time will typically not result in the same received signal y. The properties of a multipath channel are time-varying due to mobility of the entities involved. This means that the echoes in **Figure 1.4** will constantly move to new positions on the delay axis and also the magnitude of the arrows will change - all as a function of time. The number of echoes may change too, for instance if a signal propagation path is suddenly blocked by an obstructing object, again induced by dynamics in the environment. Specifically, since the delay parameters are time-varying we should write $\tau_l(t)$ to emphasize this fact.

When a model as above is assumed for a specific multipath channel scenario, it turns out to be of major importance to locate the exact delays $\tau_l(t)$, at which the different signal copies arrive **[6, 7]**. The frequency response of the channel depends on the delay parameters and if these are not located with sufficient accuracy, the effects of the channel cannot be compensated in a satisfactory manner and system performance will degrade. Existing state-of-the-art channel estimation algorithms supplied with exact knowledge of the delays hold excellent performance and on acceptable computational complexity. However, when the delays are not perfectly known, which they are not in practice, the performance degrades. Robust channel estimation algorithms exist which do not require knowledge of the delays, however, these algorithms suffer from an irreversible performance degradation due to the robust setup. Hence, when considering a trade off between performance and complexity it is suggested in **[7]**, to investigate if sufficiently accurate and low complexity algorithms can be developed for estimation of the time-varying multipath delay parameters. If so, such an algorithm could be utilized to feed the existing channel estimation algorithms with the information and time-varying parameters required. *The aim of this project is to clarify whether an algorithm, able to outperform the robust channel estimation solutions, can be developed.*

In **[12]**, a large amount of inspiration can be obtained since a more or less complete proposal of such a delay detection and estimation algorithm is provided. However, several aspects considered in the paper rely on purely theoretical backbones and do not seem to be realizable. For instance, a *minimum description length* criteria **[13]** is utilized to detect the number of channel echoes, however, this method is not appropriate in practice. Also, the simulation results presented in the paper are not concerning the actual performance of the proposed algorithm. It is justified that the proposed algorithm (in some cases) is able to detect and track the signal echoes but for the performance evaluation, exact parameter knowledge is feeded to the algorithm instead of the estimated values. Despite these issues the paper presents an approachable and lucid algorithm proposal that will be used for inspiration purposes during the forthcoming chapters.

1.5 Thesis Outline

The remaining part of the thesis is organized in three parts.

Part I

The first part deals with multipath channel modeling, multicarrier modulation (OFDM in particular) and the closing section concerns pilot assisted channel estimation in communication systems using OFDM. Basically, the entire first part introduces a wide range of background knowledge. This knowledge is useful to keep in mind when engaging the main objective of this project - detection and estimation of the time-varying delay parameters in the multipath channel model assumed. That is, in order to fully understand and analyze the project objective in its very particularities, it is considered crucial to be comfortable with the concepts described in the three chapters comprising the first part of this thesis.

Part II

The second part encompasses specifically the main objective of this project concerning delay detection and estimation. Initially, a state-of-the-art alike survey is conducted in order to initiate the problem investigation with a perspective as broad as possible. Simulation studies are included to support the primary investigations carried out and both reference scenarios and self-imposed configurations are considered. Part II is closed after discussion, conclusion and future work sections.

Part III

The third and final part of this thesis contains a number of appendices that may have already been consulted during the reading of the two primary parts of the thesis. All relevant references introduced during the different chapters are listed in the end.

Part I

OFDM and Pilot-assisted Channel Estimation

Chapter 2

Basic Concepts

This chapter provides the reader with a variety of fundamental concepts used extensively throughout the entire report. Most concepts introduced in the following could easily (and would typically) have been assumed known in advance or maybe reported in appendices. However, we present the material at this point to allow for improved readability later on and also in order to settle notational conventions.

2.1 Discrete-time Signals

Discrete-time signals (deterministic or random) are often represented mathematically as *sequences*. A real or complex-valued sequence is as mapping whose domain is the set of integer numbers $\mathbb{Z}$, or equivalently, the set of natural numbers $\mathbb{N}$. A discrete-time signal (i.e. a sequence) is denoted by

$$x = \{x_n : n \in \mathbb{Z}\},$$

where each element x_n is complex or real. In some cases it may be convenient to think of a discrete-time signal as a vector of infinite length, but in practice we consider signals where only a finite number of elements are non-zero. Accordingly, we will refer to the following N-dimensional vector

$$\boldsymbol{x} = \left[x_1, x_2, \dots, x_N\right]^\top, \tag{2.1}$$

as a discrete-time signal of length N. That is, instead of representing (2.1) as the sequence

$$x = \{\dots, 0, 0, x_1, x_2, \dots, x_N, 0, 0, \dots\},$$

we represent it as a vector $\boldsymbol{x} \in \mathbb{C}^N$. Our formal definition of a *discrete-time signal* therefore refers to a mapping

$$f : D \to \mathbb{C}^k, \quad D \subseteq \mathbb{Z}^m, \quad k, m \in \mathbb{N}.$$

Most commonly **[14, Section 1.2]**, **[15, Section 13.1]**, a discrete-time signal $\boldsymbol{x}$ is obtained by *regular sampling* of a continuous-time signal, i.e. the signal $\boldsymbol{x}$ with entries x_n is obtained from the continuous-time signal $x(t),\ t \in \mathbb{R}$, according to

$$x_n := x\big(nT_s\big), \quad n \in \mathbb{Z}, \quad T_s > 0,$$

where T_s is the *sampling time*. **Figure 2.1** illustrates the procedure of obtaining a discrete-time signal from regular sampling of a real-valued continuous-time signal.

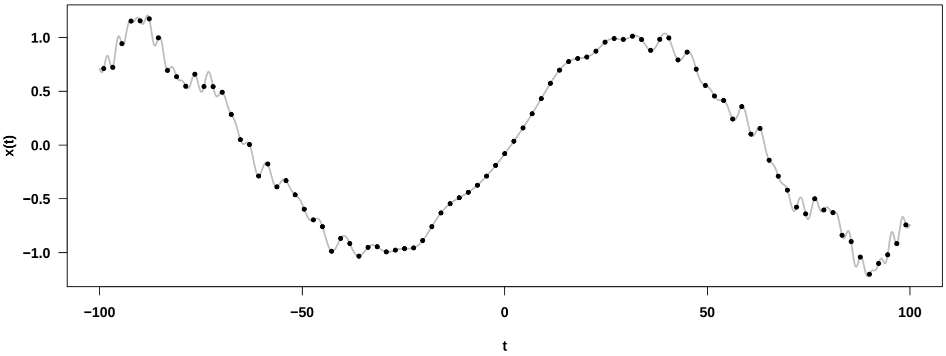


Figure 2.1: A fictitious continuous-time signal $x(t)$ sampled at regularly spaced points in time.

2.1.1 Signal Power

Most signals encountered in practical applications are random in nature (noise corrupted), e.g. future values cannot be determined exactly from present observations or repeating an experiment several times may produce different results. The mathematical tool to describe such signals is grounded on probabilistic statements in terms of *random sequences* (a formal definition is given in **Appendix A**). Typically, a *realization* $\{y_1, y_2, \ldots, y_K\}$ of a random signal, viewed as a discrete-time sequence, is not of finite *energy*, i.e.

$$\lim_{K\to\infty} \sum_{k=1}^{K} |y_k|^2 = \infty,$$

where we utilize a limit statement to describe the procedure of observing more and more data (in practice we cannot observe, store or handle an infinite amount of data points). A discrete-time sequence of infinite energy does not possess a *Fourier transform*, i.e.

$$\sum_{k=1}^{\infty} y_k \exp(-i2\pi k f) = \lim_{K\to\infty} \sum_{k=1}^{K} y_k \exp(-i2\pi k f), \quad f \in \mathbb{R},$$

does not converge in general. However, the *average power* of a random signal is usually finite **[14, 16]**, where the (probabilistic) average is taken with respect to all possible realizations of the random signal. Thus, the properties of random signals are often summarized in terms of averages, e.g. *autocorrelation* or *auto-covariance* sequences, since these often possess a Fourier transform.

2.2 The N-point Discrete-time Fourier Transform

Let $\boldsymbol{x} = \left[x_1, x_2, \dots, x_N\right]^\top$ be a real or complex-valued signal of length N. The N-point *discrete-time Fourier transform* (DFT) of $\boldsymbol{x}$ is defined as

$$\widehat{x}_k := \frac{1}{\sqrt{N}} \sum_{n=1}^{N} x_n \exp\left(-i2\pi(k-1)\tfrac{n-1}{N}\right), \quad k = 1, 2, \dots, N, \tag{2.2}$$

i.e. the N-point DFT produces a new signal $\widehat{\boldsymbol{x}} = \left[\widehat{x}_1, \widehat{x}_2, \dots, \widehat{x}_N\right]^\top$ of length N. The $\widehat{x}_k$'s characterize the frequency contents of the signal $\boldsymbol{x}$, and given $\widehat{\boldsymbol{x}}$ the original signal can be reconstructed. The N-point DFT is a linear mapping and thus the transform can be represented by a matrix. To verify the linearity of the DFT one can simply carry out the matrix-vector multiplication in (2.3) and compare with the definition in (2.2). To ease notation we introduce

$$\omega := \exp\left(-i2\pi\tfrac{1}{N}\right),$$

and the N-point DFT of the signal $\boldsymbol{x}$ is then performed according to

$$\begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \\ \vdots \\ \vdots \\ \widehat{x}_N \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_N \end{bmatrix}. \tag{2.3}$$

We denote the $N \times N$ matrix in (2.3) by $\mathbf{F}$ and refer to its elements as $\mathbf{F}(n,m)$. By indexing rows and columns from 1 to N as usual, we identify that

$$\mathbf{F}(n,m) = \frac{1}{\sqrt{N}} \exp\left(-i2\pi\tfrac{(n-1)(m-1)}{N}\right), \quad n, m \in \{1, 2, \dots N\}.$$

Note that $\mathbf{F}^\top = \mathbf{F}$, and therefore the *Hermitian* of $\mathbf{F}$, denoted $\mathbf{F}^H$, is simply obtained by complex conjugation of all entries. Defining $\mathbf{A} := \mathbf{F}\mathbf{F}^H$ we calculate the entry

$$\mathbf{A}(n,m) = \sum_{k=1}^{N} \mathbf{F}(n,k)\mathbf{F}^*(k,m) = \frac{1}{N} \sum_{k=1}^{N} \exp\left(-i2\pi\big((n-1)-(m-1)\big)\tfrac{k-1}{N}\right).$$

By rearranging the latter expression we see that the sum involved equals

$$\sum_{k=1}^{N} \exp\left(-i2\pi(n-m)\tfrac{k-1}{N}\right) = \begin{cases} N, & n = m \\ 0, & n \neq m. \end{cases}$$

These calculations show that $\mathbf{A} = \mathbf{I}_N$ and hence $\mathbf{F}^{-1} = \mathbf{F}^H$. Thus, we obtain the original signal $\boldsymbol{x}$ from its N-point DFT $\widehat{\boldsymbol{x}}$ by

$$\left[x_0, x_1, \dots, x_{N-1}\right]^\top = \mathbf{F}^H \left[\widehat{x}_1, \widehat{x}_2, \dots, \widehat{x}_N\right]^\top.$$

The matrices $\mathbf{F}$ and $\mathbf{F}^H$ are used extensively throughout the forthcoming chapters.

2.3 Digital Modulation

This section gives a brief introduction to a few selected digital modulation techniques that are commonly applied in communication systems today. For numerous reasons, systems currently being built or proposed for wireless applications are all digital **[1, Chapter 5]**. Cost, size, power and spectral efficiency, error correction coding, multiple access and security are just a few issues inheriting beneficial properties from digital implementations compared to analog equivalents. With *digital modulation* and *detection*, information in the form of bits (0's and 1's) is exchanged across a communication medium. Modulation consists of mapping the information bits into an analog signal intended for transmission over the channel. Detection consists of estimating the original information bits from the signal observed at the receiver side. A variety of modulation schemes exist and selection of a scheme for a particular application relies on a trade-off between conflicting desires such as high data rates, low bandwidth occupancy, low power consumptions and robustness against noise and channel effects. In the following we describe two modulation schemes that are both included in the 3GPP specifications for LTE. This section is brief and targeted, holding few details only, however, see **[1, Chapter 5]** and **[3]** for further insight on digital modulation and detection.

Two main categories of digital modulation exist:

- Amplitude/phase modulation (referred to as *linear modulation*), and
- frequency modulation (referred to as *non-linear modulation*).

We consider the former technique only, where information is encoded in the amplitude and/or phase of the analog signal intended for transmission. Three evident variants of amplitude/phase modulation exist:

- *Pulse amplitude modulation* (PAM) - information embedded in amplitude only,
- *phase-shift keying* (PSK) - information embedded in phase only, and
- *quadrature amplitude modulation* (QAM) - information encoded in both amplitude and phase.

Initially, when applying any of the above modulation techniques, the information bits are divided into blocks of length $K := \log_2 M$, where M (an integer power of 2) is the size of the so-called *symbol constellation*. The symbol constellation is a set of complex numbers $\{S_1, S_2, \ldots, S_M\}$ such that each of the symbols S_m, $m = 1, 2, \ldots, M$, uniquely corresponds to one of the 2^K possible blocks of information bits. Thus, we decide the number $M \in \{2, 4, 8, 16, 32, 64, \ldots\}$ of symbols in the constellation and create an invertible (constellation) mapping from the set $\{0, 1\}^K$ to the set $\{S_1, S_2, \ldots, S_M\}$.

2.3.1 Phase-Shift Keying

With PSK the bit information is encoded solely in the phase of the analog signal intended for transmission. This means that each constellation symbol S_m is located on a common circle in the complex plane, i.e.

$$S_m := r \exp\left(i2\pi\tfrac{m-1}{M}\right), \quad m = 1, 2, \ldots, M, \quad r > 0.$$

Thus, the symbols in the constellation are of equal modulus r (same power content) and equidistantly located on the circle. The constellation mapping is usually done by *Gray encoding*, meaning that two adjacent symbols in the constellation correspond to bit blocks only differing by a single bit (*Hamming distance* equal to one). Mistaking a constellation symbol for an adjacent one at the receiver side causes only a single bit error when Gray encoding is utilized. Gray encoded M-PSK constellations with $M = 4$ and $M = 8$ are shown in **Figure 2.2**.

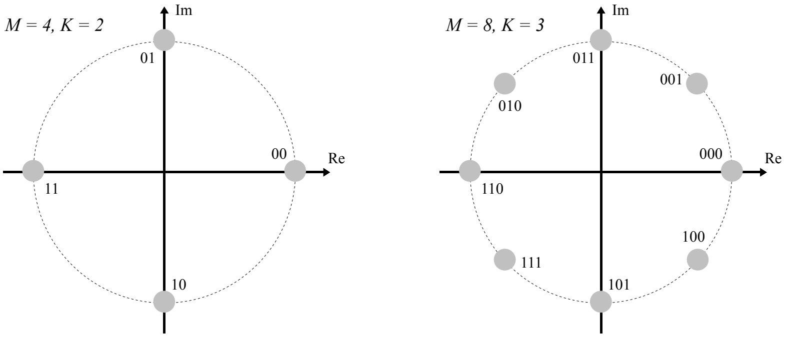


Figure 2.2: Phase-shift keying with Gray encoded constellation mappings.

When $M = 2$ the modulation scheme is referred to as *binary phase-shift keying* (BPSK) and when $M = 4$ the term *quadrature phase-shift keying* (QPSK) is used. Choosing a symbol constellation with a larger M allows for higher data rates but at the same time it makes the constellation more sensitive to deteriorating transmission effects.

2.3.2 Quadrature Amplitude Modulation

With QAM the information is embedded in both amplitude and phase of the transmitted signal. Much freedom in choosing the constellation symbols S_m is present with QAM, and the different symbols can hold different amounts of power (with PSK all symbols hold identical power content). ‘Good’ constellation mappings can be difficult to find for QAM, in particular, it is non-trivial to construct Gray code mappings. Constellations of square shape is often considered, and **Figure 2.3** depicts square shaped M-QAM constellations with $M = 16$ and $M = 64$.

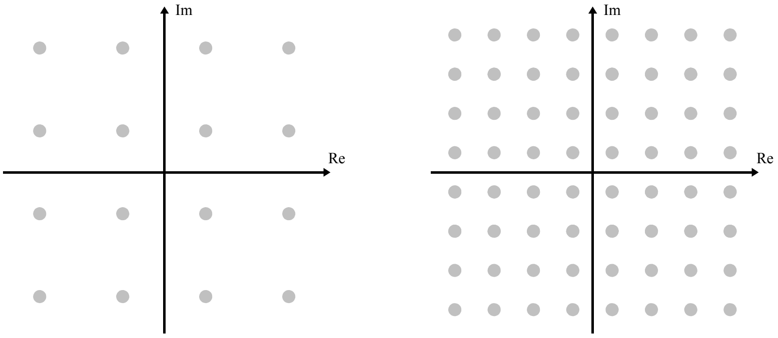


Figure 2.3: QAM square constellations with $M = 16,\ K = 4$ (left) and $M = 64,\ K = 6$ (right).

When the signal constellation is determined, the analog signal intended for transmission is constructed from *pulse shaping* of the symbols. That is, each symbol is shaped by a pulse $\Psi(t)$ of duration T_s (called the *symbol time*) implying a rate of $\frac{K}{T_s}$ bits/s. Consider a stream of information bits being appropriately mapped to constellation symbols and then pulse shaped to obtain the analog signal (referred to as a *pulse train*)

$$x(t) = \sum_{n=-\infty}^{\infty} S_{m(n)} \Psi\big(t - nT_s\big),$$

where $S_{m(n)} \in \big\{S_1, S_2, \ldots, S_M\big\}$ for each discrete time index $n \in \mathbb{Z}$. The number of bits/symbol, the symbol constellation and the choice of a shaping pulse $\Psi(t)$ determine the digital modulation scheme. We will not address the issue of choosing a particular shaping pulse. Such a discussion is too involved at this place, however, **Figure 2.4** provides an informative view of two simple pulses - a rectangular pulse and a slightly smoother version (smooth pulses hold better spectral properties).

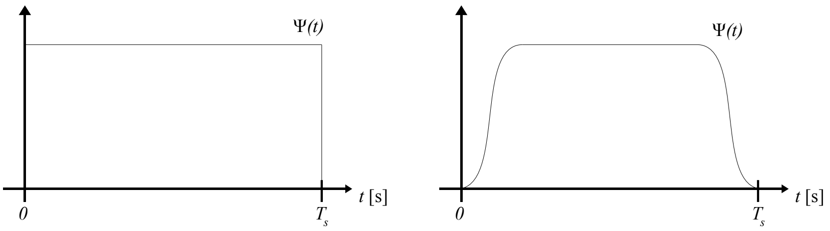


Figure 2.4: Informative examples of shaping pulses, both of duration T_s.

2.3.3 Coherent Detection

Certain applications require high data rates and symbol constellations of small sizes are not appropriate in such cases. Larger symbol constellations are more susceptible to noise and sometimes also require more advanced receiver structures. The phase of both QAM and PSK modulated signals carry information of the original bit stream. At the receiver side it is therefore crucial to be able to recover and synchronize to the phase of the transmitted signal. Increased complexity in the receiver results from phase recovery procedures, nonetheless, such alignments are needed within scenarios and applications requiring high data rates. When only low data rates are needed, symbol encoding techniques such as *differential modulation* **[1, Section 5.3.4]** (modulation with memory of prior symbols) is appropriate since no phase recovery is needed - implying simpler receiver structures and in particular no need for channel estimation. We will not elaborate further on the issues concerned with coherent demodulation and detection, only we stress the fact that coherent detection is often necessary and therefore channel estimation must be performed at the receiver side.

2.4 Dirac's Delta

In this section we present a number of details concerning the *Dirac delta*, commonly denoted $\delta(t)$. The Dirac delta is the continuous analog to the *Kronecker delta*, often (similarly) denoted by $\delta(k)$ with the variable k implicitly spelling that time is measured at discrete points and not continuously. The Kronecker delta is an indicator function located a zero, i.e.

$$\delta(k) := \mathbb{1}[k = 0].$$

Informally, the Dirac delta represents an infinitely narrow, infinitely high peak centered at zero and bounding unit area. Thus, the Dirac delta is a 'function' that is zero everywhere, except at $t = 0$ where it is infinitely large and

$$\int_{-\infty}^{\infty} \delta(t) \mathrm{d}t = 1.$$

No such function exists but formally the Dirac delta can be defined as a *distribution* or *generalized function*. Such a treatment is out of scope of this thesis, however, insight on the theory of distributions can be obtained from **[17, Appendix B]**. This appendix provides interesting aspects concerned with important operators too, e.g. the Fourier transformation on $\mathbb{R}^d$.

Alternatively, the Dirac delta may be defined as a *measure*. If so, for every $E \subseteq \mathbb{R}$ we define

$$\delta(E) := \begin{cases} 1, & \text{if } 0 \in E \\ 0, & \text{if } 0 \notin E. \end{cases}$$

Then, for any function $g : \mathbb{R} \to \mathbb{C}$ the following relationship holds

$$\int_{-\infty}^{\infty} g(t)\delta(t) \mathrm{d}t = g(0),$$

which is commonly referred to as the *sampling property* of the Dirac delta. In **Section 5.4** we shall see explicitly how this property can be appropriately utilized to mimic a sampling procedure in the frequency domain in order to facilitate certain analytical inspections.

2.4.1 Dirac's Delta as a Limit

A variety of well-known functions behave in the limit as the Dirac delta when letting certain parameters tend towards 'extreme' values. Consider for instance the pdf corresponding to a standard Gaussian distribution, and consider the behavior when letting the variance tend towards zero. Heuristically, we may define

$$\delta(t) := \lim_{\sigma \to 0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right),$$

and a few examples with selected values of σ are shown in **Figure 2.5**.

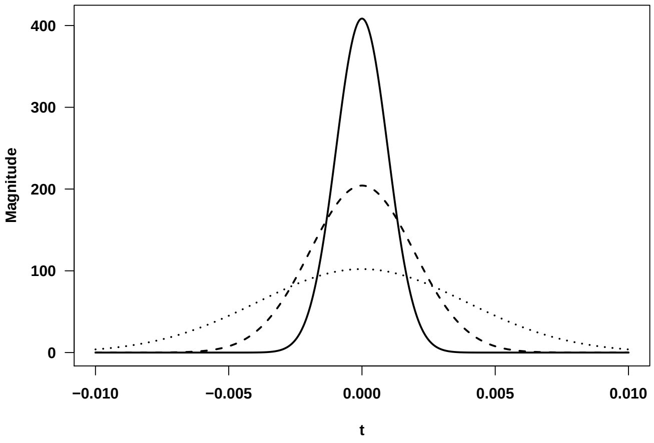


Figure 2.5: Approximations to the Dirac delta $\delta(t)$ from a Gaussian pdf with standard deviation $\sigma = 2^{-8}, 2^{-9}, 2^{-10}$ for the dotted, dashed and solid curve, respectively. Notice the narrow span on the time-axis and the wide span on the magnitude-axis.

The Dirac delta is introduced during the forthcoming chapter to facilitate the mathematical modeling of time-variant multipath channels.

Chapter 3

Multipath Channel Modeling

Consider a narrow pulse being transmitted through a *multipath channel*. Induced by a set of scatterers, the transmitted pulse will traverse multiple paths from transmitter to receiver. The received signal will therefore appear as a pulse train where each individual term in the train corresponds to different propagation paths. The first pulse received corresponds to the *line-of-sight* (LOS) component, provided that this direct path exists, and the final incoming pulse corresponds to the longest path traversed by the signal. Intermediate pulses correspond to distinct multipath components (signal echoes) associated with distinct scatterers or clusters of scatterers. A simple multipath scenario is shown in **Figure 3.1**.

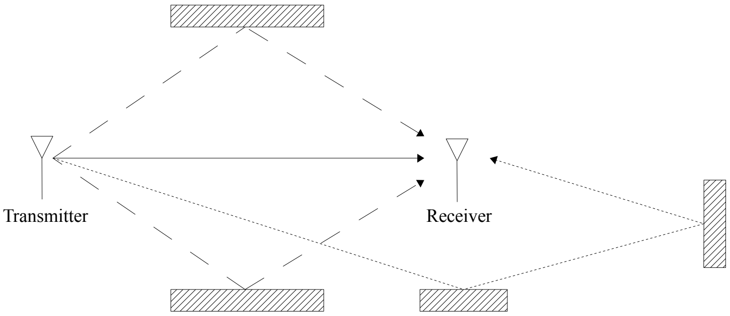


Figure 3.1: Multipath propagation scenario with a line-of-sight component (solid line) and three multipath components. Two of the multipath components arrive with equal delay (dashed lines) and the third component (tiny-dashed) arrive with a larger delay.

The properties of a multipath channel often vary with time since either the transmitter, receiver or the environment is moving. Hence, the location of reflectors and scatterers change over time and the *impulse response* of the channel is therefore time-variant. Consequently, if a moving device repeatedly transmits the exact same pulse every Δt seconds the receiver will typically experience non-identical incoming

pulse trains. The number of multipath echoes is changing from train to train and the associated delays and amplitudes varies too.
Our argument on modeling the incoming signal as a pulse train, is fundamentally grounded on approximations. During transmission of a single narrow pulse, the signal may be slightly smeared in time due to a massive amount of infinitesimal interactions from scattering, reflection and diffraction phenomenons. Hence, the incoming signal does not necessarily consists of discrete electromagnetic components but may as well constitute a continuum of contributions from the original transmitted pulse. However, we do not model this continuum - we approximate it by a set of discrete components in order to keep the model simple, while still inheriting a reasonable level of accuracy (recall the discussion about mathematical models and their purposes from **Chapter 1**).

In this chapter, we model and characterize the properties of a multipath channel based on the underlying time-varying channel impulse response. Often in literature (e.g. in **[1]**), both *narrowband* and *wideband* systems are considered since the *fading models* for each of them are important to distinguish. Due to the significant difference in bandwidth (time resolution), the wideband system will pick up fast-varying effects which are changing too fast for the narrowband system to capture. Note however, that there is no general rule spelling whether a system is wideband or narrowband and often the systems of interest are classified relative to each other. In OFDM, it is common to classify the overall system as wideband with each subchannel being narrowband (regardless whether the overall system bandwidth is 200kHz or 20MHz). Therefore, we do not seek to define what to be understood by wideband and narrowband systems since this distinction is often a relative matter. Neither will we make use of any conceptual differences between two such systems.

3.1 Multipath Signal Propagation

Consider a real *bandpass* signal $s(t)$ on *carrier frequency* f_c with equivalent *lowpass* signal $x(t)$, i.e. consider

$$s(t) = \mathrm{Re}\left\{x(t)\mathrm{e}^{i2\pi f_c t}\right\}.$$

In **Appendix B**, we briefly describe common representations of lowpass and bandpass systems. Readers who are not already familiar with these different representations may want to consult this appendix before continuing. The signal $s(t)$ is transmitted through a multipath channel and due to the inherent nature and time-varying property of the channel, the received signal is given by

$$r(t) = \mathrm{Re}\left\{\left(\sum_{l=0}^{L(t)} \alpha_l(t)x\big(t-\tau_l(t)\big)\mathrm{e}^{-i2\pi f_c \tau_l(t)}\right)\mathrm{e}^{i2\pi f_c t}\right\} + n(t). \tag{3.1}$$

The additive term $n(t)$ is a stochastic noise process which is not of primary concern at the moment, i.e. we disregard this term in the following. Instead, consider carefully the structure of the received signal $r(t)$, where t is the observation time at the receiver side. The received signal at time t is a superposition of $L(t)+1$ incoming train terms, where the l'th term

- is delayed $\tau_l(t)$ time units relative to the observation time,
- has *envelope* $\big|\alpha_l(t)x\big(t-\tau_l(t)\big)\big|$, and
- experience a *phase shift* of $-2\pi f_c\tau_l(t)$ due to the delay $\tau_l(t)$.

Hence, at each time instance the received signal depends on the following unknown entities:

- $L(t) \in \mathbb{N}$, the number of multipath components ($l = 0$ is the LOS component),
- $\big\{\alpha_l(t)\big\}_{l=0}^{L(t)}$, the associated set of complex amplitudes, and
- $\big\{\tau_l(t)\big\}_{l=0}^{L(t)}$, the set of multipath delays.

The magnitude and phase of each complex-valued coefficient $\alpha_l(t)$ depend on transmitter and receiver antennas and the directions of departure and arrival of the l'th signal component. Also, all interactions taking place during propagation across the wireless medium will impact the phase terms. Hence, the received signal also witnesses a phase shift contribution from the set of complex-valued coefficients $\big\{\alpha_l(t)\big\}$. However, in most cases this particular phase contribution is not dominant, since in (3.1) the phase shift induced from the delay $\tau_l(t)$ can be recast as

$$-2\pi f_c\tau_l(t) = -2\pi\frac{\ell_l(t)}{\lambda}, \tag{3.2}$$

where $\ell_l(t)$ denotes the length of the propagation path traversed by the l'th signal component, i.e. $\tau_l(t)$ equals $\ell_l(t)$ divided by the speed of light. Hence, if the receiver moves from the center of a sphere with radius λ (the wavelength corresponding to f_c) toward the boundary of the sphere, the l'th signal component may experience as much as an entire phase rotation. For instance with $f_c = 2$GHz the associated wavelength is $\lambda \approx 0.15$m and therefore the phase can rotate 2π if the receiver moves just 15cm. When receiver, transmitter and the environment are all moving during transmission the incoming signals will thus appear[1] as shifted in frequency. This phenomena is called a *Doppler frequency shift.* If we assume any motion in the system to happen such that the length of the l'th path can be approximated from a first order Taylor series expansion, i.e.

$$\ell_l(t) \approx \ell_l + \gamma_l t, \tag{3.3}$$

for some constant ℓ_l and some speed γ_l, then by (3.2) we may write

$$-2\pi f_c\tau_l(t) = -2\pi\frac{\ell_l(t)}{\lambda} \approx -2\pi\left(\frac{\ell_l}{\lambda} + \frac{\gamma_l}{\lambda}t\right). \tag{3.4}$$

---

[1] This is true for sufficiently narrowband signals only, since otherwise the effect of a moving transmitter or receiver (or scatters) will not be recognized as a frequency shift but as time compression or expansion.

The frequency $\frac{\gamma_l}{\lambda}$ at which the phase changes as a function of time is exactly the Doppler frequency. When only the receiver is moving and assuming it to be located far from all scatterers, indeed the approximation (3.3) is valid, and we can directly interpret the Doppler shift as follows. From the geometry in **Figure 3.2**, we see that each received signal component will appear shifted (compressed or expanded in time) by the Doppler frequency

$$\frac{\gamma_l}{\lambda} = \frac{v}{\lambda}\cos(\theta_l). \tag{3.5}$$

The receiver speed in direction of the impinging wavefront is $\gamma_l = v\cos(\theta_l)$, where

- v is the receiver speed in its own direction of motion, and
- θ_l is the angle of arrival of the signal component relative to the receiver direction of motion.

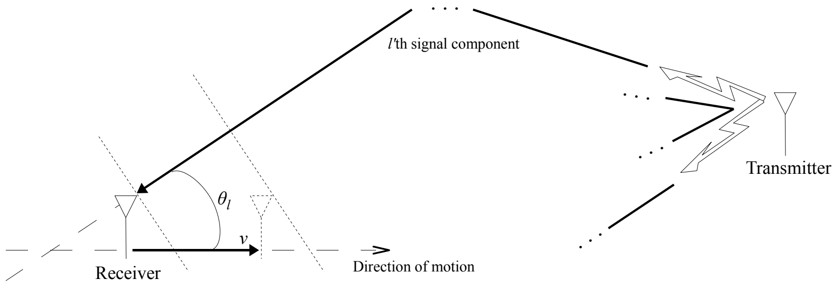


Figure 3.2: The receiver travels at speed v in a certain direction of motion. The incoming signal components appear shifted in frequency. This phenomena is called a Doppler frequency shift.

For a fixed receiver speed v, the Doppler frequency shift reaches its maximum when the receiver is moving directly toward the impinging wave. Relate this to the phenomena experienced when an ambulance is approaching fast from behind and passing straight by. The siren appears as shifted in frequency and with reference to an acoustic equivalent of the approximation in (3.3), this frequency shift is explained by the movement of the vehicle. Doppler frequency shifts of larger magnitude can appear when transmitter, receiver and the environment are all moving simultaneously. These Doppler shift considerations are crucial to keep in mind when the effects of a multipath channel are to be mimicked and interpreted from simulation studies (**Chapter 7** and **8**).

Returning our focus to the expression in (3.1) we find that each term in the sum corresponds to either a single scatter or a cluster of scatterers. Some propagation delays may appear so close that these cannot be distinguished at the receiver side due to insufficient time resolution (bandwidth). Commonly, we say that two multipath components with associated delays τ and $\widetilde{\tau}$ are *resolvable* if

$$B_x \gg \frac{1}{|\widetilde{\tau} - \tau|},$$

i.e. if the bandwidth B_x of the lowpass signal $x(t)$ significantly[2] exceeds the inverse of the delay difference. With linear modulation (e.g. PSK or QAM) the signal bandwidth is inversely proportional to the symbol duration T_s, i.e. $B_x \approx 1/T_s$. Multipath components with delays τ and $\widetilde{\tau}$ being too close cannot be distinguished at the receiver side since $x(t-\tau) \approx x(t-\widetilde{\tau})$, and hence, such components are called *non-resolvable*. Obviously, if $\tau = \widetilde{\tau}$ it is no longer a question of bandwidth - two such components cannot be separated in delay. For example, in **Figure 3.1** the multipath components marked with dashed lines are non-resolvable since they arrive with equal delay. Therefore, these two multipath components will contribute with a single term to the sum in (3.1), only. The multipath component marked with a tiny-dashed line will contribute with a term of its own since this component arrives with a delay of relatively larger magnitude and therefore can be resolved.

The above discussion on whether multipath components are resolvable or not is inconclusive and shallow. One fact concerns what actually happens in the channel and another fact concerns what can be observed by a given communication system depending on its sampling rate. These issues are recast and discussed from a frequency domain point of view in **Section 5.4**. From this perspective we are able to analyze in detail the resolution issues and we introduce the concept of *system response*.

3.1.1 Maximum Excess Delay and Channel Delay Spread

At time t, consider the integer $L(t)$ and the set of $L(t)+1$ delays $\{\tau_l(t)\}$. Without loss of generality, we assume that these delays are indexed in increasing order, i.e.

$$\tau_0(t) < \tau_1(t) < \ldots < \tau_{L(t)}(t).$$

Here, $\tau_0(t)$ corresponds to the propagation delay associated with the LOS component at time t. Consider the difference between the largest delay and the smallest delay - which is a measure of the degree of *multipath distortion* of the signal transmitted, i.e. of how much the signal is being spread in time. Since the delays change over time (as does the number of delays), the difference

$$\tau_{\max}(t) := \tau_{L(t)}(t) - \tau_0(t)$$

is a non-negative random variable for each t and is called the *maximum excess delay* of the channel. However, some multipath components hold significantly lower power than others, and in particular, some hold powers below the noise floor. Also, the demodulator at the receiver side may synchronize to a specific multipath component, e.g. the first incoming component or the average delay component. Thus, the maximum excess delay is not necessarily the most appropriate measure of how much the channel spreads the signal in time. In general, it depends on the specific application. Another measure is the so-called *channel delay spread*, defined from the *power delay profile* of the channel. Intuitively, the power delay profile explains the average signal power content as a function of multipath delay and it can be measured empirically for actual channels **[1]**. We elaborate further, and more

---

[2] This definition is quite vague due to our use of the symbol $\gg$ (or the phrase significantly).

specifically, on the power delay profile later in this chapter.

The impact of multipath distortion on the transmitted signal depends on whether the channel delay spread is large or small relative to the inverse of the signal bandwidth B_x. With a large delay spread the duration of a received signal may be significantly increased compared to the duration of the transmitted pulse. This means that delayed multipath components arriving from one pulse will interfere with subsequently transmitted pulses, and this deteriorating effect is called *inter-symbol interference* (ISI). Multipath distortion effects (such as ISI) can seriously degrade system performance and various techniques for mitigation of these effects exist. These techniques include for instance *equalization*, *multicarrier modulation* and *spread spectrum transmission*.

On the other hand, when the delay spread is small relative to the inverse of the signal bandwidth, most of the multipath signal components are non-resolvable. Thus, with the delay spread small compared to $1/B_x \approx T_s$, all the delayed signal echoes roughly correspond to the same transmitted symbol, i.e.

$$x\big(t - \tau_0(t)\big) \approx x\big(t - \tau_{L(t)}(t)\big). \tag{3.6}$$

Therefore, only a small amount of interference with subsequently transmitted pulses will be present. Thus, the distinction often emphasized in the literature between *wideband fading models* and *narrowband fading models* is reflected in the fact that the signal bandwidth and the channel delay spread determine whether the approximation in (3.6) is valid or not. In both cases, however, the received signal is a sum of copies of the original signal - with each copy individually attenuated and delayed in time as phrased by (3.1).

3.2 Time-varying Channel Impulse Response

Consider a linear time-varying bandpass channel with the real channel impulse response

$$c(\tau, t) = \mathrm{Re}\left\{g(\tau, t)\mathrm{e}^{i2\pi f_c t}\right\}.$$

For *time-invariant* channels, the equivalent lowpass channel impulse response would for an arbitrary time lag Δt satisfy

$$g(\tau, t) = g\big(\tau, t + \Delta t\big).$$

That is, the channel response at time t to an impulse launched at time $t-\tau$ coincides with the channel response at time $t+\Delta t$ to an impulse launched at time $t+\Delta t-\tau$. Specifically, this holds for $\Delta t = -t$ and hence

$$g(\tau, t) = g(\tau, t - t) = g(\tau, 0) := \widetilde{g}(\tau),$$

which is the impulse response associated with a time-invariant channel, i.e. it depends only on the delay argument τ and not on time.

When transmitting the bandpass signal $s(t)$ with equivalent lowpass signal $x(t)$, we know that the received bandpass signal $r(t)$ is given as (still, we neglect the additive noise process)

$$r(t) = \mathrm{Re}\left\{ \left(\int_{-\infty}^{\infty} g(\tau, t) x(t-\tau) \mathrm{d}\tau \right) \mathrm{e}^{i2\pi f_c t} \right\},$$

i.e. by 'convolving' $g(\tau, t)$ with $x(t)$ and then upconverting to the carrier frequency f_c. Recall from (3.1), that in the absence of noise we also have

$$r(t) = \mathrm{Re}\left\{ \left(\sum_{l=0}^{L(t)} \alpha_l(t) x\big(t - \tau_l(t)\big) \mathrm{e}^{-i2\pi f_c \tau_l(t)} \right) \mathrm{e}^{i2\pi f_c t} \right\}.$$

Comparing the two above expressions for $r(t)$, we realize that the equivalent lowpass channel impulse response must be given as

$$g(\tau, t) = \sum_{l=0}^{L(t)} \alpha_l(t) \delta\big(\tau - \tau_l(t)\big) \mathrm{e}^{-i2\pi f_c \tau_l(t)}, \tag{3.7}$$

where $\delta(t)$ is the Dirac delta and from its sampling property we get

$$\int_{-\infty}^{\infty} x(t-\tau) \delta\big(\tau - \tau_l(t)\big) \mathrm{d}\tau = x\big(t - \tau_l(t)\big), \qquad l = 0, 1, \ldots, L(t).$$

The expression (3.7) exploits the structure of the time-varying multipath channel impulse response $g(\tau, t)$ and similar or appropriately simplified equivalents are often encountered in literature **[1, 3]**.

3.3 Multipath Channel Fading Model

Due to transmitter or receiver mobility, the number of multipath components $L(t)$ is a random variable and the associated sets of amplitudes, phases and delays are considered random too. We assume that the receiver is located far from both transmitter and all scatterers. We also assume that no dominant LOS component is present, i.e. no term for $l = 0$ in (3.7). *In general, these receiver assumptions are imposed throughout the thesis.* In the following we introduce a number of simplifying assumptions on $g(\tau, t)$ in order to facilitate further characterization of the multipath channel.

Initially, we simplify the expression (3.7) by 'hiding' all the complex exponentials. We do this by redefining each of the complex amplitudes according to

$$\alpha_l(t) := \alpha_l(t) \mathrm{e}^{-i2\pi f_c \tau_l(t)}, \qquad l = 1, 2, \ldots, L(t),$$

and we model the initial phases of these redefined multipath components as i.i.d., uniformly on the interval $[-\pi, \pi)$. This assumption is reasonable, since the phases can change rapidly when the carrier frequency f_c is large, recall (3.2). The simplified version of the multipath channel impulse response reads

$$g(\tau, t) = \sum_{l=1}^{L(t)} \alpha_l(t)\delta\big(\tau - \tau_l(t)\big). \tag{3.8}$$

It is commonly suggested by several authors, e.g. **[11, 12, 18]**, to model each complex amplitude $\alpha_l(t)$, as a zero mean circular-symmetric complex Gaussian process with a variance depending on the associated delay $\tau_l(t)$. In **[1, Section 3.3]** and **[3, Section 10.1.1]**, the authors justify the Gaussian modeling from arguments grounded on the *Central Limit Theorem* **[19, Theorem 4.3.1]**. When no LOS component is present, the zero mean assumption for each echo is reasonable. However, with a dominant LOS component the corresponding complex amplitude will in general hold a non-zero mean. In this thesis, we consider only the non-LOS scenario where all complex amplitudes are of zero mean. Furthermore, it is always assumed that only the receiver is moving, i.e. the approximation in (3.3) is valid.

Even though the expression (3.8) involves a number of Dirac deltas (infinitely high and narrow spikes), the way to physically interpret its mathematical formulation is illustrated in **Figure 3.3**. We cannot depict dirac spikes and therefore we plot the squared magnitude of each complex amplitude as a function of time. The associated delays are also varying as time passes and by fixing a time instance, say $t = t_0$, we find a 2D illustration similar to the one provided in **Figure 1.4** on page **7**, i.e. an echogram.

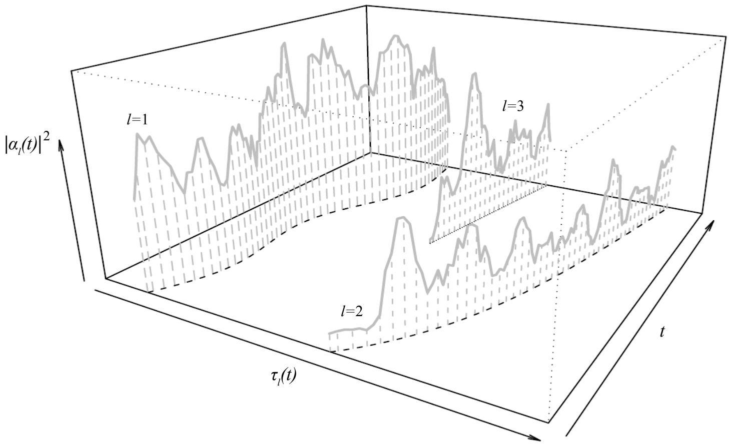


Figure 3.3: Three signal echoes illustrated as realizations of stochastic processes. At first only two echoes are present but as time passes yet another echo arises.

3.3.1 Correlation Properties

Our further investigation of the multipath channel is grounded on assumptions and properties that are commonly observed/imposed for a variety of actual channels **[1, Section 3.3]**. These particular assumptions were initially introduced by Bello **[20, Section IV. D]** and often serve as default assumptions today. In literature, this particular channel property is sometimes referred to as the WSSUS assumption **[1]**, which is an abbreviation for *wide-sense stationary* (WSS) and *uncorrelated scattering* (US). As mentioned above the complex amplitudes $\alpha_l(t)$, $l = 1, 2, \ldots, L(t)$, are modeled as zero mean complex Gaussian processes and now each of these processes are furthermore assumed to be WSS - explaining the first part of the abbreviation. The second part reflects that any two distinct multipath signal echoes are considered to have interacted with non-identical scattering objects while propagating from transmitter to receiver. Therefore, any pair of distinct complex amplitudes $\alpha_l(t)$ and $\alpha_k(t)$ corresponding to different signal echoes are assumed to comprise a pair of uncorrelated processes. In short, our assumptions so far reads:

- The time-variant amplitude $\alpha_l(t)$ comprise a zero mean circular symmetric complex Gaussian process for $l = 1, 2, \ldots, L(t)$.
- The autocorrelation function $\mathbb{E}\big[\alpha_l(t)\alpha_l^*(t+\Delta t)\big]$ depends only on the time lag Δt, while not on the particular time instance t (WSS assumption).
- The product moment $\mathbb{E}\big[\alpha_l(t_1)\alpha_k^*(t_2)\big]$ is identically zero for arbitrary time instances t_1 and t_2 whenever $l \neq k$ (US assumption).

Typically, in mobile communication links the multipath delay parameters $\tau_l(t)$, $l = 1, 2, \ldots, L(t)$, are slowly time-varying compared to the corresponding amplitude and phase terms **[12]**. If the receiver moves just a few centimeters the phases and amplitudes will change notably while the delay parameters remain practically unchanged since the electromagnetic waves travel at the speed of light. Hence, when considering only time windows of short duration we may consider $L(t) \equiv L$ and $\tau_l(t) \equiv \tau_l$ as static, deterministic terms. Paramount, we furthermore assume that all amplitude processes $\alpha_l(t)$ share the same normalized autocorrelation function $R_t(\Delta t)$. This key assumption allows us to follow the approach in **[18]**.

From appropriate Fourier transforms we can characterize the multipath channel in the frequency domain as well. Initially, we consider the time-varying *channel transfer function* (CTF) defined as

$$h(f,t) := \int_{-\infty}^{\infty} g(\tau,t)\mathrm{e}^{-i2\pi f\tau}\mathrm{d}\tau = \sum_{l=1}^{L} \alpha_l(t)\mathrm{e}^{-i2\pi f\tau_l},$$

i.e. the Fourier transform of $g(\tau,t)$ with respect to the delay variable τ. Since $\mathbb{E}\big[h(f,t)\big] = 0$, we directly calculate the autocorrelation function

$$
\begin{aligned}
R_{hh}(f_1, f_2, \Delta t) &:= \mathbb{E}\Big[h(f_1, t)h^*(f_2, t + \Delta t)\Big] \\
&= \sum_{l=1}^{L} \mathbb{E}\big[\alpha_l(t)\alpha_l^*(t + \Delta t)\big]\mathrm{e}^{-i2\pi(f_1 - f_2)\tau_l} && (3.9) \\
&= \sum_{l=1}^{L} Q(\tau_l)R_t(\Delta t)\mathrm{e}^{-i2\pi\Delta f\tau_l}, && (3.10)
\end{aligned}
$$

where $Q(\tau_l)$ denotes the average power of the l'th signal component with associated delay τ_l. More specifically, we have introduced the quantity

$$
Q(\tau_l) := \mathbb{E}\Big[|\alpha_l(t)|^2 \;\big|\tau_l\Big],
$$

which partly specifies the so-called power delay profile of the multipath channel. Notice how (3.9) is obtained using the uncorrelated scattering property. We realize from (3.10) that the US property in the delay domain is equivalent to a WSS property when recast in the frequency domian, i.e. (3.10) depends only on the frequency separation $\Delta f := f_1 - f_2$. Specifically, this means that we can rearrange the expression (3.10) and with a slight abuse of notation we reuse the name R_{hh} and write

$$
R_{hh}(\Delta f, \Delta t) = \widetilde{Q}\, R_t(\Delta t) \underbrace{\sum_{l=1}^{L} \frac{Q(\tau_l)}{\widetilde{Q}} \mathrm{e}^{-i2\pi\Delta f\tau_l}}_{R_f(\Delta f)}, \qquad (3.11)
$$

where $\widetilde{Q} := \sum Q(\tau_l)$ denotes the overall average power in the channel. Hence, the autocorrelation function associated with the time-varying channel transfer function $h(f, t)$ factorizes into two separate and normalized autocorrelation functions $R_t(\Delta t)$ and $R_f(\Delta f)$. The authors of **[18]** refer to this fact as the *separation property* of the mobile wireless channel.

3.3.2 Coherence Time and Coherence Bandwidth

The time domain autocorrelation function $R_t(\Delta t)$ depends on the speed of the receiver, transmitter and the environment, or equivalently, $R_t(\Delta t)$ depends on the Doppler frequency. Intuitively, if the receiver travels at high speed, e.g. highway or high-speed train, the multipath channel will change rapidly and the magnitude of $R_t(\Delta t)$ will decay fast as Δt increases. On the other hand, when the environment is more or less static and no movement occurs the multipath channel does not change notable over time and the magnitude of $R_t(\Delta t)$ will slowly tend towards zero as Δt increases. In **Chapter 7**, we introduce a modeling setup for a time-varying multipath channel and use the model for simulation purposes. We model each amplitude process $\alpha_l(t)$ and subsequently derive an analytical expression of the associated autocorrelation function $R_t(\Delta t)$.

The decay behavior of the frequency domain autocorrelation function $R_f(\Delta f)$ depends on the delay dispersiveness of the channel. The more dispersion introduced by the multipath channel, the more frequency selective the channel will be. Hence, when the maximum excess delay of the channel is small, the frequency response of the channel decorrelates slowly with increasing frequency separations Δf. Conversely, when the maximum excess delay of the channel is large, i.e. massive dispersion in delay, the channel transfer function will decorrelate rapidly as Δf increases.

Reconsidering the expression for $R_{hh}(\Delta f, \Delta t)$ in (3.11) we may by turns equate either Δt or Δf to zero and consider the approximate time or frequency separations needed for the channel to be 'practically' uncorrelated. That is, we define the *coherence time* of the channel to be

$$\mathcal{T}_C := \arg \min_{\Delta t > 0} \left| R_t(\Delta t) \right| = C, \qquad 0 \leq C < 1,$$

and similarly we define the *coherence bandwidth* of the channel to be

$$\mathcal{B}_C := \arg \min_{\Delta f > 0} \left| R_f(\Delta f) \right| = C, \qquad 0 \leq C < 1,$$

where the constant C will typically be chosen different from application to application (explaining our use of the phrase 'practically' uncorrelated and renders visible too, that these definitions are not very strict). Heuristically, $\mathcal{T}_C$ is a measure of the width of the main lobe of $|R_t(\Delta t)|$ and a similar interpretation is valid for $\mathcal{B}_C$.

3.3.3 Doppler Power Spectrum

Another useful and equivalent characterization of the multipath channel is available in terms of the so-called *Doppler power spectrum*. As mentioned above, the behavior of $R_t(\Delta t)$ depends on the dynamics in the communication environment, i.e. on the Doppler frequencies. If we assume that only the receiver is moving we have previously in this chapter interpreted the Doppler frequency shift from a geometrical perspective. A more general interpretation can be obtained from

$$R_D(\Delta f, \nu) := \int_{-\infty}^{\infty} R_{hh}(\Delta f, \Delta t) \mathrm{e}^{-i2\pi\nu\Delta t} \mathrm{d}\Delta t,$$

i.e. the Fourier transform of R_{hh} with respect to the time lag Δt. The function $R_D(0, \nu)$ is called the Doppler power spectrum of the channel and it characterizes how the power of the transmitted signal is distributed relative to Doppler frequency. With reference to (3.5) and assuming that only the receiver is moving (at speed v), the maximum possible Doppler frequency shift is $f_D := \frac{v}{\lambda}$, where λ denotes the signal wavelength. Hence, from our mobility assumptions we find that

$$R_D(0, \nu) = 0, \quad \text{whenever } |\nu| > f_D,$$

since no multipath signal components can hold larger Doppler frequency shifts when only the receiver is moving. Concrete examples of a few Doppler power spectrums are given in **Chapter 7**.

3.4 Distribution of Envelope and Power

Consider the random process

$$\beta(t) := \sum_{l=1}^{L(t)} \alpha_l(t),$$

which is a sum of $L(t)$ uncorrelated zero-mean complex Gaussian processes. It is easily seen that $\mathbb{E}\big[\beta(t)\big] = 0$ and using the uncorrelated scattering (US) assumption we find

$$\mathbb{V}\text{ar}\big[\beta(t)\big] = \mathbb{E}\big[\beta(t)\beta^*(t)\big] = \mathbb{E}\left[\sum_{l=1}^{L(t)} \big|\alpha_l(t)\big|^2\right].$$

Since a sum of Gaussian processes is again a Gaussian process we realize that

$$\beta(t) \sim \mathcal{CN}\big(0, \sigma^2\big),$$

where, for notational convenience, we have defined the total average power in the channel as

$$\sigma^2 := \mathbb{E}\left[\sum_{l=1}^{L(t)} \big|\alpha_l(t)\big|^2\right].$$

We write $\beta(t) = X(t) + iY(t)$ where $X(t)$ and $Y(t)$ are uncorrelated zero-mean real-valued Gaussian processes, both with variance $\sigma^2/2$. Since $X(t)$ and $Y(t)$ are uncorrelated and Gaussian distributed they are furthermore independent, and therefore

$$\frac{2}{\sigma^2}\Big(X^2(t) + Y^2(t)\Big) = \left(\frac{X(t)}{\sqrt{\sigma^2/2}}\right)^2 + \left(\frac{Y(t)}{\sqrt{\sigma^2/2}}\right)^2,$$

follows a chi-square distribution with two degrees of freedom, or equivalently, an exponential distribution with rate-parameter equal to one half. It is well-known that the square-root of an exponentially distributed random variable follows a Rayleigh distribution.

In particular, if the total channel power is normalized to unity, i.e. $\sigma^2 = 1$, then the envelope

$$\big|\beta(t)\big| = \sqrt{X^2(t) + Y^2(t)} \sim \text{Rayl}\Big(1/\sqrt{2}\Big)$$

and the power

$$\big|\beta(t)\big|^2 = X^2(t) + Y^2(t) \sim \text{Expo}\big(1\big).$$

These theoretical distributions are used in **Chapter 7** to verify channel modelings and implementations for simulation purposes. The particular channel model described with the set of assumptions imposed in this chapter in commonly referred to as a *Rayleigh fading multipath channel.*

Chapter 4

Multicarrier Modulation

Multipath channels are often to be considered with today's communication systems since consecutive information symbols are transmitted closely spaced in time. The delay dispersiveness of multipath channels cause the transmitted symbols to superimpose in time at the receiver side. Multicarrier modulation is a technique to circumvent/reduce the adverse overlapping of consecutive symbols and was first used in the late 1950s, however, the technique was far too complex for most system implementations at that time. Twenty years later the use of multicarrier modulation was ignited when fast and cheap (hardware) implementations of the discrete-time Fourier transform (DFT) were developed. The main part of this chapter is inspired by **[1, Chapter 12]**, and deals with OFDM and some of the aspects to be considered with communication systems utilizing this multicarrier modulation technique.

As mentioned shortly in the beginning of **Section 1.2**, the basic idea of multicarrier modulation is to transmit the information symbols over a relatively large set of narrowband subchannels (frequency domain). The number of subchannels is chosen large such that the symbol duration on each narrowband subchannel becomes larger than the *delay spread* of the channel. With the symbol duration prolonged the *inter-symbol interference* (ISI) is small, and in fact, ISI can be completely eliminated through the use of a *cyclic prefix*.

4.1 Orthogonal Frequency-division Multiplexing

One way to implement multicarrier modulation is by *orthogonal frequency-division multiplexing* (OFDM). Another multicarrier technique is *vector coding* **[1, Section 12.4.5]**, however, we consider only the former technique and the phrase 'orthogonal' will be explained later in this section. OFDM is implemented partly by use of DFT operations and this transform is efficiently implemented (low complexity) through the *fast Fourier transform* (FFT), see e.g. **[15, Section 8.2]** for details on this important[1] transform used extensively in a wide range of practical applications.

---

[1] Citing **[15]**: *'The FFT is possibly the second most nontrivial algorithm in practice.'*

Consider a communication system where phase/amplitude modulated symbols $\{\widehat{x}_k\}$ are transmitted with symbol duration T_s, thus occupying a bandwidth of $1/T_\mathrm{s}$. Transmission of N consecutive symbols $\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_N$ will occupy a time slot of duration NT_s at the transmitting entity, since symbols are sent one by one in a pulse train. To avoid time overlapping symbols at the receiver side, the transmitter will apply an OFDM scheme. The overall bandwidth of $1/T_\mathrm{s}$ is divided into N slots, each of bandwidth $1/(NT_\mathrm{s})$. The symbols $\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_N$ are then modulated (by an inverse DFT) to new symbols $x_1, x_2, \ldots, x_N$ such that each symbol x_n carries information about all $\widehat{x}_k$'s. Consecutive transmission of the N symbols $x_1, x_2, \ldots, x_N$ will likewise occupy a time slot of duration NT_s, however, the original symbols $\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_N$ are now individually distributed across the N narrowband slots in frequency. This entire procedure is illustrated in **Figure 4.1**, spelling the difference between whether multicarrier modulation is utilized or not.

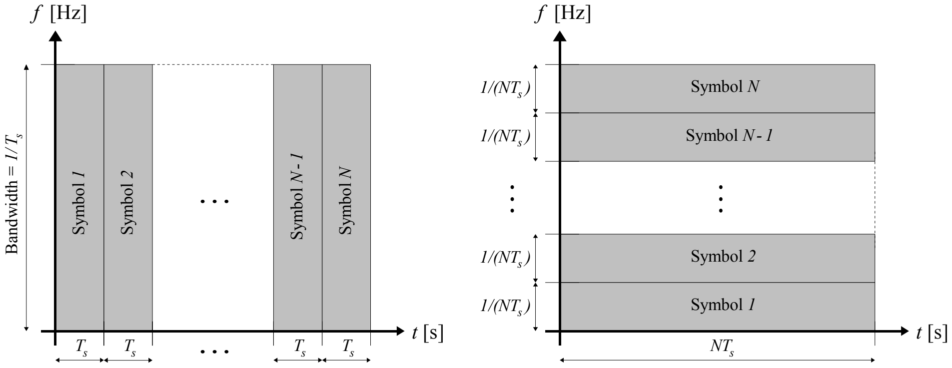


Figure 4.1: Distinction between a system where symbols are not modulated onto multiple carriers (left) and a system where multicarrier modulation is utilized (right). In both cases a time span of NT_s is used to transmit N symbols and both systems occupy a total bandwidth of $1/T_\mathrm{s}$.

With multicarrier modulation, time overlapping of the symbols $x_1, x_2, \ldots, x_N$ will only deteriorate the information about the original symbols $\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_N$ when occurring at the edges of the block. Consider for instance the transmission of three consecutive blocks of symbols

$$\underbrace{x_1, x_2, \ldots, x_N}_{\text{block 1}}\,,\ \underbrace{x_{N+1}, x_{N+2}, \ldots, x_{2N}}_{\text{block 2}}\,,\ \underbrace{x_{2N+1}, x_{2N+2}, \ldots, x_{3N}}_{\text{block 3}}\,.$$

The symbols transmitted at the end of the first block may interfere with the earliest symbols from the second block. Similarly, the last symbols from the second block may interfere with the earliest symbols form the third block. Hence, ISI will occur at the edges of the blocks but this distortion can be removed by appending a so-called cyclic prefix at the beginning of each block.

4.1.1 Mitigation of ISI by a Cyclic Prefix

Consider a discrete-time signal $\boldsymbol{x} = \left[x_1, x_2, \ldots, x_N\right]^\top$ of length N as input to a linear discrete-time delay dispersive channel. We assume that the channel is time-invariant with a *finite impulse response* (FIR) of length $L+1$ samples, and we denote these responses by $g_0, g_1, \ldots, g_L \in \mathbb{C}$. If we denote the symbol time by T_s and the maximum excess delay of the channel by $\tau_{\max}$, we see from our FIR-assumption, that

$$\tau_{\max} = LT_s.$$

The *cyclic prefix* for the signal $\boldsymbol{x}$ is defined from the last $\mu + 1$ samples of $\boldsymbol{x}$, where μ should be selected such that $\mu + 1 \geq L$. These last $\mu + 1$ samples are appended at the beginning of $\boldsymbol{x}$ to form a new discrete-time signal of length $\mu + 1 + N$ denoted by

$$\boldsymbol{x}_{\mathrm{cp}} := \left[x_{-\mu}, \ldots, x_{-1}, x_0, \boldsymbol{x}^\top\right]^\top,$$

where the subscripts less that one are nicely related to the original subscripts according to

$$\begin{aligned} x_0 &:= x_N \\ x_{-1} &:= x_{N-1} \\ &\vdots \\ x_{-\mu} &:= x_{N-\mu}. \end{aligned}$$

The cyclic prefix defined above is introduced with two main objectives:

- Like $\boldsymbol{x}_{\mathrm{cp}}$, the received signal $\boldsymbol{y}_{\mathrm{cp}}$ will be of length $\mu + 1 + N$, however, the first $\mu + 1$ samples are discarded at the receiver side since these samples have possibly suffered from severe ISI due to the delay dispersion in the multipath channel.
- The cyclic prefix turns the *linear convolution* between the input signal $\boldsymbol{x}_{\mathrm{cp}}$ and the channel into a *circular convolution* of period N with the original signal $\boldsymbol{x}$ (applying a DFT on a circular convolution in time leads to multiplication in frequency).

Thus, the cyclic prefix serves to eliminate ISI but comes with a cost since the $\mu + 1$ prefixed samples cause an overhead of $(\mu + 1)/N$. If $\mu + 1$ is chosen strictly smaller than L, then some degree of ISI will remain and if μ is chosen too large, a dispensable overhead is introduced.

4.1.2 Overlapping Subchannels and Orthogonality

Recall **Section 1.2.1** where we briefly elaborated on the frequency domain properties of the N narrowband subchannels. In OFDM the center frequencies of two adjacent subchannels are located with a spacing of $\Delta f := 1/(NT_s)$ and, as depicted in **Figure 4.2**, the range of frequencies occupied by adjacent subchannels are overlapping.

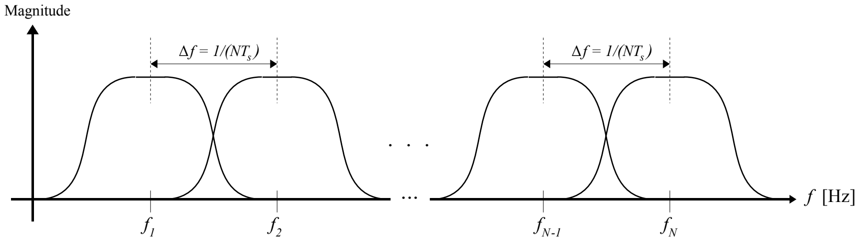


Figure 4.2: Frequency-domain view of multicarrier modulation with overlapping subchannels.

The spectrum of adjacent subchannels are overlapping since the overall system should seek to occupy no more bandwidth than needed. Obviously, the amount of frequency overlap must be confined, and this fact is partly expressed in the phrase *orthogonal* in 'OFDM'. It is common to say that the subchannels/subcarriers are orthogonal to each other, and the following lemma provides the formal meaning of this saying.

Lemma 4.1.1. (Subcarrier Orthogonality)
Let f denote a reference frequency such that $2fNT_{\text{s}} \in \mathbb{N}$, where N denotes the number of subchannels and T_{s} is the sampling time. Then for any set of phase offsets $\{\phi_k\}$, the subcarriers

$$\Big\{\psi_k(t) := \cos\Big(2\pi(f + \tfrac{k}{NT_{\text{s}}})t + \phi_k\Big) \ : \ k = 0, 1, \ldots, N-1\Big\}$$

form a set of mutually orthogonal functions on the interval $[0, NT_{\text{s}}]$.

Proof. From the formula

$$2\cos\alpha\cos\beta = \cos(\alpha - \beta) + \cos(\alpha + \beta),$$

with $n, m \in \{0, 1, \ldots, N-1\}$, we obtain

$$\begin{aligned} 2\int_0^{NT_{\text{s}}} \psi_n(t)\psi_m(t)\mathrm{d}t = & \int_0^{NT_{\text{s}}} \cos\Big(2\pi(n-m)\tfrac{t}{NT_{\text{s}}} + \phi_n - \phi_m\Big)\mathrm{d}t \\ & + \underbrace{\int_0^{NT_{\text{s}}} \cos\Big(2\pi(2fNT_{\text{s}} + n + m)\tfrac{t}{NT_{\text{s}}} + \phi_n + \phi_m\Big)\mathrm{d}t}_{(\star)}. \end{aligned}$$

For simplicity, introduce $p := 2fNT_{\text{s}} + n + m$ and $\theta := \frac{\phi_n + \phi_m}{p}$ and notice that $p \in \mathbb{N}$. By the variable change $\tau = 2\pi\frac{t}{NT_{\text{s}}} + \theta$ the integral in $(\star)$ becomes

$$\int_0^{NT_{\text{s}}} \cos\Big(2\pi p\tfrac{t}{NT_{\text{s}}} + p\theta\Big)\mathrm{d}t = \frac{NT_{\text{s}}}{2\pi}\int_\theta^{2\pi+\theta} \cos(p\tau)\mathrm{d}\tau = 0,$$

and hence we obtain that

$$\frac{2}{NT_{\mathrm{s}}}\int_0^{NT_{\mathrm{s}}}\psi_n(t)\psi_m(t)\mathrm{d}t = \mathbb{1}[n=m].$$

■

A few remarks concerned with **Lemma 4.1.1** are important to keep on mind. The assumption on $2fNT_{\mathrm{s}}$ being integer-valued can be relaxed. If instead we require only that $2fNT_{\mathrm{s}}$ is large the lemma can be reformulated as an approximation, i.e.

$$\frac{2}{NT_{\mathrm{s}}}\int_0^{NT_{\mathrm{s}}}\psi_n(t)\psi_m(t)\mathrm{d}t = \mathbb{1}[n=m]+\epsilon,$$

with $\epsilon \to 0$ as $2fNT_{\mathrm{s}} \to \infty$. That is, when the reference frequency f is large the subcarriers $\{\psi_k(t)\}$ form a set of approximately orthogonal functions on the interval $[0, NT_{\mathrm{s}}]$. Also, with frequency separations smaller than $\Delta f = \frac{1}{NT_{\mathrm{s}}}$, no set of subcarriers with arbitrary phase offsets can constitute a set of orthogonal functions on $[0, NT_{\mathrm{s}}]$. Thus, to maintain orthogonality between subcarriers the minimum frequency separation required is $\frac{1}{NT_{\mathrm{s}}}$.

4.2 Linear Algebra Representation of OFDM

In this section we present a discrete-time baseband formulation of an OFDM system. First, consider a sequence of PSK or QAM modulated information symbols, which we divide into blocks of length N. For now, the transmission of a single of these blocks is considered, and thus we arrange an arbitrary block as a column vector and denote it by

$$\widehat{\boldsymbol{x}} := \left[\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_N\right]^\top.$$

Recall the N-point IDFT matrix $\mathbf{F}^H$ from **Section 2.2** and apply it on $\widehat{\boldsymbol{x}}$ to obtain

$$\boldsymbol{x} := \left[x_1, x_2, \ldots, x_N\right]^\top = \mathbf{F}^H\widehat{\boldsymbol{x}}.$$

The signal $\boldsymbol{x}$ is appended by a cyclic prefix and $\boldsymbol{x}_{\mathrm{cp}}$ is transmitted across a linear discrete-time channel in presence of additive white noise, assumed to be *circular symmetric*[2] complex Gaussian distributed. We analyze the equivalent lowpass channel input-output relationship, and we impose a number of assumptions on the system. A discussion of these assumptions are presented later in this chapter when we consider continuous-time transmission.

Assumptions:

- The channel is time-invariant during the $(\mu + 1 + N)$ samples when $\boldsymbol{x}_{\mathrm{cp}}$ is transmitted.

---

[2] Mutually uncorrelated real and imaginary parts, all having the same variance. See **Appendix C**.

- The equivalent lowpass channel is *causal* and has a finite impulse response of length $L+1$, such that $L \leq \mu + 1 \ll N$.
- The $L+1$ channel responses $g_0, g_1, \ldots, g_L \in \mathbb{C}$ are aligned with the sampling grid.
- Perfect time and frequency synchronization between transmitter and receiver.

Assume for simplicity, that our cyclic prefix is of length $\mu + 1 = L$, i.e the minimum overhead required to eliminate ISI between consecutive blocks. The prefixed signal $\boldsymbol{x}_{\text{cp}}$ is transmitted across the linear discrete-time channel and the received signal $\boldsymbol{y}_{\text{cp}}$, of length $\mu + 1 + N$, is given by the linear convolution of $\boldsymbol{x}_{\text{cp}}$ with the channel and corrupted by additive noise. We immediately discard the first $\mu + 1$ samples of $\boldsymbol{y}_{\text{cp}}$ and the remaining N samples of interest are given by

$$\begin{bmatrix} y_1 \\ y_2 \\ \\ \vdots \\ \\ y_N \end{bmatrix} = \begin{bmatrix} g_L & \cdots & g_1 & g_0 & & & \mathrm{O} & & \\ & g_L & \cdots & g_1 & g_0 & & & \mathrm{O} & \\ & & \ddots & & & \ddots & & & \mathrm{O} \\ \mathrm{O} & & & \ddots & & & \ddots & & \\ & \mathrm{O} & & & g_L & \cdots & g_1 & g_0 & \\ & & \mathrm{O} & & & g_L & \cdots & g_1 & g_0 \end{bmatrix} \begin{bmatrix} x_{-\mu} \\ \vdots \\ x_{-1} \\ x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \\ \vdots \\ \\ w_N \end{bmatrix},$$

and this linear system of equations is compactly expressed as

$$\boldsymbol{y} = \widetilde{\mathbf{G}} \boldsymbol{x}_{\text{cp}} + \boldsymbol{w}, \tag{4.1}$$

where $\widetilde{\mathbf{G}}$ is of size $N \times (\mu + 1 + N)$ and $\boldsymbol{w}$ is the noise sequence. Notice how we have simply discarded the first $\mu + 1$ rows of the linear convolution matrix to obtain the N output samples of interest. Since the $\mu + 1$ components defining the cyclic prefix in $\boldsymbol{x}_{\text{cp}}$ are chosen such that

$$x_{-k} = x_{N-k}, \quad k = 0, 1, \ldots, \mu,$$

we can reformulate the system (4.1) and express it in terms of a square matrix $\mathbf{G}$ and the non-prefixed vector $\boldsymbol{x}$ (rather than $\boldsymbol{x}_{\text{cp}}$). This reformulated system is given by

$$\begin{bmatrix} y_1 \\ y_2 \\ \\ \vdots \\ \\ y_N \end{bmatrix} = \begin{bmatrix} g_0 & & & g_L & \cdots & g_1 \\ g_1 & g_0 & \mathrm{O} & & \ddots & \vdots \\ \vdots & & \ddots & \mathrm{O} & & g_L \\ g_L & & & \ddots & \mathrm{O} & \\ & \ddots & & & \ddots & \\ \mathrm{O} & & g_L & \cdots & g_1 & g_0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \\ \vdots \\ \\ x_N \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \\ \vdots \\ \\ w_N \end{bmatrix}, \tag{4.2}$$

and we can likewise express this linear system of equations compactly as

$$\boldsymbol{y} = \mathbf{G} \boldsymbol{x} + \boldsymbol{w}. \tag{4.3}$$

The following result is crucial for the continuation of our channel input-output characterization and allows for straightforward argumentation and simpler derivations later on.

Lemma 4.2.1.
The cyclic convolution matrix $\mathbf{G}$ in (4.3) is *normal*, i.e.

$$\mathbf{G}\mathbf{G}^H = \mathbf{G}^H\mathbf{G}.$$

Proof. Define the matrices $\mathbf{A} := \mathbf{G}\mathbf{G}^H$ and $\mathbf{B} := \mathbf{G}^H\mathbf{G}$ and we now verify that each entry $\mathbf{A}(n,m)$ is equal to $\mathbf{B}(n,m)$ for all $(n,m) \in \{1,2,\ldots,N\} \times \{1,2,\ldots,N\}$. From the cyclic convolution structure of $\mathbf{G}$ it follows that its second column is simply a one-step cyclic shift of its first column. The third column is a one-step cyclic shift of the second and so on ($\mathbf{G}$ is a so-called *circulant matrix*). Denote by $\boldsymbol{q}$ the first column of $\mathbf{G}$, i.e. let

$$\boldsymbol{q} := \left[q_{1}, q_{2}, \ldots, q_{N}\right]^\top = \left[g_0, g_1, \ldots, g_L, 0, 0, \ldots, 0\right]^\top,$$

where in our particular case the last $N-L-1$ values are assumed to be zero, however, this fact is not needed to complete the proof. Thus, we may consider an arbitrary $\boldsymbol{q} \in \mathbb{C}^N$. Consider the $N \times N$ *one-step cyclic shift matrix* $\mathbf{D}$ given by

$$\mathbf{D} := \begin{bmatrix} 0 & & & O & 1 \\ 1 & 0 & & & O & \\ & 1 & 0 & & & O \\ O & & \ddots & \ddots & & \\ & O & & 1 & 0 & \\ & & O & & 1 & 0 \end{bmatrix},$$

and notice that $\mathbf{D}$ is easily obtained from the identity matrix $\mathbf{I}_N$ (with its bottom row inserted as the first row).
With $\mathbf{D}^k$ being the k-fold matrix multiplication of $\mathbf{D}$ with itself and with the appropriate convention $\mathbf{D}^0 := \mathbf{I}_N$, we can write

$$\mathbf{G} = \begin{bmatrix}\mathbf{D}^0\boldsymbol{q} & \mathbf{D}\boldsymbol{q} & \mathbf{D}^2\boldsymbol{q} & \cdots & \mathbf{D}^{N-1}\boldsymbol{q}\end{bmatrix},$$

since the k'th column of $\mathbf{G}$ is a $(k-1)$-step cyclic shift of $\boldsymbol{q}$ for $k = 2, 3, \ldots N$. Note that

$$\mathbf{D}^{k+N} = \mathbf{D}^k \quad \text{and} \quad \left(\mathbf{D}^k\right)^\top = \mathbf{D}^{N-k}, \qquad k = 0, 1, \ldots, N-1.$$

We define a *reversal matrix* $\mathbf{J}$ to be the $N \times N$ matrix with 1's on its entire anti-diagonal and zeroes elsewhere. The transform $\mathbf{J}\boldsymbol{q}$ simply gives the vector $\boldsymbol{q}$ in reversed order. Now we can express, for instance

$$\mathbf{G}^H = \begin{bmatrix}\mathbf{D}\mathbf{J}\boldsymbol{q}^* & \mathbf{D}^2\mathbf{J}\boldsymbol{q}^* & \cdots & \mathbf{D}^{N-1}\mathbf{J}\boldsymbol{q}^* & \mathbf{D}^0\mathbf{J}\boldsymbol{q}^*\end{bmatrix},$$

but also we can express both $\mathbf{G}$ and $\mathbf{G}^H$ by their rows instead of their columns, again from appropriate transforms involving $\mathbf{D}$, $\mathbf{J}$ and $\boldsymbol{q}$. Note that $\mathbf{J}\mathbf{D}^k\mathbf{J} = \left(\mathbf{D}^k\right)^\top$ and finally we are able to obtain the two complex entries

$$\mathbf{A}(n,m) = \left(\mathbf{D}^n\mathbf{J}\boldsymbol{q}\right)^\top\left(\mathbf{D}^m\mathbf{J}\boldsymbol{q}^*\right) = \boldsymbol{q}^\top\mathbf{J}\mathbf{D}^{N-n+m}\mathbf{J}\boldsymbol{q}^* = \boldsymbol{q}^\top\left(\mathbf{D}^{N-n+m}\right)^\top\boldsymbol{q}^*$$

$$\mathbf{B}(n,m) = \left(\mathbf{D}^{n-1}\boldsymbol{q}\right)^H\left(\mathbf{D}^{m-1}\boldsymbol{q}\right) = \boldsymbol{q}^H\mathbf{D}^{N-n+m}\boldsymbol{q}.$$

Since a scalar equals its own transpose we can simply transpose the final expression for $\mathbf{A}(n,m)$ and thereby complete the proof, since

$$\mathbf{A}(n,m) = \left(\boldsymbol{q}^\top\left(\mathbf{D}^{N-n+m}\right)^\top\boldsymbol{q}^*\right)^\top = \boldsymbol{q}^H\mathbf{D}^{N-n+m}\boldsymbol{q} = \mathbf{B}(n,m).$$

■

Immediately, **Lemma 4.2.1** allows for the *complex spectral theorem* **[21, Theorem 7.9]** to be applied, and hence, the circulant square matrix $\mathbf{G}$ has an *eigenvalue decomposition*

$$\mathbf{G} = \mathbf{M}\boldsymbol{\Lambda}\mathbf{M}^H, \tag{4.4}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of the eigenvalues of $\mathbf{G}$ and $\mathbf{M}$ is a matrix whose columns constitute a set of N *orthonormal*[3] eigenvectors of $\mathbf{G}$. Consider the transpose of $\mathbf{G}$ and recall the DFT matrix $\mathbf{F}$ from **Section 2.2**, i.e. consider the two $N \times N$ matrices

$$\mathbf{G}^\top = \begin{bmatrix} g_0 & g_1 & \cdots & g_L & & \mathrm{O} \\ & g_0 & & & \ddots & \\ \mathrm{O} & & \ddots & & & g_L \\ g_L & \mathrm{O} & & \ddots & & \vdots \\ \vdots & \ddots & & \mathrm{O} & \ddots & g_1 \\ g_1 & \cdots & g_L & & & g_0 \end{bmatrix}, \quad \mathbf{F} = \frac{1}{\sqrt{N}}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega & \cdots & \omega^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \cdots & \omega^{(N-1)^2} \end{bmatrix}$$

Now we can readily verify by calculations that each column $n = 1, 2, \dots, N$, of the DFT matrix $\mathbf{F}$, is an eigenvector of $\mathbf{G}^\top$ with corresponding eigenvalue

$$\lambda_n = \sum_{k=1}^{L+1} g_{k-1}\omega^{(k-1)(n-1)} = \sum_{k=1}^{N} q_k \exp\left(-i2\pi(n-1)\tfrac{k-1}{N}\right), \tag{4.5}$$

---

[3] A list of vectors is called orthonormal if the vectors are mutually orthogonal and each vector has Euclidian norm equal to one. The columns of $\mathbf{M}$ constitute a basis for $\mathbb{C}^N$, and it fact, the rows hold the same property. Hence, $\mathbf{M}^H\mathbf{M} = \mathbf{I}_N = \mathbf{M}\mathbf{M}^H$ and a matrix $\mathbf{M}$ with this property is called *unitary*.

with $\boldsymbol{q}$ as introduced in **Lemma 4.2.1**, and where one needs to utilize the fact that

$$\omega = \exp\left(-i2\pi\tfrac{1}{N}\right) \quad \Rightarrow \quad \omega^{k+N} = \omega^{k}, \quad k \in \mathbb{Z}.$$

Hence, we have established the relationship

$$\mathbf{G}^{\top}\mathbf{F} = \mathbf{F}\boldsymbol{\Lambda}, \tag{4.6}$$

where $\boldsymbol{\Lambda} = \operatorname{diag}\big(\lambda_{1}, \ldots, \lambda_{N}\big)$. Multiply both sides of (4.6) with $\mathbf{F}^{-1}$ from the right (recall that $\mathbf{F}^{-1} = \mathbf{F}^{H}$), and transpose the result to obtain

$$\mathbf{G} = \mathbf{F}^{H}\boldsymbol{\Lambda}\mathbf{F}.$$

From the decomposition (4.4) it now follows that $\mathbf{M}^{H} = \mathbf{F}$ and therefore $\mathbf{M} = \mathbf{F}^{H}$. Thus, the frequency domain channel input-output relationship reads

$$\begin{aligned}\widehat{\boldsymbol{y}} := \mathbf{F}\boldsymbol{y} &= \mathbf{F}\big(\mathbf{G}\boldsymbol{x} + \boldsymbol{w}\big)\\ &= \mathbf{F}\big(\mathbf{M}\boldsymbol{\Lambda}\mathbf{M}^{H}\mathbf{F}^{H}\widehat{\boldsymbol{x}} + \boldsymbol{w}\big)\\ &= \boldsymbol{\Lambda}\widehat{\boldsymbol{x}} + \mathbf{F}\boldsymbol{w},\end{aligned}$$

and the above calculations show that OFDM, by the use of IDFT/DFT operations combined with the cyclic prefix, decomposes a multipath ISI channel into N orthogonal subchannels. Since the additive noise $\boldsymbol{w}$ is assumed to be white and complex Gaussian distributed, its properties are unchanged when exposed to a DFT since still the mean is zero and

$$\mathbb{V}\mathrm{ar}\big[\mathbf{F}\boldsymbol{w}\big] = \mathbf{F}\mathbb{V}\mathrm{ar}\big[\boldsymbol{w}\big]\mathbf{F}^{H} = \mathbf{F}\big(\sigma^{2}\mathbf{I}_{N}\big)\mathbf{F}^{H} = \sigma^{2}\mathbf{I}_{N}.$$

The distribution remains Gaussian too, since a linear combination of Gaussians is again Gaussian. Accordingly, and with a slight abuse of notation, we propound the coordinate-wise OFDM input-output relationship as

$$\widehat{y}_{n} = \lambda_{n}\widehat{x}_{n} + w_{n}, \quad n = 1, 2, \ldots, N, \tag{4.7}$$

and from (4.5) we directly obtain the eigenvalue λ_{n}, which we (up to normalization) recognize as the *channel transfer function* evaluated at the corresponding subcarrier frequency. That is, the complete set of eigenvalues are obtained from the N-point DFT of the first column $\boldsymbol{q}$ of $\mathbf{G}$, according to

$$\big[\lambda_{1}, \lambda_{2}, \ldots, \lambda_{N}\big]^{\top} = \sqrt{N}\mathbf{F}\boldsymbol{q}.$$

Remark that **Lemma 4.2.1** and the subsequent calculations show that the columns of the N-point DFT matrix constitute the complex eigenvector basis needed to diagonalize any circulant $N \times N$ matrix.

4.3 Continuous-time OFDM Transmission

In **Figure 4.3** a block diagram of a complete OFDM system is shown. In practice there is no transmission across a discrete-time channel as described in the previous section - instead the transmission is analog with continuous-time signals. This fact is reflected in the block diagram where the prefixed discrete-time signal $\boldsymbol{x}_{\text{cp}}$ is digital-to-analog (D/A) converted with a pulse shaping filter. The analog signal is then up-converted to a certain carrier frequency f_c before continuous-time transmission across the wireless channel. During transmission the signal is subject to multipath channel distortion and is further corrupted by additive noise, assumed to be white and Gaussian. At the receiver side the distorted signal is initially carrier demodulated (down-converted) and lowpass filtered. A filter *matched* to the shaping pulse is then applied to the signal and finally the resulting signal is sampled. This entire procedure and further details are shown in the block diagram of **Figure 4.3**.

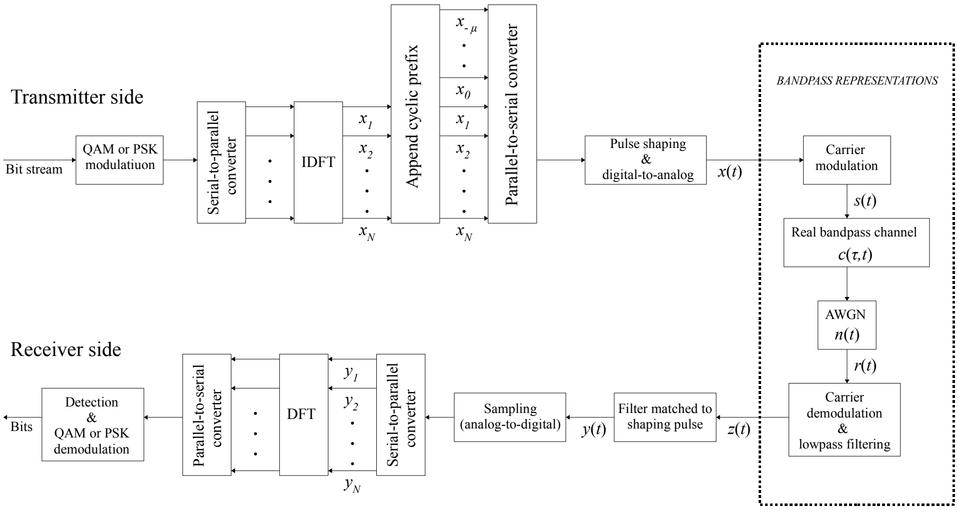


Figure 4.3: Block diagram of an OFDM system. All representations inside the tiny-dashed box are bandpass whereas all terms outside the box are complex baseband entities.

The discrete-time matrix-vector formulation presented in the previous section implied the basic ideas of OFDM and indicated how the discrete-time channel was decomposed into N orthogonal subchannels by the IDFT/DFT operations. In this section we investigate the OFDM system with continuous-time transmission on a wireless multipath channel, thus reflecting a more realistic setup compared to the simplified discrete-time system.

At this point it is worth to specifically emphasize that only a few of the blocks in **Figure 4.3** are concerned within the scope of this thesis. The figure serves to provide an overall and (to some extend) realistic picture of an OFDM system, from transmitter to receiver. In particular, what is important for the remaining parts of this thesis is the frequency domain input-output relationship obtained in (4.7). Accordingly, we now derive a similar channel input-output relationship for the scenario with continuous-time transmission.

4.3.1 Pulse Shaping and Transmission

We now consider the procedure of converting the prefixed discrete-time signal $\boldsymbol{x}_{\text{cp}}$ into an analog complex baseband signal. The $\mu+1+N$ samples of $\boldsymbol{x}_{\text{cp}}$ are converted to an analog signal $x(t)$ by *pulse shaping*. Specifically, with a pulse $\Psi(t)$ confined to the interval $[0, T_{\text{s}}]$, recall **Figure 2.4** on page **16**, the continuous-time pulse shaped signal is defined as

$$x(t) := \sum_{n=-\mu}^{N} x_n \Psi(t - nT_{\text{s}}), \quad t \in \big[-(\mu+1)T_{\text{s}}, NT_{\text{s}}\big]. \tag{4.8}$$

Expressing $x(t)$ in terms of the original symbols $\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_N$, yields

$$x(t) = \sum_{m=1}^{N} \widehat{x}_m \underbrace{\frac{1}{\sqrt{N}} \sum_{n=-\mu}^{N} \Psi(t - nT_{\text{s}}) \exp\Big(i2\pi(m-1)\tfrac{n-1}{N}\Big)}_{\Phi_m(t)}, \tag{4.9}$$

i.e. the complex baseband signal $x(t)$ is a superposition of N carrier waveforms $\Phi_m(t)$, $m = 1, 2, \ldots, N$, each scaled by the corresponding complex symbol $\widehat{x}_m$, and all defined on the same time interval of length $(\mu + 1 + N)T_{\text{s}}$. The distinction between the two different interpretations of $x(t)$ given in (4.8) and (4.9) is obtained by recalling **Figure 4.1**. The expression in (4.8) clarifies the transmitted signal as a pulse train while the expression in (4.9) clarifies the transmitted signal as a multicarrier waveform.

As depicted in **Figure 4.3** we only consider the equivalent lowpass representations of the entire system, i.e. the real bandpass representations inside the tiny-dashed box in the figure are not of primary concern (only their complex baseband representations). Hence, for the multipath channel we consider the equivalent lowpass time-varying channel impulse response $g(\tau, t)$, which was introduced and described in detail in the previous chapter.

Assumptions:

- The channel is linear.
- $g(\tau, t) = g(\tau, 0)$ for all $t \in \big[-(\mu+1)T_{\text{s}}, NT_{\text{s}}\big]$, i.e. the channel is time-invariant over the period of duration $(\mu+1+N)T_{\text{s}}$ when $x(t)$ is transmitted. Makeshift, and with a slight abuse of notation we let $g(\tau) := g(\tau, 0)$.
- The impulse response is non-zero for at most a period of duration $(\mu + 1)T_{\text{s}}$ and more specifically, $g(\tau) = 0$ whenever $\tau < 0$ or $\tau > (\mu + 1)T_{\text{s}}$.

- Perfect time and frequency synchronization between transmitter and receiver.

The above assumptions are similar to those imposed on the discrete-time system in the previous section. We assume a linear channel such that the output signal, in the absence of noise, equals the convolution of the input signal with the 'temporally' time-invariant channel impulse response. Furthermore, we assume to have selected a cyclic prefix of sufficient duration such that no ISI is present. Finally, synchronization issues are not within the scope of this thesis.

The tiny-dashed box in **Figure 4.3** reflects the procedure of transmitting signals in practice, i.e. transmission of real signals on carrier frequency f_c (and not complex baseband signals). However, we skip this entire part and investigate the received signal after carrier demodulation and lowpass filtering, i.e. we consider

$$z(t) := \{x * g\}(t) + w(t),$$

where all terms are in accordance with those given in **Figure 4.3**. The transmitted baseband signal $x(t)$ is convolved with the equivalent lowpass channel impulse response $g(\tau)$ and complex circular symmetric white Gaussian noise is added.

4.3.2 Matched Filtering and Sampling

To maximize the signal-to-noise ratio (SNR) before sampling of the received baseband signal, yet another filter is applied.

Definition 4.3.1. (Matched filter)
Let $\Psi(t)$ be a signal confined to the interval $[0, T_{\text{s}}]$. A filter with impulse response $f(\tau) = \Psi\big(T_{\text{s}} - \tau\big)$ is called a *matched filter* to the signal $\Psi(t)$.

Define the mirrored pulse $\Psi_{\text{mf}}(t) := \Psi\big(T_{\text{s}} - t\big)$ and consider the convolution

$$\big\{\Psi * \Psi_{\text{mf}}\big\}(\tau) = \int_{-\infty}^{\infty} \Psi(t)\Psi_{\text{mf}}(\tau - t)\mathrm{d}t = \underbrace{\int_{-\infty}^{\infty} \Psi(t)\Psi\big(t + T_{\text{s}} - \tau\big)\mathrm{d}t}_{(\star)}.$$

We may consider the integral $(\star)$ as an autocorrelation of the pulse $\Psi(t)$ evaluated at lag $T_{\text{s}} - \tau$. Although no stochastic terms are involved we still refer to $(\star)$ as an autocorrelation, and we define

$$R_{\Psi\Psi}(\tau) := \int_{-\infty}^{\infty} \Psi(t)\Psi\big(t + \tau\big)\mathrm{d}t.$$

Notice that $R_{\Psi\Psi}(0)$ reports the power content of the pulse $\Psi(t)$ and furthermore, since $\Psi(t)$ is confined to the interval $[0, T_{\text{s}}]$, it follows that $R_{\Psi\Psi}(\tau) = 0$ whenever $|\tau| \geq T_{\text{s}}$.

A filter with impulse response $\Psi_{\mathrm{mf}}(t)$, i.e. a filter matched to the shaping pulse $\Psi(t)$, is applied to the baseband signal $z(t)$ and the output signal is

$$\begin{aligned} y(t) &:= \big\{x * g * \Psi_{\mathrm{mf}}\big\}(t) + v(t), \qquad v(t) := \{w * \Psi_{\mathrm{mf}}\}(t) \\ &= \sum_{n=-\mu}^{N} x_n \cdot \big\{g * R_{\Psi\Psi}\big\}\big(t - nT_{\mathrm{s}}\big) + v(t). \end{aligned}$$

Finally, the signal $y(t)$ is sampled at integer multiples of the sampling time T_{s}. The first $\mu+1$ samples corresponding to the cyclic prefix are discarded and the remaining samples of interest are for $k = 1, 2, \ldots, N$, given by

$$y_k := y(kT_{\mathrm{s}}) = \sum_{n=-\mu}^{N} x_n \cdot \big\{g * R_{\Psi\Psi}\big\}\big((k-n)T_{\mathrm{s}}\big) + v(kT_{\mathrm{s}}). \tag{4.10}$$

A structural exploration of (4.10) reveals (not surprisingly) that we are dealing with a convolution. Hence, by assigning the function $\tilde{g}(t) := \big\{g{*}R_{\Psi\Psi}\big\}(t)$ we may similarly to (4.1) in the previous section, express the sample-wise input-output relationship (4.10) in matrix-vector form according to

$$\boldsymbol{y} = \widetilde{\mathbf{G}}\boldsymbol{x}_{\mathrm{cp}} + \boldsymbol{v}. \tag{4.11}$$

The concatenated response $\tilde{g}(t) = \big\{g{*}R_{\Psi\Psi}\big\}(t)$ is a slightly stretched and altered version of the 'true' channel impulse response. Heuristically, the 'true' channel has been 'convolutional shaped' by the autocorrelation function $R_{\Psi\Psi}(\tau)$ associated with the pulse used to shape the symbols intended for transmission. An illustrative example of this channel shaping phenomena is provided by **Figure 4.4**.

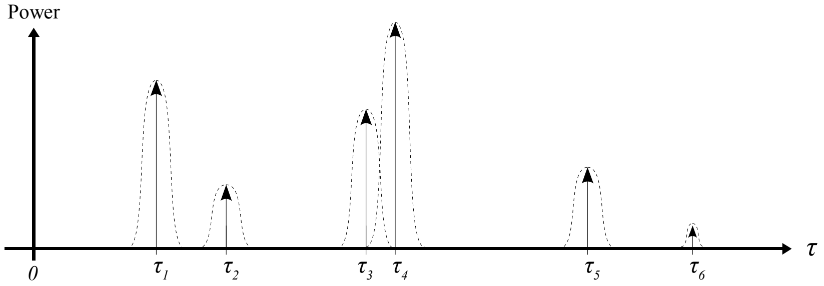


Figure 4.4: Signal echoes of a multipath channel (arrows) being shaped by an autocorrelation pulse (dashed function).

Applying exact similar arguments as in the previous section allow us to manipulate the matrix-vector relationship (4.11) and reshape the linear convolution matrix $\widetilde{\mathbf{G}}$ into a circulant square matrix $\mathbf{G}_{\Psi}$. The reformulated linear system of equations reads

$$\boldsymbol{y} = \mathbf{G}_{\Psi}\boldsymbol{x} + \boldsymbol{v}, \tag{4.12}$$

and conceptually there is no difference between (4.12) and the expression (4.3) obtained in the previous section. Only, we have emphasized by the subscript attached on $\mathbf{G}_{\Psi}$, that the channel we 'observe' has been altered by certain properties of the shaping pulse employed in the communication system. In particular, the frequency domain properties of $R_{\Psi\Psi}(\tau)$ turns out to be crucial - a fact that we elaborate on in the forthcoming chapter. **Lemma 4.2.1** applies directly to $\mathbf{G}_{\Psi}$ as well, and we can essentially repeat all arguments from previous section (explaining the amount of attention payed to this discrete-time scenario). Specifically, we can write $\mathbf{G}_{\Psi}$ in terms of its eigenvalue decomposition and since the matrix is circulant, the N-point DFT matrix $\mathbf{F}$ plays the exact same role in this case. Hence, applying $\mathbf{F}$ to the observed $\boldsymbol{y}$ at the receiver side yields, equivalent to (4.7), the frequency domain input-output relationship

$$\widehat{y}_n = \lambda_n \widehat{x}_n + v_n, \quad n = 1, 2, \ldots, N. \tag{4.13}$$

Each eigenvalue λ_n can be calculated from a Fourier transform relationship to the first column of $\mathbf{G}_{\Psi}$, as was shown in the previous section. Therefore, letting

$$\boldsymbol{q} = \Big[\tilde{g}(0), \tilde{g}(T_{\mathrm{s}}), \tilde{g}(2T_{\mathrm{s}}), \ldots, \tilde{g}\big((N-1)T_{\mathrm{s}}\big)\Big]^{\top}$$

denote the first column of $\mathbf{G}_{\Psi}$, it follows that

$$\big[\lambda_1, \lambda_2, \ldots, \lambda_N\big]^{\top} = \sqrt{N}\mathbf{F}\boldsymbol{q},$$

and recall that we defined $\tilde{g}(t) = \big\{g * R_{\Psi\Psi}\big\}(t)$, i.e. a convolution of two functions. Hence, each eigenvalue λ_n, is given as the product between the frequency response of the channel and the frequency response of the autocorrelation pulse, since the sampled version of the convolution in $\boldsymbol{q}$ will turn into a product when exposed to the N-point DFT.

4.4 OFDM Drawbacks

We close this chapter by pointing out a few selected drawbacks of OFDM. This is merely to emphasize in brief manners that a variety of technical issues need to be managed in practice. Recall the time domain OFDM signal $x(t)$ given in (4.9). Since this signal is a superposition of complex-valued symbols modulated onto orthogonal carrier waves, the signal envelope $|x(t)|$ is subject to strong and notable variations **[4, Section 2.1.2]**. Consider, for instance, a time instant where all terms in the superposition add constructively in phase. Compared to the average signal power level during transmission, this constructive adding will cause a significant peak in the instant transmission power. Such sudden and extreme fluctuations in the signal power may cause severe signal distortion due to saturation of the amplifier in the transmitter. If the amplifier is saturated, signal power may be introduced at frequencies outside the bandwidth where the system was supposed to operate. The above mentioned issue is commonly referred to as the problem of high *peak-to-average power ratio*. *Pre-coding* the sequences of data symbols may be used to deal with the problem. The coding allows only sequences of symbols known to have a relatively low peak-to-average power ratio, however, the cost of this power controlling is a redundant overhead introduced by the coding.

Another important issue, that in particular has to be considered for OFDM, is synchronization. Misalignments between transmitter and receiver may degrade notably the transmission quality **[4, Section 2.3.3]**, implying an unacceptable *bit-error-rate* (BER) at the receiver side. Consider, for instance, a misalignment in the overall carrier frequency f_c of the system. In such a case the subcarrier orthogonality (recall **Lemma 4.1.1**) is no longer maintained and the information transmitted across the different subchannels will interfere. This effect is commonly referred to as *inter-carrier interference* (ICI). Time domain synchronization and *phase noise* effects are important aspects too (see **[4]**), however, we do not consider such issues in this thesis. Overall, we always assume no presence of ISI and ICI, i.e. the input-output relationship in (4.13) is valid.

Chapter 5

Channel Estimation

In this chapter we describe a number of state-of-the-art channel estimation methods and algorithms for communication systems using OFDM. All methods presented are based on *pilot symbol* transmissions, i.e. systematic transmission of symbols known to both transmitter and receiver. From the pilot symbol data the channel can be estimated and such approaches are referred to as *pilot assisted channel estimation* (PACE). Critical issues concerned with the channel estimation emerge from so-called *power leakage* effects and special attention is therefore payed to this subject. Simulation results concerning performance evaluation of a few selected PACE algorithms constitute the closing part of this chapter.

In general, channel estimation techniques can employ both frequency domain and time domain properties of the channel. Solely, we point our emphasis towards estimation of the channel in the frequency domain and we make no use of time direction interpolation or filtering. Channel estimation techniques which are not based on pilot symbol transmissions exist, however, such approaches are not within the scope of this thesis. The interested reader may consult **[7, Section 5.1]** and the references therein for further insight on methods which do not employ pilot symbols.

5.1 Unused Subchannels

As mentioned earlier, multicarrier modulation schemes are beneficial due to their flexibility in usage of bandwidth. Even though an OFDM system is designed with a total of N subchannels evenly spaced in frequency, not all subchannels need to be employed simultaneously. Any set of subchannels can be made ‘inactive’ whenever required, hereby adjusting the effective usage of electromagnetic spectrum. This flexibility is advantageous since spectral allocations may differ from country to country and also it allows for operators to squeeze new systems into frequency bands already appointed or financially acquired.

Consider an OFDM system designed with a total of N subchannels where only a subset of $N_u \leq N$ subchannels are used for actual transmissions. The remaining $N - N_u$ subchannels are ‘shut down’, simply by not transmitting on these frequency bands, i.e. in (4.9), the corresponding symbols from the digital constellation are replaced by zeroes (no power). Often, a number of subchannels at each edge of the system bandwidth are shut down by default to serve as *guard bands* **[12, 18]**. The pulse shaping filter $\Psi(t)$ in (4.8) will typically introduce some attenuation at the edges of the bandwidth, and this attenuation depends on the frequency domain 'cut off' properties of $R_{\Psi\Psi}(\tau)$. For a given system, specific choices of $\widehat{R}_{\Psi\Psi}(f)$ are numerous, however, this autocorrelation pulse must satisfy the *Nyquist criterion* to avoid ISI prior to sampling of the received signal. *Raised cosine pulses* or *root-raised cosine pulses* are often used in practice but this fact is relatively non-essential for our objective. Hence, we will not elaborate further on this subject, only we stress the fact that an appropriate number of subchannels are kept inactive at both edges of the overall system bandwidth. The consequences of this are revealed later on when the channel transfer function is to be estimated from pilot symbol data.

In the following we explain in detail how pilot symbols are transmitted and subsequently used for channel estimation at the receiver. Although the setup is presentable in its full generality, we specifically mind our exposition towards the 3GPP setup proposed for LTE. This choice is made to improve readability, insight and interpretations, especially for readers somewhat unfamiliar with OFDM and channel estimation in general.

5.2 LTE System Setup

The system parameters presented in the following are based on the 3GPP technical specifications **[22]** and **[23]**. In order to get a feeling of the possible dimensionality of the estimation task, we consider the LTE setup proposed with parameters as given in **Table 5.1** below. We do not discuss the selections and the associated justifications for the actual parameter values since indeed, such a discussion is out of scope of this thesis.

Sampling time	System bandwidth	DFT size	Subcarrier spacing	Active subchannels	Cyclic prefix duration
T_s 32.55ns	B 30.72MHz	N 2048	Δf 15kHz	N_u 1200	$(\mu+1)T_s$ $144T_s$

Table 5.1: Common set of system parameters for downlink OFDM in LTE.

As can be seen from the above table, only 58.6% of the $N = 2048$ subchannels are active, i.e. the effective number of subchannels used is $N_u = 1200$. A frequency domain view of the entire system is provided by **Figure 5.1**, and notice how the bandwidth is centered around $f = 0$ (baseband). Hence, when upconverting to the carrier frequency f_c for actual transmission, f_c corresponds to the center frequency of the spectrum occupied by the OFDM system in practice.

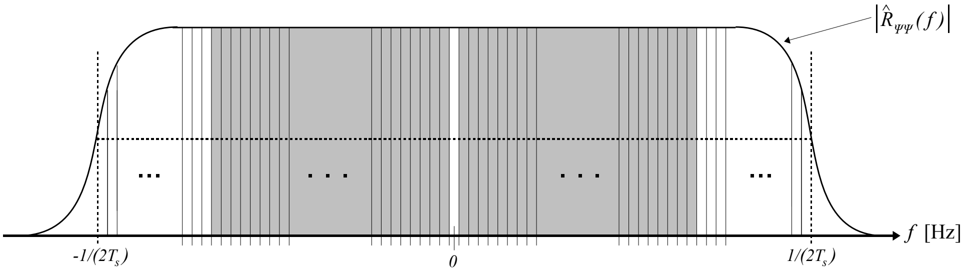


Figure 5.1: Frequency domain view of the overall system with $N_u/2 = 600$ active subchannels on each side of the zero frequency (grey-shaded areas). An example showing the 'cut off' properties of a shaping filter autocorrelation pulse is given too.

As revealed in **[22]**, the subchannel with baseband center frequency equal to zero is preserved inactive for technical reasons. Hence, 600 active subchannels are present to the left of zero, and similarly, 600 active subchannels to the right. This means that the effective bandwidth of the system is $\widetilde{B} \approx 18\text{MHz}$. More subchannels can be shut down if needed, e.g. to fit a 5MHz or 10MHz band set aside for a given application. Furthermore, as already apparent from above, the effective spectrum used need not be coherent/contiguous. Hence, multicarrier systems are quite flexible in usage of bandwidth.

5.2.1 Baseband Reference Frequency

As mentioned in the previous section, the baseband reference frequency $f = 0$ informs the midpoint of the frequency band in use. According to our definition of the DFT in (2.2) on page **13**, the reference frequency $f = 0$ corresponds to the leftmost point of the band. Hence, when upconverting to the carrier frequency f_c in this case, the particular value f_c does not correspond to the midpoint of the frequency band used in practice. Adapting to the LTE setup we now identify the frequency shift (phase rotation) needed in order to make the two references coincide.

Consider the discrete-time signal $\boldsymbol{x} = \left[x_1, x_2, \ldots, x_N\right]^\top$ and our previous definition of the DFT in (2.2). From a straightforward change of summation indices we notice that

$$\widehat{x}_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} x_n \exp\left(-i2\pi(k-1)\tfrac{n-1}{N}\right), \quad k = 1, 2, \ldots, N$$

$$= \frac{1}{\sqrt{N}} \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x_{m+\frac{N}{2}+1} \exp\left(-i2\pi(k-1)\tfrac{m}{N}\right) \underbrace{\mathrm{e}^{-i\pi(k-1)}}_{(\star)},$$

where the complex exponential term ($\star$) can be seen as $k-1$ concatenated phase rotations of π, i.e. the term simply alternates between 1 and -1 as a function of k. In general, a shift in frequency results in a phase rotation and in our particular case the phase rotation is quite simple since the shift in frequency is by half the bandwidth. Alternatively, we may apply the DFT from (2.2) on the shifted signal $\left[x_{N/2+1}, \ldots, x_N, x_1, \ldots, x_{N/2}\right]^\top$ instead of applying

$$\widehat{x}_k := \frac{1}{\sqrt{N}} \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x_{m+\frac{N}{2}+1} \exp\left(-i2\pi(k-1)\frac{m}{N}\right), \quad k = 1, 2, \ldots, N. \tag{5.1}$$

The redefined DFT in (5.1) is forthright implemented in matrix form according to

$$\mathbf{F} := \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & \omega^{-N/2} & \omega^{-2N/2} & \cdots & \omega^{-(N-1)N/2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{-1} & \omega^{-2} & \cdots & \omega^{-(N-1)} \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^{1} & \omega^{2} & \cdots & \omega^{(N-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{N/2-1} & \omega^{2(N/2-1)} & \cdots & \omega^{(N-1)(N/2-1)} \end{bmatrix}, \tag{5.2}$$

and notice how this ‘new’ DFT matrix is nothing but a row-shifted version of the matrix originally defined in (2.3). One can argue that we could as well have introduced the DFT as above from the very beginning, and this is true indeed. However, to our opinion the commonly used definition (2.2) is easier to interpret at first glance and the redefining made in this section should be straightforward to comprehend at this point.

5.2.2 Pilot Symbol Pattern

In LTE a certain number M of the N_u active subchannels are utilized for pilot symbol transmissions. The pilot symbols are distributed in both time and frequency with a configuration depending on the specific application. If the multipath channel is highly frequency selective, the pilot symbols need to be closely spaced in the frequency domain in order to capture the rapid variations of the channel transfer function. Similarly, if the receiver is highly mobile and the environment is changing fast, the pilot symbols need to be closely spaced in time. Obviously, there are numerous ways to distribute the pilot symbols simultaneously in these two domains. Many configurations have been investigated and the 3GPP proposal for LTE is a diamond alike pattern which we elaborate on later in this section.

Consecutive transmissions of OFDM symbols are arranged in so-called *radio frames*. The duration of one radio frame is ten milliseconds. Each radio frame consists of ten *sub frames* and each sub frame consists of two *slots*. Every slot contains seven OFDM symbols including cyclic prefixes. The entire structure of a radio frame is depicted in **Figure 5.2**.

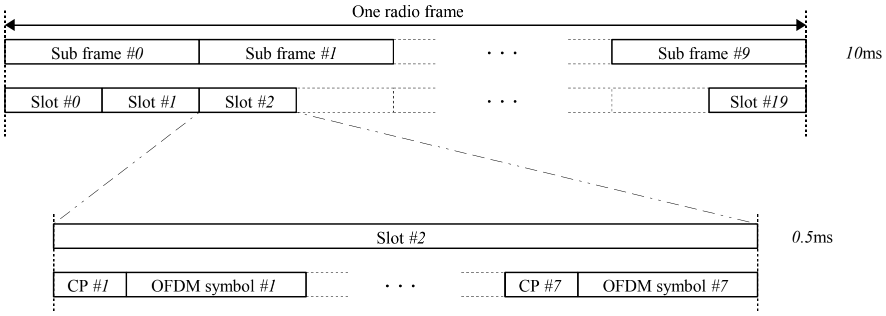


Figure 5.2: The structure of a radio frame.

The duration of a sub frame is one millisecond and the duration of a slot is therefore half a millisecond. The first OFDM symbol in every slot holds a cyclic prefix of slightly prolonged duration ($160T_s$), while the remaining six OFDM symbols all hold a cyclic prefix of duration $144T_s$ as stated in **Table 5.1**. This prolonging is merely a matter of making the durations sum up to fit the radio frame structure. We may verify that indeed the duration of each slot is half a millisecond since

$$\big(160 + 2048 + 6 \cdot (144 + 2048)\big) \cdot \underbrace{32.55\text{ns}}_{T_s} = 0.000499968\text{s} \approx 0.5\text{ms}.$$

Hence, a transfer of 14000 OFDM symbols per second is possible with the system configuration presented here. With $N_u = 1200$ active subchannels and for instance $M = 200$ of these used for pilot symbols, we are left with 1000 subchannels carrying non-redundant data. Assume for simplicity this pilot symbol configuration for every single OFDM symbol. Use of QPSK modulation, i.e. two bits per constellation symbol, then yields a theoretical data rate of 28Mbits/s and if 64-QAM is utilized, the theoretical rate is 84Mbits/s. The reason for stating these numbers is simply to suggest the aims and rigorous demands of LTE. Employing additional subchannels and usage of multiple-input-multiple-output (MIMO) techniques imply even higher data rates.

The diamond alike pilot symbol pattern suggested for LTE is shown in the time-frequency plot in **Figure 5.3**. Partly, the figure illustrates the joint configuration of the pilot symbols in time and frequency, and furthermore the figure suggests the configuration when both transmitter and receiver are equipped with two antennas (i.e. MIMO). The left-hand side depicts the pilot symbol configuration at the first transmit antenna port. The right-hand side depicts the pilot symbol configuration at the second transmit antenna port. Notice carefully how the second antenna port is not allowed to utilize the time-frequency allocations occupied by pilot symbols from the first antenna port and vice versa **[23]**.

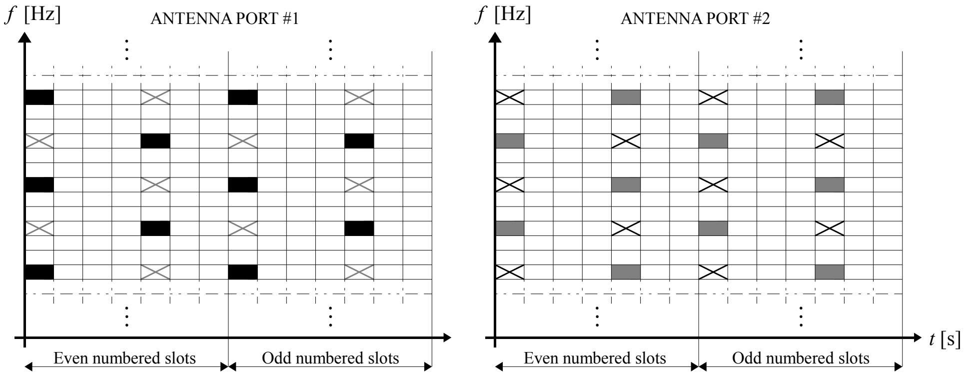


Figure 5.3: Pilot symbol patterns. Black and grey boxes indicate pilot symbols while the white boxes indicate non-redundant data symbols. Crossed boxes indicate time-frequency allocations which are not allowed to be occupied by the current antenna port.

When at most two antenna ports are used, the temporal gaps between OFDM symbols containing pilot symbols are fixed as shown in **Figure 5.3**. That is, for both even numbered and odd numbered slots, pilot symbols are contained only in the first and the fifth OFDM symbol. When three or four antenna ports are used, the temporal configuration is slightly altered at these additional ports, however, we do not exemplify more than two antenna ports in this thesis. The spacing between adjacent pilot symbols in the frequency direction is six subchannels by default. Since $N_u = 1200$ subchannels are active it means that $M = 200$ of these are equipped with pilot symbols if the considered OFDM symbol is supposed to carry pilot symbols (first or fifth in every slot).

Recall from **Figure 5.1** that the baseband center frequencies of the N_u active subchannels are

$$-\tfrac{N_u}{2}\Delta f, \ldots, -2\Delta f, -\Delta f \ , \ \Delta f, 2\Delta f, \ldots, \tfrac{N_u}{2}\Delta f,$$

with the subcarrier spacing $\Delta f = 1/(NT_{\rm s}) = 15\text{kHz}$. We define the following ordered set of indices

$$\mathcal{I} := \Big(-\tfrac{N_u}{2}, \ldots, -2, -1 \ , \ 1, 2, \ldots, \tfrac{N_u}{2} \Big),$$

and refer to its elements as $\mathcal{I}(1), \mathcal{I}(2), \ldots, \mathcal{I}\big(N_u\big)$. In practical accordance with **[23]**, we define the pilot symbol indices

$$p(m) := \mathcal{I}\Big(\big[\eta + \eta_{\rm shift} \ \mathrm{mod}\ 6\big] + 6\cdot(m-1) + 1\Big), \quad m = 1, 2, \ldots, M, \qquad (5.3)$$

where in (5.3) the term $\big[\eta+\eta_{\rm shift} \ \mathrm{mod}\ 6\big] \in \{0,1,2,3,4,5\}$ is a technical displacement mechanism defined such that

$$\eta := \begin{cases} 0, & \text{at the first OFDM symbol in any slot at antenna port \#1} \\ 3, & \text{at the fifth OFDM symbol in any slot at antenna port \#1} \\ 3, & \text{at the first OFDM symbol in any slot at antenna port \#2} \\ 0, & \text{at the fifth OFDM symbol in any slot at antenna port \#2,} \end{cases}$$

and where $\eta_{\text{shift}} := \big[\text{'Cell ID' mod } 6\big]$ depends specifically on the current cell identification and the base station from where the signal originated. Hence, from (5.3) it can be seen that the grid of pilot symbols is systematically exposed to a cell-dependent offset such that, in practice, the pilot symbols from two adjacent cells do not employ the same frequency bands. For simplicity we assume that η_{shift} stay fixed, say $\eta_{\text{shift}} = 2$, i.e. we assume that the receiver does not leave the current cell during transmission. With $\eta = 0$ and $\eta_{\text{shift}} = 2$ we collect the indices from (5.3) in the set

$$\mathcal{P} := \Big\{p(m) \mid m = 1, 2, \ldots, M\Big\},$$

which specifically reads

$$\mathcal{P} = \Big\{\underbrace{-598, -592, -586, \ldots, -10, -4}_{\text{100 negative indices}}\ , \underbrace{3, 9, 15, \ldots, 585, 591, 597}_{\text{100 positive indices}}\Big\}.$$

Also, we should consider the above set with the number 3 added to all elements, corresponding to $\eta = 3$ and $\eta_{\text{shift}} = 2$. At the location in $\mathcal{P}$ where the sign changes it is crucial to remark the aberrant index gap of seven (and not the common spacing of six). This occurrence is a result of the inactive subchannel with baseband center frequency equal to zero. In **[23]** this irregular pilot symbol gap is seemingly not specified/selected for any particular reason, however, we shall see later on that its presence comes to pose technical difficulties. Nonetheless, we are now through with the technical LTE parameter descriptions, i.e we are well-prepared to consult the task of pilot assisted channel estimation (PACE) for OFDM.

5.3 PACE Framework

The remaining sections of this chapter are based on the approaches and modeling suggestions in **[9, 10, 11, 12, 18]**. Common for these state-of-the-art references concerned with PACE for OFDM is a similar and comparative framework. In particular, their baseband multipath channel model coincide, with only a few variations in individual designs. Overall, the primary channel model is the same and we have chosen to compare with a reference channel model specified by 3GPP **[22, Annex B.2]**. The objective with the following sections is to provide a throughout insight on state-of-the-art PACE for OFDM. In particular, we have aimed towards replication of the results reported in **[7]**.

5.3.1 Baseband Models

In **Chapter 3** we considered a Rayleigh fading multipath channel with an associated model of the channel impulse response reading

$$g(\tau, t) = \sum_{l=1}^{L(t)} \alpha_l(t)\delta\big(\tau - \tau_l(t)\big), \qquad \alpha_l(t) \in \mathbb{C}.$$

For each echo the associated time-varying coefficient $\alpha_l(t)$ comprise a WSS zero-mean complex Gaussian process and the $L(t)$ different processes are mutually uncorrelated (the US property). We now adapt to the state-of-the-art references given above and consider the slightly more static model

$$g(\tau, t) = \sum_{l=1}^{L} \alpha_l(t)\delta\big(\tau - \tau_l T_{\mathrm{s}}\big).$$

Here, the number L of channel echoes is considered fixed and known a priori and the delay parameters $\tau_1 T_{\mathrm{s}}, \tau_2 T_{\mathrm{s}}, \ldots, \tau_L T_{\mathrm{s}}$ are static too[1]. The complex amplitudes $\alpha_l(t)$ are still considered time-variant and some of the above mentioned authors impose no correlation structure in the time direction while others do. The latest incoming signal echo with associated delay $\tau_L T_{\mathrm{s}}$ is assumed to always arrive no later than $(\mu + 1)T_{\mathrm{s}}$, i.e. the duration of the cyclic prefix. Hereby, no ISI will occur between adjacent OFDM symbols. The time-varying channel transfer function of the multipath channel is given as the Fourier transform of $g(\tau, t)$ with respect to the delay variable τ, i.e.

$$h(f, t) = \int_{-\infty}^{\infty} g(\tau, t)\mathrm{e}^{-i2\pi f\tau}\mathrm{d}\tau = \sum_{l=1}^{L} \alpha_l(t)\mathrm{e}^{-i2\pi f \tau_l T_{\mathrm{s}}}, \qquad f \in \mathbb{R}.$$

When data symbols are transmitted across the N_u active subchannels of the OFDM system, we know from **Chapter 4**, that each individual data symbol is distorted by the channel frequency response evaluated at the corresponding subcarrier frequency. Hence, the N_u coefficients

$$h_n(t) := h(f, t)\Big|_{f=\mathcal{I}(n)\Delta f} = \sum_{l=1}^{L} \alpha_l(t)\mathrm{e}^{-i2\pi \frac{\mathcal{I}(n)}{N}\tau_l}, \qquad n = 1, 2, \ldots, N_u, \tag{5.4}$$

are the complex numbers that need to be estimated in order to recover the transmitted data symbols. Recall the frequency domain input-output relationship (4.13), and from this equation we state the corresponding relationship for our LTE system under consideration as

$$\widehat{y}_n = \widehat{x}_n h_n(t) + w_n, \quad n = 1, 2, \ldots, N_u.$$

Notice that only N_u symbols are involved since $N - N_u$ subchannels serve as guard bands as explained in the beginning of this chapter. We assume that all N_u active subcarriers are in the 'flat region' of the shaping filter frequency response as depicted in **Figure 5.1**. Without loss of generality, we assume this flat response to hold unit gain and therefore it has been omitted above. The point is that the shaping filters comprise a design issue of the communication system, and hence their flat gain in the frequency domain will be known.

---

[1] Notice how the delay parameters are now modeled in terms of the sampling time T_{s}, as also suggested in **[9, 11]**. This modeling turns out to be appropriate later on.

Since the channel is assumed to be time-invariant during the transmission time of each OFDM symbol, we omit the time-dependency on the channel coefficients from this point and on. The subchannel-wise relationship above then allows for the overall matrix-vector formulation

$$\widehat{\boldsymbol{y}} = \widehat{\mathbf{X}}\boldsymbol{h} + \boldsymbol{w}, \tag{5.5}$$

where the input symbols (the information symbols) have been appropriately collected in the matrix $\widehat{\mathbf{X}} := \operatorname{diag}\big(\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_{N_u}\big)$. Notice carefully, that for each OFDM symbol transmitted we use the same model (5.5) over and over again.

5.3.2 Pilot Symbol Observations

The model spelling the input-output relationship at the pilot symbol positions $\mathcal{P}$, consists of an appropriate subset of the complete signal model (5.5) and is given by

$$\widehat{\boldsymbol{y}}_{\mathrm{p}} = \widehat{\mathbf{X}}_{\mathrm{p}}\mathbf{T}\boldsymbol{\alpha} + \boldsymbol{w}_{\mathrm{p}}.$$

The subscript ‘p’ is present as abbreviation for ‘pilots’ and furthermore

- $\widehat{\boldsymbol{y}}_{\mathrm{p}} = \big[\widehat{y}_{p(1)}, \widehat{y}_{p(2)}, \ldots, \widehat{y}_{p(M)}\big]^{\top}$ is a vector containing all observations from the M pilot symbol positions.
- $\widehat{\mathbf{X}}_{\mathrm{p}} = \operatorname{diag}\big(\widehat{x}_{p(1)}, \widehat{x}_{p(2)}, \ldots, \widehat{x}_{p(M)}\big)$, i.e. a diagonal matrix of the transmitted pilot symbols. Notice that this matrix is known to the receiver.
- $\mathbf{T}$ is an $M \times L$ matrix depending on the delays $\boldsymbol{\tau} = [\tau_1, \tau_2, \ldots, \tau_L]^{\top}$ and the pilot symbol positions $\mathcal{P}$ in such a way that its (m, l)’th entry reads

$$\mathbf{T}(m, l) := \exp\Big(-i2\pi\tfrac{p(m)}{N}\tau_l\Big). \tag{5.6}$$

- $\boldsymbol{\alpha} = \big[\alpha_1, \alpha_2, \ldots, \alpha_L\big]^{\top}$ is a vector of the complex channel amplitudes.
- $\boldsymbol{w}_{\mathrm{p}} = \big[w_{p(1)}, w_{p(2)}, \ldots, w_{p(M)}\big]^{\top}$ contains as components the complex AWGN contributions at the pilot symbol positions.

Initially, the observed vector $\widehat{\boldsymbol{y}}_{\mathrm{p}}$ is multiplied by the inverse of $\widehat{\mathbf{X}}_{\mathrm{p}}$, thereby obtaining initial estimates of the channel transfer function at the M subcarriers where pilot symbols have been transmitted. These initial estimates, sometimes referred to as *zero-forcing* estimates **[9]**, yields the ‘true values’ additively embedded in zero mean complex Gaussian noise, i.e.

$$\boldsymbol{h}_{\mathrm{zf}} := \big(\widehat{\mathbf{X}}_{\mathrm{p}}\big)^{-1}\widehat{\boldsymbol{y}}_{\mathrm{p}} = \mathbf{T}\boldsymbol{\alpha} + \big(\widehat{\mathbf{X}}_{\mathrm{p}}\big)^{-1}\boldsymbol{w}_{\mathrm{p}}.$$

The power (variance) of each noise component is scaled by the inverse power of the corresponding pilot symbol - suggesting that the noise power can be forced toward zero by boosting the power content of the pilot symbols. *However, we assume throughout the thesis that all pilot symbols hold unit power, and therefore the noise statistics remain unchanged.* Finally, we end up with the crucial observation model

$$\boldsymbol{h}_{\text{zf}} = \mathbf{T}\boldsymbol{\alpha} + \boldsymbol{w}, \tag{5.7}$$

where a slight abuse of notation is introduced by omitting the subscript 'p' at the noise vector. This is merely to make the system of linear equations appear as simple and familiar as possible - namely a vector of observations expressed in terms of a matrix-vector product embedded in noise. The left-hand side of (5.7) is known to the receiver and from these M complex numbers the PACE task is essentially to estimate the $N_u - M$ remaining complex numbers associated with the subcarrier positions of the non-redundant data symbols.
The PACE framework has now been presented and with the observation model (5.7) at hand, we proceed by elaborating on the *amount of channel knowledge* that can actually be extracted from observing the frequency response at selected positions.

5.4 Power Leakage Effects

Any communication system applied in practice is operating with some finite resolution in the time-domain, or equivalently, with limited bandwidth. No system can realize the theoretical tool in use when applying, for instance, a continuous-time Fourier transform. That is, no system can operate with infinite bandwidth since it would require continuous resolution in the time direction.
For the OFDM system considered in this chapter, with LTE parameters as given in **Table 5.1**, the time resolution appear in terms of the sampling time T_{s}. Initially at the receiver side, a continuous-time signal is sampled into N time-domain samples (represented on a T_{s}-sample aligned grid). An N-point DFT is then applied to this discrete-time signal and due to the guard band subcarriers, only N_u useful samples are produced. In this section we elaborate on the effects introduced from the fact that the sampled signal is represented on the discrete-time T_{s}-spaced grid. In **Chapter 4** we obtained a minor interpretation of these effects in terms of **Figure 4.4** where we depicted how the multipath channel is inevitably being altered by the shaping pulses employed in the communication system. We now investigate in detail the nature of these so-called *leakage effects* **[9]**.

Still, denote by $h(f,t)$ the frequency response of the multipath channel. Consider the normalized 'sampling function' defined by

$$\widehat{q}\big(f;A\big) := \frac{1}{\sqrt{N|A|}} \sum_{n \in A} \delta(f - n\Delta f) = \frac{1}{\sqrt{N|A|}} \sum_{n \in A} \delta\Big(f - \tfrac{n}{NT_{\text{s}}}\Big),$$

where the set A represents some specific subcarriers in use and $|A|$ denotes the number of elements in A. For instance, we may select $A = \mathcal{P}$ or perhaps $A = \big\{\mathcal{I}(1), \mathcal{I}(2), \ldots, \mathcal{I}(N_u)\big\}$, but for now it just represents an arbitrary subset of the N potential subcarriers (the guard bands and the zero-subcarrier stay unused by choice). To imitate a noise free sampling procedure of $h(f,t)$ at the selected subcarrier frequencies specified in A, we consider

$$\widehat{s}\big(f,t;A\big) := h(f,t)\widehat{q}\big(f;A\big).$$

Analyzing the behavior of the Fourier inverse of $\widehat{s}(f, t; A)$ reveals the time-domain multipath channel the receiver is ideally able to 'see' on the T_s-spaced sampling grid.

5.4.1 System Response

In **Section 2.4** we mentioned the sampling property of the Dirac delta, and the approach taken here illustrates an application of this property. Applying the Fourier inverse transform on $\widehat{s}(f, t; A)$ we obtain

$$\begin{aligned} s\big(\tau, t; A\big) &= \big\{g(\cdot, t) * q(\cdot\,; A)\big\}(\tau) \\ &= \frac{1}{\sqrt{N|A|}} \sum_{l=1}^{L} \alpha_l(t) \sum_{n \in A} \exp\Big(i2\pi \tfrac{n}{NT_s}\big(\tau - \tau_l T_s\big)\Big). \end{aligned}$$

The crucial part is to analyze the time-domain function $s\big(\tau, t; A\big)$ sampled at integer multiples of the sampling time T_s. More specifically, we consider the N samples

$$s\big(0, t; A\big), s\big(T_s, t; A\big), s\big(2T_s, t; A\big), \ldots, s\big((N-1)T_s, t; A\big), \tag{5.8}$$

and we refer to this sampled version of $s\big(\tau, t; A\big)$ as the *system response*. We illustrate only a few special cases with different choices of the set A. Each choice is deliberately selected to emphasize a crucial fact.

Recall **Section 4.2** where we assumed a discrete-time multipath channel[2] with all responses perfectly aligned on the T_s-spaced sampling grid. In such a case, all multipath delay parameters τ_l, $l = 1, 2, \ldots, L$, are assumed integer-valued and the latest delay τ_L, is assumed inside the cyclic prefix duration. Now consider the system response in the extreme case where all N subcarriers are employed, i.e. with

$$A = \Big\{ -1024, -1023, \ldots\ldots, 1022, 1023 \Big\}, \qquad |A| = N.$$

In this case, with all L delays being integer-valued, the receiver will observe

$$s\big(kT_s, t; A\big) = \frac{1}{N} \sum_{l=1}^{L} \alpha_l(t) \sum_{n=-1024}^{1023} \exp\Big(i2\pi \tfrac{n}{N}\big(k - \tau_l\big)\Big) = \begin{cases} 0\;, & k \neq \tau_l \\ \alpha_l(t), & k = \tau_l. \end{cases}$$

Hence, the receiver observes perfectly all L channel amplitudes at their respective delay positions (recall that we consider no additive noise effects). If no channel component is present at the grid point kT_s, the receiver will simply observe a zero. Consider then the case where just one single delay, e.g. τ_2, is not integer-valued. For any of the grid points the sum of complex exponentials

$$\sum_{n=-1024}^{1023} \exp\Big(i2\pi \tfrac{n}{N}\big(k - \tau_2\big)\Big) \neq 0,$$

---

[2] See [1, **Section 3.4**] and [3, **Section 10.1.1**] for other examples of discrete-time channels.

since $k - \tau_2$ is not integer-valued. Hence, the system response (5.8) is non-zero at all N samples. Among the state-of-the-art references given in the beginning of **Section 5.3**, **[9]** was the first to consider this phenomena, commonly referred to as the *leakage effect*. In practice the channel is of course not nicely aligned with the T_s-spaced sampling grid. Hence, the leakage effect will always be present and the extent of the leakage depends critically on the set A.

By shutting down subcarriers (starting from both edges of the overall system bandwidth) the effective bandwidth of the OFDM system is reduced, e.g. when $A = \{\mathcal{I}(1), \mathcal{I}(2), \ldots, \mathcal{I}(N_u)\}$. Hereby, the leakage effect becomes more conspicuous in the neighborhood of any true channel echo. In fact, channel components of larger power may easily mask channel components of lower power if two such echoes are closely spaced in delay.
The leakage effect becomes even more conspicuous when the effective bandwidth of the OFDM system is not just reduced, but also the samples are taken more distant. This corresponds to the case when $A = \mathcal{P}$. In this case a peaky, periodic behavior can be observed in the system response.

5.5 Channel Estimation Algorithms

With the leakage effect in mind and given a zero-forcing estimate $\boldsymbol{h}_{\mathrm{zf}}$ from (5.7), we now describe three selected state-of-the-art PACE algorithms suggested for OFDM systems. A unified algorithm framework is given in **[7, Section 6.2-6.4]**, and we therefore omit a variety of details since a full overview can be obtained by consulting this reference. The first algorithm presented is included only to illustrate that the leakage effect has to be compensated or circumvented in practice. The remaining two algorithms constitute the methods, recommended in **[7]**, to be used within LTE.

5.5.1 Sample-aligned DFT-based Method

A crude, but very simple channel estimation approach is to ignore the leakage effect and carry out a least-squares estimation procedure of a carefully chosen subset of the system response. More precisely, the particular subset of grid points corresponding to the duration of the cyclic prefix is selected. Provided that the extent of the leakage has not been 'too massive', this subset of the system response still explains a decent amount of the overall channel behavior. Hence, ignoring (5.7) involving only L complex-valued channel amplitudes and L delay parameters (L, $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$ unknown of course), the observation model is summarily replaced by

$$\boldsymbol{h}_{\mathrm{zf}} = \mathbf{F}_{\mathcal{P}} \boldsymbol{\alpha}_{\mathrm{sr}} + \boldsymbol{w}, \tag{5.9}$$

where $\mathbf{F}_{\mathcal{P}}$ denotes an appropriate $M \times (\mu+1)$ sub-matrix extracted from the N-point DFT matrix in (5.2). Furthermore, $\boldsymbol{\alpha}_{\mathrm{sr}} \in \mathbb{C}^{\mu+1}$ represents the system response ('sr') at the first $\mu+1$ grid points, corresponding to the cyclic prefix duration. The matrix $\mathbf{F}_{\mathcal{P}}$ consists, more precisely, of an un-normalized row-column subset of the $N \times N$

DFT matrix $\mathbf{F}$, such that

$$\mathbf{F}_{\mathcal{P}}(m,k) := \exp\Big(-i2\pi \tfrac{p(m)}{N} k \Big), \qquad \begin{matrix} m = 1, 2, \ldots, M \\ k = 0, 1, \ldots, \mu. \end{matrix}$$

From the new, purposely ignorant observation model (5.9), a least squares estimate of $\boldsymbol{\alpha}_{\text{sr}}$ is calculated. Usually, such a least squares estimate would be expressed by

$$\hat{\boldsymbol{\alpha}}_{\text{sr}} = \big(\mathbf{F}_{\mathcal{P}}^H \mathbf{F}_{\mathcal{P}}\big)^{-1} \mathbf{F}_{\mathcal{P}}^H \boldsymbol{h}_{\text{zf}},$$

however, the square matrix $\mathbf{F}_{\mathcal{P}}^H \mathbf{F}_{\mathcal{P}}$ is subject to numerical instability issues (even though its invertibility can be guaranteed by theoretical means, due to the *Vandermonde*-alike structure of $\mathbf{F}_{\mathcal{P}}$). To reduce the numerical issues, the least squares estimate is therefore obtained indirectly from solving[3] the linear system of equations

$$\mathbf{F}_{\mathcal{P}}^H \mathbf{F}_{\mathcal{P}} \boldsymbol{\alpha}_{\text{sr}} = \mathbf{F}_{\mathcal{P}}^H \boldsymbol{h}_{\text{zf}}.$$

From the least squares estimate $\hat{\boldsymbol{\alpha}}_{\text{sr}}$, the frequency response at all subcarrier positions employed in the system is estimated from

$$\hat{\boldsymbol{h}}_{\text{DFT}} := \mathbf{F}_{\mathcal{I}} \hat{\boldsymbol{\alpha}}_{\text{sr}}, \tag{5.10}$$

where $\mathbf{F}_{\mathcal{I}}$ denotes the $N_u \times (\mu + 1)$ sub-matrix straightforwardly obtained by extracting the N_u rows of the complete DFT matrix $\mathbf{F}$, corresponding to the active subcarrier positions. Of course, only the first $\mu + 1$ columns of $\mathbf{F}$ are extracted.

In **[7, Section 6.3]**, an estimator corresponding to (5.10) is described. However, in this reference the underlying channel is assumed to be nicely aligned with the sampling grid. Our approach is somewhat different, since we purposely and falsely assume the system response to be non-zero only at the first $\mu + 1$ grid points. Indeed, the interested reader should consult **[7, Section 6.3]** and the original references given therein, while also an interesting numerical study is conducted in **[7, Appendix C.3]** concerning the ill-conditioning behavior of the matrix $\mathbf{F}_{\mathcal{P}}^H \mathbf{F}_{\mathcal{P}}$.

5.5.2 Enhanced Noise Reduction Algorithm

The following channel estimator was proposed in **[12]** and in **[7, Section 6.3.2]** it is referred to as the *Enhanced Noise Reduction Algorithm* (ENRA). Indeed it can be discussed whether this calling is appropriate or not, however, we will reuse the name anyway. We follow the lines of **[12]** in the description below.

The ENRA is based on the observation model (5.7), which we recast for clarity

$$\boldsymbol{h}_{\text{zf}} = \mathbf{T}\boldsymbol{\alpha} + \boldsymbol{w}.$$

In particular, the assumptions for the ENRA are critical. The number L of channel components is assumed to be known, and also perfect knowledge of the delay

---

[3] When carried out, for instance in Matlab, the Gaussian elimination algorithms still produce warning messages, reporting a badly scaled linear system.

parameters $\boldsymbol{\tau} = [\tau_1, \tau_2, \dots, \tau_L]^\top$ is assumed. Hence, only the comlex-valued channel amplitudes $\boldsymbol{\alpha}$ and the noise $\boldsymbol{w}$ are unknown terms. The channel amplitudes are assumed to comprise a zero-mean complex Gaussian vector with covariance matrix

$$\mathbf{P} := \mathbb{E}\big[\boldsymbol{\alpha}\boldsymbol{\alpha}^H\big] = \operatorname{diag}\big(\sigma_1^2, \sigma_2^2, \dots, \sigma_L^2\big).$$

The individual power terms σ_l^2 are considered unknown, and notice that the uncorrelated scattering (US) assumption is implicitly invoked. The assumptions for the noise term reads

$$\boldsymbol{w} \sim \mathcal{CN}\big(\mathbf{0}, \sigma_w^2 \mathbf{I}_M\big).$$

The channel amplitude vector $\boldsymbol{\alpha}$ and the noise vector $\boldsymbol{w}$ are assumed to be independent. By taking all our assumptions into account we easily find

$$\mathbf{R} := \mathbb{E}\Big[\boldsymbol{h}_{\text{zf}}\boldsymbol{h}_{\text{zf}}^H\Big] = \mathbf{T}\mathbf{P}\mathbf{T}^H + \sigma_w^2 \mathbf{I}_M,$$

and

$$\mathbf{C} := \mathbb{E}\Big[\boldsymbol{\alpha}\boldsymbol{h}_{\text{zf}}^H\Big] = \mathbf{P}\mathbf{T}^H.$$

Applying the theory of *linear minimum mean squared error estimation*, we obtain the following estimator of the channel amplitude vector

$$\hat{\boldsymbol{\alpha}} := \mathbf{C}\mathbf{R}^{-1}\boldsymbol{h}_{\text{zf}} = \Big(\mathbf{T}^H\mathbf{T} + \sigma_w^2\mathbf{P}^{-1}\Big)^{-1}\mathbf{T}^H\boldsymbol{h}_{\text{zf}}.$$

From this estimate of the channel amplitude vector, the frequency response at all subcarrier positions employed in the system is estimated from

$$\hat{\boldsymbol{h}}_{\text{ENRA}} := \mathbf{T}_{\mathcal{I}}\hat{\boldsymbol{\alpha}} = \mathbf{T}_{\mathcal{I}}\Big(\mathbf{T}^H\mathbf{T} + \sigma_w^2\mathbf{P}^{-1}\Big)^{-1}\mathbf{T}^H\boldsymbol{h}_{\text{zf}}, \tag{5.11}$$

where we have introduced the $N_u \times L$ matrix $\mathbf{T}_{\mathcal{I}}$, with entries given by

$$\mathbf{T}_{\mathcal{I}}(n,l) = \exp\Big(-i2\pi\frac{\mathcal{I}(n)}{N}\tau_l\Big), \qquad \begin{array}{c} n = 1, 2, \dots, N_u \\ l = 1, 2, \dots, L. \end{array}$$

As mention above, the channel power terms in $\mathbf{P} = \operatorname{diag}\big(\sigma_1^2, \sigma_2^2, \dots, \sigma_L^2\big)$ are not known, and in general the noise variance σ_w^2 is unknown as well. The authors of **[12]** refer to the results obtained by **[11]** and **[18]**, suggesting that a uniformly distributed power-delay profile and a value of the noise variance corresponding to a high SNR level, should be substituted. These choices comprise a robust design and the estimation performance of (5.11) degrades only slightly with deviations from the true parameters.

5.5.3 Robust Wiener Filter

As mentioned in **[7, Section 6.3]**, the linear minimum mean squared error estimator given in terms of a 'classical' Wiener Filter (WF), would straightforwardly read

$$\hat{\boldsymbol{h}}_{\text{WF}} := \mathbb{E}\big[\boldsymbol{h}_{\mathcal{I}}\boldsymbol{h}_{\mathcal{P}}^H\big]\Big(\mathbb{E}\big[\boldsymbol{h}_{\mathcal{P}}\boldsymbol{h}_{\mathcal{P}}^H\big] + \sigma_w^2\mathbf{I}_M\Big)^{-1}\boldsymbol{h}_{\text{zf}}. \tag{5.12}$$

In (5.12), $\boldsymbol{h}_{\mathcal{I}}$ denotes the 'true' channel transfer function at the N_u active subcarrier positions, similar for $\boldsymbol{h}_{\mathcal{P}}$ but only at the M pilot symbol positions and $\sigma_w^2 \mathbf{I}_M$ reports the noise covariance matrix. Implicitly, the channel transfer function is assumed to comprise a zero-mean complex Gaussian process with certain correlation properties and the assumptions for the noise statistics are 'as usual'.

The two matrices in (5.12) which are expressed only in terms of their definitions, are typically not known in practice. Imposing a robust choice or design for these correlation matrices may therefore seem evident. To this end, we follow closely the lines of **[11]**. The sampled version of the channel transfer function arranged in the vector $\boldsymbol{h}_{\mathcal{I}}$ is modeled as earlier in (5.4). For the sake of clarity we recast the expression as

$$h_n = \sum_{l=1}^{L} \alpha_l \mathrm{e}^{-i2\pi \frac{\mathcal{I}(n)}{N} \tau_l}, \qquad n = 1, 2, \ldots, N_u.$$

In **[11]**, it is suggested to model the L delay parameters as

$$\tau_l \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0, \mu + 1), \qquad l = 1, 2, \ldots, L,$$

i.e. mutually independent and uniformly distributed along the duration of the cyclic prefix. It is furthermore suggested to let the total power in the channel be uniformly distributed among the L amplitude components in $\boldsymbol{\alpha}$. This choice was shown to be robust to mismatches in **[18]** (also used with the ENRA). Hence, considering a channel with total average power normalized to unity, the power assumptions for the complex-valued channel amplitudes are

$$\mathbb{E}\Big[\left|\alpha_l\right|^2\Big] = \frac{1}{L}, \qquad l = 1, 2, \ldots, L.$$

With these different design choices/assumptions, the entries of the two correlation matrices in (5.12) can be calculated. Specifically, using the uncorrelated scattering assumption, we calculate

$$\begin{aligned}
\mathbb{E}\big[h_n h_k^*\big] &= \frac{1}{L} \sum_{l=1}^{L} \mathbb{E}\Big[\exp\Big(-i2\pi \tfrac{\mathcal{I}(n) - \mathcal{I}(k)}{N} \tau_l \Big)\Big], \qquad 1 \le n, k \le N_n, \\
&= \frac{1}{\mu + 1} \int_0^{\mu+1} \exp\Big(-i2\pi \tfrac{\mathcal{I}(n) - \mathcal{I}(k)}{N} \tau \Big) \mathrm{d}\tau \\
&= \begin{cases} \dfrac{1 - \exp\Big(-i2\pi \frac{\mathcal{I}(n) - \mathcal{I}(k)}{N} (\mu+1) \Big)}{i2\pi \frac{\mathcal{I}(n) - \mathcal{I}(k)}{N} (\mu+1)} & , \quad n \neq k \\ 1 & , \quad n = k. \end{cases}
\end{aligned}$$

It is interesting to observe that the number L of channel components, does not appear in the above expression. Plugging this particular correlation structure into the classical Wiener Filter expression (5.12), we obtain the third and final channel estimator to be considered in this thesis. In **[7, Section 6.3.2]**, this type of estimator is referred to as the *Robust Wiener Filter* (RWF). A value of the noise variance σ_w^2 corresponding to a high SNR level may be substituted if not known a priori, just as was suggested for the ENRA in the previous section.

5.5.4 Performance Evaluation

In this section, we compare the performance of the three PACE algorithms just described. The comparison is not fair since the individual assumptions for the three algorithms differ. However, the objective is merely to reproduce the results reported in **[7, Section 6.4]**. The following results are therefore presented in terms of bit-error-rate (BER) performance as a function of signal-to-noise ratio (SNR). Throughout this thesis, the SNR will be expressed as signal-to-noise ratio per bit and always the modulation scheme to be used, is a unit symbol power QPSK constellation with Gray encoding (recall **Section 2.3**). Furthermore, the channel considered is a Rayleigh fading multipath channel with average power normalized to unity. Hence, both transmit symbol power and average channel power are normalized to unity. As explained in **[7, Section 6.4]**, the signal-to-noise ratio per bit is then given according to

$$\text{SNR (dB)} = 10 \log_{10} \left(\frac{1}{2\sigma_w^2} \right),$$

where σ_w^2 denotes the variance of the complex-valued, additive white Gaussian noise.

Recall the left-hand side of **Figure 5.3**, where the pilot symbol pattern for a single transmit antenna is shown. In the following we simulate transmission only of OFDM symbols carrying pilot symbols, i.e. OFDM symbols in the first and fifth time allocations of every slot. We draw channel realizations from a highly unrealistic multipath channel, more specifically, the realizations are uncorrelated from OFDM symbol to OFDM symbol. The channel chosen for this verification study constantly includes nine echoes with static delay parameters. Overall, the particular channel is not of primary concern at the moment, however in **Chapter 7** the situation is completely refashioned. The simulations are carried out directly in the frequency domain according to (5.5), i.e. we assume that all ISI have been consumed by the (discarded) samples in the cyclic prefix.

The performance of the three different PACE algorithms are reported in **Figure 5.4**. The ENRA algorithm is fed with exact knowledge of the multipath delay parameters and also the noise variance is known to it. The noise variance is likewise fed to the Robust Wiener Filter. Overall, all three algorithm perform under ideal conditions in terms of their individual assumptions. As the SNR increases, the performance of the sample-aligned DFT-based method drifts even more away from the theoretical BER-curve. The Robust Wiener Filter associates a constant and irreducible BER degradation in the entire SNR-range due to its robust design. The ENRA performs very close to the theoretical limit, but notice that this algorithm is also provided with perfect a priori channel information.

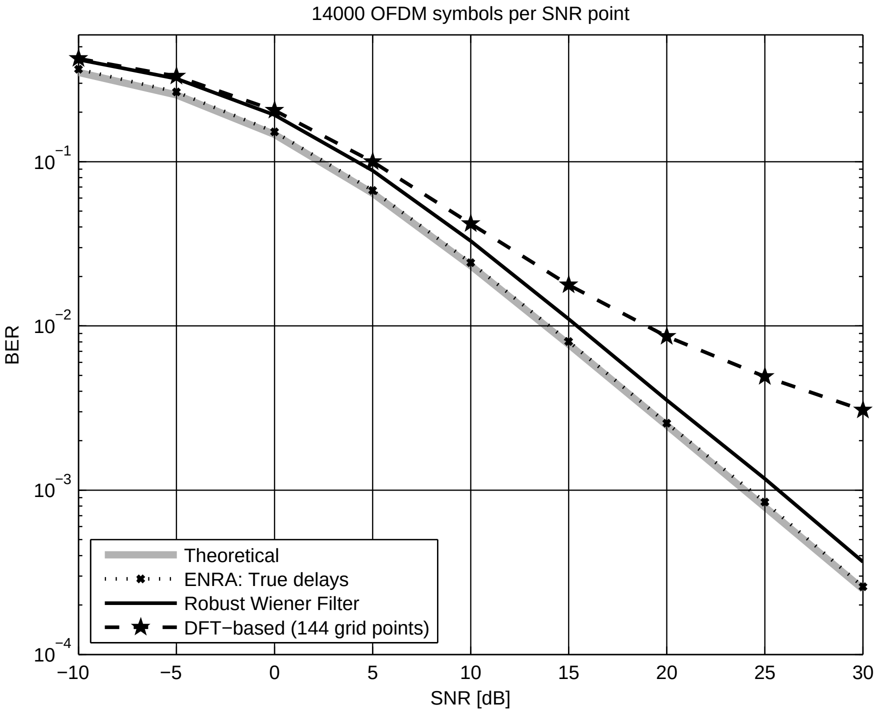


Figure 5.4: Performance of the ENRA when feeded with true delay parameters. Robust Wiener Filter performance and the sample-aligned DFT-based method are shown too. Theoretical BER-curve shown for comparison reasons. For each SNR point, the curves are generated from simulation of 14000 OFDM symbols - corresponding to one second of real-time transmission.

The multipath delay parameters being directly provided to the ENRA are not known in practice. Therefore, these have to be estimated if the ENRA comprise the channel estimation algorithm selected for use. The ENRA performance degrades notably with only slightly erroneous delay parameters provided **[7, Section 6.4.3]**. This fact is critical, in particular at high SNR-levels.

We now point our attention towards the main objective in this thesis - estimation of the multipath propagation delay parameters.

Part II

Multipath Propagation Delay Estimation

Chapter 6

Array Signal Processing Survey

Given the zero-forcing estimates $\boldsymbol{h}_{\text{zf}}$ we seek to estimate the delay vector $\boldsymbol{\tau}$. With the particular observation model (5.7) at hand, this estimation task is similar to that of *direction of arrival* estimation within the field of *array signal processing.* Accordingly, we set off this second part of the thesis by investigating a variety of estimation methods directly adopted from well-established literature on array signal processing.

6.1 Framework

A variety of techniques for direction of arrival estimation from array signal processing exist, including for instance *beamforming* and *subspace fitting* approaches. Another approach is *maximum likelihood* (ML) estimation and two such methods have been proposed in the literature **[24]**, stated as *deterministic* and *stochastic* ML estimation. Both ML methods are grounded on a parametric signal model identical to (5.7), however, the fundamental difference between the two methods rely on whether the underlying 'signal source' is considered deterministic or random. The subspace fitting methods are parametric approaches relying on the observation model (5.7) with the underlying signal source considered random. The structure of the theoretical covariance matrix associated with the observations is then utilized to separate the observation space into a so-called *signal subspace* and a *noise subspace.* The beamforming approaches are fundamentally simple, non-parametric methods assuming no particular model for the underlying signal. Spatially white noise is the only assumption for the different beamformers.

With the parametric approach we consider the observation model (5.7), which we recast as

$$\boldsymbol{h}_{\text{zf}}(k) = \mathbf{T}\big(\boldsymbol{\tau}(k)\big)\boldsymbol{\alpha}(k) + \boldsymbol{w}(k), \qquad k = 1, 2, \ldots, K.$$

The index k is used to emphasize the temporal sampling procedure performed at the pilot symbol positions (twice every slot, four times every sub-frame or forty times every radio frame). The receiver makes generic use of a memory of maximum K zero-forcing estimates denoted by $\boldsymbol{h}_{\text{zf}}(1), \boldsymbol{h}_{\text{zf}}(2), \ldots, \boldsymbol{h}_{\text{zf}}(K)$. In a generic fashion these K vector observations are appropriately collected in the $M \times K$ matrix

$$\mathbf{H}_{\text{zf}} := \begin{bmatrix} | & | & & | \\ \boldsymbol{h}_{\text{zf}}(1) & \boldsymbol{h}_{\text{zf}}(2) & \cdots & \boldsymbol{h}_{\text{zf}}(K) \\ | & | & & | \end{bmatrix}.$$

Throughout the following sections we impose a crucial assumption which does *not* allow the number $L(t)$ of attending channel echoes to vary at all. That is, we assume $L(t) \equiv L$ stay fixed during the entire time span under consideration. The reason for this is easily seen when the different methods are presented (simply, the dimension of the estimation problem must stay fixed). However, the individual delays may of course vary over time, i.e. in-between the times of observation of the data in $\mathbf{H}_{\text{zf}}$. This fact turns out to be very critical, since some methods suffers greatly when the parameters we aim to estimate may happen to slightly fluctuate during the estimation procedure.

As mentioned above, two overall cases (deterministic/stochastic) are considered and the assumption for these are as follows:

- *Deterministic case:* the underlying 'signal source', i.e. the channel amplitude vector $\boldsymbol{\alpha}(k)$, is considered unknown but fixed/deterministic.
- *Stochastic case:* the channel amplitude vector $\boldsymbol{\alpha}(k)$ is considered random.

With the deterministic approach, the statistics of the vector observations read

$$\boldsymbol{h}_{\text{zf}}(k) \sim \mathcal{CN}\Big(\mathbf{T}\big(\boldsymbol{\tau}(k)\big)\boldsymbol{\alpha}(k), \sigma^2\mathbf{I}_M\Big), \qquad k = 1, 2, \ldots, K,$$

i.e. the only random contribution encompasses the complex Gaussian noise, assumed to be spatially and temporally uncorrelated.

With the stochastic approach the source signals $\boldsymbol{\alpha}(k)$, $k = 1, 2, \ldots, K$, are considered random too - more specifically being complex Gaussian distributed with zero mean and with unknown but time-invariant covariance structure

$$\mathbb{E}\Big[\boldsymbol{\alpha}(k)\boldsymbol{\alpha}^H(n)\Big] = \mathbb{1}[k = n]\mathbf{P}, \qquad 1 \leq k, n \leq K.$$

From the uncorrelated scattering (US) assumption we see that $\mathbf{P}$ is a full rank diagonal matrix. Notice carefully how it is assumed that the temporal lags between consecutive sampling points are of sufficient extent such that all channel components have decorrelated during the time in-between. Hence, with the above assumptions on $\boldsymbol{\alpha}(k)$ the statistical model in the stochastic case reads

$$\boldsymbol{h}_{\text{zf}}(k) \sim \mathcal{CN}\Big(\mathbf{0},\, \mathbf{T}\big(\boldsymbol{\tau}(k)\big)\mathbf{P}\mathbf{T}^H\big(\boldsymbol{\tau}(k)\big) + \sigma^2\mathbf{I}_M\Big), \qquad k = 1, 2, \ldots, K.$$

With both approaches the snapshots observed over time $\boldsymbol{h}_{\text{zf}}(1), \boldsymbol{h}_{\text{zf}}(2), \ldots, \boldsymbol{h}_{\text{zf}}(K)$ are mutually independent, which is seen as a consequence of the random terms involved being assumed Gaussian distributed and temporally uncorrelated.

6.1.1 Analogy to Direction of Arrival Estimation

Recasting the above introduced delay estimation framework into its ‘usual’ array signal processing setup we explicitly come to realize the striking similarities as

$$\underset{\underset{\boldsymbol{x}(k)}{\uparrow}}{\boldsymbol{h}_{\text{zf}}(k)} = \underset{\underset{\mathbf{A}(\boldsymbol{\theta})}{\uparrow}}{\mathbf{T}\big(\boldsymbol{\tau}(k)\big)}\,\underset{\underset{\boldsymbol{s}(k)}{\uparrow}}{\boldsymbol{\alpha}(k)} + \underset{\underset{\boldsymbol{n}(k)}{\uparrow}}{\boldsymbol{w}(k)}\,, \qquad k = 1, 2, \dots, K,$$

where the notation used for the lower mentioned terms are adopted directly from **[24]**. In this notation, $\mathbf{A}(\boldsymbol{\theta})$ is a $M \times L$ matrix depending on an unknown parameter $\boldsymbol{\theta} = \big[\theta_1, \theta_2, \dots, \theta_L\big]^\top$. The columns of $\mathbf{A}(\boldsymbol{\theta})$ are so-called *steering vectors* and structured such that

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} | & | & & | \\ \boldsymbol{a}(\theta_1) & \boldsymbol{a}(\theta_2) & \cdots & \boldsymbol{a}(\theta_L) \\ | & | & & | \end{bmatrix},$$

where it is assumed that whenever the L components of $\boldsymbol{\theta}$ are mutually different, the columns of $\mathbf{A}(\boldsymbol{\theta})$ are linearly independent (ambiguity condition). The unknown parameter $\boldsymbol{\theta}$ contains the L directions of arrival to be estimated from the snapshot vectors $\boldsymbol{x}(k)$, observed at times $k = 1, 2, \dots, K$ by the M sensors comprising the array. Hence, each individual component of $\boldsymbol{x}(k)$ corresponds to an individual sensor in the array. Typically, the M sensors are identical and arranged uniformly on a common straight line with fixed inter-sensor distance $d < \lambda/2$, where λ is the wavelength of the incoming signal. Such a construction is called an *uniform linear array*, see **Figure 6.1**.

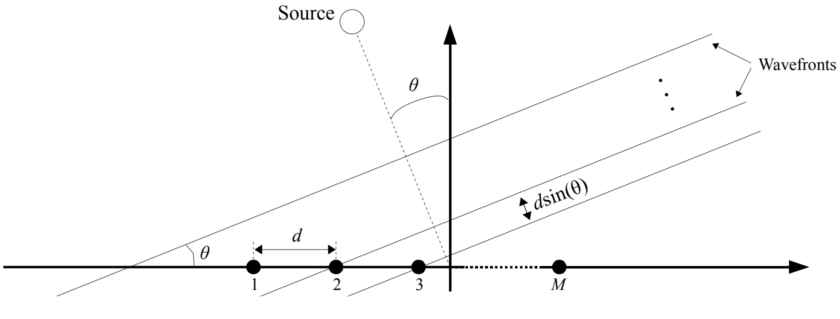


Figure 6.1: Uniform linear array of M sensors.

The steering vector associated with an uniform linear array construction is given by

$$\boldsymbol{a}(\theta) = \Big[\, 1 \quad \mathrm{e}^{-i2\pi d \sin(\theta)/\lambda} \quad \dots \quad \mathrm{e}^{-i2\pi(M-1)d\sin(\theta)/\lambda} \,\Big]^\top, \qquad \theta \in \big[-\tfrac{\pi}{2}, \tfrac{\pi}{2}\big).$$

The constraints $d < \lambda/2$ and $-\pi/2 \leq \theta < \pi/2$ are ambiguity conditions and comprise the limitations of the uniform linear array (interpreted as a spatial variant of the *Nyquist criterion*).

In (5.6) the entries of $\mathbf{T}\big(\boldsymbol{\tau}(k)\big)$ are defined and from these we immediately find the 'steering vector' associated within our delay estimation framework to be

$$\boldsymbol{t}(\tau) := \begin{bmatrix} \mathrm{e}^{-i2\pi p(1)\tau/N} & \mathrm{e}^{-i2\pi p(2)\tau/N} & \dots & \mathrm{e}^{-i2\pi p(M)\tau/N} \end{bmatrix}^{\top}, \qquad \tau \geq 0.$$

From this definition we may appropriately write

$$\mathbf{T}\big(\boldsymbol{\tau}(k)\big) = \begin{bmatrix} | & | & & | \\ \boldsymbol{t}\big(\tau_1(k)\big) & \boldsymbol{t}\big(\tau_2(k)\big) & \cdots & \boldsymbol{t}\big(\tau_L(k)\big) \\ | & | & & | \end{bmatrix}.$$

Later on, we shall see how we may divide the matrix $\mathbf{T}\big(\boldsymbol{\tau}(k)\big)$ into smaller submatrices and relate these by rotational transformations. Lowering the matrix dimensions gives rise to reduced amounts of computations, while consequently we sacrifice on the level of estimation accuracy.

6.2 Maximum Likelihood Approaches

In the following we describe two ML estimation approaches and for notational convenience we omit the delay dependency with $\mathbf{T}\big(\boldsymbol{\tau}(k)\big)$ and write $\mathbf{T}$, only. However, remember that the delays may fluctuate over time.

6.2.1 Deterministic ML Estimation

The joint pdf of the complex M-variate snapshots $\boldsymbol{h}_{\mathrm{zf}}(1), \boldsymbol{h}_{\mathrm{zf}}(2), \dots, \boldsymbol{h}_{\mathrm{zf}}(K)$ is given by

$$p\big(\boldsymbol{h}_{\mathrm{zf}}(1), \dots, \boldsymbol{h}_{\mathrm{zf}}(K)\big) = \prod_{k=1}^{K} \left(\pi\sigma^2\right)^{-M} \exp\left(-\frac{1}{\sigma^2}\left\|\boldsymbol{h}_{\mathrm{zf}}(k) - \mathbf{T}\boldsymbol{\alpha}(k)\right\|^2\right).$$

The negative log-likelihood function, normalized by K, is therefore

$$-\frac{1}{K}\ell\big(\boldsymbol{\tau}, \sigma^2, \boldsymbol{\alpha}(1), \dots, \boldsymbol{\alpha}(K)\big) = C + M \log \sigma^2 + \frac{1}{\sigma^2 K}\sum_{k=1}^{K}\left\|\boldsymbol{h}_{\mathrm{zf}}(k) - \mathbf{T}\boldsymbol{\alpha}(k)\right\|^2, \tag{6.1}$$

where C is a constant not depending on the unknown parameters to estimate. This negative log-likelihood function is to be minimized with respect to the unknown parameters, in particular we seek the ML estimate of $\boldsymbol{\tau}$. The joint estimation task of all unknown parameters allows for some degree of separability, i.e. for $\boldsymbol{\tau}$ and σ^2 fixed, the minima of the non-negative quantities $\left\|\boldsymbol{h}_{\mathrm{zf}}(k) - \mathbf{T}\boldsymbol{\alpha}(k)\right\|^2$ are attained at

$$\hat{\boldsymbol{\alpha}}(k) := \arg\min_{\boldsymbol{\alpha}(k)} \left\|\boldsymbol{h}_{\mathrm{zf}}(k) - \mathbf{T}\boldsymbol{\alpha}(k)\right\|^2 = \left(\mathbf{T}^H\mathbf{T}\right)^{-1}\mathbf{T}^H\boldsymbol{h}_{\mathrm{zf}}(k), \tag{6.2}$$

i.e. the unique linear combinations yielding the orthogonal projections of the snapshots $\boldsymbol{h}_{\text{zf}}(k)$ onto the L-dimensional subspace spanned by the linearly independent columns of $\mathbf{T}$. To ease notation we define the orthogonal projection operators

$$\boldsymbol{\Pi}_{\mathbf{T}} := \mathbf{T}\big(\mathbf{T}^H\mathbf{T}\big)^{-1}\mathbf{T}^H \qquad \text{and} \qquad \boldsymbol{\Pi}_{\mathbf{T}}^{\perp} := \mathbf{I}_M - \boldsymbol{\Pi}_{\mathbf{T}},$$

where $\boldsymbol{\Pi}_{\mathbf{T}}^{\perp}$ is the projection mapping onto the null space of $\mathbf{T}^H$. Notice that any orthogonal projection operator is *idempotent* and *self-adjoint*, e.g. $\boldsymbol{\Pi}_{\mathbf{T}}\boldsymbol{\Pi}_{\mathbf{T}} = \boldsymbol{\Pi}_{\mathbf{T}}$ and $\boldsymbol{\Pi}_{\mathbf{T}}^H = \boldsymbol{\Pi}_{\mathbf{T}}$ which can be readily verified. Furthermore, the trace of an orthogonal projection operator equals the dimension of the subspace it projects onto, and its eigenvalues are members of the set $\{0, 1\}$, only.

With $\boldsymbol{\tau}$ and $\hat{\boldsymbol{\alpha}}(1), \hat{\boldsymbol{\alpha}}(2), \ldots, \hat{\boldsymbol{\alpha}}(K)$ fixed, the minimum of the negative log-likelihood function (6.1) with respect to the noise variance σ^2 is obtained from the derivative

$$-\frac{1}{K}\frac{\partial}{\partial\sigma^2}\ell\big(\boldsymbol{\tau}, \sigma^2, \hat{\boldsymbol{\alpha}}(1), \ldots, \hat{\boldsymbol{\alpha}}(K)\big) = \frac{M}{\sigma^2} - \frac{1}{\sigma^4 K}\sum_{k=1}^{K}\big\|\boldsymbol{\Pi}_{\mathbf{T}}^{\perp}\boldsymbol{h}_{\text{zf}}(k)\big\|^2. \tag{6.3}$$

Equating the right-hand side of (6.3) to zero and solving for σ^2 yields

$$\hat{\sigma^2} = \frac{1}{MK}\sum_{k=1}^{K}\big\|\boldsymbol{\Pi}_{\mathbf{T}}^{\perp}\boldsymbol{h}_{\text{zf}}(k)\big\|^2 = \frac{1}{M}\text{tr}\big(\boldsymbol{\Pi}_{\mathbf{T}}^{\perp}\mathbf{S}\big), \tag{6.4}$$

where we have introduced the *sample covariance matrix*

$$\mathbf{S} := \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{h}_{\text{zf}}(k)\boldsymbol{h}_{\text{zf}}^H(k) = \frac{1}{K}\mathbf{H}_{\text{zf}}\mathbf{H}_{\text{zf}}^H, \tag{6.5}$$

and used that $\text{tr}(z) = z$ for any scalar $z \in \mathbb{C}$, and also that $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ provided conformable matrix dimensions. Substituting (6.2) and (6.4) back into the negative log-likelihood (6.1) yields the ML estimate

$$\hat{\boldsymbol{\tau}} = \arg\min_{\boldsymbol{\tau}} \text{tr}\big(\boldsymbol{\Pi}_{\mathbf{T}}^{\perp}\mathbf{S}\big). \tag{6.6}$$

In (6.6), the scalar expression to be minimized is a non-linear function of its L-dimensional argument $\boldsymbol{\tau}$. Analytical optimization is unfeasible but a variety of numerical search procedures have been suggested and examined in the literature **[24]**.

6.2.2 Stochastic ML Estimation

With the stochastic approach the likelihood function now depends on the parameters $\boldsymbol{\tau}$, $\mathbf{P}$ and σ^2. We denote by $\mathbf{R} := \mathbf{TPT}^H + \sigma^2\mathbf{I}_M$ the true covariance matrix and still $\mathbf{S}$ denotes the sample covariance matrix. The negative log-likelihood function, normalized by K, now takes the form

$$-\frac{1}{K}\ell\big(\boldsymbol{\tau}, \mathbf{P}, \sigma^2\big) = C + \log\det(\mathbf{R}) + \text{tr}\big(\mathbf{R}^{-1}\mathbf{S}\big). \tag{6.7}$$

For $\boldsymbol{\tau}$ and σ^2 fixed, minimization of (6.7) with respect to $\mathbf{P}$ calls for the matrix equation

$$\mathbf{T}^H(\mathbf{S}-\mathbf{R})\mathbf{T} = \big[\,\mathbf{0}\,\big]_{L\times L}$$

to hold, see **[25]** for details. Plugging in $\mathbf{R} = \mathbf{T}\mathbf{P}\mathbf{T}^H + \sigma^2\mathbf{I}_M$ and solving for $\mathbf{P}$ yields

$$\hat{\mathbf{P}} = \big(\mathbf{T}^H\mathbf{T}\big)^{-1}\mathbf{T}^H\big(\mathbf{S}-\sigma^2\mathbf{I}_M\big)\mathbf{T}\big(\mathbf{T}^H\mathbf{T}\big)^{-1}. \tag{6.8}$$

Immediately, from the estimate (6.8) of the $L \times L$ source covariance matrix we let

$$\begin{aligned}\hat{\mathbf{R}} &:= \mathbf{T}\hat{\mathbf{P}}\mathbf{T}^H + \sigma^2\mathbf{I}_M & (6.9)\\ &= \mathbf{\Pi}_{\mathbf{T}}\mathbf{S}\mathbf{\Pi}_{\mathbf{T}} + \sigma^2\mathbf{\Pi}_{\mathbf{T}}^{\perp}. & (6.10)\end{aligned}$$

A common matrix inversion formula applied to (6.9) yields

$$\hat{\mathbf{R}}^{-1} = \frac{1}{\sigma^2}\mathbf{I}_M - \frac{1}{\sigma^2}\mathbf{T}\big(\hat{\mathbf{P}}\mathbf{T}^H\mathbf{T} + \sigma^2\mathbf{I}_L\big)^{-1}\hat{\mathbf{P}}\mathbf{T}^H,$$

and using this expression[1] and a few matrix manipulations, one can show (yet, see **[25]** for details) that

$$\operatorname{tr}\big(\hat{\mathbf{R}}^{-1}\mathbf{S}\big) = \frac{1}{\sigma^2}\operatorname{tr}\big(\mathbf{\Pi}_{\mathbf{T}}^{\perp}\mathbf{S}\big) + L.$$

With $\boldsymbol{\tau}$ fixed and with $\hat{\mathbf{R}}$ as given by (6.10), we calculate the derivative

$$\begin{aligned}-\frac{1}{K}\frac{\partial}{\partial\sigma^2}\ell\big(\boldsymbol{\tau},\hat{\mathbf{P}},\sigma^2\big) &= \frac{\partial}{\partial\sigma^2}\Big(C + \log\det(\hat{\mathbf{R}}) + \frac{1}{\sigma^2}\operatorname{tr}\big(\mathbf{\Pi}_{\mathbf{T}}^{\perp}\mathbf{S}\big) + L\Big)\\ &= \operatorname{tr}\Big(\hat{\mathbf{R}}^{-1}\mathbf{\Pi}_{\mathbf{T}}^{\perp}\Big) - \frac{1}{\sigma^4}\operatorname{tr}\big(\mathbf{\Pi}_{\mathbf{T}}^{\perp}\mathbf{S}\big)\\ &= \frac{1}{\sigma^2}\operatorname{tr}\big(\mathbf{\Pi}_{\mathbf{T}}^{\perp}\big) - \frac{1}{\sigma^4}\operatorname{tr}\big(\mathbf{\Pi}_{\mathbf{T}}^{\perp}\mathbf{S}\big).\end{aligned}$$

Equating to zero and solving for σ^2 implies

$$\hat{\sigma^2} = \frac{1}{M-L}\operatorname{tr}\big(\mathbf{\Pi}_{\mathbf{T}}^{\perp}\mathbf{S}\big). \tag{6.11}$$

Substitution of (6.10) and (6.11) back into the negative and normalized log-likelihood function (6.7) implies

$$\hat{\boldsymbol{\tau}} = \arg\min_{\boldsymbol{\tau}}\log\det\big(\mathbf{\Pi}_{\mathbf{T}}\mathbf{S}\mathbf{\Pi}_{\mathbf{T}} + \hat{\sigma^2}\mathbf{\Pi}_{\mathbf{T}}^{\perp}\big). \tag{6.12}$$

As with the deterministic ML case, the functional relationship in (6.12) is ‘highly’ non-linear in the L-dimensional argument $\boldsymbol{\tau}$. To summarize this section we have

---

[1] The matrix $\hat{\mathbf{P}}\mathbf{T}^H\mathbf{T} + \sigma^2\mathbf{I}_L = \big(\mathbf{T}^H\mathbf{T}\big)^{-1}\mathbf{T}^H\mathbf{S}\mathbf{T}$ is non-singular whenever $\mathbf{S}$ is of full rank.

now derived the two useful expressions (6.6) and (6.12), which under their individual assumptions are optimal in ML sense.

As mentioned in the beginning of this section, the whole theory and the two ML estimation methods just described originate from the field of array signal processing. Nonetheless, the striking similarities with our multipath delay estimation task allows for straightforward handing-over of the principles. We now describe a method belonging to the category of subspace fitting approaches.

6.3 ESPRIT

The method presented in this section is commonly referred to as the ESPRIT method, which is an abbreviation for *estimation of signal parameters by rotational invariance techniques* **[14, Section 4.7]**. In **[12]**, the ESPRIT method is proposed to serve as an acquisition tool when estimating multipath delay parameters for OFDM systems. Initially, some matrix manipulations are required and we simply reuse our previously introduced notation for the true covariance matrix and write

$$\mathbf{R} = \mathbf{T}\mathbf{P}\mathbf{T}^H + \sigma^2\mathbf{I}_M. \tag{6.13}$$

In this section, we assume that $\mathbf{P}$ has always full rank equal to L. Notice first, that any vector $\boldsymbol{v} \in \mathrm{Null}\big(\mathbf{T}^H\big)$ is an eigenvector of $\mathbf{R}$ with associated eigenvalue σ^2. The dimension of the eigenspace associated with the eigenvalue σ^2 is $M - L$, i.e. the null space of $\mathbf{T}^H$ is spanned by $M - L$ linearly independent vectors. Secondly, any eigenvector $\boldsymbol{u} \notin \mathrm{Null}\big(\mathbf{T}^H\big)$ will hold an associated eigenvalue strictly greater than σ^2. This eigenspace partitioning comprise the very core of the following derivation and since $\mathbf{R}$ is Hermitian and positive definite we can write

$$\mathbf{R} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H = \mathbf{U}_{\mathrm{s}}\boldsymbol{\Lambda}_{\mathrm{s}}\mathbf{U}_{\mathrm{s}}^H + \sigma^2\mathbf{U}_{\mathrm{n}}\mathbf{U}_{\mathrm{n}}^H,$$

where the M overall eigenvectors have been partitioned in two sets according to the eigenvalue discussion above. That is, $\mathbf{U}_{\mathrm{s}}$ is an $M \times L$ matrix containing the eigenvectors associated with the L eigenvalues strictly greater than σ^2 and $\mathbf{U}_{\mathrm{n}}$ is an $M \times (M - L)$ matrix of the remaining eigenvectors. The subscripts 's' and 'n' serve as abbreviation for *signal* and *noise*, respectively. Since the eigenvector basis is orthonormal it follows that $\mathbf{U}_{\mathrm{s}}\mathbf{U}_{\mathrm{s}}^H + \mathbf{U}_{\mathrm{n}}\mathbf{U}_{\mathrm{n}}^H = \mathbf{I}_M$, which can be used to manipulate the relationship

$$\mathbf{T}\mathbf{P}\mathbf{T}^H + \sigma^2\Big(\mathbf{U}_{\mathrm{s}}\mathbf{U}_{\mathrm{s}}^H + \mathbf{U}_{\mathrm{n}}\mathbf{U}_{\mathrm{n}}^H\Big) = \mathbf{U}_{\mathrm{s}}\boldsymbol{\Lambda}_{\mathrm{s}}\mathbf{U}_{\mathrm{s}}^H + \sigma^2\mathbf{U}_{\mathrm{n}}\mathbf{U}_{\mathrm{n}}^H.$$

From a few calculations and rearrangements we obtain

$$\mathbf{U}_{\mathrm{s}} = \mathbf{T}\underbrace{\mathbf{P}\mathbf{T}^H\mathbf{U}_{\mathrm{s}}\big(\boldsymbol{\Lambda}_{\mathrm{s}} - \sigma^2\mathbf{I}_L\big)^{-1}}_{\mathbf{Q}} = \mathbf{T}\mathbf{Q}, \tag{6.14}$$

where the diagonal matrix $\boldsymbol{\Lambda}_{\mathrm{s}} - \sigma^2\mathbf{I}_L$ is non-singular indeed, and the $L \times L$ matrix $\mathbf{Q}$ is therefore well-defined and of full rank.

The ESPRIT method is now based on a paramount structural requirement of the sensor array (the set $\mathcal{P}$ of pilot symbol subcarriers in our framework) - it must contain two identical sub-arrays that are shifted by a known displacement factor. To illustrate this requirement, we define two sub-matrices $\mathbf{T}_1(\boldsymbol{\tau})$ and $\mathbf{T}_2(\boldsymbol{\tau})$, both of dimension $M/2 \times L$, according to

$$\mathbf{T}(\boldsymbol{\tau}) = \begin{bmatrix} | & & | \\ \boldsymbol{t}(\tau_1) & \cdots & \boldsymbol{t}(\tau_L) \\ | & & | \end{bmatrix} = \begin{bmatrix} \mathbf{T}_1(\boldsymbol{\tau}) \\ --- \\ \mathbf{T}_2(\boldsymbol{\tau}) \end{bmatrix}. \tag{6.15}$$

When the shift or displacement between the sub-arrays $\mathbf{T}_1(\boldsymbol{\tau})$ and $\mathbf{T}_2(\boldsymbol{\tau})$ is known, e.g. by construction, we can relate these according to the rotational transformation

$$\mathbf{T}_2(\boldsymbol{\tau}) = \mathbf{T}_1(\boldsymbol{\tau})\mathbf{D}. \tag{6.16}$$

In our framework one can readily verify that the unitary matrix $\mathbf{D}$, is given by

$$\mathbf{D} = \operatorname{diag}\Big(\mathrm{e}^{-i2\pi\frac{(6\cdot M/2+1)}{N}\tau_1}, \ldots, \mathrm{e}^{-i2\pi\frac{(6\cdot M/2+1)}{N}\tau_L}\Big),$$

i.e., a displacement of $6 \cdot M/2 + 1 = 601$ subcarrier positions is needed in order to shift/rotate the sub-array $\mathbf{T}_1(\boldsymbol{\tau})$ into the sub-array comprised by the matrix $\mathbf{T}_2(\boldsymbol{\tau})$. Notice how the matrix $\mathbf{D}$ actually depends on time, since indeed, if the delay parameters change over time the rotational transform changes too.

Now, estimation of the delay parameters $\tau_1, \tau_2, \ldots, \tau_L$ can be done from knowing the eigenvalues of the matrix $\mathbf{D}$. Denote by $\mathbf{U}_{s,1}$ and $\mathbf{U}_{s,2}$ the associated sub-matrices obtained from (6.14) when invoking the sub-array relationship in (6.15). By definition we have

$$\mathbf{U}_{s,1} = \mathbf{T}_1\mathbf{Q} \qquad \text{implying} \qquad \mathbf{T}_1 = \mathbf{U}_{s,1}\mathbf{Q}^{-1},$$

and using the relationship in (6.16) we also find

$$\mathbf{U}_{s,2} = \mathbf{T}_2\mathbf{Q} = \mathbf{T}_1\mathbf{D}\mathbf{Q} = \mathbf{U}_{s,1}\mathbf{Q}^{-1}\mathbf{D}\mathbf{Q}.$$

Defining the matrix $\boldsymbol{\Phi} := \mathbf{Q}^{-1}\mathbf{D}\mathbf{Q}$, we immediately see that the eigenvalues of $\boldsymbol{\Phi}$ and $\mathbf{D}$ coincide. Implicitly, the matrix $\boldsymbol{\Phi}$ is obtained as the solution to the matrix equation

$$\mathbf{U}_{s,1}\boldsymbol{\Phi} = \mathbf{U}_{s,2}.$$

Of course, the true covariance matrix $\mathbf{R}$ is not available but rather an estimate of it, e.g. the sample covariance matrix $\mathbf{S}$. Hence, only estimates of the eigenvectors can be made available and straightforwardly, we denote the relevant estimates by $\hat{\mathbf{U}}_{s,1}$ and $\hat{\mathbf{U}}_{s,2}$.

To summarize, the ESPRIT algorithm takes as input an estimate $\hat{\mathbf{R}}$ of the true covariance matrix and the number L of signal dimensions to collect. An eigenvalue decomposition of $\hat{\mathbf{R}}$ is performed and the eigenvectors associated with the L largest eigenvalues are then used to build the two sub-matrices $\hat{\mathbf{U}}_{s,1}$ and $\hat{\mathbf{U}}_{s,2}$. The matrix equation

$$\hat{\mathbf{U}}_{s,1}\mathbf{\Phi} = \hat{\mathbf{U}}_{s,2},$$

is solved yielding a solution which we denote by $\hat{\mathbf{\Phi}}$. Finally, the eigenvalues of $\hat{\mathbf{\Phi}}$ are calculated and we denote these by $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_L$. The known sub-array displacement factor is utilized to recover/obtain estimates of the delay parameters according to

$$\hat{\tau}_l = \frac{\arg\big(\hat{\lambda}_l^*\big)N}{2\pi(6 \cdot M/2 + 1)}, \qquad l = 1, 2, \ldots, L,$$

where $\arg(\cdot) \in [0, 2\pi)$ returns the phase angle of its complex argument.

6.3.1 Assumptions and Limitations of ESPRIT

We consider it crucial to be fully aware of the limitations of the ESPRIT method and therefore we emphasize the assumptions in more compact form. The ESPRIT algorithm was derived using the fact that

- the theoretical covariance matrix has the particular form (6.13).
- $\mathbf{P}$ must be non-singular and the noise must be white and of identical power among all sensors.
- the number L of signal sources has to be known in order to properly collect the eigenvectors needed.
- the array considered must contain two identical sub-arrays and their displacement factor is required known too.

When ESPRIT is used for estimation of delay parameters in our framework, yet another requirement is present - and this requirement has been deliberately violated during the above derivation (in order to specifically emphasize the limitation). Recall the ambiguity conditions for the uniform linear array, stated slightly below **Figure 6.1**. Then, consider again the particular eigenvalues of the rotation matrix $\mathbf{D}$. Emphasizing their dependency on time, these read

$$\exp\Big(-i2\pi\tfrac{601}{N}\tau_l(t)\Big), \qquad l = 1, 2, \ldots, L,$$

and since the complex exponential is periodic in nature, we cannot distinguish whether the phase have already rotated 2π or not. That is, whenever a true delay parameter appears such that $601\tau_l(t) > N$, we face an ambiguity. The point is that ambiguity constraints are present on the displacement factor as well. Simply, the product between the displacement factor and the largest possible delay parameter must be less than N. Hence, the particular way we split up the matrix $\mathbf{T}(\boldsymbol{\tau})$ in (6.15) cannot be used directly. However, due to the property of wide-sense-stationarity (WSS) in the frequency-domain (inherited from the uncorrelated scattering (US) assumption in the delay-domain), the observations associated with the two sub-matrices $\mathbf{T}_1(\boldsymbol{\tau})$ and $\mathbf{T}_2(\boldsymbol{\tau})$ can be averaged. That is, instead of building the usual $M \times M$ sample covariance matrix $\mathbf{S}$ in (6.5), we can spilt up the zero-forcing estimates $\boldsymbol{h}_{\text{zf}}(k)$ in halves (obtaining twice as many observations) and build a $M/2 \times M/2$ sample covariance matrix instead. This matrix of lowered dimension inherits the exact same properties as if only $M/2$ pilot symbol positions had been employed in the system (still with an inter-pilot spacing of six subcarriers).

Essentially, this means that only the sub-matrix $\mathbf{T}_1(\boldsymbol{\tau})$ is involved now. However, this matrix holds identical sub-arrays too, e.g. the first $M/2-1$ rows can be displaced, *by one single row position*, directly into the last $M/2-1$ rows of the matrix (displacement factor of six subcarrier positions). This is the common formulation of the ESPRIT algorithm when encountered in literature **[14, 24]**. At this point, recall the aberrant pilot symbol subcarrier shift of seven positions induced by the zero-subcarrier being preserved inactive in LTE (**Section 5.2**). One may notice that this shift actually forces us to spilt up the zero-forcing estimates $\boldsymbol{h}_{\text{zf}}(k)$ in halves. We simply cannot displace any sub-array across this point.

In literature, a variety of other subspace fitting methods can be encountered, including for instance the *MUSIC* algorithm, again see **[14, 24]**. For numerous reasons, we focus only at the ESPRIT algorithm. A majority of the subspace fitting methods are all eigenvalue decomposition based, and hence, they basically suffer in equally manners given the circumstances in a particular scenario.

6.4 Beamforming

Historically, beamforming techniques comprised the first attempts of automatic signal source localization by use of antenna/sensor arrays **[24]**. With M sensors deployed, the procedure is to ‘steer’ the array into one direction at a time while measuring the output power. The L locations returning the L largest power levels are selected to yield the direction of arrival estimates. The array responses are calculated from linear combinations of the M sensor outputs and within our delay estimation framework this procedure reads

$$\boldsymbol{c}^H \boldsymbol{h}_{\text{zf}}(k), \qquad k = 1, 2, \ldots, K,$$

where $\boldsymbol{c} \in \mathbb{C}^M$ is a vector of coefficients to be selected. The empirical average output power is then calculated as

$$\frac{1}{K} \sum_{k=1}^{K} \left| \boldsymbol{c}^H \boldsymbol{h}_{\text{zf}}(k) \right|^2 = \boldsymbol{c}^H \mathbf{S} \boldsymbol{c} = \frac{1}{K} \left\| \boldsymbol{c}^H \mathbf{H}_{\text{zf}} \right\|^2. \tag{6.17}$$

Different beamforming approaches correspond to different choices of the weighting vector $\boldsymbol{c}$. We describe two such choices in the following.

6.4.1 Conventional Beamformer

From the set of pilot symbol positions $\mathcal{P}$, i.e. by construction of the array, we know that the associated steering vector reads

$$\boldsymbol{t}(\tau) = \Big[\mathrm{e}^{-i2\pi p(1)\tau/N} \quad \mathrm{e}^{-i2\pi p(2)\tau/N} \quad \dots \quad \mathrm{e}^{-i2\pi p(M)\tau/N} \Big]^{\top}, \qquad \tau \geq 0.$$

Fixing a particular value of the delay variable τ, we desire to maximize the output power in terms of this single delay, only. If indeed a single-echo channel with complex-valued amplitude $\widetilde{\alpha}$ holds associated delay $\widetilde{\tau}$, the zero-forcing output at the M pilot symbol positions is

$$\boldsymbol{h}_{\mathrm{zf}} = \boldsymbol{t}(\widetilde{\tau})\widetilde{\alpha} + \boldsymbol{w},$$

where the assumptions for the noise term $\boldsymbol{w}$ are ‘as usual’, i.e. $\boldsymbol{w} \sim \mathcal{CN}\big(\mathbf{0}, \sigma^2 \mathbf{I}_M\big)$. Generally, the power maximization criterion is formulated as

$$\begin{aligned} \max_{\boldsymbol{c}:\|\boldsymbol{c}\|=1} \mathbb{E}\Big[\big|\boldsymbol{c}^H \boldsymbol{h}_{\mathrm{zf}}\big|^2\Big] &= \max_{\boldsymbol{c}:\|\boldsymbol{c}\|=1} \boldsymbol{c}^H \Big(\mathbb{E}\big[|\alpha|^2\big]\boldsymbol{t}(\tau)\boldsymbol{t}^H(\tau) + \mathbb{E}\big[\boldsymbol{w}\boldsymbol{w}^H\big]\Big)\boldsymbol{c} \\ &= \max_{\boldsymbol{c}:\|\boldsymbol{c}\|=1} \big|\boldsymbol{c}^H \boldsymbol{t}(\tau)\big|^2. \end{aligned}$$

The maximizing choice of $\boldsymbol{c}$ is therefore found to be

$$\boldsymbol{c}_{\mathrm{bf}} := \arg \max_{\boldsymbol{c}:\|\boldsymbol{c}\|=1} \big|\boldsymbol{c}^H \boldsymbol{t}(\tau)\big|^2 = \frac{\boldsymbol{t}(\tau)}{\|\boldsymbol{t}\|} = \frac{\boldsymbol{t}(\tau)}{\sqrt{M}}$$

Plugging the particular choice $\boldsymbol{c}_{\mathrm{bf}}$ into (6.17), we obtain the expression for the conventional beamformer

$$\mathrm{BF}(\tau) := \frac{\boldsymbol{t}^H(\tau)\mathbf{S}\boldsymbol{t}(\tau)}{M} = \frac{\big\|\boldsymbol{t}^H(\tau)\mathbf{H}_{\mathrm{zf}}\big\|^2}{MK}.$$

As can be seen above, the choice of weights only depends on the sensor array and not on the actual observations.

6.4.2 Capon Beamformer

Other choices for the weighting vector $\boldsymbol{c}$ can be obtained by enforcement of different optimization criterions. In this section we describe a choice resulting in a technique commonly referred to as the *Capon beamformer* **[14, 24]**. Recall the empirical average output power expression (6.17) and consider the minimization criterion

$$\min_{\boldsymbol{c}} \ \boldsymbol{c}^H \mathbf{S} \boldsymbol{c} \qquad \text{subject to} \qquad \boldsymbol{c}^H \boldsymbol{t}(\tau) = 1 \tag{6.18}$$

which has a rather intuitive interpretation. The minimization criterion (6.18) dictates that a fixed gain should be maintained for the particular delay τ, while simultaneously, the average output power should be minimized for all other delay values. The weighting vector $\boldsymbol{c}_{\mathrm{capon}}$ fulfilling the requirement (6.18) can be found using, for instance, the technique of *Lagrange multipliers* **[26, Theorem 11.63]**, **[27, Section 13.9]**. It is given by

$$\boldsymbol{c}_{\text{capon}} := \frac{\mathbf{S}^{-1}\boldsymbol{t}(\tau)}{\boldsymbol{t}^{H}(\tau)\mathbf{S}^{-1}\boldsymbol{t}(\tau)},$$

and when inserted into (6.17) the expression for the Capon beamformer reads

$$\text{CAPON}(\tau) := \frac{1}{\boldsymbol{t}^{H}(\tau)\mathbf{S}^{-1}\boldsymbol{t}(\tau)}.$$

The weights associated with the Capon beamformer depends on the sensor array but also on the actual observations (in contrast to the conventional beamformer). The additional gain obtained from this fact is present in terms of very 'peaky' behavior around the true delay parameters. Hence, the ability to separate the L largest power levels is notably improved when using CAPON(τ) compared to BF(τ). However, a major drawback of the Capon Beamformer is that it requires a matrix inverse and we have quietly assumed its existence. In **Appendix C.1** we discuss the criteria for the sample covariance matrix $\mathbf{S}$ being non-singular.

6.5 Sequential Beamforming Algorithm

The non-linear L-dimensional optimization procedures inherited from either of the two ML approaches are computationally exhaustive to carry out. Matrix inversions are required to build up the projection operators and the optimization steps are potentially subject to numerical instability issues unless appropriate constraints are assigned to the unknown parameter $\boldsymbol{\tau}$. More specifically, one needs to ensure that the absolute difference between any two individual components of $\boldsymbol{\tau}$ are not getting 'too small' during the optimization process (recall the ambiguity condition imposed on the steering vectors building up the matrix $\mathbf{T}$).

In the following we describe a 'deterministic ML and beamforming'-merged approach inspired by **[28, 29]**. Basically, a projection operator is recursively constructed one subspace dimension at a time from consecutive beamforming searches. No L-dimensional optimization procedure is needed since instead L consecutive one-dimensional searches are performed. Also, no matrix inversions are required. With this approach the computational complexity is notably reduced but consequently we cannot guarantee ML optimality. Recall the deterministic case ML optimization criterion (6.6) which we recast as

$$\hat{\boldsymbol{\tau}} = \arg\min_{\boldsymbol{\tau}} \operatorname{tr}\big(\boldsymbol{\Pi}_{\mathbf{T}}^{\perp}\mathbf{S}\big) = \arg\max_{\boldsymbol{\tau}} \operatorname{tr}\big(\boldsymbol{\Pi}_{\mathbf{T}}\mathbf{S}\big) = \arg\max_{\boldsymbol{\tau}} \frac{1}{K}\sum_{k=1}^{K} \big\|\boldsymbol{\Pi}_{\mathbf{T}}\boldsymbol{h}_{\text{zf}}(k)\big\|^{2}.$$

The latter expression holds a rather intuitive interpretation - namely that we seek the L-dimensional subspace (spanned by steering vectors) explaining the largest amount of average power contained in the signals $\boldsymbol{h}_{\text{zf}}(1), \boldsymbol{h}_{\text{zf}}(2), \ldots, \boldsymbol{h}_{\text{zf}}(K)$. Define the initial projection operator

$$\boldsymbol{\Pi}_{1}(\tau) := \frac{\boldsymbol{t}(\tau)\boldsymbol{t}^{H}(\tau)}{\|\boldsymbol{t}(\tau)\|^{2}}, \qquad 0 \leq \tau \leq \mu + 1,$$

and consider the search

$$\max_{\tau} \operatorname{tr}\{\mathbf{\Pi}_1(\tau)\mathbf{S}\} = \max_{\tau}\left\{\frac{\boldsymbol{t}^H(\tau)\mathbf{S}\boldsymbol{t}(\tau)}{M}\right\} = \max_{\tau}\left\{\frac{\left\|\boldsymbol{t}^H(\tau)\mathbf{H}_{\mathrm{zf}}\right\|^2}{KM}\right\}. \tag{6.19}$$

Hence, in (6.19) we seek a single delay τ which in terms of the subspace spanned by the steering vector $\boldsymbol{t}(\tau)$ explains the largest amount of average power in the signals at hand. The maximizing argument in (6.19) is non-trivial to determine analytically and therefore the delay may be approximated from direct numerical examination, e.g. a high-resolution grid search. From the output of the search we let

$$\hat{\tau}_1 := \arg\max_{\tau}\left\{\frac{\boldsymbol{t}^H(\tau)\mathbf{S}\boldsymbol{t}(\tau)}{M}\right\}.$$

This value stay fixed while the following steps are performed in a recursive manner. For notational convenience we let

$$\widetilde{\boldsymbol{t}}(\tau) := \boldsymbol{t}(\tau) - \mathbf{\Pi}_1(\hat{\tau}_1)\boldsymbol{t}(\tau),$$

i.e. $\widetilde{\boldsymbol{t}}(\tau)$ is the projection of the steering vector $\boldsymbol{t}(\tau)$, onto the orthogonal complement of the current and fixed projection operator $\mathbf{\Pi}_1(\hat{\tau}_1)$. Define a new projection operator by

$$\mathbf{\Pi}_2(\tau;\hat{\tau}_1) := \mathbf{\Pi}_1(\hat{\tau}_1) + \frac{\widetilde{\boldsymbol{t}}(\tau)\widetilde{\boldsymbol{t}}^H(\tau)}{\left\|\widetilde{\boldsymbol{t}}(\tau)\right\|^2}, \qquad 0 \le \tau \le \mu + 1,$$

and similarly, we search for yet another delay explaining as much additional average power as possible, i.e. we let

$$\hat{\tau}_2 := \arg\max_{\tau} \operatorname{tr}\{\mathbf{\Pi}_2(\tau;\hat{\tau}_1)\mathbf{S}\} = \arg\max_{\tau}\left\{\frac{\left\|\widetilde{\boldsymbol{t}}^H(\tau)\mathbf{H}_{\mathrm{zf}}\right\|^2}{K\left\|\widetilde{\boldsymbol{t}}(\tau)\right\|^2}\right\}.$$

This procedure is repeated in an obvious manner, until L delay estimates have been collected. As in **[29]**, we refer to the above method as the *sequential beamforming algorithm* (SBA).

6.5.1 Potential Improvements

The recursive structure of the SBA allows (heuristically, but potentially) for a variety of refinements. We list two ideas that immediately seems evident:

- Instead of running the recursive procedure for a fixed and predetermined number of iterations, it may be possible to define an 'on-the-fly' stopping criteria.
- Possibly more than one delay estimate can be extracted during each iteration, in particular if the candidates appear significantly separated (subspaces approximatively orthogonal).

The suggested stopping criteria could possibly be implemented from concurrent noise power estimation. More specifically, suppose that $\hat{\tau}_j$ has just been fixed such that $\mathbf{\Pi}_j := \mathbf{\Pi}_j\left(\hat{\tau}_j; \hat{\tau}_{j-1}, \ldots, \hat{\tau}_1\right)$ is the current projection operator in the recursion. Inspired by the noise variance estimator in (6.11) we may, additional to the sequence of delay estimates $\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_j$, also calculate a sequence of estimated noise powers $\hat{\sigma_1^2}, \hat{\sigma_2^2}, \ldots, \hat{\sigma_j^2}$. Hence, at each recursion depth we calculate

$$\hat{\sigma_j^2} := \frac{1}{M-j}\mathrm{tr}\big(\mathbf{\Pi}_j^{\perp}\mathbf{S}\big), \qquad j = 1, 2, 3, \ldots.$$

Obviously, a decreasing sequence of noise powers is obtained, since as j increases, more and more average signal power is explained in terms of the projection operator $\mathbf{\Pi}_j$. If the SNR is known a priori to the SBA, this information could possibly be utilized to determine a proper stopping criteria of the recursion. Indeed, the current noise power estimate $\hat{\sigma_j^2}$ should not go below the magnitude anticipated from the SNR level.

The options mentioned in this brief section have not been investigated further. Only, they serve to point out the flexible structure and behavior of the SBA and we find it evident to keep these suggestions in mind for future work (**Chapter 9**).

6.6 Chapter Summary

The algorithms investigated during this chapter are summarized in **Table 6.1**.

Method	Computations
Deterministic ML	L-dimensional optimization, matrix inversions
Stochastic ML	L-dimensional optimization, matrix inversions, determinant
ESPRIT	Eigenvalue decomposition
Beamformer	One grid search
Capon	Matrix inverse, one grid search
SBA	L grid searches

Table 6.1: Algorithm overview

In the remaining part of this thesis, we consider only the ESPRIT algorithm and the SBA. The two ML methods are considered too computationally demanding, and in particular, the assumptions for the stochastic ML method will obviously be violated in practice. That is, the zero-forcing observations are not temporally uncorrelated. The Capon method requires the sample covariance matrix to be invertible. Hence, an unrealistic amount of observations need to be stored and the delay parameters may have changed notably while collecting all this data at the receiver.

Chapter 7

Channel Selection and Comparison

In this chapter we elaborate on details concerning two selected Rayleigh fading multipath channel models, in particular, their individual construction and implementation for simulation purposes. These channel models are included in the simulation studies conducted in **Chapter 8** and for comparison reasons, we include as the first model a reference channel configuration proposed by the 3GPP. Typically, the multipath channel models proposed for simulation and performance evaluations are too simplified and do not reflect a satisfactory degree of behavior and properties observable in practice. The number of channel echoes is usually proposed fixed and also the multipath delay parameters are static, i.e. no changes over time. Only the complex amplitudes are changing over time, however, the physical interpretation of the proposed modelings are difficult to accept. Therefore we investigate and propose a more dynamic, time-varying multipath channel model in order to capture channel effects that will be present indeed, when utilizing the wireless channel in dynamic environments in practice. Details in **Chapter 8** then reveal how performance evaluations may critically depend on the particular channel model in use.

7.1 LTE Reference Channel

In **[22, Annex B.2]**, three different multipath channel profiles are given and these serve as reference models for performance evaluations. The first one called *Extended Pedestrian A* is a short channel with a maximum excess delay of 410ns, the second one has a maximum excess delay of 2510ns and is called *Extended Vehicular A* and the last profile called *Extended Typical Urban*, is a long channel with a maximum excess delay of 5000ns. The latter profile is deliberately chosen longer than the duration of the cyclic prefix in order to mimic scenarios where some degree of ISI is present.

We point our attention only at the second profile mentioned, i.e. the Extended Vehicular A is the only reference channel profile considered in this thesis. It consists of nine multipath echoes together with corresponding relative average power terms, see **Table 7.1**.

$\tau_l T_s$ [ns]	0	30	150	310	370	710	1090	1730	2510
$Q(\tau_l)$ [dB]	0.0	−1.5	−1.4	−3.6	−0.6	−9.1	−7.0	−12.0	−16.9

Table 7.1: 3GPP reference profile: Extended Vehicular A.

The above channel consists of $L(t) \equiv 9$ multipath echoes and the delay parameters are all static. The profile corresponds to a multipath channel model as specified in **Chapter 3**. The first delay parameter (excess delay of 0ns) reflects the synchronization procedure carried out at the receiver, i.e. the receiver will synchronize to the first dominant multipath signal echo and perfect synchronization is assumed. In practice, a guard interval of a certain span is enforced during the synchronization procedure. Specifically, the entire excess delay profile in **Table 7.1** is displaced a certain amount to right of zero in order for the synchronization mechanism to react on sudden, distinctive changes in the channel. To imitate this guard interval we choose to shift the entire Extended Vehicular A profile by an amount of ten T_s-samples to the right. The relative power contributions remain unchanged and this altered channel profile is summarized in **Table 7.2**.

$\tau_l T_s$ [T_s]	10.00	10.92	14.61	19.52	21.37	31.81	43.49	63.15	87.11
$Q(\tau_l)$ [dB]	0.0	−1.5	−1.4	−3.6	−0.6	−9.1	−7.0	−12.0	−16.9

Table 7.2: Extended Vehicular A profile shifted ten samples to the right of zero. Delay parameter accuracies given to two decimal places only.

Since the entire profile is shifted evenly to the right, the maximum excess delay of the channel remains unchanged. Hence, no loss of generality and no actual change in the channel is inherited from imitating this synchronization guard interval.

7.1.1 Correlation Properties

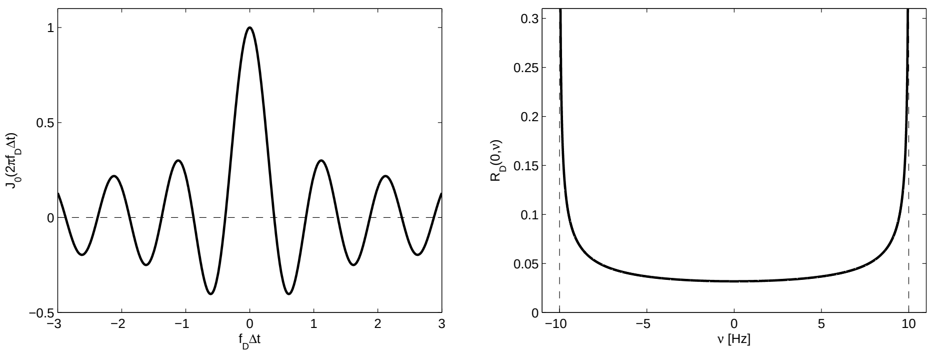
As mentioned in **[22, Annex B.2]**, all nine multipath echoes hold the same normalized Doppler power spectrum (recall **Section 3.3.3**) such that

$$R_D\big(0,\nu;f_D\big) = \frac{1}{\pi f_D}\,\frac{\mathbb{1}\big[|\nu| \leq f_D\big]}{\sqrt{1-(\nu/f_D)^2}}, \tag{7.1}$$

and this Doppler power spectrum is commonly referred to as the *Jakes' spectrum*. Its underlying assumption was first introduced by Clarke, further developed by Jakes and has the advantage of being mathematically tractable but at the same time it has a major drawback in being physically unrealistic. The particular Doppler power spectrum (7.1) is derived based on the so-called *uniform scattering environment* assumption, which is explained in detail in **[1, Section 3.2.1]** including the original references to the work of Clarke and Jakes. Basically, a multipath echo with associated delay τ_l is assumed to consist of a large amount of equally powered sub-components arriving uniformly on a circle centered around the mobile receiver. Hence, it is assumed that a huge number of signal components arrive from all directions, all equally delayed relative to transmission time, all of equal power, and together they form a single echo in the multipath channel. From a physical point of view this modeling is very hard to accept, since indeed it is impossible for signal reflections to arrive from all directions but still at equal delay relative to the time the signal was launched into the wireless channel. Nonetheless, with the uniform scattering assumption imposed on every of the complex-valued channel amplitude processes, their common and normalized time-domain autocorrelation function $R_t(\Delta t)$ can be derived with little difficulty. Referring again to the details in **[1, Section 3.2.1]**, one can show that the normalized autocorrelation function reads

$$R_t\big(\Delta t; f_D\big) = \frac{1}{2\pi}\int_0^{2\pi} \cos\big(2\pi f_D \Delta t \cos\theta\big)\mathrm{d}\theta = J_0\big(2\pi f_D \Delta t\big), \tag{7.2}$$

where $J_0(\cdot)$ denotes the zeroth-order *Bessel function* of the first kind. From the Fourier transform relationships explained in **Section 3.3** one may identify that (7.2) is indeed the Fourier inverse of (7.1) and vice versa. **Figure 7.1** illustrates the Bessel function and the associated Doppler power spectrum.



(a) Bessel function

(b) Jakes' Doppler spectrum

Figure 7.1: Fourier transform pairs with maximum Doppler frequency f_D = 10Hz. Notice how the Doppler power spectrum is symmetric around zero due to the fact that the Bessel function is real (not complex-valued).

The peaky behavior at $\pm f_D$ in **Figure 7.1(b)** clarifies the Doppler properties inherited from the non-linear cosine transformation 'acting' on the uniform scattering environment (only the receiver is assumed to move - recall **Figure 3.2** if necessary).

7.1.2 Simulation Aspects

To approximate/mimic the uniform scattering environment assumed for every single channel echo, we introduce a fixed number R of individual *sub-components* for each of the nine dominant echoes. That is, to capture the time-domain correlation properties imposed in the previous section, we model each complex-valued channel amplitude process according to

$$\alpha_l(t) = \sqrt{\frac{Q(\tau_l)}{R}} \sum_{r=1}^{R} \exp\big(i2\pi f_D \cos(\theta_{l,r})t + i\varphi_{l,r}\big), \qquad l = 1, 2, \ldots, 9,$$

where the statistics of the sub-component *angle of arrivals* $\theta_{l,r}$ and *initial phases* $\phi_{l,r}$, respectively, are given as

$$\theta_{l,r} \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-\pi, \pi), \qquad \varphi_{l,r} \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-\pi, \pi), \qquad r = 1, 2, \ldots, R,$$

The associated channel impulse response reads

$$g(\tau, t) = \sum_{l=1}^{9} \alpha_l(t)\delta\big(\tau - \tau_l T_{\mathrm{s}}\big),$$

where the average power term $Q(\tau_l)$ and delay parameter $\tau_l T_{\mathrm{s}}$ can be extracted from **Table 7.2** for $l = 1, 2, \ldots, 9$. Moreover, the average power terms $Q(\tau_l)$ may be normalized such that the overall channel holds unit average power (appropriate when running simulations). The integer R can be chosen somewhat arbitrarily, however, it should not be 'too small' since the uniform scattering environment is implicitly mimicked through this parameter.

7.2 Dynamic Multipath Channel

The reference channel model presented in the previous section, Extended Vehicular A, is too unrealistic and does not reflect a satisfactory degree of behavior and properties observable in practice. The delay parameters are static, the number of channel echoes is fixed and the time direction correlation properties of the complex-valued amplitude processes are difficult to accept from a physical point of view. Therefore we propose the following more dynamic multipath channel model where inspiration has been gained through the modeling suggestions in **[30, 31]**.

7.2.1 Amplitude Processes and Correlation Properties

From the framework of **Chapter 3** we propose a time-varying multipath channel impulse response of the form

$$g(\tau, t) = \sum_{l=1}^{L(t)} \alpha_l(t)\delta\big(\tau - \tau_l(t)T_{\mathrm{s}}\big),$$

where each complex-valued amplitude process is now modeled as

$$\alpha_l(t) = \sqrt{\frac{Q\big(\tau_l(t)\big)}{R}} \sum_{r=1}^{R} \exp\big(i2\pi f_D \cos(\theta_{l,r})t + i\varphi_{l,r}\big), \qquad l = 1, 2, \ldots, L(t).$$

The overall structure of each channel amplitude is similar to the one presented in the previous section, however, some statistical properties and the average power configuration have been refined. Setting aside the uniform scattering environment we model instead each channel echo from *azimuth* excited sub-components centered around a nominal angle of arrival (wavefront incidence direction). More specifically, we model the nominal angle of arrivals as *hyper parameters* according to

$$\bar{\theta}_l \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-\pi, \pi), \qquad l = 1, 2, \ldots, L(t),$$

and given these, we model

$$\theta_{l,r} | \bar{\theta}_l \overset{\text{i.i.d.}}{\sim} \mathrm{vM}\big(\bar{\theta}_l, \kappa\big), \qquad r = 1, 2, \ldots, R,$$

where $\mathrm{vM}\big(\bar{\theta}_l, \kappa\big)$ denotes the *von Mises distribution* with location parameter $\bar{\theta}_l$ and concentration parameter κ. The von Mises distribution is sometimes referred to as the *circular normal distribution* and when $X \sim \mathrm{vM}\big(\bar{\theta}_l, \kappa\big)$ the associated probability density function reads

$$f_X(x) = \frac{\mathrm{e}^{\kappa \cos(x - \bar{\theta}_l)}}{2\pi I_0(\kappa)}, \qquad \bar{\theta}_l \in [-\pi, \pi), \qquad \kappa \geq 0,$$

where $I_0(\cdot)$ denotes the zeroth order *modified Bessel function*. Choosing $\kappa = 0$ simply gives the uniform distribution on the interval $[-\pi, \pi)$ while choosing κ large the von Mises distribution becomes highly concentrated about the angle $\bar{\theta}_l$ (it approximates a Gaussian distribution with variance $1/\kappa$). From an analytical point of view the von Mises distribution is easier to handle compared to the Gaussian distribution, as we shall soon come to realize. The hyper parameter $\bar{\theta}_l$ is considered fixed/deterministic but any angular value is as likely as another since indeed, a signal wavefront may arrive from any direction in a non-line-of-sight scenario (but notice the crucial difference compared to the assumptions of the uniform scattering environment). The initial phases $\varphi_{l,r}$ are modeled as i.i.d. $\mathcal{U}(-\pi, \pi)$ as in the previous section. In order to carry out the forthcoming calculations, this is a paramount assumption indeed.

From the assumptions made until this point we are straightforwardly able to calculate the normalized time-direction autocorrelation function $R_t\big(\Delta t; f_D, \kappa, \bar{\theta}_l\big)$. Assume for simplicity that the average power term $Q\big(\tau_l(t)\big) \equiv 1$ and consider

$$\begin{aligned} R_t\big(\Delta t; f_D, \kappa, \bar{\theta}_l\big) &:= \mathbb{E}\Big[\alpha_l(t)\alpha_l^*\big(t + \Delta t\big)\Big] \\ &= \frac{1}{R}\sum_{r=1}^{R} \mathbb{E}\Big[\exp\big(-i2\pi f_D \cos(\theta_{l,r})\Delta t\big)\Big] + 0 \\ &= \frac{1}{2\pi I_0(\kappa)} \int_{-\pi}^{\pi} \exp\Big(-i2\pi f_D \cos(x)\Delta t + \kappa\cos(x - \bar{\theta}_l)\Big)\mathrm{d}x, \end{aligned}$$

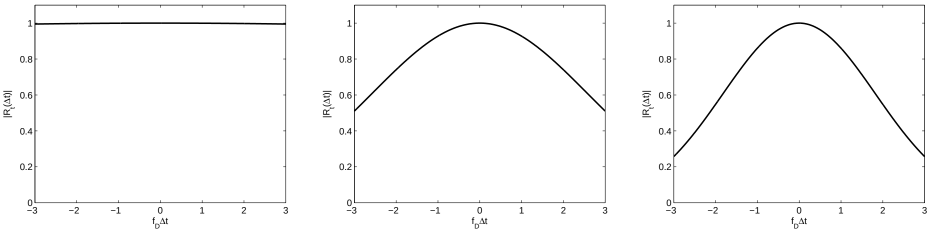
and the latter expression clarifies why the von Mises distribution is convenient compared to a Gaussian equivalent. Applying a few trigonometric identities and invoking the formula

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\exp\Big(z\cos(x)+w\sin(x)\Big)\mathrm{d}x = I_0\Big(\sqrt{z^2+w^2}\Big), \qquad z,w\in\mathbb{C},$$

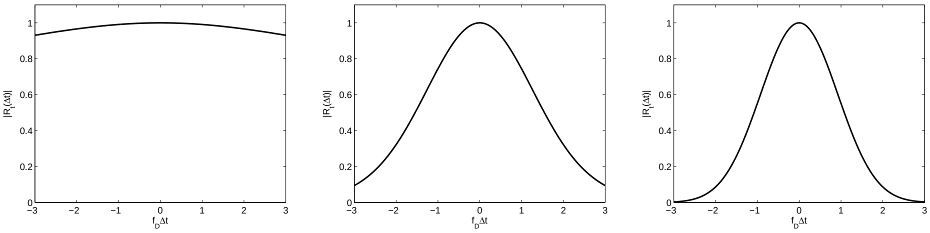
we end up with the closed-form expression

$$R_t\big(\Delta t; f_D,\kappa,\bar{\theta}_l\big) = \frac{I_0\Big(\sqrt{\kappa^2-\big(2\pi f_D\Delta t\big)^2 - i4\pi\kappa\cos(\bar{\theta}_l)f_D\Delta t}\Big)}{I_0(\kappa)} \tag{7.3}$$

which is also obtained from similar calculations in **[31]**. Opposite to the real-valued autocorrelation function obtained in the previous section, the expression (7.3) is complex-valued in general - except when $\bar{\theta}_l = \pm\frac{\pi}{2}$. **Figure 7.2** depicts the shape of $\big|R_t(\Delta t)\big|$ for two selected values of the concentration parameter κ.



(a) $\bar{\theta}_l = 0,\ \kappa = 131$ (b) $\bar{\theta}_l = \frac{\pi}{4},\ \kappa = 131$ (c) $\bar{\theta}_l = \frac{\pi}{2},\ \kappa = 131$



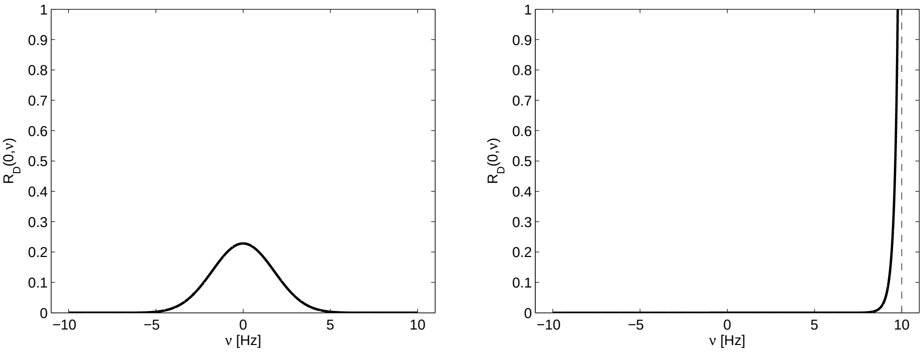
(d) $\bar{\theta}_l = 0,\ \kappa = 33$ (e) $\bar{\theta}_l = \frac{\pi}{4},\ \kappa = 33$ (f) $\bar{\theta}_l = \frac{\pi}{2},\ \kappa = 33$

Figure 7.2: Decorrelation properties for different nominal angle of arrivals $\bar{\theta}_l$. Azimuth spreads of 5° and 10° correspond to $\kappa = 131$ and $\kappa = 33$, respectively. Maximum Doppler frequency $f_D = 10$Hz.

Interesting observations can be made from studying the behavior of the curves in **Figure 7.2(a)** - **Figure 7.2(f)**. For a fixed azimuth (wavefront incidence direction) spread, we observe that the magnitude-wise greatest decorrelation occur when wavefronts impinge perpendicular upon the direction of motion of the receiver. This is also an intuitively appealing conclusion since the non-linear cosine transformation will assign Doppler shifts of the largest possible span exactly when $\bar{\theta}_l = \pm\frac{\pi}{2}$. Hence, due to the assumed azimuth spreading, both negative and positive Doppler shifts will be assigned in this case (compare with the case when $\bar{\theta}_l = 0$ or $\bar{\theta}_l = \pi$). In **[31]** an analytical expression of the Doppler power spectrum $R_D(0,\nu)$ is given, i.e. the Fourier transform of (7.3) with respect to the time-lag variable Δt. We omit the details and state only the ruling formula which reads

$$R_D\left(0,\nu;f_D,\kappa,\bar{\theta}_l\right) = \frac{\exp\left(\kappa\cos(\bar{\theta}_l)\nu/f_D\right)\cosh\left(\kappa\sin(\bar{\theta}_l)\sqrt{1-(\nu/f_D)^2}\right)}{I_0(\kappa)\pi f_D\sqrt{1-(\nu/f_D)^2}} \tag{7.4}$$

where due to the common receiver mobility assumption, the Doppler frequency ν satisfies $-f_D \leq \nu \leq f_D$. It is interesting to note that when choosing $\kappa = 0$ the expression (7.4) reduces to Jakes' Doppler power spectrum in (7.1). The general Doppler behavior intuitively justified from the study of **Figure 7.2** is directly interpretable from **Figure 7.3** showing two particularly interesting special cases of incidence direction.



(a) $\bar{\theta}_l = \frac{\pi}{2}$, $\kappa = 33$ (b) $\bar{\theta}_l = 0$, $\kappa = 33$

Figure 7.3: Normalized Doppler power spectra - maximum Doppler frequency $f_D = 10$Hz.

In the above derivations we assumed for simplicity the average power term $Q(\tau_l(t))$ to be constantly equal to one. We remain to elaborate on these power terms but first we introduce a dynamic model for the delay parameters $\tau_1(t), \tau_2(t), \ldots, \tau_{L(t)}(t)$.

7.2.2 Poisson Line Processes

When a mobile receiver is traveling, e.g. through an urban area, it is likely that the number of multipath channel echoes changes over time. Transitions occur where echoes arise and disappear, the propagation delay parameters drift or fluctuate over time and the incidence directions of the signal wavefronts varies too **[30]**. All effects occur due to mobility and we have already accounted for the fluctuations in incidence directions by the azimuth spread modeling in the previous section. In order to incorporate the remaining mentioned behavior of the multipath channel we follow the lines of **[30]** and incorporate a mathematically tractable Poisson process approach.

Consider a receiver traveling along a regular trajectory. Assume the transitions of echoes arising in the channel to occur according to a *homogeneous Poisson process* with occurrence rate λ_B. Assume furthermore that the time until disappearance of any echo (its lifetime basically) is exponentially distributed with rate parameter λ_S. If the initial number of echoes in the channel is Poisson distributed with mean λ_B/λ_S, then at any given time t, the number $L(t)$ of attending echoes in the channel is a Poisson distributed random variable with mean $\bar{L} := \lambda_B/\lambda_S$. That is, at any time instant the integer $L(t) \sim \mathrm{Pois}\big(\bar{L}\big)$, see **[32, Section 6.10.2]**.

When the l'th echo arises in the channel at time $t_{l,0}$, say, it holds an initial propagation delay $\tau_{l,0}$ with associated Doppler shift

$$\nu_l := f_D \cos(\bar{\theta}_l).$$

This nominal Doppler shift cause the propagation delay to vary over time. For simplicity and due to our receiver mobility assumption, it is convenient to model this variation from straight line advancement[1], i.e.

$$\tau_l(t) = \tau_{l,0} + \frac{\nu_l}{f_c}(t - t_{l,0}), \qquad t \geq t_{l,0},$$

where f_c denotes the carrier frequency (recall (3.4) on page **21**). For the sake of simplicity we furthermore let the initial delay parameters $\tau_{1,0}, \tau_{2,0}, \ldots, \tau_{L(t),0}$ be i.i.d., uniformly on some appropriate interval $[a, b]$, in principle from zero to the end of the cyclic prefix duration. Hence, the pairs $(t_{l,0}, \tau_{l,0})$ of arise times and initial delays form a two-dimensional homogeneous Poisson process on $\mathbb{R} \times [a, b]$. Appropriately, we may think of the associated echo lifetimes and Doppler shifts as *marks* attached to the points, i.e. a marked Poisson point process **[33, Section 3.3]**.

7.2.3 Average Channel Power

Typically, the power delay profile is modeled according to an exponentially decaying function **[1, Section 3.3.1]**, see also **[11, Appendix A]**. Hence, the average power contribution of each channel echo explained as a function of its associated multipath delay reads

$$Q\big(\tau_l(t); C, \rho\big) := \mathbb{E}\Big[|\alpha_l(t)|^2 \,\big|\, \tau_l(t)\Big] = C \exp\big(-\tau_l(t)/\rho\big), \qquad l = 1, 2, \ldots, L(t),$$

where $C > 0$ is a scaling factor and $\rho > 0$ specifies the decay rate of the exponential term. The functional form of Q does not depend on the index l and for notational convenience we therefore write

$$Q\big(\tau; C, \rho\big) = \mathbb{E}\Big[|\alpha|^2 \,\big|\, \tau\Big] = C \exp\big(-\tau/\rho\big).$$

In order to carry out the following average channel power calculations, we assume that all Doppler shifts ν_l are zero. Hereby no delays are drifting over time - they

---

[1] In practice, when both the receiver and the scatterers are moving simultaneously, the drift of each propagation delay is more likely to follow an elliptic curve.

all stay fixed at their assigned initial delay $\tau_{l,0}$. Then, at any given time t, the delays $\tau_1(t), \tau_2(t), \ldots, \tau_{L(t)}(t)$ are mutually independent and identically distributed according to the uniform distribution on $[a, b]$, where $0 \leq a < b \leq \mu + 1$ and $\mu + 1$ is the length of the cyclix prefix (measured in T_s-samples). Notice carefully that the delays $\tau_1(t), \tau_2(t), \ldots, \tau_{L(t)}(t)$ would not be uniformly distributed if the straight line advancement slopes were non-zero. Since the channel amplitude processes are mutually uncorrelated and the associated propagation delays are mutually independent, the joint distribution of $\alpha_l(t)$ and $\tau_l(t)$ does not depend on the index l and neither on $L(t)$. One should carefully notice the variety of assumptions invoked in order to obtain this crucial fact.

The point is that we desire a channel with average power normalized to unity and therefore we need to tune the scaling parameter C such that

$$\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\tau},L(t)}\left[\sum_{l=1}^{L(t)} |\alpha_l(t)|^2\right] = 1.$$

Obviously, the required value of C depends on the average number of channel components $\bar{L}$ and the decay rate ρ. Furthermore, C depends critically on the joint distribution of the delay parameters, in this case (due to our simplifying assumptions) on the edge points a and b. To determine the required value of C we calculate

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\tau},L(t)}\left[\sum_{l=1}^{L(t)} |\alpha_l(t)|^2\right] &= \mathbb{E}_{L(t)}\left[\sum_{l=1}^{L(t)} \mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\tau}}\Big[|\alpha_l(t)|^2\Big] \,\Bigg|\, L(t)\right] \\
&= \mathbb{E}_{L(t)}\left[\sum_{l=1}^{L(t)} \mathbb{E}_{\tau}\Big[\mathbb{E}_{\alpha|\tau}\Big[|\alpha_l(t)|^2 \,|\, \tau_l(t)\Big]\Big] \,\Bigg|\, L(t)\right] \\
&= \mathbb{E}_{L(t)}\Big[L(t) \cdot \mathbb{E}_{\tau}\Big[Q\big(\tau; C, \rho\big)\Big]\Big]. \qquad (7.5)
\end{aligned}$$

Due to our simplifying assumption, any delay $\tau \sim \mathcal{U}[a, b]$ and the innermost expectation in (7.5) reads

$$\mathbb{E}_{\tau|L(t)}\Big[Q\big(\tau; C, \rho\big)\Big] = \frac{C}{b-a}\int_a^b \exp(-\tau/\rho)\mathrm{d}\tau,$$

and therefore we end up with the overall expression

$$\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\tau},L(t)}\left[\sum_{l=1}^{L(t)} |\alpha_l(t)|^2\right] = \bar{L} \cdot \frac{C\rho}{b-a}\big(\exp(-a/\rho) - \exp(-b/\rho)\big).$$

Equating to unity and solving for C yields

$$C\big(\bar{L}, \rho, a, b\big) = \frac{b-a}{\rho\bar{L}\big(\exp(-a/\rho) - \exp(-b/\rho)\big)}, \qquad (7.6)$$

and remark carefully how this outcome depends critically on our assumptions made and also on different parameter selections. First of all, the two event rates λ_B and λ_S must be specified, or equivalently $\bar{L}$ together with one of the two rates. We model the initial multipath delays as i.i.d. uniformly on $[a, b]$ and the final parameter to specify is therefore the decay rate ρ. The scaling factor C can then be calculated from (7.6). If in practice the slopes of the straight line delay advancements are of notable magnitudes (i.e. high Doppler shifts), the above average power calculation will not comprise an acceptable approximation anymore.

7.2.4 Simulation Aspects

The von Mises distribution was introduced in **Section 7.2.1** for analytical convenience, specifically in order for us to evaluate a certain integral. However, for simulation purposes the Gaussian equivalent approximated by the von Mises distribution is much more convenient to apply. Also, the cosine transformation applied to the angular values will do the job of wrapping the Gaussian realizations around the circle (due to its 2π-periodicity), and therefore no modifications need to be made.

Another interesting and challenging aspect concerned with the proposed dynamic channel model is present in terms of its overall implementation. To this end it should come as no surprise why Poisson and exponential assumptions have been invoked to model the time evolution of the attending channel echoes. In fact, the dynamic channel can be updated 'on the fly' according to certain Markov properties inherited from the *memoryless property* of the exponential distribution. From a simulation technical point of view this turns out to be notably advantageous.

Consider the channel state being observed at time t' and suppose the current number of attending echoes in the channel is $L' := L(t')$. At future discrete-point times

$$t' + \Delta t,\ t' + 2\Delta t,\ t' + 3\Delta t, \ldots\ldots,$$

the channel is likewise to be observed/simulated while in-between (continuous time) the channel has changed, i.e. phases have rotated, delays have drifted and so on. The current channel state is identified in terms of

(i) the delay parameters $\tau_1(t'), \tau_2(t'), \ldots, \tau_{L'}(t')$

(ii) the nominal incidence directions $\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_{L'}$

(iii) the $L' \cdot R$ azimuth excited angle of arrivals $\{\theta_{l,r}\}$, and finally

(iv) the $L' \cdot R$ complex exponentials (all current amplitude sub-components)

$$\Big\{ \exp\left(i 2\pi f_D \cos(\theta_{l,r}) t' + i\varphi_{l,r}\right) \Big\}, \qquad \begin{array}{l} l = 1, 2, \ldots, L' \\ r = 1, 2, \ldots, R. \end{array}$$

If no transitions in the channel occur during the time period from t' to $t' + \Delta t$, the present channel state can be directly updated. Using the information in **(i)** and **(ii)**, the delay parameters are updated according to

$$\tau_l\big(t' + \Delta t\big) = \tau_l(t') + \frac{f_D}{f_c}\cos(\bar{\theta}_l)\Delta t, \qquad l = 1, 2, \ldots, L',$$

while each of the $L' \cdot R$ amplitude sub-components in **(iv)** are updated using the information in **(iii)**, simply from multiplication with $\exp\big(i2\pi f_D \cos(\theta_{l,r})\Delta t\big)$, i.e. phase rotations.

At any point in continuous time, new echoes may arise in the channel or current ones may disappear. Due to the memoryless property of the exponential distribution the future channel states become independent of the previous states when given the current channel state. Therefore, to simulate 'on the fly' realizations of the channel we just have to locate the points in continuous time where transitions occur. Since all time evolution terms involved are assumed mutually independent and exponentially distributed, the actual transition points can be located from realizations of a single exponential variable. **Figure 7.4** depicts the scenario considered, showing how mutually independent, exponentially distributed clocks $E_1, E_2, \ldots, E_{L'}$ are being set at time t'. If some of these L' clocks sound before the next observation time $t' + \Delta t$, it means immediate disappearance of the corresponding channel echoes.

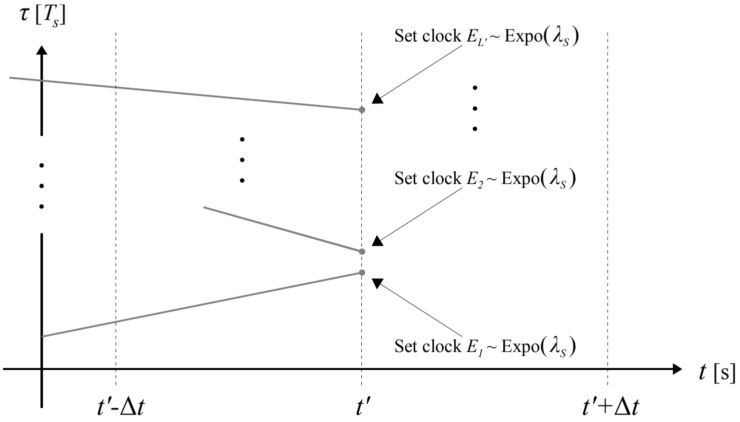


Figure 7.4: At observation time t', an exponential clock is set/reset for each of the currently attending channel echoes.

An additional independent clock $E_B \sim \text{Expo}(\lambda_B)$ is furthermore set at time t'. If this particular clock sounds before time $t' + \Delta t$, it means that a new echo arises in the channel. Instead of keeping track of all these exponential clocks simultaneously, we can simply do by keeping track of

$$X := \min\big\{E_B, E_1, E_2, \ldots, E_{L'}\big\}.$$

When the first clock sounds the remaining ones can be reset since they are all memoryless. We easily realize that

$$X \sim \mathrm{Expo}\big(\lambda_B + L'\lambda_S\big),$$

since

$$\begin{aligned} F_X(x) := P(X \leq x) &= 1 - P\Big(\min\big\{E_B, E_1, E_2, \ldots, E_{L'}\big\} > x\Big) \\ &= 1 - \exp\big(-\lambda_B x\big)\prod_{l=1}^{L'}\exp\big(-\lambda_S x\big) \\ &= 1 - \exp\big(-(\lambda_B + L'\lambda_S)x\big), \end{aligned}$$

which is recognized as the cdf of an exponential distribution with rate-parameter $\lambda_B + L'\lambda_S$. Hence, the realizations of X determine the inter-transition times in the channel and since X inherits a memoryless property too, this ‘combined clock’ may similarly be reset every time no events have occurred. When a transition occurs, its underlying purpose is determined according to the probability distribution

$$\big(p_{\text{arise}}, p_{\text{disappear}}\big) = \left(\frac{\lambda_B}{\lambda_B + L'\lambda_S}, \frac{L'\lambda_S}{\lambda_B + L'\lambda_S}\right),$$

and when disappearance is the outcome, a single of the currently attending echoes is removed according to the uniform probability distribution among the L' candidates.

7.3 Channel Verification

In this section we verify by simulations that the dynamic channel introduced in **Section 7.2** is indeed a Rayleigh fading multipath channel. To this end our choice of implementation is partly verified too. For later on comparison reasons, we desire to imitate (to the extent possible) the power delay profile configuration of the LTE reference channel introduced in **Section 7.1**. Specifically this means that we choose the edge points a and b and the decay rate ρ in such a way that the power delay profile from **Table 7.2** is reflected in an appropriate manner.

From the numbers in **Table 7.2** we choose $a = 10$ and $b = 88$. In this way the dynamic channel holds approximately the same maximum excess delay as the LTE channel. Furthermore, we choose the decay rate ρ such that the relative power relationship between the earliest echo (0.0dB) and the latest echo (−16.9dB) is maintained. Solving for ρ in the following equation of ratios

$$\frac{C\exp\big(-88/\rho\big)}{C\exp\big(-10/\rho\big)} = \frac{\exp_{10}\big(-16.9/10\big)}{\exp_{10}\big(0.0/10\big)}$$

yields a particular value of $\rho = 20$. **Figure 7.5** shows the profile from **Table 7.2** together with the fitted exponential curve (average power terms have been normalized such that the first echo in the fixed LTE profile is unit-reference).

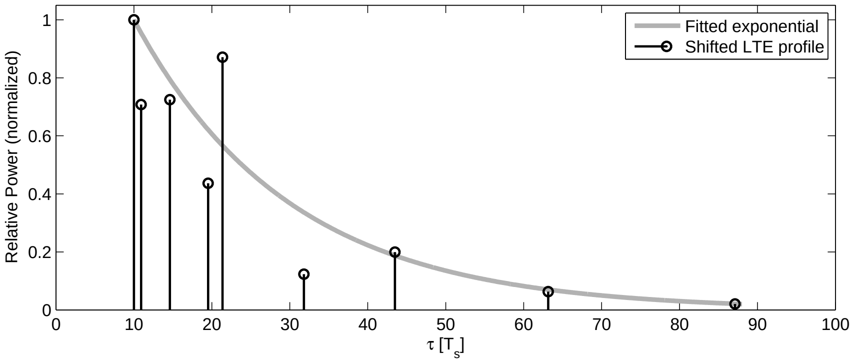


Figure 7.5: Shifted LTE profile and the fitted, exponentially decaying power delay profile.

Beside the parameters a, b and ρ we have to select the values of λ_B and λ_S such that the average number of echoes in the channel is specified. To this end we face a crucial drawback of the dynamic behavior introduced in the channel model. Since at any given time the number $L(t)$ follows a Poisson distribution, inevitably, the event of $L(t)$ being equal to zero is associated with strictly positive probability. Of course, all signal reflections may happen to be blocked in practice (connection loss), however, this fact is not very appreciated from a simulation practical point of view. A straightforward choice for the average number of channel echoes would be $\bar{L} = \lambda_B/\lambda_S = 9$, since the LTE channel constantly includes nine echoes. However, the probability of the zero-echo event would then be $\exp(-9) \approx 0.1234 \cdot 10^{-3}$, which is too large to be considered negligible for simulation purposes. Somewhat heuristically, we therefore choose the average lifespan of each channel echo to be one second, i.e. $\lambda_S = 1$, and from this value we choose $\lambda_B = 15$ such that the average number of channel echoes becomes $\bar{L} = 15$. The probability of the zero-echo event then reads $\exp(-15) \approx 0.3 \cdot 10^{-6}$ and therefore the channel has to be observed (on average) more than three million times before its recurrence. The remaining parameters selected for the following simulation study are given in **Table 7.3**.

Carrier frequency	Maximum Doppler frequency	Number of sub-components	Azimuth spread
f_c	f_D	R	κ
2GHz	10Hz	100	131

Table 7.3: Parameter selections.

A 2GHz carrier frequency is a midpoint compromise of the range where LTE systems are likely to be deployed. On this carrier frequency the wavelength is approximately 15cm and a maximum Doppler shift of 10Hz corresponds to a receiver moving at about five kilometers per hour (walking speed of human beings). The choice of $R = 100$ is rather heuristical, as is the choice of $\kappa = 131$ which corresponds to a standard deviation of 5° in the approximative Gaussian distribution.

In the following, two different realizations of the dynamic channel are considered - partly for implementation verification and partly for comparison evaluation. First we simulate a single, long-lasting (contiguous) realization of the channel. Afterwards, we explain how to initialize the channel in states of equilibria. Optimally, from such perfect initializations we could simulate multiple, non-contiguous channel realizations of short durations. However, due to time constraints, the correct/perfect initialization procedure has not been implemented. Therefore we just approximate the states of equilibria by crude initializations. This crude procedure has also been utilized during the simulation studies conducted in the forthcoming chapter. Despite these imperfect initializations, the average behavior of the channel can be captured from a notably reduced amount of overall realization time from this non-contiguous procedure.

7.3.1 Contiguous Channel Realization

In **Figure 7.6**, a time vs. delay illustration of the straight line advancements of the propagation delays are shown. For illustration purposes, the maximum Doppler frequency have been selected notably larger than 10Hz, and also the initial delays have been drawn from an interval corresponding to the entire duration of the cyclic prefix. The maximum Doppler frequency and the interval $[a, b]$ are immediately reassigned to the values suggested in the previous section.

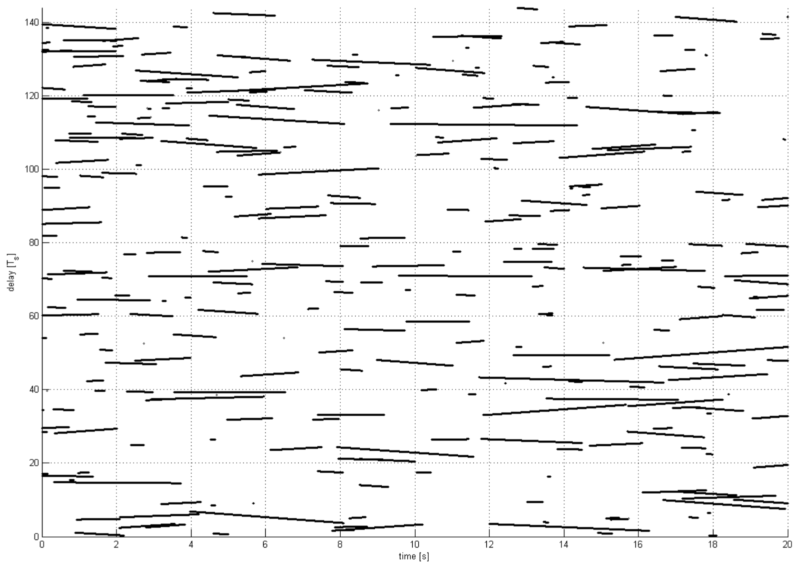
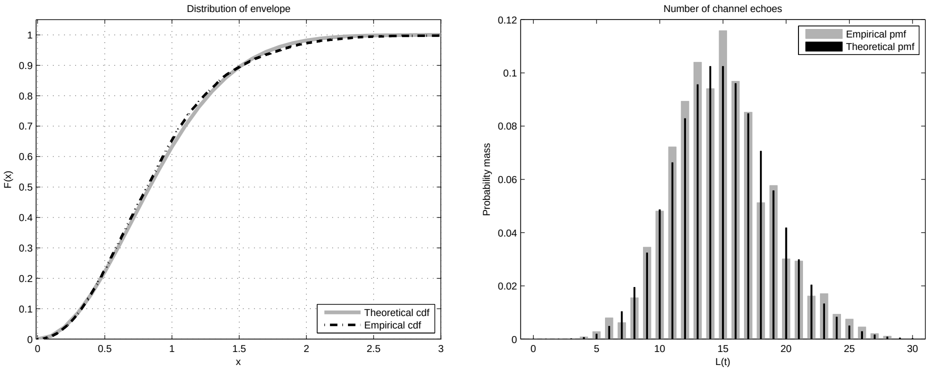


Figure 7.6: A twenty seconds realization of the Poisson line processes.

Figure 7.7 summarizes a five minutes coherent realization of the dynamic channel. The channel has been sampled at discrete times corresponding to every single OFDM symbol, hence 14000 samples every realized second of the channel. The instant envelope and the number of attending echoes in the channel have been monitored during the realization.



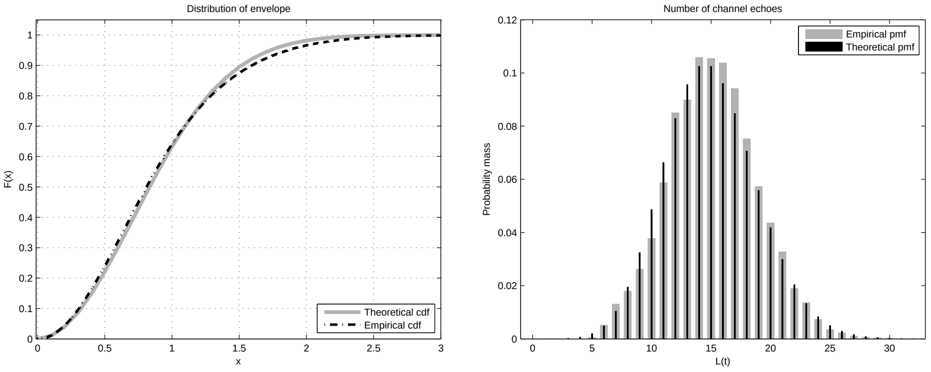
(a) Envelope $|\beta(t)|$ (b) Echo count $L(t)$

Figure 7.7: Channel realization (contiguous) - five minutes duration (30.000 radio frames).

As can be seen, the empirical distribution of the (Rayleigh distributed) envelope $|\beta(t)|$ and the (Poisson distributed) non-negative integer $L(t)$, appear both in reasonable correspondence with the theoretical distributions, recall **Section 3.4** if necessary.

7.3.2 Non-contiguous Channel Realizations

In contrast to **Figure 7.7**, we have summarized in **Figure 7.8**, a 25 seconds non-coherent realization of the dynamic channel. The channel has simply been re-initiated over and over again while observed only for durations of five radio frames. Also in this case the empirical distributions appear in reasonable correspondence with the theoretical distributions.



(a) Envelope $|\beta(t)|$ (b) Echo count $L(t)$

Figure 7.8: Channel realization (non-contiguous) - 25 seconds in total ($500 \cdot 5$ radio frames).

In order to correctly initiate the channel in a state of equilibrium, the first to do is to draw a Poisson distributed number (mean $\bar{L}$) of attending echoes at time $t = 0$. Given this number, we have to simulate backwards in time to locate the particular moments when the uniformly distributed initial delays arose in the channel. These backward recurrence times are i.i.d. exponentials with rate λ_S. Given these echo arise times and the uniformly distributed initial delays, we simply have to simulate forward in time until time zero while letting the delays drift according to their assigned Doppler shifts.

Chapter 8

Simulation Study

In this chapter a simulation study is carried out based on the two channel models described in the previous chapter. We demonstrate the delay estimation performance of both the ESPRIT algorithm and the sequential beamforming algorithm (SBA) within these two multipath channel scenarios. More specifically, we use these two algorithms to provide the delay parameters presupposed by the ENRA.

8.1 Preliminaries and Chapter Outline

In **[12]**, the ESPRIT algorithm is proposed to serve as initial multipath delay acquisition tool for pilot-assisted OFDM systems. For complexity reasons the ESPRIT algorithm is periodically employed for re-acquisitions only. In-between the acquisitions a *delay locked loop* (DLL) tracking scheme is proposed. Based on the proposal in **[12]** we are going to demonstrate a number of crucial facts concerned with the applicability of ESPRIT. In particular, we demonstrate how the ESPRIT algorithm completely fails to provide satisfactory delay estimates when used in the dynamic multipath channel. This severe performance degradation of ESPRIT is however not too surprisingly, considering its underlying assumptions.

Furthermore, and very interestingly, we demonstrate that the SBA is a promising tool for estimation of time-varying multipath propagation delays. However, more insight and refinements are needed in order to comprise a decent competitor against the Robust Wiener Filter.

8.1.1 Simulator Make-up

The scripts forming the simulator have been developed in Matlab and the overall scheme coincide for both scenarios, i.e. for the LTE reference scenario and the scenario including the dynamic multipath channel. The general simulator framework is presented below in a step-by-step pseudo-code formulation and two of the involved steps are of particular importance. These two steps comprise the very essential and non-trivial parts of the simulator and the time consumption levels associated with the development of these parts have been particularly dominant.

The major distinction between the two channel scenarios considered is of course the fact that $L(t)$ and $\tau_1(t), \tau_2(t), \dots, \tau_{L(t)}(t)$, are time-varying in the one case, while fixed/static in the other case. Furthermore, the time direction correlation properties of the complex channel amplitudes are not the same for the two scenarios.

SIMULATOR

```
Fix SNR level and select $\mathcal{T}_{\mathrm{rf}} \in \mathbb{N}$ (radioframes per Monte Carlo run).

Carry out $\mathcal{X}$ independent Monte Carlo runs of the following OFDM transmission scheme.

    (i) Initialize multipath channel (at time $t = 0$).
for $t = 1, 2, 3, \dots, \mathcal{T}_{\mathrm{rf}} \cdot 140$, carry out step (1) to (8).
    (1) Draw $2N_u$ bits and modulate these to $N_u$ symbols (QPSK with Gray-coding).
    (2) for $l = 1, 2, \dots, L(t)$, update the multipath channel parameter $\big[\alpha_l(t), \tau_l(t)\big]$
        given the previous parameter $\big[\alpha_l(t-1), \tau_l(t-1)\big]$.
    (3) Compute the frequency response of the channel and distort the $N_u$ transmitted
        symbols by channel and noise effects (frequency domain).
    (4) if $(t \bmod 7) \in \{1, 5\}$ then collect the $M$ pilot symbol observations, calculate
        $\boldsymbol{h}_{\mathrm{zf}}$ and initiate step (5). if $(t \bmod 7) \notin \{1, 5\}$ then goto step (7).
    (5) Estimate the delay parameters of the channel with ESPRIT or SBA using the
        collection of zero-forcing estimates stored in memory.
    (6) Estimate the channel transfer function (CTF) by ENRA using the delay esti-
        mates obtained in step (5).
    (7) Undo the channel effects using the current estimate of the CTF, thereby obtain-
        ing estimates of the transmitted data symbols.
    (8) if $(t \bmod 7) \in \{1, 5\}$ then demodulate the $N_u - M$ non-redundant data symbols
        and count erroneous bits. if $(t \bmod 7) \notin \{1, 5\}$ then demodulate all $N_u$ non-
        redundant data symbols and count erroneous bits.
```

The above simulation scheme applies for both channel scenarios, but obviously step **(2)** inherits notably dissimilar execution depending on the actual channel under consideration. Step **(2)** and **(5)** comprise the two non-trivial parts of the simulator, in particular step **(2)** when using the dynamic channel. Obviously, a lot of details are omitted in the stepwise pseudo-code given above and a variety of parameters[1] must be defined before execution of the script can take place.

A crucial and important remark can be obtained from a detailed inspection of the

---

[1] Including for instance: N, N_u, M, μ, R, f_D, f_c, K, $\mathcal{I}$, $\mathcal{P}$ and many more.

simulation scheme above. The current estimate of the channel transfer function is only updated twice every slot, i.e. when pilot symbols are present in an OFDM symbol (first and fifth OFDM symbol in every slot). We include no time direction interpolation schemes for in-between improvement of the channel transfer function estimate. Furthermore, we do not make any use of previously obtained delay estimates, i.e. the delay estimation is carried out from scratch every single time.

Initially, we carry out a variety of simulations within the Extended Vehicular A channel scenario. Thereafter, we carry out a similar (but smaller) collection of simulations within the dynamic multipath channel scenario.

8.2 Extended Vehicular A

The true number of channel echoes, i.e. $L = 9$, is fed to the ESPRIT algorithm in all cases considered below. Hence, whenever a delay parameter eigenvalue is plain mistaken by a noise eigenvalue, the corresponding delay parameter will simply not be collected. However, if the ESPRIT algorithm was deliberately forced to overestimate the number of channel components, such a mistaken delay parameter could possibly be collected even though. **Figure 8.1** show the performance of the ENRA when feeded with delay estimates from ESPRIT. Four different values of the snapshot memory parameter K have been used.

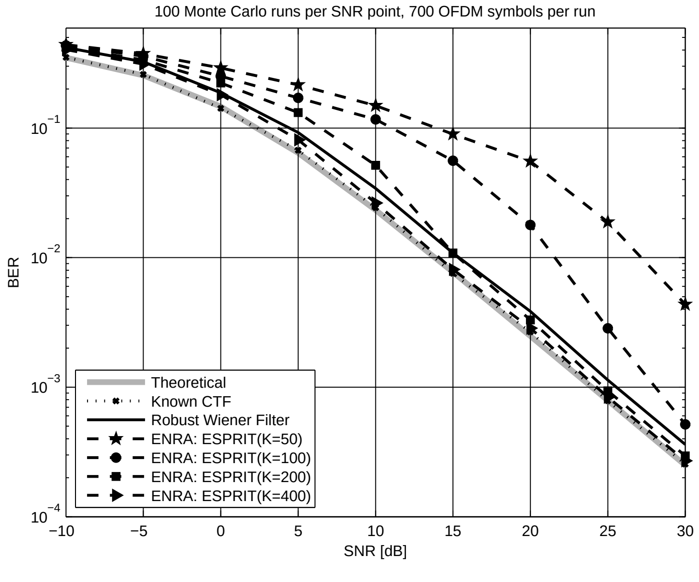


Figure 8.1: Performance of the ENRA when feeded with delay estimates from ESPRIT using persymmetric covariance matrix estimates based on different amounts of snapshots. Known CTF and RWF performance are shown for comparison reasons.

Notice how a memory of $K = 400$ snapshots are needed in order to obtain better performance than the Robust Wiener Filter, when considering the entire SNR-range. This is a highly unrealistic number of observations to store at the receiver. Furthermore, there seems to be a SNR threshold at which the ESPRIT starts to improve its performance. Even the $K = 50$ improves a lot at the very high SNR levels. In order to get a better insight on the actual performance of ESPRIT the delay estimation errors have been monitored by a *Levy metric* comparison, **Figure 8.2** below.

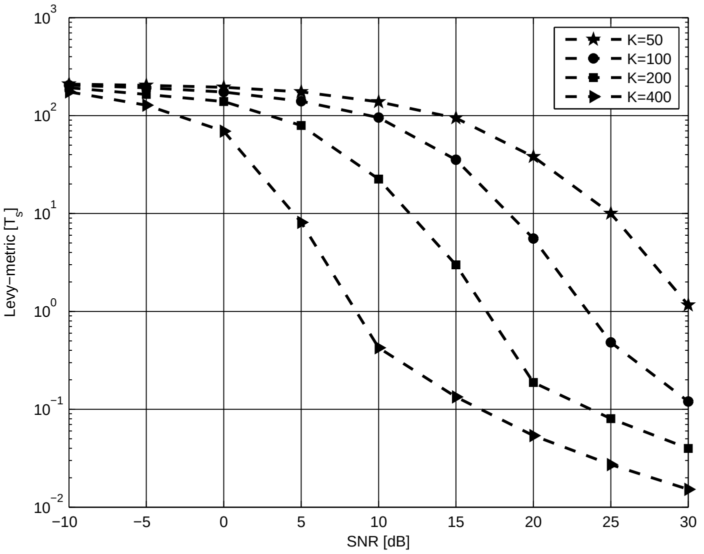


Figure 8.2: Estimation performance of ESPRIT expressed in terms of the Levy-metric.

To reduce the size of the snapshot memory K, the Forward-Backward (FB) approach used in **[12]** has been investigated also. This approach is a heuristical method to obtain more snapshots by sacrificing overall dimensionality. More specifically, a sliding window is introduced and used to cut the observations into smaller pieces, see **[12]** and **[24]**. The performance is shown in **Figure 8.3**.

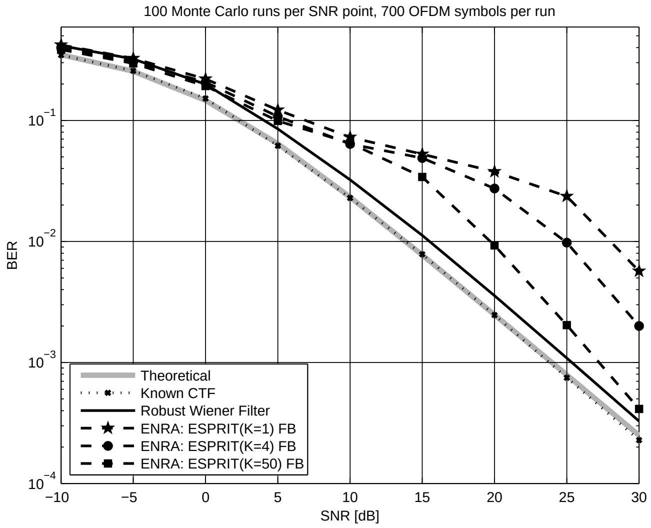


Figure 8.3: Performance of the ENRA when fed with delay estimates from ESPRIT using Forward-Backward averaged persymmetric covariance matrix estimates (50×50). Known CTF and RWF performance shown for comparison reasons.

Also in this case the Levy-metric has been invoked to clarify the actual performance of ESPRIT, shown in **Figure 8.4**. Notice how a notably BER performance improvement is obtained for the curve with $K = 50$.

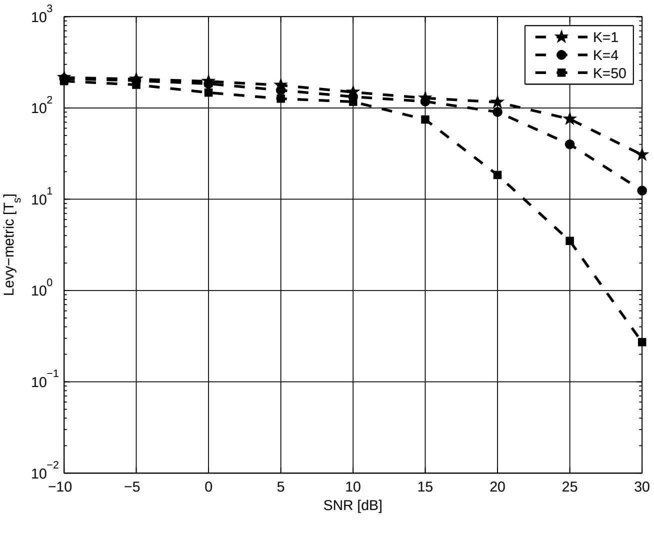


Figure 8.4: Estimation performance of ESPRIT using Forward-Backward approaches, expressed in terms of the Levý-metric.

Hence, provided that enough snapshots are used, the ESPRIT algorithm is able to estimate the delays in a satisfactory manner. We now look at the performance of the SBA. Initial results are shown in **Figure 8.5**.

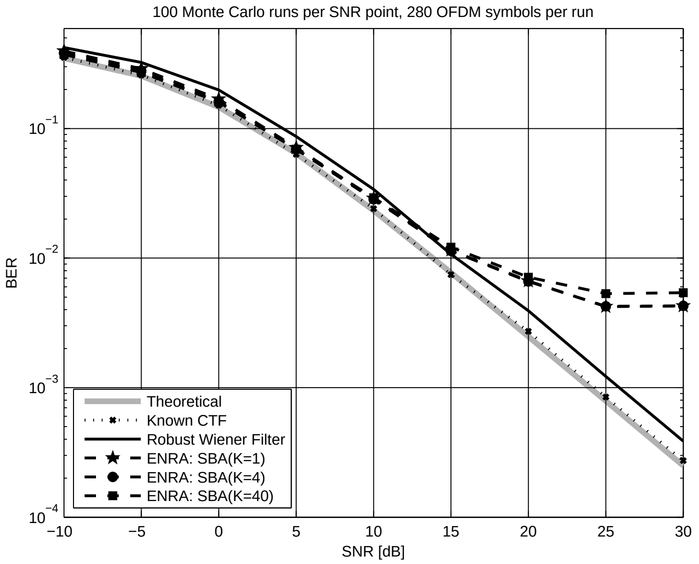


Figure 8.5: Performance of the ENRA when feeded with delay estimates from SBA and the true number of channel echoes. Grid resolution: 10^{-2} of the sampling time.

Notice how the performance of the SBA algorithm seems to be almost identical for the three selected values of the snapshot memory K. In **Figure 8.6**, the number of channel echoes have been deliberately overestimated. The performance of the SBA can be observed to improve notably.

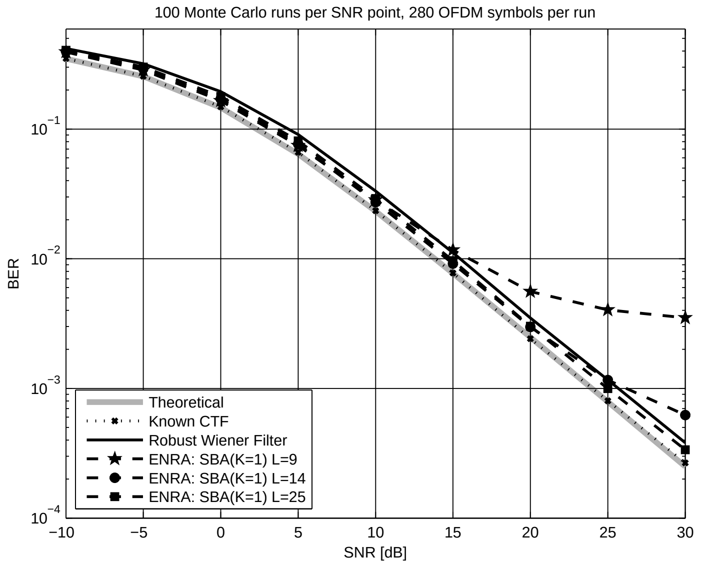


Figure 8.6: Performance of the ENRA when feeded with delay estimates from SBA while overestimating the number of channel echoes in two cases. Grid resolution: 10^{-2} of the sampling time.

Notice that only a single snapshot ($K = 1$) is used above and that by overestimating the number of channel echoes, the BER performance is better than the Robust Wiener Filter in the case when $L = 25$ channel echoes are collected (the true number is nine).

8.3 Dynamic Multipath Channel

The methods investigated for the LTE reference channel are now tested in the dynamic environment. Performance of the ESPRIT algorithm changes dramatically when the delay parameters are varying over time. **Figure 8.7** indicates this clearly, and notice also how the lack of time direction filtering is much more notable at high SNR. This is due to the fact that the delays are drifting.

In this case the ESPRIT algorithm seems non-applicable unless a decent covariance matrix estimate can be obtained from a small amount of snapshots. **Figure 8.8** shows the performance when the Forward-Backward method is invoked to improve the covariance matrix estimate. However, conclusions remain the same - ESPRIT cannot be used to estimate the delays when they fluctuate over time.

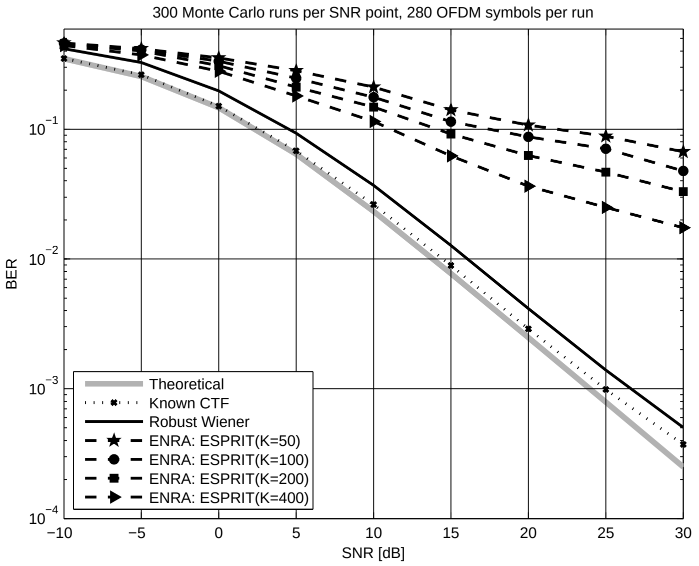


Figure 8.7: Performance of the ENRA when feeded with delay estimates from ESPRIT using persymmetric covariance matrix estimates based on different amounts of snapshots. Known CTF and RWF performance shown for comparison reasons.

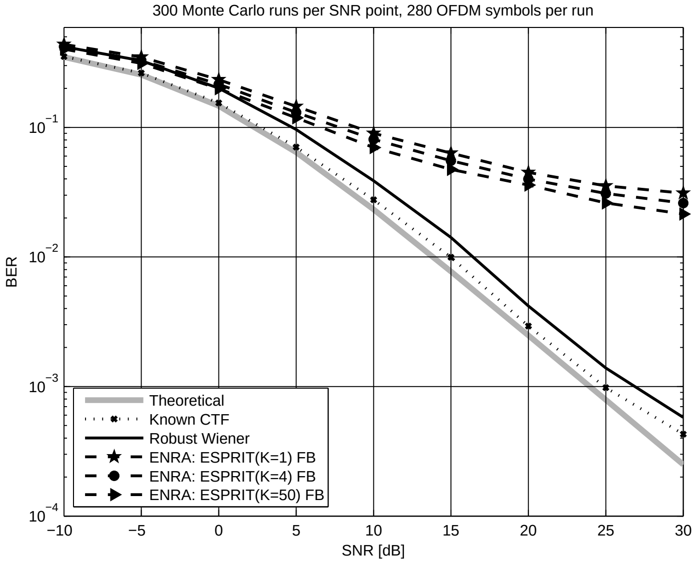


Figure 8.8: Performance of the ENRA when feeded with delay estimates from ESPRIT using Forward-Backward averaged persymmetric covariance matrix estimates based on different amounts of snapshots. Known CTF and RWF performance shown for comparison reasons.

Opposite the the ESPRIT algorithm, the SBA does not suffer from the fact that the delays are moving. This mainly due to the fact that it performs very nice even though only a single snapshot is processed. The performance of the SBA is shown in **Figure 8.8**, here with three different amounts of channel echoes to collect.

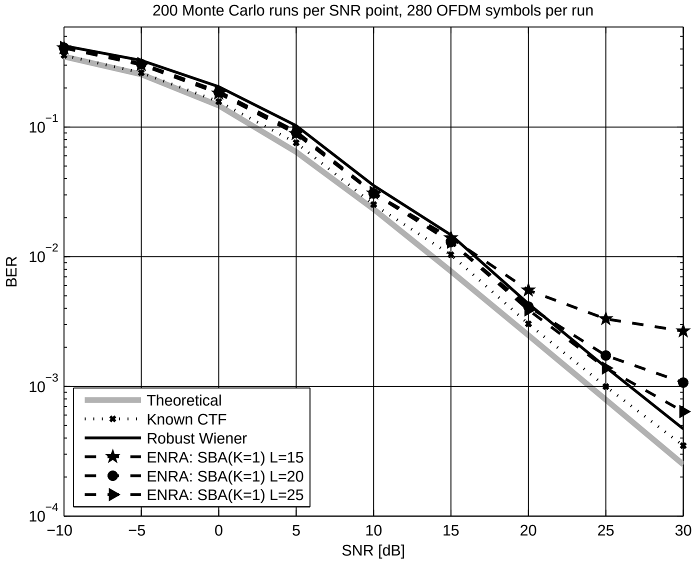


Figure 8.9: Performance of the ENRA when feeded with delay estimates from SBA. Known CTF and RWF performance shown for comparison reasons.

Chapter 9

Conclusion and Future Work

9.1 Conclusions

It has been clearly demonstrated that the performance level of the ESPRIT algorithm depends critically on the channel model assumed. In **[12]**, the ESPRIT algorithm was proposed to serve solely as an initial multipath delay acquisition tool. However, if the delays are drifting over time the algorithm simply cannot be used, not even for acquisitions. Also, it seems that even though the channel model is assumed to hold static delays, still a large amount of observations are needed in order to achieve a satisfactory covariance matrix estimate. This turns out to impact the BER performance for a wide SNR-range (except at very high SNR).

When using the ENRA at high SNR levels, overestimation of the number of echoes (increased complexity due to matrix inverse) seems to be a very acceptable compromise (convergence of the ENRA towards RWF performance). In the case where just a single true delay is missed in the estimation procedure, notably performance degradation can be observed.

A more complete and throughout simulation study could possibly have demonstrated the above conclusions more evidently. However, due to time constraints this have not been possible.

The sequential beamforming algorithm (SBA) exhibits satisfactory estimation accuracy, and it only requires the current observations, i.e. no memory needed. It seems to be a promising tool for estimation of multipath propagation delays. To our knowledge, the SBA has not been proposed in such a context earlier in literature.

The dynamic channel model investigated is considered notably more realistic than the common reference channels proposed in literature and standards, e.g. those specified by the 3GPP. Since algorithm performance can be shown to depend very critically on the channel model in use, we propose to use more realistic channel models in general.

9.2 Future Work

The sequential beamforming algorithm appears as a promising delay estimation tool. However, more insight is needed (stopping criteria, distinguishing side lobes from main lobes, etc.). Recall for instance the discussion in **Section 6.5.1**. The implementation used in this project is rather crude and has not been optimized or refined in any way. Therefore, we sense a great potential with this method, indeed.

The dynamic channel model investigated can absolutely be made even more realistic. For instance, the transitions in the channel are on/off-alike and it seems obvious to invoke more soft transitions. Also, it would be interesting if the zero-echo event problematic could be solved (we simply avoided it by choosing the mean parameter in the Poisson distribution large enough).

Part III

Appendices

Appendix A

Measure Theoretical Details and Stochastic Processes

This appendix presents a selection of mathematical tools and results used throughout the report. Measure theoretical details, random variables and stochastic processes are treated in the following and knowledge of basic probability theory is assumed. Readers who are less interested in the mathematical definitions and aspects may prefer to skip this appendix.

A.1 Sigma-algebra, Measures and Random Variables

This section is inspired by **[34]** and **[33, Appendix B]**. Denote by Ω an arbitrary non-empty set and let $\mathcal{A}$ be a *paving* on Ω, i.e. a set of subsets of Ω.

Definition A.1.1. (Sigma-algebra)
A *sigma-algebra* on Ω is a non-empty paving $\mathcal{E}$ on Ω, with $\varnothing \in \mathcal{E}$ and such that $\mathcal{E}$ is closed under complements and countable unions of its members.

Hence, if $\mathcal{E}$ is a sigma-algebra on Ω, then $\Omega \in \mathcal{E}$ since $\mathcal{E}$ is non-empty. By *De Morgans Laws* **[35, Theorem 1.2.5]**, it follows that $\mathcal{E}$ is closed under countable intersections too. Basically, a sigma-algebra is closed under all common set operations and in general, a sigma-algebra is a very large family of subsets. The smallest possible sigma-algebra is $\{\varnothing, \Omega\}$ and the largest possible sigma-algebra is $\mathcal{P}(\Omega)$, the *powerset* of Ω, i.e. the set of all subsets of Ω. A sigma-algebra $\mathcal{E}$ is said to be *generated* by a non-empty paving $\mathcal{A}$ if $\mathcal{E}$ is the smallest sigma-algebra such that $\mathcal{A} \subseteq \mathcal{E}$. Notice that a smallest sigma-algebra containing $\mathcal{A}$ exists since the intersection of any two sigma-algebras containing $\mathcal{A}$ is also a sigma-algebra containing $\mathcal{A}$, and at least one sigma-algebra contains the paving $\mathcal{A}$, namely the powerset $\mathcal{P}(\Omega)$.

The *Borel* sigma-algebra on $\mathbb{R}^n$, denoted $\mathcal{B}(\mathbb{R}^n)$, is the smallest sigma-algebra containing all open subsets of $\mathbb{R}^n$. By the properties of a sigma-algebra, $\mathcal{B}(\mathbb{R}^n)$ contains all closed subsets of $\mathbb{R}^n$ and hence all half-open subsets too. However, the sigma-algebra $\mathcal{B}(\mathbb{R}^n)$ does not contain all subsets of $\mathbb{R}^n$ (even though it may seem difficult to imagine a subset not included). It turns out that very bizarre subsets of $\mathbb{R}^n$ can be constructed - sets that are never encountered in practice. For technical reasons such sets are desirable to completely disregard, and this is done by considering only the sets in $\mathcal{B}(\mathbb{R}^n)$.

The elements of a sigma-algebra are called *measurable sets* and the pair $(\Omega, \mathcal{E})$, where Ω is a non-empty set and $\mathcal{E}$ a sigma-algebra on Ω, is called a *measurable space.*

Definition A.1.2. (Measure)
Let $(\Omega, \mathcal{E})$ be a measurable space. A *measure* on $(\Omega, \mathcal{E})$ is a mapping $\mu : \mathcal{E} \to [0, \infty]$ such that $\mu(\varnothing) = 0$ and for mutually disjoint $E_1, E_2, \ldots \in \mathcal{E}$ the mapping satisfies

$$\mu\Big(\bigcup_{k=1}^{\infty} E_k\Big) = \sum_{k=1}^{\infty} \mu(E_k). \tag{A.1}$$

The property (A.1) is referred to as *countable additivity* or *sigma-additivity* of the measure. Two important examples of measures are *Lebesgue-measure* and *counting-measure.* The former example is the standard way of assigning length, area and volume to subsets of Euclidian spaces and the latter is a measure taking on integer-values such that $\mu(E) = |E|$, i.e. the number of elements in E and $\mu(E) = \infty$ if the set contains an infinite number of elements. A measurable space $(\Omega, \mathcal{E})$ equipped with a measure μ such that $\mu(\Omega) = 1$, is called a *probability space.* Since $\mu(\Omega) = 1$ the measure is, for obvious reasons, called a *probability measure* and often it its denoted by P instead of μ. Hence, when referring to $(\Omega, \mathcal{E}, P)$ as a probability space it means that $P(\Omega) = 1$ and that we are able to assign probabilities to all elements (appropriate subsets of Ω) in the sigma-algebra $\mathcal{E}$.

Definition A.1.3. (Measurable mapping)
Let $(\Omega_1, \mathcal{E}_1)$ and $(\Omega_2, \mathcal{E}_2)$ denote two measurable spaces and let $f : \Omega_1 \to \Omega_2$. The mapping f is said to be $\mathcal{E}_1$-$\mathcal{E}_2$ *measurable*, if for every $E \in \mathcal{E}_2$ the set

$$\big\{\omega \in \Omega_1 : f(\omega) \in E\big\} \in \mathcal{E}_1,$$

i.e., if the pre-image (under f) of every set in $\mathcal{E}_2$ is a set in $\mathcal{E}_1$.

The familiar and sometimes casual notion of a random variable is made precise with a formal definition of random variables as measurable mappings.

Definition A.1.4. (Random variable)
Denote by $(\Omega, \mathcal{E}, P)$ a latent probability space. Let $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ denote one-

dimensional Euclidian space (the real line) equipped with the Borel sigma-algebra on $\mathbb{R}$. A real-valued *random variable* is a $\mathcal{E}$-$\mathcal{B}(\mathbb{R})$ measurable mapping.

Thus, when expressing probability statements regarding real-valued random variables, we are utilizing properties of the underlying (latent) probability space $(\Omega, \mathcal{E}, P)$, which we do not have any direct access to. For instance, the well-known properties of the *cumulative distribution function* (cdf) F_X of an ordinary real-valued random variable X are directly inherited from the underlying probability space and the measurability of $X : \Omega \to \mathbb{R}$, i.e.

$$F_X(x) := P(X \leq x) = P\Big(\big\{\omega \in \Omega : X(\omega) \in (-\infty, x]\big\}\Big),$$

and the probability statement is well-defined since $(-\infty, x]$ is a half-open interval in $\mathbb{R}$ and hence $(-\infty, x] \in \mathcal{B}(\mathbb{R})$. Since X is a $\mathcal{E}$-$\mathcal{B}(\mathbb{R})$ measurable mapping the pre-image of the interval $(-\infty, x]$ is therefore a measurable subset of Ω, i.e $\big\{\omega \in \Omega : X(\omega) \in (-\infty, x]\big\} \in \mathcal{E}$. Since P is a probability measure it follows directly that $0 \leq F_X(x) \leq 1$ for any $x \in \mathbb{R}$. Furthermore, for $x_1, x_2 \in \mathbb{R}$ with $x_1 \leq x_2$ we calculate

$$\begin{aligned}
F_X(x_2) &= P\Big(\big\{\omega \in \Omega : X(\omega) \in (-\infty, x_2]\big\}\Big) \\
&= P\Big(\big\{\omega \in \Omega : X(\omega) \in (-\infty, x_1]\big\} \cup \big\{\omega \in \Omega : X(\omega) \in (x_1, x_2]\big\}\Big) \\
&= P\Big(\big\{\omega \in \Omega : X(\omega) \in (-\infty, x_1]\big\}\Big) + P\Big(\big\{\omega \in \Omega : X(\omega) \in (x_1, x_2]\big\}\Big),
\end{aligned}$$

where we have used the sigma-additivity of P, which we can do since the sets $(-\infty, x_1]$ and $(x_1, x_2]$ are disjoint, and hence their pre-images are disjoint. It follows that

$$\begin{aligned}
F_X(x_2) &= F_X(x_1) + P\big(x_1 < X \leq x_2\big) \\
&\geq F_X(x_1),
\end{aligned}$$

and the above calculations show that the cdf of a random variable is a non-decreasing function, which is a well-known property for any cdf.

A.2 Stochastic processes

A stochastic (or random) process is a collection of waveforms denoted $X(t, \Lambda)$, where t represents time and Λ is a variable representing outcomes in a sample space S of some underlying random experiment. Usually, the sample space S is a subset of $\mathbb{R}^n$ or $\mathbb{C}^n$. For each specific outcome $\Lambda = \lambda$ a deterministic function of time $x(t, \lambda)$ is associated. These deterministic functions are called *sample functions* or *realizations* of the process (sometimes these cannot be expressed in closed form). Formally, the random variable Λ is defined on a latent or underlying *probability space* $(\Omega, \mathcal{E}, P)$, and specifically, the random variable Λ is a *measurable mapping* $\Lambda : \Omega \to S$.

When neither of t and Λ are fixed, then $X(t,\Lambda)$ is a collection of functions of time. When both are fixed, say $t = t_0$ and $\Lambda = \lambda$ then $X(t_0,\lambda) = x(t_0,\lambda)$ is a single numerical value. If only $t = t_0$ is fixed, $X(t_0,\Lambda)$ is a random variable, however, when only $\Lambda = \lambda$ is fixed, $X(t,\lambda) = x(t,\lambda)$ is a deterministic function of time. Thus $X(t,\Lambda)$ may denote four different quantities and with a slight abuse of notation a stochastic process $X(t,\Lambda)$ is often just denoted by $X(t)$. Then, of course, the interpretation of $X(t)$ as one of the above mentioned quantities should be clear from the context. A formal definition of a stochastic process is as follows **[36]**.

Definition A.2.1. (Stochastic process)
Let S be a sample space of a random experiment and let $\mathcal{T} \subseteq \mathbb{R}$. A real-valued *stochastic process* X is a measurable mapping $X : \mathcal{T} \times S \to \mathbb{R}$ (measurable on S for every $t \in \mathcal{T}$). If $\mathcal{T} \subseteq \mathbb{Z}$ then X is called a *stochastic sequence.*

A complex-valued stochastic process is defined in a straightforward way as

$$Z(t) = X(t) + iY(t),$$

where $X(t)$ and $Y(t)$ are real-valued stochastic processes. Specifically, we can think of $Z(t)$ as a two-dimensional stochastic process $\big(X(t), Y(t)\big)$. Among several others, the following quantities (first and second order moments) are used for the characterization of stochastic processes.

Definition A.2.2.
Let $X(t)$ be a stochastic process. The *mean* of $X(t)$ is defined as

$$\mu_X(t) := \mathbb{E}\big[X(t,\Lambda)\big] = \mathbb{E}\big[X(t)\big],$$

and the *autocorrelation* of $X(t)$ is given by

$$R_{XX}(t_1,t_2) := \mathbb{E}\big[X(t_1,\Lambda)X^*(t_2,\Lambda)\big] = \mathbb{E}\big[X(t_1)X^*(t_2)\big].$$

Note how the abuse of notation is inherited in the last expressions for the mean and the autocorrelation by neglecting the stochastic argument Λ.

Definition A.2.3. (Stationarity)
A stochastic process $X(t)$ is said to be *wide-sense stationary* (WSS) if the mean is constant (does not depend on time) and if the autocorrelation only depends on the time difference. That is, if $\mu_X(t) = \mu_X$ and $R_{XX}(t,t+\tau) = \tilde{R}_{XX}(\tau)$.
A stochastic process $X(t)$ is said to be *strict-sense stationary* (SSS) if all the distribution functions of the process are invariant under an arbitrary translation of time. That is, if for all $k \in \mathbb{N}$ and for all $t_1,\dots,t_k, t_1+\tau,\dots,t_k+\tau \in \mathcal{T}$

$$P\big(X(t_1) \le x_1,\dots,X(t_k) \le x_1\big) = P\big(X(t_1+\tau) \le x_1,\dots,X(t_k+\tau) \le x_k\big).$$

Strict-sense stationarity implies wide-sense stationarity but the converse is not true in general. In order for a stochastic process to be strict-sense stationary it is required that the process is defined on the entire real line (because of the arbitrary translation of time).

Definition A.2.4. (Ergodicity)
A stationary stochastic process $X(t)$ is called *ergodic* if the time-averaged (n, m)'th autocorrelation (product moment) is constant for all realizations of $X(t)$, i.e. if

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t,\lambda)^n x^*(t+\tau,\lambda)^m \mathrm{d}t = K, \quad n, m \in \{0, 1, 2, \ldots\},$$

for all realizations $x(t, \lambda)$ of $X(t)$, where K is a constant, not depending on λ.

Specifically, ergodicity of the process $X(t)$ ensures the time-averaged product moments of any orders to equal their corresponding probabilistic product moments, i.e.

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t,\lambda)^n x^*(t+\tau,\lambda)^m \mathrm{d}t = \mathbb{E}\big[X(t,\Lambda)^n X^*(t+\tau,\Lambda)^m\big],$$

for all realizations of $X(t)$. Other (weaker) forms of ergodicity are often introduced with the above time-averaging criteria relaxed to hold only for product moments up to a certain order, e.g. only first and second order (called wide-sense ergodic processes).

Appendix B

Signal Representations

Today, continuous-time *bandpass signals* are frequently used to model transmitted and received signals in wireless communication systems. A bandpass signal typically results from *carrier modulation* of a baseband signal, i.e. from converting a signal of low frequency (baseband) to a signal of much higher frequency level (bandpass). Bandpass signals are partly used for wireless communication because of their improved propagation properties compared to baseband signals at low frequencies. Also, high frequency implies small wavelength since frequency and wavelength are related by their product being equal to the speed of light (a constant). Hence, to pick up a high frequency signal (short wavelength) requires only a small antenna.

Signal modulators and transmission facilities are built from oscillators that generate real sines and cosines, and therefore bandpass signals are real-valued (not complex-valued). When a signal is transmitted through different environments the wireless channel will inevitably introduce changes in amplitude and phase at each frequency of the signal. In this appendix we partly describe how the analysis of bandpass signals is facilitated through their equivalent lowpass representations. The following section is mainly based on **[1, Appendix A]**.

B.1 Lowpass Representation of Bandpass Signals

We shall see that several useful ways of representing signals are available. A common way to express a continuous-time bandpass signal $s(t)$, at carrier frequency f_c, is

$$s(t) = s_{\mathrm{I}}(t)\cos(2\pi f_c t) - s_{\mathrm{Q}}(t)\sin(2\pi f_c t),$$

where both $s_{\mathrm{I}}(t)$ and $s_{\mathrm{Q}}(t)$ are real lowpass (baseband) signals of bandwidth much lesser than the carrier level. We refer to $s_{\mathrm{I}}(t)$ and $s_{\mathrm{Q}}(t)$ as the *in-phase* and *quadrature* component of $s(t)$, respectively. If we define the complex lowpass signal

$$x(t) := s_{\mathrm{I}}(t) - i s_{\mathrm{Q}}(t), \tag{B.1}$$

we immediately see that we can express the signal $s(t)$ according to

$$\begin{aligned} s(t) &= \mathrm{Re}\left\{x(t)\right\}\cos(2\pi f_c t) - \mathrm{Im}\left\{x(t)\right\}\sin(2\pi f_c t) \\ &= \mathrm{Re}\left\{x(t)\mathrm{e}^{i2\pi f_c t}\right\}. \end{aligned} \tag{B.2}$$

The expression (B.2) is referred to as the complex lowpass representation of the bandpass signal $s(t)$, and the complex baseband signal $x(t)$ is called the *equivalent lowpass signal*[1] for $s(t)$. Alternatively, in (B.1), we can also express the lowpass signal according to

$$x(t) = \beta(t)\mathrm{e}^{i\varphi(t)},$$

where

$$\beta(t) = \left|x(t)\right| = \sqrt{s_{\mathrm{I}}^2(t) + s_{\mathrm{Q}}^2(t)}$$

is referred to as the *envelope* of $x(t)$ and

$$\varphi(t) = \arctan\left(\frac{s_{\mathrm{Q}}(t)}{s_{\mathrm{I}}(t)}\right), \quad s_{\mathrm{I}}(t) \neq 0,$$

is called the *phase* of $x(t)$. Thus, $x(t)$ is called the complex envelope of $s(t)$ since the magnitude of $x(t)$ and $s(t)$ coincide.

A *bandpass channel* is similar to a bandpass signal in the sense that it has a real impulse response $c(t)$ which likewise has a lowpass representation, referred to as the *equivalent lowpass channel*. Now consider a real bandpass signal $s(t)$ as input to a real and linear bandpass channel $c(t)$. Denote by $x(t)$ the equivalent lowpass signal for $s(t)$ and denote by $g(t)$ the equivalent lowpass impulse response for $c(t)$. The received signal $r(t) = \{s * c\}(t)$ is real and since both $s(t)$ and $c(t)$ are bandpass, so is $r(t)$. The received signal therefore has an equivalent lowpass representation, say

$$r(t) = \mathrm{Re}\left\{v(t)\mathrm{e}^{i2\pi f_c t}\right\},$$

where $v(t)$ is the complex envelope of $r(t)$. Omitting certain details, and in particular, calculations of corresponding Fourier transforms of $s(t)$, $c(t)$ and $r(t)$, it can be shown that (see **[1, Appendix A]**)

$$\begin{aligned} v(t) &= \left\{x * g\right\}(t) \\ &= \left\{(s_{\mathrm{I}} - js_{\mathrm{Q}}) * (c_{\mathrm{I}} - ic_{\mathrm{Q}})\right\}(t). \end{aligned} \tag{B.3}$$

From (B.3) we see that, even if $x(t)$ is real[2], by the presence of a quadrature component in $g(t)$, the equivalent lowpass output signal $v(t)$ will be complex. In many wireless channels, e.g. *frequency selective fading channels*[3], the quadrature component $g_{\mathrm{Q}}(t) = \mathrm{Im}\left\{g(t)\right\}$ is present and this imply that the frequency response

---

[1] Sometimes, also referred to as the *complex envelope* of $s(t)$.

[2] i.e. the equivalent lowpass input signal has no quadrature component.

[3] A channel being frequency selective is equivalent to state that the channel is *delay dispersive*, i.e. signals transmitted across the channel is spread in delay, see **Chapter 3**.

$S_c(f)$, of the bandpass channel $c(t)$, is non-symmetric around the carrier frequency.

Hence, through this section we have seen that the received signal can be written

$$r(t) = \mathrm{Re}\left\{\{x * g\}(t)\mathrm{e}^{i2\pi f_c t}\right\},$$

and the primary reason to introduce the equivalent lowpass representations is that it partly removes the dependency on the carrier frequency f_c in the analysis of bandpass signals.

Appendix C

The Complex Gaussian Distribution

Consider a real-valued random vector (X, Y) holding a bivariate Gaussian distribution with mean $(\mu_X, \mu_Y) \in \mathbb{R}^2$ and covariance matrix $\frac{\sigma^2}{2}\mathbf{I}_2$. For short we write

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_2\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \frac{\sigma^2}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right),$$

The pdf of this bivariate Gaussian variable (using generic notation for pdf's) is given by

$$p(x, y) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{1}{\sigma^2} \Big[\big(x - \mu_X\big)^2 + \big(y - \mu_Y\big)^2 \Big] \right), \qquad (x, y) \in \mathbb{R}^2.$$

We define the complex-valued random variable $Z := X + iY$, and in a straightforward manner we define its mean as

$$\mathbb{E}[Z] := \mathbb{E}[X] + i\mathbb{E}[Y] = \mu_X + i\mu_Y,$$

which analogously, we denote by $\mu_Z \in \mathbb{C}$. Then, we define the variance of Z by

$$\mathbb{V}\mathrm{ar}[Z] := \mathbb{E}\Big[\big(Z - \mu_Z\big)\big(Z - \mu_Z\big)^* \Big] = \mathbb{E}\Big[\big|Z - \mu_Z\big|^2 \Big] = \sigma^2.$$

Directly inherited from the above pdf, we find the corresponding pdf of Z to be

$$p(z) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{1}{\sigma^2} \big|z - \mu_Z\big|^2 \right), \qquad z \in \mathbb{C}.$$

We write $Z \sim \mathcal{CN}\big(\mu_Z, \sigma^2\big)$ and say that Z holds a *circular symmetric complex Gaussian distribution* with mean μ_Z and variance σ^2. The phrase 'circular symmetric' refers to the fact of uncorrelated (independent) real and imaginary parts with equal variances (half of the variance of Z).

More generally, we consider a complex M-variate Gaussian distributed vector

$$\boldsymbol{z} = \big[Z_1, Z_2, \dots, Z_M\big]^\top \in \mathbb{C}^M$$

with mean $\boldsymbol{\mu}_{\boldsymbol{z}} \in \mathbb{C}^M$ and positive definite covariance matrix

$$\boldsymbol{\Sigma} := \mathbb{E}\Big[\big(\boldsymbol{z} - \boldsymbol{\mu}_{\boldsymbol{z}}\big)\big(\boldsymbol{z} - \boldsymbol{\mu}_{\boldsymbol{z}}\big)^H\Big].$$

As in the complex scalar case presented above, we can view the complex M-variate vector $\boldsymbol{z}$ as the vector $\big(\mathrm{Re}\,\{\boldsymbol{z}\}\,, \mathrm{Im}\,\{\boldsymbol{z}\}\,\big)$ of double dimensionality, i.e. a real-valued jointly Gaussian random vector of dimension $2M$.

C.1 Persymmetric Covariance Matrices

Let $\boldsymbol{y}(1), \boldsymbol{y}(2), \dots, \boldsymbol{y}(K)$ be an observed sequence of i.i.d. complex M-variate Gaussian vectors with zero mean and unknown covariance matrix $\mathbf{R}$. From the i.i.d. assumption it follows that the negative (and normalized) log-likelihood function is given by

$$-\frac{1}{K}\ell\big(\mathbf{R}; \boldsymbol{y}(1), \dots, \boldsymbol{y}(K)\big) = C + \log\det(\mathbf{R}) + \mathrm{tr}\big(\mathbf{R}^{-1}\mathbf{S}\big), \tag{C.1}$$

where $\mathbf{S}$ is the sample covariance matrix defined as

$$\mathbf{S} := \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{y}(k)\boldsymbol{y}^H(k).$$

Minimization of (C.1) with respect to a Hermitian and positive definite matrix yields the maximum likelihood estimate of the unknown covariance matrix $\mathbf{R}$, i.e.

$$\hat{\mathbf{R}} = \arg\min_{\mathbf{R}>0}\Big\{\log\det(\mathbf{R}) + \mathrm{tr}\big(\mathbf{R}^{-1}\mathbf{S}\big)\Big\}. \tag{C.2}$$

For the moment, suppose that the true covariance matrix $\mathbf{R}$ is *persymmetric*, i.e.

$$\mathbf{RJ} = \big(\mathbf{RJ}\big)^\top \quad\Longleftrightarrow\quad \mathbf{R} = \mathbf{J}\mathbf{R}^\top\mathbf{J}, \tag{C.3}$$

where $\mathbf{J}$ is the $M \times M$ *reversal matrix* with 1's on its entire anti diagonal and zeros elsewhere (also introduced in the proof of **Lemma 4.2.1**). As can be readily verified from the property (C.3), a persymmetric matrix is symmetric in its anti diagonal. If by theoretical means, the true covariance matrix $\mathbf{R}$ is not just Hermitian and positive definite but also persymmetric, it is certainly reasonable to optimize in (C.2) under this additional constraint. The following general result turns out to be useful in particular.

Lemma C.1.1.
Let $\mathbf{A}$ be a fixed, positive definite $M \times M$ matrix. Then, the matrix function

$$f\big(\mathbf{B}; \mathbf{A}\big) = \log\det(\mathbf{B}) + \mathrm{tr}\big(\mathbf{B}^{-1}\mathbf{A}\big),$$

is minimized uniquely with respect to a positive definite matrix at $\mathbf{B} = \mathbf{A}$.

Proof. Assume that $\mathbf{B}$ is positive definite, and hence the matrix $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ is positive definite too. Let $\lambda_1, \lambda_2, \ldots, \lambda_M > 0$ be the eigenvalues of $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ and notice that these are also the eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$. Now consider

$$\begin{aligned} f\big(\mathbf{B};\mathbf{A}\big) - f\big(\mathbf{A};\mathbf{A}\big) &= \log\det(\mathbf{B}) + \operatorname{tr}\big(\mathbf{B}^{-1}\mathbf{A}\big) - \log\det(\mathbf{A}) - M \\ &= -\log\det\big(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\big) + \operatorname{tr}\big(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\big) - M \\ &= \sum_{m=1}^{M} \Big(-\log\lambda_m + \lambda_m - 1\Big) \geq 0, \end{aligned}$$

since $\log x \leq x - 1$ for all $x \geq 0$, where equality holds if and only if $x = 1$. Hence, $f\big(\mathbf{B};\mathbf{A}\big) \geq f\big(\mathbf{A};\mathbf{A}\big)$ with equality exactly when $\lambda_m = 1$ for $m = 1, 2, \ldots, M$. Since these were the eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$ we see that $f\big(\mathbf{B};\mathbf{A}\big)$ is minimized uniquely at $\mathbf{B} = \mathbf{A}$. ■

Assume that the sample covariance matrix $\mathbf{S}$ is positive definite, which will be true with probability one if and only if $K \geq M$. When $K < M$ the sample covariance matrix is rank deficient and therefore certainly not positive definite. When $K = M$ the sample covariance matrix is potentially of full rank and we need to calculate the probability of $\mathbf{S}$ still being rank deficient, i.e. that a least one of the observations $\boldsymbol{y}(k)$ can be written as a linear combination of the $K-1 = M-1$ remaining observations. Such a linear combination comprise a hyperplane of dimension $M-1$ in $\mathbb{C}^M$ and therefore

$$P\Big(\big\{\mathbf{S}\text{ has rank strictly less than } M\big\}\Big) = 0.$$

The interested reader should consult **[37]** for a formal, lucid and very short proof of the above discussed positive definiteness of the sample covariance matrix.

In terms of **Lemma C.1.1** the unconstrained maximum likelihood estimate of $\mathbf{R}$ is given by the sample covariance matrix $\mathbf{S}$. However, if we restrict our attention to a persymmetric estimate we can make straightforward use of the fact that $\mathbf{R} = \mathbf{J}\mathbf{R}^\top\mathbf{J}$ and rewrite

$$\operatorname{tr}\big(\mathbf{R}^{-1}\mathbf{S}\big) = \operatorname{tr}\Big(\tfrac{1}{2}\mathbf{R}^{-1}\mathbf{S}\Big) + \operatorname{tr}\Big(\tfrac{1}{2}\big(\mathbf{J}\mathbf{R}^\top\mathbf{J}\big)^{-1}\mathbf{S}\Big) = \operatorname{tr}\Big(\mathbf{R}^{-1}\tfrac{1}{2}\big(\mathbf{S} + \mathbf{J}\mathbf{S}^\top\mathbf{J}\big)\Big).$$

This rewritten expression together with **Lemma C.1.1** yields the minimizing argument in (C.2), i.e. the persymmetric maximum likelihood estimate

$$\widetilde{\mathbf{S}} := \frac{1}{2}\Big(\mathbf{S} + \mathbf{J}\mathbf{S}^\top\mathbf{J}\Big).$$

Notice, that if the true covariance matrix $\mathbf{R}$ is *Toeplitz* then in particular, it is persymmetric. Hence, if the true covariance matrix is Toeplitz we would intuitively expect the persymmetric estimate $\widetilde{\mathbf{S}}$ to be 'more accurate' or 'better' than the estimate provided by the sample covariance matrix $\mathbf{S}$. Heuristically, we could furthermore replace the persymmetric estimate by a crude/naive Toeplitz estimate, however, as mentioned in **[14, Section 4.8]**, we should not. The reason for this is seen from theoretical considerations with a finite set of 'ideal' (noiseless) observations. For instance, the complex M-variate observations $\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots, \boldsymbol{y}(K)$ could be realizations from

$$\boldsymbol{y}(k) = \mathbf{A}\boldsymbol{s}(k) + \boldsymbol{w}(k),$$

e.g. a case of L uncorrelated narrowband signal contributions embedded in complex AWGN (see (5.7) on page **55**). Now consider a crude Toeplitz estimate of the true covariance matrix obtained by averaging the elements along each diagonal of $\mathbf{S}$ or $\widetilde{\mathbf{S}}$. If the K observations available are completely noiseless (infinite signal-to-noise ratio), both $\mathbf{S}$ and $\widetilde{\mathbf{S}}$ have rank equal to L, however, the averaged Toeplitz estimate will be a full rank matrix in general **[14]**. Hence, a finite number of observations collected in a noiseless environment give rise to a Toeplitz estimate not reflecting the true dimension of the underlying signal.

No general closed-form solution for Toeplitz constrained ML estimation of the true covariance matrix $\mathbf{R}$ is known. However, approximative methods have been suggested in the literature **[38]**, where in particular, we find the following expectation maximization (EM) approach interesting, at least from a conceptual point of view.

C.2 Toeplitz Constrained Approximation

We describe the following EM algorithm only in brief manners. Preliminary simulation studies carried out in an early period of the project showed the EM approach rather unfeasible from a computational point of view. Hence, we present it in this appendix only to mention that the method has actually been investigated during the project - with inspiration gained from **[38]**.

Denote by $\boldsymbol{s} := \left[s_1, s_2, \ldots, s_M\right]^\top$ the first column of the $M \times M$ sample covariance matrix $\mathbf{S}$ defined in the previous section. In the following we construct a $(2M-1) \times (2M-1)$ *circularly extended* version of $\mathbf{S}$, denoted by $\mathbf{S}_{\text{ext}}$, and we consider this circular matrix for two different reasons. We know from **Lemma 4.2.1** that any circular matrix is normal, and hence, the complex spectral theorem is directly applicable. Furthermore, we also know that the $(2M-1)$-point DFT matrix comprise the eigenvector basis needed to carry out the diagonalization.

We define the first column of the circularly extended matrix $\mathbf{S}_{\text{ext}}$ as

$$\boldsymbol{s}_{\text{ext}} := \left[s_1, s_2, \ldots, s_{M-1}, s_M, s_M^*, s_{M-1}^*, \ldots, s_3^*, s_2^*\right]^\top \in \mathbb{C}^{2M-1},$$

which is the original first column $\boldsymbol{s}$ concatenated with a reversed and complex conjugated version of itself (but with the very last component discarded). The circularly extended $(2M-1) \times (2M-1)$ matrix $\mathbf{S}_{\text{ext}}$ is then build from its first column $\boldsymbol{s}_{\text{ext}}$ and the $2M-2$ subsequent one-step circularly shifts of $\boldsymbol{s}_{\text{ext}}$. By construction, $\mathbf{S}_{\text{ext}}$ is Hermitian and the original sample covariance matrix $\mathbf{S}$ is present in its upper left corner. Immediately, it follows that

$$\mathbf{S}_{\text{ext}} = \mathbf{F}^H \boldsymbol{\Lambda} \mathbf{F},$$

where $\mathbf{F}$ is the $(2M-1)$-point DFT matrix as defined in **Section 2.2** and where

$$\mathbf{\Lambda} = \mathrm{diag}\Big((2M-1)^{1/2}\mathbf{F}\boldsymbol{s}_{\mathrm{ext}}\Big),$$

i.e. a diagonal matrix of the $2M-1$ real eigenvalues of $\mathbf{S}_{\mathrm{ext}}$. These eigenvalues hold all the information about the original sample covariance matrix $\mathbf{S}$. Specifically, by defining the $(2M-1)\times M$ matrix $\mathbf{F}_{\mathrm{sub}}$ to consist of the first M columns of $\mathbf{F}$, we immediately find

$$\mathbf{S} = \mathbf{F}_{\mathrm{sub}}^{H}\mathbf{\Lambda}\mathbf{F}_{\mathrm{sub}}.$$

The above considerations are now utilized to form a sequence of Toeplitz constrained approximations to the true covariance matrix $\mathbf{R}$. For some $k > 0$, denote by $\hat{\mathbf{R}}^{(k)}$ the current Toeplitz constrained estimate of $\mathbf{R}$. Denote by $\hat{\mathbf{R}}_{\mathrm{ext}}^{(k)}$ the corresponding circularly extended matrix and let $\hat{\boldsymbol{\lambda}}^{(k)}$ be a vector holding its $2M-1$ real eigenvalues such that

$$\hat{\mathbf{\Lambda}}^{(k)} := \mathrm{diag}\big(\hat{\boldsymbol{\lambda}}^{(k)}\big).$$

EM Algorithm

Initialize $\hat{\mathbf{R}}^{(0)} = \frac{\mathrm{tr}(\mathbf{S})}{M}\mathbf{I}_M$, and for $k = 0, 1, 2, \ldots$

E-step: $$\mathbf{Q} = \hat{\mathbf{\Lambda}}^{(k)}\mathbf{F}_{\mathrm{sub}}\big(\hat{\mathbf{R}}^{(k)}\big)^{-1}\Big(\mathbf{S}-\hat{\mathbf{R}}^{(k)}\Big)\big(\hat{\mathbf{R}}^{(k)}\big)^{-1}\mathbf{F}_{\mathrm{sub}}^{H}\hat{\mathbf{\Lambda}}^{(k)}$$

$$\hat{\boldsymbol{\lambda}}^{(k+1)} = \hat{\boldsymbol{\lambda}}^{(k)} + \mathrm{diag}\big(\mathbf{Q}\big)$$

M-step: $$\hat{\mathbf{R}}^{(k+1)} = \mathbf{F}_{\mathrm{sub}}^{H}\hat{\mathbf{\Lambda}}^{(k+1)}\mathbf{F}_{\mathrm{sub}}$$

The recursion is terminated when two consecutive matrix estimates are close within an appropriate metric, e.g. in terms of a *Frobenius norm* threshold.

Bibliography

[1] Andrea Goldsmith.
Wireless Communications.
Cambridge University Press, 2005. ISBN 978-0-521-83716-3.

[2] Andrew S. Tanenbaum.
Computer Networks.
Pearson Education, Inc., 2003. ISBN 0-13-038488-7.

[3] John G. Proakis and Masoud Salehi.
Communication Systems Engineering.
Prentice Hall, 2002. ISBN 0-13-095007-6.

[4] Carles Navarro Manchón.
Next Generation of Mobile Systems:
The Channel Estimation Problematic in Downlink OFDM.
Department of Communication Technology, Aalborg University, 2006. Master Thesis.

[5] Erik Dahlman, Stefan Parkvall, Johan Sköld and Per Beming.
3G Evolution: HSPA and LTE for Mobile Broadband.
Elsevier, 2007. ISBN 978-0-12-372533-2.

[6] Christian Rom, Carles Navarro Manchón, Luc Deneire, Troels Bundgaard Sørensen and Preben Mogensen.
Unification of Frequency Direction PACE Algorithms for OFDM.
In Proceedings of Wireless Personal Multimedia Communications, 2007. pp. 213-217.

[7] Christian Rom.
Physical Layer Parameter and Algorithm Study in a Downlink OFDM-LTE Context.
Department of Electronic Systems, Aalborg University, 2008. Ph.D. Thesis.

[8] David Slepian.
On Bandwidth.
Proceedings of the IEEE, Vol. 64, No. 3, 1976. pp. 292-300.

[9] Jan-Jaap van de Beek, O. Edfors, M. Sandell, S. K. Wilson and P. O. Börjesson.
On Channel Estimation in OFDM Systems.
In Proceedings of IEEE Vehicular Technology Conference, Vol. 2, 1995. pp. 815-819.

[10] Peter Hoeher, Stefan Kaiser and Patrick Robertson.
Two-dimensional Pilot-symbol-aided Channel Estimation by Wiener Filtering.
IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 3, 1997. pp. 1845-1848.

[11] Ove Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson and P. O. Börjesson.
OFDM Channel Estimation by Singular Value Decomposition.
IEEE Transactions on Communications, Vol. 46, No. 7, 1998. pp. 931-939.

[12] Baoguo Yang, Khaled Ben Letaief, Roger S. Cheng and Zhigang Cao.
Channel Estimation for OFDM Transmission in Multipath Fading Channels Based on Parametric Channel Modeling.
IEEE Transactions on Communications, Vol. 49, No. 3, 2001. pp. 467-479.

[13] Mati Wax and Thomas Kailath.
Detection of Signals by Information Theoretic Criteria.
IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 33, No. 2, 1985. pp. 387-392.

[14] Petre Stoica and Randolph Moses.
Spectral Analysis of Signals.
Pearson Prentice Hall, 2005. ISBN 0-13-1113956-8.

[15] Joachim von zur Gathen and Jürgen Gerhard.
Modern Computer Algebra.
Cambridge University Press, 2003. ISBN 0-521-82646-2.

[16] Alan V. Oppenheim, Ronald W. Schafer with John R. Buck.
Discrete-Time Signal Processing.
Prentice-Hall, Inc., 1999. ISBN 0-13-083443-2.

[17] Peter D. Lax.
Functional Analysis.
John Wiley & Sons, 2002. ISBN 0-471-55604-1.

[18] Ye (Geoffrey) Li, Leonard J. Cimini, Jr. and Nelson R. Sollenberger.
Robust Channel Estimation for OFDM Systems with Rapid Dispersive Fading Channels.
IEEE Transactions on Communications, Vol. 46, No. 7, 1998. pp. 902-915.

[19] Peter Olofsson.
Probability, Statistics, and Stochastic Processes.
John Wiley & Sons, 2005. ISBN 0-471-67969-0.

[20] Philip A. Bello.
Characterization of Randomly Time-Variant Linear Channels.
IEEE Transactions on Communications Systems, 1963. pp. 360-393.

[21] Sheldon Axler.
Linear Algebra Done Right.
Springer, 1997. ISBN 0-387-98258-2.

[22] 3rd Generation Partnership Project (3GPP) Technical Specification.
Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Radio Transmission and Reception. TS 36.104 V8.4.0 (Release 8), December 2008.
`http://www.3gpp.org/ftp/specs/html-info/36104.htm`.

[23] 3rd Generation Partnership Project (3GPP) Technical Specification.
Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation. TS 36.211 V8.5.0 (Release 8), December 2008.
`http://www.3gpp.org/ftp/specs/html-info/36211.htm`.

[24] Hamid Krim and Mats Viberg.
Two Decades of Array Signal Processing Research: The Parametric Approach.
IEEE Signal Processing Magazine, Vol. 13, No. 4, 1996. pp. 67-94.

[25] Amin G. Jaffer.
Maximum Likelihood Direction Finding of Stochastic Sources: A Separable Soultion.
In Proceedings of ICASSP, Vol. 5, 1988. pp. 2893-2896.

[26] William R. Wade. *An Introduction to Analysis.*
Pearson Prentice Hall, 2004. ISBN 0-13-124683-6.

[27] C. Henry Edwards and David E. Penney.
Calculus (6e).
Prentice Hall, 2002. ISBN 0-13-095006-8.

[28] Ilan Ziskind and Mati Wax.
Maximum Likelihood Localization of Multiple Sources by Alternating Projection.
IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 36, No. 10, 1988. pp. 1553-1560.

[29] James A. Cadzow.
Multiple Source Location - The Signal Subspace Approach.
IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 38, No. 7, 1990. pp. 1110-1125.

[30] Ralf Heddergott, Urs P. Bernhard and Bernard H. Fleury.
Stochastic Radio Channel Model For Advanced Indoor Mobile Communication Systems.
Proceedings of the 8th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Vol. 1, 1997. pp. 140-144.

[31] Ali Abdi, Janet A. Barger and Mostafa Kaveh.
A Parametric Model for the Distribution of the Angle of Arrival and the Associated Correlation Function and Power Spectrum at the Mobile Station.
IEEE Transactions on Vehicular Technonlgy, Vol. 51, No. 3, 2002. pp. 425-434.

[32] Janine Illian, Antti Penttinen, Helga Stoyan and Dietrich Stoyan.
Statistical Analysis and Modelling of Spatial Point Patterns.
Wiley, 2008. ISBN 978-0-470-01491-2.

[33] Jesper Møller and Rasmus Plenge Waagepetersen.
Statistical Inference and Simulation for Spatial Point Processes.
Chapman & Hall/CRC, 2004. ISBN 1-58488-265-4.

[34] Christian Berg and Tage Gutmann Madsen.
Mål- og integralteori.
Københavns Universitet, 2001. ISBN 87-91180-15-5.

[35] Graeme Cohen.
A Course in Modern Analysis and its Applications.
Cambridge University Press, 2003. ISBN 0-521-52627-2.

[36] K. Sam Shanmugan and Arthur M. Breipohl.
Random Signals: Detection, Estimation and Data Analysis.
John Wiley & Sons, 1988. ISBN 0-471-81555-1.

[37] Richard L. Dykstra.
Establishing the Positive Definiteness of the Sample Covariance Matrix.
The Annals of Mathematical Statistics, Vol. 41, No. 6, 1970. pp. 2153-2154.

[38] Lars P. B. Christensen.
An EM-algorithm for Band-Toeplitz Covariance Matrix Estimation.
IEEE International Conference on Acoustics, Speech and Signal Processing, 2007.