Joint Communication and Positioning based on Interleave-Division Multiplexing

Dissertation

zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
(Dr.-Ing.)
der Technischen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Kathrin Schmeink

Kiel 2012

Tag der Einreichung: 29. Mai 2012
Tag der Disputation: 9. November 2012

Berichterstatter: Prof. Dr.-Ing. Peter Adam Höher
Prof. Dr.-Ing. Reinhold Häb-Umbach

Abstract

Interest in joint communication and positioning is steadily increasing because the combination of both techniques offers a wide range of advantages. On the one hand, synergy effects between communication and positioning like enhanced resource allocation can be exploited. On the other hand, new applications are enabled. Examples comprise a wide area of interest and include the automated localisation of emergency calls, tracking and guiding fire fighters or policemen on a mission, monitoring people with special needs in a hospital or a nursing home, asset tracking, location-based services and so forth. However, it is a challenging task to combine communication and positioning because their prerequisites are quite different. On the one hand, high data rates with little training overhead and low bit error rate are desirable for communication. On the other hand, localisation aims at precise position estimates. Much training is typically spent for that purpose. Given a single transmit signal supporting communication as well as positioning, it is very difficult to fulfil all requirements at the same time. Hence, a flexible configuration is desirable for a joint communication and positioning system with a unified signal structure in order to adjust the tradeoff between both parts to the instantaneous needs.

In this thesis, a new system concept for joint communication and positioning with a unified signal structure is proposed and investigated. The system concept is based on interleave-division multiplexing (IDM) in combination with pilot layer aided channel estimation (PLACE) and multilateration via the time of arrival (TOA). On the one hand, IDM seems to be a suitable candidate for a joint communication and positioning system because of its flexible but simple transmitter structure. On the other hand, multilateration via the TOA enables precise localisation. The connection between the communication and the positioning part is accomplished via an enhanced PLACE unit. Through the incorporation of a channel parameter estimator, not only the channel coefficients of the equivalent discrete-time channel model, that are needed for data detection, but also parameters of the physical channel, that are required for positioning, can be estimated. A priori information about pulse shaping and receive filtering is exploited for that purpose.

The main aim of this thesis is to show the feasibility of the proposed joint communication and positioning system. Hence, a fundamental system setup is analysed systematically. Since many applications of joint communication and positioning are located in urban or indoor environments, a very high positioning accuracy in the centimetre region is desirable. Unfortunately, positioning is most challenging in these environments due to severe multipath propagation. In order to achieve the required accuracies, the positioning part of the proposed system concept can be complemented by other localisation sources like GPS/Galileo and/or motion sensors via sensor fusion. However, the stand-alone per-

formance of the proposed joint communication and positioning system is evaluated by means of Monte Carlo simulations in this thesis. The achieved results are compared to performance limits in terms of Cramer-Rao lower bounds. In order to improve the overall system performance and to enable sensor fusion, soft information with respect to the parameter as well as the position estimates is taken into account. The accuracy of the soft information is analysed with the help of curvature measures. Altogether, promising results are obtained.

Keywords: communication, positioning, navigation, radiolocation, parameter estimation, channel estimation, optimisation, interleave-division multiplexing, pilot layer aided channel estimation, time of arrival, multilateration, soft information.

Kurzfassung

Das Interesse an gemeinsamer Kommunikation und Positionierung nimmt aufgrund vieler Vorteile stetig zu: Durch die Kombination beider Techniken können Synergieeffekte wie beispielsweise eine verbesserte Ressourcenverteilung ausgenutzt werden. Des Weiteren werden neue Anwendungen in den unterschiedlichsten Bereichen ermöglicht: Notrufe können automatisch lokalisiert werden, Feuerwehrmänner und Polizisten im Einsatz können durch eine Verfolgung ihrer Position und gegebenenfalls eine Überwachung ihrer Vitalwerte besser angeleitet und koordiniert werden, Patienten mit speziellen Bedürfnissen in Krankenhäusern können durch ein effizientes Monitoring besser versorgt werden, Ein- und Auslagerungsprozesse in Warenhäusern können erleichtert werden, positionsbezogene Dienste können realisiert werden und vieles anderes mehr. Aufgrund der verschiedenen Anforderungen von Kommunikations- und Positionierungsdiensten ist es schwierig, diese beiden Bereiche zu vereinen. Einerseits sollen große Datenraten mit geringem Trainingsaufwand als auch geringen Bitfehlerraten erreicht werden. Andererseits ist eine hohe Positionierungsgenauigkeit erwünscht, die einen großen Trainingsaufwand erfordert. In einem Systementwurf mit einer einheitlichen Signalstruktur ist es schwer, alle Anforderungen gleichzeitig zu erfüllen. Daher ist ein flexibler Systementwurf von Vorteil, um den Abtausch zwischen Kommunikation und Positionierung an die aktuellen Bedürfnisse anpassen zu können.

Im Rahmen dieser Arbeit wird ein neues gemeinsames Kommunikations- und Positionierungssystem mit einer einheitlichen Signalstruktur vorgeschlagen und untersucht. Der Systementwurf basiert auf Interleave-Division Multiplexing (IDM) in Kombination mit einer Pilotlayer basierten Kanalschätzung und Multilateration mit Hilfe der Signalankunftszeit, im Folgenden Time of Arrival (TOA) genannt. Einerseits ist IDM aufgrund seiner flexiblen, jedoch einfachen Senderstruktur gut für ein gemeinsames Kommunikations- und Positionierungssystem geeignet. Andererseits ermöglicht eine Multilateration mit Hilfe der TOA hohe Positionierungsgenauigkeiten. Die Verbindung zwischen beiden Komponenten wird durch eine erweiterte Pilotlayer basierte Kanalschätzung erreicht: Durch die Verwendung eines Kanalparameterschätzers können sowohl die Kanalkoeffizienten des äquivalenten zeitdiskreten Ersatzkanalmodells, die für die Datendetektion benötigt werden, als auch Parameter des physikalischen Kanals, die für die Lokalisierung erforderlich sind, geschätzt werden. A priori Information bezüglich des Pulsformungs- und Empfangsfilters werden hierfür ausgenutzt.

Das Hauptziel dieser Arbeit ist es, die Realisierbarkeit des vorgeschlagenen gemeinsamen Kommunikations- und Positionierungssystems zu zeigen. Daher wird ein grundlegender Systementwurf systematisch analysiert. Da viele Anwendungen von gemeinsamer

Kommunikation und Positionierung innerhalb von Städten oder Gebäuden angesiedelt sind, ist eine sehr hohe Positionierungsgenauigkeit im Zentimeter-Bereich wünschenswert. Unglücklicherweise ist es in diesen Gebieten aufgrund von starker Mehrwegeausbreitung besonders schwer, die Position eines Objektes zu bestimmen. Allerdings kann die Positionierungskomponente durch andere Lokalisierungsquellen wie beispielsweise GPS/Galileo und/oder Bewegungssensoren mittels Sensorfusion ergänzt werden, um die erforderlichen Genauigkeiten zu erreichen. In Rahmen dieser Arbeit wird jedoch nur die eigenständige Leistungsfähigkeit des vorgeschlagenen Systementwurfs mit Hilfe von Monte Carlo Simulationen untersucht. Die Simulationsergebnisse werden mit Leistungsgrenzen in Form von Cramer-Rao Untergrenzen verglichen. Dabei wird Zuverlässigkeitsinformation bezüglich der geschätzten Parameter und der geschätzten Position berücksichtigt, um die gesamte Systemleistung zu verbessern und Sensorfusion zu ermöglichen. Die Genauigkeit der Zuverlässigkeitsinformation wird mit Hilfe von Krümmungsmaßen analysiert. Insgesamt werden vielversprechende Ergebnisse erzielt.

Stichwörter: Kommunikation, Positionierung, Navigation, Parameterschätzung, Kanalschätzung, Optimierungsverfahren, Interleave-Division Multiplexing, Pilotlayer basierte Kanalschätzung, Signalankunftszeit, Time of Arrival, Multilateration, Zuverlässigkeitsinformation.

Contents

Chapter 1

Introduction

1.1 Motivation

Communication as well as navigation have always been of great importance to humanity. Traditionally, both fields developed independently, but especially during the last decade, the interest in combining both techniques has increased enormously and is still increasing. Comprehensive research projects like the *wireless hybrid enhanced mobile radio estimators (WHERE)* project governed by the German Aerospace Center [Gera] or the *FeuerWhere* project governed by the Freie Univerität Berlin [Fre] have been initiated. Actually, the WHERE project was already completed in 2010, but its research is continued by the WHERE2 project until 2013 [Gerb].

The fundamental question is: *Why* should communication and positioning be combined? The answer is that *joint* communication and positioning offers various benefits and enables new applications. On the one hand, synergy effects between communication and positioning can be exploited. With the help of position information, resource allocation in mobile wireless communication systems can be improved, e.g. handover, power control, beamforming as well as interference management are facilitated. On the other hand, there is a wide variety of potential applications including an automated localisation of emergency calls, tracking and guiding fire fighters or policemen on a mission as well as location-based services (e.g. where is the next hotel, restaurant or bus station). There are different approaches to combine communication and localisation: (i) existing systems like GSM[1]/UMTS[2]/LTE[3] and GPS[4]/Galileo can be combined in a hybrid receiver [HWD+07], (ii) existing communication systems like GSM/UMTS/LTE/WLAN[5] can be extended to provide a position estimate [AN07] or (iii) new systems with a unified signal structure

[1] GSM is the acronym for *global system for mobile communications.*
[2] UMTS is the acronym for *universal mobile telecommunications system.*
[3] LTE is the acronym for *long term evolution.*
[4] GPS is the acronym for *global positioning system.*
[5] WLAN is the acronym for *wireless local area network.*

for communication and positioning can be designed [HS07]. Each approach has its advantages and disadvantages. Using an existing system is cost saving because it is not necessary to built up a new infrastructure. But the fixed system parameters may restrict the performance. The advantage of a new system design is its flexibility to match the requirements of a desired application.

Since the prerequisites for communication and positioning are quite different, it is a challenging task to combine both techniques. On the one hand, high data rates with little training overhead and low bit error rates are desirable for communication. On the other hand, positioning aims at precise location estimates. Much training is typically spent for that purpose. In general, it is very difficult to fulfil *all* requirements at the same time. Hence, a flexible configuration is advantageous for a potential joint communication and positioning system because the tradeoff between communication and positioning is facilitated. Furthermore, adaptability with respect to changing channel conditions and/or user requirements is provided by a flexible system configuration.

Many applications of joint communication and positioning are located in urban or indoor environments. A very high localisation accuracy in the centimetre region is desirable in these environments. Unfortunately, positioning is most challenging in urban and indoor environments due to severe multipath propagation. Even pure positioning systems may fail under these conditions. The required accuracies can only be obtained by combining *all* possible localisation sources via sensor fusion [WKAR06], [HROW07], [RRA10]. Any contribution to the overall positioning accuracy is beneficial. This means that a joint communication and positioning system can be complemented by other localisation sources such as GPS/Galileo and/or motion sensors in order to achieve the required accuracies. Or, from another point of view, the exploration of any further localisation source is desirable. This means that the design and implementation of *new* joint communication and positioning systems is highly motivated.

1.2 Scope and Aim

In this thesis, a new system concept for joint communication and positioning with a unified signal structure is proposed and investigated. The system proposal is based on interleave-division multiplexing (IDM) [HSF08] in combination with pilot layer aided channel estimation (PLACE) [SH06] and positioning based on the time of arrival (TOA) [PL05, Chap. 13]. IDM seems to be a suitable candidate for a joint communication and positioning system because of its flexible yet simple transmitter structure. Data rates as well as the tradeoff between communication and positioning can be easily adapted to the instantaneous needs. In case of an emergency call, emphasis can be laid on the positioning part since the communication requirements with respect to pure speech transmission are rather low. In case of file downloads or video streaming, that demand high data rates, the communication payload can be increased. The main aim of this thesis is to show the feasibility of the proposed system concept. Hence, a fundamental system setup is analysed

where the focus is laid on the positioning part. For the communication part, standard solutions available in the literature are applied after slight modifications, i.e. code and/or receiver design for IDM are out of the scope of this thesis. As already mentioned above, the positioning part of a joint communication and positioning system may be complemented by other localisation sources via sensor fusion in order to improve the localisation accuracy. However, the stand-alone performance of the system proposal, that should be as good as possible, is studied in this thesis.

The connection between the communication and the positioning part is accomplished via an enhanced PLACE unit. On the one hand, channel coefficients of the equivalent discrete-time channel model are required for data detection. On the other hand, parameters of the physical channel like the TOA are needed for positioning. This means that channel estimation is mandatory for both parts even though with different requirements. Hence, *channel parameter estimation*, that delivers parameter estimates as well as channel estimates, is the core part of the proposed joint communication and positioning system and much effort is spent on its analysis. A priori information about pulse shaping and receive filtering is exploited for channel parameter estimation. Therefore, improved channel estimates are obtained compared to standard channel estimation approaches like least-squares channel estimation. This means that the communication part may benefit from the incorporation of channel parameter estimation, that is necessary for positioning purposes, i.e. synergy may be created. If synergy is really created is investigated in this thesis.

1.3 Thesis Outline

This thesis is organised as follows: In Chapter 2, the mobile radio channel is introduced. The basic phenomena that characterise radio propagation are described, statistical channel modelling is discussed and the classification of small-scale channel models is explained. Furthermore, three specific channel models are presented that are applied throughout this thesis in order to assess the performance of the proposed joint communication and positioning system.

In Chapter 3, the system proposal for joint communication and positioning is introduced in detail. At first, several radiolocation methods with their benefits and drawbacks are discussed and the choice of multilateration via the TOA is explained. Afterwards, the communication part, that is based on IDM, is examined. The principles of IDM are presented and potential receiver structures are introduced. Finally, the connection between the communication and the positioning part is explained: The pilot sequence, that is superimposed to the data symbols, can be exploited for communication as well as for positioning purposes. For the incorporation of PLACE, some changes of the basic transmitter and receiver structure are required that are briefly illustrated. In the remaining chapters, the system performance is investigated systematically.

In Chapter 4, channel parameter estimation, that is the core part of the system proposal, is examined in detail. In order to exclude negative effects that may be caused by data layer interference, pure training transmission is assumed. At first, fundamentals needed for channel parameter estimation are summarised. The underlying channel model is reviewed and the employed block fading assumption is justified. Two different maximum-likelihood estimators, that are either based on received samples or on so-called pre-stage channel estimates, are derived. It is shown that both estimators are equivalent. The estimator based on the pre-stage channel estimates is applied throughout this thesis because it is less complex. Before examining the performance of the proposed channel parameter estimator, performance limits in terms of Cramer-Rao lower bounds (CRLBs) are derived. The CRLBs for all three channel models are examined and compared. Since the metric of the maximum-likelihood estimator is nonlinear, it can not be solved in closed form and optimisation algorithms are required in order to find a parameter estimate. Several local and global optimisation algorithms are discussed and their performance for channel parameter estimation is examined. Finally, soft information for channel parameter estimation is studied. The definition of soft information is summarised and its relationship to confidence regions is explained. Based on these fundamentals, two methods to obtain soft information are proposed. So-called curvature measures are introduced that can be used to analyse the accuracy of the proposed soft information methods.

In Chapter 5, localisation based on the obtained channel parameters is investigated. As in Chapter 4, pure training transmission is assumed. Again, a maximum-likelihood estimator as well as performance limits in terms of CRLBs are derived. In case of positioning, the geometric dilution of precision (GDOP) is of great importance because the CRLB depends not only on the pseudorange errors but also on the geometrical configuration. Two different scenarios with small and large average GDOP are considered in this thesis. Different positioning algorithms are presented and their performance in conjunction with the proposed channel parameter estimator is evaluated. Furthermore, soft information with respect to position estimates is discussed. The same methods as already applied for channel parameter estimation are considered and their accuracy is analysed with the help of curvature measures.

In Chapter 6, the overall system performance including data transmission is studied for all three channel models. Especially, the tradeoff between communication and positioning as well as potential synergy effects are of main interest in this chapter. Based on the insights obtained throughout the preceding chapters, the basic receiver structure for IDM with PLACE is adjusted in order to provide the best possible performance for communication as well as positioning purposes. Again, two different geometric scenarios with small and large average GDOP are considered for the performance evaluation. Promising results are obtained.

In Chapter 7, the results obtained throughout this thesis are summarised and conclusions are drawn. Furthermore, potential topics for future work are presented.

Parts of this thesis have been published in [HS07], [SH08], [SBKH10], [SAH11a] and [SAH11b] and [SAH12].

Chapter 2

The Mobile Radio Channel

A signal that is transmitted from a transmitter to a receiver is generally distorted by the channel over which it is transmitted. In order to design effective receiver structures and to assess their performance, it is necessary to characterise and model the distortions introduced by the channel. In case of the mobile radio channel, this is a challenging task because its characteristics depend on many different parameters like the topology and morphology of the surroundings, which additionally change over time due to mobility. In the majority of cases, it is not feasible to model the mobile radio channel deterministically. Thus, statistical channel models are common, which are derived from extensive measurements. Similar propagation environments are summarised in specific scenarios such as rural area, hilly terrain, urban area or indoor scenarios. The basic phenomena that characterise radio propagation are *path loss*, *shadowing*, *multipath propagation* and *Doppler shift*. The first two phenomena influence the (average) received signal strength over large distances of several hundred or thousand metres. Hence, the effects caused by path loss and shadowing are called *large-scale* or *long-term*. Additionally, the received signal strength fluctuates around its mean value due to small changes in distance (in the order of few wavelengths). These fluctuations due to multipath propagation and Doppler shift are called *small-scale* or *short-term* effects. Figure 2.1 shows an example for the received signal strength in dB over the distance between transmitter and receiver in logarithmic scale. Often, large-scale and small-scale effects are modelled separately: Large-scale models are typically used to predict the coverage of a system, whereas small-scale models are needed to assess and compare the performance of different receiver structures. Hence, this thesis mainly focusses on small-scale models. The small-scale behaviour of a channel depends, furthermore, on signal parameters like bandwidth and symbol duration. Based on the relationship between channel and signal parameters, different types of channel models can be classified: *flat fading* (narrowband/non-dispersive) versus *frequency-selective fading* (wideband/dispersive) based on the multipath characteristics and *fast fading* versus *slow fading* based on the Doppler characteristics.

The remainder of this chapter is organised as follows: Since the mobile radio channel is time-variant, the most important fundamentals on linear time-variant systems are intro-

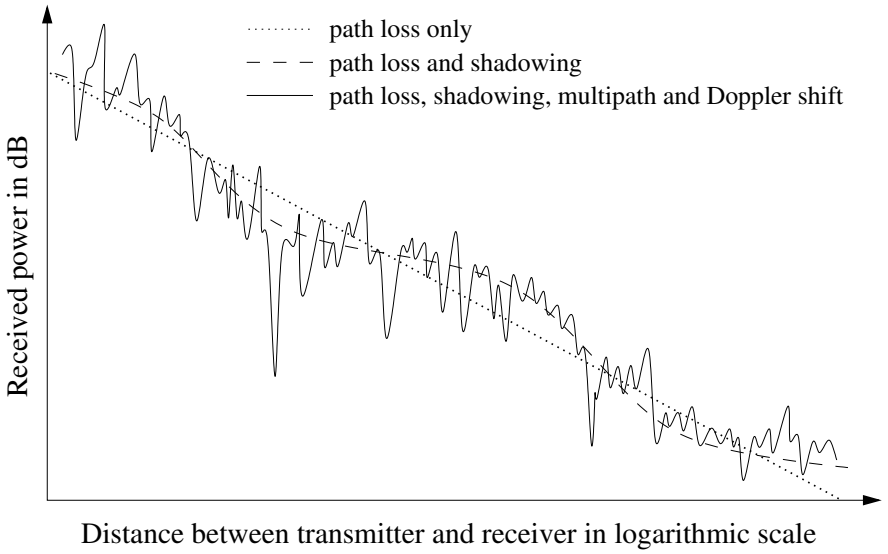


Figure 2.1: Example for the received power (in dB) over the distance between transmitter and receiver (in logarithmic scale) [PL05, p. 95], [Gol05, p. 28].

duced in Section 2.1. In Section 2.2 and Section 2.3, large-scale and small-scale channel modelling is described, respectively. Since this thesis focusses on small-scale channel models, large-scale channel modelling is only explained briefly for the sake of completeness. Small-scale channel modelling is examined in more detail: At first, the phenomena of multipath propagation and Doppler shift are shortly reviewed. In order to facilitate statistical modelling of the mobile radio channel, the assumptions of *wide-sense stationarity (WSS)* and *uncorrelated scattering (US)* are introduced. Afterwards, the classification of small-scale channel models, as already mentioned above, is explained in more detail. For computer simulations, a discrete-time representation of the continuous-time channel is needed. Thus, the *equivalent discrete-time channel model* is introduced in Section 2.4. In Section 2.5, specific channel models that are applied throughout this thesis are presented. Finally, the most important aspects of this chapter are summarised in Section 2.6.

2.1 Fundamentals on Linear Time-Variant Systems

The theory of linear time-invariant (LTI) systems is well known. In contrast, the more sophisticated theory on linear time-variant (LTV) systems is less common. Often, a simplified and, thus, misleading nomenclature is used in the literature concerning time-variant channel modelling. Hence, the most important fundamentals on LTV systems are introduced in this section. Bello presented a comprehensive study of LTV channels in [Bel63], which is beyond the scope of this thesis. Only the most important aspects of [Bel63] as summarised in [Hub96] and [Höh11, Ch. 21.3.1] are presented here in order to clarify the nomenclature.

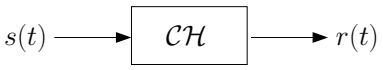


Figure 2.2: Linear time-variant channel.

Let t denote the time and let $s(t)$ be the input/transmit signal, that is transformed by the LTV system/channel $\mathcal{CH}$ into the output/received signal (see Figure 2.2):

$$s(t) \overset{\mathcal{CH}}{\longmapsto} r(t). \tag{2.1}$$

In case of an LTV channel, the impulse response $c_0(t_E, t)$ does not only depend on the time t, but also on the excitation time t_E:

$$\delta(t - t_E) \overset{\mathcal{CH}}{\longmapsto} c_0(t_E, t)\,. \tag{2.2}$$

Given this time-variant channel impulse response, the received signal can be written as

$$r(t) = \int_{-\infty}^{\infty} s(t_E) c_0(t_E, t)\,\mathrm{d}\,t_E. \tag{2.3}$$

The main difference to an LTI channel is that the output $r(t)$ can *not* be expressed as the convolution of the input $s(t)$ and the impulse response $c_0(t_E, t)$. In order to describe the input/output behaviour of the channel by a convolution, a new function $c(\tau, t)$ called *weight function* needs to be introduced:

$$r(t) = \int_{-\infty}^{\infty} s(t - \tau) c(\tau, t)\,\mathrm{d}\,\tau, \tag{2.4}$$

where τ denotes the delay of the transmitted signal due to the channel. This representation of the channel corresponds to an input-delayline structure shown in Figure 2.3, which is typically used to characterise LTV channels. Often, the time-variant weight function $c(\tau, t)$ is misleadingly denoted as time-variant impulse response in the literature. Substituting

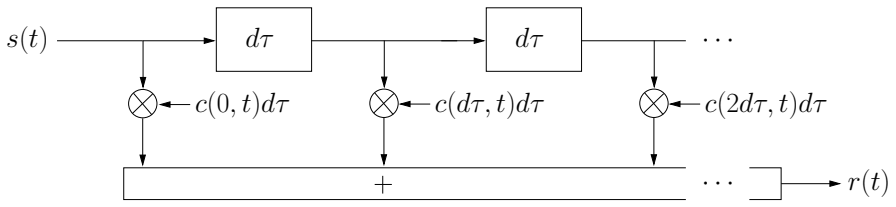


Figure 2.3: Representation of a linear time-variant channel with an input-delayline structure [Hub96, p. 4], [Höh11, p. 445].

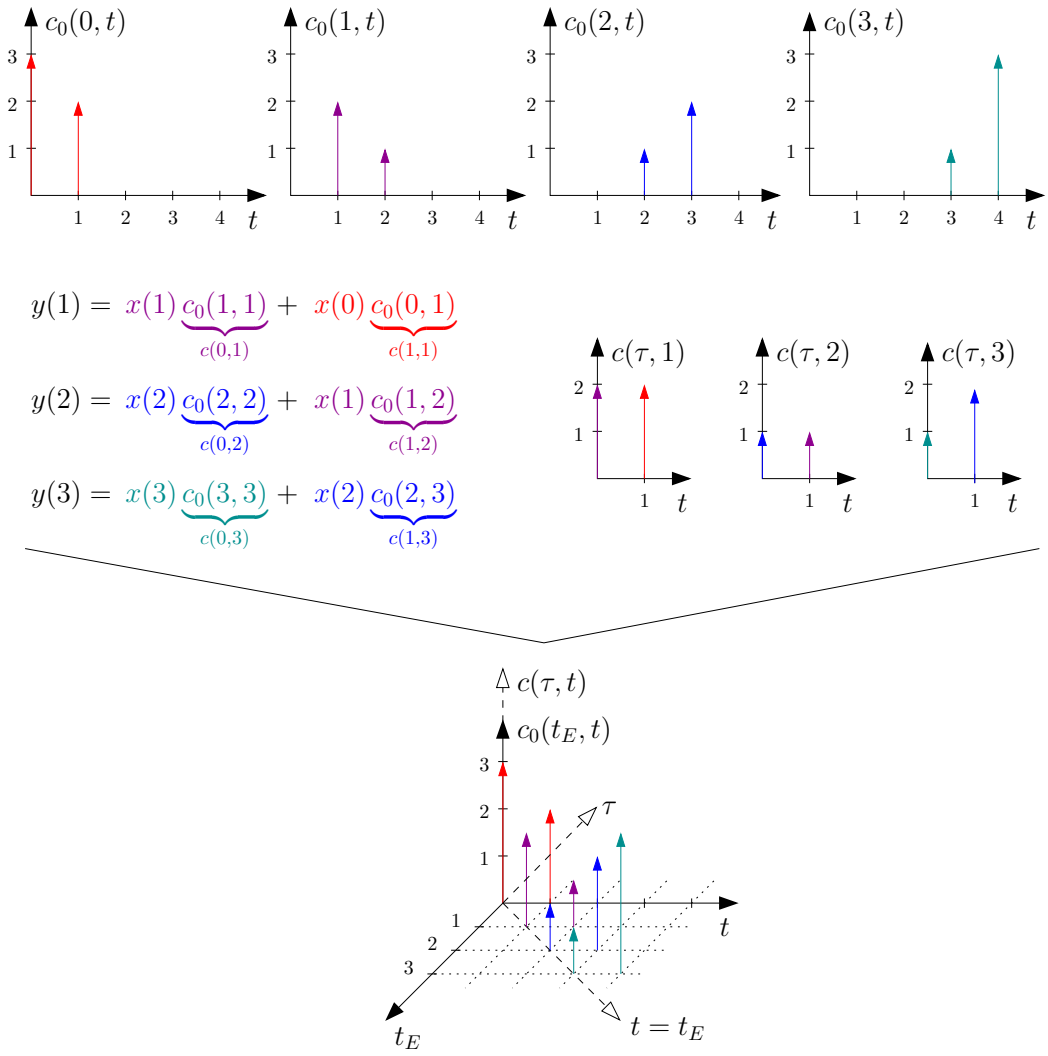


Figure 2.4: Example for the relationship between the impulse response $c_0\,(t_E, t)$ and the weight function $c\,(\tau, t)$. In the three-dimensional plot, the solid coordinate system belongs to the impulse response $c_0\,(t_E, t)$ and the dashed coordinate system belongs to the weight function $c\,(\tau, t)$.

$t_E{=}t{-}\tau$ in (2.3) leads to

$$r(t) = \int_{-\infty}^{\infty} s(t-\tau) c_0\,(t-\tau, t)\,\mathrm{d}\,\tau. \tag{2.5}$$

Comparing (2.4) and (2.5), the following relationships between the weight function and the impulse response are obtained:

$$c\,(\tau, t) = c_0\,(t-\tau, t) \qquad \text{and} \qquad c_0\,(t_E, t) = c\,(t-t_E, t)\,. \tag{2.6}$$

It is important to notice that both functions are completely different and that one function can *not* be approximated by the other as illustrated in Figure 2.4 for a simplified example.

Only in case of time-invariance, both functions are the same. For the remainder of this thesis, the weight function $c(\tau, t)$ and its Fourier transforms

$$C(f,t) := \mathcal{F}_\tau \{c(\tau,t)\} = \int_{-\infty}^{\infty} c(\tau,t)\, \mathrm{e}^{-\mathrm{j}\,2\pi f \tau}\, \mathrm{d}\,\tau \tag{2.7}$$

and

$$S(\tau, f_D) := \mathcal{F}_t \{c(\tau,t)\} = \int_{-\infty}^{\infty} c(\tau,t)\, \mathrm{e}^{-\mathrm{j}\,2\pi f_D t}\, \mathrm{d}\,t \tag{2.8}$$

called *time-variant transfer function* and *delay-Doppler function*, respectively, are used to describe the time-variant mobile radio channel [Hub96].

2.2 Large-Scale Channel Modelling

As already mentioned in the introduction of this chapter, large-scale channel models are typically used to predict the average received power at a certain distance d between transmitter and receiver. Hence, large-scale channel modelling is necessary for network planning in the context of communication systems. Furthermore, large-scale models can be applied to roughly estimate the distance d between transmitter and receiver from the received signal strength (RSS) in the context of positioning (see Section 3.1). According to *path loss* models, the received signal strength at different locations with the same distance d is equal. In order to account for variations in the topology/morphology of the surroundings at different locations, *shadowing* is introduced. Based on [Rap96, Ch. 3], [PL05, Ch. 4], [Gol05, Ch. 2] and [Höh11, Ch. 21], path loss models are introduced in the following, which are afterwards extended to include shadowing.

In general, the received signal strength decreases with increasing distance between transmitter and receiver due to path loss (see dotted line in Figure 2.1). The simplest scenario to describe the path loss is free-space propagation, i.e. the signal propagates only on the *line-of-sight (LOS)* path from transmitter to receiver since there are no obstacles in the surroundings, that may cause reflection, diffraction or scattering. For an isotropic radiator, the transmitted power P_T is spread uniformly over the surface of a sphere. Thus, the received power P_R depends on the distance d between transmitter and receiver. In the far-field ($d \gg \lambda$), the received power is given by

$$P_R = P_T \left(\frac{\lambda}{4\pi d}\right)^2, \tag{2.9}$$

where λ is the wavelength of the signal. The ratio of transmitted power to received power

$$L_{\mathrm{FS}} = \frac{P_T}{P_R} = \left(\frac{4\pi d}{\lambda}\right)^2 \tag{2.10}$$

is called *free-space path loss*. If directional antennas with antenna gains G_T and G_R are employed at the transmitter and the receiver side, respectively, the received power is given by the so-called *Friis transmission formula*:

$$P_R = P_T G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2 . \tag{2.11}$$

The assumption of free-space propagation is valid for example for satellite communications, but typically not for terrestrial communications. Measurements have shown that the path loss can be generally modelled by

$$L = \frac{P_T}{P_R} = \frac{1}{G_T G_R}\left(\frac{4\pi d}{\lambda}\right)^p , \tag{2.12}$$

in which the exponent p depends on the morphology and topology of the surroundings. As mentioned above, similar propagation environments are summarised in specific scenarios. The values of p range from approximately 1.5 to 6.5 [Rap96, p. 104], [PL05, p. 97], [Gol05, p. 47], but typically p lies in the interval between 2 and 4 [Höh11, p. 446]. The path loss of (2.12) is often expressed in dB, which leads to

$$L[\text{dB}] = 10\,\log\left(L\right) = \underbrace{-10\,\log\left(G_T G_R \left(\frac{\lambda}{4\pi}\right)^p\right)}_{L_0} + 10\,p\,\log\left(d\right), \tag{2.13}$$

where L_0 represents a reference path loss at a distance of 1 m. Plotting the average received power in dB over the distance d in logarithmic scale according to $P_R[\text{dB}]{=}P_T[\text{dB}]{-}L[\text{dB}]$ leads to a straight line with negative slope as illustrated in Figure 2.1 (dotted line). Due to changes in the morphology/topology of the surroundings, the average received power is not equal at different locations with the same distance d. In order to predict the average received power more realistically, shadowing is introduced, which is commonly modelled by *log-normal shadowing*: Equation (2.13) can be simply extended by introducing a third summand S_σ, which is a zero mean Gaussian distributed random variable with variance σ^2:

$$L[\text{dB}] = L_0 + 10\,p\,\log\left(d\right) + S_\sigma . \tag{2.14}$$

Since S_σ is normally distributed in the logarithmic domain, it is log-normally distributed in the linear domain. If shadowing is considered, a behaviour similar to the dashed line in Figure 2.1 is obtained.

2.3 Small-Scale Channel Modelling

Due to the phenomena of multipath propagation and Doppler shift, the received signal strength fluctuates around its mean value as shown by the solid line in Figure 2.1. Thus, accurate small-scale channel models are required to assess and compare the performance

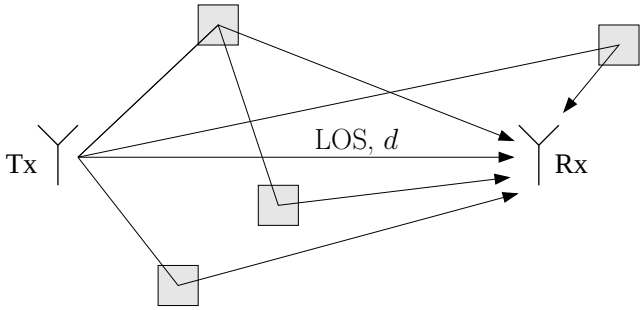


Figure 2.5: Schematic illustration of multipath propagation from transmitter (Tx) to receiver (Rx) [PL05, p. 56], [Gol05, p. 33], [Höh11, p. 438]. The LOS path corresponds to the distance d between transmitter and receiver.

of different receiver structures. Based on [Hub96], [Rap96, Ch. 4], [PL05, Ch. 3 - 6], [Gol05, Ch. 3] and [Höh11, Ch. 21], small-scale channel modelling is introduced in this section.

As mentioned above, free-space propagation can rarely be assumed because there are typically obstacles in the surroundings of the transmitter and the receiver, that cause reflection, diffraction or scattering. Hence, the transmitted signal propagates to the receiver on several paths in addition to the LOS path as illustrated schematically in Figure 2.5. Generally, the first arriving path is the LOS path as it is the shortest path corresponding to the distance d between transmitter and receiver. It might happen that the LOS path is blocked. This scenario is called *non LOS (NLOS)*. In case of communication systems, it is not relevant if the LOS path exists or not as long as the received signal strength is above a certain threshold. In the context of positioning based on the time of arrival or the angle of arrival, NLOS propagation causes significant positioning errors since the first arriving multipath does not correspond to the distance d between transmitter and receiver. The time-variant weight function for multipath propagation over N_{MP}[1] paths can be expressed as follows:

$$c(\tau, t) = \sum_{\mu=1}^{N_{\mathrm{MP}}} f_\mu(t)\, \delta\left(\tau - \tau_\mu(t)\right), \tag{2.15}$$

where $f_\mu(t)=\alpha_\mu(t)\ \mathrm{e}^{\mathrm{j}\,\phi_\mu(t)}$ and $\tau_\mu(t)$ are the complex amplitude and the propagation delay of the μth path, respectively. Replicas of the transmitted signal, that arrive on paths with the same[2] / a similar delay $\tau_i(t)\approx\tau_j(t)$, $i\neq j$, interfere with each other. The complex amplitudes of these *non-resolvable paths* add vectorially, which results in one path whose amplitude fluctuates rapidly. If two paths are resolvable or not, depends on the bandwidth of the transmitted signal. This is related to the classification of small-scale models as explained in Section 2.3.2 in more detail (flat fading vs. frequency-selective fading). If the

[1] The number of multipath components N_{MP} may also change with time.

[2] If only single reflections are considered, reflectors that lie on an ellipse (with transmitter and receiver in its foci) cause paths with the same delay.

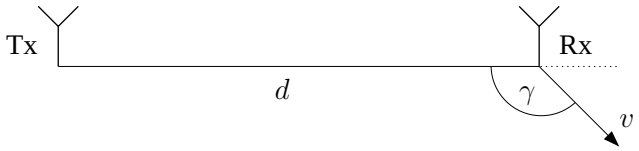


Figure 2.6: Example of Doppler shift when the transmitter (Tx) is static and the receiver (Rx) is moving [Höh11, p. 440].

transmitter, the receiver and the obstacles in their surroundings are static, the multipath structure does not change and as a result the time dependency can be omitted in (2.15). In case of the mobile radio channel, at least the transmitter or the receiver is moving. Typically, the obstacles in the surroundings of transmitter and receiver do also move (e.g. cars in urban areas). Due to these movements, the Doppler effect occurs. In Figure 2.6, a simple example is shown: The transmitter (Tx) is static and the receiver (Rx) is moving with velocity v in the direction indicated by γ. This results in a Doppler shift given by

$$f_D = \frac{v}{\lambda} \cos \gamma, \tag{2.16}$$

where λ is the wavelength of the signal and $v \cos \gamma$ is the relative velocity between transmitter and receiver.

Deterministic modelling of the mobile radio channel is usually not feasible. For very simple settings of moderate size whose geometry is known (e.g. indoor scenarios), *ray-tracing* can be applied in order to obtain deterministic channel models. But even for these simple settings, ray-tracing is very complex from a computational point of view. Hence, statistical channel models are commonly used. The computational complexity is generally reduced in comparison to ray-tracing methods, but the relationship between the time-variant weight function and the recent location is lost. This does not matter for the evaluation of communication systems, but in the context of positioning a relation between the time-variant weight function and the recent location is desirable. However, this thesis concentrates on statistical channel models.
Most of the literature concerning statistical channel modelling refers to [Bel63], which gives a comprehensive overview on randomly time-variant channels. There is a confusing number of interrelated functions, that can be used to characterise a randomly time-variant channel. Unfortunately, there is no unique nomenclature/naming convention in the literature. Thus, one can easily lose track when reading about statistical channel modelling. The nomenclature/naming convention utilised in this thesis is mainly based on [Hub96].

If a channel varies randomly, the time-variant weight function $c(\tau, t)$ and its Fourier transforms $C(f, t)$ and $S(\tau, f_D)$ (with respect to τ and t, respectively) are random processes. Thus, a description via autocorrelation functions is commonly applied:

$$\varphi_{cc}(\tau_1, \tau_2, t_1, t_2) = \mathrm{E}\left\{c(\tau_1, t_1)\, c^*(\tau_2, t_2)\right\}, \tag{2.17}$$

$$\varphi_{CC}(f_1, f_2, t_1, t_2) = \mathrm{E}\left\{C(f_1, t_1)\, C^*(f_2, t_2)\right\}, \tag{2.18}$$

$$\varphi_{SS}(\tau_1, \tau_2, f_{D1}, f_{D2}) = \mathrm{E}\left\{S(\tau_1, f_{D1})\, S^*(\tau_2, f_{D2})\right\}. \tag{2.19}$$

Since there are many free parameters to characterise a randomly time-variant channel via (2.17) to (2.19), statistical channel models based on these autocorrelation functions are still not really manageable. In order to further simplify the modelling of the mobile radio channel, the assumptions of *wide-sense stationarity (WSS)* and *uncorrelated scattering (US)* are introduced.

2.3.1 Wide-Sense Stationary Uncorrelated Scattering

First, the assumptions of wide-sense stationarity and uncorrelated scattering and their influence on the autocorrelation functions (2.17) to (2.19) are explained separately. Then, based on the combination of both assumptions, the *power delay profile* and the *Doppler power spectrum* are introduced.

If the correlation properties of a channel do not change with time t, the channel is *wide-sense stationary (WSS)*. Hence, the autocorrelation function (2.17) only depends on the time difference $\Delta t{=}t_1{-}t_2$. With $t_2{=}t$ follows [Hub96]:

$$\varphi_{cc}(\tau_1,\tau_2,t+\Delta t,t) =: \Psi_{cc}(\tau_1,\tau_2,\Delta t), \qquad \forall t\in\mathbb{R}, \tag{2.20}$$

$$\varphi_{CC}(f_1,f_2,t+\Delta t,t) =: \Psi_{CC}(f_1,f_2,\Delta t), \qquad \forall t\in\mathbb{R}, \tag{2.21}$$

$$\begin{aligned}\varphi_{SS}(\tau_1,\tau_2,f_D+\Delta f_D,f_D) &=: \Psi_{SS}(\tau_1,\tau_2,f_D)\,\delta(\Delta f_D), \\ &= \mathcal{F}_{\Delta t}\left\{\Psi_{cc}(\tau_1,\tau_2,\Delta t)\right\}\,\delta(\Delta f_D).\end{aligned} \tag{2.22}$$

If multipath components with different propagation delays $\tau_i{\neq}\tau_j$, $i{\neq}j$, are caused by different scatterers, they can be assumed to be uncorrelated. For channels with *uncorrelated scattering (US)*, the autocorrelation function (2.17) is zero for all $\Delta\tau{=}\tau_1{-}\tau_2{\neq}0$. Thus, it follows with $\tau_2{=}\tau$ [Hub96]:

$$\varphi_{cc}(\tau+\Delta\tau,\tau,t_1,t_2) =: \Phi_{cc}(\tau,t_1,t_2)\,\delta(\Delta\tau), \tag{2.23}$$

$$\begin{aligned}\varphi_{CC}(f+\Delta f,f,t_1,t_2) &=: \Phi_{CC}(\Delta f,t_1,t_2), \qquad \forall f\in\mathbb{R}, \\ &= \mathcal{F}_{\tau}\left\{\Phi_{cc}(\tau,t_1,t_2)\right\},\end{aligned} \tag{2.24}$$

$$\varphi_{SS}(\tau+\Delta\tau,f_{D1},f_{D2}) =: \Phi_{SS}(\tau,f_{D1},f_{D2})\,\delta(\Delta\tau). \tag{2.25}$$

If both assumptions are combined, a *wide-sense stationary uncorrelated scattering (WSSUS)* channel is obtained. In this case, the autocorrelation functions (2.17) to (2.19) are simplified significantly [Hub96]:

$$\varphi_{cc}(\tau+\Delta\tau,\tau,t+\Delta t,t) =: \Theta_{cc}(\tau,\Delta t)\,\delta(\Delta\tau), \qquad \forall t\in\mathbb{R}, \tag{2.26}$$

$$\begin{aligned}\varphi_{CC}(f+\Delta f,f,t+\Delta t,t) &=: \Theta_{CC}(\Delta f,\Delta t), \qquad \forall t,f\in\mathbb{R}, \\ &= \mathcal{F}_{\tau}\left\{\Theta_{cc}(\tau,\Delta t)\right\},\end{aligned} \tag{2.27}$$

$$\begin{aligned}\varphi_{SS}(\tau+\Delta\tau,\tau,f_D+\Delta f_D,f_D) &=: \Theta_{SS}(\tau,f_D)\,\delta(\Delta\tau)\,\delta(\Delta f_D), \\ &= \mathcal{F}_{\Delta t}\left\{\Theta_{cc}(\tau,\Delta t)\right\}\,\delta(\Delta\tau)\,\delta(\Delta f_D).\end{aligned} \tag{2.28}$$

The functions $\Theta_{cc}(\tau,\Delta t)$ and $\Theta_{CC}(\Delta f,\Delta t)$ are called *delay-time-correlation function* and *time-frequency-correlation function*, respectively. Commonly, the *delay-Doppler power*

density spectrum $\Theta_{SS}(\tau, f_D)$, which is also known as *scattering function*, is used to characterise a randomly time-variant channel. If the scattering function is integrated over τ, the so-called *Doppler power spectrum*

$$D(f_D) = \int_{-\infty}^{\infty} \Theta_{SS}(\tau, f_D)\ \mathrm{d}\,\tau \tag{2.29}$$

is obtained. An integration over f_D results in the so-called *power delay profile*

$$M(\tau) = \int_{-\infty}^{\infty} \Theta_{SS}(\tau, f_D)\ \mathrm{d}\,f_D = \Theta_{cc}(\tau, 0), \tag{2.30}$$

which corresponds to the delay-time-correlation function at Δt=0. If the time and frequency components of the scattering function are independent, the scattering function can be factorised as

$$\Theta_{SS}(\tau, f_D) = M(\tau) \cdot D(f_D). \tag{2.31}$$

In this case, the channel is fully characterised by the power delay profile and the Doppler power spectrum, i.e. realisations of the weight function $c\,(\tau, t)$ can be generated if $M(\tau)$ and $D(f_D)$ are known. In which way realisations of the channel weight function can be generated, is explained in Section 2.3.3. Prior to that, the classifications of small-scale channel models are presented in the following section.

2.3.2 Classification of Small-Scale Channel Models

The small-scale behaviour of the mobile radio channel depends on the symbol duration T_s and the bandwidth $B_s \approx 1/T_s$ of the transmitted signal, respectively. Different types of small-scale channel models can be classified depending on the relationship between these signal parameters and parameters that characterise the channel such as the *delay spread* or the *Doppler spread*, which are defined below. On the one hand, the multipath characteristics (delay spread) of the channel influence the time dispersion of the signal, which leads to *flat* or *frequency-selective fading*. On the other hand, the frequency dispersion of the signal is affected by the Doppler characteristics (Doppler spread) of the channel, which leads to *slow* or *fast fading*. In the following, the classification of small-scale channel models is presented in more detail.

Flat Fading vs. Frequency-Selective Fading

The multipath characteristics of the mobile radio channel are described by the power delay profile $M(\tau)$ and can be summarised by the delay spread T_M. Unfortunately, there is no unique definition of the delay spread. Basically, it describes the range over which $M(\tau)$

has significant values, i.e. values which are above the noise level. In order to quantify the delay spread T_M, the mean

$$\mu_\tau = \frac{\int_{-\infty}^{\infty} \tau\, M(\tau)\, \mathrm{d}\,\tau}{\int_{-\infty}^{\infty} M(\tau)\, \mathrm{d}\,\tau} \tag{2.32}$$

and variance

$$\sigma_\tau^2 = \frac{\int_{-\infty}^{\infty} (\tau - \mu_\tau)^2\, M(\tau)\, \mathrm{d}\,\tau}{\int_{-\infty}^{\infty} M(\tau)\, \mathrm{d}\,\tau} \tag{2.33}$$

of the power delay profile can be determined. The delay spread is usually defined as a multiple of the standard deviation, e.g. $T_M{=}\sigma_\tau$ or $T_M{=}3\sigma_\tau$ [Gol05, p. 86]. Often, the first definition, $T_M{=}\sigma_\tau$, which is also called *root mean square (RMS)* delay spread, is applied. If the delay spread is much smaller than the symbol duration, $T_M{\ll}T_s$, the distortion introduced by the channel is non-dispersive, i.e. the shape of a symbol is preserved. This type of channel is called a *flat fading* channel. In the opposite case, $T_M \gtrapprox T_s$, replicas of the same symbol arrive with considerable delay at the receiver and overlap, i.e. the symbol shape is distorted and *inter-symbol interference (ISI)* occurs. This dispersive type of channel is called *frequency-selective*. Equivalently, the channel can be classified in the frequency domain: The bandwidth over which the mean frequency-correlation function $\Theta_{CC}(\Delta f, 0){=}\mathcal{F}_\tau\{\Theta_{cc}(\tau, 0)\}{=}\mathcal{F}_\tau\{M(\tau)\}$ does not change significantly is called *coherence bandwidth* B_M of the channel. The coherence bandwidth is approximately the inverse of the delay spread: $B_M{\approx}1/T_M$. A channel is frequency-selective if its coherence bandwidth is smaller than or approximately equal to the signal bandwidth, $B_M \lessapprox B_s$, which is usually the case for wideband signals. In the opposite case, $B_M{\gg}B_s$, the channel is flat, which is commonly true for narrowband signals.

This classification of flat and frequency-selective channels is related to the issue of resolvable and non-resolvable multipath components as mentioned in conjunction with (2.15). Two multipath components, which are separated by $\Delta\tau{=}\tau_1{-}\tau_2$, are resolvable if the absolute value of their separation is larger than or approximately equal to the symbol duration: $|\Delta\tau| \gtrapprox T_s$. This constraint is similar to the condition for frequency-selective fading. Thus, in case of flat fading, *all* multipath components are non-resolvable, i.e. all paths have a similar propagation delay $\tau_i{\approx}\tau_j$, $\forall i, j$, such that their amplitudes add vectorially. This results in *one* path whose amplitude fluctuates rapidly. In case of frequency-selective fading, the power delay profile consists of several resolvable paths, where each path usually represents a cluster of non-resolvable paths. Thus, flat fading models are not only required to describe flat channels, but also to model each path of a frequency-selective channel. *Rayleigh* and *Rician fading* are common flat fading models [Rap96, p. 172ff.], [PL05, p. 111ff.], [Gol05, p. 73ff.], [Höh11, p. 279f., 447ff.]. For Rayleigh fading, two-dimensional isotropic scattering without a dominant LOS path is assumed, i.e. the incident angles of all multipath components are uniformly distributed. These assumptions lead to the well known Jakes spectrum [Rap96, p. 181], [PL05, p. 113],

[Gol05, p. 75], [Höh11, p. 450]. If a LOS path is added to the Rayleigh fading model, Rician fading is obtained. The power ratio between the LOS component and the scattered components is given by the Rician factor K_R.

Fast Fading vs. Slow Fading

The Doppler characteristics of the mobile radio channel are described by the Doppler power spectrum $D(f_D)$ and can be summarised by the Doppler spread B_D. As in the case of the delay spread, there is no unique definition of the Doppler spread. Basically, it describes the range over which $D(f_D)$ has non-zero values. Comparable to (2.32) and (2.33), the mean

$$\mu_{f_D} = \frac{\int\limits_{-\infty}^{\infty} f_D\, D(f_D)\, \mathrm{d}\, f_D}{\int\limits_{-\infty}^{\infty} D(f_D)\, \mathrm{d}\, f_D} \tag{2.34}$$

and variance

$$\sigma_{f_D}^2 = \frac{\int\limits_{-\infty}^{\infty} (f_D - \mu_{f_D})^2\, D(f_D)\, \mathrm{d}\, f_D}{\int\limits_{-\infty}^{\infty} D(f_D)\, \mathrm{d}\, f_D} \tag{2.35}$$

of the Doppler power spectrum can be utilised to quantify the Doppler spread. Similar to the RMS delay spread, the RMS Doppler spread $B_D{=}\sigma_{f_D}$ is commonly applied. Alternatively, the Doppler spread can be simply equated to the maximum Doppler shift: $B_D{=}f_{D,\max}{=}v/\lambda$. If the Doppler spread is much smaller than the bandwidth of the transmitted signal, $B_D{\ll}B_s$, the channel varies slowly (*slow fading*). In the opposite case, $B_D \gtrapprox B_s$, *fast fading* occurs. For a classification in the time domain, the *coherence time* T_D of the channel has to be compared to the symbol duration T_s. The coherence time is approximately the inverse of the Doppler spread, $T_D{\approx}1/B_D$, and corresponds to the time duration over which the mean time-correlation function $\Theta_{CC}(0,\Delta t){=}\int \Theta_{cc}(\tau,\Delta t)\, \mathrm{d}\,\tau$ does not change significantly. The channel changes fast if the coherence time is smaller than or approximately equal to the symbol duration: $T_D \lessapprox T_s$. In the opposite case, $T_D{\gg}T_s$, slow fading occurs.

In the context of slow and fast fading, the so-called *normalised Doppler spread* $B_D{\cdot}T_s$ is often applied in order to classify the channel. Since the symbol duration is approximately the inverse of the signal bandwidth, slow fading is obtained for $B_D{\cdot}T_s{\ll}1$, whereas the channel changes fast for $B_D{\cdot}T_s \gtrapprox 1$. This classification declares if the channel changes over the symbol duration or not. For communications systems, data is typically transmitted/processed burst-wise. Hence, it is rather important to know if the channel changes over the burst duration or not. If K' denotes the number of symbols per burst, the channel changes slowly over the burst if $B_D{\cdot}K'T_s{\ll}1$. For $B_D{\cdot}K'T_s \gtrapprox 1$, the channel changes fast over the burst. This aspect is especially interesting in the context of channel estimation: In order to detect the transmitted data, the channel needs to be known at the receiver

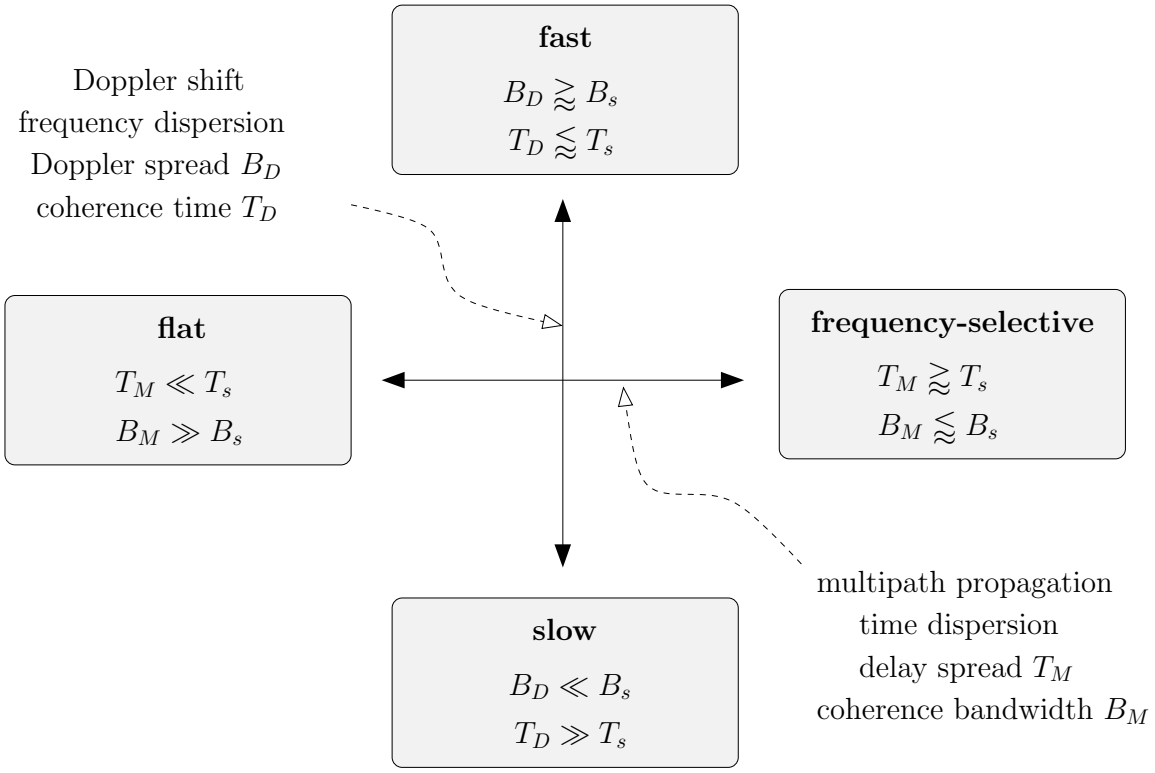


Figure 2.7: Classification of the mobile radio channel.

side. Typically, training symbols are inserted into the burst in order to estimate the channel. On the one hand, it is important to know whether the channel changes over the length of the training sequence. On the other hand, it is important to know whether the channel changes after it has been estimated. Least-squares channel estimation (LSCE) is a famous method to estimate the channel, that assumes *block fading*, which means that the channel is constant over the training sequence. LSCE is for example popular in GSM, where the training sequence is placed in the middle of the burst (midamble) in order to minimise the error between the true and the estimated channel at the edges of the burst [Kam04, p. 551-556], [Höh11, p. 396f].

The classification of the mobile radio channel is summarised in Figure 2.7. In this thesis, slow frequency-selective channels are considered. It should be mentioned here that the above classifications are also valid for deterministic channel modelling. In this case, the delay and Doppler spread are not determined from the power delay profile and the Doppler power spectrum, but from the channel weight function and its Fourier transforms.

2.3.3 Realisations of the Channel Weight Function

As mentioned before, realisations of the channel weight function $c(\tau, t)$ can be generated if the power delay profile $M(\tau)$ and the Doppler power spectrum $D(f_D)$ are known. In general, the power delay profile $M(\tau)$ is continuous and can be of any form, but typically it is specified as a weighted sum of delayed Dirac impulses:

$$M(\tau) = \sum_{c=1}^{C} P_c \, \delta(\tau - \tau_c), \tag{2.36}$$

where P_c is the average power of the cth path and τ_c is the propagation delay of the cth path. Often, each path in the power delay profile represents a cluster of several non-resolvable paths, that add vectorially. Thus, each path in the power delay profile can be modelled via flat fading, which is determined by the Doppler power spectrum $D(f_D)$. In order to distinguish between resolvable and non-resolvable paths, the naming convention of [KMH+08] is applied in the following: resolvable paths in the power delay profile are called *clusters* and non-resolvable paths, that interfere with each other, are called *rays*. A realisation of the channel weight function is obtained according to

$$c(\tau, t) = \sum_{c=1}^{C} \underbrace{\left(A_c \cdot \frac{1}{\sqrt{R_c}} \sum_{r=1}^{R_c} \exp\left(\mathrm{j}\left(2\pi f_{D,c,r} t + \Phi_{c,r}\right)\right) \right)}_{f_c(t)} \delta(\tau - \tau_c), \tag{2.37}$$

where

$$A_c = \sqrt{P_c \Big/ \sum_{c=1}^{C} P_c} \tag{2.38}$$

is the normalised amplitude of the corresponding cluster and R_c is the number of rays that contribute to cluster c. Each ray has a specific Doppler shift $f_{D,c,r}$, which is determined according to the Doppler power spectrum $D(f_D)$, and a randomly generated starting phase $\Phi_{c,r}$ in the interval $[0,2\pi[$. In case of Rayleigh fading, the number of rays per cluster goes to infinity, $R_c \to \infty$, and the Doppler shifts are given by $f_{D,c,r} = f_{D,\max} \cdot \cos\gamma_{c,r} = v/\lambda \cdot \cos\gamma_{c,r}$, where $\gamma_{c,r}$ is a randomly generated angle in the interval $[0,2\pi[$. For simulations on a computer, a finite number of rays has to be applied. According to the central limit theorem, reasonable values for R_c lie between 10 and 20.

2.4 Equivalent Discrete-Time Channel Model

Assuming a WSSUS channel with independent time and frequency components of the scattering function, realisations of the channel weight function $c(\tau, t)$ can be generated as explained above. In order to simulate a transmission link on a computer, a discrete-time representation of the continuous-time channel is necessary. For this purpose, all

continuous-time elements of the transmission link are joined in the so-called *equivalent discrete-time channel model*. Thus, the equivalent discrete-time channel model includes the pulse shaping filter $g_{\mathrm{Tx}}(\tau)$, the physical channel $c(\tau, t)$, additive white Gaussian noise (AWGN), the receive filter $g_{\mathrm{Rx}}(\tau)$ as well as sampling. The input/output behaviour of the equivalent discrete-time channel model equals *exactly* the input/output behaviour of the continuous-time channel as shown in Figure 2.8. The delay elements z^{-1} of the equivalent discrete-time channel model correspond to the sampling period $\mathcal{T}$, which is typically given as a fraction of the symbol duration $\mathcal{T}=T_s/J$, where J is called *oversampling factor*. For *symbol-rate sampling*, the oversampling factor equals to one ($J=1$) and the sampling period corresponds to the symbol duration ($\mathcal{T}=T_s$). The factor ε accounts for sampling time offsets and is called *sampling phase*. The channel coefficients $h_l[k]$ are samples of the overall (continuous-time) channel weight function $h(\tau, t)$, which is given by the convolution of the pulse shaping filter $g_{\mathrm{Tx}}(\tau)$, the physical channel $c(\tau, t)$ and the receive filter $g_{\mathrm{Rx}}(\tau)$:

$$h(\tau, t) = g_{\mathrm{Tx}}(\tau) * c(\tau, t) * g_{\mathrm{Rx}}(\tau) = c(\tau, t) * \underbrace{g_{\mathrm{Tx}}(\tau) * g_{\mathrm{Rx}}(\tau)}_{g(\tau)}. \tag{2.39}$$

If the weight function of (2.37) is considered, the channel coefficients are given by

$$h_l[k] = \sum_{c=1}^{C} f_c[k] \cdot g(l\mathcal{T} + \varepsilon - \tau_c) \tag{2.40}$$

with

$$f_c[k] = A_c \cdot \frac{1}{\sqrt{R_c}} \sum_{r=1}^{R_c} \exp\left(\mathrm{j}\left(2\pi f_{D,c,r}\left(k\mathcal{T} + \varepsilon\right) + \Phi_{c,r}\right)\right). \tag{2.41}$$

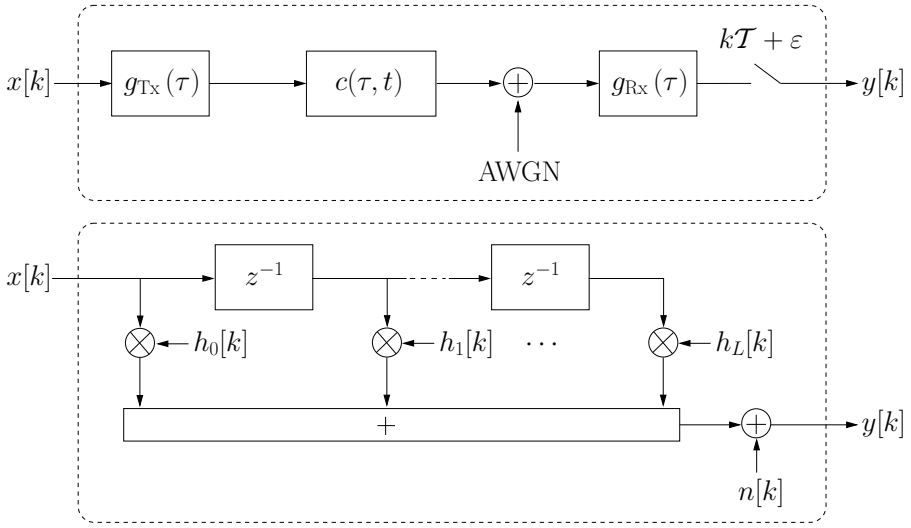


Figure 2.8: Continuous-time channel and equivalent discrete-time channel model.

Assuming a causal weight function $h(\tau,t)$ of finite length, the channel coefficients $h_l[k]$ have zero values for $l<0$ and $l>L$, where L is called *channel memory length*.

As mentioned earlier, burst-wise data transmission is commonly applied for communications. Let $x'[\kappa]$, $0\leq\kappa<K'$, denote the κth symbol of a coded and modulated burst of length K'. If oversampling is applied, this sequence is upsampled to a burst of length of $K=JK'$, whose kth symbol is given according to

$$x[k] = \begin{cases} x'[k/J] & \text{if } k \bmod J = 0, \\ 0 & \text{else.} \end{cases} \tag{2.42}$$

In case of symbol-rate sampling, both sequences are the same: $x[k]=x'[\kappa]$ with $K=K'$. The received sample $y[k]$ at time index k is given by

$$y[k] = \sum_{l=0}^{L} h_l[k] \cdot x[k-l] + n[k], \qquad 0 \leq k < K+L. \tag{2.43}$$

The noise process $n[k]$ is generally coloured because the additive white Gaussian noise with zero mean and variance σ_n^2 is filtered by $g_{\mathrm{Rx}}(\tau)$ in the continuous-time channel (see Figure 2.8). Thus, the sampled autocorrelation function of $n[k]$ is given by

$$\varphi_{nn}[\Delta k] = \sigma_n^2 \cdot \psi_{\mathrm{Rx}}(\Delta k \mathcal{T}) \tag{2.44}$$

with $\Delta k=k_1-k_2$ and where $\psi_{\mathrm{Rx}}(\tau)=g_{\mathrm{Rx}}(\tau)*g_{\mathrm{Rx}}(-\tau)$ denotes the autocorrelation function of the receive filter. If a square-root Nyquist pulse is applied at the receiver, the noise remains white for symbol-rate sampling.

In general, the channel memory length is $L=0$ for flat fading channels. The distortion introduced by the channel is solely multiplicative: $y[k]=h_0[k]\cdot x[k]+n[k]$. For $L>0$, intersymbol interference (ISI) occurs. It is important to note that ISI may also occur for a flat fading physical channel since the channel coefficients comprise pulse shaping and receive filtering in addition to the physical channel. ISI can not be avoided by any means if the pulse shape of $g(\tau)$ extends over several symbol durations T_s. Thus, the channel memory length L is influenced by $c(\tau,t)$ as well as by $g(\tau)$. If T_g denotes the effective width of $g(\tau)$, the channel memory length L is given by[3]

$$L = J \underbrace{\left\lceil \frac{\tau_C + T_g}{T_s} \right\rceil}_{L'}, \tag{2.45}$$

where τ_C is the maximum propagation delay. Since the propagation delay of the first path τ_1 is usually greater than zero, the leading channel coefficients for $l<J(\tau_1-\varepsilon)/T_s$ are

[3]Actually, a smaller memory length could be obtained in some cases if the following formula $L=\lceil(\tau_C+T_g)/\mathcal{T}\rceil=\lceil J(\tau_C+T_g)/T_s\rceil$ would be applied instead of (2.45). The advantage of (2.45) lies in the fact, that the channel memory length for oversampling is a multiple of the channel memory length for symbol-rate sampling L'. Thus, up- and down-sampling as applied in Chapter 6 are facilitated.

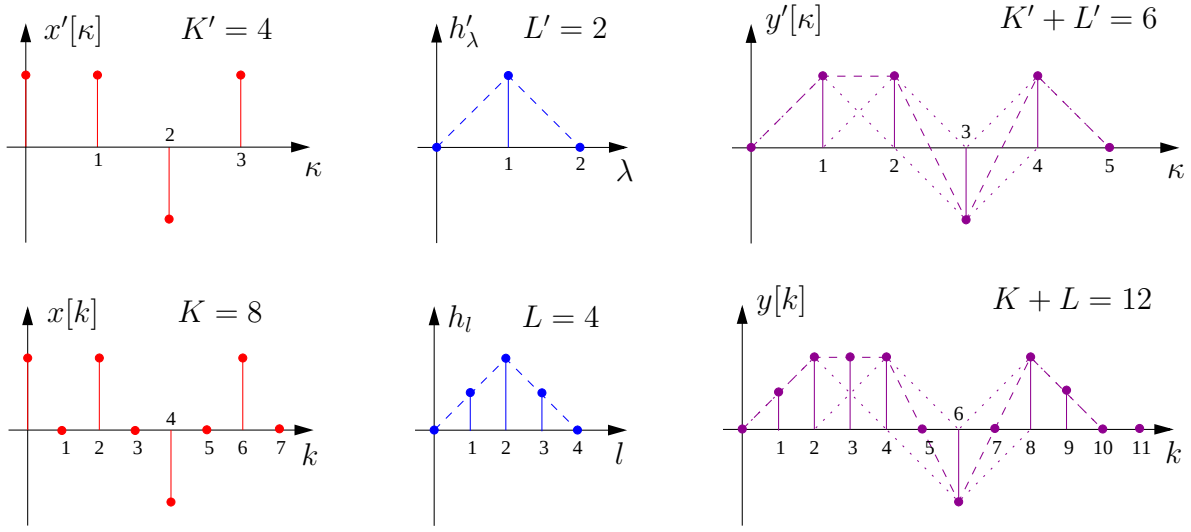


Figure 2.9: A simple real-valued example for oversampling with J=2.

zero and can be neglected. Thus, the delay of the first path τ_1 is typically eliminated and only *excess delays* ν_c=τ_c−τ_1 are considered, which results in a normalisation of the power delay profile in (2.36):

$$M(\nu) = \sum_{c=1}^{C} P_c\, \delta(\nu - \nu_c) = M(\tau - \tau_1). \tag{2.46}$$

This normalisation does not influence the classification of the channel: As the calculation of the variance in (2.33) is invariant with respect to delay shifts, the delay spread of the normalised power delay profile $M(\nu)$ is equal to the delay spread of the original power delay profile $M(\tau)$. Now, the channel coefficients are samples of the overall normalised channel weight function $h\,(\nu, t)$ =$h\,(\tau$−$\tau_1, t)$ =$c\,(\tau$−$\tau_1, t) * g\,(\tau)$ and equation (2.40) is changed to

$$h_l[k] = \sum_{c=1}^{C} f_c[k] \cdot g\,(l\mathcal{T} + \varepsilon - \nu_c)\,, \tag{2.47}$$

in which $f_c[k]$ is still determined by (2.41). The leading channel coefficients of zero value are eliminated and the channel memory length reduces to

$$L = J \underbrace{\left\lceil \frac{\nu_C + T_g}{T_s} \right\rceil}_{L'}, \tag{2.48}$$

where ν_C is the maximum excess delay. Figure 2.9 illustrates the difference between symbol-rate sampling and oversampling for a (simple) real-valued example. The upper row shows the transmit sequence, the channel coefficients and the received sequence for symbol-rate sampling, whereas the same quantities for oversampling with J=2 are depicted in the lower row. The sequences for symbol-rate sampling can be obtained from

the oversampled sequences by taking only every Jth sample into account (down-sampling).

2.5 Specific Channel Models Applied in this Thesis

The channel models, that are applied in this thesis in order to evaluate the performance of the proposed joint communication and positioning system, are presented in this section. For the evaluation of communication systems, it is sufficient to simulate a single link between transmitter and receiver. Furthermore, perfect synchronisation is typically assumed, i.e. the sampling phase ε is zero, which implies that the propagation delay of the first arriving path τ_1 is perfectly known. These assumptions are not applicable in this thesis since the positioning part of the proposed joint communication and positioning system is based on the *time of arrival (TOA)* (see Chapter 3 and references therein). In order to understand, why these assumptions are not applicable, localisation based on the TOA is shortly explained at this point: The location of a mobile station (MS) is determined relative to so-called reference objects (ROs) whose positions are known. In a first step, the propagation delay of the first arriving path (TOA) is estimated for each link between MS and a RO (*parameter estimation*). In a second step, the position of the MS is determined by combing the estimated parameters of all links and by exploiting the spatial relationship between the MS and ROs (*position estimation*). For parameter estimation, the propagation delay of the first arriving path needs to be estimated and can not be assumed to be known. Thus, the assumption of perfect synchronisation is not applicable. Therefore, a coarse synchronisation, $\check{\tau}_1{=}\tau_1{+}\varepsilon$, that eliminates the propagation delay of the first arriving path only approximately, is considered in the following. This leads to a non-zero sampling phase ε. The synchronisation accuracy is given by the range of the the sampling phase $\varepsilon{\in}\,[-\Delta\tau_1, +\Delta\tau_1]$. Hence, the estimation of the TOA basically corresponds to the estimation of the sampling phase $\hat{\varepsilon}$ in this thesis (see Chapter 4): $\hat{\tau}_1{=}\check{\tau}_1{-}\hat{\varepsilon}{=}\tau_1{+}\varepsilon{-}\hat{\varepsilon}$. Furthermore, several links that are geometrically related need to be simulated in order to assess the positioning accuracy. As mentioned earlier, the relationship between the channel weight function and the recent location is lost for statistical channel modelling. Since statistical channel models are applied in this thesis, a spatial relationship has to be re-introduced. For this purpose, scenarios with a certain geometry are introduced in Chapter 5. The positions of the ROs are fixed and the location of the MS is generated randomly within a given area. The propagation delay of the first arriving path τ_1 is calculated according to the given geometry for each link and the sampling phase ε is generated randomly in the interval $[-\Delta\tau_1, +\Delta\tau_1]$. Another aspect that has to be taken into account is the existence of a LOS path: For communication systems, it does not matter if a LOS path exists as long as the received signal strength is above a certain threshold. For localisation based on the TOA, it is assumed that the propagation delay of the first arriving path τ_1 corresponds to the distance d between transmitter and receiver. This is only true if a LOS path exists. In urban or indoor environments, the LOS path is often blocked. In such NLOS scenarios, the modelling error reduces the positioning accuracy significantly. Additionally, positioning based on the TOA typically suffers from a

bias introduced by multipath propagation *even* if a LOS path exists. This thesis focusses on the multipath mitigation ability of the proposed joint communication and positioning system. Hence, it is assumed in the following that a LOS path exists.

Basically, there are three abstraction levels for the channel models: At first, pure LOS transmission is considered, which is quite unlikely to occur in terrestrial radio systems. Nevertheless, the *LOS channel model* is taken into account since it is the most favourable scenario and, thus, provides a lower bound for all other channel models. The second abstraction level comprises simple *two-path channel models*. Based on these channel models, the influence of changing channel parameters such as the separation of the multipath components (excess delay) or the power ratio of both components is studied. At last, the *wireless world initiative new radio (WINNER) channel models* reported in [KMH+08] are applied to evaluate the system performance in a realistic scenario.

Before the specific channel models are presented, the equations that are required for their description are summarised here: All channel models are based on the normalised power delay profile in (2.46),

$$M(\nu) = \sum_{c=1}^{C} P_c \, \delta(\nu - \nu_c),$$

which leads to the channel coefficients in (2.47),

$$h_l[k] = \sum_{c=1}^{C} f_c[k] \cdot g\left(l\mathcal{T} + \varepsilon - \nu_c\right),$$

where $f_c[k]$ is determined via (2.41):

$$f_c[k] = A_c \cdot \frac{1}{\sqrt{R_c}} \sum_{r=1}^{R_c} \exp\left(\mathrm{j}\left(2\pi f_{D,c,r}\left(k\mathcal{T} + \varepsilon\right) + \Phi_{c,r}\right)\right).$$

The normalised amplitudes given in (2.38) are calculated as

$$A_c = \sqrt{P_c \Big/ \sum_{c=1}^{C} P_c},$$

the Doppler shifts $f_{D,c,r}$ are determined according to the Doppler power spectrum $D(f_D)$ and the starting phases $\Phi_{c,r}$ are generated randomly between 0 and 2π. In case of block fading ($B_D{\cdot}K'T_s \ll 1$), the fading factor $f_c[k]$ becomes time-invariant and simplifies to

$$f_c \approx A_c \cdot \frac{1}{\sqrt{R_c}} \sum_{r=1}^{R_c} \exp\left(\mathrm{j}\,\Phi_{c,r}\right).$$

The important parameters that describe a specific channel model are, therefore, the number of clusters C, the average power of each cluster P_c, the excess delay of each cluster ν_c, the number of rays per cluster R_c and the way in which the Doppler shifts $f_{D,c,r}$ are generated. The average power of each cluster P_c is typically given in dB, where the largest average power is normalised to 0 dB. The excess delay of the first arriving path is $\nu_1{=}0$ by definition.

2.5.1 LOS Channel Model

The simplest channel model is the LOS channel model, whose normalised power delay profile is given by a single cluster (C=1, P_1[dB]=0),

$$M(\nu) = \delta(\nu), \tag{2.49}$$

with a single ray (R_1=1), which leads to the following channel coefficients:

$$h_l[k] = \exp\left(\mathrm{j}\left(2\pi f_D\left(k\mathcal{T}+\varepsilon\right)+\Phi\right)\right)\, g\left(l\mathcal{T}+\varepsilon\right). \tag{2.50}$$

The Doppler shift is given by (2.16), f_D=$v/\lambda\cdot\cos\gamma$, where γ is generated randomly between 0 and 2π for each run of a Monte Carlo simulation. Thus, all Doppler shifts between $\pm v/\lambda$ are possible. Since the cluster consists of a single ray, the amplitude of that cluster does not fade, only the phase is changing with time. For block fading, the channel coefficients become time-invariant and simplify to

$$h_l \approx \exp\left(\mathrm{j}\,\Phi\right)\, g\left(l\mathcal{T}+\varepsilon\right). \tag{2.51}$$

2.5.2 Two-Path Channel Models

The two-path channel model consists of two clusters (C=2):

$$M(\nu) = P_1\,\delta(\nu) + P_2\delta(\nu-\nu_2), \tag{2.52}$$

where the excess delay ν_2 and the power ratio $\mathcal{P}$=P_1/P_2 are specified for each simulation. For $\mathcal{P}$>1, a minimum phase channel is obtained (P_1[dB]=0, P_2[dB]= $-\mathcal{P}$[dB]) and $\mathcal{P}\leq$1 results in a maximum phase channel (P_2[dB]=0, P_1[dB]= $-\mathcal{P}$[dB]). Each cluster consists of a single ray (R_1=R_2=1):

$$\begin{aligned} h_l[k] = {} & A_1 \exp\left(\mathrm{j}\left(2\pi f_{D,1}\left(k\mathcal{T}+\varepsilon\right)+\Phi_1\right)\right)\, g\left(l\mathcal{T}+\varepsilon\right) \\ & + A_2 \exp\left(\mathrm{j}\left(2\pi f_{D,2}\left(k\mathcal{T}+\varepsilon\right)+\Phi_2\right)\right)\, g\left(l\mathcal{T}+\varepsilon-\nu_2\right), \end{aligned} \tag{2.53}$$

in which the Doppler shifts are generated similar to the LOS channel: $f_{D,c}$=$v/\lambda\cdot\cos\gamma_c$. The angles γ_c are generated independently between 0 and 2π for each run of a Monte Carlo simulation. Thus, the clusters do not fade and their power ratio always equals $\mathcal{P}$. For block fading, the channel coefficients become time-invariant and simplify to

$$h_l \approx A_1 \exp\left(\mathrm{j}\,\Phi_1\right)\, g\left(l\mathcal{T}+\varepsilon\right) + A_2 \exp\left(\mathrm{j}\,\Phi_2\right)\, g\left(l\mathcal{T}+\varepsilon-\nu_2\right). \tag{2.54}$$

2.5.3 WINNER Channel Models

In order to evaluate the performance of the proposed joint communication and positioning system in a realistic environment, the WINNER channel models presented in [KMH+08] are employed. Many different propagation scenarios are considered in [KMH+08] including rural, suburban and urban as well as indoor scenarios. There are two types of channel models: The *generic* model is suited for system level simulations, whereas the *clustered delay line (CDL)* model is reduced in complexity for fast link level simulations. The parameters of the generic model "are determined stochastically, based on statistical distributions extracted from channel measurement" [KMH+08, p. 2]. In contrast, the parameters of the CDL models, which "are based on expectation values of the generic models" [KMH+08, p. 2], are fixed and tabulated. Due to complexity reasons, only the CDL models are applied in this thesis. Furthermore, the WINNER channel models are suited for the evaluation of multiple-input multiple-output (MIMO) systems. However, only the case of single-input single-output (SISO) is considered here.

All channel models are applicable for the frequency range from 2 GHz to 6 GHz. For the CDL models, the normalised power delay profiles $M(\nu)$ are given according to (2.46) and consist of C=8 to C=20 clusters. Each cluster is composed of R_c=20 rays. For most scenarios, a LOS as well as a NLOS channel model are available. If a LOS path exists, the first cluster consists of R_1=21 rays. Since the WINNER channel models are suited for MIMO systems, not only the excess delay ν_c and average power P_c of each cluster are tabulated, but also the angle of departure (AOD) and the angle of arrival (AOA). Figure 2.10 shows how a single link is modelled in [KMH+08]: Each cluster is assigned a AOD α_c and a AOA β_c. Given the azimuth spread of departure angles (ASD) $\Delta\alpha$ and the azimuth spread of arrival angles (ASA) $\Delta\beta$, the AOD $\alpha_{c,r}$ and AOA $\beta_{c,r}$ of each ray

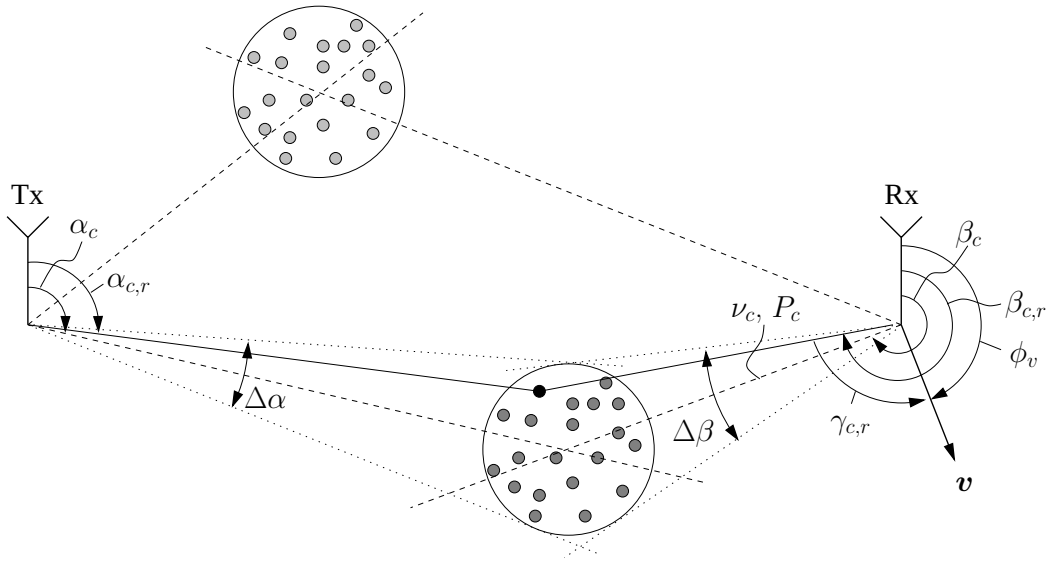


Figure 2.10: Single link of the WINNER channel models [KMH+08, p. 29].

can be calculated as:

$$\alpha_{c,r} = \alpha_c + \Delta\alpha \cdot \vartheta_r, \tag{2.55}$$
$$\beta_{c,r} = \beta_c + \Delta\beta \cdot \vartheta_r, \tag{2.56}$$

where ϑ_r is a fixed ray offset as shown in Table 2.1. If a LOS path is present, it has zero offset. In case of SISO transmission, only the AOA $\beta_{c,r}$ is required for generating the channel coefficients since it influences the Doppler shift of the corresponding ray according to:

$$f_{D,c,r} = v/\lambda \cos\gamma_{c,r} = v/\lambda \cos\left(\beta_{c,r} - \phi_v\right), \tag{2.57}$$

in which v and ϕ_v are the magnitude and the angle of the velocity vector $\boldsymbol{v}$, respectively. For simulations, v is fixed and ϕ_v is generated randomly between 0 and 2π. The Doppler power spectra $D(f_D)$ correspond to sections of the Jakes spectrum since scattering with a preferred direction is assumed instead of two-dimensional isotropic scattering.

Three WINNER channel models are considered in this thesis: *typical urban micro-cell* (B1), *large indoor hall* (B3) and *typical urban macro-cell* (C2). The exact description of these scenarios including the acceptable velocity range are summarised in Table 2.2. The relevant parameters including excess delays ν_c and cluster powers P_c of all three scenarios (LOS version) are given in Table 2.3 to Table 2.5. There is one major difference to the tables given in [KMH+08]. In [KMH+08], the two strongest clusters are divided into three sub-clusters: The first sub-cluster is composed of ten rays and has a zero delay offset, the second sub-cluster consists of six rays and has a delay offset of 5 ns and the last sub-cluster comprises four rays with a delay offset of 10 ns [KMH+08, p. 41]. This division is neglected here: All 20 rays have the same delay, where the original delay is offset by 3.5 ns (mean offset).

Ray r	Ray offset ϑ_r [°]
1,2	±0.0447
3,4	±0.1413
5,6	±0.2492
7,8	±0.3715
9,10	±0.5129
11,12	±0.6797
13,14	±0.8844
15,16	±1.1481
17,18	±1.5195
19,20	±2.1551

Table 2.1: Ray offsets of WINNER channel models [KMH+08, p. 40].

B1 urban micro-cell (0-70 km/h)	"In urban micro-cell scenarios the height of both the antenna at the BS [base station] and at the MS is assumed to be well below the tops of surrounding buildings. Both antennas are assumed to be outdoors in an area where streets are laid out in a Manhattan-like grid. The streets in the coverage area are classified as 'the main street', where there is the LOS from all locations to the BS, with the possible exception in cases where the LOS is temporarily blocked by traffic (e.g. trucks and busses) on the street. Streets that intersect the main street are referred to as perpendicular streets, and those that run parallel to it are referred to as parallel streets. This scenario is defined for both the LOS and the NLOS cases. Cell shapes are defined by the surrounding buildings, and energy reaches NLOS streets as a result of the propagation around corners, through buildings, and between them." [KMH+08, p. 17]
B3 large indoor hall (0-5 km/h)	"Scenario B3 represents the propagation conditions pertinent to operation in a typical indoor hotspot, with wide, but non-ubiquitous coverage and low mobility (0-5 km/h). Traffic of high density would be expected in such scenarios, as for example, in conference halls, factories, train stations and airports, where the indoor environment is characterised by larger open spaces, where ranges between a BS and a MS or between two MS can be significant. Typical dimensions of such areas could range from 20 m x 20 m up to more than 100 m in length and width and up to 20 m in height. Both LOS and NLOS propagation conditions could exist." [KMH+08, p. 17]
C2 urban macro-cell (0-120 km/h)	"In typical urban macro-cell [scenarios a] mobile station is located outdoors at street level and [a] fixed base station [is located] clearly above surrounding building heights. As for propagation conditions, non- or obstructed line-of-sight is a common case, since street level is often reached by a single diffraction over the rooftop. The building blocks can form either a regular Manhattan type of grid, or have more irregular locations. Typical building heights in urban environments are over four floors. Buildings height and density in typical urban macro-cell are mostly homogenous." [KMH+08, p. 19]

Table 2.2: WINNER channel models applied in this thesis.

Cluster c	Excess delay ν_c [ns]	Cluster power P_c [dB]	Ray power $P_{c,r}$ [dB]	AOA β_c [°]	ASA $\Delta\beta$ [°]	Rician factor K_R [dB]
1	0	0.0	-0.31 / -24.7	0	18	3.3
2	33.5	-7.5	-20.5	45		
3	55	-14.8	-27.8	63		
4	63.5	-10.6	-23.6	-69		
5	105	-13.9	-26.9	61		
6	115	-17.8	-30.8	-69		
7	250	-19.6	-32.6	-73		
8	460	-31.4	-44.4	92		

Table 2.3: WINNER channel models: B1-LOS scenario [KMH+08, p. 63].

Cluster c	Excess delay ν_c [ns]	Cluster power P_c [dB]	Ray power $P_{c,r}$ [dB]	AOA β_c [°]	ASA $\Delta\beta$ [°]	Rician factor K_R [dB]
1	0	0.0	-0.32 / -24.5	0	5	2.0
2	3.5	-6.6	-19.6	-53		
3	15	-14.5	-27.6	-79		
4	25	-12.8	-25.8	-74		
5	40	-13.7	-26.7	76		
6	43.5	-11.1	-24.1	80		
7	90	-12.6	-25.6	-73		
8	130	-15.2	-28.2	80		
9	185	-23.3	-36.3	-100		
10	280	-27.7	-40.7	-108		

Table 2.4: WINNER channel models: B3-LOS scenario [KMH+08, p. 65].

Cluster c	Excess delay ν_c [ns]	Cluster power P_c [dB]	Ray power $P_{c,r}$ [dB]	AOA β_c [°]	ASA $\Delta\beta$ [°]	Rician factor K_R [dB]
1	0	0.0	-0.08 / -30.6	0	12	7.0
2	3.5	-13.2	-26.2	-120		
3	30	-15.3	-28.3	129		
4	85	-16.7	-29.7	-135		
5	148.5	-15.2	-28.2	-129		
6	150	-18.2	-31.2	141		
7	160	-15.3	-28.3	-129		
8	220	-23.1	-36.1	-158		

Table 2.5: WINNER channel models: C2-LOS scenario [KMH+08, p. 67].

2.6 Chapter Summary

In this chapter, the mobile radio channel is characterised. Large-scale channel models are briefly presented, whereas small-scale channel modelling is discussed in more detail: Based on the phenomena of multipath propagation and Doppler shifts, common statistical small-scale channel models are presented and classified. Afterwards, the equivalent discrete-time channel model is introduced in order to simulate a continuous-time channel on a computer. Finally, the channel models that are applied throughout this thesis are described.

Chapter 3

System Proposal for Joint Communication and Positioning

In Chapter 1, the motivation for and the benefits of joint communication and positioning have been introduced and different approaches to combine communication with positioning have been presented. On the one hand, existing systems can be combined or extended. On the other hand, new systems can be designed to offer communication and positioning services given a unified signal design. The latter approach is investigated in this thesis. Thus, the first question is: Which general conditions should be fulfilled by such a future joint communication and positioning system? Basically, the system should provide *high data rates* with *low bit error rate (BER)* for the communication part and a *high localisation accuracy* for the positioning part. Since the system requirements for communication and positioning are quite different, as already mentioned in Chapter 1, it is quite challenging to fulfil both conditions at the same time. Thus, a flexible configuration of the system is desirable in order adjust the tradeoff between communication and positioning to the instantaneous needs: In case of an emergency call, emphasis can be laid on the positioning part, whereas the communication payload can be increased in case of data transmissions like file downloads or video streaming. Furthermore, a flexible configuration allows adapting the transmission scheme to changing channel conditions in order to fulfil a certain quality of service (QOS). In addition to flexibility, robustness with respect to noise and interference is desirable.

According to the above mentioned conditions, interleave-division multiplexing (IDM) in combination with pilot layer aided channel estimation (PLACE) and multilateration via the time of arrival (TOA) seems to be a suitable candidate for a joint communication and positioning system as will be explained in the following: The positioning part is examined in Section 3.1. At first, several radiolocation methods with their benefits and drawbacks are introduced. Afterwards, it is explained why multilateration via the TOA is chosen for the proposed joint communication and positioning system. In Section 3.2, the principles of IDM, which are the basis of the communication part, are explained. Both parts are connected via PLACE as presented in Section 3.3. The pilot layer, which carries

only training data, is exploited for estimating the channel coefficients of the equivalent discrete-time channel model, that are needed for data detection, as well as for estimating the TOA. Thus, the pilot layer is essential for the system proposal. Finally, the most important aspects of this chapter are summarised in Section 3.4.

3.1 Positioning Part: Multilateration via the Time of Arrival

There are many different approaches to determine the position of a mobile device, e.g. dead reckoning via motion sensors, sonic or optical positioning as well as radiolocation. Since the joint communication and positioning system investigated in this thesis is based on radio transmission, only radiolocation methods are considered in the following: The position of a *mobile station (MS)* is estimated based on radio signals that are transmitted to or from so-called *reference objects (ROs)* whose positions are known as already mentioned in Section 2.5. There are two basic radiolocation concepts, namely *mobile-centric-* or *self-positioning* and *network-centric-* or *remote-positioning* [Caf00, Ch. 3], [STK05], [GG05], [Gez08]. In case of self-positioning, the MS estimates its own location, i.e. the signals are transmitted from the ROs to the MS. In case of remote-positioning, the location of the MS is determined by the network, i.e. the signals are transmitted from the MS to the ROs. For both concepts, the same positioning algorithms can be applied. Typically, localisation is performed in two steps as shown in Figure 3.1 [PLM02], [SCGL05], [STK05], [Gez08]: At first, position-related parameters are extracted from the received signals, which are exploited for position estimation in the second step. Common parameters that are used for localisation are the *received signal strength (RSS)*, the *time of arrival (TOA)* or *time difference of arrival (TDOA)* and the *angle of arrival (AOA)* [CS98], [Caf00], [PLY+00], [PLM02], [STK05], [GG05], [Gez08]. Based on these parameters, the position estimation can be performed either by *pattern recognition* methods or by *geometrical/statistical* approaches [Gez08].

The pattern recognition techniques are also called *fingerprinting*: In an offline phase, position-related parameters are measured for all locations inside the service area, which are stored in a database. In the online phase, the recent estimated parameters are compared to the database and the position with the best fit is selected for example by the nearest neighbour method or via artificial neural networks [PLM02], [SCGL05], [Gez08]. For fingerprinting, not only the above mentioned parameters like RSS or TOA are used,

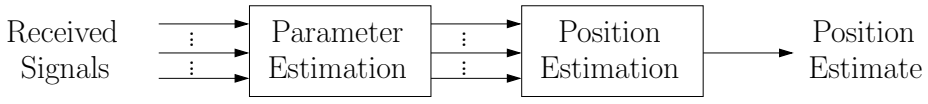
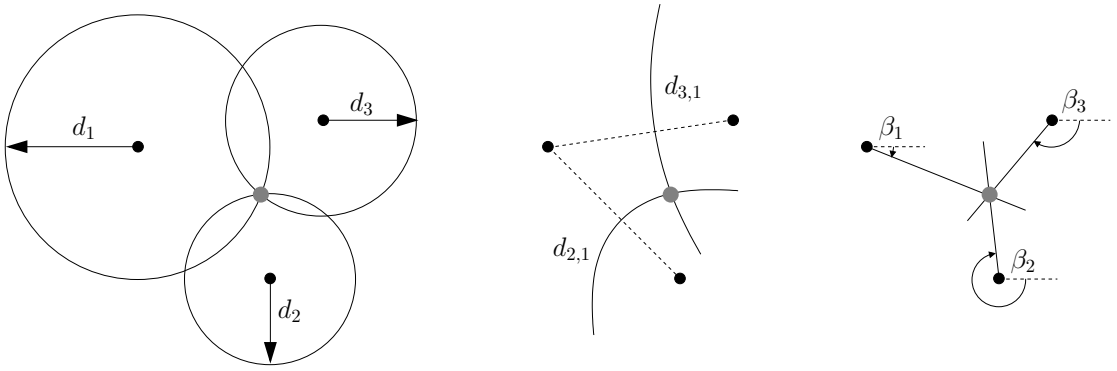


Figure 3.1: Two-step positioning [Gez08, p. 2].

but also the power delay profile or the channel weight function can be exploited [SCGL05], [Gez08]. An advantage of fingerprinting is that it performs well even in the case of multipath and NLOS propagation. The major drawback of fingerprinting is the need of a large database, that has to be updated frequently if the environment changes. Thus, fingerprinting is best suited to and has been mainly applied for quasi-static or slowly varying environments of moderate size (e.g. indoor scenarios) [PLM02], [SCGL05], [Gez08]. However, due to the advance of storage devices and with the help of large companies like Apple or Google, fingerprinting is not restricted to moderate service areas anymore. For example, the coverage area offered by Skyhook, a company that applies Wi-Fi fingerprinting as well as GPS for positioning, comprises North America and Europe amongst others [Sky].

In contrast to fingerprinting, positioning based on geometrical/statistical approaches does not require a database, but, unfortunately, these methods are susceptible to multipath and NLOS propagation. The geometrical/statistical approaches rely on the geometrical relationship between the MS and the ROs and can be subdivided into *multilateration* and *multiangulation* methods, which are based on distance estimates (RSS, TOA, TDOA) and direction estimates (AOA), respectively. From a geometrical point of view and in case of two-dimensional positioning, all position-related parameters lead to so-called *position lines*, that intersect at the MS's position if the true parameters are considered (i.e. for noiseless LOS transmission) [Caf00, Ch. 3], [Gez08]. The shape of the position lines depends on the corresponding parameter as shown in Figure 3.2: The RSS and TOA correspond to the distance d_i between the ith RO and the MS, which leads to a circle with radius d_i, that is centred at the RO's position (see Figure 3.2(a)). In order to obtain a unique intersection, a minimum number of three ROs is required. In case of TDOA, distance differences $d_{j,i}{=}d_j{-}d_i$ instead of distances are considered, where one RO (here: i) is chosen as a reference. The distance differences define hyperbolas with foci at the corresponding ROs (see Figure 3.2(b)). Two hyperbolas lead to a unique intersection. Thus, at



(a) RSS and TOA. (b) TDOA. (c) AOA.

Figure 3.2: Position lines for RSS, TOA, TDOA and AOA [Gez08, p. 11f.]. The black dots denote the positions of the ROs, whereas the slightly larger grey dot denotes the MS's location.

least three ROs are needed in order to determine the MS's position. If the AOA is applied for localisation, only two ROs are required since the position lines defined by the AOA are straight lines, that pass the MS and the corresponding RO (see Figure 3.2(c)). If TOA and AOA measurements are combined, it is possible to determine the MS's position with a single RO. For three-dimensional positioning, the position lines translate to *position surfaces*, namely spheres, hyperboloids and half-planes. Hence, a further RO is necessary in order to determine a unique intersection. A general problem of the geometrical approach is that the position lines/surfaces only intersect at a unique point if the true parameters are considered. Due to noise and other impairments like multipath or NLOS propagation, estimation errors occur such that the position lines/surfaces intersect at several points. Hence, statistical methods are applied in order to determine the MS's position in the presence of estimation errors. Basically, a set of nonlinear equations has to be solved. Further details for multilateration based on the TOA are provided in Chapter 5.

After introducing the basic mechanisms of radiolocation, it is discussed in the following which methods are promising for a joint communication and positioning system. Of course, the best selection can be done if a specific application is already known. Since this is not the case here, more abstract considerations have to be taken into account: Since fingerprinting requires a large database and is, thus, best suited for slowly varying environments of moderate size, the more general geometrical/statistical methods are considered subsequently. The next matter is, which position-related parameters should be utilised? To answer this question, it is useful to have a closer look on the available parameters again.
As already mentioned in Section 2.2, large-scale channel models can be applied to relate the RSS to the distance between transmitter and receiver. Small-scale fading effects should be mitigated by averaging the RSS over a certain observation interval [Gez08]. Knowing the transmit power, the distance between transmitter and receiver can be determined via (2.13) or (2.14) if path loss only or path loss and shadowing are considered, respectively. It does not matter, whether a LOS path is present or not. However, only rough distance estimates can be obtained by the RSS since the path loss and shadowing models are just approximate. Hence, the achievable positioning accuracy is quite poor. For that reason, the RSS is ignored in the following.
All remaining position-related parameters are based on the assumption that a LOS path exists. Thus, their accuracy degrades significantly if this is not the case (NLOS). Furthermore, they are all susceptible to multipath propagation as will be explained below. Similar to the RSS, the TOA estimates the distance between transmitter and receiver. The most prominent application employing the TOA is the global positioning system (GPS) [Kap96], [EM99], [BvD99]. If the time of departure (TOD)[1] is known, the propagation delay τ_i of the signal can be determined by measuring the TOA. The distance between transmitter and receiver is related to the propagation delay by the speed of light $c \approx 3 \cdot 10^8$ m/s according to $d_i = c\,\tau_i$. Hence, small errors in propagation delay translate to large errors in distance. In order to estimate the propagation delay as accurate as possi-

[1] The TOD is typically transmitted together with the ranging signal, i.e. the signal is time-stamped by the transmitter [GTG+05]. This method is for example applied for GPS [BvD99].

ble, the clocks at the ROs and the MS need to be synchronised. Clock synchronisation is a challenging task and, typically, it is only feasible to synchronise the clocks at the ROs since they can be linked by a network. The MS is usually equipped with a simple, unsynchronised clock, which leads to a common bias in all distance estimates. This bias is cancelled automatically if TDOAs are considered [PLY+00], [GG05]. Another possibility to mitigate the clock bias is to include it as a nuisance parameter into the set of nonlinear equations [GG05] as it is done for example for GPS [Kap96, Ch. 2], [BvD99]. In this case, an additional RO is needed to determine the set of nonlinear equations. That means that a minimum number of four instead of three and five instead of four ROs is required for two- and three-dimensional positioning, respectively. Another source of error is multipath propagation as already mentioned above. In order to understand why this is the case, it is examined how the TOAs are commonly estimated: Basically, estimating the TOA corresponds to a timing synchronisation, which is typically performed by means of correlation or matched filtering [CS98], [PLM02], [GTG+05], [Gez08]. A known training sequence is transmitted, which is time delayed by the channel. The received signal is correlated to a replica of the training sequence and the correlation peak corresponds to the TOA in case of pure LOS transmission. In case of multipath propagation, the received signal consists of several superimposed replicas of the training sequence and, thus, the correlation function has several peaks instead of a single one. If the multipath components are *resolvable*, i.e. their separation is larger than the symbol duration T_s (see Section 2.3.2), then the peaks can be distinguished and the TOA can be estimated by finding the first peak. In contrast, the TOA estimator suffers if the multipath components are *non-resolvable*. In this case, the correlation function is distorted such that the peaks can not be distinguished anymore. The first peak is shifted to a larger delay, which leads to a bias in the TOA estimate [CS98], [Caf00, Ch. 3], [GTG+05]. The TDOA estimates suffer in the same manner from multipath propagation since they are either obtained directly by cross-correlating two received signals or by subtracting two TOA measurements. Since the resolvability of multipath components depends on the signal bandwidth, the accuracy of the TOA/TDOA estimates can be improved by increasing the bandwidth. Hence, ultra wideband (UWB) systems are promising for high resolution positioning in severe multipath propagation environments (e.g. indoors) [PLM02], [GTG+05], [Gez08]. However, the coverage of UWB signals is limited, which means that a dense network of ROs is required in that case.
The last position-related parameter is the AOA, which is estimated by directional antennas or antenna arrays [CS98], [Caf00, Ch. 3], [GG05], [GTG+05], [Gez08]. The accuracy of the AOA estimates depends among others on the number of antenna elements and their spacing [Gez08]. Hence, in order to obtain a high resolution, cost intensive antenna arrays have to be applied. On the one hand, an advantage of AOA based positioning is that only two ROs are required. On the other hand, a major drawback lies in the fact that the positioning accuracy degrades with increasing distance between transmitter and receiver given a certain AOA resolution [CS98], [Caf00, Ch. 3]. Furthermore, the AOA estimates are very susceptible to multipath propagation [CS98], [Caf00, Ch. 3], [GTG+05].
Taking the above considerations into account, the TOA appears to be a good choice for a joint communication and positioning system since it enables precise localisation:

> "The time-based radiolocation methods seem to have the greatest potential for providing high accuracy location estimates. Unlike the AoA methods, the accuracy of time-based methods does not degrade with increasing MS-BS [mobile station-base station] distances, and it is not necessary to place large antenna arrays at the cell sites. Moreover, it is generally considered easier to obtain accurate time measurements as opposed to accurate angle measurements." [Caf00, p. 33]

This choice is further encouraged by the fact that timing synchronisation is needed for communication purposes as well even though with different requirements: For communication, sampling should be performed such that the signal-to-noise ratio (SNR) is maximised at the receiver side. Thus, the highest correlation peak should always be chosen instead of the first one. Nevertheless, synchronisation has to be performed anyway. Hence, multilateration based on the TOA is applied in the remainder of this thesis. However, it should be kept in mind that it is always advantageous to combine several positioning techniques via sensor fusion [WKAR06], [HROW07], [RRA10]. Especially in urban and indoor environments, which are very challenging concerning radiolocation due to multipath propagation and NLOS conditions, sensor fusion is essential if high positioning accuracies are desired. Thus, the positioning part of the system proposed in this thesis can be understood as a contribution to the overall positioning process using sensor fusion as already mentioned in Chapter 1.

3.2 Communication Part: Interleave-Division Multiplexing

It has been stated in the beginning of this chapter that a *flexible configuration* is desirable for a joint communication and positioning system. According to this aspect *interleave-division multiplexing (IDM)* seems to be a suitable candidate for joint communication and positioning. IDM is closely related to *interleave-division multiple access (IDMA)* since both techniques are based on the same principle. IDMA is used to distinguish signals of different users, whereas IDM is employed to separate several data streams of a single user. Due to the similarity of IDM and IDMA, research results obtained for one technique can be usually applied to the other as well. IDMA was first mentioned by the group of Li Ping in [PLWL02]. Similar ideas had been published previously/simultaneously in [CT97], [FOO00], [BSGS00] and [MP02]. The term IDM was introduced independently by the group of Li Ping and the group of Peter A. Höher with slightly different meaning. On the one hand, IDM in conjunction with space-time codes was proposed by Li Ping's group in [PLWL02]. In this case, coding is performed for a single data stream that is afterwards divided into several data streams that are interleaved separately. On the other hand, pure IDM, as applied in this thesis, was proposed by Peter A. Höher's group in [SH04]. In that case, coding *and* interleaving are performed for each data stream separately. The combination of IDM and IDMA is also called *multi layer IDMA* [HSF08], [Sch08].

In order to illustrate the basic concepts of IDM concisely, only symbol-rate sampling (J=1) is considered in the subsequent sections. This means that the sampling period $\mathcal{T}$ corresponds to the symbol duration T_s and that the transmitted sequence $x[k]$ equals the coded and modulated sequence $x'[\kappa]$ with burst length K=K'. Therefore, the notation using the prime is omitted at this point and the coded and modulated sequence is denoted by $x[k]$ directly.

3.2.1 Transmitter Structure of Interleave-Division Multiplexing

The basic principle of IDM is illustrated in Figure 3.3: Several data streams of a user, which are called *layers* in the following, are linearly superimposed. In contrast to traditional multiplexing schemes like time-division multiplexing (TDM) and frequency-division multiplexing (FDM), all layers are transmitted at the same time on the same carrier frequency. IDM is a special case of code-division multiplexing (CDM), where the layers are distinguished by layer-specific interleavers. Let $u_m[n]$, $0 \leq n < N$, $1 \leq m \leq M$, denote the nth bit of the mth layer. Each layer is encoded with code rate R=N/K (ENC), where K the length of the transmission burst. After encoding, every layer is interleaved by a layer-specific interleaver (π_m) and mapped onto the complex plane via binary phase shift keying (BPSK), which leads to the layer-wise symbols $x_m[k]$. Before all layers are summed up, an adequate power and phase allocation with complex weighting factors $a_m\,\mathrm{e}^{\mathrm{j}\,\xi_m}$ is performed. Thus, the kth symbol of the transmission burst of length K is given by

$$x[k] = \sum_{m=1}^{M} a_m\,\mathrm{e}^{\mathrm{j}\,\xi_m} \cdot x_m[k]. \tag{3.1}$$

Each symbol $x[k]$ carries $\mathcal{B}$=RM bits. Hence, $\mathcal{B}$ is also called *bit load* [HSF08]. Since all layers employ the same encoders in combination with BPSK mapping, the transmitter structure of IDM is very simple. However, IDM offers a *flexible configuration* because the data rate can be easily adapted by changing the number of layers M instead of changing the modulation scheme [SH04], [SHF05]. Furthermore, (layer-wise) unequal

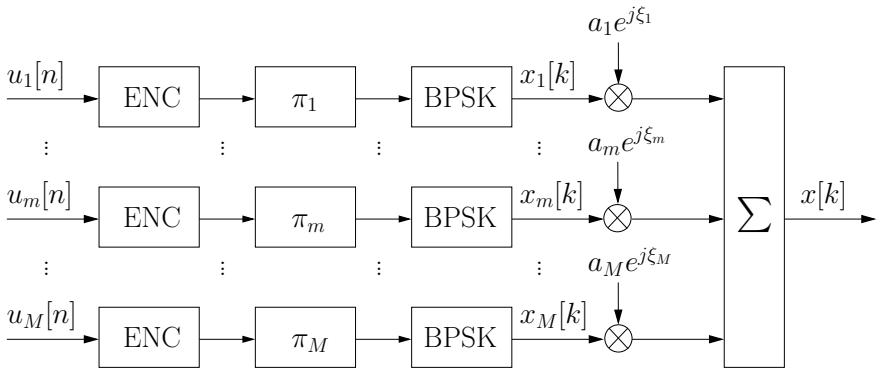


Figure 3.3: Transmitter structure of IDM.

error protection (UEP) can be easily achieved by assigning different amplitude levels to different layers [HSF08], [Sch08]. Of course, more sophisticated transmitter structures are possible, e.g. with layer-specific encoders or higher-order modulation schemes instead of BPSK. Nevertheless, only the transmitter structure shown in Figure 3.3 is considered throughout this thesis due to its simplicity and flexibility.

The bits of the different layers $u_1[n], \cdots, u_M[n]$ may be obtained by different, independent sources (e.g. voice transmission or file download) or by a serial-to-parallel conversion of a single data stream with bits $u[n]$. In the latter case, IDM is also called *multi-level coding with sigma mapping* [MP04], *superposition coding* [TP06], [TPM06] or *superposition coded modulation* [TPM09]. Another technique that is closely related to IDM is *superposition mapping* in combination with *bit-interleaved coded modulation (BICM)* [WH10a], [WNHH10] as shown in Figure 3.4: In contrast to IDM, encoding and interleaving is performed before serial-to-parallel conversion. Thus, only BPSK mapping and power and phase allocation are performed in each layer before superposition. BICM with superposition mapping (BICM-SM) had previously been investigated in [MP04] where it is called *single-level coding with sigma mapping*. IDM and BICM-SM provide similar benefits and drawbacks: Due to the superposition of several binary layers, the symbols $x[k]$ are Gaussian-like[2]. This is a significant advantage from a theoretical point of view since the channel capacity of an AWGN channel can only be achieved by Gaussian distributed channel outputs according to Shannon [Sha48], which in turn can only be achieved by Gaussian distributed channel inputs. From a practical point of view, Gaussian-like symbols are disadvantageous since they cause a high peak-to-average power ratio (PAPR). Thus, sophisticated/expensive power amplifiers with a large linear range are required in order to avoid nonlinear distortions. If simple power amplifiers are applied, the PAPR should be reduced in order to mitigate nonlinear distortions. Clipping can be applied for that purpose as investigated in [TP06], [TPM06], [TPM09] and [HH08]. Alternatively, the PAPR can be decreased by an adequate power and phase allocation [NHW11]. Another aspect of IDM and BICM-SM is that the overall mapping from bit tuples to symbols

[2]Constellation diagrams for BICM-SM are shown for example in [WH10a] and [WNHH10].

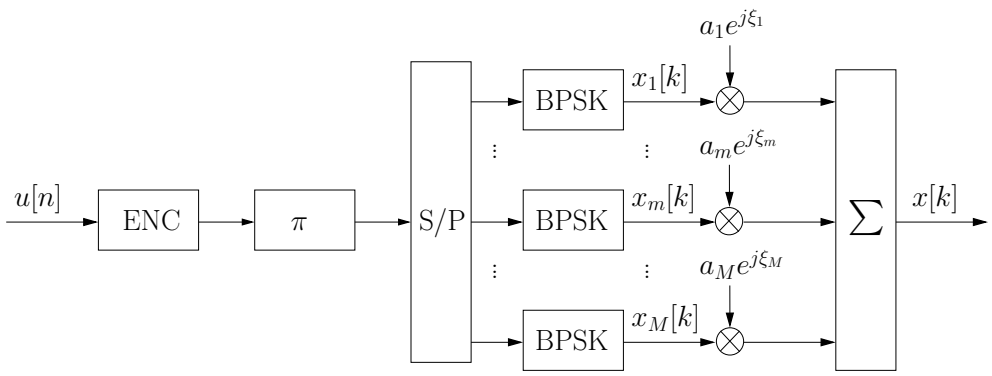


Figure 3.4: BICM with superposition mapping.

can be *non-bijective* depending on the power and phase allocation [WH10a], [WNHH10]. This means that there is no one-by-one relationship between bit tuples and symbols as it is the case for traditional mapping schemes like amplitude shift keying (ASK), phase shift keying (PSK) or quadrature amplitude modulation (QAM). A simple example is the superposition of two layers with equal power and phase ($a_1\,\mathrm{e}^{\mathrm{j}\,\xi_1} = a_2\,\mathrm{e}^{\mathrm{j}\,\xi_2} = 1$): There are four possible bit tuples, (-1,-1), (+1,-1), (-1,+1), (+1,+1), but the symbol alphabet consists only of three symbols: $\{-2, 0, 2\}$. For the symbols ± 2, the demapping can be done directly. In order to resolve the mapping ambiguity of the last symbol, channel coding and interleaving in combination with an iterative turbo type receiver are mandatory [WH10a], [WH10b], [WNHH10]. In [WH10b], different channel codes for BICM-SM are investigated by means of extrinsic information transfer (EXIT) chart analysis. It is shown that a *repetition (REP)* code is better suited to resolve the mapping ambiguities than a *forward error correction (FEC)* code:

> "The key task of channel coding is to introduce redundancy to the originally independent information bits. Without loss of generality, the way of adding redundancy can be categorised into two classes: repetition and parity bits. Conventionally, the latter is favourable since it brings coding gain. For systems with bijective uniform mapping schemes, e.g. ASK, the only task of channel coding is to combat the additive noise. Now, with a non-bijective non-uniform mapping scheme, the situation becomes different. For systems employing SM-EPA [superposition mapping with equal power allocation], the primary task of channel coding is to ensure a perfect separation of superimposed chips, while combating the additive noise becomes the secondary task. We will show that repetition is much more efficient than parity bits for the first task. Meanwhile, parity bits are still necessary because of the second task." [WH10b, p. 326]

Typically, a serial concatenation of an FEC and an REP code is applied. In the author's opinion, this concatenated code design is rather motivated by historical reasons than by the findings in [WH10b]: In the beginning, research was mainly focussed on IDMA, which was derived from direct sequence code-division multiple access (DS-CDMA). The

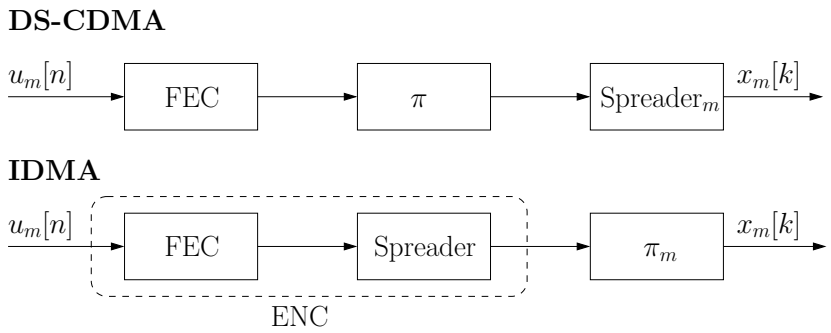


Figure 3.5: DS-CDMA versus IDMA [HSF08, p. 2], [Sch08, p. 17].

interest in IDM and BICM-SM emerged a little bit later. IDMA and DS-CDMA are both special forms of code-division multiple access (CDMA) and, thus, they are closely related. The basic difference between both multiple access schemes is that the order and the role of the interleaver and the spreader are interchanged as shown in Figure 3.5: In DS-CDMA, the data of each user is first encoded by an FEC code, then interleaved and finally multiplied with a user-specific spreading sequence. The bits after the spreader are also called *chips*. In contrast, the FEC encoder is followed by a common spreader in IDMA. User separation is done by applying user-specific interleavers on chip-level. The FEC code and the spreader can be combined to an overall encoder (ENC). Spreading corresponds to an REP code in combination with scrambling. Thus, the frequent usage of a serial concatenated code (FEC+REP) in the framework of IDMA, IDM and BICM-SM is, according to the author's opinion, mainly motivated by the original design of IDMA. If only an REP code is applied for IDMA, IDM or BICM-SM, these schemes are also called *uncoded* since an REP code does not provide any coding gain in case of an AWGN channel. Otherwise, they are called *coded*. This naming convention was first introduced in [PLWL02].

3.2.2 Receiver Structure of Interleave-Division Multiplexing

After introducing and discussing the transmitter structure of IDM, possible receivers are presented in the following. Generally, turbo type receivers are applied in the framework of IDMA, IDM and BICM-SM. In Figure 3.6, a turbo type receiver for IDM is shown: Basically, the receiver consists of a *multi layer detector (MLD)*[3] and a bank of layer-wise

[3]In case of IDMA, the MLD corresponds to a multi user detector (MUD).

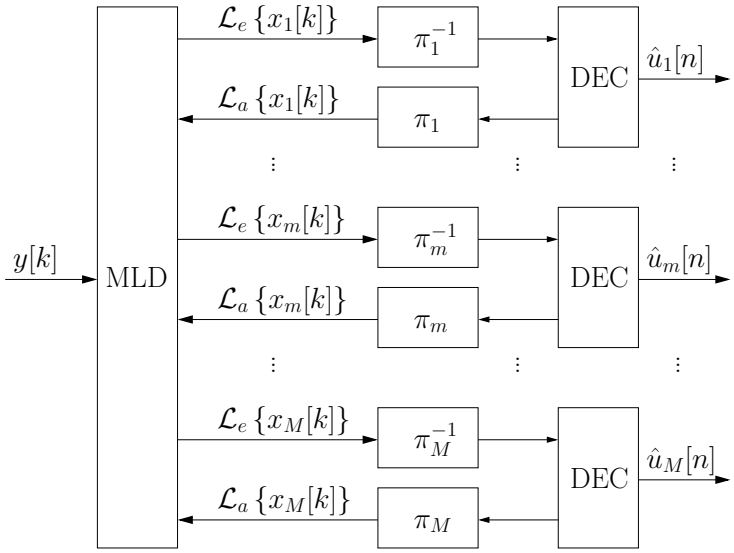


Figure 3.6: Receiver structure of IDM.

decoders (DEC). In the MLD, only the multiplexing (multiple access/mapping) and the channel constraint are considered ignoring the coding constraint. In contrast, the layer-wise decoders take only the coding constraint into account. According to the turbo principle, *extrinsic information* is exchanged iteratively between the MLD and the decoders: The extrinsic information generated by the MLD is deinterleaved for each layer and acts as a priori information to the corresponding decoder and vice versa. Typically, the extrinsic information is exchanged in form of *log-likelihood ratios (LLRs)*. The LLR of the kth symbol in the mth layer, $x_m[k]$, given the received burst $\boldsymbol{y}=[y[0],\ldots,y[K+L-1]]^T$, is defined as follows [Hub02]:

$$\mathcal{L}\{x_m[k]\} := \ln \frac{P(x_m[k]=+1|\boldsymbol{y})}{P(x_m[k]=-1|\boldsymbol{y})}. \tag{3.2}$$

According to Bayes rule, the LLR can be separated into the *extrinsic LLR* $\mathcal{L}_e\{x_m[k]\}$ and the *a priori LLR* $\mathcal{L}_a\{x_m[k]\}$:

$$\mathcal{L}\{x_m[k]\} = \underbrace{\ln \frac{p(\boldsymbol{y}|x_m[k]=+1)}{p(\boldsymbol{y}|x_m[k]=-1)}}_{\mathcal{L}_e\{x_m[k]\}} + \underbrace{\ln \frac{P(x_m[k]=+1)}{P(x_m[k]=-1)}}_{\mathcal{L}_a\{x_m[k]\}}. \tag{3.3}$$

The extrinsic LLRs can be either generated directly or they can be obtained by subtracting the a priori LLRs from the overall LLRs: $\mathcal{L}_e\{\cdot\}=\mathcal{L}\{\cdot\}-\mathcal{L}_a\{\cdot\}$. The first method is usually applied at the MLD and the second method is typically used at the decoder side, where maximum a posteriori probability (MAP) decoding is commonly performed. At the MLD, different detection strategies can be applied: MAP detection is optimal in the sense that the a posteriori probability of the layer-wise symbols $x_m[k]$ is maximised, but its complexity is generally prohibitive. Thus, MAP detection at the MLD is only investigated for AWGN channels with perfect channel knowledge as in [CRE09], [NWH10], [HH10], [WH10a], [WH10b] and [WNHH10] or for a frequency-selective channel in combination with orthogonal frequency-division multiplexing (OFDM) as in [SWH10]. It is shown in [NWH10], [WH10a] and [WNHH10] that arbitrary high bit loads $\mathcal{B}$ can be supported in this case. Alternative detection strategies comprise minimum mean squared error (MMSE) detection [KDUB07], tree-based approaches [WNHH10], detection based on a Gaussian approximation (GA) [PLWL02], [LLP03], [LP04], [PLWL06], [CRE09] and hybrid schemes [HH10]. The suboptimal GA-based detectors are very popular due to their reduced complexity, especially for IDMA, but their performance is limited: For *equal power allocation* (a_m=a, $\forall m$), the maximum bit load $\mathcal{B}$, that can be supported by a GA-based detector, is limited to approximately 2.085 per signal dimension for the noiseless case [Sch08, p. 111]. For realistic scenarios, the achievable bit loads are even lower. Hence, low rate codes should be applied if a large number of layers M has to be supported ($\mathcal{B}$=RM). The reason for the limited performance is that the GA is poor if the signal-to-noise ratio (SNR) is high and the a priori information $\mathcal{L}_a\{x_m[k]\}$ is low as it is the case in the beginning of the iterative process [HH10]. Bit loads higher than 2.085 per signal dimension can only be achieved by *unequal power allocation* at the cost of decreased power efficiency [Sch08, p. 125], [HSF08].

Nevertheless, a GA-based MLD is applied in this thesis since the complexity of all other detection strategies is prohibitive for the considered frequency-selective channel models. Thus, the GA-based MLD is examined in more detail in the following based on [PLWL02], [LLP03], [LP04] and [PLWL06]. Due to didactic reasons, the GA-based approach is first introduced for flat channels. The extension to frequency-selective channels is presented afterwards. Furthermore, block fading and perfect channel knowledge are assumed for clarity reasons.

GA-based MLD for flat channels

Given a flat channel, the kth symbol of the mth layer, $x_m[k]$, is contained in the kth received sample $y[k]$ only:

$$y[k] = \sum_{m=1}^{M} \underbrace{h_0\, a_m\, \mathrm{e}^{\mathrm{j}\,\xi_m}}_{g_m} \cdot x_m[k] + n[k]. \tag{3.4}$$

Thus, the extrinsic LLR is given by

$$\mathcal{L}_e\left\{x_m[k]\right\} = \ln \frac{p(y[k]|x_m[k]=+1)}{p(y[k]|x_m[k]=-1)}. \tag{3.5}$$

The received sample in (3.4) can be rewritten as

$$y[k] = g_m x_m[k] + \underbrace{\sum_{\substack{m'=1 \\ m'\neq m}}^{M} g_{m'} x_{m'}[k] + n[k]}_{\zeta_m[k]}, \tag{3.6}$$

where the interference caused by other layers, which is also called *multi layer interference (MLI)*, and the noise sample can be combined to an *effective noise term* $\zeta_m[k]$. According to the central limit theorem, the effective noise term $\zeta_m[k]$ can be approximated by a Gaussian variable with mean $\mu_{\zeta_m[k]} = \mathrm{E}\left\{\zeta_m[k]\right\}$ and variance $\sigma^2_{\zeta_m[k]} = \mathrm{Var}\left\{\zeta_m[k]\right\}$ for large M, which leads to the following probability density function

$$p(y[k]|x_m[k]=\pm 1) = \frac{1}{\pi\sigma^2_{\zeta_m[k]}} \exp\left(-\frac{1}{\sigma^2_{\zeta_m[k]}}\left|y[k] \mp g_m - \mu_{\zeta_m[k]}\right|^2\right). \tag{3.7}$$

By substituting (3.7) into (3.5), the extrinsic LLR can be simplified as

$$\mathcal{L}_e\left\{x_m[k]\right\} = \frac{4}{\sigma^2_{\zeta_m[k]}}\, \mathrm{Re}\left\{g_m^* \cdot \left(y[k] - \mu_{\zeta_m[k]}\right)\right\}. \tag{3.8}$$

In order to calculate the extrinsic LLR, the mean and variance of the effective noise term have to be estimated:

$$\mu_{\zeta_m[k]} = \mathrm{E}\left\{y[k]\right\} - g_m\, \mathrm{E}\left\{x_m[k]\right\}, \tag{3.9}$$

$$\sigma^2_{\zeta_m[k]} = \mathrm{Var}\left\{y[k]\right\} - |g_m|^2\, \mathrm{Var}\left\{x_m[k]\right\}, \tag{3.10}$$

with

$$\mathrm{E}\{y[k]\} = \sum_{m=1}^{M} g_m \,\mathrm{E}\{x_m[k]\}, \tag{3.11}$$

$$\mathrm{Var}\{y[k]\} = \sum_{m=1}^{M} |g_m|^2 \,\mathrm{Var}\{x_m[k]\} + \sigma_n^2, \tag{3.12}$$

and

$$\mathrm{E}\{x_m[k]\} = \tanh\left(\frac{\mathcal{L}_a\{x_m[k]\}}{2}\right), \tag{3.13}$$

$$\mathrm{Var}\{x_m[k]\} = 1 - |\,\mathrm{E}\{x_m[k]\}\,|^2. \tag{3.14}$$

Equations (3.9) to (3.12) only apply if all $x_m[k]$ are independent. This is approximately true due to the interleavers π_m, which break the correlations introduced by the encoders. At the MLD, the equations (3.8) to (3.14) are evaluated in reversed order: (i) At first, the mean and variance of the layer-wise symbols $x_m[k]$ are updated based on the a priori LLRs $\mathcal{L}_a\{x_m[k]\}$, (ii) then, the mean and variance of the received samples $y[k]$ are calculated and, (iii) finally, the mean and variance of the effective noise terms $\zeta_m[k]$ are determined, which are needed to (iv) compute the extrinsic LLRs $\mathcal{L}_e\{x_m[k]\}$. The second step is shared by all layers. In the beginning of the iterative process, all a priori LLRs $\mathcal{L}_a\{x_m[k]\}$ are initialised to zero since no a priori knowledge about the layer-wise symbols is available yet.

GA-based MLD for frequency-selective channels

For frequency-selective channels, the procedure at the MLD is very similar to that presented above. The main difference is that the kth symbol of the mth layer, $x_m[k]$, is now contained in $L{+}1$ adjacent received samples $y[k], \cdots, y[k{+}L]$ due to ISI:

$$y[k] = \sum_{l=0}^{L} \sum_{m=1}^{M} \underbrace{h_l \, a_m \, \mathrm{e}^{\mathrm{j}\,\xi_m}}_{g_{m,l}} \cdot x_m[k-l] + n[k]. \tag{3.15}$$

Thus, the extrinsic LLR is now given by

$$\mathcal{L}_e\{x_m[k]\} = \ln \frac{p(y[k], \cdots, y[k+L] | x_m[k] = +1)}{p(y[k], \cdots, y[k+L] | x_m[k] = -1)}. \tag{3.16}$$

The effective noise term additionally depends on l now since it comprises ISI as well as the MLI and the noise sample:

$$y[k+l] = g_{m,l} x_m[k] + \underbrace{\sum_{\substack{l'=0 \\ l' \neq l}}^{L} g_{m,l'} x_m[k-l'] + \sum_{l'=0}^{L} \sum_{\substack{m'=1 \\ m' \neq m}}^{M} g_{m',l'} x_{m'}[k-l'] + n[k]}_{\zeta_{m,l}[k]}, \tag{3.17}$$

The effective noise terms $\zeta_{m,0}[k], \cdots, \zeta_{m,L}[k]$, which correspond to $x_m[k]$, are generally correlated due to ISI. This means that the covariances between them have to be taken into account. The calculation of the extrinsic LLR $\mathcal{L}_e\{x_m[k]\}$ can be considerably simplified if the effective noise terms $\zeta_{m,0}[k], \cdots, \zeta_{m,L}[k]$ are assumed to be uncorrelated. In this case, the probability density function can be factorised according to

$$p(y[k], \cdots, y[k+L] | x_m[k] = \pm 1) = \prod_{l=0}^{L} p(y[k+l] | x_m[k] = \pm 1). \tag{3.18}$$

Hence, the extrinsic LLR simplifies as

$$\mathcal{L}_e\{x_m[k]\} = \sum_{l=0}^{L} \underbrace{\frac{4}{\sigma^2_{\zeta_{m,l}[k]}} \operatorname{Re}\left\{ g^*_{m,l} \cdot \left(y[k+l] - \mu_{\zeta_{m,l}[k]} \right) \right\}}_{\mathcal{L}^l_e\{x_m[k]\}}. \tag{3.19}$$

The mean $\mu_{\zeta_{m,l}[k]} = \mathrm{E}\{\zeta_{m,l}[k]\}$ and variance $\sigma^2_{\zeta_{m,l}[k]} = \mathrm{Var}\{\zeta_{m,l}[k]\}$ of the effective noise term are determined similar to (3.9) to (3.12):

$$\mu_{\zeta_{m,l}[k]} = \mathrm{E}\{y[k+l]\} - g_{m,l}\,\mathrm{E}\{x_m[k]\}, \tag{3.20}$$

$$\sigma^2_{\zeta_{m,l}[k]} = \mathrm{Var}\{y[k+l]\} - |g_{m,l}|^2\,\mathrm{Var}\{x_m[k]\}, \tag{3.21}$$

with

$$\mathrm{E}\{y[k]\} = \sum_{l=0}^{L} \sum_{m=1}^{M} g_{m,l}\,\mathrm{E}\{x_m[k-l]\}, \tag{3.22}$$

$$\mathrm{Var}\{y[k]\} = \sum_{l=0}^{L} \sum_{m=1}^{M} |g_{m,l}|^2\,\mathrm{Var}\{x_m[k-l]\} + \sigma^2_n. \tag{3.23}$$

The mean and variance of the layer-wise symbols $x_m[k]$ are updated according to (3.13) and (3.14). The detection approach described above is also called *soft rake* approach [LLP03] or *LLR combining* approach [LP04], [PLWL06]. The latter term is used subsequently. The MLD based on LLR combining (LLRC) offers a good tradeoff between performance and complexity [LP04]: For moderate bit loads, it performs nearly optimal, but its performance degrades for increased bit loads and an error floor occurs. In order to obtain an improved performance at increased bit loads, other combining strategies can be applied at the cost of increased complexity. The most complex alternative, that shows the best performance in the context of GA-based detection, is *joint Gaussian combining (JGC)*, which takes the correlation between the effective noise terms $\zeta_{m,0}[k], \cdots, \zeta_{m,L}[k]$ into account [LLP03], [LP04]. Another option is to use a *maximum ratio combining (MRC)* approach, where the order of estimation and combining is interchanged in comparison to the LLRC approach [LP04]: First, the received signal is filtered according the MRC principle for each layer:

$$z_m[k] = \sum_{l=0}^{L} g^*_{m,l} y[k+l]. \tag{3.24}$$

Afterwards, the extrinsic LLRs $\mathcal{L}_e\{x_m[k]\}$ are determined similar to (3.8) based on the filtered received samples $z_m[k]$. For moderate bit loads, all three approaches show a similar performance [LLP03], [LP04]. Throughout this thesis, LLRC in combination with moderate bit loads is applied since it offers nearly optimal performance at the lowest complexity in this case.

3.3 Connection between both Parts: Pilot Layer Aided Channel Estimation

In the preceding sections, it has been demonstrated that multilateration via the TOA as well as IDM are suitable techniques for a joint communication and positioning system. The last question that needs to be answered is: How can these techniques be combined? On the one hand, position-related parameters like the TOA shall be estimated from the radio signals. On the other hand, unknown data shall be transmitted via the radio signals from one location to another. The connection between both parts lies in the fact that the mobile radio channel needs to be known in order to detect the unknown transmitted data. That means that the channel coefficients of the equivalent discrete-time channel model have to be estimated. Blind channel estimation is possible, but, typically, known training data is inserted into the transmission burst since it significantly simplifies channel estimation. Traditionally, training is multiplexed to the data in the time and/or frequency domain. This can be done block-wise as pre-, mid- or postamble like it is the case for GSM [Kam04, p. 551-556], [Höh11, p. 396f.] or applying a periodic pattern like it is the case for UMTS [Höh11, p. 397ff.] and OFDM based systems like terrestrial digital video broadcasting (DVB-T) or long term evolution (LTE) [HKR97], [KSHA10], [KHTA11]. Alternatively, training and data can be superimposed [HT99]. This concept has been proposed for IDM/IDMA where it is called *pilot layer aided channel estimation (PLACE)* [SH05], [SH06], [Sch08, Ch. 8]: An additional layer is superimposed onto the M data layers, which carries only training symbols $x_0[k]$ as shown in Figure 3.7. This pilot layer can not only be exploited for channel estimation purposes, but also for the estimation of position-related parameters like the TOA. Hence, the pilot layer is essential for the proposed joint communication and positioning system. The tradeoff between communication and positioning purposes can be regulated via an adequate power allocation: The ratio of the pilot layer power to the total power

$$\rho = \frac{a_0^2}{\sum_{m=0}^{M} a_m^2} \tag{3.25}$$

can be varied between 0 to 1, where ρ=0 and ρ=1 correspond to no training at all and pure training, respectively. The more power is spent on the pilot layer, the better the channel coefficients as well as the TOA can be estimated. However, with respect to the communication payload, the design parameter ρ should be kept as small as possible. Fortunately, the transmission scheme can be adapted to the instantaneous requirements by changing

ρ accordingly. Therefore, the proposed joint communication and positioning system is highly flexible. Not all data layers have to be devoted to communication purposes. Some of them may also be used to transmit auxiliary information for localisation purposes such as the TOD or the position of the corresponding ROs in case of self-positioning.

If channel estimation is performed based on a superimposed pilot layer, some modifications are required at the receiver side as shown in Figure 3.8. First, equations (3.8) to (3.12) and (3.19) to (3.23) concerning the MLD have to be adjusted replacing the true (perfectly known) channel coefficients h_l by their estimates $\hat{h}_l$, which leads to $\hat{g}_m{=}\hat{h}_0{\cdot}a_m\,\mathrm{e}^{\mathrm{j}\,\xi_m}$ and $\hat{g}_{m,l}{=}\hat{h}_l{\cdot}a_m\,\mathrm{e}^{\mathrm{j}\,\xi_m}$, respectively. Furthermore, the pilot layer does not have to be detected by the MLD because it only contains known training symbols. In order to mitigate the inference caused by the pilot layer, the training symbols $x_0[k]$ should be subtracted from the received samples $y[k]$ before they are fed into the MLD. Thus, *pilot layer interference cancellation (PIC)* is performed previous to the MLD. Furthermore, the received samples are fed into a PLACE unit, where channel estimation is accomplished based on the a priori knowledge about the pilot layer. Additionally, the a priori LLRs $\mathcal{L}_a\{x_m[k]\}$, $1{\leq}m{\leq}M$, which are fed back from the layer-wise decoders, can be exploited to improve channel estimation: They can be either used as virtual training for *semi-blind channel estimation* or in order to perform *data layer interference cancellation (DIC)* before training-based channel estimation [SH06], [Sch08, Ch. 8]. Thus, PLACE is included into the iterative detection and decoding process. In each iteration, updated channel estimates $\hat{h}_l[k]$ are fed to the PIC and the MLD. If DIC is applied before channel estimation, the estimated channel coefficients of the previous iteration are utilised. In contrast to pure communication systems, position-related parameters like the TOA are estimated in addition to the channel coefficients from the pilot layer. Hence, channel and parameter estimation are performed simultaneously. Therefore, the estimation process performed by the PLACE unit is called *channel parameter estimation (CPE)* in the following. CPE based on the pilot layer is the core part of the proposed joint communication and posi-

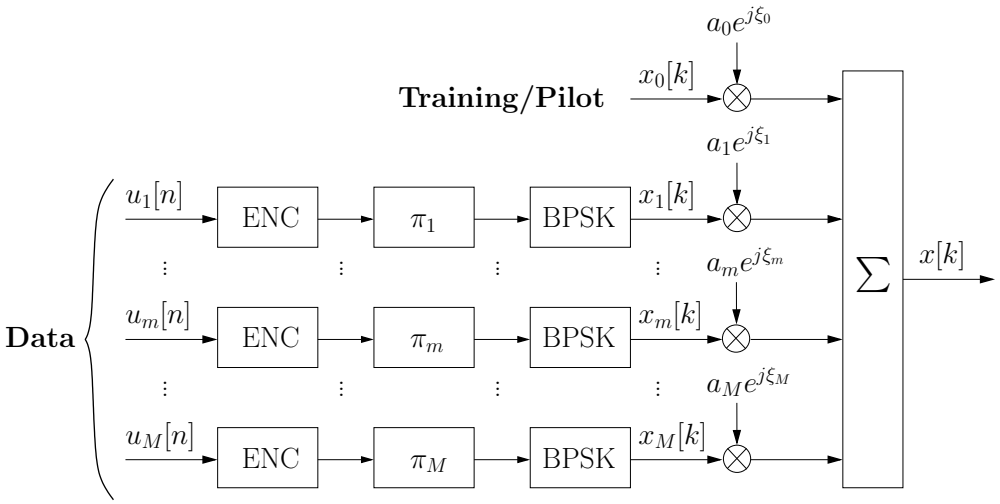


Figure 3.7: Transmitter structure of IDM with PLACE.

tioning system and is, thus, a major concern of this thesis. Details on how CPE can be performed are provided in Chapter 4.

3.4 Chapter Summary

In the beginning of this chapter, it is deduced that a basic requirement for a joint communication and positioning system is a *flexible configuration* since it is very challenging to reliably provide high data rates with low BER and accurate positioning at the same time. Furthermore, it is demonstrated that *multilateration via the TOA* and *IDM* are suitable techniques for a joint communication and positioning system: TOA is chosen since it enables precise localisation and IDM assures a flexible configuration while offering a simple transmitter structure. The data rates can be easily adapted by changing the number of layers M and unequal error protection can be performed by assigning different amplitude levels to different layers. Furthermore, IDM is potentially capacity-achieving due to Gaussian-like symbols. The connection between communication and positioning is accomplished via PLACE: A pilot layer, that carries only training symbols, is superimposed to the data layers. Based on this pilot layer, channel estimation as well as TOA estimation can be performed. Since the channel coefficients and the position-related TOA are estimated simultaneously, the overall estimation process is called channel parameter

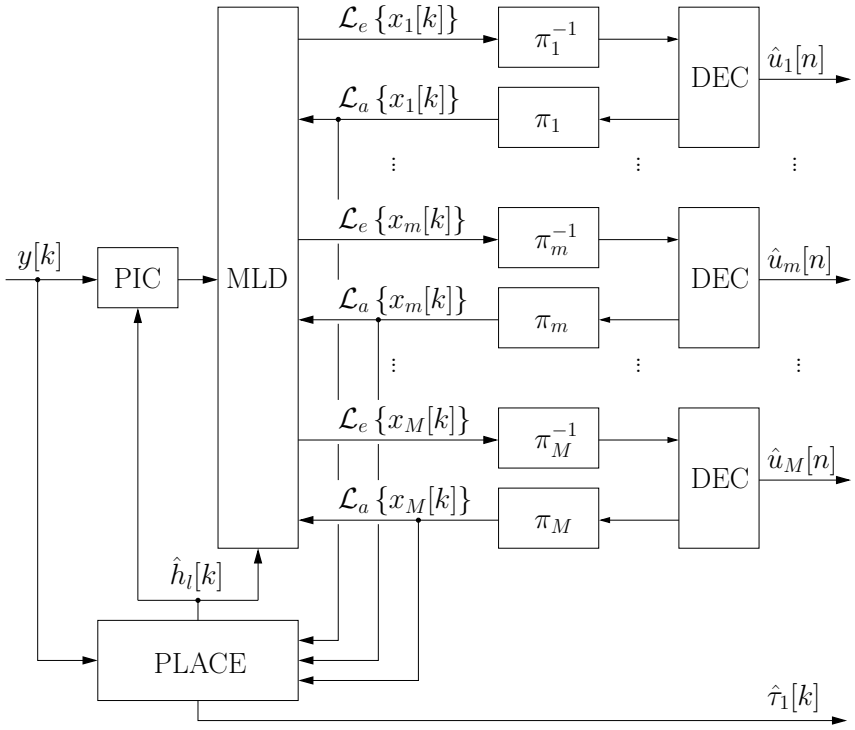


Figure 3.8: Receiver structure of IDM with PLACE.

estimation. The tradeoff between communication and positioning can be controlled by the parameter ρ, which denotes the ratio between the pilot layer power to the total power. Furthermore, not all data layers have to be devoted to communication purposes. Some of them may also be used to transmit auxiliary information for localisation purposes such as the TOD or the position of the corresponding ROs in case of self-positioning. Therefore, the proposed joint communication and positioning system is highly flexible. This system proposal has already been published in [HS07] and [SH08].

Chapter 4

Channel Parameter Estimation

As already mentioned in Section 3.3, the estimation process performed in the PLACE unit is called *channel parameter estimation* since channel and parameter estimation are performed simultaneously. The channel estimates of the equivalent discrete-time channel model are needed for data detection and the TOA is needed for positioning. Hence, channel parameter estimation is the core part of the proposed joint communication and positioning system. This and the following chapter focus on *positioning only*, where this chapter is devoted to the first localisation step, namely parameter estimation, and Chapter 5 deals with the second localisation step, namely position estimation (see also Figure 3.1). In these chapters, effects that may be caused by interference from the data layers are excluded. Therefore, data transmission and detection are neglected such that pure training is transmitted (ρ=1). This corresponds to the best possible scenario for positioning. However, not only the accuracy of the parameter estimates, but also the accuracy of the channel estimates is investigated in the following. The overall system performance *including* data detection is examined in Chapter 6.

Generally, it is a challenging task to combine communication and positioning, but fortunately channel and parameter estimation are closely related: The functional relationship between the parameters of the physical channel (propagation/excess delays and complex amplitudes) and the channel coefficients of the equivalent discrete-time channel model is described by (2.40) and (2.47). If the overall pulse shape $g(\tau)$ (including pulse shaping and receive filtering) is known, the channel coefficients can be computed directly from the channel parameters. This means that, under the condition that a priori information about pulse shaping and receive filtering is available, channel estimates are obtained inherently if the parameters of the physical channel are estimated. Therefore, channel parameter estimation enables positioning *and* data detection. Typically, only *one* aspect of channel parameter estimation is considered: On the one hand, channel parameter estimation is well known in the context of channel sounding [FW88], [FTH+99], [RLT03]. In this case, the parameters of the physical channel are of interest, while the channel coefficients of the equivalent discrete-time channel model, which are available as well, are not further processed. On the other hand, the usage of a priori information about pulse shaping and

receive filtering has already been suggested in [KRBK97], [LNCP97], [LP98] for improved channel estimation in communication systems. In this case, the information about the physical channel is discarded. For joint communication and positioning, *both* aspects of channel parameter estimation are exploited.

The channel parameter estimator considered in this thesis is based on the *maximum-likelihood (ML)* principle. There are two equivalent approaches: On the one hand, the channel parameters can be estimated directly from the received samples as it is often the case for channel sounding [FW88], [FTH+99], [RLT03]. Similar estimators have been investigated for example in the context of rake receivers in code-division multiple access [FSRSM02] or for pure navigation purposes [ANU08]. On the other hand, channel parameter estimation can be performed in two steps: First, standard channel estimation (without a priori information about pulse shaping and receive filtering) is applied in order to obtain a preliminary estimate of the channel coefficients. Based on this pre-stage channel estimates, the parameters of the physical channel are estimated and *enhanced* channel estimates are obtained. This two-step approach is for example considered in [vdVVP98]. Similar to the channel estimation approaches in [KRBK97], [LNCP97], [LP98], the channel estimates which are available after channel parameter estimation are typically more reliable than the pre-stage channel estimates due to the exploitation of a priori information about pulse shaping and receive filtering. Both approaches are considered subsequently and their relationship is explained.

The remainder of this chapter is organised as follows: In Section 4.1, the fundamentals needed for channel parameter estimation are summarised. The underlying channel model is reviewed and the required equations are presented and adjusted to channel parameter estimation. The proposed ML estimator, which can be either based on the received samples or the pre-stage channel estimates, is introduced in Section 4.2. For each approach, a nonlinear metric is derived and the relationship between both metrics is discussed. The best possible performance of the proposed estimator is given by the Cramer-Rao lower bound (CRLB), which is determined in Section 4.3. The CRLBs of the different channel models introduced in Section 2.5 are compared and the influence of oversampling is investigated. Since the derived metrics are nonlinear, they can not be solved in closed form. Hence, several optimisation methods are presented in Section 4.4. If soft information about the parameter estimates is available, it can be exploited to improve the position estimate. In Section 4.5, the concept of soft information is shortly reviewed and different methods to obtain soft information concerning the parameter estimates are presented. Finally, the most important aspects of this chapter are summarised in Section 4.6.

4.1 Fundamentals

It has already been discussed in Section 2.5 that the assumption of perfect synchronisation, which is often applied for the simulation of communication systems, is not applicable in the framework of the proposed joint communication and positioning system: The prop-

agation delay of the first arriving path (TOA) needs to be estimated and can not be assumed to be known. Thus, a coarse synchronisation, which is modelled by a non-zero sampling phase $\varepsilon \in [-\Delta\tau_1, +\Delta\tau_1]$, is considered, i.e. a coarse pre-estimate of the TOA, $\check{\tau}_1 = \tau_1 + \varepsilon$, is obtained during the synchronisation process, which is subtracted from the true propagation delays ($\tau_c - \check{\tau}_1 = \nu_c - \varepsilon$). This leads to the channel coefficients described by (2.47), which depend on the excess delays ν_c and the sampling phase ε:

$$h_l[k] = \sum_{c=1}^{C} f_c[k] \cdot g\left(l\mathcal{T} + \varepsilon - \nu_c\right).$$

In order to obtain a refined TOA estimate $\hat{\tau}_1$, the sampling phase needs to be estimated: $\hat{\tau}_1 = \check{\tau}_1 - \hat{\varepsilon} = \tau_1 + \varepsilon - \hat{\varepsilon}$. Thus, the estimation of the TOA basically corresponds to the estimation of the sampling phase. For the purpose of sampling phase estimation, the channel model in (2.47) is reformulated by combining the excess delays with the sampling phase to auxiliary parameters $\varrho_c = \nu_c - \varepsilon$, which are termed *coarse excess delays* in the following. Furthermore, *block fading*[1] is assumed, i.e. the parameters of the physical channel do not change over the transmission burst (see Section 2.3.2). In this case, the channel coefficients do not depend on the time index k anymore:

$$h_l = \sum_{c=1}^{C} f_c \cdot g\left(l\mathcal{T} - \varrho_c\right). \tag{4.1}$$

In order to emphasise the functional relationship between the parameters of the physical channel and the channel coefficients of the equivalent discrete-time channel model, the channel parameters are stacked in a vector

$$\boldsymbol{\theta} = \left[\mathrm{Re}\left\{f_1\right\}, \mathrm{Im}\left\{f_1\right\}, \varrho_1, \cdots, \mathrm{Re}\left\{f_C\right\}, \mathrm{Im}\left\{f_C\right\}, \varrho_C\right]^T = \left[\theta_1, \ldots, \theta_P\right]^T \tag{4.2}$$

of length $P = 3C$. Each propagation path is characterised by three parameters: the real and imaginary part of the complex amplitude, $\mathrm{Re}\{f_c\}$ and $\mathrm{Im}\{f_c\}$, and the coarse excess delay ϱ_c. Given the parameter vector $\boldsymbol{\theta}$, the channel coefficients can be determined according to

$$h_l(\boldsymbol{\theta}) = \sum_{\substack{p=1 \\ p+=3}}^{P} \left(\theta_p + j\theta_{p+1}\right) g\left(l\mathcal{T} - \theta_{p+2}\right). \tag{4.3}$$

Together with the assumption that only training is transmitted ($\rho = 1$), the received samples are described by

$$y[k] = \sum_{l=0}^{L} h_l(\boldsymbol{\theta}) \cdot x_0[k-l] + n[k], \qquad 0 \le k < K + L. \tag{4.4}$$

Based on these observations, the parameter vector $\boldsymbol{\theta}$ can be estimated via the *maximum-likelihood (ML)* approach exploiting a priori information about the pilot symbols $x_0[k]$

[1] A brief discussion on the validity of the block fading assumption is given further down.

and the overall pulse shape $g(\tau)$ (including pulse shaping and receive filtering). For the derivation of the ML estimator, it is useful to express (4.4) in vector/matrix notation:

$$\boldsymbol{y} = \boldsymbol{X}_0 \boldsymbol{h}(\boldsymbol{\theta}) + \boldsymbol{n}, \tag{4.5}$$

where $\boldsymbol{y} = [y[L], \ldots, y[K-1]]^T$ is the observation vector containing the received samples and $\boldsymbol{X}_0$ is the training matrix with Toeplitz structure:

$$\boldsymbol{X}_0 = \begin{bmatrix} x_0[L] & x_0[L-1] & \cdots & x_0[0] \\ x_0[L+1] & x_0[L] & \cdots & x_0[1] \\ \vdots & \vdots & \ddots & \vdots \\ x_0[K-1] & x_0[K-2] & \cdots & x_0[K-L-1] \end{bmatrix}. \tag{4.6}$$

Furthermore, $\boldsymbol{h}(\boldsymbol{\theta})=[h_0(\boldsymbol{\theta}), \cdots, h_L(\boldsymbol{\theta})]^T$ and $\boldsymbol{n}=[n[L], \ldots, n[K-1]]^T$ denote the channel coefficient vector and a zero mean Gaussian noise vector with covariance matrix $\boldsymbol{C_n}$, respectively. Generally, the entries of the noise covariance matrix are determined according to the autocorrelation function of the noise process given in (2.44):

$$[\boldsymbol{C_n}]_{i,j} = \sigma_n^2 \, \psi_{\mathrm{Rx}} \left((i-j) \mathcal{T} \right). \tag{4.7}$$

In case of white noise, the covariance matrix corresponds to a scaled identity matrix: $\boldsymbol{C_n}=\boldsymbol{\mathcal{I}}\sigma_n^2$.

Equations (4.1) to (4.5) are based on the assumption of block fading. Of course, the question arises if this assumption is adequate or not. Therefore, the block fading assumption is examined in more detail at this point: As defined in Section 2.3.2, block fading can be assumed for $B_D \cdot K' T_s \ll 1$, where B_D is the Doppler spread, K' denotes number of symbols per burst and T_s is the symbol duration. The Doppler spread depends on the mobility of the investigated scenario and can be equated with the maximum Doppler shift $B_D=f_{D,\max}=v/\lambda$. In this case, v represents the maximum possible velocity. The limit for block fading is assumed to be $B_D \cdot K' T_s \leq 0.01$ in this thesis. In that case, the maximum possible velocity that can be supported by a system to ensure block fading can be computed as:

$$v \leq 0.01 \frac{\lambda}{K' T_s} = 0.01 \frac{c}{f_0 K' T_s}, \tag{4.8}$$

where f_0 is the carrier frequency.[2] The maximum supportable velocity increases with decreasing system parameters (f_0, K', T_s) as illustrated in Figure 4.1, in which the carrier frequency is fixed to f_0=2 GHz and the symbol duration T_s as well as the burst length K' are varied. For T_s=100 ns and K'=500, velocities up to 108 km/h can be supported, whereas only velocities up to 5 km/h are possible if T_s=900 ns and K'=1200. A prominent example is UMTS, in which block fading can be assumed for velocities smaller than approximately 8.1 km/h since T_s=260 ns and K'=2560. In this thesis, mainly indoor and urban scenarios are of interest. Hence, the maximum velocities that typically occur lie

[2] Please note that this general relationship does not dependent on the oversampling factor J due to $K \cdot \mathcal{T} = (JK') \cdot (T_s/J) = K' \cdot T_s$.

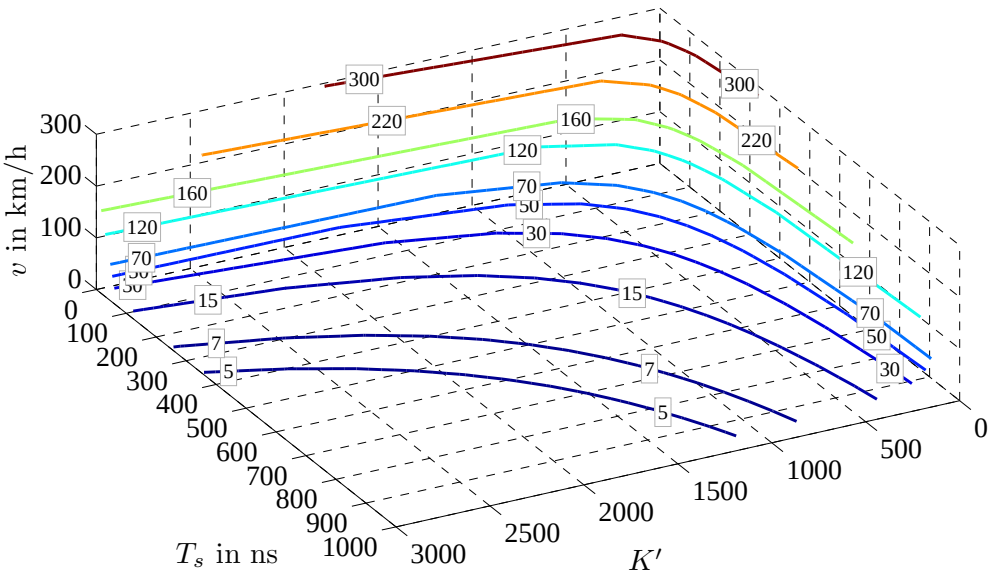


Figure 4.1: Maximum possible velocity v to ensure block fading over burst length K' and symbol duration T_s with carrier frequency f_0=2 GHz (λ=15 cm).

between 7 km/h and 50 km/h. Thus, a wide range of reasonable parameter combinations exists, for which the block fading assumption is valid. If the parameters are fixed to specific values, that can not be changed, and the block fading assumption is violated, the transmission burst can be subdivided into smaller blocks, for which the block fading assumption is valid again. In this case, channel parameter estimation can be performed block- instead of burst-wise (sliding window approach, see also [SH06]). Consequently, the assumption of block fading is adequate and hardly restricts the applicability of the proposed channel parameter estimator.

As a concluding remark, it should be noted here that the *only* parameter, that is needed for positioning, is $\hat{\theta}_3{=}\hat{\varrho}_1{=}{-}\hat{\varepsilon}$. All remaining $P{-}1$ parameters act as *nuisance parameters* and are not relevant for positioning. However, the whole parameter estimate $\hat{\boldsymbol{\theta}}$ is needed to obtain channel estimates $\hat{\boldsymbol{h}}{=}\boldsymbol{h}(\hat{\boldsymbol{\theta}})$.

4.2 Maximum-Likelihood Estimator

The maximum-likelihood (ML) approach is very popular in statistical signal processing and related fields. An overview about ML estimation is for example given in [Sch91, Chap. 6] and [Kay93, Chap. 7], on which the following section is mainly based. ML estimators exhibit some nice properties, which justify their popularity. Basically, ML estimators are asymptotically *optimal* (*unbiased* and *efficient*), i.e. they achieve the best

possible performance given by the CRLB (see also Section 4.3) for a large number of observations or at high signal-to-noise ratios. For linear models, the ML estimator is the *minimum variance unbiased (MVU)* estimator [Kay93, p. 186] or more generally: "If an MVU estimator exists, the maximum-likelihood procedure will produce it" [Kay93, p. 482]. Furthermore, the ML estimator corresponds to the least-squares (LS) estimator in case of Gaussian noise [Kay93, p. 254, 483]. In the following, the ML estimator for channel parameter estimation is examined based on the signal model (4.5),

$$\boldsymbol{y} = \boldsymbol{X}_0 \boldsymbol{h}(\boldsymbol{\theta}) + \boldsymbol{n},$$

in which $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{C_n})$ with $\boldsymbol{C_n}$ given by (4.7). Two nonlinear metrics are derived and their interrelationship is explained.

4.2.1 Metric with Received Samples

According to [Kay93], the ML estimate $\hat{\boldsymbol{\theta}}$ is given by the value of $\boldsymbol{\theta}$ that maximises the *likelihood function* $p(\boldsymbol{y}; \boldsymbol{\theta})$. As the logarithm is monotonically increasing, the maximisation of the *log-likelihood function* $\ln p(\boldsymbol{y}; \boldsymbol{\theta})$ is equivalent to maximising the likelihood function:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\tilde{\boldsymbol{\theta}}} \left\{ p(\boldsymbol{y}, \tilde{\boldsymbol{\theta}}) \right\} = \arg\max_{\tilde{\boldsymbol{\theta}}} \left\{ \ln p(\boldsymbol{y}, \tilde{\boldsymbol{\theta}}) \right\}. \tag{4.9}$$

Since the received samples are distorted by complex Gaussian noise with covariance matrix $\boldsymbol{C_n}$, the likelihood function, that needs to be maximised, is given by a multivariate complex Gaussian distribution:

$$p(\boldsymbol{y}, \boldsymbol{\theta}) = \frac{1}{\pi^{K-L} \det(\boldsymbol{C_n})} \exp\left(- \left(\boldsymbol{y} - \boldsymbol{X}_0 \boldsymbol{h}(\boldsymbol{\theta}) \right)^H \boldsymbol{C_n}^{-1} \left(\boldsymbol{y} - \boldsymbol{X}_0 \boldsymbol{h}(\boldsymbol{\theta}) \right) \right). \tag{4.10}$$

Taking the logarithm of the above function leads to

$$\ln p(\boldsymbol{y}, \boldsymbol{\theta}) = - \ln\left(\pi^{K-L} \det(\boldsymbol{C_n}) \right) - \left(\boldsymbol{y} - \boldsymbol{X}_0 \boldsymbol{h}(\boldsymbol{\theta}) \right)^H \boldsymbol{C_n}^{-1} \left(\boldsymbol{y} - \boldsymbol{X}_0 \boldsymbol{h}(\boldsymbol{\theta}) \right), \tag{4.11}$$

where only the second part of the right hand side depends on the parameter vector $\boldsymbol{\theta}$. Thus, the first part of the right hand side is not relevant for maximisation and can be neglected. By changing the leading minus sign into a plus sign, maximisation is transformed to minimisation. Therefore, the ML estimate of the parameter vector is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} \left\{ \left(\boldsymbol{y} - \boldsymbol{X}_0 \boldsymbol{h}(\tilde{\boldsymbol{\theta}}) \right)^H \boldsymbol{C_n}^{-1} \left(\boldsymbol{y} - \boldsymbol{X}_0 \boldsymbol{h}(\tilde{\boldsymbol{\theta}}) \right) \right\} = \arg\min_{\tilde{\boldsymbol{\theta}}} \left\{ \Omega_{\boldsymbol{y}}(\tilde{\boldsymbol{\theta}}) \right\}, \tag{4.12}$$

where $\Omega_{\boldsymbol{y}}(\boldsymbol{\theta})$ denotes the metric that has to be minimised. Due to the pulse shape $g(\tau)$, the function $\boldsymbol{h}(\boldsymbol{\theta})$ is nonlinear and, thus, the overall metric $\Omega_{\boldsymbol{y}}(\boldsymbol{\theta})$ is nonlinear as well. This means that the minimisation of $\Omega_{\boldsymbol{y}}(\boldsymbol{\theta})$ can not be solved in closed form as in the linear case. Exhaustive search is prohibitive because the search space (parameter space) is continuous and of high dimension (P=3C). Hence, an optimisation method has to be applied. Further details on optimisation algorithms are provided in Section 4.4.

4.2.2 Metric with Pre-Stage Channel Estimates

In addition to the metric $\Omega_{\boldsymbol{y}}(\boldsymbol{\theta})$, which is based on the received samples, a second metric can be derived for channel parameter estimation. In this case, channel parameter estimation is performed in two steps: At first, *pre-stage channel estimates* $\check{\boldsymbol{h}}$ are obtained via a standard channel estimation algorithm. In a second step, the parameters of the physical channel $\boldsymbol{\theta}$ are estimated by fitting the model function $\boldsymbol{h}(\boldsymbol{\theta})$ to these pre-stage channel estimates. Since block fading is assumed, *(weighted) least-squares channel estimation (LSCE)* can be performed. The signal model in (4.5) is linear concerning the channel coefficients. Thus, the channel estimates can be obtained in closed form according to [Kay93, p. 225f.]:

$$\begin{aligned}\check{\boldsymbol{h}} &= \left(\boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{X}_0\right)^{-1} \boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{y} \\ &= \underbrace{\left(\boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{X}_0\right)^{-1} \boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{X}_0}_{\boldsymbol{\mathcal{I}}} \boldsymbol{h}(\boldsymbol{\theta}) + \underbrace{\left(\boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{X}_0\right)^{-1} \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{X}_0^H \boldsymbol{n}}_{\boldsymbol{\eta}},\end{aligned} \tag{4.13}$$

where $\boldsymbol{\eta} \sim \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\eta}}\right)$ is the channel estimation error with covariance matrix

$$\boldsymbol{C}_{\boldsymbol{\eta}} = \left(\boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{X}_0\right)^{-1}. \tag{4.14}$$

Based on this simplified signal model, the parameter vector $\boldsymbol{\theta}$ can be estimated according to the ML approach by maximising the likelihood function $p(\check{\boldsymbol{h}}, \boldsymbol{\theta})$ or the log-likelihood function $\ln p(\check{\boldsymbol{h}}, \boldsymbol{\theta})$:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\tilde{\boldsymbol{\theta}}} \left\{ p(\check{\boldsymbol{h}}, \tilde{\boldsymbol{\theta}}) \right\} = \arg\max_{\tilde{\boldsymbol{\theta}}} \left\{ \ln p(\check{\boldsymbol{h}}, \tilde{\boldsymbol{\theta}}) \right\}. \tag{4.15}$$

Again, the likelihood function is given by a multivariate complex Gaussian distribution:

$$p(\check{\boldsymbol{h}}, \boldsymbol{\theta}) = \frac{1}{\pi^{L+1} \det(\boldsymbol{C}_{\boldsymbol{\eta}})} \exp\left(- \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\boldsymbol{\theta})\right)^H \boldsymbol{C}_{\boldsymbol{\eta}}^{-1} \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\boldsymbol{\theta})\right) \right). \tag{4.16}$$

Hence, the parameter estimate can be determined according to

$$\hat{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} \left\{ \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)^H \boldsymbol{C}_{\boldsymbol{\eta}}^{-1} \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right) \right\} = \arg\min_{\tilde{\boldsymbol{\theta}}} \left\{ \Omega_{\check{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}}) \right\} \tag{4.17}$$

by minimising the metric $\Omega_{\check{\boldsymbol{h}}}(\boldsymbol{\theta})$, that is based on the pre-stage channel estimates. This metric is also nonlinear and has to be solved by an optimisation algorithm as presented in Section 4.4.

4.2.3 Relation Between Both Metrics

Two different metrics, $\Omega_{\boldsymbol{y}}(\boldsymbol{\theta})$ and $\Omega_{\check{\boldsymbol{h}}}(\boldsymbol{\theta})$, have been derived above. The question is whether these metrics are related and, if this is the case, how they are related? The answer is that both metrics are equivalent and lead to the same parameter estimates as will be explained

in the following. At first, both metrics are rearranged to show their similarity. Here, $\tilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$ denote the hypothesis of the parameter vector and its true value, respectively. The first metric can be represented as

$$\begin{aligned}\Omega_{\boldsymbol{y}}(\tilde{\boldsymbol{\theta}}) &= \left(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)^H \boldsymbol{C}_{\boldsymbol{n}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)\\ &= \left(\boldsymbol{X}_0\boldsymbol{h}(\boldsymbol{\theta})+\boldsymbol{n}-\boldsymbol{X}_0\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)^H \boldsymbol{C}_{\boldsymbol{n}}^{-1}\left(\boldsymbol{X}_0\boldsymbol{h}(\boldsymbol{\theta})+\boldsymbol{n}-\boldsymbol{X}_0\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)\\ &= \left(\boldsymbol{X}_0\left(\boldsymbol{h}(\boldsymbol{\theta})-\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)+\boldsymbol{n}\right)^H \boldsymbol{C}_{\boldsymbol{n}}^{-1}\left(\boldsymbol{X}_0\left(\boldsymbol{h}(\boldsymbol{\theta})-\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)+\boldsymbol{n}\right)\end{aligned} \tag{4.18}$$

and the second metric via

$$\begin{aligned}\Omega_{\check{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}}) &= \left(\check{\boldsymbol{h}}-\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)^H \boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\left(\check{\boldsymbol{h}}-\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)\\ &= \left(\boldsymbol{h}(\boldsymbol{\theta})+\boldsymbol{\eta}-\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)^H \boldsymbol{X}_0^H\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{X}_0\left(\boldsymbol{h}(\boldsymbol{\theta})+\boldsymbol{\eta}-\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)\\ &= \left(\boldsymbol{X}_0\left(\boldsymbol{h}(\boldsymbol{\theta})-\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)+\boldsymbol{X}_0\boldsymbol{\eta}\right)^H \boldsymbol{C}_{\boldsymbol{n}}^{-1}\left(\boldsymbol{X}_0\left(\boldsymbol{h}(\boldsymbol{\theta})-\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)+\boldsymbol{X}_0\boldsymbol{\eta}\right).\end{aligned} \tag{4.19}$$

The comparison of (4.18) and (4.19) shows that both metrics only differ in their distortion part, $\boldsymbol{n}$ and $\boldsymbol{X}_0\boldsymbol{\eta}$, respectively. To understand why both metrics lead to the same parameter estimate, it is helpful to describe the estimation process from a geometrical point of view. Due to the Gaussian noise, the ML estimator corresponds to the LS estimator such that the geometrical interpretation of LS problems as given in [BW80] and [Kay93, p. 226-232] can be applied. In the following, white noise is considered for simplicity. However, the obtained results also hold for the general case with coloured Gaussian noise since the insertion of the (coloured) covariance matrix $\boldsymbol{C}_{\boldsymbol{n}}$ corresponds to a whitening approach [Kay93, p. 94f.]. In case of white Gaussian noise, the noise covariance matrix $\boldsymbol{C}_{\boldsymbol{n}}$ is a (scaled) identity matrix and both metrics can be rewritten as

$$\Omega_{\boldsymbol{y}}(\tilde{\boldsymbol{\theta}}) = \left|\left|\boldsymbol{X}_0\left(\boldsymbol{h}(\boldsymbol{\theta})-\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)+\ \boldsymbol{n}\ \right|\right|^2 = \left|\left|\ \ \boldsymbol{y}\ -\boldsymbol{X}_0\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right|\right|^2, \tag{4.20}$$

$$\Omega_{\check{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}}) = \left|\left|\boldsymbol{X}_0\left(\boldsymbol{h}(\boldsymbol{\theta})-\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)+\boldsymbol{X}_0\boldsymbol{\eta}\right|\right|^2 = \left|\left|\boldsymbol{X}_0\check{\boldsymbol{h}}-\boldsymbol{X}_0\boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right|\right|^2. \tag{4.21}$$

Geometrically, (4.20) and (4.21) describe the distance between the observations, given either by $\boldsymbol{y}$ or $\boldsymbol{X}_0\check{\boldsymbol{h}}$, and the hypothesis $\boldsymbol{X}_0\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ in the $(K-L)$-dimensional *observation space*[3]. If the hypothesis of the parameter vector $\tilde{\boldsymbol{\theta}}$ is changed in the P-dimensional *parameter space*[4], the vector $\boldsymbol{X}_0\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ traces a P-dimensional subspace in the $(K-L)$-dimensional observation space. This subspace is called *solution locus*. Hence, the function $\boldsymbol{X}_0\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ maps all feasible parameters in the P-dimensional parameter space to the P-dimensional solution locus in the $(K-L)$-dimensional observation space. Due to the distortion by $\boldsymbol{n}$ or $\boldsymbol{X}_0\boldsymbol{\eta}$, the observations typically do not lie on the solution locus, but anywhere in the observation space. The parameter estimate $\hat{\boldsymbol{\theta}}$ corresponds to the point on the solution locus, $\boldsymbol{X}_0\boldsymbol{h}(\hat{\boldsymbol{\theta}})$, with the smallest distance to the observations [BW80],

[3] The observation space is called *sample space* as well.

[4] The parameter space is called *search space* as well.

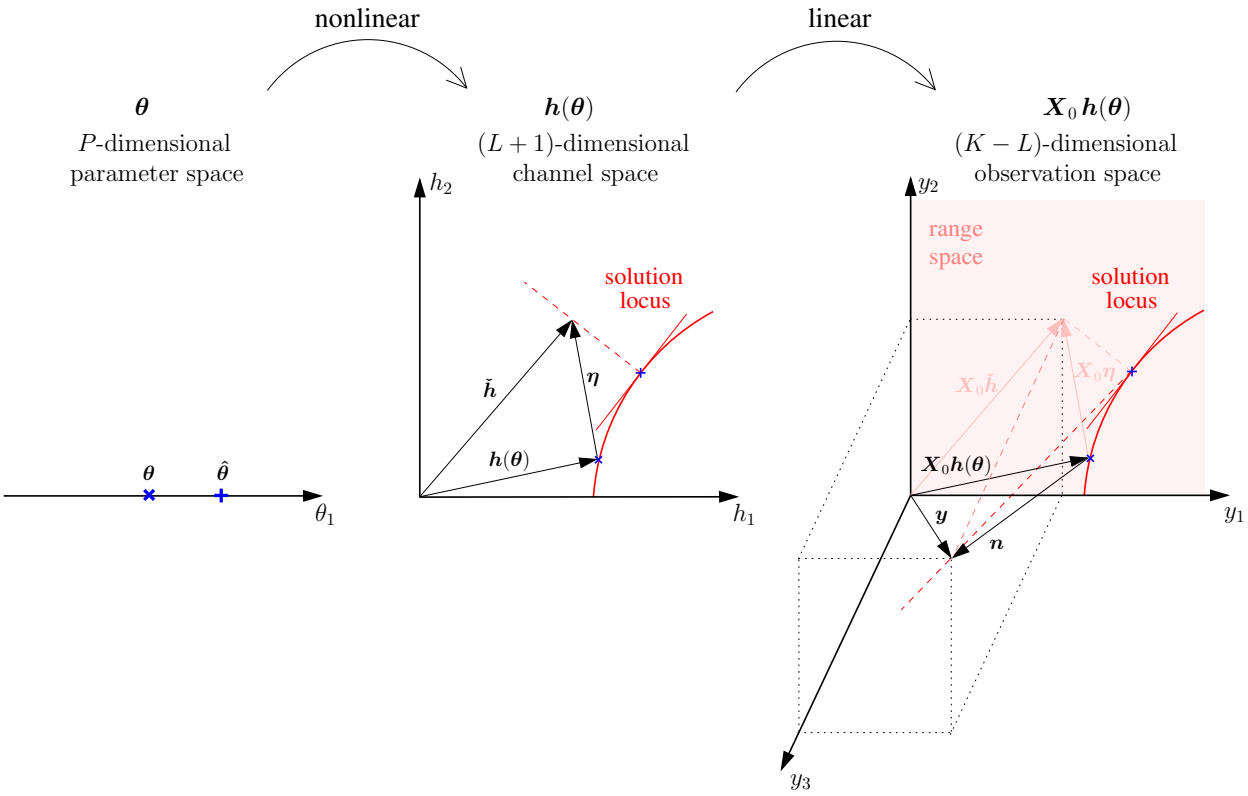


Figure 4.2: Simplified example for the transformation from $\boldsymbol{\theta}$ over $\boldsymbol{h}(\boldsymbol{\theta})$ to $\boldsymbol{X}_0\boldsymbol{h}(\boldsymbol{\theta})$.

[Kay93, p. 227]. That implies that the error vectors, $(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{h}(\hat{\boldsymbol{\theta}}))$ and $(\boldsymbol{X}_0\check{\boldsymbol{h}}-\boldsymbol{X}_0\boldsymbol{h}(\hat{\boldsymbol{\theta}}))$, are *orthogonal* to the solution locus [Kay93, p. 228].

The overall transformation from the parameter space to the observation space is nonlinear due to the pulse shape $g(\tau)$. Thus, the solution locus will not be planar but curved. However, the overall transformation can be divided into a linear part described by the matrix $\boldsymbol{X}_0$ and a nonlinear part determined by the function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$. Thus, an intermediate state can be observed in the $(L+1)$-dimensional space of the channel coefficients, which is called *channel space* in the following. At first, the function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ maps all feasible parameters to a solution locus in the channel space. Afterwards, the channel space is mapped into the observation space. That means that the solution locus is a P-dimensional subspace of the channel space, which is itself an $(L+1)$-dimensional subspace of the $(K-L)$-dimensional observation space. Since the dimensionality of the different spaces is typically larger than three (e.g. $P=6$, $L+1=11$ and $K-L=90$), it is difficult to visualise the above concepts. Therefore, a simplified example is illustrated in Figure 4.2.

At this point, the second transformation between the channel space and the observation space, which corresponds to LSCE, is examined in more detail: All feasible channel coefficient vectors are mapped onto an $(L+1)$-dimensional subspace in the observation space, which is the solution locus for channel estimation. This solution locus is planar since

the mapping from the channel space to the observation space is linear. In this case, the pre-stage channel estimates in the observation space given by $\boldsymbol{X}_0\check{\boldsymbol{h}}=\boldsymbol{X}_0(\boldsymbol{X}_0^H\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0^H\boldsymbol{y}$ are the *orthogonal projection* of $\boldsymbol{y}$ into the solution locus of channel estimation, where the matrix

$$\boldsymbol{P}_0 = \boldsymbol{X}_0 \left(\boldsymbol{X}_0^H \boldsymbol{X}_0\right)^{-1} \boldsymbol{X}_0^H \tag{4.22}$$

is called *projection matrix* [Kay93, p. 228ff.]. Projection matrices have some special properties: They are *symmetric* ($\boldsymbol{P}_0^T=\boldsymbol{P}_0$), *idempotent* ($\boldsymbol{P}_0^2=\boldsymbol{P}_0$) and *singular* (i.e. the projection is non-bijective such that several observations are mapped to the same estimate) [GMW91, p. 244ff.], [Kay93, p. 231]. The solution locus for channel estimation corresponds to the *range space* of $\boldsymbol{X}_0$, whose $(K-2L-1)$-dimensional orthogonal complement is called *null space* [GMW91, p. 245f.]. The error vector $(\boldsymbol{y}-\boldsymbol{X}_0\check{\boldsymbol{h}})$ is orthogonal to the range space and, hence, lies in the null space. Every vector in the observation space, including the noise vector $\boldsymbol{n}$, can be decomposed into a range space and a null space component:

$$\boldsymbol{n} = \boldsymbol{n}_R + \boldsymbol{n}_N = \boldsymbol{P}_0\,\boldsymbol{n} + (\boldsymbol{\mathcal{I}} - \boldsymbol{P}_0)\,\boldsymbol{n}. \tag{4.23}$$

The range space component of the noise vector corresponds to the channel estimation error (see also Figure 4.2):

$$\boldsymbol{P}_0\,\boldsymbol{n} = \boldsymbol{X}_0 \left(\boldsymbol{X}_0^H \boldsymbol{X}_0\right)^{-1} \boldsymbol{X}_0^H \boldsymbol{n} = \boldsymbol{X}_0 \boldsymbol{\eta}. \tag{4.24}$$

As could be seen from the comparison of (4.18) and (4.19), both metrics only differ in their distortion part, $\boldsymbol{n}=\boldsymbol{n}_R+\boldsymbol{n}_N$ and $\boldsymbol{X}_0\boldsymbol{\eta}=\boldsymbol{n}_R$, respectively. Now, it is clear why both metrics lead to the same estimate: The null space component of the noise vector $\boldsymbol{n}_N$ is by definition orthogonal to the range space of $\boldsymbol{X}_0$ and, hence, it is always orthogonal to the solution locus of parameter estimation since this solution locus is a subspace of the range space of $\boldsymbol{X}_0$ (see also Figure 4.2). Only the range space component of the noise $\boldsymbol{n}_R$ influences the parameter estimate $\hat{\boldsymbol{\theta}}$. As a result, both metrics lead to the same point on the solution locus but with different residuals. The expectation of the residuals for $\Omega_{\boldsymbol{y}}(\tilde{\boldsymbol{\theta}})$ is given by

$$\mathrm{E}\left\{\boldsymbol{n}^H\boldsymbol{n}\right\} = (K-L)\,\sigma_n^2. \tag{4.25}$$

With the assumption that $L+1$ and $K-L$ are large, the expectation of the residuals of $\Omega_{\check{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}})$ results in

$$\begin{aligned}
\mathrm{E}\left\{\boldsymbol{\eta}^H\boldsymbol{X}_0^H\boldsymbol{X}_0\boldsymbol{\eta}\right\} &= \mathrm{E}\left\{\boldsymbol{n}^H \underbrace{\boldsymbol{P}_0^H\boldsymbol{P}_0}_{\boldsymbol{P}_0^2=\boldsymbol{P}_0}\boldsymbol{n}\right\} \\
&= \mathrm{E}\left\{\boldsymbol{n}^H\boldsymbol{X}_0\big(\underbrace{\boldsymbol{X}_0^H\boldsymbol{X}_0}_{(K-L)\boldsymbol{\mathcal{I}}}\big)^{-1}\boldsymbol{X}_0^H\boldsymbol{n}\right\} \\
&= \frac{1}{K-L}\,\mathrm{E}\left\{\boldsymbol{n}^H \underbrace{\boldsymbol{X}_0\boldsymbol{X}_0^H}_{(L+1)\boldsymbol{\mathcal{I}}}\boldsymbol{n}\right\} \\
&= \frac{L+1}{K-L}\,\mathrm{E}\left\{\boldsymbol{n}^H\boldsymbol{n}\right\} = (L+1)\,\sigma_n^2.
\end{aligned} \tag{4.26}$$

If $K-L$ equals $L+1$, the observation space and the channel space have the same dimension, i.e. the range space of $\boldsymbol{X}_0$ corresponds to the whole observation space. In this case, $\boldsymbol{X}_0$ is a square matrix such that the distortion terms as well as the residuals are the same for both metrics:

$$\boldsymbol{X}_0\boldsymbol{\eta} = \boldsymbol{X}_0 \left(\boldsymbol{X}_0^H \boldsymbol{X}_0\right)^{-1} \boldsymbol{X}_0^H \boldsymbol{n} = \underbrace{\boldsymbol{X}_0 \left(\boldsymbol{X}_0\right)^{-1}}_{\boldsymbol{\mathcal{I}}} \underbrace{\left(\boldsymbol{X}_0^H\right)^{-1} \boldsymbol{X}_0^H}_{\boldsymbol{\mathcal{I}}} \boldsymbol{n} = \boldsymbol{n}. \tag{4.27}$$

Since both metrics are equivalent, it is sufficient to consider only one of them in the following. As already mentioned above, both metrics are nonlinear and have to be solved by an optimisation method, which typically performs the minimisation/maximisation in an iterative manner (see Section 4.4). This means that the metric has to be computed in every iteration, i.e. the metric whose computation is less complex is advantageous from a complexity point of view. In general, $(L+1)$ is smaller than $(K-L)$. Thus, the vectors and matrices involved in $\Omega_{\check{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}})$ are smaller than those involved in $\Omega_{\boldsymbol{y}}(\tilde{\boldsymbol{\theta}})$, i.e. less multiplications and additions have to be performed in order to compute the first metric. The complexity of pre-stage channel estimation is negligible since the pre-stage channel estimates $\check{\boldsymbol{h}}$ are obtained in closed form. Furthermore, the pseudoinverse $(\boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{X}_0)^{-1} \boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1}$ can be computed in advance since the pilot symbols $x_0[k]$ and the receive filter $g_{\mathrm{Rx}}(\tau)$ are known. Therefore, if not stated otherwise, $\Omega(\tilde{\boldsymbol{\theta}}) = \Omega_{\check{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}})$ is considered in the remainder of this thesis.

4.3 Performance Limits: Cramer-Rao Lower Bound

The CRLB corresponds the best performance that any *unbiased* estimator can achieve. This means that the covariance matrix of the estimator $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}$ is greater than or equal to the inverse of the *Fisher information matrix* $\boldsymbol{I}^{-1}(\boldsymbol{\theta})$ [Sch91, Sec. 6.4], [Kay93, Chap. 4]:

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} - \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \geq \boldsymbol{0}, \tag{4.28}$$

i.e. the matrix $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} - \boldsymbol{I}^{-1}(\boldsymbol{\theta})$ is positive semidefinite. If only a single parameter θ_p, $1 \leq p \leq P$, is considered, the variance of this parameter, which corresponds to the *mean squared error (MSE)* in case of an unbiased estimator, is greater than or equal to the corresponding diagonal entry of the inverse Fisher information matrix:

$$\mathrm{MSE}\left(\hat{\theta}_p\right) = \left[\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}\right]_{p,p} \geq \left[\boldsymbol{I}^{-1}(\boldsymbol{\theta})\right]_{p,p} = \mathrm{CRLB}\left(\theta_p\right). \tag{4.29}$$

Since both estimators derived above are equivalent, they *must* lead to the same Fisher information matrix and, hence, to the same CRLB. Therefore, the Fisher information matrix is determined for both metrics, $\Omega_{\boldsymbol{y}}(\boldsymbol{\theta})$ and $\Omega_{\check{\boldsymbol{h}}}(\boldsymbol{\theta})$, separately in order to confirm their equivalence.

4.3.1 Metric with Received Samples

According to [Sch91, p. 223], the Fisher information matrix is defined as

$$\boldsymbol{I}(\boldsymbol{\theta}) = -\,\mathrm{E}\left\{\left(\frac{\partial \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{T}\right\}, \tag{4.30}$$

where the partial derivative of the log-likelihood function $\ln p(\boldsymbol{y},\boldsymbol{\theta})$ corresponds to

$$\frac{\partial \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left[\frac{\partial \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \theta_1}, \cdots, \frac{\partial \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \theta_P}\right]^{T}. \tag{4.31}$$

Therefore, each entry of the Fisher information matrix is given by

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{p,q} = -\,\mathrm{E}\left\{\frac{\partial \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \theta_p}\cdot\frac{\partial \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \theta_q}\right\} = -\,\mathrm{E}\left\{\frac{\partial^2 \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q}\right\} \tag{4.32}$$

as defined in [Kay93, p. 44]. In order to obtain the second partial derivative of the log-likelihood function, its first partial derivative is computed previously:

$$\begin{aligned}\frac{\partial \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \theta_p} &= -\frac{\partial}{\partial \theta_p}\left(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{h}(\boldsymbol{\theta})\right)^{H}\boldsymbol{C}_{\boldsymbol{n}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{h}(\boldsymbol{\theta})\right)\\ &= \left(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{h}(\boldsymbol{\theta})\right)^{H}\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{X}_0\frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \theta_p} + \frac{\partial \boldsymbol{h}^{H}(\boldsymbol{\theta})}{\partial \theta_p}\boldsymbol{X}_0^{H}\boldsymbol{C}_{\boldsymbol{n}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{h}(\boldsymbol{\theta})\right)\\ &= 2\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{h}^{H}(\boldsymbol{\theta})}{\partial \theta_p}\boldsymbol{X}_0^{H}\boldsymbol{C}_{\boldsymbol{n}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{h}(\boldsymbol{\theta})\right)\right\}.\end{aligned} \tag{4.33}$$

Based on (4.33), the second derivative of $\ln p(\boldsymbol{y},\boldsymbol{\theta})$ can be determined as

$$\begin{aligned}\frac{\partial^2 \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q} &= \frac{\partial}{\partial \theta_q} 2\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{h}^{H}(\boldsymbol{\theta})}{\partial \theta_p}\boldsymbol{X}_0^{H}\boldsymbol{C}_{\boldsymbol{n}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{h}(\boldsymbol{\theta})\right)\right\}\\ &= 2\,\mathrm{Re}\left\{\frac{\partial^2 \boldsymbol{h}^{H}(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q}\boldsymbol{X}_0^{H}\boldsymbol{C}_{\boldsymbol{n}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{h}(\boldsymbol{\theta})\right) - \frac{\partial \boldsymbol{h}^{H}(\boldsymbol{\theta})}{\partial \theta_p}\boldsymbol{X}_0^{H}\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{X}_0\frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \theta_q}\right\}\\ &= 2\,\mathrm{Re}\left\{\frac{\partial^2 \boldsymbol{h}^{H}(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q}\boldsymbol{X}_0^{H}\boldsymbol{C}_{\boldsymbol{n}}^{-1} \qquad \boldsymbol{n} \qquad - \frac{\partial \boldsymbol{h}^{H}(\boldsymbol{\theta})}{\partial \theta_p}\boldsymbol{X}_0^{H}\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{X}_0\frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \theta_q}\right\}.\end{aligned} \tag{4.34}$$

If the expectation of (4.34) is taken, only its second part remains since the noise is assumed to be zero mean. Hence, the entries of the Fisher information matrix are given by

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{p,q} = -\,\mathrm{E}\left\{\frac{\partial^2 \ln p(\boldsymbol{y},\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q}\right\} = 2\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{h}^{H}(\boldsymbol{\theta})}{\partial \theta_p}\boldsymbol{X}_0^{H}\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{X}_0\frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \theta_q}\right\}. \tag{4.35}$$

Employing the *Jacobian matrix* of the channel function $\boldsymbol{h}(\boldsymbol{\theta})$, which is defined as

$$\boldsymbol{J}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{T}} = \left[\frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \theta_1}, \cdots, \frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \theta_P}\right] = \begin{bmatrix}\frac{\partial h_0}{\partial \theta_1} & \cdots & \frac{\partial h_0}{\partial \theta_P}\\ \vdots & \ddots & \vdots\\ \frac{\partial h_L}{\partial \theta_1} & \cdots & \frac{\partial h_L}{\partial \theta_P}\end{bmatrix}, \tag{4.36}$$

results in the following compact expression for the Fisher information matrix:

$$\boldsymbol{I}(\boldsymbol{\theta}) = 2\,\mathrm{Re}\left\{\boldsymbol{J}^H(\boldsymbol{\theta})\boldsymbol{X}_0^H\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{X}_0\boldsymbol{J}(\boldsymbol{\theta})\right\}. \tag{4.37}$$

Due to the equivalence of both derived estimators, the estimator based on the pre-stage channel estimates must lead to the same expression.

4.3.2 Metric with Pre-Stage Channel Estimates

Similar to (4.32), the entries of the Fisher information matrix for the second estimator, which is based on the pre-stage channel estimates, are defined as

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{p,q} = -\,\mathrm{E}\left\{\frac{\partial \ln p(\check{\boldsymbol{h}},\boldsymbol{\theta})}{\partial\theta_p}\cdot\frac{\partial \ln p(\check{\boldsymbol{h}},\boldsymbol{\theta})}{\partial\theta_q}\right\} = -\,\mathrm{E}\left\{\frac{\partial^2 \ln p(\check{\boldsymbol{h}},\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q}\right\}. \tag{4.38}$$

The first derivative of the log-likelihood function results in

$$\begin{aligned}\frac{\partial \ln p(\check{\boldsymbol{h}},\boldsymbol{\theta})}{\partial\theta_p} &= -\frac{\partial}{\partial\theta_p}\left(\check{\boldsymbol{h}}-\boldsymbol{h}(\boldsymbol{\theta})\right)^H\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\left(\check{\boldsymbol{h}}-\boldsymbol{h}(\boldsymbol{\theta})\right)\\ &= 2\,\mathrm{Re}\left\{\frac{\partial\boldsymbol{h}^H(\boldsymbol{\theta})}{\partial\theta_p}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\left(\check{\boldsymbol{h}}-\boldsymbol{h}(\boldsymbol{\theta})\right)\right\}\end{aligned} \tag{4.39}$$

and the second derivative is given by

$$\begin{aligned}\frac{\partial^2 \ln p(\check{\boldsymbol{h}},\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q} &= \frac{\partial}{\partial\theta_q}2\,\mathrm{Re}\left\{\frac{\partial\boldsymbol{h}^H(\boldsymbol{\theta})}{\partial\theta_p}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\left(\check{\boldsymbol{h}}-\boldsymbol{h}(\boldsymbol{\theta})\right)\right\}\\ &= 2\,\mathrm{Re}\left\{\frac{\partial^2\boldsymbol{h}^H(\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\left(\check{\boldsymbol{h}}-\boldsymbol{h}(\boldsymbol{\theta})\right) - \frac{\partial\boldsymbol{h}^H(\boldsymbol{\theta})}{\partial\theta_p}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\frac{\partial\boldsymbol{h}(\boldsymbol{\theta})}{\partial\theta_q}\right\}\\ &= 2\,\mathrm{Re}\left\{\frac{\partial^2\boldsymbol{h}^H(\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\quad\boldsymbol{\eta}\quad - \frac{\partial\boldsymbol{h}^H(\boldsymbol{\theta})}{\partial\theta_p}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\frac{\partial\boldsymbol{h}(\boldsymbol{\theta})}{\partial\theta_q}\right\}.\end{aligned} \tag{4.40}$$

Again, taking the expectation of (4.40), only its second part remains since the channel estimation error is zero mean as well. Thus, the entries of the Fisher information matrix are described by

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{p,q} = -\,\mathrm{E}\left\{\frac{\partial^2 \ln p(\check{\boldsymbol{h}},\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q}\right\} = 2\,\mathrm{Re}\left\{\frac{\partial\boldsymbol{h}^H(\boldsymbol{\theta})}{\partial\theta_p}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\frac{\partial\boldsymbol{h}(\boldsymbol{\theta})}{\partial\theta_q}\right\}, \tag{4.41}$$

with leads to

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\theta}) &= 2\,\mathrm{Re}\left\{\boldsymbol{J}^H(\boldsymbol{\theta})\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\boldsymbol{J}(\boldsymbol{\theta})\right\}\\ &= 2\,\mathrm{Re}\left\{\boldsymbol{J}^H(\boldsymbol{\theta})\boldsymbol{X}_0^H\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{X}_0\boldsymbol{J}(\boldsymbol{\theta})\right\}.\end{aligned} \tag{4.42}$$

As predicted, both Fisher information matrices given in (4.37) and (4.42) are the same. Thus, the equivalence of both metrics, $\Omega_{\boldsymbol{y}}(\tilde{\boldsymbol{\theta}})$ and $\Omega_{\check{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}})$, is confirmed.

4.3.3 Comparison of CRLBs for Different Channel Models

In this section, the CRLBs for different channel models are compared. Especially, the influence of oversampling is investigated. Parts of this section have been published in [SAH12].

As already mentioned earlier, the estimation of the TOA $\hat{\tau}_1$ corresponds to the estimation of the sampling phase $\hat{\varepsilon}$ and, thus, the only parameter of interest for positioning is $\hat{\theta}_3{=}\hat{\varrho}_1{=}{-}\hat{\varepsilon}$. Given a certain parameter vector $\boldsymbol{\theta}$, only the corresponding CRLB according to (4.29) with $p{=}3$ is considered in the following:

$$\mathrm{MSE}\left(\hat{\tau}_1\right) = \mathrm{MSE}\left(\hat{\varepsilon}\right) = \left[\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}\right]_{3,3} \geq \left[\boldsymbol{I}^{-1}(\boldsymbol{\theta})\right]_{3,3} = \mathrm{CRLB}\left(\varepsilon\right).$$

Since there are many possible parameter sets $\boldsymbol{\theta}$, the CRLBs are determined semi-analytically by means of Monte Carlo simulations. In each run of a Monte Carlo simulation, a different channel realisation with a different parameter vector $\boldsymbol{\theta}$ is generated and the corresponding Fisher information matrix is determined according to (4.42). The overall CRLB is given by the *expectation* of the inverse Fisher information matrices

$$\mathrm{CRLB}\left(\varepsilon\right) = \mathrm{E}\left\{\left[\boldsymbol{I}^{-1}(\boldsymbol{\theta})\right]_{3,3}\right\}, \tag{4.43}$$

where the expectation is taken with respect to the parameter vector $\boldsymbol{\theta}$. For all channel models, the CRLBs are determined for different oversampling factors (J = 1, 2, 4 and 8) over the signal-to-noise ratio (SNR) in dB. The following simulation setup is applied if not stated otherwise: A burst length of $K'{=}100$ symbol durations ($K{=}JK'$) is assumed and a pseudo-random sequence of BPSK symbols is used as training. A Gaussian pulse shape

$$p(\tau) = \exp\left(-(\tau/T_s)^2\right) \tag{4.44}$$

is applied. In order to obtain a causal pulse shape, the Gaussian pulse is shifted by a certain amount s. That means that the overall pulse shape is given by

$$g\left(\tau\right) = g_{\mathrm{Tx}}\left(\tau\right) * g_{\mathrm{Rx}}\left(\tau\right) = p(\tau - s), \tag{4.45}$$

which is equally distributed among the pulse shaping and the receive filter. Hence, the autocorrelation of each filter corresponds to a Gaussian pulse: $\psi_{\mathrm{Tx}}\left(\tau\right){=}\psi_{\mathrm{Rx}}\left(\tau\right){=}p(\tau)$. In this case, the noise covariance matrix is given by

$$\left[\boldsymbol{C}_{\boldsymbol{n}}\right]_{i,j} = \sigma_n^2 \;\, p\big((i-j)\mathcal{T}\big) = \sigma_n^2 \;\, \exp\left(-\left(\frac{i-j}{J}\right)^2\right). \tag{4.46}$$

An effective pulse width of $T_g{=}8T_s$ and a shift of $s{=}T_g/2{=}4T_s$ are assumed subsequently. In each run of a Monte Carlo simulation, a random sampling phase ε is generated uniformly distributed in the interval $[-T_s/2, +T_s/2]$. The remaining parameters of $\boldsymbol{\theta}$ are generated according to the channel model. In all figures below, the quantities concerning timing or delay measures are normalised with respect to the symbol duration, e.g. the CRLB of ε is normalised to T_s^2.

Before presenting the numerical results, the applied power normalisation is shortly discussed at this point. Generally, the SNR is described by

$$\gamma_s = \frac{\mathrm{E}\left\{\left|\sum_{l=0}^{L} h_l(\boldsymbol{\theta}) \cdot x_0[k-l]\right|^2\right\}}{\mathrm{E}\left\{|n[k]|^2\right\}}. \tag{4.47}$$

As the noise is zero mean, the noise power in the denominator of (4.47) corresponds to the noise variance, which is given by the value of the sampled autocorrelation function at Δk=0:

$$\mathrm{E}\left\{|n[k]|^2\right\} = \varphi_{nn}[0] = \sigma_n^2 \ \exp\left(-(0/J)^2\right) = \sigma_n^2, \tag{4.48}$$

i.e. the noise power remains the same for all oversampling factors. Hence, the signal power must also remain constant for different oversampling factors. The following power normalisation is applied for symbol-rate sampling: The average power of the transmit symbols is normalised to one. In case of BPSK with equiprobable symbols, this is fulfilled automatically:

$$\mathrm{E}\left\{|x_0[k]|^2\right\} = \frac{1}{2}(+1)^2 + \frac{1}{2}(-1)^2 = 1. \tag{4.49}$$

Furthermore, the average power of the channel is normalised to one:

$$\mathrm{E}\left\{\sum_{l=0}^{L} |h_l(\boldsymbol{\theta})|^2\right\} \stackrel{!}{=} 1. \tag{4.50}$$

In case of oversampling, the power of the channel coefficients corresponds to the oversampling factor J. At the same time, the average power of the transmit symbols reduces to $1/J$ since $J-1$ zeros are inserted between neighbouring symbols. Thus, the average signal power in the numerator of (4.47) is one for all oversampling factors with the above power normalisation. In this case, the SNR is given as

$$\gamma_s = \frac{1}{\sigma_n^2}. \tag{4.51}$$

LOS Channel Model

The LOS channel, which is described in Section 2.5.1, is taken into account since it is the best possible case for positioning and, thus, provides a lower bound for all other channel models. As block fading is considered, equation (2.51) applies. With the assumed Gaussian pulse shape it follows

$$h_l = \exp(\mathrm{j}\,\Phi)\ \exp\left(-\left(\frac{l\mathcal{T} + \varepsilon - s}{T_s}\right)^2\right) = f\ \exp\left(-\alpha_{l,J}^2\right) \tag{4.52}$$

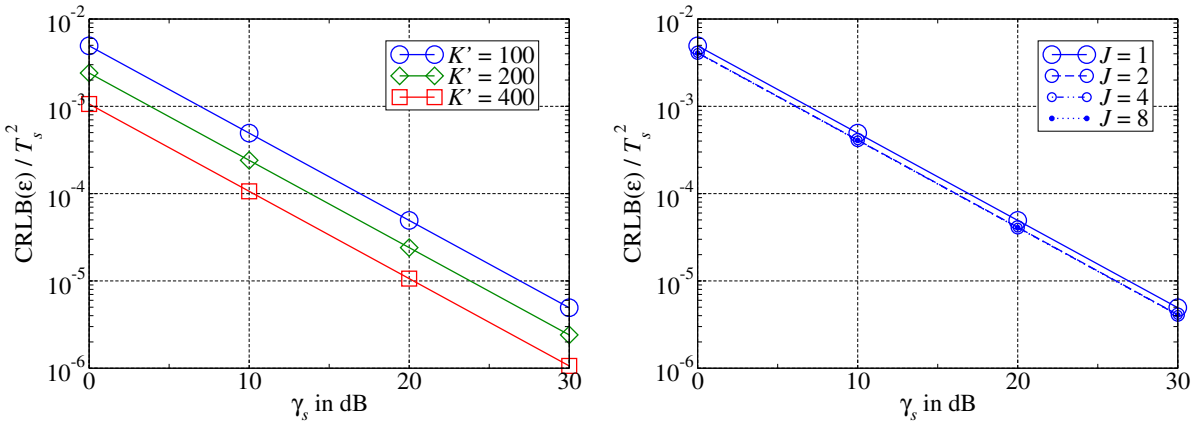
with $f = \exp(\mathrm{j}\,\Phi)$, $\alpha_{l,J} = l/J + (\varepsilon - s)/T_s$. Since there is only a single path, there are no excess delays and, thus, the channel memory length results in L'=9 ($L = JL'$). The starting

phase Φ is generated randomly between 0 and 2π. In Figure 4.3, the normalised CRLB of ε for the LOS channel is plotted over the SNR. In Figure 4.3(a), the influence of the burst length K' is shown for symbol-rate sampling (J=1), whereas the influence of the oversampling factor J is illustrated in Figure 4.3(b) for a burst length of K'=100. In all cases, the normalised CRLB decreases with the SNR and is much smaller than 1, which corresponds to a small fraction of the squared symbol duration T_s^2. This means that the estimation error is much smaller than the symbol duration T_s such that the TOA estimate is improved significantly in comparison to the coarse pre-synchronisation. The larger the burst length, the better the performance: The CRLB improves by approximately 3 dB if the burst length is doubled (see Figure 4.3(a)). The same influence of the burst length is observed in all other channel models. Hence, only burst lengths of K'=100 are considered in the following if not stated otherwise. Furthermore, oversampling provides a slight performance gain as shown in Figure 4.3(b): For all oversampling factors $J \geq 2$, the CRLBs are improved by approximately 0.8 dB in comparison to symbol-rate sampling. Thus, oversampling with J=2 should be applied. The question arises for what reason oversampling provides a performance gain? What is the difference between symbol-rate sampling and oversampling? In order to answer these questions, the CRLB for the LOS channel is examined in more detail. The starting point is the Fisher information matrix given in (4.37) and (4.42). In order to keep the following investigations manageable, two assumptions are applied that simplify the determination of the Fisher information matrix. First, white noise is assumed for all oversampling factors and, second, the burst length is assumed to be large. In this case, the inverse covariance matrix of the channel estimation error can be approximated by a scaled identity matrix

$$\boldsymbol{C}_{\boldsymbol{\eta}}^{-1} = \boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{X}_0 \approx \gamma \cdot \boldsymbol{\mathcal{I}} \tag{4.53}$$

with scaling factor $\gamma=(K'-L')/(J\sigma_n^2)$. This leads to an approximate Fisher information matrix described by

$$\boldsymbol{I}(\boldsymbol{\theta}) \approx 2\gamma \cdot \mathrm{Re}\left\{\boldsymbol{J}^H(\boldsymbol{\theta})\boldsymbol{J}(\boldsymbol{\theta})\right\}. \tag{4.54}$$



(a) Influence of the burst length K' (J=1).

(b) Influence of the oversampling factor J (K'=100).

Figure 4.3: CRLB of ε normalised to T_s^2 over SNR for the LOS channel.

Hence, the Jacobian matrix as defined in (4.36) needs to be determined. The partial derivatives of the channel coefficients are given by

$$\frac{\partial h_l}{\partial \theta_1} = \exp\left(-\alpha_{l,J}^2\right), \qquad \frac{\partial h_l}{\partial \theta_2} = \mathrm{j}\exp\left(-\alpha_{l,J}^2\right), \qquad \frac{\partial h_l}{\partial \theta_3} = 2f \cdot \exp\left(-\alpha_{l,J}^2\right) \cdot \alpha_{l,J} \tag{4.55}$$

Multiplying the Hermitian conjugate of the Jacobian matrix with the Jacobian matrix itself and taking the real part of the result leads to a Fisher information matrix of the following form:

$$\boldsymbol{I}(\boldsymbol{\theta}) \approx 2\gamma \cdot \begin{bmatrix} A & 0 & B \\ 0 & A & C \\ B & C & D \end{bmatrix} \tag{4.56}$$

with

$$A = \sum_{l=0}^{L} \left\{\exp\left(-\alpha_{l,J}^2\right)\right\}^2, \tag{4.57}$$

$$B = 2\,\mathrm{Re}\left\{f\right\} \sum_{l=0}^{L} \left\{\exp\left(-\alpha_{l,J}^2\right)\right\}^2 \cdot \alpha_{l,J}, \tag{4.58}$$

$$C = 2\,\mathrm{Im}\left\{f\right\} \sum_{l=0}^{L} \left\{\exp\left(-\alpha_{l,J}^2\right)\right\}^2 \cdot \alpha_{l,J}, \tag{4.59}$$

$$D = 4|f|^2 \sum_{l=0}^{L} \left\{\exp\left(-\alpha_{l,J}^2\right)\right\}^2 \cdot \alpha_{l,J}^2. \tag{4.60}$$

Inverting the matrix in (4.56) results in

$$\boldsymbol{I}^{-1}(\boldsymbol{\theta}) \approx \frac{1}{2\gamma} \cdot \frac{1}{A(AD - B^2 - C^2)} \begin{bmatrix} AD - C^2 & BC & -AB \\ BC & AD - B^2 & -AC \\ -AB & -AC & A^2 \end{bmatrix}. \tag{4.61}$$

The CRLB of the sampling phase corresponds to the third main diagonal entry of the above matrix

$$\mathrm{CRLB}\left(\varepsilon\right) = \left[\boldsymbol{I}^{-1}(\boldsymbol{\theta})\right]_{3,3} \approx \frac{1}{2\gamma} \cdot \frac{A}{AD - B^2 - C^2} \tag{4.62}$$

$$= \frac{1}{8\gamma|f|^2} \cdot \frac{\sum_{l=0}^{L} \left\{\exp\left(-\alpha_{l,J}^2\right)\right\}^2}{\left[\sum_{l=0}^{L} \left\{\exp\left(-\alpha_{l,J}^2\right)\right\}^2\right] \left[\sum_{l=0}^{L} \left\{\exp\left(-\alpha_{l,J}^2\right)\right\}^2 \cdot \alpha_{l,J}^2\right] - \left[\sum_{l=0}^{L} \left\{\exp\left(-\alpha_{l,J}^2\right)\right\}^2 \cdot \alpha_{l,J}\right]^2}$$

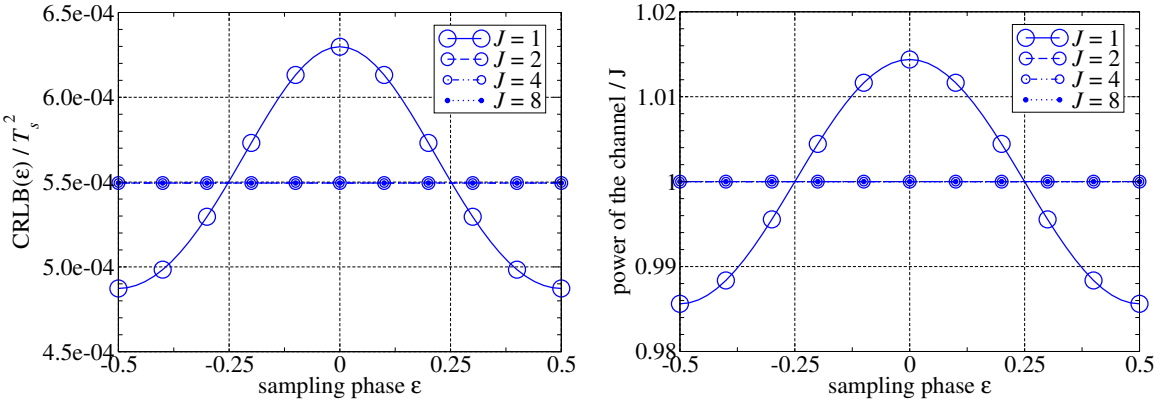
Since $|f|^2$ is *always* one in the LOS channel, the difference between symbol-rate sampling and oversampling must depend somehow on the sums over the exponential terms and, thus, on the sampling phase itself. In Figure 4.4, the influence of the sampling phase on the approximate CRLB is illustrated: In Figure 4.4(a), the approximate CRLB is plotted over the sampling phase ε for different oversampling factors J with γ_s=10 dB.

For $J \geq 2$, the approximate CRLB is constant and does *not* depend on ε. Only for symbol-rate sampling, the approximate CRLB varies with the sampling phase. Thus, there is a considerable difference between symbol-rate sampling and oversampling. The same behaviour can be observed for the average channel power divided by the oversampling factor,

$$\frac{1}{J}\,\mathrm{E}\left\{\sum_{l=0}^{L}\left|h_l\left(\boldsymbol{\theta}\right)\right|^2\right\} = \frac{1}{J}\,\mathrm{E}\left\{\sum_{l=0}^{L}\left\{\exp\left(\alpha_{l,J}^2\right)\right\}^2\right\}, \tag{4.63}$$

which corresponds to the numerator in (4.62). The average channel power according to (4.63) is plotted in Figure 4.4(b): For $J \geq 2$, the power of the channel is constant and, thus, independent of the sampling phase. For symbol-rate sampling, in contrast, the power of the channel is a function of the sampling phase: The highest power is obtained for ε=0 (perfect synchronisation), while the smallest power occurs for $|\varepsilon|$=0.5. This means that non-perfect synchronisation leads to a power loss concerning symbol-rate sampling. Similar results have already been reported in [SH08] for a rectangular and a root raised cosine pulse shape.

Now, it is clear that there is a considerable difference between symbol-rate sampling and oversampling. But the question remains: *How* does this difference lead to the observed performance gain of oversampling? In order to answer this question, it is necessary to return to the exact Fisher information matrix described by (4.37) and (4.42) and the corresponding exact CRLB. How does the behaviour of the CRLB change, if the exact instead of the approximate version is taken into account? In Figure 4.5, both types of the CRLB are compared for γ_s=10 dB. The curves are labelled with "A" for approximate and "E" for exact. On the left hand side of Figure 4.5, both types of the CRLBs are plotted over the sampling phase ε for K'=100 (Figure 4.5(a)) and K'=400 (Figure 4.5(c)). In order to have a closer look at the different behaviour of the approximate and the exact CRLB, the normalised difference between the symbol-rate sampled and the oversampled



(a) Approximate CRLB of ε normalised to T_s^2 over the sampling phase ε with γ_s=10 dB.

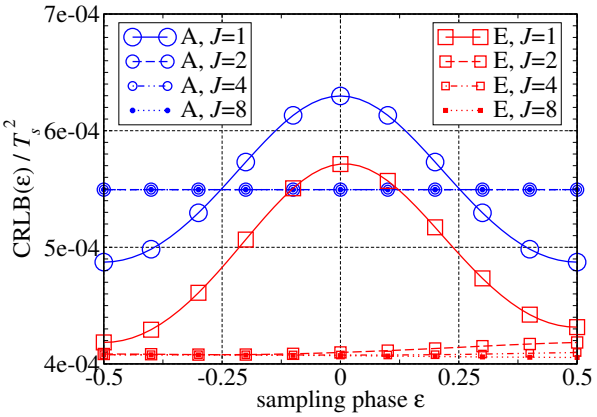
(b) Average channel power divided by the oversampling factor over the sampling phase.

Figure 4.4: Influence of the sampling phase ε on the approximate CRLB.

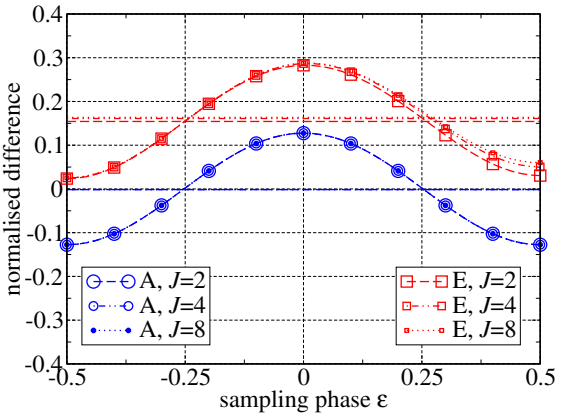
CRLBs according to

$$\frac{\mathrm{CRLB}_1(\varepsilon) - \mathrm{CRLB}_J(\varepsilon)}{\mathrm{CRLB}_1(\varepsilon)}, \tag{4.64}$$

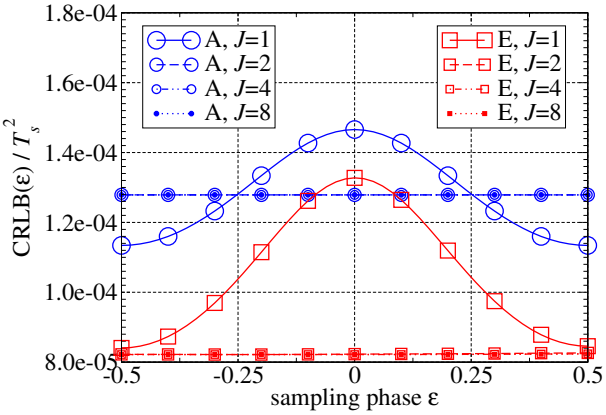
is plotted for each type on the right hand side of Figure 4.5 for K'=100 (Figure 4.5(b)) and K'=400 (Figure 4.5(d)) as well. The shape of the curves is basically the same for both types. The difference between the approximate CRLB and the exact CRLB lies in the fact that the curves of the exact CRLB are shifted downward with the shift for symbol-rate sampling being smaller than the shift for oversampling. This means that the curves of the normalised difference between the CRLBs are shifted upward. For K'=100, the exact CRLBs for $J\geq 2$ and the corresponding normalised differences do not coincide over the whole range of the sampling phase, but differ slightly for $\varepsilon\geq 0$ (see Figure 4.5(a) and 4.5(b)). This effect is due to a short burst length since it is not present anymore if burst length is increased to K'=400 (see Figure 4.5(c) and 4.5(d)). The most interesting aspect is revealed by the *mean* of the normalised difference, which is shown as a straight line in Figure 4.5(b) and 4.5(d) as well. If white noise is assumed for



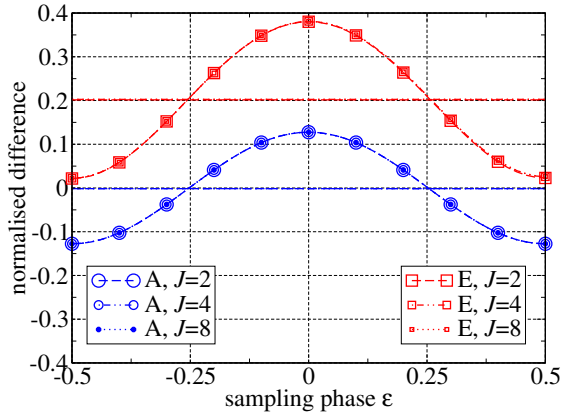
(a) CRLB of ε normalised to T_s^2 with K'=100.



(b) Normalised difference of CRLBs with K'=100.



(c) CRLB of ε normalised to T_s^2 with K'=400.

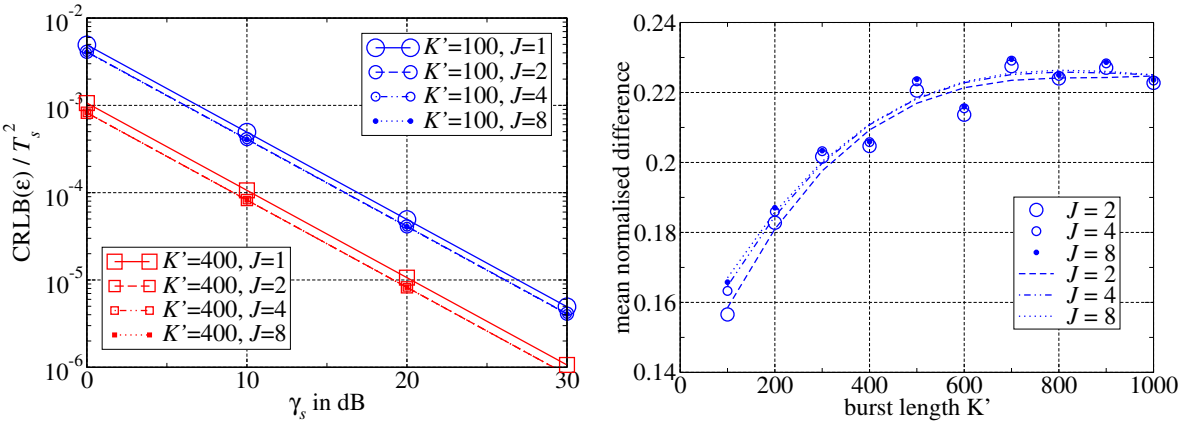


(d) Normalised difference of CRLBs with K'=400.

Figure 4.5: Comparison of approximate and exact CRLBs over the sampling phase ε with γ_s=10 dB. The curves are labelled with "A" for approximate and "E" for exact.

all oversampling factors ("A"), then the mean of the normalised difference (4.64) equals zero in all cases. That means, that *on average* there is no difference between symbol-rate sampling and oversampling such that the CRLBs determined via Monte Carlo simulations (with a random sampling phase) are the same. In case of coloured noise ("E"), the mean of the normalised difference (4.64) is approximately 0.16 for K'=100 and 0.2 for K'=400, i.e. the exact CRLBs for symbol-rate sampling and oversampling differ on average by a small amount. This difference leads to the oversampling gain which has been observed in Figure 4.3(b).

Last but not least, it is interesting to note that the mean of the normalised difference varies with the burst length K'. This implies that the performance gain of oversampling varies with the burst length as well. In Figure 4.6, the influence of the burst length K' is examined in more detail. In order to check the above hypothesis about the performance gain, the CRLB of ε is plotted for different oversampling factors and the two burst length of K'=100 and K'=400 in Figure 4.6(a). As already mentioned before, the performance gain of oversampling is approximately 0.8 dB for K'=100. For K'=400, this gain is increased to approximately 1 dB. The ratio of these performance gains corresponds to the ratio of the mean normalised differences stated above: 0.8 dB/1 dB=0.8=0.16/0.2. Thus, the mean of the normalised difference translates into the performance gain with a positive factor. Finally, Figure 4.6(b) shows the mean of the normalised difference of the CRLBs over the burst length K'. The circles denote the simulated values, while the lines are the corresponding fitted curves. The absolute value of the mean normalised difference increases with the burst length and saturates around 0.223, i.e the mean normalised difference is constant from approximately K'=800 upwards. Hence, the maximum achievable oversampling gain is approximately 1.115 dB in a LOS channel.



(a) CRLB of ε normalised to T_s^2 over SNR for different oversampling factors and burst length.

(b) Mean normalised difference of CRLBs over burst length K'.

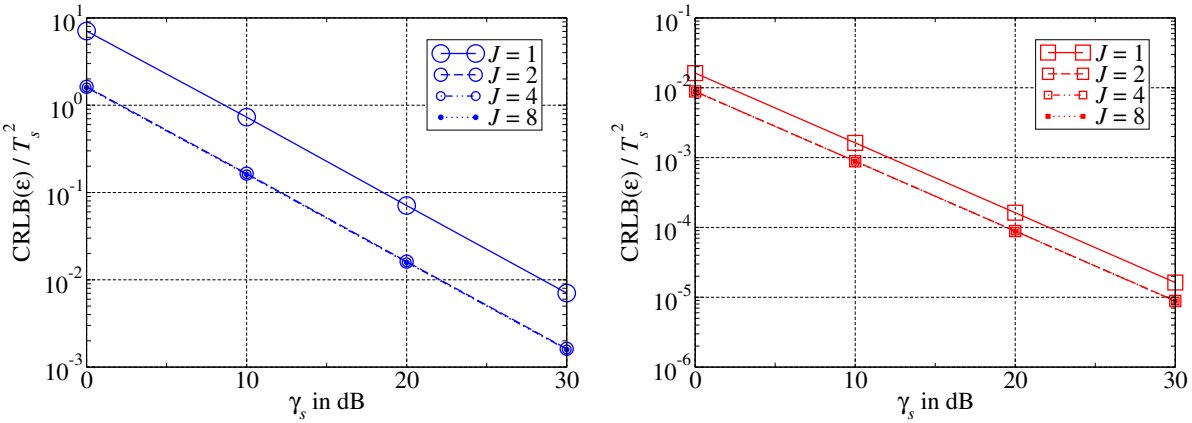
Figure 4.6: Influence of the burst length K' on the performance gain of oversampling.

Two-Path Channel Model

The two-path channel, which is described in Section 2.5.2, represents the simplest form of a multipath channel with only two propagation paths. Since block fading is assumed, equation (2.54) fully characterises the two-path channel model. With the assumed Gaussian pulse shape it follows

$$\begin{aligned} h_l = A_1 \exp(\mathrm{j}\,\Phi_1) \exp\left(-\left(\frac{l\mathcal{T}+\varepsilon-s}{T_s}\right)^2\right) \\ + A_2 \exp(\mathrm{j}\,\Phi_2) \exp\left(-\left(\frac{l\mathcal{T}+\varepsilon-s-\nu_2}{T_s}\right)^2\right). \end{aligned} \tag{4.65}$$

By means of this two-path channel model, the influence of different channel parameters such as the excess delay ν_2, power ratio $\mathcal{P}=A_1^2/A_2^2$ and phase offset $\Delta\Phi=\Phi_2-\Phi_1$ between the two propagation paths can be investigated. The maximum possible excess delay is fixed to $\nu_2^{\max}=2T_s$, which leads to a channel memory length of $L'=11$ ($L=JL'$). At first, the CRLBs are examined over SNR, where the power ratio and the phase offset are generated randomly in the intervals [0.1, 10] and [0,2π[, respectively. Concerning the excess delay, two different scenarios are considered: one with small excess delays ($\nu_2/T_s\in[0.05,1]$) and one with large excess delays ($\nu_2/T_s\in[1,2]$). The corresponding CRLBs for different oversampling factors are shown in Figure 4.7(a) and Figure 4.7(b), respectively. The overall performance decreases with respect to the LOS channel (as expected), but the oversampling gain increases. Both, the overall performance and the oversampling gain, depend on the channel model: The smaller the excess delay, the worse is the CRLB and the larger is the oversampling gain. For the two-path channel with small excess delay, the gain is approximately 6.5 dB, whereas it is only 2.6 dB for the channel with large excess delay. In order to have a closer look at this dependence, the CRLBs are determined over the excess delay for a fixed SNR (γ_s=10 dB). Similarly, the influence of the phase offset $\Delta\Phi$ and the power ratio $\mathcal{P}$ is investigated.



(a) Two-path channel with $\nu_2/T_s\in[0.05,1]$.

(b) Two-path channel with $\nu_2/T_s\in[1,2]$.

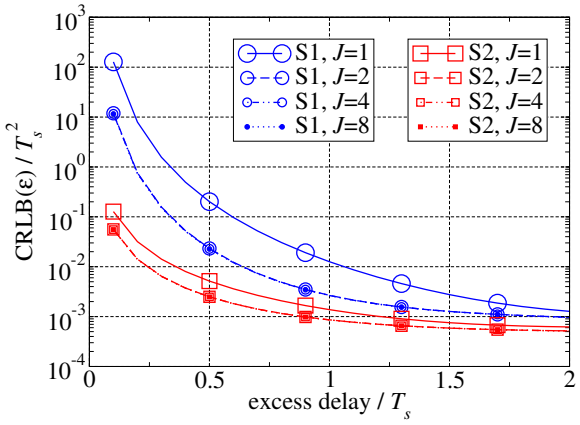
Figure 4.7: CRLB of ε normalised to T_s^2 over SNR for the two-path channels.

In contrast to the LOS channel, the approximate Fisher information matrix according to (4.54) and the corresponding approximate CRLB are not considered here in closed form. Due to the increased size of the Fisher information matrix (6×6 instead of 3×3), the resulting formula for the approximate CRLB consists of many different terms and is, therefore, less concise than in the case of the LOS channel. Hence, only simulation results are presented in the following. Two basic simulation setups are applied, whose parameters are tabulated in Table 4.1. The parameter of interest is varied, while the other parameters are fixed to the values given in Table 4.1. The corresponding CRLBs and their normalised differences according to (4.64) are shown on the left hand side and on the right hand side of Figure 4.8, respectively. The curves are labelled with "S1" for the first setup and "S2" for the second setup. The mean normalised difference of the LOS channel (for K'=100) is given for comparison in Figure 4.8(b), 4.8(d) and 4.8(f) since it corresponds to a lower bound concerning the normalised difference of the two-path channels. In the first row of Figure 4.8, the influence of the excess delay ν_2 is illustrated for both setups, while the influence of the phase offset $\Delta\Phi$ and the power ratio $\mathcal{P}$ are shown in the second and third row, respectively. The overall performance and the oversampling gain strongly depend on the simulation setup and the channel characteristic of interest. It is obvious that the first setup corresponds to a kind of worst case concerning parameter estimation, whereas the second setup represents a kind of best case scenario. As mentioned above, the normalised difference translates into the oversampling gain with a positive factor. Thus, the larger the normalised difference, the larger is the oversampling gain. For the first setup, oversampling can provide a significant gain over symbol-rate sampling. This gain is *not* influenced by the power ratio (see Figure 4.8(f)), while it *mainly* depends on the excess delay ν_2 (see Figure 4.8(b)): The smaller the excess delay, the larger is the oversampling gain. Hence, oversampling proves especially helpful in dense multipath scenarios. Furthermore, the phase offset $\Delta\Phi$ has a significant impact on the oversampling gain, which is highest when both propagations paths have the same or the opposite phase (see Figure 4.8(d)).

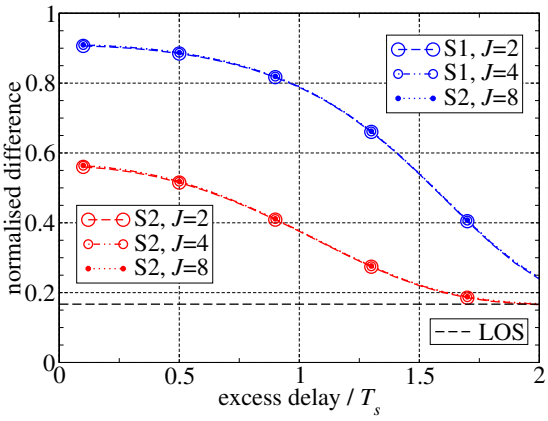
Figure 4.8 does not only illustrate the dependence of the oversampling gain on the channel characteristics, but gives also an insight into the relation between these characteristics and the overall performance: With decreasing excess delay ν_2, the CRLB increases as it becomes more difficult to separate the two propagation paths. Similarly, the CRLB is worst if the paths are in phase ($\Delta\Phi$=0°) or have an opposite phase ($\Delta\Phi$=180°). The smaller the power ratio $\mathcal{P}$, the smaller is the amplitude of the first path in comparison to the second path and the more likely it is that the delay of the first path, namely the sampling phase, is estimated wrongly. Thus, the influence of the channel characteristics

Parameter	Setup 1	Setup 2
ν_2/T_s	0.5	1.5
$\mathcal{P}$	1.0	5.0
$\Delta\Phi$	0°	90°

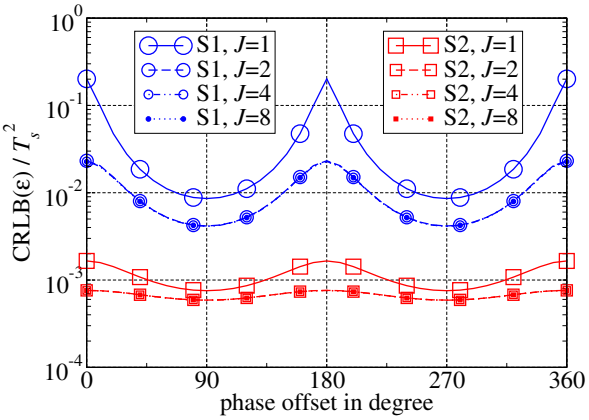
Table 4.1: Basic simulation setups for the two-path channel models.



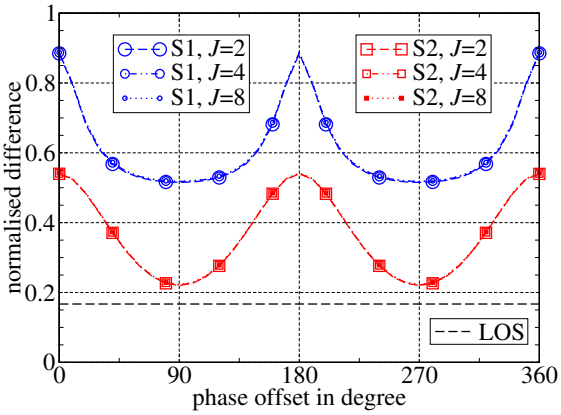
(a) CRLB of ε normalised to T_s^2 over excess delay.



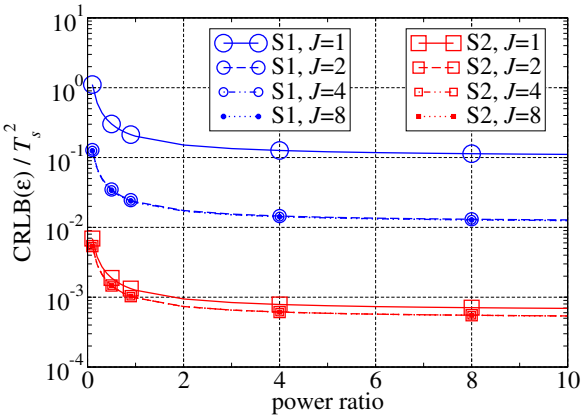
(b) Normalised difference over excess delay.



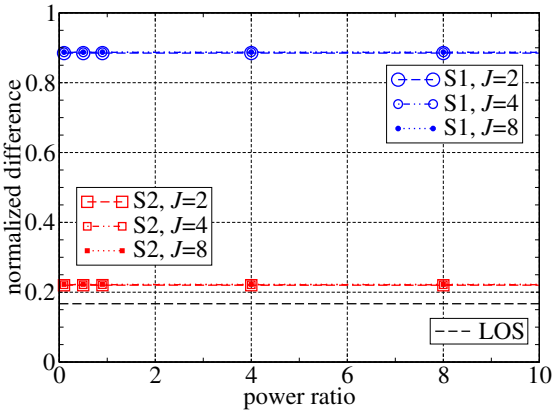
(c) CRLB of ε normalised to T_s^2 over phase offset.



(d) Normalised difference over phase offset.



(e) CRLB of ε normalised to T_s^2 over power ratio.



(f) Normalised difference over power ratio.

Figure 4.8: Comparison of CRLBs and the normalised difference between CRLBs for different channel characteristics with γ_s=10 dB. The curves are labelled with "S1" for the first setup and "S2" for the second setup.

on the overall performance can be summarised as follows: Dense multipath scenarios with similar or opposite phases and a small power ratio are most challenging. The more challenging it is to estimate the sampling phase, the higher is the gain due to oversampling. Hence, the application of oversampling (with J=2) is highly recommended.

WINNER Channel Models

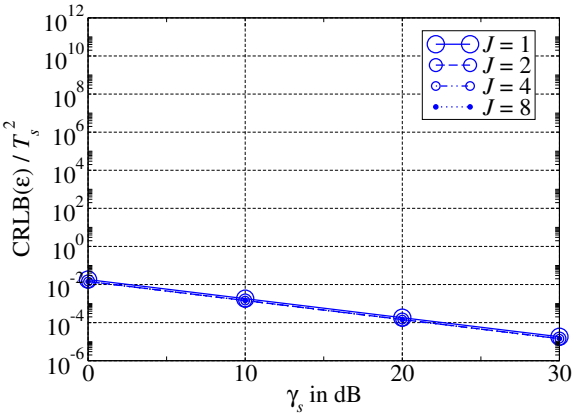
The WINNER channels, which are described in Section 2.5.3, represent realistic scenarios in order to evaluate the system performance. In the WINNER channel models, parameter estimation is much more challenging in comparison to the two-path channel models since a much larger number of propagation paths C>2 is considered. Thus, the dimension of the estimation process P=3C increases significantly. However, the relationships observed for the two-path channels are the basis for more complex channel models with an arbitrary number of propagation paths like the WINNER channel models. In this case, the mutual relationship between all paths determines the performance, where the relationship between the first and the second path is of special importance. The most important parameters of the WINNER channel models are summarised in the upper part of Table 4.2. In addition to the number of clusters C and the number of parameters P, the Rician factor K_R as well as the excess delays ν_2 and ν_C are listed. Furthermore, the minimum delay difference, $\min\{\nu_c - \nu_{c-1}\}$, and the average delay difference, $\mathrm{E}\{\nu_c - \nu_{c-1}\}$,

	Parameter	B1-LOS (urban micro-cell)	B3-LOS (large indoor hall)	C2-LOS (urban macro-cell)
	C	8	10	8
	P	**24**	**30**	**24**
	K_R [dB]	3.3	2.0	7.0
	ν_2 [ns]	33.5	3.5	3.5
	ν_C [ns]	460.0	280.0	220.0
	$\min\{\nu_c - \nu_{c-1}\}$ [ns]	(c=4) 8.5	(c=2;6) 3.5	(c=6) 1.5
	$\mathrm{E}\{\nu_c - \nu_{c-1}\}$ [ns]	65.7	31.1	32.85
	L'	**55**	**37**	**31**
	ν_2/T_s	3.35	0.35	0.35
T_s=10 ns	ν_C/T_s	46.00	28.00	22.00
	$\min\{\nu_c - \nu_{c-1}\}/T_s$	0.85	0.35	0.15
	$\mathrm{E}\{\nu_c - \nu_{c-1}\}/T_s$	6.57	3.11	3.285
	L'	**19**	**15**	**14**
	ν_2/T_s	0.67	0.07	0.07
T_s=50 ns	ν_C/T_s	9.20	5.60	4.40
	$\min\{\nu_c - \nu_{c-1}\}/T_s$	0.17	0.07	0.03
	$\mathrm{E}\{\nu_c - \nu_{c-1}\}/T_s$	1.314	0.622	0.657

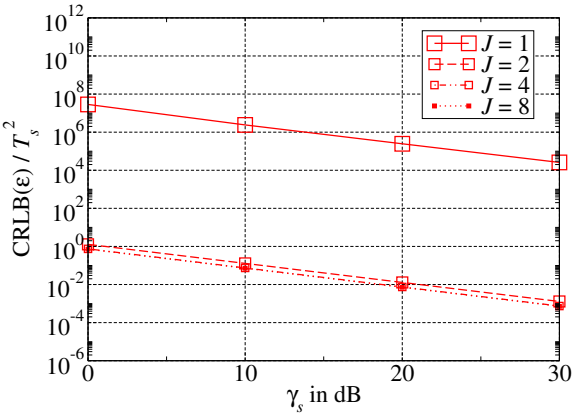
Table 4.2: Parameters of the WINNER channels.

between neighbouring paths is determined. For the minimum delay difference, the cluster number for which this difference occurs is given. Until now, only normalised excess delays ν_2/T_s have been considered. Thus, the preceding results are valid for any choice of the symbol duration T_s. In case of the WINNER channel models, the excess delays are given as absolute values in ns. Hence, the simulation results depend on the choice of the symbol duration T_s. Thus, two different scenarios with T_s=10 ns and T_s=50 ns are investigated in the following. In the lower part of Table 4.2, the channel memory length L' for symbol-rate sampling (L=JL') as well as the relative excess delays and relative delay differences are tabulated for both scenarios. Comparing the number of parameters P with the channel memory length L' (both are highlighted in Table 4.2) reveals a potential problem: For T_s=50 ns, the dimension of the parameter space P is larger than the dimension of the channel space L'+1, i.e. there are more unknowns that observed samples such that the estimation problem is underdetermined. Hence, the performance is expected to deteriorate in this case. A further performance prognosis can be given based on the results obtained for the two-path channel models: Dense multipath scenarios with similar or opposite phases and a small power ratio are most challenging. Due to the superposition of several rays per cluster with random starting phases, the phase offsets and power ratios of all paths are random and can not be influenced. Nevertheless, the Rician factor is a valuable indicator since it is defined as the average power ratio of the LOS path and the scattered components. The smaller the Rician factor, the smaller is the average LOS power and the worse the performance should be. From this point of view, the B3-LOS channel is most challenging. Furthermore, the relative excess delays and delay differences are of interest. Especially, the relative excess delay of the first multipath component ν_2/T_s is important since the corresponding pulse overlaps the most with the LOS component. For T_s=10 ns, the relative excess delays of the first multipath component are in a reasonable range, while they reduce to very small values in case of T_s=50 ns. The same is valid for the relative minimum delay differences. Very small delay differences lead to ill-conditioned (or even rank-deficient) Fisher information matrices, that are difficult to invert. In order to avoid a complete failure of the matrix inversion, singular value decomposition as described in [PTVF02, p. 62ff.] is applied for that purpose. Again, the B3-LOS is most challenging with respect to delay differences. Taking all parameters into account, the worst performance is expected for the B3-LOS channel (large indoor hall), whereas the best performance is predicted for the B1-LOS channel (urban micro-cell). In all cases, the performance for T_s=10 ns should be much better than for T_s=50 ns. In Figure 4.9, the CRLBs of the different WINNER channel models are plotted over the SNR. The curves for the problematic case of T_s=50 ns (underdetermined problem and very small delay differences) are shown on the right hand side of Figure 4.9. For symbol-rate sampling, and in the case of the B3-LOS channel also for oversampling, the CRLBs take on values larger than 10^4 that are meaningless high[5]. This means that the coarse synchronisation, which is performed before channel parameter estimation, can not

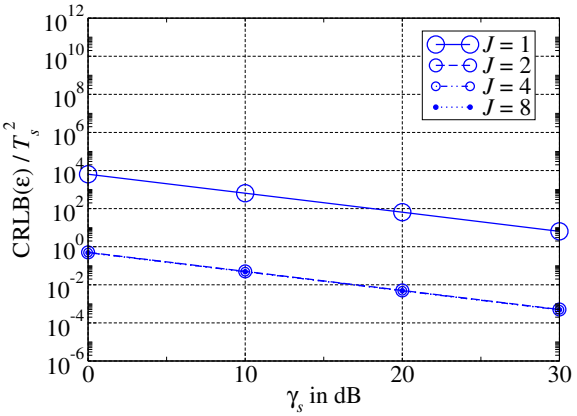
[5]The CRLB is normalised with respect to the squared symbol duration. If the non-normalised square root of the CRLB is taken into account, this issue becomes even more clear: In case of T_s=50 ns, a normalised CRLB of 10^4 translates into an a non-normalised root CRLB of 5 μs, which corresponds to a distance error of about 1.5 km.



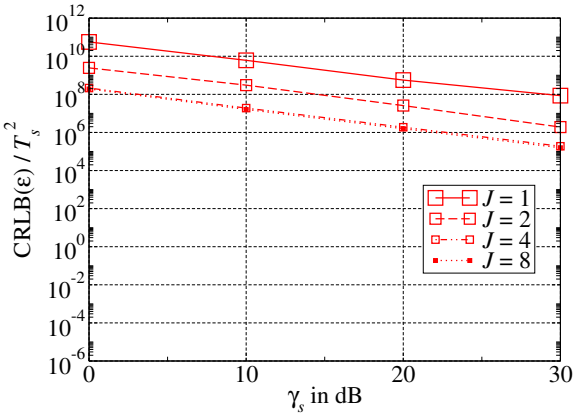
(a) B1-LOS with T_s=10 ns.



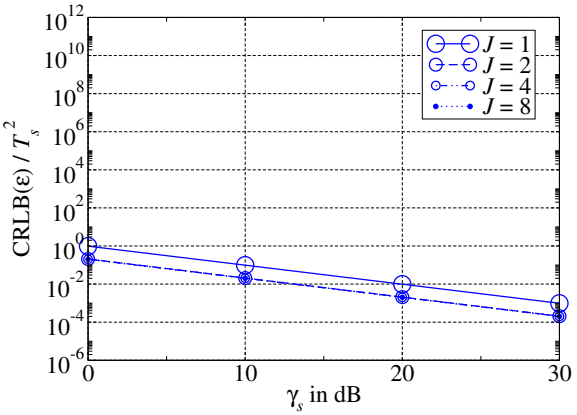
(b) B1-LOS with T_s=50 ns.



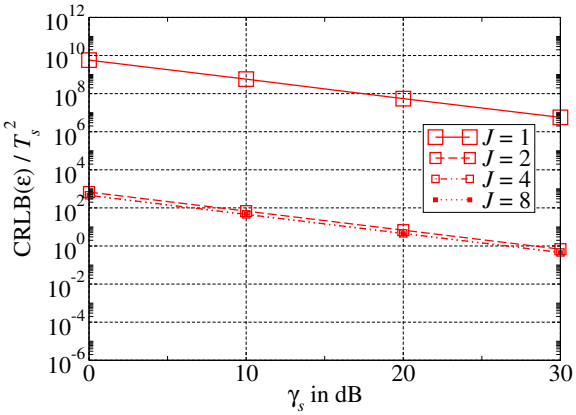
(c) B3-LOS with T_s=10 ns.



(d) B3-LOS with T_s=50 ns.



(e) C2-LOS with T_s=10 ns.



(f) C2-LOS with T_s=50 ns.

Figure 4.9: CRLB of ε normalised to T_s^2 over SNR for the WINNER channels.

be improved. Thus, channel parameter estimation is futile in such a case. If oversampling is applied, the problem of too less observations can be partly circumvented. Nevertheless, the problem may remain underdetermined because the samples, which are added by oversampling, are typically not independent of the previous samples but correlated to them. If the problem is well determined and the normalised excess delays are not too small, channel parameter estimation is able to improve the TOA estimate significantly as shown on the left hand side of Figure 4.9 for T_s=10 ns. The above performance prediction is met by all simulation results: The best and worst performance is obtained for the B1-LOS and the B3-LOS channel, respectively. Oversampling gains larger than 30 dB are possible as in the case of the B3-LOS channel (see Figure 4.9(c)). Hence, it is highly recommended to apply oversampling (with J=2) because accuracies well below the symbol duration can only be obtained for oversampling.

The above results can be summarised as follows: Oversampling is necessary for accurate parameter estimation. The number of parameters to estimate P should be smaller than the symbol-rate memory length L'. Thus, multipath components with similar delays may be combined to a single component in order to reduce the number of parameters.

4.4 Optimisation Algorithms for the ML Estimator

Before introducing and discussing potential optimisation algorithms, a short reminder is included at this point why optimisation algorithms are required: In the previous sections, two ML estimators with different metrics $\Omega_{\boldsymbol{y}}(\tilde{\boldsymbol{\theta}})$ and $\Omega_{\check{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}})$ have been derived for channel parameter estimation and it has been shown that both estimators are equivalent. As the second metric is advantageous from a complexity point of view, only this metric is considered subsequently:

$$\Omega(\tilde{\boldsymbol{\theta}}) = \Omega_{\check{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}}) = \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)^H \boldsymbol{C}_{\boldsymbol{\eta}}^{-1} \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)$$

with

$$\check{\boldsymbol{h}} = \left(\boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{X}_0\right)^{-1} \boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{y}$$

and

$$\boldsymbol{C}_{\boldsymbol{\eta}} = \left(\boldsymbol{X}_0^H \boldsymbol{C}_{\boldsymbol{n}}^{-1} \boldsymbol{X}_0\right)^{-1}.$$

Due to the nonlinearity of the function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$, the minimisation of the metric $\Omega(\tilde{\boldsymbol{\theta}})$ can not be solved in closed form. An exhaustive search is prohibitive because the search space is continuous and of high dimension (P=3C). Therefore, an optimisation algorithm is required in order to minimise $\Omega(\tilde{\boldsymbol{\theta}})$.

Since optimisation problems occur frequently in science and practice, research about numerical optimisation has always been of great interest and importance. There is a wide variety of optimisation algorithms that are tailored to the particular needs of different kinds of optimisation problems. A concise overview about optimisation can be found in

[PTVF02, Chap. 10], while more detailed information on optimisation is provided for example in [DS83], [GMW97] or [NW99]. In order to find a suitable optimisation method, several aspects of the problem at hand have to be taken into account and a tradeoff depending on the requirements needs to be found. The function to optimise is called *fitness function* in the following. According to [NW99, Chap. 1], optimisation can be characterised by four main categories: (i) *stochastic* versus *deterministic* optimisation, (ii) *discrete* versus *continuous* optimisation, (iii) *unconstrained* versus *constrained* optimisation and (iv) *local* versus *global* optimisation. The first category refers to the underlying model of the problem: If it is not possible to fully characterise the fitness function because it depends on quantities that may change and are not known yet, the problem is called stochastic. Such problems occur for example in the field of economics. If the fitness function can be fully characterised, the problem is deterministic as it is the case for channel parameter estimation. The second and third category refer to the domain of the fitness function (i.e. the search space): If the parameters on which the fitness function depend can only take on integer values, the problem at hand is discrete. Otherwise, it is continuous. If there are no constraints on the parameters, the problem is unconstrained, i.e. the search space corresponds to $\mathbb{R}^P$ and all parameter combinations are feasible. Typically, this is not the case since there are some natural constraints on the parameters. On the one hand, the values of a single parameter may be bounded (e.g. a propagation delay can not be negative). On the other hand, there may be constraints linking several parameters with each other (e.g. the sum of all parameters must be smaller than a certain value). If there are only bound constraints, unconstrained optimisation methods may be applied to constrained problems as well. The last category mentioned in [NW99, Chap. 1] refers to the fitness function itself. Often, a fitness function has several local optima in addition to the global optimum as illustrated in Figure 4.10, where minimisation is considered without loss of generality. A global optimisation method should be able to find the global optimum, whereas local optimisation algorithms are only able to find a local optimum in the neighbourhood of a predefined starting point called *initial guess*. It is much easier

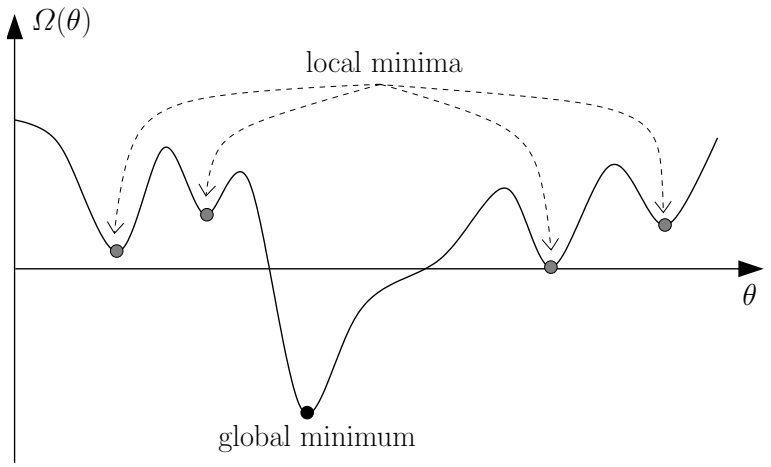


Figure 4.10: Example of a function with several local minima and one global minimum.

to find a local optimum instead of the global one. Unfortunately, only the global optimum is of interest in many applications. Sometimes, a priori information about the fitness function is available, that can be exploited to generate a good initial guess in the neighbourhood of the global optimum. In this case, a local optimisation method can be applied to a global optimisation problem. A further category, which is not mentioned in [NW99, Chap. 1], is related to the number of objective functions that need to be optimised simultaneously: If there is only a single objective function, the optimisation problem is called *single-objective*. Otherwise, it is denoted as *multi-objective*.

According to the above categories, channel parameter estimation is a single-objective, deterministic, continuous, bound constrained, global minimisation problem. Thus, a global optimisation method is required in general. However, *global* as well as *local* optimisation methods are considered in the following. For example, local optimisation methods can be effectively applied if the channel parameters vary *slowly* from burst to burst. In this case, the parameter estimate of the preceding burst can be used as initial guess for the current burst and a global method is only required for initialisation. This approach corresponds to a kind of tracking. But even in a one-shot estimation, local methods can be helpful. Actually, local methods might be sufficient if the fitness function is "good-natured", i.e. the function exhibits only few optima that are far away from each other. Hence, it is generally worth to consider global as well as local optimisation methods. Due to didactical reasons, local optimisation algorithms are discussed first in Section 4.4.1, while global methods are presented afterwards in Section 4.4.2. Without loss of generality, only minimisation is considered in the remainder of this chapter.

4.4.1 Local Optimisation Algorithms

As already mentioned above, local optimisation methods are only able to find a local optimum in the vicinity of an initial guess. The solution is improved iteratively until a stopping criterion is met. Many local optimisation algorithms are related to the *Newton method* [DS83], [GMW97], [NW99], [PTVF02, Chap. 10]. In case of the Newton method, the fitness function $\Omega(\tilde{\boldsymbol{\theta}})$ is approximated by a second order Taylor series according to

$$\Omega(\tilde{\boldsymbol{\theta}}) \approx \Omega(\tilde{\boldsymbol{\theta}}_{\text{cur}}) + \left(\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_{\text{cur}}\right)^T \boldsymbol{J}_\Omega(\tilde{\boldsymbol{\theta}}_{\text{cur}}) + \frac{1}{2}\left(\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_{\text{cur}}\right)^T \boldsymbol{H}_\Omega(\tilde{\boldsymbol{\theta}}_{\text{cur}}) \left(\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_{\text{cur}}\right), \quad (4.66)$$

where $\tilde{\boldsymbol{\theta}}_{\text{cur}}$ denotes the current hypothesis of the parameter vector. $\boldsymbol{J}_\Omega(\tilde{\boldsymbol{\theta}})$ and $\boldsymbol{H}_\Omega(\tilde{\boldsymbol{\theta}})$ are the Jacobian and the Hessian of the fitness function $\Omega(\tilde{\boldsymbol{\theta}})$, respectively. With this approximation, an improved hypothesis can be found based on

$$\frac{\partial \Omega(\tilde{\boldsymbol{\theta}})}{\partial \tilde{\boldsymbol{\theta}}} \approx \boldsymbol{J}_\Omega(\tilde{\boldsymbol{\theta}}_{\text{cur}}) + \boldsymbol{H}_\Omega(\tilde{\boldsymbol{\theta}}_{\text{cur}}) \underbrace{\left(\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_{\text{cur}}\right)}_{\Delta\tilde{\boldsymbol{\theta}}} \stackrel{!}{=} 0, \quad (4.67)$$

which leads to a *linear* set of equations with unknown difference vector $\Delta\tilde{\boldsymbol{\theta}}$, that can be solved in closed form:

$$\Delta\tilde{\boldsymbol{\theta}} = -\boldsymbol{H}_\Omega^{-1}(\tilde{\boldsymbol{\theta}}_{\text{cur}}) \cdot \boldsymbol{J}_\Omega(\tilde{\boldsymbol{\theta}}_{\text{cur}}). \quad (4.68)$$

Based on this solution, a new hypothesis can be determined:

$$\tilde{\boldsymbol{\theta}}_{\text{new}} = \tilde{\boldsymbol{\theta}}_{\text{cur}} + \Delta\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}_{\text{cur}} - \boldsymbol{H}_{\Omega}^{-1}(\tilde{\boldsymbol{\theta}}_{\text{cur}}) \cdot \boldsymbol{J}_{\Omega}(\tilde{\boldsymbol{\theta}}_{\text{cur}}). \tag{4.69}$$

This procedure is repeated iteratively until an optimum point $\hat{\boldsymbol{\theta}}$ is found. A general problem of the Newton method is that it can not intrinsically distinguish between minimum and maximum points [DS83, p. 101]: Generally, the Jacobian at the optimum point is required to be zero: $\boldsymbol{J}_{\Omega}(\hat{\boldsymbol{\theta}})=0$. In order to determine whether the optimum point is a minimum or a maximum, the properties of the Hessian at the optimum point, $\boldsymbol{H}_{\Omega}(\hat{\boldsymbol{\theta}})$, are relevant: In case of minimisation, it is a necessary condition that the Hessian is *positive semidefinite*, while it is a sufficient condition when the Hessian is *positive definite* [DS83, p. 81]. Actually, the Hessian should be positive definite during the whole iterative process in order to ensure that the update step in (4.69) is taken in a *descent* direction [DS83, p. 101], [GMW97, p. 106], [NW99, p. 52]. This requirement is typically fulfilled if the initial guess is sufficiently close to a local minimum. But, in general, the Newton method can not be prevented from taking an update step in an uphill direction. Another drawback of the Newton method is that first *and* second order derivatives of the fitness function $\Omega(\tilde{\boldsymbol{\theta}})$ need to be known. Often, it is neither desirable nor feasible to determine second order derivatives of the fitness function (i.e. the Hessian). Therefore, *quasi-Newton* methods are commonly applied, which are also called *variable metric methods* [DS83, Sec. 6.1], [GMW97, Sec. 4.5.2], [NW99, Chap. 8], [PTVF02, Sec. 10.7]. For quasi-Newton methods, only the Jacobian of the fitness function is required, while the Hessian is approximated. A significant advantage of quasi-Newton methods is that the Hessian can be coerced to be positive definite such that each update step is forced in a descent direction. Famous examples for quasi-Newton methods are the steepest descent and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [NW99], [PTVF02, Chap. 10].

Special algorithms have been developed for the class of *nonlinear least-squares (NLS)* problems. Since ML estimators correspond to LS estimators in the case of Gaussian noise, these methods are well suited for channel parameter estimation. The form of the fitness function is in principle the same for all NLS problems: A model function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ is fitted to observations $\check{\boldsymbol{h}}$ by minimising the (squared) difference between both, usually taking the covariance matrix $\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}$ of the distortion into account:

$$\Omega(\tilde{\boldsymbol{\theta}}) = \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)^{H} \boldsymbol{C}_{\boldsymbol{\eta}}^{-1} \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right).$$

Hence, general expressions for the Jacobian and the Hessian can be derived in case of NLS problems: The Jacobian is a column vector of length P and the Hessian is a symmetric matrix of size $P \times P$, where P denotes the number of parameters to estimate. The pth entry of the Jacobian is defined as

$$\begin{aligned}\left[\boldsymbol{J}_{\Omega}(\tilde{\boldsymbol{\theta}})\right]_p &= \frac{\partial\Omega(\tilde{\boldsymbol{\theta}})}{\partial\tilde{\theta}_p} = \frac{\partial}{\partial\tilde{\theta}_p}\left\{\left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)^{H} \boldsymbol{C}_{\boldsymbol{\eta}}^{-1} \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)\right\} \\ &= -2\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{h}^{H}(\tilde{\boldsymbol{\theta}})}{\partial\tilde{\theta}_p} \boldsymbol{C}_{\boldsymbol{\eta}}^{-1} \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)\right\}.\end{aligned} \tag{4.70}$$

Based on (4.70), a single entry of the Hessian can be determined according to

$$\begin{aligned}\left[\boldsymbol{H}_{\Omega}(\tilde{\boldsymbol{\theta}})\right]_{pq} &= \frac{\partial^2 \Omega(\tilde{\boldsymbol{\theta}})}{\partial\tilde{\theta}_p \partial\tilde{\theta}_q} = -\frac{\partial}{\partial\tilde{\theta}_q} 2\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{h}^H(\tilde{\boldsymbol{\theta}})}{\partial\tilde{\theta}_p}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)\right\} \\ &= 2\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{h}^H(\tilde{\boldsymbol{\theta}})}{\partial\tilde{\theta}_p}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\frac{\partial \boldsymbol{h}(\tilde{\boldsymbol{\theta}})}{\partial\tilde{\theta}_q} - \frac{\partial^2 \boldsymbol{h}^H(\tilde{\boldsymbol{\theta}})}{\partial\tilde{\theta}_p \partial\tilde{\theta}_q}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)\right\}.\end{aligned} \tag{4.71}$$

Similar calculations have been performed in Section 4.3.2, where the Fisher information matrix has been determined. The main difference between the expressions in Section 4.3.2 and the above expressions is that the model function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ and its derivatives are examined at an arbitrary parameter hypothesis $\tilde{\boldsymbol{\theta}}$ here instead of the true parameter vector $\boldsymbol{\theta}$. In (4.40), second order derivatives similar to (4.71) have been derived. The summand with the second order derivatives of the model function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ and the difference between the observations and this function has been omitted in (4.40) due to taking the expectation: The difference between the observations, $\check{\boldsymbol{h}}$, and the model function evaluated at the true parameters, $\boldsymbol{h}(\boldsymbol{\theta})$, corresponds to the measurement error, $\boldsymbol{\eta}$, that is by definition zero mean. Similarly, the second part in (4.71) is neglected in case of the *Gauss-Newton* method:

$$\left[\boldsymbol{H}_{\Omega}(\tilde{\boldsymbol{\theta}})\right]_{pq} \approx 2\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{h}^H(\tilde{\boldsymbol{\theta}})}{\partial\tilde{\theta}_p}\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\frac{\partial \boldsymbol{h}(\tilde{\boldsymbol{\theta}})}{\partial\tilde{\theta}_q}\right\}. \tag{4.72}$$

The above approximation is reasonable if the second part in (4.71) is small compared to the first part. This assumption is true in two distinct cases: Either in case of a *nearly linear* model function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ (i.e. the second order derivatives tend to be zero) or in case of a so-called *small-residual* problem [DS83, p. 222], [GMW97, p. 134], [NW99, p. 260] [PTVF02, p. 688]. A small-residual problem is defined as follows: The value of the fitness function at the minimum, $\Omega(\hat{\boldsymbol{\theta}})$, needs to be smaller than the small eigenvalues of the approximated Hessian in (4.72) evaluated at the parameter estimate, $\boldsymbol{H}_{\Omega}(\hat{\boldsymbol{\theta}})$. In other words: The difference between the observations, $\check{\boldsymbol{h}}$, and model function evaluated at the parameter estimate, $\boldsymbol{h}(\hat{\boldsymbol{\theta}})$, needs to be small. In case of a large residual problem, the value of the fitness function at the minimum, $\Omega(\hat{\boldsymbol{\theta}})$, exceeds the small eigenvalues of the approximated Hessian $\boldsymbol{H}_{\Omega}(\hat{\boldsymbol{\theta}})$ [GMW97, p. 137]. In this case, the second part in (4.71) is not negligible anymore such that the Gauss-Newton method, which is based on the approximated Hessian, may fail. Fortunately, many practical NLS problems exhibit small residuals. However, it should be kept in mind that a large residual might be the cause whenever a problem in conjunction with the Gauss-Newton method occurs.

Practically, it is often advantageous if (4.72) instead of (4.71) is considered because the probability that the Hessian is positive definite is increased. In fact, the approximate Hessian is always positive semidefinite [NW99, p. 264]. However, it can not be ensured that the approximated Hessian in (4.72) is positive definite. Therefore, modified versions of the Gauss-Newton are often applied to NLS problems. A standard method for solving NLS problems is the *Levenberg-Marquardt (LM)* method [DS83, p. 227f.], [GMW97, p. 136f.] [NW99, p. 262-266], [PTVF02, p. 688-693]. Basically, the LM method is a combination of the Gauss-Newton and the steepest descent method [PTVF02, p. 688]. Using the

Gauss-Newton method, a new hypothesis can be obtained based on a current hypothesis as already described in (4.69):

$$\tilde{\boldsymbol{\theta}}_{\text{new}} = \tilde{\boldsymbol{\theta}}_{\text{cur}} - \boldsymbol{H}_{\Omega}^{-1}(\tilde{\boldsymbol{\theta}}_{\text{cur}}) \cdot \boldsymbol{J}_{\Omega}(\tilde{\boldsymbol{\theta}}_{\text{cur}}), \tag{4.73}$$

in which the Jacobian and the Hessian are given by (4.70) and (4.72), respectively. More compact expressions for the Jacobian and the Hessian of the fitness function are obtained with the Jacobian $\boldsymbol{J}(\tilde{\boldsymbol{\theta}})$ of the model function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ as defined in (4.36):

$$\boldsymbol{J}_{\Omega}(\tilde{\boldsymbol{\theta}}) = -2\,\text{Re}\left\{\boldsymbol{J}^{H}(\tilde{\boldsymbol{\theta}})\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)\right\}, \tag{4.74}$$

$$\boldsymbol{H}_{\Omega}(\tilde{\boldsymbol{\theta}}) \approx \quad 2\,\text{Re}\left\{\boldsymbol{J}^{H}(\tilde{\boldsymbol{\theta}})\boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\boldsymbol{J}(\tilde{\boldsymbol{\theta}})\right\}. \tag{4.75}$$

It is worth noting again that the Hessian in (4.75) is of the same form as the Fisher information matrix in (4.42). Near a local minimum, the underlying approximation according to the second order Taylor series in (4.66) is typically good and the Hessian is positive definite. In this case, the Gauss-Newton method is a very effective way to approach the minimum. But, far away from a local minimum, the approximation in (4.66) may be poor and the Hessian may not be positive definite. In that case, it is usually better to choose a new hypothesis along the gradient according to the steepest descent method:

$$\tilde{\boldsymbol{\theta}}_{\text{new}} = \tilde{\boldsymbol{\theta}}_{\text{cur}} - \text{constant} \cdot \boldsymbol{J}_{\Omega}(\tilde{\boldsymbol{\theta}}_{\text{cur}}). \tag{4.76}$$

The constant in (4.76) determines how large the step along the gradient is and should be chosen with care. If the step length is too short, it takes many iterations in order to reach the local minimum. If the step length is too large, the algorithm jumps over the region with downhill gradient and, thus, over the local minimum. The Hessian contains information about the curvature of the fitness function, that can be used to choose an appropriate step length. Here lies the connection between the Gauss-Newton and the steepest descent method, which is exploited in the LM method: Comparing (4.73) and (4.76) shows that the Gauss-Newton and steepest descent method only differ in the factor with which the Jacobian is multiplied. In the former case, it is the inverse of the Hessian and, in the latter case, it is a constant that can be chosen adequately with the help of the Hessian. Against this background, the update step can be generalised as

$$\tilde{\boldsymbol{\theta}}_{\text{new}} = \tilde{\boldsymbol{\theta}}_{\text{cur}} - \Big(\underbrace{\boldsymbol{H}_{\Omega}(\tilde{\boldsymbol{\theta}}_{\text{cur}}) + \omega\boldsymbol{\mathcal{I}}}_{\boldsymbol{\mathcal{H}}_{\Omega}(\tilde{\boldsymbol{\theta}}_{\text{cur}})}\Big)^{-1} \cdot \boldsymbol{J}_{\Omega}(\tilde{\boldsymbol{\theta}}_{\text{cur}}), \tag{4.77}$$

where ω is called *tuning factor*. With the help of the tuning factor, the diagonal entries of the Hessian are manipulated: If the tuning factor ω is large enough, the modified Hessian $\boldsymbol{\mathcal{H}}_{\Omega}(\tilde{\boldsymbol{\theta}})$ becomes positive definite, i.e. the update step is forced in a descent direction. If the tuning factor ω is close to zero, it is obvious that the update in (4.77) corresponds to the Gauss-Newton method in (4.73). If the tuning factor ω is very large, the matrix $\boldsymbol{\mathcal{H}}_{\Omega}(\tilde{\boldsymbol{\theta}})$ becomes diagonally dominant. In this case, the update in (4.77) corresponds to the steepest descend method in (4.76). With this knowledge, the LM method can be summarised as shown in Algorithm 4.1 [PTVF02, p. 689ff.]. The LM method can be

Algorithm 4.1 Levenberg-Marquardt method.

1. **Initialisation:**

   Choose an initial guess $\tilde{\boldsymbol{\theta}}_0$ and a moderate value for the tuning factor (e.g. ω_0=0.001).

   Compute the metric $\Omega(\tilde{\boldsymbol{\theta}}_0)$ as well as its Jacobian $\boldsymbol{J}_\Omega(\tilde{\boldsymbol{\theta}}_0)$ and Hessian $\boldsymbol{H}_\Omega(\tilde{\boldsymbol{\theta}}_0)$.

2. **Repeat until a stopping criterion is met:**

   Find a new hypothesis via (4.77): $\tilde{\boldsymbol{\theta}}_{n+1} = \tilde{\boldsymbol{\theta}}_n - \left(\boldsymbol{H}_\Omega(\tilde{\boldsymbol{\theta}}_n) + \omega_n \boldsymbol{\mathcal{I}}\right)^{-1} \cdot \boldsymbol{J}_\Omega(\tilde{\boldsymbol{\theta}}_n)$.

   (a) If $\Omega(\tilde{\boldsymbol{\theta}}_{n+1}) \geq \Omega(\tilde{\boldsymbol{\theta}}_n)$:
   - Increase the tuning factor by a factor of 10: ω_{n+1}=10·ω_n.
   - Return to the old hypothesis: $\tilde{\boldsymbol{\theta}}_{n+1}$=$\tilde{\boldsymbol{\theta}}_n$.

   (b) If $\Omega(\tilde{\boldsymbol{\theta}}_{n+1}) < \Omega(\tilde{\boldsymbol{\theta}}_n)$:
   - Decrease the tuning factor by a factor of 10: ω_{n+1}=0.1·ω_n.
   - Accept the new hypothesis: $\tilde{\boldsymbol{\theta}}_{n+1} = \tilde{\boldsymbol{\theta}}_n - \left(\boldsymbol{H}_\Omega(\tilde{\boldsymbol{\theta}}_n) + \omega_n \boldsymbol{\mathcal{I}}\right)^{-1} \cdot \boldsymbol{J}_\Omega(\tilde{\boldsymbol{\theta}}_n)$.
   - Determine the new Jacobian $\boldsymbol{J}_\Omega(\tilde{\boldsymbol{\theta}}_{n+1})$ and Hessian $\boldsymbol{H}_\Omega(\tilde{\boldsymbol{\theta}}_{n+1})$.

stopped if the fitness improves (i.e. case (b)) and the absolute value of the improvement is smaller than a certain threshold (e.g. $\left|\Omega(\tilde{\boldsymbol{\theta}}_{n+1}) - \Omega(\tilde{\boldsymbol{\theta}}_n)\right| < 10^{-3}$). This criterion can be intensified if it is required to be fulfilled for several times in a sequence (e.g. two or five times) [PTVF02, p. 689ff.]. Of course, other stopping criteria can be chosen as well. Often, the maximum number of iterations is additionally limited to a reasonable number.

When the algorithm has stopped, the final estimate $\hat{\boldsymbol{\theta}}$ and the corresponding fitness $\Omega(\hat{\boldsymbol{\theta}})$ are returned. Furthermore, a covariance matrix for the parameter estimate $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}$ can be determined. Remembering the similarity between the Fisher information matrix in (4.42) and the Hessian in (4.75), it suggests itself to determine the inverse of the Hessian to compute a covariance matrix:

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} \approx \boldsymbol{H}_\Omega^{-1}(\hat{\boldsymbol{\theta}}). \tag{4.78}$$

Since the Hessian in (4.75) is only approximate, the covariance matrix in (4.78) is only approximate as well. Therefore, it should be used with caution. More information about this topic is provided in Section 4.5. Since the LM method is a local optimiser, its performance depends on the choice of the initial guess: If the initial guess is close to the global optimum, the LM algorithm will be able to find the global optimum. Otherwise, it will converge to a local optimum in the neighbourhood of the initial guess. The LM algorithm is one of the methods that is applied to channel parameter estimation in the remainder of this thesis.

Another local optimisation algorithm, that is considered in this thesis, is the *expectation-maximisation (EM)* algorithm and its generalisation, the *space-alternating*

generalised expectation-maximisation (SAGE) algorithm. The EM algorithm is applicable to ML problems and was first introduced in a general[6] form by Dempster, Laird and Rubin in 1977 [DLR77]. 27 years later, Fessler and Hero published a further generalised version of the EM method, which they named SAGE algorithm [FH94]. Since the EM and the SAGE algorithm are applicable to ML problems, they have been used extensively in many fields including parameter estimation. As will be explained later, both algorithms are especially suitable for parameter estimation of superimposed signals [FW88], [FDHT96], [FTH+99], [RLT03]. Hence, they are well suited to the channel parameter estimation problem considered in this thesis. Therefore, both algorithms are discussed in more detail subsequently.

At first, the notion of *incomplete* and *complete* data needs to be introduced: Incomplete data corresponds to the observations, on which the estimation process is based. Incomplete data is obtained from complete data by a non-bijective (many-to-one) mapping [DLR77]. Hence, complete data is *not* directly observable. Through the incorporation of complete data, the estimation process can be simplified because a high-dimensional problem is split into several (parallel) subproblems of lower dimensionality. As explicitly noted in [FW88], there is no unique choice of the complete data: All combinations of data sets that can be mapped onto the incomplete data via a non-bijective mapping are feasible. However, in case of superimposed signals, there is a *natural* mapping provided by the superposition itself: Each summand corresponds to a single complete data set and all sets are mapped onto the incomplete data via the summation [FW88], [FDHT96], [FTH+99]. For this reason, the EM and SAGE algorithm are especially suitable for parameter estimation of superimposed signals. In case of the channel parameter estimator considered in this thesis, the observed pre-stage channel estimates $\check{\boldsymbol{h}}$ (incomplete data) are split into individual multipath components $\check{\boldsymbol{h}}_c$ (complete data) according to the natural mapping:

$$\check{\boldsymbol{h}} = \sum_{c=1}^{C} \check{\boldsymbol{h}}_c \quad \text{with} \quad \check{\boldsymbol{h}}_c = \boldsymbol{h}_c(\boldsymbol{\theta}_c) + \boldsymbol{\eta}_c, \tag{4.79}$$

where C denotes the number of multipath components. Each multipath component $\check{\boldsymbol{h}}_c$ depends on a part $\boldsymbol{\theta}_c$ of the full parameter vector $\boldsymbol{\theta}$ only:

$$\boldsymbol{\theta} = \big[\underbrace{\mathrm{Re}\{f_1\}, \mathrm{Im}\{f_1\}, \varrho_1}_{\boldsymbol{\theta}_1^T}, \cdots, \underbrace{\mathrm{Re}\{f_C\}, \mathrm{Im}\{f_C\}, \varrho_C}_{\boldsymbol{\theta}_C^T}\big]^T. \tag{4.80}$$

In accordance with (4.79), the model function $\boldsymbol{h}(\boldsymbol{\theta})$ and the estimation error $\boldsymbol{\eta}$ are split into multipath components as well:

$$\boldsymbol{h}(\boldsymbol{\theta}) = \sum_{c=1}^{C} \boldsymbol{h}_c(\boldsymbol{\theta}_c) \quad \text{and} \quad \boldsymbol{\eta} = \sum_{c=1}^{C} \boldsymbol{\eta}_c. \tag{4.81}$$

[6] Dempster, Laird and Robin explicitly mention that the EM algorithm had been proposed before, but only under certain conditions: "The EM algorithm has been proposed many times in special circumstances" [DLR77, p. 3].

Algorithm 4.2 Expectation-maximisation method.

1. **Initialisation:**

   Choose an initial guess $\hat{\boldsymbol{\theta}}^0$.

2. **Repeat until a stopping criterion is met:**

   E-step: For all $c{=}1,\cdots,C$ do

   – $\check{\boldsymbol{h}}_c^n = \boldsymbol{h}_c(\hat{\boldsymbol{\theta}}_c^n) + \beta_c^n \left(\check{\boldsymbol{h}} - \sum_{c'=1}^{C} \boldsymbol{h}_{c'}(\hat{\boldsymbol{\theta}}_{c'}^n)\right)$ with $\beta_c^n{\geq}0$ and $\sum_{c=1}^{C} \beta_c^n \overset{!}{=} 1$.

   M-step: For all $c{=}1,\cdots,C$ do

   – $\hat{\boldsymbol{\theta}}_c^{n+1} = \arg\min_{\tilde{\boldsymbol{\theta}}_c} \left\{ \left(\check{\boldsymbol{h}}_c^n - \boldsymbol{h}_c(\tilde{\boldsymbol{\theta}}_c)\right)^H \left(\beta_c^n \boldsymbol{C}_{\boldsymbol{\eta}}\right)^{-1} \left(\check{\boldsymbol{h}}_c^n - \boldsymbol{h}_c(\tilde{\boldsymbol{\theta}}_c)\right) \right\}$.

The splitting of the model function $\boldsymbol{h}(\boldsymbol{\theta})$ is specified by the parameter vector part $\boldsymbol{\theta}_c$,

$$h_{c,l}(\boldsymbol{\theta}_c) = f_c \cdot g\left(l\mathcal{T} - \varrho_c\right), \tag{4.82}$$

while the estimation error $\boldsymbol{\eta}$ can be split arbitrarily. Typically, the components $\boldsymbol{\eta}_c$ are assumed to be statistically independent, zero mean Gaussian processes with covariance matrix $\boldsymbol{C}_{\boldsymbol{\eta}_c}{=}\beta_c{\cdot}\boldsymbol{C}_{\boldsymbol{\eta}}$, where $\beta_c{\geq}0$ and $\sum_{c=1}^{C} \beta_c \overset{!}{=} 1$ [FW88]. Based on the above decompositions, the EM method can be formulated as summarised in Algorithm 4.2: In each iteration n, two steps are performed. Given the current parameter estimate $\hat{\boldsymbol{\theta}}^n$, the conditional expectation of each complete data set is determined (*E-step*):

$$\check{\boldsymbol{h}}_c^n = \boldsymbol{h}_c(\hat{\boldsymbol{\theta}}_c^n) + \boldsymbol{\eta}_c^n = \boldsymbol{h}_c(\hat{\boldsymbol{\theta}}_c^n) + \beta_c^n \left(\check{\boldsymbol{h}} - \sum_{c'=1}^{C} \boldsymbol{h}_{c'}(\hat{\boldsymbol{\theta}}_{c'}^n)\right). \tag{4.83}$$

Afterwards, a new parameter estimate $\hat{\boldsymbol{\theta}}^{n+1}$ is obtained by optimising each component separately using the expectations $\check{\boldsymbol{h}}_c^n$ (*M-step*):

$$\hat{\boldsymbol{\theta}}_c^{n+1} = \arg\min_{\tilde{\boldsymbol{\theta}}_c} \left\{ \left(\check{\boldsymbol{h}}_c^n - \boldsymbol{h}_c(\tilde{\boldsymbol{\theta}}_c)\right)^H \left(\beta_c^n \boldsymbol{C}_{\boldsymbol{\eta}}\right)^{-1} \left(\check{\boldsymbol{h}}_c^n - \boldsymbol{h}_c(\tilde{\boldsymbol{\theta}}_c)\right) \right\}. \tag{4.84}$$

The involved optimisation of the subproblems is of lower dimensionality compared to the overall problem and is, thus, simpler. However, the metric of the subproblems is still nonlinear due to the nonlinear model functions $\boldsymbol{h}_c(\tilde{\boldsymbol{\theta}}_c)$. Therefore, an optimisation algorithm needs to be applied to the subproblems as well. In this thesis, the LM method is utilised for this purpose. Of course, any other optimisation method is applicable as well.

The iterative process of the EM method is repeated until a stopping criterion is met. Similar to the LM method, the EM algorithm can be stopped if the overall fitness improvement is smaller than a certain threshold (e.g. $\left|\Omega(\hat{\boldsymbol{\theta}}^{n+1}){-}\Omega(\hat{\boldsymbol{\theta}}^n)\right|{<}10^{-5}$). Again, the

stopping criterion can be intensified if it is required to be fulfilled for several times in a sequence (e.g. five or ten times). It should be noted here that the notation in conjunction with the EM (and SAGE) algorithm is slightly changed compared to the LM method: Actually, a parameter estimate is *first* obtained when the iterative process is stopped. During the iterations, the encountered parameter vectors are rather hypotheses than estimates. However, in the context of the EM (and SAGE) algorithm, the parameter hypotheses encountered in the *E-step* are denoted as (current) parameter estimates $\hat{\boldsymbol{\theta}}_c^n$ in order distinguish them from the parameter hypotheses $\tilde{\boldsymbol{\theta}}_c$, which appear in the optimisation process of the subproblems during the *M-step*.

A significant characteristic of the EM method is that the overall fitness $\Omega(\hat{\boldsymbol{\theta}}^n)$ decreases monotonically over the iterations n such that the likelihood is monotonically increased [DLR77], [FW88]. Furthermore, it is guaranteed that the EM algorithm converges to a stationary point of the fitness function. However, the EM method is still a local optimiser and does not necessarily converge to the global optimum. Its performance depends on the choice of the initial guess as it is the case for the LM method. As already mentioned above, an advantage of the EM method is that the overall optimisation process can be simplified because a high-dimensional problem is split into several low-dimensional subproblems. This advantage is pronounced the simpler the subproblems are: Ideally, the subproblems are one-dimensional and/or can be solved in closed form. In this thesis, the $3C$-dimensional channel parameter estimation problem is split into C three-dimensional subproblems. Further simplifications are possible, but are not considered here because the focus lies on the basic differences between the (full) LM method and the (separated) EM method with respect to different performance criteria like the robustness against the IG or the convergence rate.

With respect to the convergence rate, the EM algorithm may be disadvantageous because the simplification of the optimisation process comes at the cost of a rather slow convergence rate according to [FW88] and [FH94]. This is one motivation for the SAGE algorithm, whose convergence rate is improved compared to the EM algorithm [FH94].[7] Fessler and Hero introduce so-called *admissible hidden data sets*, which are defined as those data sets that depend on a parameter part only and are complete if the remaining parameters are assumed to be known [FH94], [FDHT96], [FTH+99]. For channel parameter estimation as proposed in this thesis, each admissible hidden data set corresponds to a single multipath component $\check{\boldsymbol{h}}_c^n$ that depends on the parameter part $\hat{\boldsymbol{\theta}}_c^n$ only. In contrast to the EM algorithm, the admissible hidden data sets are updated *sequentially* rather than parallelly. This means that the *E-step* and *M-step* are performed separately for each admissible hidden data set, while the remaining parameter estimates are kept constant as shown in Algorithm 4.3. A full iteration cycle is accomplished when the *E-* and *M-steps* have been performed for *all* multipath components. Due to the sequential structure of the SAGE method, an auxiliary variable $\boldsymbol{\Psi}^n{=}\hat{\boldsymbol{\theta}}^n$ is introduced, whose parts are updated after each *M-step*. This means that $\boldsymbol{\Psi}^n$ contains all $c{-}1$ updated parameter

[7] A second motivation for the SAGE algorithm is the desired or even necessary usage of so-called smoothness penalties, that "usually render intractable the maximisation steps of classical EM methods" [FH94, p. 1]. Here, only the first motivation is relevant.

Algorithm 4.3 Space-alternating generalised expectation-maximisation method.

1. **Initialisation:**

   Choose an initial guess $\hat{\boldsymbol{\theta}}^0$.

2. **Repeat until a stopping criterion is met:**

   Initialise auxiliary variable: $\boldsymbol{\Psi}^n{=}\hat{\boldsymbol{\theta}}^n$.

   For all $c{=}1,\cdots,C$ do

   - *E-step:* $\check{\boldsymbol{h}}_c^n = \boldsymbol{h}_c(\boldsymbol{\Psi}_c^n) + \left(\check{\boldsymbol{h}} - \sum_{c'=1}^{C} \boldsymbol{h}_{c'}(\boldsymbol{\Psi}_{c'}^n)\right)$
   - *M-step:* $\hat{\boldsymbol{\theta}}_c^{n+1} = \arg\min_{\tilde{\boldsymbol{\theta}}_c} \left\{ \left(\check{\boldsymbol{h}}_c^n - \boldsymbol{h}_c(\tilde{\boldsymbol{\theta}}_c)\right)^H \boldsymbol{C}_{\boldsymbol{\eta}}^{-1} \left(\check{\boldsymbol{h}}_c^n - \boldsymbol{h}_c(\tilde{\boldsymbol{\theta}}_c)\right) \right\}$
   - Update corresponding part of the auxiliary variable: $\boldsymbol{\Psi}_c^n{=}\hat{\boldsymbol{\theta}}_c^{n+1}$.

estimates for the *E-step* of the cth multipath component:

$$\boldsymbol{\Psi}^n = \left[\left(\hat{\boldsymbol{\theta}}_1^{n+1}\right)^T, \cdots, \left(\hat{\boldsymbol{\theta}}_{c-1}^{n+1}\right)^T, \left(\hat{\boldsymbol{\theta}}_c^{n}\right)^T, \left(\hat{\boldsymbol{\theta}}_{c+1}^{n}\right)^T, \cdots, \left(\hat{\boldsymbol{\theta}}_C^{n}\right)^T \right]^T. \tag{4.85}$$

As for the EM algorithm, the overall fitness $\Omega(\hat{\boldsymbol{\theta}}^n)$ decreases monotonically over iterations n and it is guaranteed that the SAGE method converges to a stationary point of the fitness function. According to Fessler and Hero, the convergence rate of the EM algorithm is "inversely related to the Fisher information of its complete data space" [FH94, p. 1]. This means that less informative complete data is advantageous with respect to the convergence rate. In case of the SAGE algorithm, the Fisher information of the admissible hidden data is maximised because the total estimation error is considered for each multipath component:

$$\boldsymbol{\eta}_c^n = \beta_c^n \cdot \boldsymbol{\eta}^n \overset{\beta_c^n=1}{=} \boldsymbol{\eta}^n = \check{\boldsymbol{h}} - \sum_{c'=1}^{C} \boldsymbol{h}_{c'}(\boldsymbol{\Psi}_{c'}). \tag{4.86}$$

Therefore, the convergence rate of the SAGE algorithm is improved compared to the EM method. The performance of the LM, EM and SAGE algorithm for channel parameter estimation is examined in Section 4.4.3. Previous to this performance analysis, global optimisation algorithms are discussed in the following section.

4.4.2 Global Optimisation Algorithms

The simplest approach to perform global optimisation, besides an exhaustive search, is to use a local optimiser with many different starting points and then choosing the best solution with the smallest fitness. This means that the local optimisation algorithms described above may be applied for global optimisation in conjunction with a *good* initial guess or

a *certain number* of different initial guesses. This approach is indeed simple, but not necessarily satisfactory. Many different global optimisation algorithms have been invented and investigated in the literature. Many of these methods involve function evaluations only, i.e. derivative information is usually *not* required. Global optimisation algorithms can be roughly divided into *deterministic* and *probabilistic/stochastic* approaches [PV02], [Wei09]. This classification is based on the fact whether the algorithm contains random elements or not. In [Neu04], Neumaier proposes a more elaborate classification into *incomplete*, *asymptotically complete*, *complete* and *rigorous* global optimisation methods based on the level of sophistication. Incomplete methods are rather simple and fast, but they are least reliable since they can not be prevented from premature convergence to a local optimum. Asymptotically complete methods are able to definitely find the global optimum assuming an infinite/unlimited number of iterations, but they are not able to decide whether the global optimum has already been found or not. Complete and rigorous algorithms are more sophisticated since they can additionally determine when an approximate global optimum has been found (after a finite number of iterations) given some predefined tolerances. These two types of algorithms only differ in the fact that exact calculations are assumed for complete methods, while rigorous algorithms can also handle inaccurate calculations due to rounding errors. Obviously, there is a tradeoff between time consumption/complexity and reliability. Deterministic methods can be of any kind, while probabilistic methods are typically incomplete. Even though the incomplete methods may perform quite well, there is *no guarantee* that the global optimum has been found. Deterministic approaches are often based on the concept of *divide and conquer/branching* [Neu04], [Wei09]. Furthermore, deterministic approaches may exploit some special properties of the optimisation problem that simplify the optimisation process (e.g. convex optimisation [BV04] or Lipschitzian optimisation [Pin96]). Examples for (complete) deterministic methods comprise the *branch and bound (B&B)* algorithm [LD60] for discrete global optimisation as well as the *dividing rectangles (DIRECT)* algorithm [JPS93] and the *multilevel coordinate search (MCS)* [HN99] for continuous global optimisation. Even though these complete algorithms exist, incomplete probabilistic methods are widely applied. The reason for the success of probabilistic methods is probably that they perform quite well and are comparably fast: Complete methods become intractable for very complicated and/or high-dimensional optimisation problems. Therefore, the loss of a global convergence guarantee in case of incomplete methods becomes acceptable. Or as Weise states: "..., a solution a little bit inferior to the best possible one is better than one which needs 10^{100} years to be found ..." [Wei09, p. 22]. Since channel parameter estimation is of high dimension, only probabilistic methods are considered in this thesis. A concise overview about probabilistic methods can be found for example in [Neu04, Sec. 7] or [CZ08, Chap. 14], while a more detailed discussion on probabilistic methods is provided in [Wei09]. Many probabilistic algorithms are inspired by optimisation processes in nature. Three famous examples are *simulated annealing (SA)*, *genetic algorithms (GAs)* and *particle swarm optimisation (PSO)*. In this thesis, PSO is considered for global optimisation. However, the basic concepts of all three methods are presented subsequently in order to justify the application of the PSO method.

SA has been proposed by Kirkpatrick, Gelatt and Vecchi in 1983 [KGV83]. Origi-

nally, SA was designed for combinatorial/discrete optimisation problems like the travelling salesman problem. In conjunction with the downhill simplex method by Nelder and Mead [NM65], it can also be applied to continuous problems [PTVF02, Sec. 10.9].[8] SA is inspired by thermodynamics/statistical mechanics: The state in which a liquid or a metal ends up when cooled down depends on the way the material is annealed. If the material is cooled *slowly*, the atoms are able to arrange themselves in a pure crystal, which corresponds to a state of minimum energy. If the material is cooled too fast, the atoms are not able to reach this minimum energy state. As a result, a polycrystalline or amorphous structure is obtained. Therefore, slow annealing is preferable. But how can this insight be applied to optimisation? For minimisation, it is always desirable to move in a downhill direction in order to decrease the fitness from iteration to iteration. But an algorithm that is trapped in a local minimum will not be able to escape from this state if only downhill steps are allowed. Uphill steps are needed for this purpose. Since uphill steps are not always reasonable, they should be allowed selectively. For SA, an uphill step is allowed based on the Boltzmann probability distribution, which depends on the current temperature. The lower the temperature, the smaller is the probability that an uphill step is taken. At the beginning, the temperature should be high such that the algorithm is able to explore the overall search space. Over iterations, the temperature is slowly decreased in order to limit the number of uphill steps. Hence, in a later stage, the SA algorithm acts like a pure downhill method in order to improve the current solution until a certain stopping criterion is met. In this way, the global optimum (corresponding to the minimum energy state) is supposed to be found.

In contrast to SA, GAs and PSO are both population based optimisation methods, i.e. several "entities" that interact with each other are searching for the global optimum at the same time. The interaction suggests a somehow faster and/or better exploration of the global search space. GAs are far older than PSO and date back until the mid-1950s [Wei09, Chap. 3]. They are inspired by natural evolution, i.e. "the survival of the fittest"[9]. A single "entity" of the GA population is called *chromosome* or *genome*. New potential solutions are obtained by a reproduction of the fittest solution candidates involving mutation and crossover effects. Chromosomes that lead to bad fitness values are rejected (and "die"). Hence, all chromosomes interact in a competitive way. In contrast, the interaction involved in PSO is of a social kind. PSO was first proposed by Kennedy and Eberhart in 1995 and is inspired by the social behaviour of fish schools or bird flocks [KE95].[10] In the context of PSO, a single "entity" of the population is denoted as *particle* and the population itself is called *swarm*. The particles explore the search space by moving in a randomised manner through it. The movement of a single particle is not only influenced by the interaction with other particles, but also by the history of the particle itself. Hence, memory is introduced for PSO. At each position, every particle evaluates

[8] Other combinations are possible as well: In [CZ08, Chap. 14], SA is for example combined with a naive random search.

[9] Actually, the origin for the name "fitness function" lies in the area of GAs.

[10] Essentially, Kennedy and Eberhart wanted to model and simulate human social behaviour and found out that their method is suitable for global optimisation as well: "The method was discovered through simulation of a simplified social model; ..." [KE95, p. 1].

its fitness and compares it to the fitness of the best position it has visited so far, which it always remembers. This position is called *individual best (IB)* position. If the current position leads to a better fitness than the remembered one, the IB position is updated to the current position. Furthermore, each particle can communicate with a certain number of other particles. Through this interaction, the particles exchange information about their best positions so far: Each particle chooses the best solution of its *neighbourhood*, i.e. the set of particles with which it can communicate. This position is called *local best (LB)* position. When a particle has to choose a new position, it considers the IB *and* the LB position by which it is attracted. In this way, the particles are able to find a good solution together, which (hopefully) corresponds to the global optimum.

All three methods presented above are incomplete and suffer from premature convergence. According to the "no free lunch" theorems for optimisation by Wolpert and Macready [WM97], all optimisation algorithms perform equally well over the whole range of possible optimisation problems. An improved performance in a certain area is paid by a worse performance in another area. As a consequence, a certain optimisation algorithm can only outperform others over a limited class of problems. This also means that the performance of each algorithm is problem-dependent. If possible, the optimisation method of choice should be "tuned" to the problem at hand. In case of the above, nature-inspired algorithms, "tuning" can be understood as the adaption of some algorithm-specific design parameters such as the "cooling schedule" for SA, the "mutation rate" for GAs or the "neighbourhood size" for PSO.

Even though it is not possible to a priori choose the best optimisation algorithm for a given problem, PSO is chosen as global optimisation method in the remainder of this thesis due to several reasons: The search space for channel parameter estimation is continuous and of high dimension (P=3C). The fitness function $\Omega(\tilde{\boldsymbol{\theta}})$ has many local optima due to the superposition of several multipath components. Therefore, population based approaches seem to be better suited to channel parameter estimation due to their improved ability to explore the search space. Furthermore, PSO is rather simple to implement and depends on less design parameters compared to GAs. Hence, it is easier to "tune" the PSO method to the problem at hand if necessary. As a consequence, PSO appears to be a good choice for channel parameter estimation. Therefore, it is examined in more detail subsequently.

Since its invention in 1995, PSO has been studied extensively and many different variants of PSO aiming at an improved performance have been suggested. A concise summary on these advances is for example given in [BK07] or [BL08, Sec. 3]. In [BK07], Bratton and Kennedy suggest a standard for PSO incorporating the most important changes to the original PSO. The PSO implementation considered in this thesis is mainly based on [BK07]. As already mentioned before, the particles move in a randomised manner through the search space being attracted by their IB and LB position. This means that every particle i has a specific position $\tilde{\boldsymbol{\theta}}_i^n$ and a certain velocity $\boldsymbol{v}_i^n$ at the nth iteration. Both, the position and the velocity, are initialised randomly inside the search space. A new position $\tilde{\boldsymbol{\theta}}_i^{n+1}$ is found by first updating the velocity and then adding the new velocity

to the old position:

$$\boldsymbol{v}_i^{n+1} = \chi \left\{ \boldsymbol{v}_i^n + \lambda_1 \boldsymbol{\epsilon}_1^n \odot \left(\tilde{\boldsymbol{\theta}}_{\mathrm{IB},i}^n - \tilde{\boldsymbol{\theta}}_i^n \right) + \lambda_2 \boldsymbol{\epsilon}_2^n \odot \left(\tilde{\boldsymbol{\theta}}_{\mathrm{LB},i}^n - \tilde{\boldsymbol{\theta}}_i^n \right) \right\}, \tag{4.87}$$

$$\tilde{\boldsymbol{\theta}}_i^{n+1} = \tilde{\boldsymbol{\theta}}_i^n + \boldsymbol{v}_i^{n+1}, \tag{4.88}$$

where $\tilde{\boldsymbol{\theta}}_{\mathrm{IB},i}^n$ and $\tilde{\boldsymbol{\theta}}_{\mathrm{LB},i}^n$ denote the IB and LB position of the ith particle, respectively. χ is a constant called *constriction factor* and λ_1 and λ_2 are constants called *learning factors*. The vectors $\boldsymbol{\epsilon}_1^n$ and $\boldsymbol{\epsilon}_2^n$ are independent random vectors with elements uniformly distributed between $[0, 1]$. Furthermore, $\odot$ denotes the element-wise vector multiplication. Based on the analysis in [CK02], it is derived that the constriction factor should be chosen according to

$$\chi = \frac{2}{|2 - \phi - \sqrt{\phi^2 - 4\phi}|} \quad \text{with} \quad \phi = \lambda_1 + \lambda_2 \overset{!}{>} 4 \tag{4.89}$$

in order to assure convergence to a stationary point. For standard PSO, the following values are chosen: ϕ=4.1 with λ_1=λ_2=2.05, which leads to $\chi \approx 0.72984$ [BK07]. After performing the updates in (4.87) and (4.88) for all particles, the fitness of each particle $\Omega(\tilde{\boldsymbol{\theta}}_i^{n+1})$ is evaluated and new IB and LB positions are found. The IB position is updated if the current position provides a better fitness:

$$\tilde{\boldsymbol{\theta}}_{\mathrm{IB},i}^{n+1} = \begin{cases} \tilde{\boldsymbol{\theta}}_i^{n+1} & \text{if } \Omega(\tilde{\boldsymbol{\theta}}_i^{n+1}) < \Omega(\tilde{\boldsymbol{\theta}}_{\mathrm{IB},i}^n), \\ \tilde{\boldsymbol{\theta}}_{\mathrm{IB},i}^n & \text{else.} \end{cases} \tag{4.90}$$

The LB position is determined from the IB positions of the particles in a certain neighbourhood $\mathcal{S}$. In case of PSO, the term neighbourhood does not denote particles that are near to each other in the sense of Euclidean distance. The word neighbourhood is rather defined by the numbering of the particles. In [BK07], a *ring structure* is recommended: The ith particle communicates with the $(i-1)$th and the $(i+1)$th particle:

$$\tilde{\boldsymbol{\theta}}_{\mathrm{LB},i}^{n+1} = \tilde{\boldsymbol{\theta}}_{\mathrm{IB},b}^{n+1} \quad \text{with} \quad b = \arg\min_{\tilde{b} \in \mathcal{S}} \left\{ \Omega(\tilde{\boldsymbol{\theta}}_{\mathrm{IB},\tilde{b}}^{n+1}) \right\}, \mathcal{S} = \{i-1,\, i,\, i+1\}. \tag{4.91}$$

Other neighbourhood topologies are discussed in [KM06]. Originally, a fully informed swarm with connections between all particle was used. In case of a fully informed swarm, the LB position is denoted as *global best (GB)* position. Generally, swarms with a large neighbourhood size converge faster, but have a higher probability to converge prematurely to a local optimum. If the information exchange is limited by a smaller neighbourhood size, the probability of premature convergence can be reduced at the cost of a slower convergence rate. Since premature convergence should be avoided by any means (if possible), the ring topology with limited information exchange is used for standard PSO [BK07].

The above equations (4.87) to (4.91) describe the basic PSO method as summarised in Algorithm 4.4. A last aspect that needs to be considered is the behaviour of the algorithm when a particle leaves the bound constraint search space. Different bound handling strategies are investigated in [HW07]. It is shown that particles are initialised *with overwhelming probability* close to a search space border for high-dimensional optimisation

Algorithm 4.4 Particle swarm optimisation.

1. **Initialisation:**

   Initialise the positions $\tilde{\boldsymbol{\theta}}_i^0$ and velocities $\boldsymbol{v}_i^0$ of all particles i randomly inside the search space.

   Set the IB position to $\tilde{\boldsymbol{\theta}}_{\mathrm{IB},i}^0 = \tilde{\boldsymbol{\theta}}_i^0$ and determine the LB position according to (4.91).

2. **Repeat until a stopping criterion is met:**

   - Update velocity and position of all particles:
     $$\boldsymbol{v}_i^{n+1} = \chi\left\{\boldsymbol{v}_i^n + \lambda_1 \boldsymbol{\epsilon}_1^n \odot \left(\tilde{\boldsymbol{\theta}}_{\mathrm{IB},i}^n - \tilde{\boldsymbol{\theta}}_i^n\right) + \lambda_2 \boldsymbol{\epsilon}_2^n \odot \left(\tilde{\boldsymbol{\theta}}_{\mathrm{LB},i}^n - \tilde{\boldsymbol{\theta}}_i^n\right)\right\},$$
     $$\tilde{\boldsymbol{\theta}}_i^{n+1} = \tilde{\boldsymbol{\theta}}_i^n + \boldsymbol{v}_i^{n+1},$$
   - Evaluate fitness $\Omega(\tilde{\boldsymbol{\theta}}_i^{n+1})$ of all particles and determine new IB position $\tilde{\boldsymbol{\theta}}_{\mathrm{IB},i}^{n+1}$ and LB position $\tilde{\boldsymbol{\theta}}_{\mathrm{LB},i}^{n+1}$ according to (4.90) and (4.91), respectively.

problems. This means that the importance of bound handling increases with dimensionality. Bratton and Kennedy propose to "let the particles fly", i.e. the particles are allowed to leave the search space, but their fitness is not evaluated and further considered as long as they are outside the search space [BK07]. This approach can be generalised to the so-called infinity method in [HW07]: If a particle leaves the search space, its fitness is not evaluated, but set to "bad" value. In case of minimisation, "bad" means a very high value ($\rightarrow \infty$). With this bound handling, particles outside the search space are not further considered automatically due to their high fitness values. Sooner or later, the particles that left the search space will return into it due to the attraction of their IB and LB positions. In addition to bound handling, the velocities $\boldsymbol{v}_i^{n+1}$ should be limited by a certain threshold $\boldsymbol{v}_{\max}$ in order to reduce the probability that a particle leaves the search space. This means that in each (parameter) dimension p the absolute value of the updated velocity $v_{p,i}^{n+1}$ is clipped to the corresponding value defined by $v_{p,\max}$.

Bratton and Kennedy do not suggest a certain stopping criterion for standard PSO. For uniformity and comparability, the stopping criterion already suggested for the local methods is applied here for PSO as well. This means that the fitness of the global best position $\tilde{\boldsymbol{\theta}}_{\mathrm{GB}}^n$ is additionally determined in each iteration, which is updated if a better fitness has been found by any of the particles. By this means, the overall fitness improvement of PSO can be monitored. If the fitness improvement is below a certain threshold (e.g. $|\Omega(\tilde{\boldsymbol{\theta}}_{\mathrm{GB}}^{n+1}) - \Omega(\tilde{\boldsymbol{\theta}}_{\mathrm{GB}}^n)| < 10^{-5}$) for several times in a sequence, the algorithm is stopped. Further stopping criteria are for example discussed in [ZL07]: Zielinski and Laur explain that the above described method might not be the best one because "improvement often occurs irregularly in evolutionary algorithms" [ZL07, p. 54]. The choice of the above stopping criterion is based on the fact that, *on average*, the fitness improvement at an early stage of the iterative process is larger than at a later stage. In case of local methods, that

apply derivative information, this is also true for a single realisation of the algorithm. For global probabilistic methods, the fitness improvement of a single realisation is not necessarily related to the iterations due to the incorporated randomness. That means that small fitness improvements may already occur at an early stage of the iterative process. Thus, it may happen that the algorithm is stopped too early. However, the probability for small fitness improvements increases over iterations. As a consequence, the fitness improvement should be smaller than a predefined threshold for *many* times (e.g 30 or 50 times) in order to improve the reliability of the stopping criterion by preventing the algorithm from stopping too early [ZL07].[11] Finally, a suitable swarm size I needs to be selected. Bratton and Kennedy suggest to use a swarm size of I=50 particles, independent of the problem dimensionality [BK07]. In this thesis, the swarm size is chosen in relation to the dimension of the parameter estimation problem: If not stated otherwise, ten particles per dimension are used. This means that a swarm consists of I=30·C particles for a channel with C multipath components (i.e. 30 particles for the LOS channel and 60 particles for the two-path channels). The reason for this choice will be explained below.

In addition to the standard PSO, a further PSO variant called *cooperative PSO (CPSO)* is considered in this thesis [vdB01], [vdBE04]. The main idea behind CPSO is to divide the overall optimisation problem into smaller subproblems similar to the EM and SAGE algorithm. The motivation why van den Bergh and Engelbrecht suggest this separation lies in the fact that standard PSO exhibits a weakness that they summarise as "taking two steps forward, and one step back" [vdBE04, p. 228]. This means that an already good solution in a single dimension may be discarded in favour of an improvement in other dimensions because the fitness is only evaluated after updating all dimensions. Hence, valuable information may be lost that way. Consequently, the fitness should be evaluated more frequently in order to exploit all available information. For that purpose, van den Bergh and Engelbrecht suggest to divide the overall swarm into subswarms that search for the best solution in subdimensions of the overall optimisation problem only. If P denotes the dimension of the overall optimisation problem and S is the number of subswarms, then P_S denotes the number of dimensions per subswarm. One iteration cycle comprises the separate updates of all S subswarms according to the standard PSO rules: First, the velocity and the position of each particle i for a certain subswarm s are updated:

$$\boldsymbol{v}_{s,i}^{n+1} = \chi \left\{ \boldsymbol{v}_{s,i}^{n} + \lambda_1 \boldsymbol{\epsilon}_{s,1}^{n} \odot \left(\tilde{\boldsymbol{\theta}}_{\mathrm{IB},s,i}^{n} - \tilde{\boldsymbol{\theta}}_{s,i}^{n} \right) + \lambda_2 \boldsymbol{\epsilon}_{s,2}^{n} \odot \left(\tilde{\boldsymbol{\theta}}_{\mathrm{LB},s,i}^{n} - \tilde{\boldsymbol{\theta}}_{s,i}^{n} \right) \right\}, \tag{4.92}$$

$$\tilde{\boldsymbol{\theta}}_{s,i}^{n+1} = \tilde{\boldsymbol{\theta}}_{s,i}^{n} + \boldsymbol{v}_{s,i}^{n+1}, \tag{4.93}$$

where all vectors are of dimension P_S now. The constriction factor χ and the learning factors λ_1 and λ_2 have the same values as for standard PSO. After this update, the fitness of each particle is evaluated and new IB and LB positions are determined similar to (4.90) and (4.91). Since the evaluation of the fitness function requires a full parameter vector of dimension P, a so-called *context vector* $\boldsymbol{\Upsilon}$ is introduced by which the subswarms are

[11] Alternatively, PSO could be stopped if the velocity of all particles is below a certain threshold or if the distance of all particles to the global best is smaller than a given threshold.

Algorithm 4.5 Cooperative particle swarm optimisation.

1. **Initialisation:**

   Initialise the positions $\tilde{\boldsymbol{\theta}}^0_{s,i}$ and velocities $\boldsymbol{v}^0_{s,i}$ of all particles i in each subswarm s randomly inside the corresponding subspace.

   Set the IB positions to $\tilde{\boldsymbol{\theta}}^0_{\mathrm{IB},s,i}{=}\tilde{\boldsymbol{\theta}}^0_{s,i}$ and initialise the context vector $\boldsymbol{\Upsilon}$ randomly inside the search space.

   Determine the LB positions $\tilde{\boldsymbol{\theta}}^0_{\mathrm{LB},s,i}$ of all particles in each swarm with the help of the context vector and update the corresponding part of the context vector to the global best of this subswarm ($\boldsymbol{\Upsilon}_s{=}\tilde{\boldsymbol{\theta}}^0_{\mathrm{GB},s}$).

2. **Repeat until a stopping criterion is met:**

   For all subswarms $s{=}1,\cdots,S$ do:

   - Update velocity and position of all particles:
     $$\boldsymbol{v}^{n+1}_{s,i} = \chi\left\{\boldsymbol{v}^n_{s,i} + \lambda_1\boldsymbol{\epsilon}^n_{s,1}\odot\left(\tilde{\boldsymbol{\theta}}^n_{\mathrm{IB},s,i} - \tilde{\boldsymbol{\theta}}^n_{s,i}\right) + \lambda_2\boldsymbol{\epsilon}^n_{s,2}\odot\left(\tilde{\boldsymbol{\theta}}^n_{\mathrm{LB},s,i} - \tilde{\boldsymbol{\theta}}^n_{s,i}\right)\right\},$$
     $$\tilde{\boldsymbol{\theta}}^{n+1}_{s,i} = \tilde{\boldsymbol{\theta}}^n_{s,i} + \boldsymbol{v}^{n+1}_{s,i},$$
   - Evaluate fitness of all particles with the help of the context vector $\boldsymbol{\Upsilon}$ and determine new IB position $\tilde{\boldsymbol{\theta}}^{n+1}_{\mathrm{IB},s,i}$ and LB position $\tilde{\boldsymbol{\theta}}^{n+1}_{\mathrm{LB},s,i}$.
   - Update corresponding part of the context vector: $\boldsymbol{\Upsilon}_s{=}\tilde{\boldsymbol{\theta}}^{n+1}_{\mathrm{GB},s}$.

enabled to cooperate [vdBE04]. The context vector is composed of the current global best positions of all subswarms and is updated after each subswarm update. Hence, the context vector after updating $(s{-}1)$ subswarms in iteration cycle n is given by

$$\boldsymbol{\Upsilon} = \left[\left(\tilde{\boldsymbol{\theta}}^{n+1}_{\mathrm{GB},1}\right)^T,\cdots,\left(\tilde{\boldsymbol{\theta}}^{n+1}_{\mathrm{GB},s-1}\right)^T,\left(\tilde{\boldsymbol{\theta}}^{n}_{\mathrm{GB},s}\right)^T,\left(\tilde{\boldsymbol{\theta}}^{n}_{\mathrm{GB},s+1}\right)^T,\cdots,\left(\tilde{\boldsymbol{\theta}}^{n}_{\mathrm{GB},S}\right)^T\right]^T \tag{4.94}$$

similar to the auxiliary variable in (4.85) for the SAGE method. The fitness of each particle in the sth subswarm can now be evaluated by temporarily replacing the corresponding part of the context vector with the updated position of this particle: $\boldsymbol{\Upsilon}_s{=}\tilde{\boldsymbol{\theta}}^{n+1}_{s,i}$. A summary of CPSO is given in Algorithm 4.5. Apart form the subswarm-wise updates, the basic PSO algorithm is unchanged, e.g. the bound handling strategy and the stopping criterion remain the same. Hence, PSO is a special case of CPSO with only one subswarm. If more than one subswarm is used, the CPSO method is expected to converge faster than the PSO method due to more frequent fitness evaluations. According to this property, CPSO appears to be a better choice for global optimisation than PSO. Unfortunately, so-called *pseudo-minima* may be created due to the division of the overall search space into subspaces. This means that the subswarms may be trapped in a state in which they are not able to further improve their fitness although they have not even reached a local minimum. This event is denoted as *stagnation* [vdBE04]. The occurrence of stagnation

for CPSO is problem-dependent. Therefore, it can not be predicted whether the PSO or CPSO method is the better choice for global optimisation. Hence, the performance of both algorithms is investigated in Section 4.4.3. Typically, many different possibilities exist to divide the overall swarm into several subswarms. In order to distinguish between different CPSO versions, the number of subswarms is appended to the acronym: CPSO-S. For a two-path channel with C=2 and P=6, there are three possible versions of the CPSO method besides the full swarm version denoted by PSO: CPSO-2, CPSO-3 and CPSO-6. In the last case, the fitness is evaluated most frequently. In order to enable a fair comparison between different CPSO versions, the number of particles per dimension is fixed. As already mentioned above, ten particles per dimension are assumed in this thesis. This means that each subswarm consists of ten particles in case of CPSO-6, while 30 particles per subswarm are used for CPSO-2.

4.4.3 Performance Comparison of the Proposed Optimisation Algorithms

In this section, the performance of the ML channel parameter estimator is investigated in conjunction with the local and global optimisation algorithms proposed above, namely the LM, EM and SAGE method as well as the PSO and CPSO method. One-shot estimations without tracking or averaging are considered. Similar to Section 4.3.3, the performance is examined for the LOS channel, several two-path channels and three different WINNER channels. The LOS channel is a trivial case for optimisation because there is a unique optimum. Hence, local methods are sufficient in this case. However, the performance of local as well as global methods is investigated for the LOS channel. For the two-path and the WINNER channel models, an inferior performance of the local methods is expected because the fitness function exhibits many local minima due to multipath propagation.

For all simulations, the same setup as already used in Section 4.3.3 is applied. However, the basic setup is summarised at this point again: A burst length of K'=100 symbol durations is assumed and a pseudo-random sequence of BPSK symbols is used as training. According to the results in Section 4.3.3, the oversampling factor is fixed to J=2 (K=2K'=200). The Gaussian pulse shape of (4.44) is applied with an effective pulse width of T_g=8T_s and a shift of s=4T_s. The sampling phase ε is generated uniformly distributed in the interval $[-T_s/2, +T_s/2]$. The remaining parameters of $\boldsymbol{\theta}$ are generated according to the current channel model. For all optimisation methods, the MSE of the estimated sampling phase $\hat{\varepsilon}$,

$$\mathrm{MSE}\left(\hat{\varepsilon}\right) = \mathrm{E}\left\{\left|\hat{\varepsilon} - \varepsilon\right|^2\right\}, \tag{4.95}$$

is compared to the CRLB derived in Section 4.3. Furthermore, the MSEs of the channel estimates for LSCE,

$$\mathrm{MSE}\left(\check{\boldsymbol{h}}\right) = \mathrm{E}\left\{\frac{1}{L+1}\sum_{l=0}^{L}\left|\check{h}_l - h_l\right|^2\right\}, \tag{4.96}$$

and the channel estimates for channel parameter estimation,

$$\mathrm{MSE}\left(\hat{\boldsymbol{h}}\right) = \mathrm{E}\left\{\frac{1}{L+1}\sum_{l=0}^{L}\left|\hat{h}_l - h_l\right|^2\right\}, \tag{4.97}$$

are compared. Due to the exploitation of a priori information about the pulse shaping and the receive filter, the MSE for channel parameter estimation in (4.97) is supposed to be better than the MSE of LSCE in (4.96). In all figures presented below, the quantities concerning timing or delay measures are normalised with respect to the symbol duration.

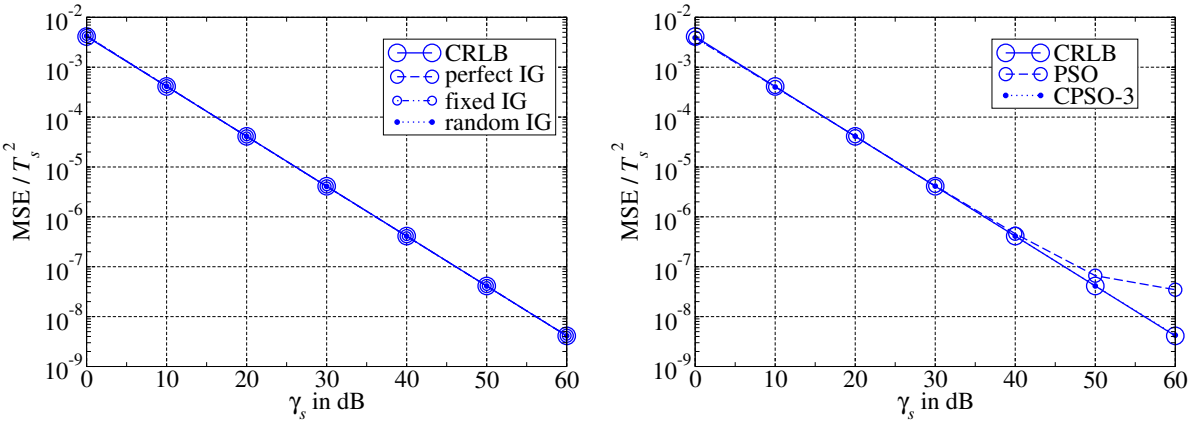
LOS Channel Model

In case of the LOS channel, the LM, EM and SAGE method are the same because there is only a single propagation path. Hence, the only local method that is considered for the LOS channel is the LM algorithm. Concerning the global methods, a single CPSO version besides the PSO exists, namely the CPSO-3 with three one-dimensional subswarms. Consequently, the LM, the PSO and the CPSO-3 method are investigated for the LOS channel. For all algorithms, the same basic stopping criterion as described above is applied, but with different values: Generally, the algorithms are stopped if the fitness improvement is below a threshold T for several times $N_{\min}$ in a sequence. For the LM method, the threshold is set to T=10^{-3} and the minimum number, for which the fitness improvement must be smaller than this threshold, is set to $N_{\min}$=2. For the global methods, the fitness improvement threshold is set to T=10^{-5} and the minimum number is set to $N_{\min}$=30. The LM method requires an initial guess (IG) as input. Three different types of IGs are considered here: a perfect IG (pIG), a fixed IG (fIG) and a random IG (rIG). The values of the IGs are determined as follows:

$$\tilde{\boldsymbol{\theta}}_{0,\mathrm{pIG}} = \boldsymbol{\theta}, \qquad \tilde{\boldsymbol{\theta}}_{0,\mathrm{fIG}} = [1,0,0]^T, \qquad \tilde{\boldsymbol{\theta}}_{0,\mathrm{rIG}} = [\cos(\tilde{\Phi}_0), \sin(\tilde{\Phi}_0), -\tilde{\varepsilon}_0]^T$$

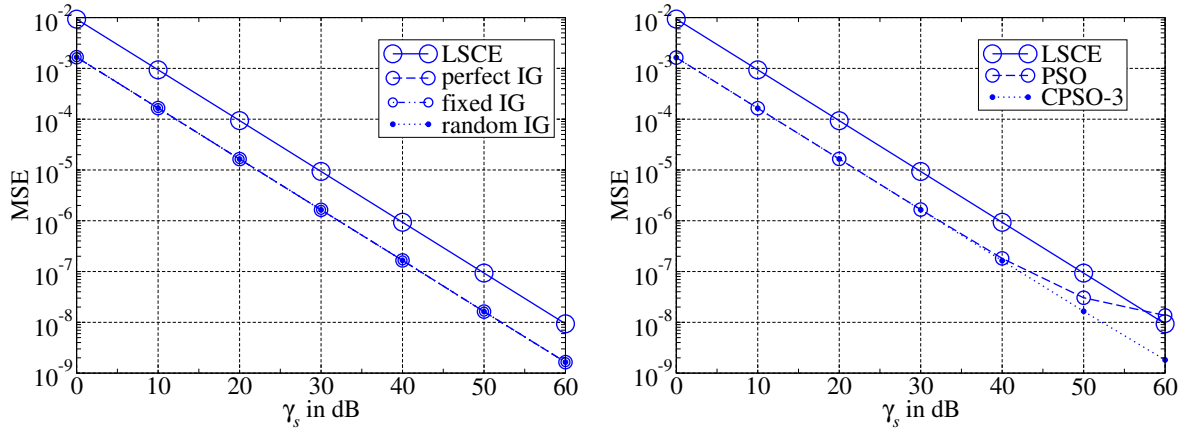
with $\tilde{\Phi}_0 \sim \mathcal{U}([0, 2\pi[)$, $\tilde{\varepsilon}_0 \sim \mathcal{U}([-0.5T_s, 0.5T_s])$ and where $\mathcal{U}([a,b])$ denotes a uniform distribution in the interval $[a,b]$. Since there is a unique optimum in case of the LOS channel, it is expected that the LM method is able to find this optimum independent of the IG. For the global methods, bounds of the search space need to be determined: The real and imaginary part of the amplitude vary in the interval [-1,1], while the sampling phase is bounded in the interval [-0.5T_s,0.5T_s]. The MSE of $\hat{\varepsilon}$ normalised to T_s^2 and the MSE of the channel estimates are shown for all three algorithms in Figure 4.11 and Figure 4.12, respectively. On the left hand side, the results for the LM method are displayed, while the results for the global methods are shown on the right hand side. As expected, the MSE performance of the LM method is independent of the IG in case of the LOS channel: For all three types of IGs, the MSE of $\hat{\varepsilon}$ normalised to T_s^2 lies exactly on the CRLB over the whole SNR range, i.e. the estimator corresponds to the minimum variance unbiased (MVU) estimator. The same is true for the CPSO-3 method, while the PSO method shows an error floor at high SNR. A similar behaviour occurs for the MSE of the channel estimates shown in Figure 4.12. As already supposed above, channel parameter estimation leads to improved channel estimates in comparison to LSCE. For the LM and the

CPSO-3 method, a gain of approximately 7.8 dB is obtained over the whole SNR range, while the MSE of the PSO method saturates at high SNR. The error floor for the PSO method occurs because the algorithm is stopped too early. The dependence of the error floor on the values of T and $N_{\min}$ for the global methods is illustrated in Figure 4.13 and Figure 4.14. On the left hand side, the MSEs are shown for different thresholds T with $N_{\min}$=10. On the right hand side, the threshold is fixed to T=10^{-5} and the minimum number $N_{\min}$, for which the fitness improvement must be smaller than the threshold, is varied. The larger the threshold T and the smaller the number $N_{\min}$, the more likely it is that the algorithm is stopped too early and the higher is the error floor. The influence of the minimum number $N_{\min}$ is more significant than the influence of the threshold T. Even the CPSO-3 method shows an error floor if the stopping criterion is badly chosen, but it outperforms the PSO method in all cases. As a consequence and in accordance with [ZL07], the minimum number $N_{\min}$ should not be chosen too small for the global



(a) LM method. (b) PSO and CPSO-3 method.

Figure 4.11: MSE of $\hat{\varepsilon}$ normalised to T_s^2 for the LOS channel.

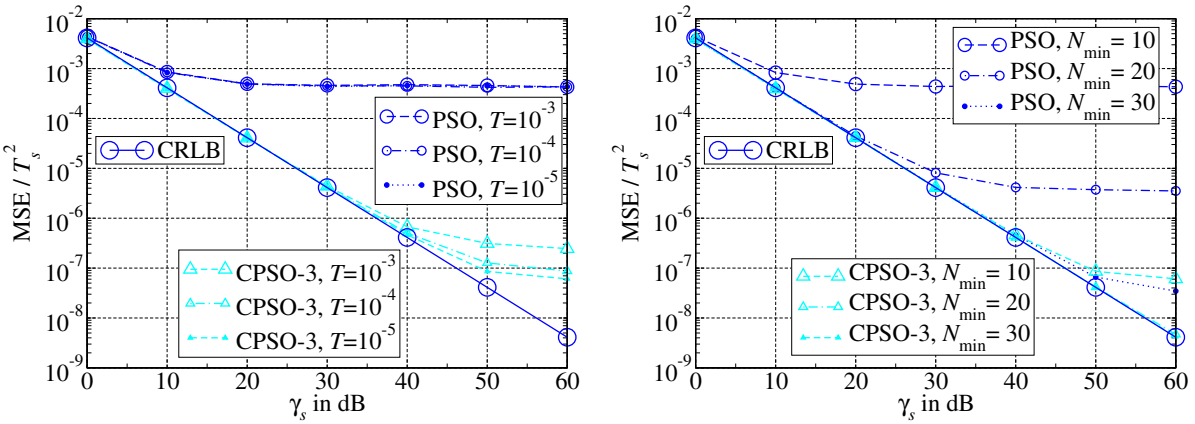


(a) LM method. (b) PSO and CPSO-3 method.

Figure 4.12: MSE of channel estimates for the LOS channel.

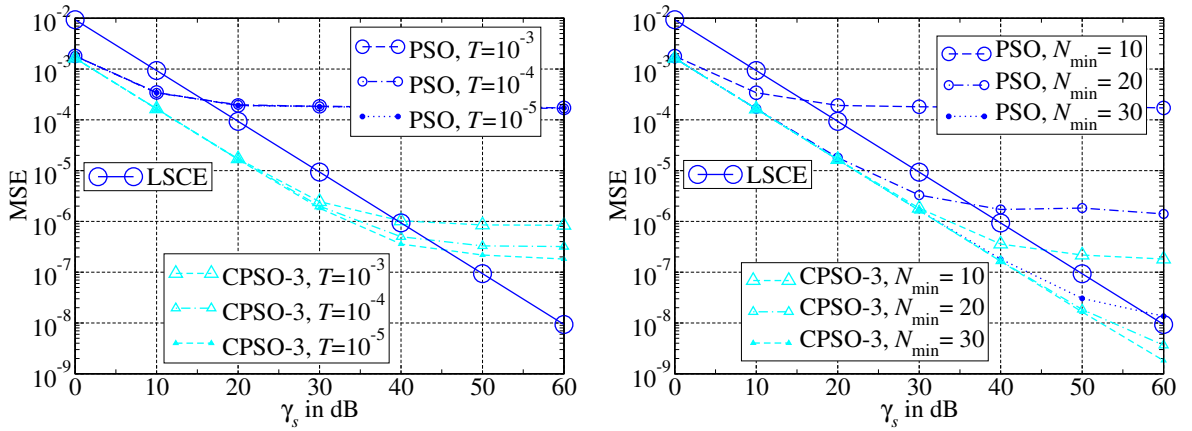
methods. The larger $N_{\min}$, the more iterations are required until the algorithms stop. Hence, $N_{\min}$ should not be chosen too large as well in order limit the average number of iterations. The initial choice of T=10^{-5} and $N_{\min}$=30 is already quite good because the error floor for PSO occurs late and with a small normalised MSE value of approximately $3{\cdot}10^{-8}$. For symbol durations smaller than a microsecond ($T_s{\leq}1$ μs), this value corresponds to a distance error in the low centimetre region and below. Furthermore, the SNR range of interest for many applications lies approximately between 0 and 30 dB, where all algorithms show the best possible performance. Hence, the error floor of the PSO for the stopping criterion with T=10^{-5} and $N_{\min}$=30 is acceptable.

This choice is further justified by Figure 4.15 and Figure 4.16. Again, the results for the LM method are shown on the left hand side, while the results for the global



(a) Variation of the threshold T with $N_{\min}$=10. (b) Variation of $N_{\min}$ with threshold T=10^{-5}.

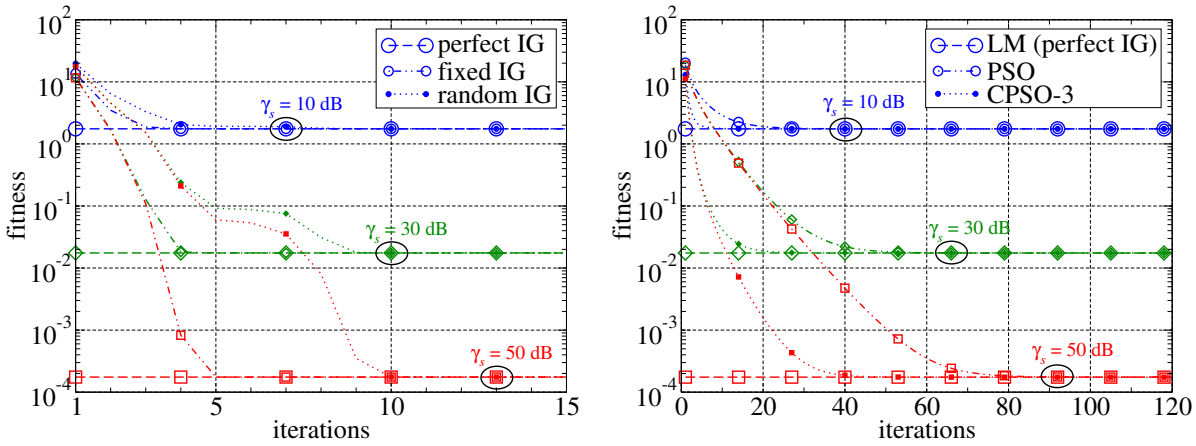
Figure 4.13: MSE of $\hat{\varepsilon}$ normalised to T_s^2 for the global methods with different stopping criteria in the LOS channel.



(a) Variation of the threshold T with $N_{\min}$=10. (b) Variation of $N_{\min}$ with threshold T=10^{-5}.

Figure 4.14: MSE of channel estimates for the global methods with different stopping criteria in the LOS channel.

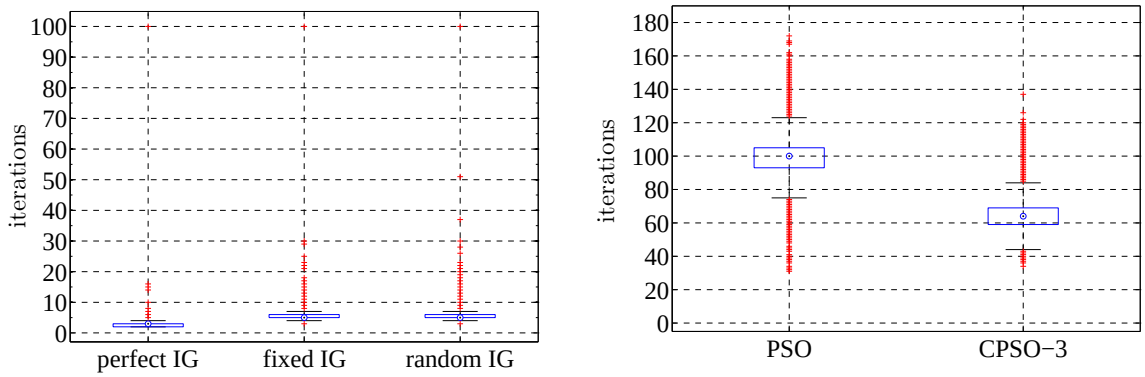
methods are illustrated on the right hand side. In Figure 4.15, the average fitness history is shown, i.e. the way the fitness evolves over iterations. In case of the fitness history, the algorithms are not stopped until a certain number of iterations have been performed. Hence, the *required* number of iterations can be determined with the help of the fitness history. This number can be compared to the *performed* number of iterations in order to judge whether the algorithms are stopped adequately. For the fitness histories in Figure 4.15, the iterations are limited to 100 and 1000 iterations for the LM method and the global methods, respectively. The LM method with perfect IG provides a lower bound on the achievable fitness and is, thus, also plotted in Figure 4.15(b) for comparison. Three different SNRs are considered: γ_s=10 dB (blue), γ_s=30 dB (green) and γ_s=50 dB (red). In general, more iterations are needed until convergence with increasing SNR because the best achievable fitness decreases with increasing SNR. Concerning the LM method, a fixed IG enables a faster convergence than a random IG: At 50 dB, the LM method



(a) LM method.

(b) PSO and CPSO-3 method.

Figure 4.15: Fitness history for the LOS channel.



(a) LM method.

(b) PSO and CPSO-3 method.

Figure 4.16: Boxplots of iterations for the LOS channel.

with fixed IG requires five iterations, while the LM method with random IG requires ten iterations. Regarding the global methods, the CPSO-3 converges much faster than the PSO method due to a more frequent fitness evaluation. The PSO needs approximately twice the number of iterations the CPSO-3 requires: At 50 dB, the CPSO-3 requires 40 iterations to approach the best possible fitness, while the PSO requires 80 iterations. In order to judge whether the algorithms are stopped adequately or too early, the number of required iterations needs to be compared to the number of performed iterations using the stopping criteria with T and $N_{\min}$. The number of performed iterations for the initial choice of the stopping criteria (LM: T=10^{-3}, $N_{\min}$=2; PSO/CPSO-3: T=10^{-5}, $N_{\min}$=30) are illustrated in Figure 4.16 with the help of boxplots. By using boxplots, a more detailed insight into the iterative behaviour of the algorithms is possible because the statistics of a data set can be summarised in a convenient and robust way [Ros04, p. 24-27], [Hen06, p. 35f.]. Any boxplot consists of a box with whiskers. The box is always defined by the *first quartile* Q_1 (25% percentile) and the *third quartile* Q_3 (75% percentile), where the length of the box, Q_3-Q_1, is called *interquartile range* [Ros04, p. 24-27], [Hen06, p. 35f.]. The *median* or *second quartile* Q_2 (50% percentile) divides the box into two parts. The whiskers can be defined differently: For example, they can denote the smallest and the largest sample of a data set. Often, the whiskers are defined as a multiple of the interquartile length or, more precisely, by the most extreme sample that is inside this length. Hence, the lower whisker is given by the smallest sample that is larger than $Q_1-m\cdot(Q_3-Q_1)$ and the upper whisker by the largest sample that is smaller than $Q_3+m\cdot(Q_3-Q_1)$ with $m>0$. For the boxplots shown in this thesis, m=1.5 is used. Samples outside the whiskers are denoted as outliers. In Figure 4.16, the median is marked by a circle with a dot inside and the outliers are marked by a red plus. The underlying data sets comprise the performed iterations for all SNRs from 0 dB to 60 dB in 10 dB steps. Naturally, the LM method performs much less iterations than the global methods. Of course, least iterations are performed for the LM method with perfect IG. The median lies at 3 iterations for the perfect IG, while a median of 6 iterations is obtained for the fixed and the random IG. These medians are in the range of the required iterations. Together with the perfect MSE performance of the LM method, it can be concluded that the stopping criterion for the LM method is adequate. Interestingly, there is an extreme outlier at the maximum allowed number of iterations for the LM method, that is *independent* of the IG. This means that the LM algorithm is not always able to stop based on the stopping criterion with T=10^{-3} and $N_{\min}$=2 before the maximum allowed number of iterations is reached. Since the MSE performance of the LM method is optimal over the whole SNR range in the LOS channel, the fitness presumably oscillates around the optimum in that case such that the changes in fitness alternate between positive and negative values. This means that the fitness first improves a little bit and then degrades a little bit. Hence, a fitness improvement for two times in a sequence does not occur. However, this situation only appears rarely: At 10 dB, in 0.05% of all cases, at 30 dB, in 0.63% of all cases and, at 50 dB, in 2.43% of all cases for a random IG. For a fixed IG, approximately the same values are obtained, while the perfect IG shows this behaviour only at a maximum of 0.12% of all cases at 20 dB. This behaviour of the LM method would not have been discovered if only the mean number of iterations instead of boxplots were considered.

For the global methods, the algorithms are generally stopped before the maximum number of 1000 iterations is reached as shown in Figure 4.16(b): The number of performed iterations ranges from 30 to 170 iterations and from 30 to 140 in case of the PSO and CPSO-3 method, respectively. The median for the PSO method lies at 100 iterations, while a median of 63 iterations is obtained for the CPSO-3 method. These medians are larger than the required iterations according to the fitness histories shown in Figure 4.15(b). This implies that the global methods are stopped adequately. Since the medians are not much larger than the required iterations until fitness convergence, the stopping criterion with T=10^{-5} and $N_{\min}$=30 is a reasonable choice.

The above results can be summarised as follows: In the LOS channel, local methods are sufficient and global probabilistic methods are not required in general. However, if global methods are considered, it is advantageous to apply a cooperative approach since the CPSO-3 clearly outperforms the PSO due to a more frequent fitness evaluation. A different behaviour is expected for the channel models with multipath propagation since the LOS channel is a trivial case for channel parameter estimation. In case of multipath propagation, the topology of the fitness function becomes much more complicated and exhibits local optima, in which the algorithms might get stuck in. Corresponding results are presented in the following.

Two-Path Channel Models

In case of the two-path channel models, six different algorithms are considered: The LM, EM and SAGE method as well as the PSO, CPSO-2 and CPSO-6 method. The LM and the PSO algorithm work on the overall six-dimensional search space, while the EM, SAGE and CPSO-2 method divide the overall problem into two three-dimensional subproblems. The CPSO-6 method considers six one-dimensional subproblems. For the LM method and the global methods, the same stopping criteria as for the LOS channel are assumed (LM: T=10^{-3}, $N_{\min}$=2; PSO/CPSO-2/CPSO-6: T=10^{-5}, $N_{\min}$=30). For the EM and the SAGE algorithm, T=10^{-5} and $N_{\min}$=5 are applied. The maximum number of iterations is limited to 2000 iterations and 5000 iterations for the local and global methods, respectively. If not stated otherwise, all local methods are initialised with a random IG:

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_0^{\mathrm{LM}} &= [\tilde{A}_{1,0}\cdot\cos(\tilde{\Phi}_{1,0}),\ \tilde{A}_{1,0}\cdot\sin(\tilde{\Phi}_{1,0}),\ -\tilde{\varepsilon}_0,\ \tilde{A}_{2,0}\cdot\cos(\tilde{\Phi}_{2,0}),\ \tilde{A}_{2,0}\cdot\sin(\tilde{\Phi}_{2,0}),\ \tilde{\nu}_{2,0}-\tilde{\varepsilon}_0]^T \\ &= \hat{\boldsymbol{\theta}}_{\mathrm{EM}}^0 = \hat{\boldsymbol{\theta}}_{\mathrm{SAGE}}^0\end{aligned}$$

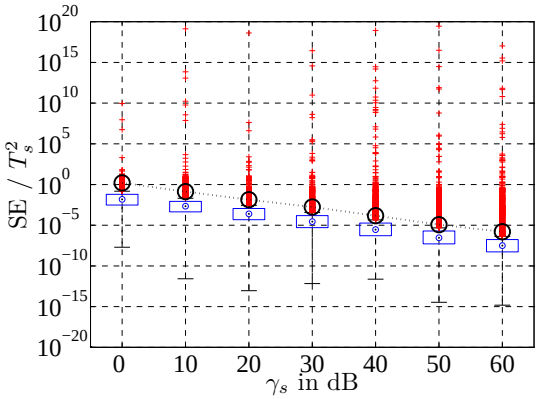
with

$$\tilde{A}_{1,0} = \sqrt{\frac{\tilde{\mathcal{P}}_0}{1+\tilde{\mathcal{P}}_0}}, \qquad \tilde{A}_{2,0} = \sqrt{\frac{1}{1+\tilde{\mathcal{P}}_0}}$$

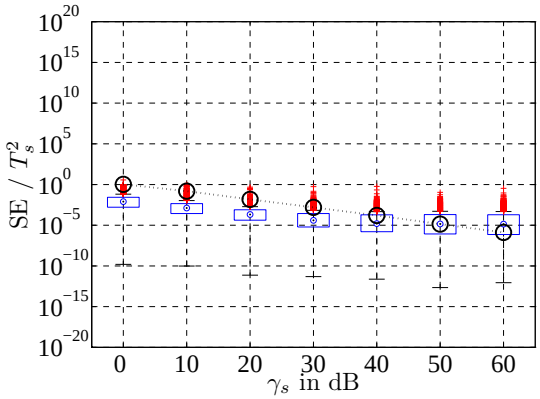
and where $\tilde{\mathcal{P}}_0 \sim \mathcal{U}([0.5, 2])$, $\tilde{\Phi}_{1,0} \sim \mathcal{U}([0, 2\pi[)$, $\tilde{\Phi}_{2,0} \sim \mathcal{U}([0, 2\pi[)$, $\tilde{\varepsilon}_0 \sim \mathcal{U}([-0.5T_s, 0.5T_s])$, $\tilde{\nu}_{2,0} \sim \mathcal{U}([0.1T_s, \nu_2^{\max}])$. For the global methods, the following bounds on the search space are utilised: The real and imaginary parts of the amplitudes vary in the interval [-1,1],

while the coarse excess delays are bounded in the interval [-0.5T_s, $\nu_2^{\max}$]. The maximum possible excess delay is fixed to $\nu_2^{\max}$=2T_s as in Section 4.3.3, which leads to a channel memory length of L'=11 (L=2L'=22). Furthermore, the same two-path channels as in Section 4.3.3 are examined. This means that the power ratio $\mathcal{P}$ and the phase offset $\Delta\Phi$ are generated randomly in the intervals [0.1, 10] and [0, 2π[, respectively, and either small excess delays ($\nu_2/T_s \in [0.05, 1]$) or large excess delays ($\nu_2/T_s \in [1, 2]$) are considered. As already expected above, the general performance of all algorithms differs from the the results obtained for the LOS channel. Due to multipath propagation, the fitness function $\Omega(\tilde{\boldsymbol{\theta}})$ exhibits local optima, in which the algorithms might get stuck in: The local methods converge to local optima in the neighbourhood of the initial guess, that may be far away from the global optimum, and the global methods may suffer from premature convergence or stagnation (in case of the cooperative approach). Very large estimation errors (outliers) may occur because the local methods are unconstrained. Since the MSE is not robust against outliers, the squared error (SE) performance of all algorithms is initially analysed with the help of boxplots. In Figure 4.17 and Figure 4.18, boxplots of the SE of $\hat{\varepsilon}$ normalised to T_s^2 are shown for the two-path channels with small and large excess delays, respectively. For comparison, the corresponding CRLBs are plotted, which are denoted by large black circles. On the left hand side of each figure, the results for the local methods are shown, while the performance of the global methods is illustrated on the right hand side. As expected above, outliers with very large values ($>10^5$) occur for the unconstrained local methods because local optima outside the search space and, hence, far away from the global optimum are sometimes found. With respect to outliers, the LM method is worst because it shows the largest range of outliers for both channel models. The EM and SAGE method are more robust against the IG than the LM method because they show much less extreme outliers. For the global methods, the values of outliers are generally bounded because the search space is bounded. From an outlier point of view, the PSO method outperforms all remaining methods because it shows the smallest range of outliers. This is especially obvious for the channel with large excess delay in Figure 4.18: For both CPSO methods, outliers up to a normalised SE of $8{\cdot}10^0$ are present over the whole SNR range, while the maximum values of the PSO outliers decrease with SNR. The reason for the worse performance of the CPSO methods is stagnation: By the division of the overall search space into smaller subspaces, pseudo-minima are created, in which the CPSO methods are sometimes stuck in.

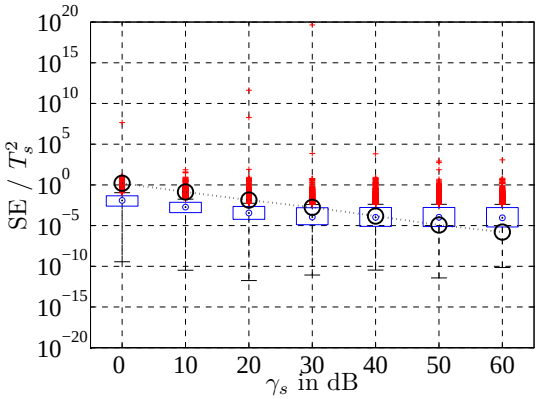
Other performance measures, besides the outlier range, are the position of the median and the size of the box, which contains the middle 50% of the data set and corresponds to the interquartile range. The interquartile range is a measure of statistical dispersion like the variance or standard deviation. But in comparison to the variance and the standard deviation, the interquartile range is more robust against outliers. The smaller the interquartile range, the smaller is the dispersion of the data. Hence, a small box is desirable. In case of the two-path channel with large excess delay in Figure 4.18, the boxes for all algorithms and at all SNRs have approximately the same size. Furthermore, the medians are generally located inside the large black circles denoting the CRLB. Only for the EM method, the median at 60 dB lies slightly above the CRLB, i.e. an error floor occurs at high SNR for the EM method in the channel with large excess delays.



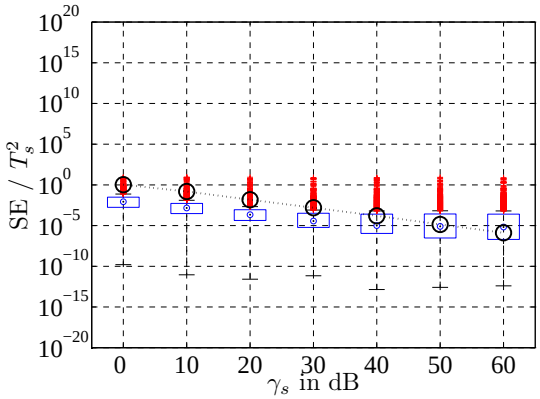
(a) LM method.



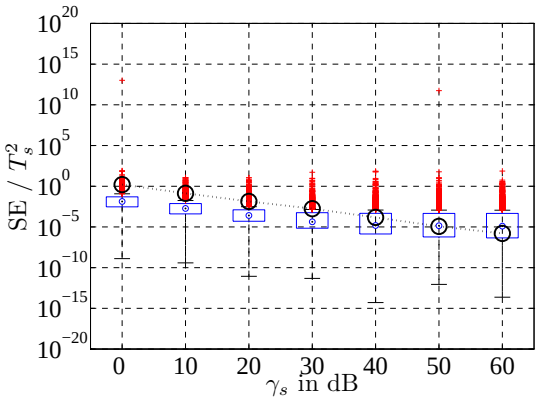
(b) PSO method.



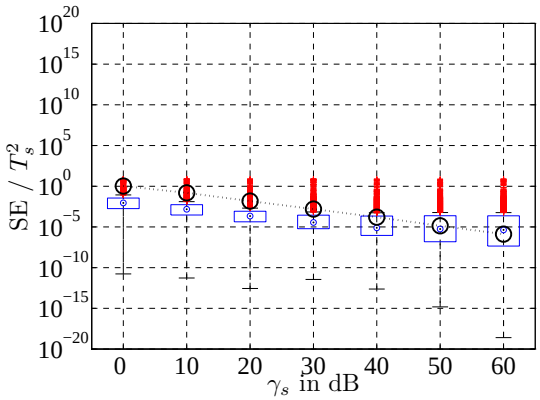
(c) EM method.



(d) CPSO-2 method.

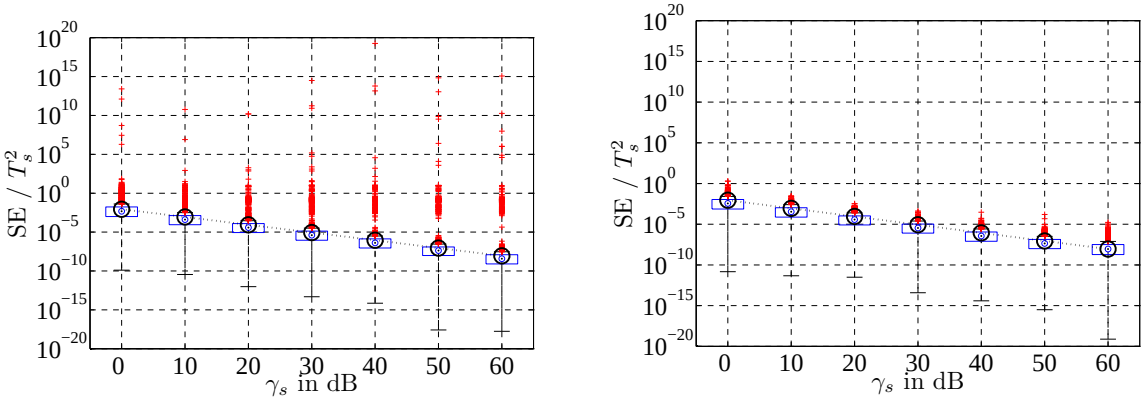


(e) SAGE method.



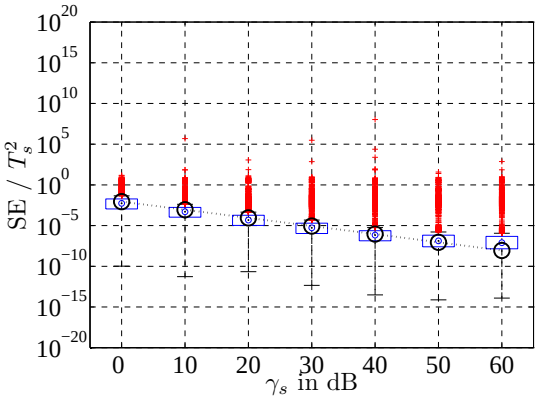
(f) CPSO-6 method.

Figure 4.17: Boxplots of the SE of $\hat{\varepsilon}$ normalised to T_s^2 for the two-path channel with $\nu_2/T_s \in [0.05, 1]$. For comparison, the corresponding CRLBs are shown, which are denoted by large black circles.

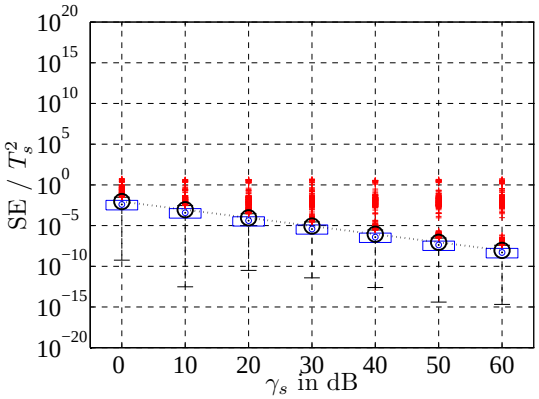


(a) LM method.

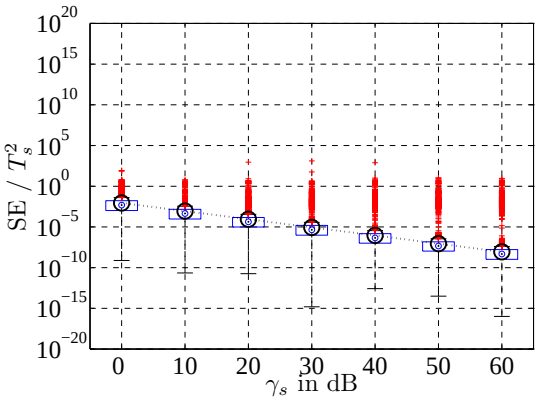
(b) PSO method.



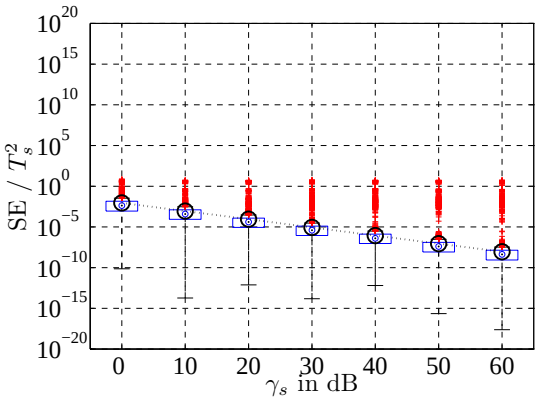
(c) EM method.



(d) CPSO-2 method.



(e) SAGE method.

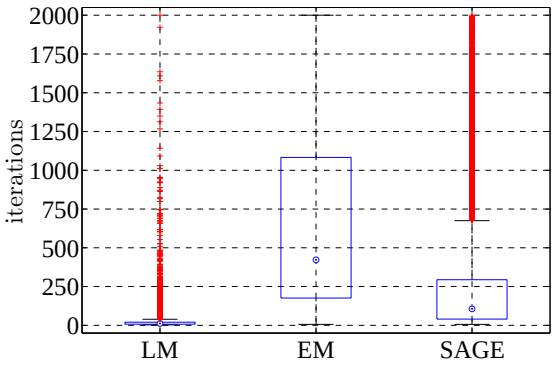


(f) CPSO-6 method.

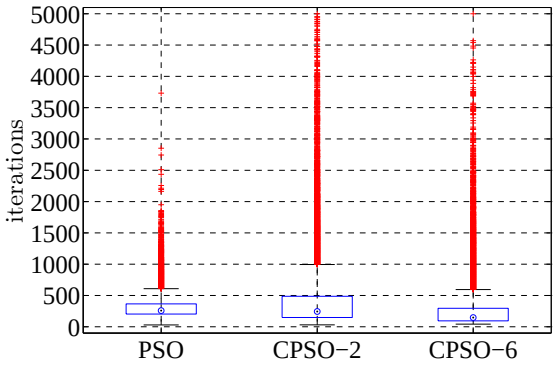
Figure 4.18: Boxplots of the SE of $\hat{\varepsilon}$ normalised to T_s^2 for the two-path channel with $\nu_2/T_s \in [1, 2]$. For comparison, the corresponding CRLBs are shown, which are denoted by large black circles.

Similarly, the medians of all algorithms, except the LM method, saturate at high SNR for the channel with small excess delay in Figure 4.17. Additionally, the interquartile range increases with increasing SNR, i.e the dispersion of the estimates increases. Only for the LM method, the size of the boxes remains approximately the same over the whole SNR range. From this point of view, the LM method outperforms the remaining algorithms since it exhibits the smallest interquartile ranges and does not run into an error floor. In addition to the SE performance, the performance with respect to the number of iterations is of interest.

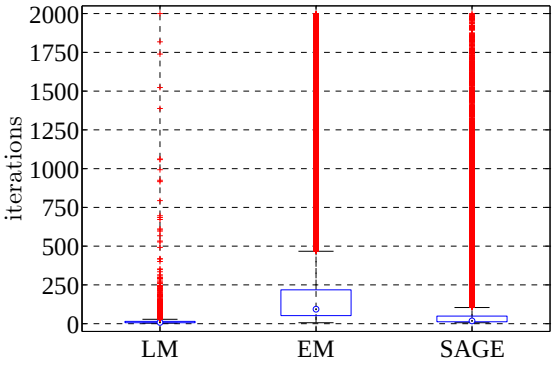
In Figure 4.19, boxplots of the performed iterations are shown for both two-path channels. Again, the underlying data sets contain all SNR points from 0 dB to 60 dB. In the upper part of Figure 4.19, the boxplots for the channel with small excess delay are shown and, in the lower part, the results for the channel with large excess delay are illustrated. With respect to the performed iterations, the LM method outperforms all other algorithms because it has the lowest median number of iterations and the smallest interquartile range. The remaining methods perform much more iterations, where the



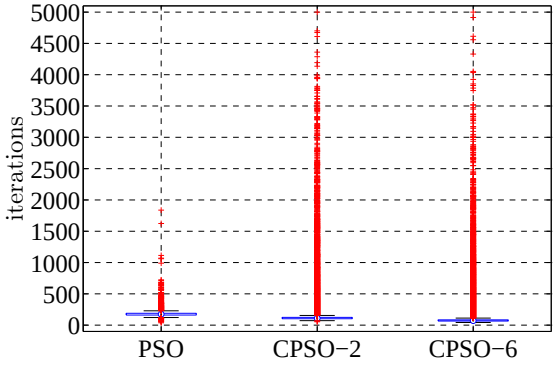
(a) Local methods with $\nu_2/T_s \in [0.05, 1]$.



(b) Global methods with $\nu_2/T_s \in [0.05, 1]$.



(c) Local methods with $\nu_2/T_s \in [1, 2]$.

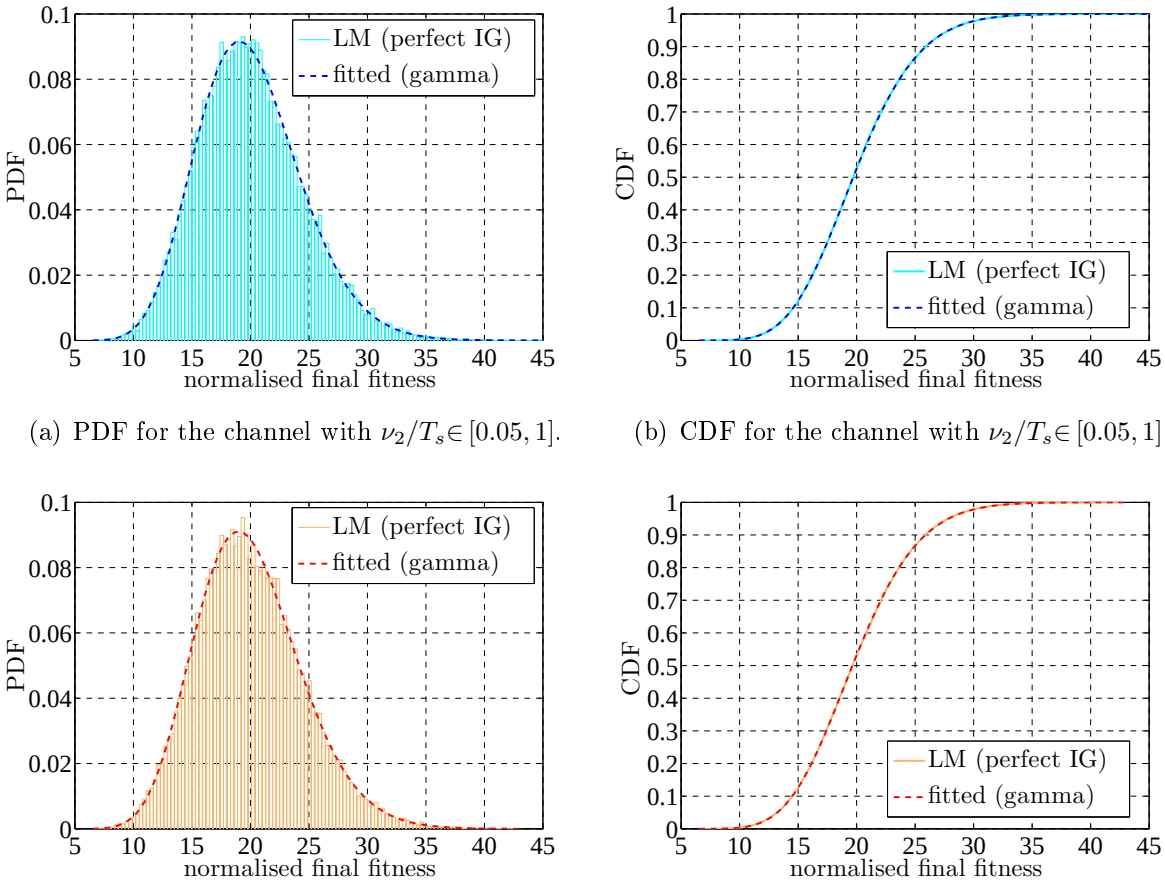


(d) Global methods with $\nu_2/T_s \in [1, 2]$.

Figure 4.19: Boxplots of iterations for the two-path channels.

EM algorithm is by far the worst method. It has the highest median number of iterations and a very large interquartile range with a span of almost 1000 iterations for the channel with small excess delay. Due to this (already predicted) very slow convergence rate, the EM algorithm is neglected in the remainder of this thesis. In general, more iterations are required for the two-path channel with small excess delay since it is much more challenging to separate the multipath components in this case.

Based on the above results, the PSO method is a good choice for multipath channels because it offers a good tradeoff between performance and time consumption. Without any modifications, the local methods and especially the LM algorithm are not suitable for multipath channels due to extreme outliers. However, if the number of outliers is small and if it is possible to detect such outliers, the local methods might still be applicable for multipath channels. It is generally desirable to detect outliers in order remove a potential error floor. Hence, important questions are: Is it possible to detect outliers? And, if yes, how can these outliers be detected? If the global optimum is not a priori known,



(a) PDF for the channel with $\nu_2/T_s \in [0.05, 1]$.

(b) CDF for the channel with $\nu_2/T_s \in [0.05, 1]$.

(c) PDF for the channel with $\nu_2/T_s \in [1, 2]$.

(d) CDF for the channel with $\nu_2/T_s \in [1, 2]$.

Figure 4.20: PDFs and CDFs of the final fitness in the two-path channels for the LM method with perfect initial guess.

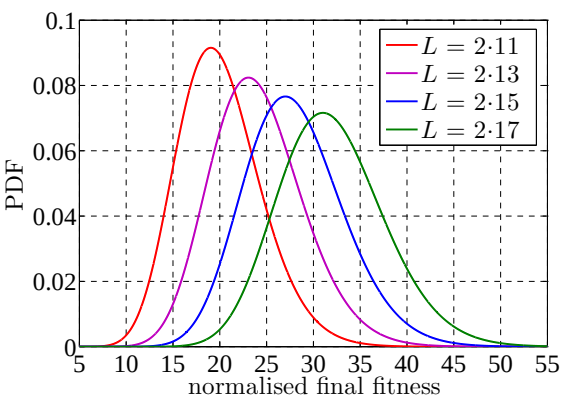
it is very challenging to decide whether it has been found or not. Therefore, it is quite difficult to detect outliers due to premature convergence. However, it *is* possible as will be explained below. The outlier detection method proposed in this thesis is based on the final fitness, i.e. the fitness of the parameter estimate $\Omega(\hat{\boldsymbol{\theta}})$. Since the final fitness is *always* available, this method is universally applicable. At first, the distribution of the final fitness is analysed for the LM method with perfect IG. In general, the final fitness decreases with increasing SNR. In order to obtain fitness values in the same range for all SNRs, the final fitness is normalised with respect to the SNR in linear scale: $\Omega(\hat{\boldsymbol{\theta}})\cdot\gamma_s$. In Figure 4.20, the normalised final fitness for the LM method with perfect IG is shown for both two-path channels. On the left hand side, histograms and fitted probability density functions (PDFs) are plotted and, on the right hand side, the corresponding cumulative distribution functions (CDFs) are shown. For both channel models, almost the same distribution of the final fitness is obtained, i.e. the optimal fitness distribution does not depend on the channel parameters. The histograms in Figure 4.20 are best fitted by a gamma distribution with shape parameter $a{\approx}20.2755$ and scale parameter $b{\approx}0.9883$. The PDF of a gamma distribution is given by

$$p_\Gamma(x|a,b) = \frac{1}{b^a\Gamma(a)}x^{a-1}\exp\left(\frac{-x}{b}\right) \tag{4.98}$$

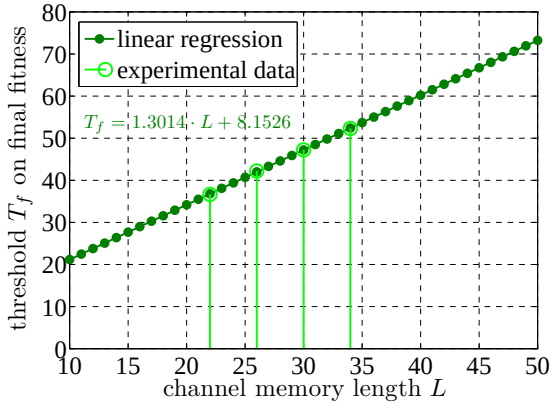
with $x{\geq}0$, $a{>}0$, $b{>}0$ and where

$$\Gamma(a) = \int_0^\infty t^{a-1}\,\mathrm{e}^{-t}\,\mathrm{d}\,t \tag{4.99}$$

is the gamma function. The gamma distributions in Figure 4.20 are slightly skewed to the right, i.e. the right tail is longer than the left tail. Hence, final fitness values larger than the mean $\mu_\Gamma{=}a{\cdot}b{\approx}20.0383$ are more probable than smaller fitness values. Based on the assumption that the final fitness follows the above defined gamma distribution *only* if



(a) Fitted PDFs of the final fitness.



(b) Fitness threshold T_f obtained by linear regression.

Figure 4.21: On the determination of the fitness threshold T_f.

the global optimum has been found, a fitness threshold T_f can be determined in order to detect outliers. If the normalised final fitness is below this fitness threshold, the parameter estimate is accepted. Otherwise, the parameter estimate is rejected and denoted as outlier. The fitness threshold is determined by integrating the fitted gamma distribution: The area under a PDF is by definition one. The fitness threshold T_f is defined as the fitness value for which the area under the PDF equals 0.999. This means that 99.9% of all normalised fitness values at the global optimum are below this threshold. For the two-path channels with memory length L=$2L'$=22 ($\nu_2^{\max}$=$2T_s$), a threshold of T_f=36.6376 is obtained. However, the distribution of the normalised final fitness changes with the channel memory length L as shown in Figure 4.21(a). Larger channel memory lengths are obtained by increasing the maximum possible excess delay $\nu_2^{\max}$. With increasing channel memory length L, the mean of the gamma distribution is shifted to the right and the PDF flattens. As a consequence, the fitness threshold T_f depends on the channel memory length L. Based on the fitness thresholds determined for L=2·11, L=2·13, L=2·15 and L=2·17, the following relationship between the channel memory length and the fitness threshold is obtained via a linear regression as illustrated in Figure 4.21(b):

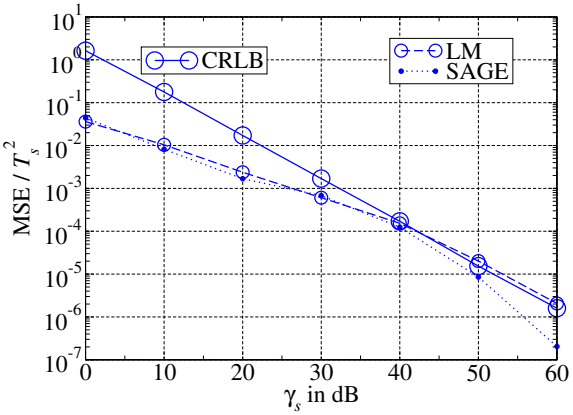
$$T_f = 1.3014 \cdot L + 8.1526. \tag{4.100}$$

Subsequently, outliers are detected based on (4.100): If

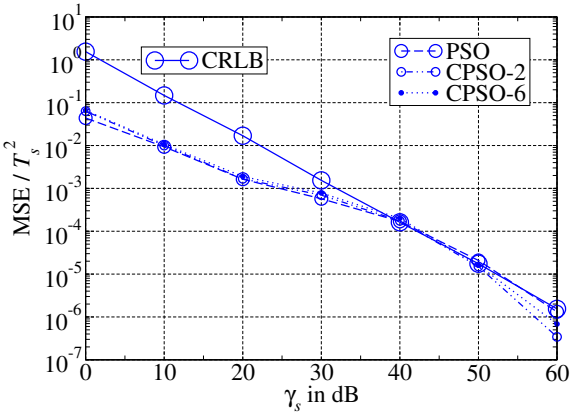
$$\Omega(\hat{\boldsymbol{\theta}}) \cdot \gamma_s > T_f, \tag{4.101}$$

the parameter estimate $\hat{\boldsymbol{\theta}}$ is rejected and not considered for the calculation of the MSE. The above detection criterion is sufficient for the global methods that work on a bounded search space. For the unconstrained local methods, an additional detection criterion is required: Even though the outlier detection based on (4.100) and (4.101) performs well for the local methods at high SNR, extreme outliers may still occur at low SNR if a local optimum with good fitness but outside the search space and, hence, far away from the global optimum is found (see Figure C.1 in Appendix C). Hence, for the local methods, it is additionally checked if the parameter estimate $\hat{\boldsymbol{\theta}}$ lies inside a bounded space. If this is the case, the parameter estimate is accepted. Otherwise, it is rejected. By means of this double check, outliers can be reliably detected over the whole SNR range for the local methods as well. Different bounds for the second detection criterion can be chosen: Of course, the search space bounds applied for the global methods are a reasonable choice. However, broader bounds are also possible. In conjunction with outlier detection, the *outage rate* in percent is a further performance measure in addition to MSEs and the number of iterations. The outage rate should be as small as possible, while outliers are reliably detected. Using the search space bound of the global methods (real and imaginary part of amplitudes in [-1,1], coarse excess delays in [-0.5T_s, $\nu_2^{\max}$]), a quite high outage rate is obtained at low SNR (see Figure C.2 in Appendix C). By broadening the bounds on the real and imaginary part of the amplitudes, the outage rate at low SNR can be decreased, while retaining a reliable outlier detection (see Figure C.1 and Figure C.2 in Appendix C). The following outlier bounds are applied subsequently: The real and imaginary parts of the amplitudes must be in the interval [-2,2], while the coarse excess delays are still bounded in the interval [-0.5T_s, $\nu_2^{\max}$].

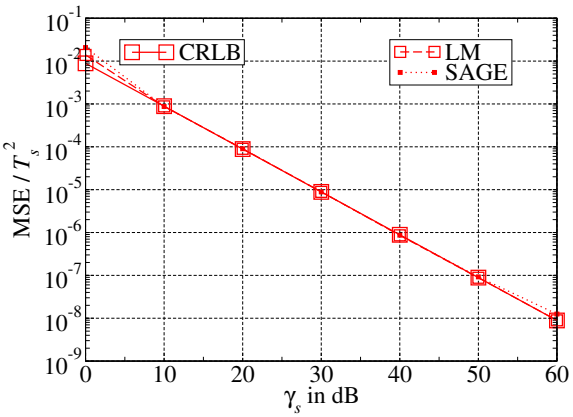
In Figure 4.22, the MSE of $\hat{\varepsilon}$ normalised to T_s^2 is shown for the local as well as the global methods in conjunction with the proposed outlier detection. Very good results are obtained for the channel with large excess delay shown in the lower part of Figure 4.22. The MSEs of all algorithms coincide with the CRLB almost over the whole SNR range. Only at very low or very high SNR, the MSEs are slightly larger than the CRLB. In case of the more challenging channel with small excess delays, a quite different behaviour is observed. The MSEs of all algorithms are mostly *below* the CRLB. For the LM and PSO method, the MSE of $\hat{\varepsilon}$ normalised to T_s^2 lies on the CRLB for SNRs larger than 40 dB. The MSEs of the remaining algorithms only coincide with the CRLB over a small SNR range between 40 dB and 50 dB. For higher SNRs, the MSEs are again below the CRLB. At first sight, this behaviour seems strange because the CRLB is a *lower* bound on the MSE of any unbiased estimator. On second thoughts, the above behaviour can be clearly explained. There are two different reasons for the deviations at low and high SNR, respectively. At first, the deviation at low SNR, that is present for all methods, is explained: In Section 4.2, an important property of ML estimators has been mentioned: ML estimators are *asymptotically* optimal, i.e. for a large number of observations or at high SNR they



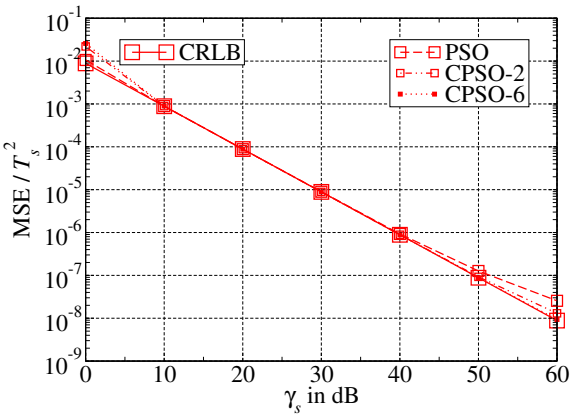
(a) Local methods with $\nu_2/T_s \in [0.05, 1]$.



(b) Global methods with $\nu_2/T_s \in [0.05, 1]$.



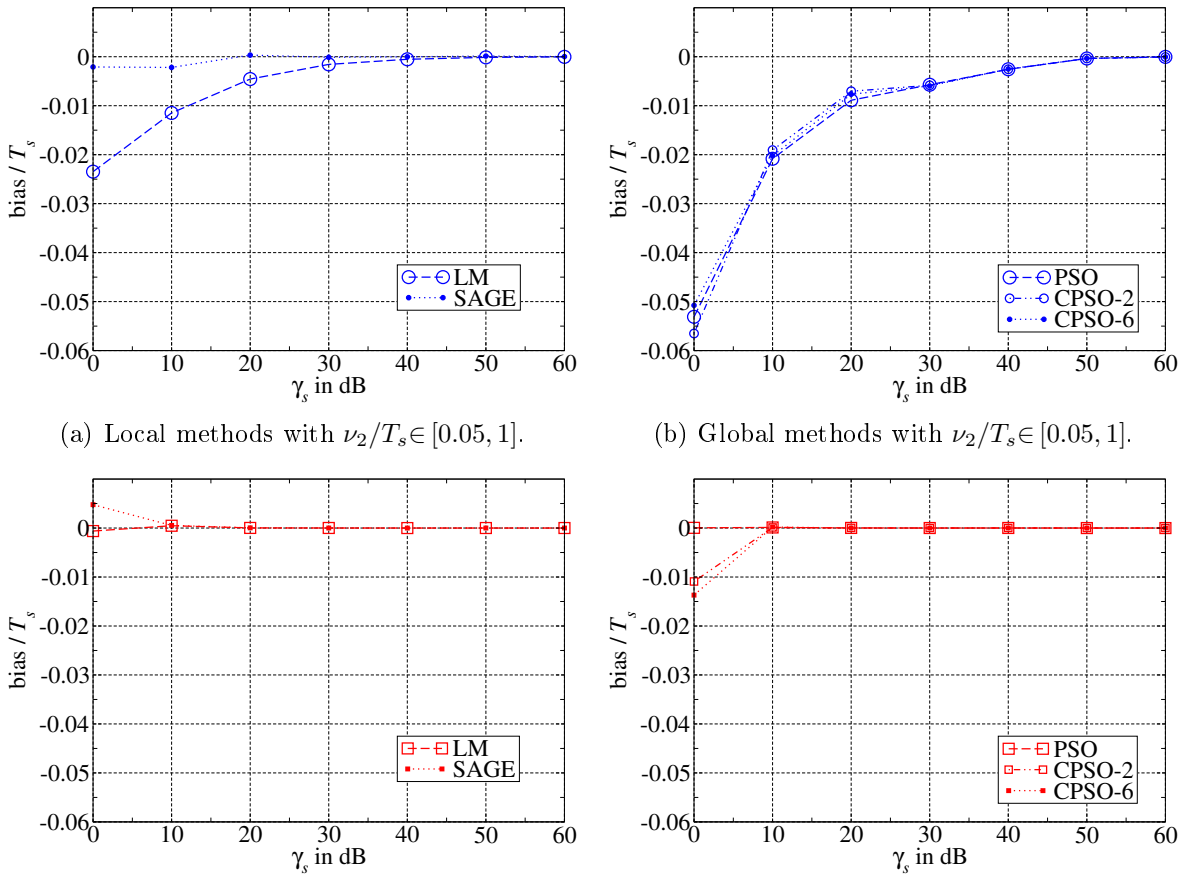
(c) Local methods with $\nu_2/T_s \in [1, 2]$.



(d) Global methods with $\nu_2/T_s \in [1, 2]$.

Figure 4.22: MSE of $\hat{\varepsilon}$ normalised to T_s^2 for the two-path channels.

are unbiased and efficient. Conversely, ML estimators may be biased at low SNR or for a small number of observations. In Figure 4.23, the bias of the estimated sampling phase ($\mathrm{E}\{\hat{\varepsilon}-\varepsilon\}$) normalised to the symbol duration T_s is shown for both channel models. In case of the two-path channel with large excess delay, the estimators are unbiased over almost the whole SNR range. Only at 0 dB, a small bias is observable for the SAGE algorithm and the two CPSO methods. In case of the two-path channel with small excess delay, the estimators are clearly biased up to 40 dB. This means that the CRLB is not valid in this region such that MSEs below the CRLB are possible. Hence, the behaviour of the estimators at low SNR is caused by a bias. However, the fact that the MSEs are smaller than the CRLB at high SNR can not be explained by a bias because all estimators are unbiased at high SNR. The deviations at high SNR are due to large outage rates as illustrated in Figure 4.24 and Figure 4.25. In Figure 4.24, the outage rate in percent is shown for both channel models. For the two-path channel with large excess delay, the outage rates are always below 14% and are nearly constant over an SNR range from 20 dB to 50 dB. For the local methods, higher outage rates are observed at low SNR, while the outage rates of the global methods are very small at low SNR. The PSO method exhibits the best outage rate except for very high SNR. In case of the two-path channel with



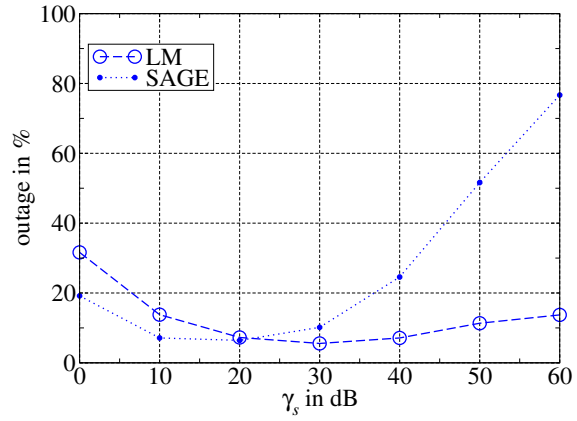
(a) Local methods with $\nu_2/T_s \in [0.05, 1]$.

(b) Global methods with $\nu_2/T_s \in [0.05, 1]$.

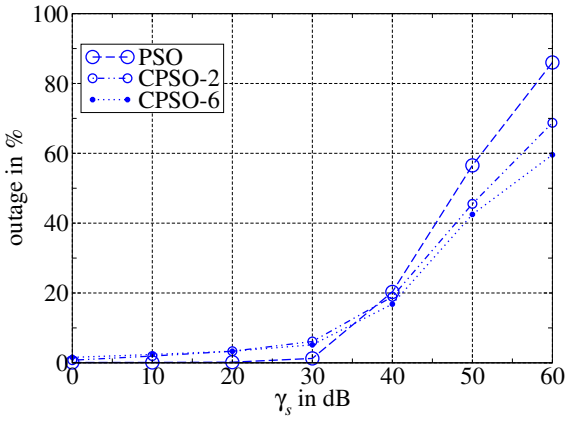
(c) Local methods with $\nu_2/T_s \in [1, 2]$.

(d) Global methods with $\nu_2/T_s \in [1, 2]$.

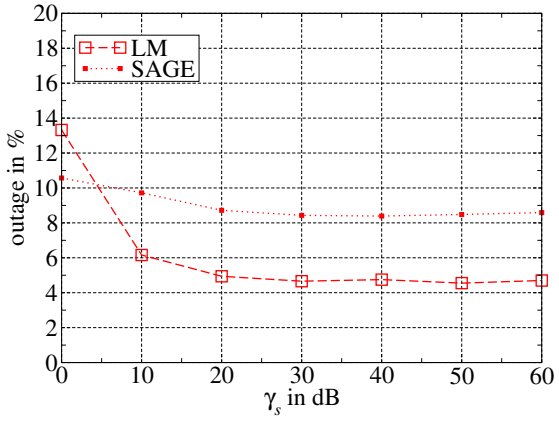
Figure 4.23: Bias of $\hat{\varepsilon}$ normalised to T_s for the two-path channels.



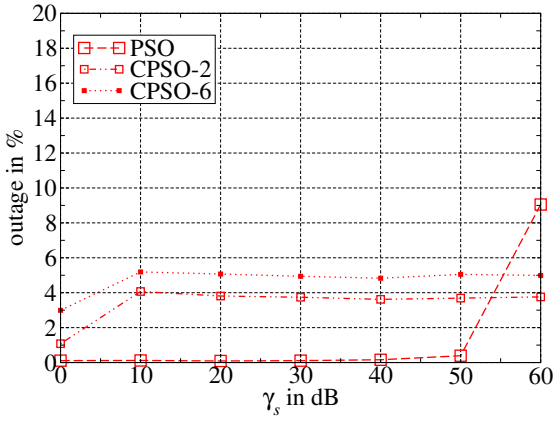
(a) Local methods with $\nu_2/T_s \in [0.05, 1]$.



(b) Global methods with $\nu_2/T_s \in [0.05, 1]$.

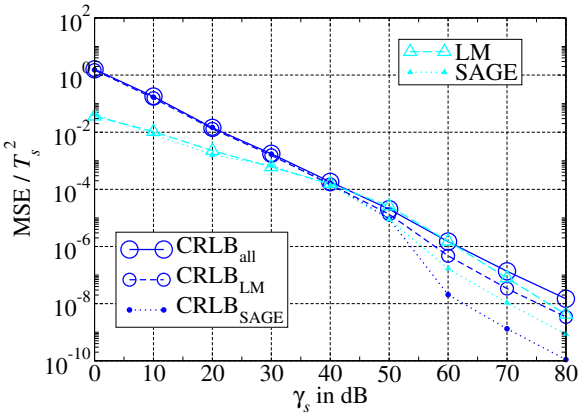


(c) Local methods with $\nu_2/T_s \in [1, 2]$.

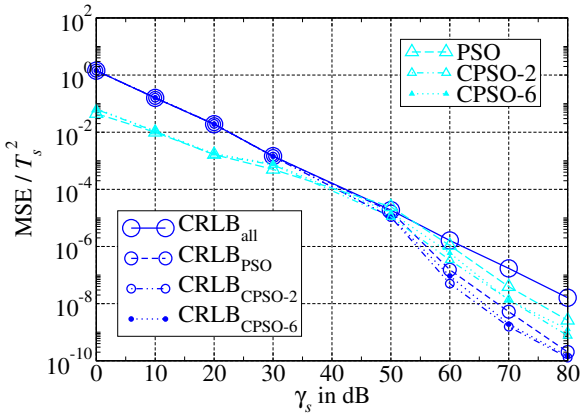


(d) Global methods with $\nu_2/T_s \in [1, 2]$.

Figure 4.24: Outage based on the fitness threshold T_f for the two-path channels.



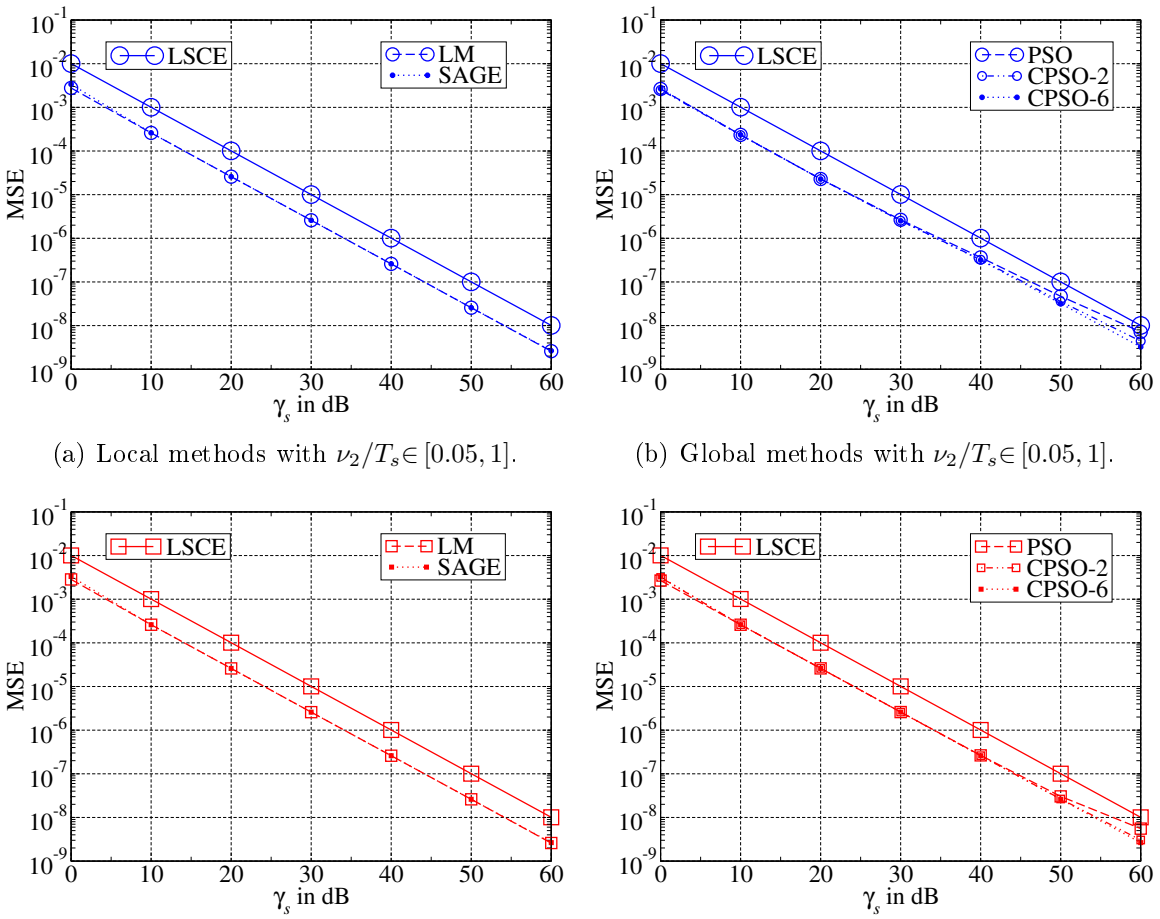
(a) Local methods.



(b) Global methods.

Figure 4.25: MSE of $\hat{\varepsilon}$ normalised to T_s^2 for the two-path channel with $\nu_2/T_s \in [0.05, 1]$ compared to individual CRLBs.

small excess delay, much larger outage rates occur. Generally, the outage rate increases from 30 dB upwards. At 60 dB, outage rates up to 80% are possible. This means that only in 20% of all cases a valid parameter estimate is found. Hence, a comparison of the corresponding MSE to the CRLB, that is averaged over *all* channel realisations, is not adequate anymore. The MSEs after outage detection should be rather compared to a CRLB that is averaged over the channel realisations for *successful* channel parameter estimation. This means that for each algorithm a different CRLB needs to be determined as shown in Figure 4.25 for the two-path channel with small excess delay: If the outage rate is below 50%, the difference between the overall CRLB and the individual CRLBs is negligible. For outage rates larger than 50%, the individual CRLBs are lower than the overall CRLB. Outage rates larger than 50% occur only at high SNR from 50 dB upwards. In order to get a comprehensive impression of this behaviour, the CRLBs and the MSEs of $\hat{\varepsilon}$ normalised to T_s^2 are plotted over an SNR range from 0 dB up to 80 dB in Figure 4.25. The MSEs are indeed smaller than the overall CRLB at high SNR, but they are *always* larger than or equal to the individual CRLBs. The only method that is able to reach the individual CRLB is the LM algorithm.



(a) Local methods with $\nu_2/T_s \in [0.05, 1]$.

(b) Global methods with $\nu_2/T_s \in [0.05, 1]$.

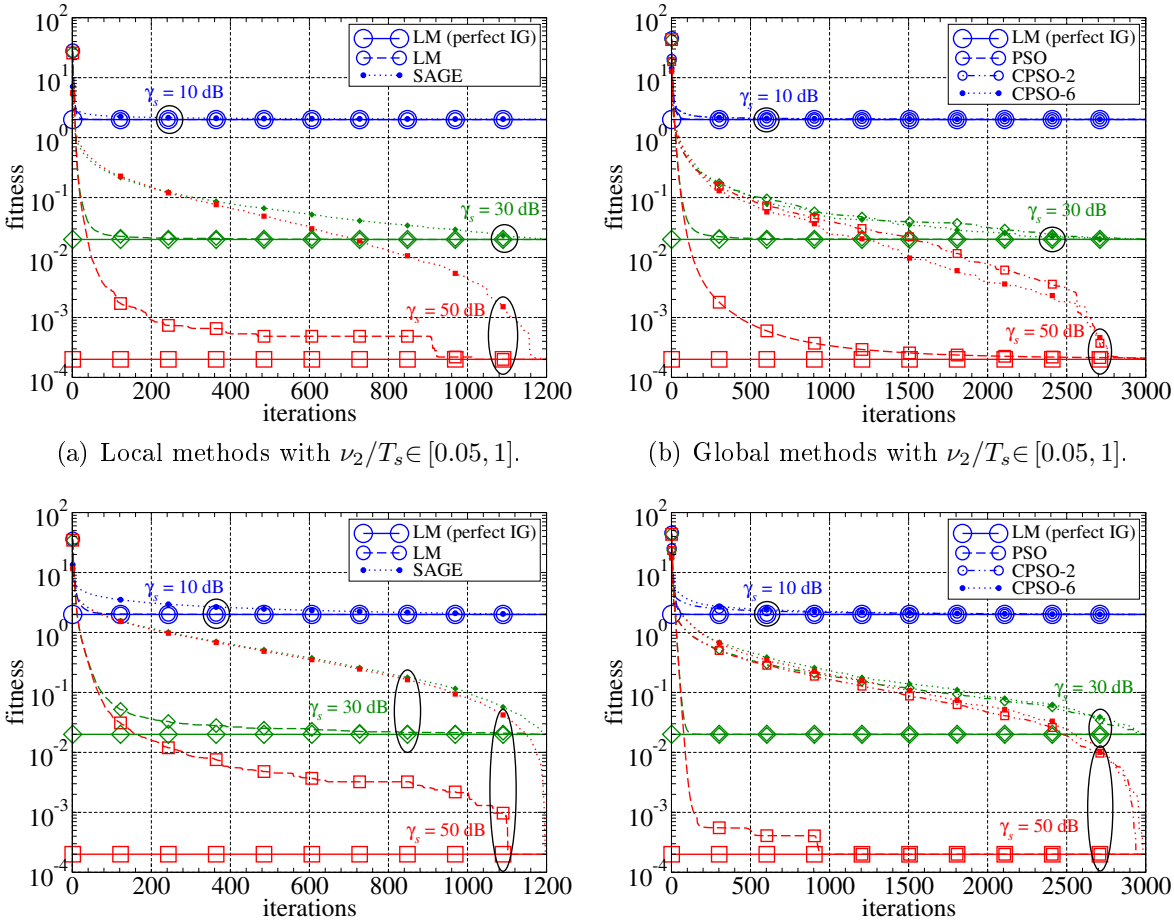
(c) Local methods with $\nu_2/T_s \in [1, 2]$.

(d) Global methods with $\nu_2/T_s \in [1, 2]$.

Figure 4.26: MSE of channel estimates for the two-path channels.

In summary, the deviation from the (overall) CRLB at low SNR is caused by a bias, while the deviation at high SNR is due to high outage rates. Similar to the error floor of the global methods in case of the LOS channel, the outage rates are very large at high SNR because the algorithms are stopped too early. Before discussing this aspect in more detail, the MSEs of the channel estimates for the two-path channels are presented in Figure 4.26. Generally, a gain of approximately 5.7 dB is obtained for channel parameter estimation compared to LSCE. The achievable gain is decreased by approximately 2.1 dB in comparison to the LOS channel because six instead of three parameters need to be estimated here. For almost all algorithms, the achievable gain is constant over the whole SNR range. Only for the PSO method, the MSE starts to saturate at very high SNR despite outlier detection. This saturation is again caused by a too early interruption of the algorithm and can be prevented by decreasing the threshold T and/or by increasing $N_{\min}$.

In order to judge how many iterations are required until convergence, the fitness histories for all methods with outlier detection is shown in Figure 4.27. As already mentioned



(a) Local methods with $\nu_2/T_s \in [0.05, 1]$.

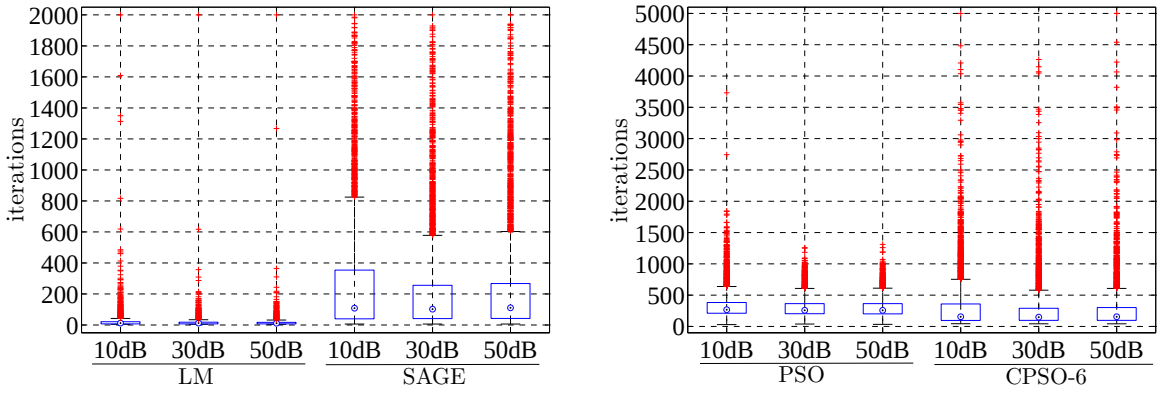
(b) Global methods with $\nu_2/T_s \in [0.05, 1]$.

(c) Local methods with $\nu_2/T_s \in [1, 2]$.

(d) Global methods with $\nu_2/T_s \in [1, 2]$.

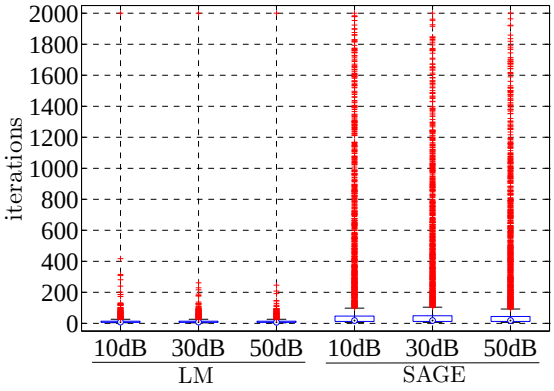
Figure 4.27: Fitness histories for the two-path channels.

above, the maximum number of iterations is limited to 2000 iterations and 5000 iterations for the local and global methods, respectively. In all subfigures, the fitness history for the LM method with perfect IG is plotted for comparison. Again, three different SNRs are considered: γ_s=10 dB (blue), γ_s=30 dB (green) and γ_s=50 dB (red). Similar to the LOS channel, the number of required iterations increases with SNR because the best achievable fitness decreases with SNR. Obviously, much more iterations are required by all algorithms compared to the LOS channel. With respect to the convergence rate, the LM method performs better than the SAGE algorithm and the PSO method outperforms the two CPSO methods, that show a similar behaviour. For an SNR of 50 dB, approximately 1000 to 3000 iterations are required until the fitness converges to the best possible value. For comparison, boxplots of the performed iterations are shown for the LM and SAGE method as well as the PSO and CPSO-6 method at 10 dB, 30 dB and 50 dB in Figure 4.28. It is observed that the required iterations in Figure 4.27 and the performed iterations in Figure 4.28 do not match. Considering a single algorithm, the median number of iterations in Figure 4.28 is approximately the same for all SNRs, while the interquartile range

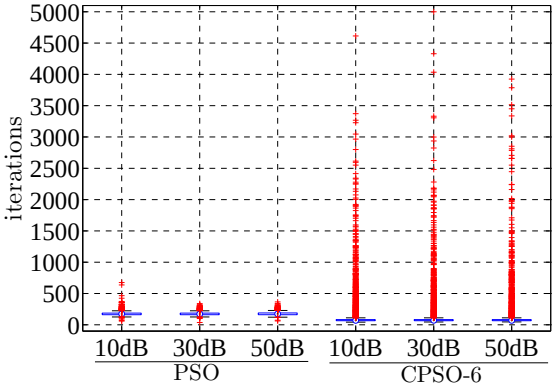


(a) Local methods with $\nu_2/T_s \in [0.05, 1]$.

(b) Global methods with $\nu_2/T_s \in [0.05, 1]$.



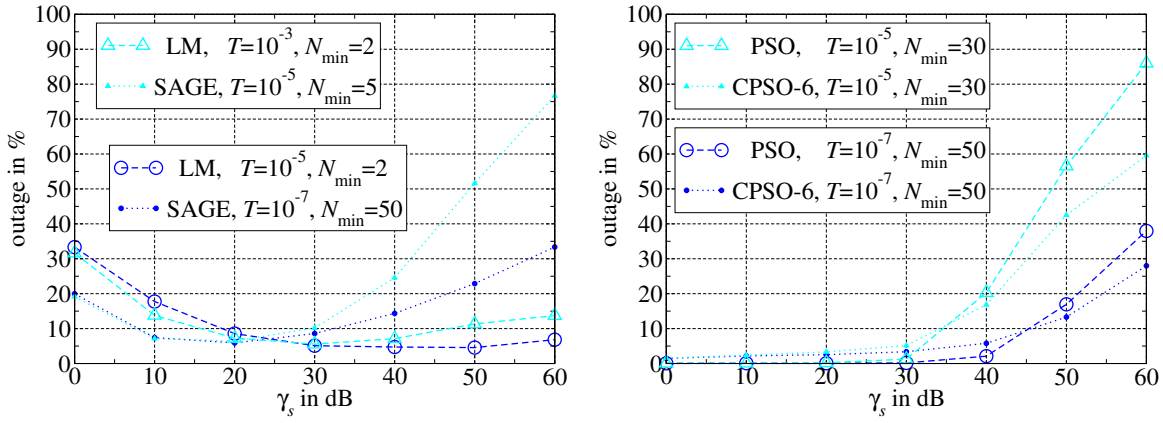
(c) Local methods with $\nu_2/T_s \in [1, 2]$.



(d) Global methods with $\nu_2/T_s \in [1, 2]$.

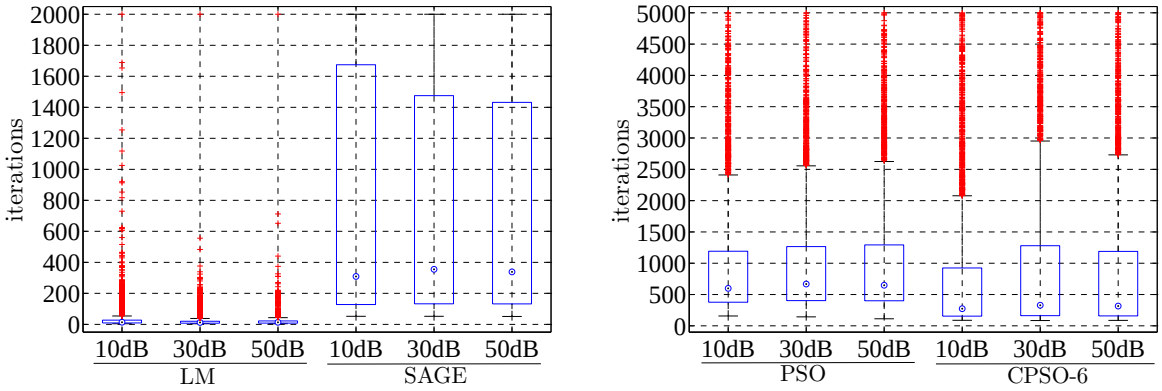
Figure 4.28: Boxplots of iterations for the two-path channels with γ_s= 10 dB, 30 dB and 50 dB.

as well as the outlier range are larger at low SNR. This behaviour is contrary to the fact that less iterations are required at low SNR. Furthermore, all median values are below 300 iterations and all upper whiskers are below 900 iterations which is considerably less than the required number of iterations at high SNR. Due to this mismatch, high outage rates occur at high SNR. In order to reduce the outage rates at high SNR, the stopping criteria can be tightened by decreasing the threshold T and/or increasing $N_{\min}$. In Figure 4.29, the outage rates for the original stopping criteria are compared to outage rates for tightened stopping criteria (LM: T=10^{-5}, $N_{\min}$=2; SAGE/PSO/CPSO-6: T=10^{-7}, $N_{\min}$=50). Exemplarily, the results for the two-path channel with small excess delay are considered at this point. The outage rates at high SNR are significantly reduced, e.g. the outage rate of the SAGE method at 60 dB decreases from 76.7% to 33.4%. Unfortunately, the outage rates are improved at the cost of a higher number of performed iterations over the *whole*



(a) Local methods. (b) Global methods.

Figure 4.29: Outage with different stopping criteria for the two-path channel with $\nu_2/T_s \in [0.05, 1]$.



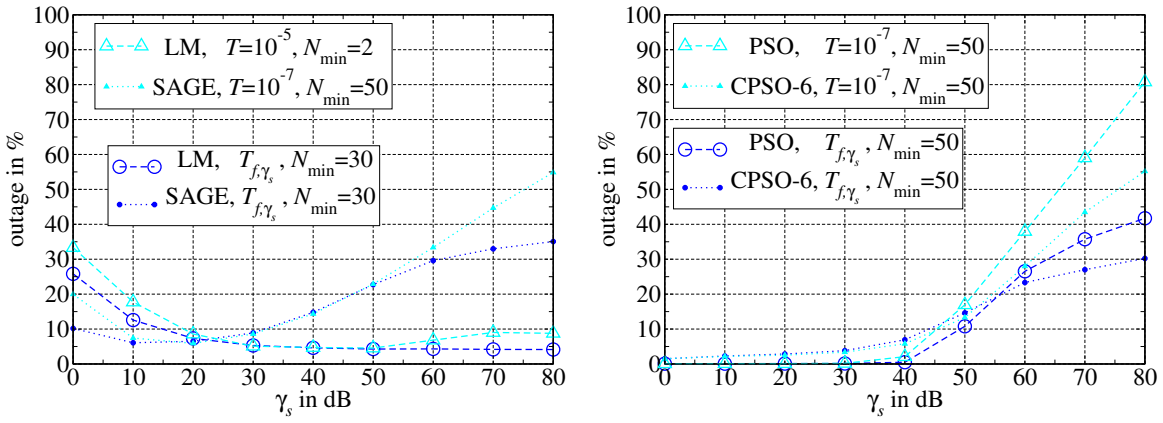
(a) Local methods. (b) Global methods.

Figure 4.30: Boxplots of iterations with tightened stopping criteria for the two-path channel with $\nu_2/T_s \in [0.05, 1]$.

SNR range as shown in Figure 4.30. Hence, there is a strong tradeoff between the outage rate and the number of performed iterations. Ideally, less iterations should be performed by *all* algorithms with decreasing SNR according to the fitness histories in Figure 4.27.

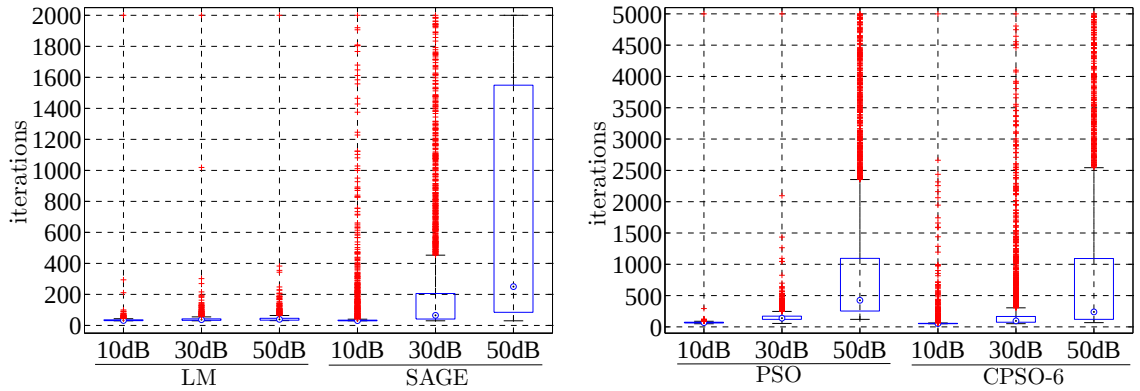
Based on the derived fitness threshold in (4.100), a new stopping criterion can be applied for channel parameter estimation. The new criterion depends on the *absolute* fitness at a certain iteration instead of the fitness improvement. For a given channel with channel memory length L, the fitness threshold T_f is fixed if the normalised final fitness is considered:

$$\Omega(\hat{\boldsymbol{\theta}}) \cdot \gamma_s \overset{!}{\leq} T_f. \tag{4.102}$$



(a) Local methods. (b) Global methods.

Figure 4.31: Outage with stopping criteria based on the fitness threshold T_{f,γ_s} for the two-path channel with $\nu_2/T_s \in [0.05, 1]$. For comparison, the outage rates of the corresponding algorithms for the tightened stopping criterion are plotted.



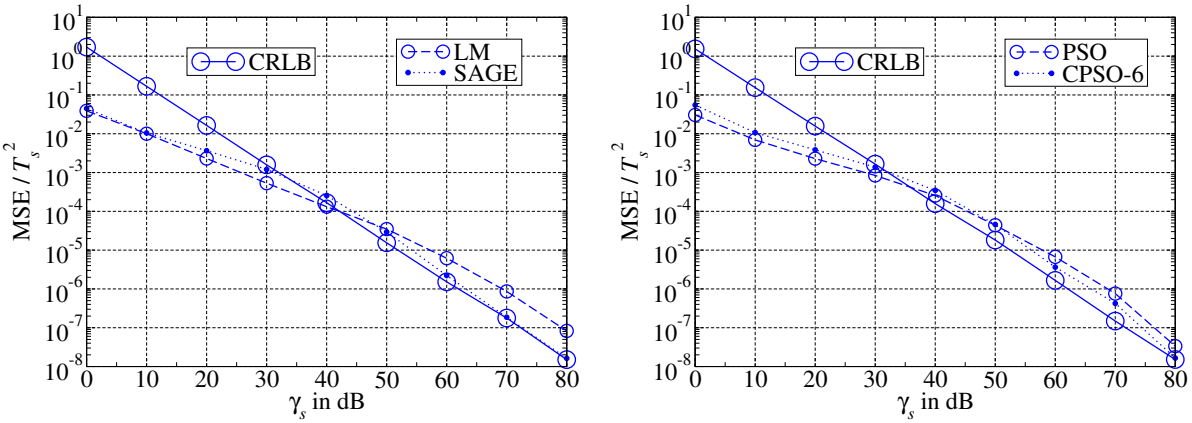
(a) Local methods. (b) Global methods.

Figure 4.32: Boxplots of iterations with stopping criteria based on the fitness threshold T_{f,γ_s} for the two-path channel with $\nu_2/T_s \in [0.05, 1]$.

If the unnormalised final fitness is considered, the threshold decreases with SNR:

$$\Omega(\hat{\boldsymbol{\theta}}) \overset{!}{\leq} \frac{T_f}{\gamma_s} = T_{f,\gamma_s}. \tag{4.103}$$

If the fitness falls below this threshold T_{f,γ_s} during the iterative process, only few further iterations are required on average until a fitness close to the best possible fitness is achieved. Hence, the algorithms can be stopped when the fitness $\Omega(\tilde{\boldsymbol{\theta}})$ is below the threshold T_{f,γ_s} for several times $N_{\min}$ in a sequence. In Figure 4.31, outage rates for the local and global methods applying the new stopping criterion with $N_{\min}$=30 and $N_{\min}$=50, respectively, are shown. For comparison, the outage rates of the corresponding algorithms with the tightened stopping criterion are plotted in Figure 4.31 as well. It is observed that the outage rates can be further decreased with the help of the proposed stopping criterion based on the fitness threshold T_{f,γ_s}. Furthermore, the number of performed iterations depends on the SNR for the fitness threshold based criterion as illustrated in Figure 4.32. For all algorithms, the median number of iterations as well as the interquartile range and the upper whiskers increase with increasing SNR in accordance with the fitness histories in Figure 4.27. But, as usual, the improved performance with respect to the outage rate and the number of performed iterations comes at a certain cost: The MSE performance at high SNR slightly degrades for the new stopping criterion based on T_{f,γ_s} as shown in Figure 4.33. This means that the MSEs of all estimators do not lie exactly *on* the CRLB but slightly *above* the CRLB. Consequently, the stopping criterion has a strong impact on the performance of the algorithms. Hence, the stopping criterion should be chosen with care and in accordance with the requirements of the considered application. In the remainder of this thesis, the new stopping based on T_{f,γ_s} is considered, if not stated otherwise, because it offers the best tradeoff between outage rate, number of performed iterations and the MSE in the interesting SNR region from 0 dB to 40 dB. Especially, the decreased number of iterations at low SNR is advantageous because rapid estimation is generally beneficial.



(a) Local methods.

(b) Global methods.

Figure 4.33: MSE of $\hat{\varepsilon}$ normalised to T_s^2 with stopping criteria based on the fitness threshold T_{f,γ_s} for the two-path channel with $\nu_2/T_s \in [0.05, 1]$.

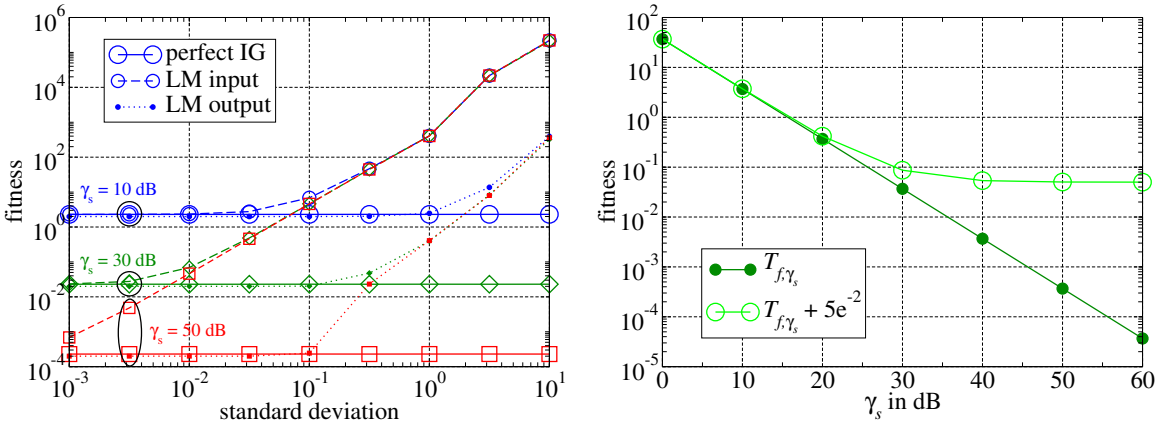
A further performance improvement can be obtained if global and local methods are combined: A global method is used to explore the search space, while a local method is applied for exploitation, i.e. the global method is used to generate a *good* IG for the local method. Subsequently, a combination of the PSO and the LM method is considered. In Figure 4.34(a), the robustness of the LM method against the quality of the IG is examined for γ_s=10 dB (blue), γ_s=30 dB (green) and γ_s=50 dB (red). The input of the LM method, which is denoted by dashed lines in Figure 4.34(a), corresponds to a perfect IG with a random error $\boldsymbol{\delta}$ of a certain standard deviation $\sigma_{\boldsymbol{\delta}}$:

$$\tilde{\boldsymbol{\theta}}_0 = \boldsymbol{\theta} + \boldsymbol{\delta} \qquad \text{with} \qquad \boldsymbol{\delta} \sim \mathcal{N}\left(\mathbf{0}, \sigma_{\boldsymbol{\delta}}^2 \boldsymbol{\mathcal{I}}\right). \tag{4.104}$$

The higher the standard deviation $\sigma_{\boldsymbol{\delta}}$, the larger is the fitness of the IG, i.e. the worse is quality of the IG. Without any outlier detection, the LM method is only able to achieve the best possible fitness, given by the perfect IG, up to a standard deviation of $\sigma_{\boldsymbol{\delta}}$=$10^{-1}$. For larger standard deviations, the final fitness of the LM method increases with decreasing quality of the IG, i.e. the probability of outliers increases. Consequently, the outage rates of the LM method can be kept low if the IG fulfils a certain quality. Based on this knowledge, the PSO and the LM method can be combined in a sophisticated way: The PSO method searches for a parameter estimate up to a certain quality. This rough estimate is used as IG for the LM method, which refines it. That way, the number of iterations can be further limited and the outage rates can be kept low, while retaining a good MSE performance. In Figure 4.34(b), the different fitness thresholds for the combination of the PSO and the LM method are shown. The threshold for the PSO method is given by the threshold T_{f,γ_s} plus a constant:

$$T_{\text{PSO}} = T_{f,\gamma_s} + c, \tag{4.105}$$

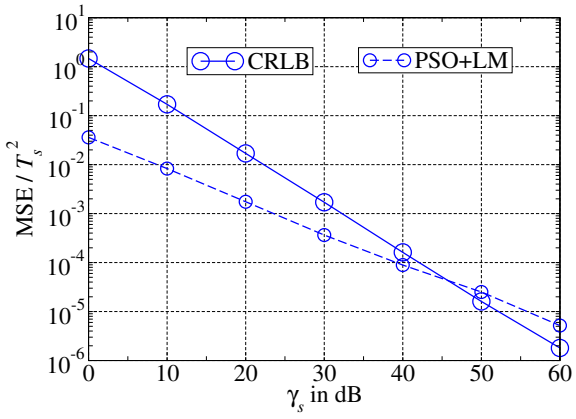
where c is a constant that defines the quality of the PSO parameter estimate because the threshold T_{PSO} saturates at the value given by c. Here, c=$5{\cdot}10^{-2}$ is chosen in accordance



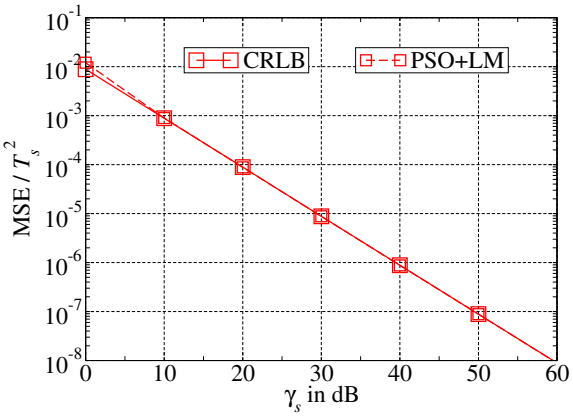
(a) Robustness of the LM method against the quality of the IG for different SNRs.

(b) Different fitness thresholds for the combination of the PSO and the LM method.

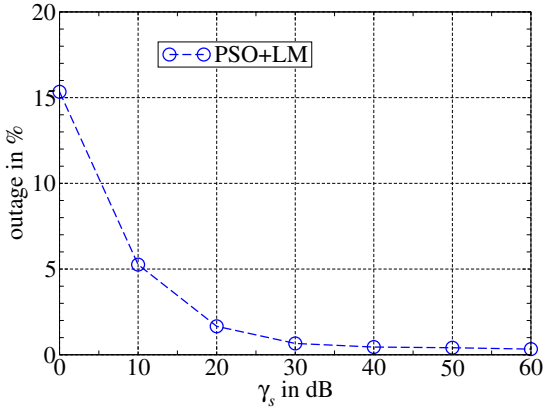
Figure 4.34: On the combination of the PSO and the LM method.



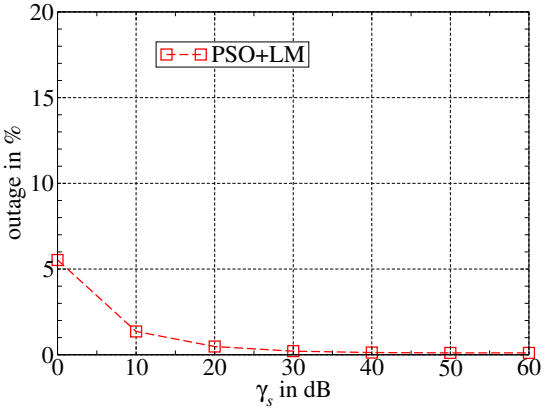
(a) MSE of $\hat{\varepsilon}$ normalised to T_s^2 for $\nu_2/T_s \in [0.05, 1]$.



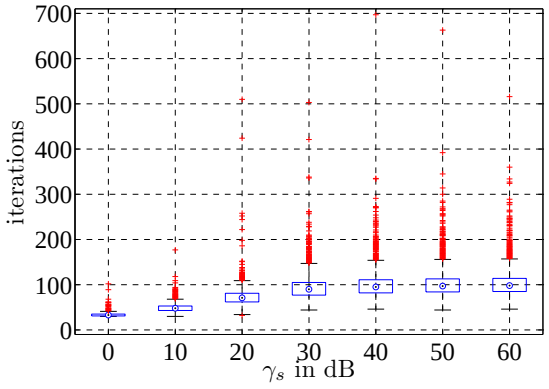
(b) MSE of $\hat{\varepsilon}$ normalised to T_s^2 for $\nu_2/T_s \in [1, 2]$.



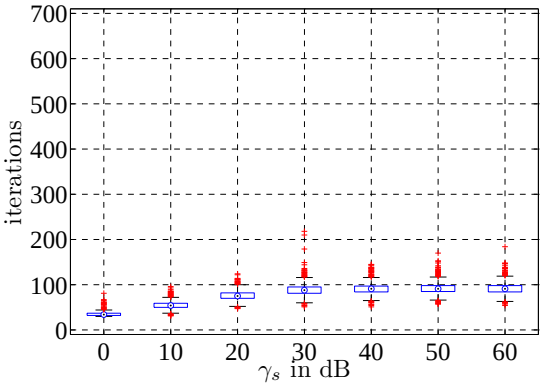
(c) Outage rate for $\nu_2/T_s \in [0.05, 1]$.



(d) Outage rate for $\nu_2/T_s \in [1, 2]$.



(e) Boxplots of iterations for $\nu_2/T_s \in [0.05, 1]$.



(f) Boxplots of iterations for $\nu_2/T_s \in [1, 2]$.

Figure 4.35: Results for the combination of the PSO and the LM method for the two-path channels.

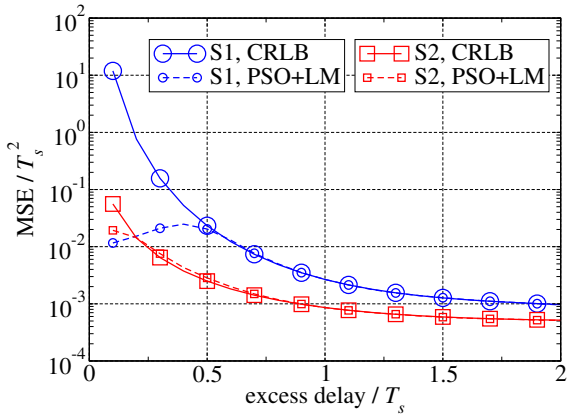
with Figure 4.34(a) since the LM method, which uses the PSO estimate as IG, is able to converge to the best possible fitness for an IG quality of $\sigma_{\boldsymbol{\delta}} \leq 10^{-1}$. For the LM method, the original fitness threshold is applied:

$$T_{\mathrm{LM}} = T_{f,\gamma_s}. \tag{4.106}$$

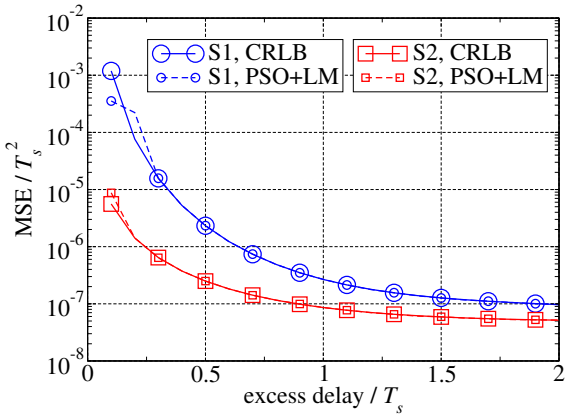
Due to the saturation of the fitness threshold T_{PSO}, the number of performed PSO iterations is limited even for very high SNR. Based on the IG from the PSO, the LM method is able to find the global optimum while performing much less iterations than the PSO method would need to find a final estimate of the same quality. Corresponding results with the thresholds given by (4.105) and (4.106) in combination with $N_{\min}$=20 and $N_{\min}$=10 for the PSO and the LM method, respectively, are shown in Figure 4.35. For both two-path channel models, the normalised MSE of $\hat{\varepsilon}$, the outage rate and the number of performed iterations are illustrated. As expected, the MSE performance of the LM method is retained, while the outage rates can be further reduced. In any case, the outage rate for the combination of the PSO and the LM method is below 16%. For SNRs larger than 30 dB, the outage rates are negligible. At the same time, the number of performed iterations is kept as low as possible even in the challenging case of small excess delays ($\nu_2/T_s \in [0.05, 1]$). For both channel models, the median values of performed iterations increase with increasing SNR, but saturate around 100 iterations for $\gamma_s \geq 30$ dB. Hence, fast and accurate estimation is enabled by the combination of the PSO and the LM method.

Before examining the performance of the proposed estimators for the WINNER channel models, the influence of the excess delay ν_2, the phase offset $\Delta\Phi$ and the power ratio $\mathcal{P}$ on the parameter estimates is studied similar to the investigations in Section 4.3.3. Again, the two basic simulations setups defined in Table 4.1 are applied: The parameter of interest is varied, while the remaining parameters are fixed to the values given in Table 4.1. The first setup (S1) corresponds to a worst case scenario and the second setup (S2) represents a best case scenario. As the performance of the proposed estimator depends on the SNR, γ_s=10 dB and γ_s=50 dB are considered subsequently. In all cases, the above combination of the PSO and the LM method is utilised. In Figure 4.36 to Figure 4.38, the results for the variation of the excess delay ν_2, the phase offset $\Delta\Phi$ and the power ratio $\mathcal{P}$ are shown. In each figure, the normalised MSE of $\hat{\varepsilon}$, the normalised bias and the outage rate are shown. On the left hand side, the results for γ_s=10 dB are illustrated, while the results for γ_s=50 dB are plotted on the right hand side. On average, 40 to 130 iterations are performed in all cases.

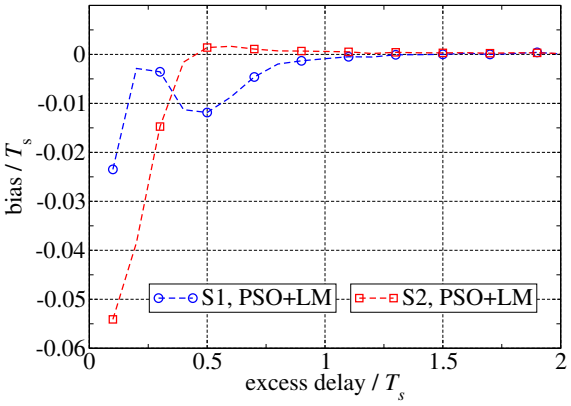
For γ_s=50 dB, the estimator is almost always optimal because the MSEs lie on the CRLBs and the bias as well as the outage rates are negligible. Only for very small excess delays, slight deviations from the optimal performance occur. For γ_s=10 dB, different results are obtained. Considerable deviations from the optimal performance occur, i.e. MSEs below and above the CRLBs are possible and the bias as well as the outage rates are not negligible anymore. The smaller the excess delay, the more difficult it is to separate the propagations paths. For excess delays smaller than a symbol duration, the estimator becomes biased. The bias increases with decreasing excess delay. Hence, MSEs below



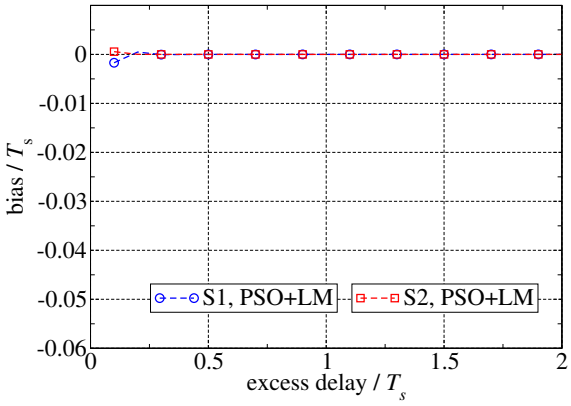
(a) MSE of $\hat{\varepsilon}$ normalised to T_s^2 at 10 dB.



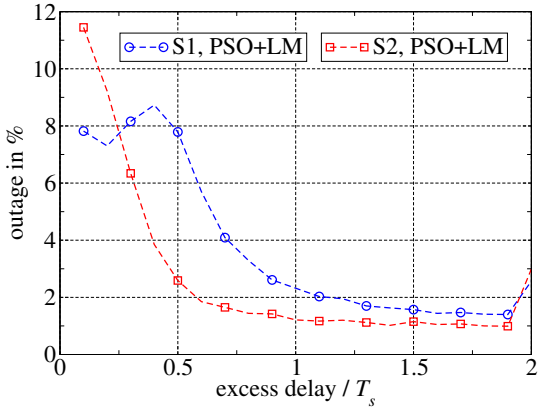
(b) MSE of $\hat{\varepsilon}$ normalised to T_s^2 at 50 dB.



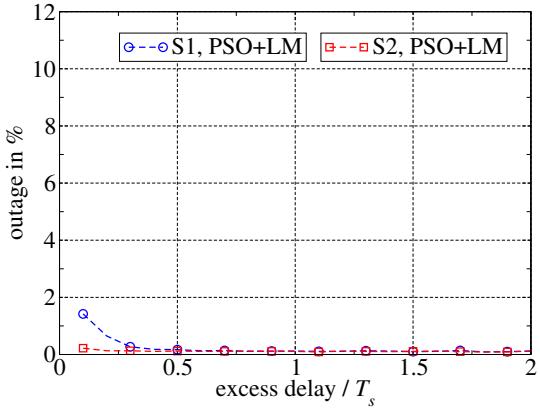
(c) Bias of $\hat{\varepsilon}$ normalised to T_s at 10 dB.



(d) Bias of $\hat{\varepsilon}$ normalised to T_s at 50 dB.

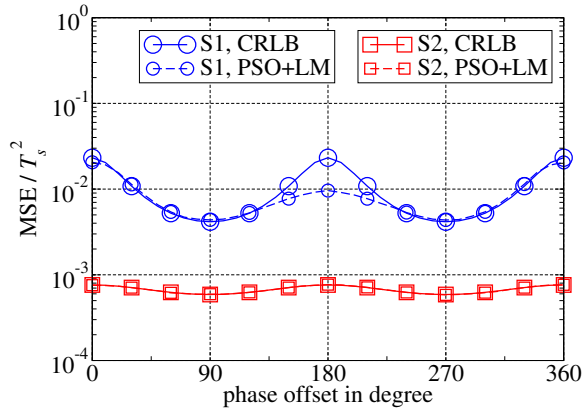


(e) Outage at 10 dB.

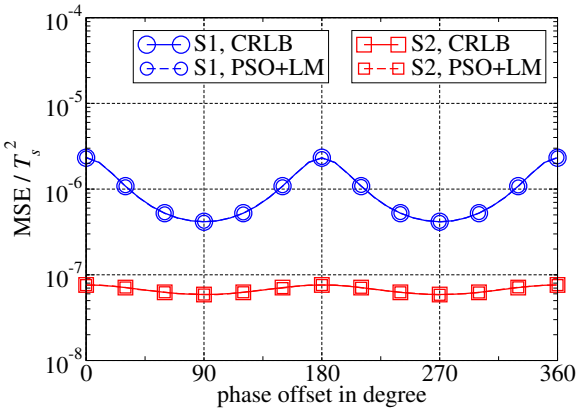


(f) Outage at 50 dB.

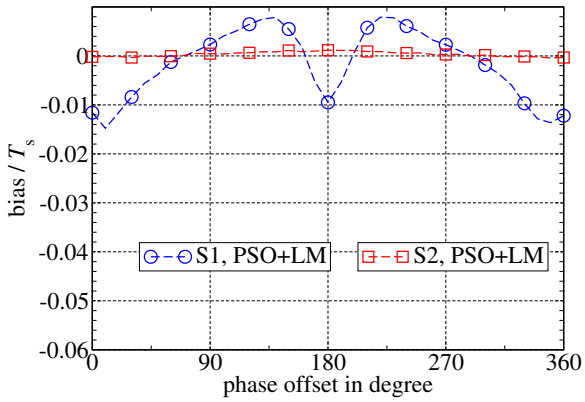
Figure 4.36: Results for the combination of the PSO and the LM method over excess delay. The curves are labelled with "S1" for the first setup and "S2" for the second setup.



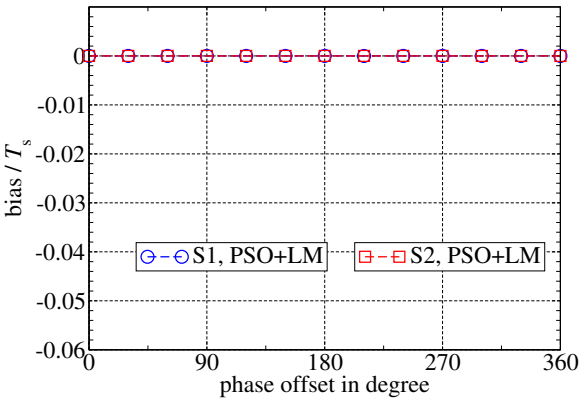
(a) MSE of $\hat{\varepsilon}$ normalised to T_s^2 at 10 dB.



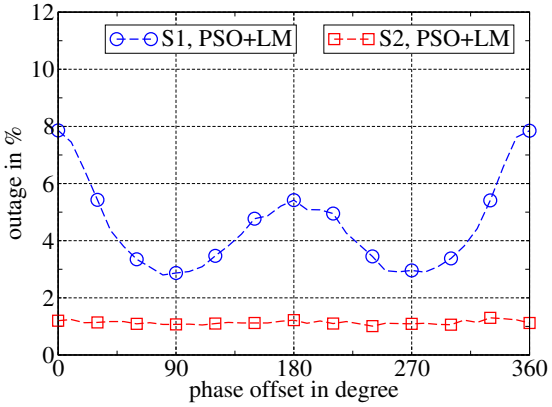
(b) MSE of $\hat{\varepsilon}$ normalised to T_s^2 at 50 dB.



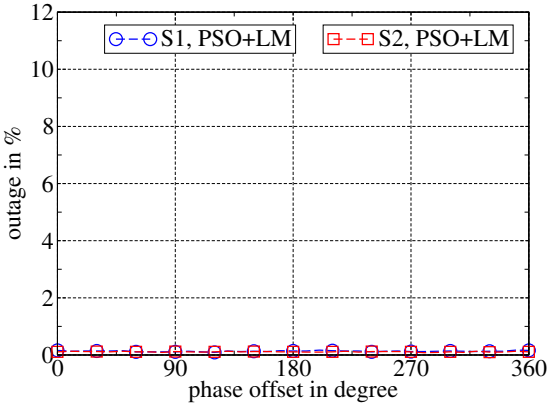
(c) Bias of $\hat{\varepsilon}$ normalised to T_s at 10 dB.



(d) Bias of $\hat{\varepsilon}$ normalised to T_s at 50 dB.

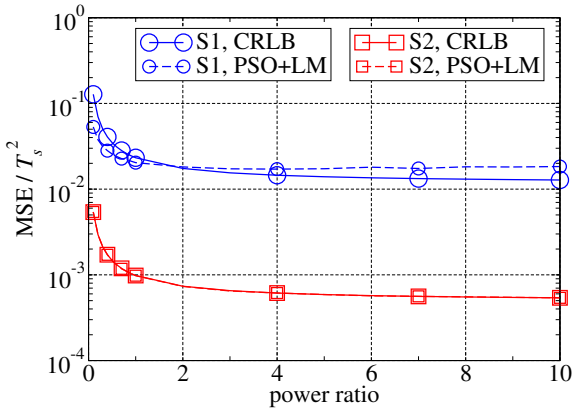


(e) Outage at 10 dB.

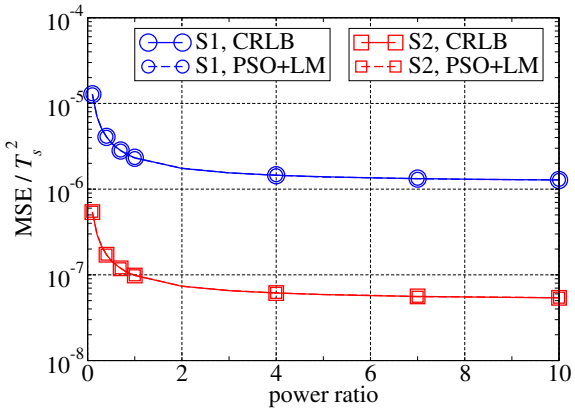


(f) Outage at 50 dB.

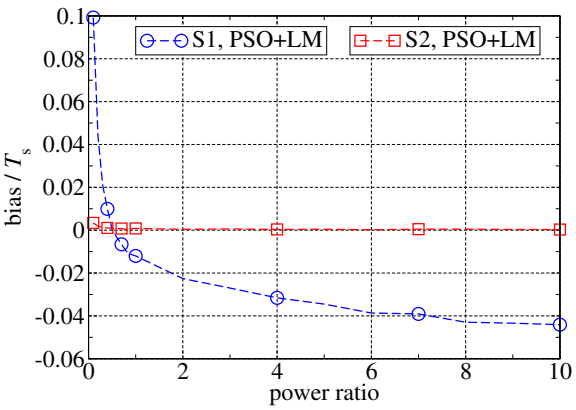
Figure 4.37: Results for the combination of the PSO and the LM method over phase offset. The curves are labelled with "S1" for the first setup and "S2" for the second setup.



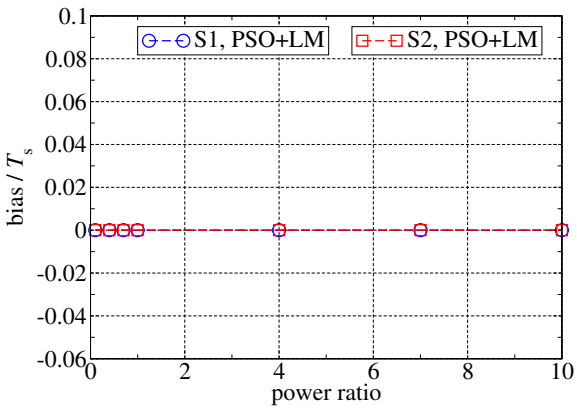
(a) MSE of $\hat{\varepsilon}$ normalised to T_s^2 at 10 dB.



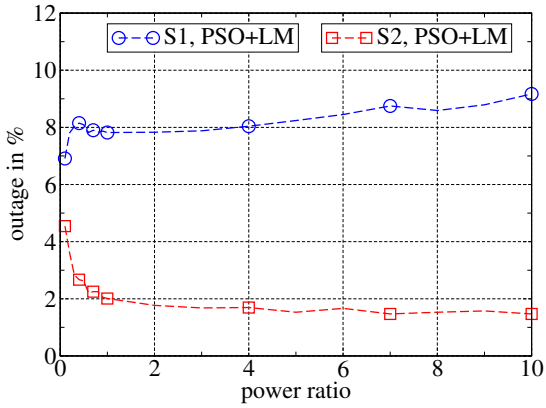
(b) MSE of $\hat{\varepsilon}$ normalised to T_s^2 at 50 dB.



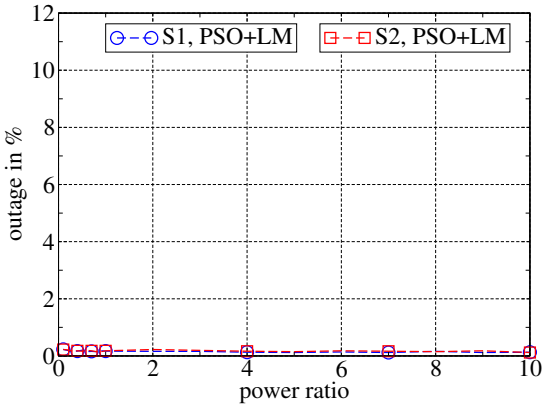
(c) Bias of $\hat{\varepsilon}$ normalised to T_s at 10 dB.



(d) Bias of $\hat{\varepsilon}$ normalised to T_s at 50 dB.



(e) Outage at 10 dB.



(f) Outage at 50 dB.

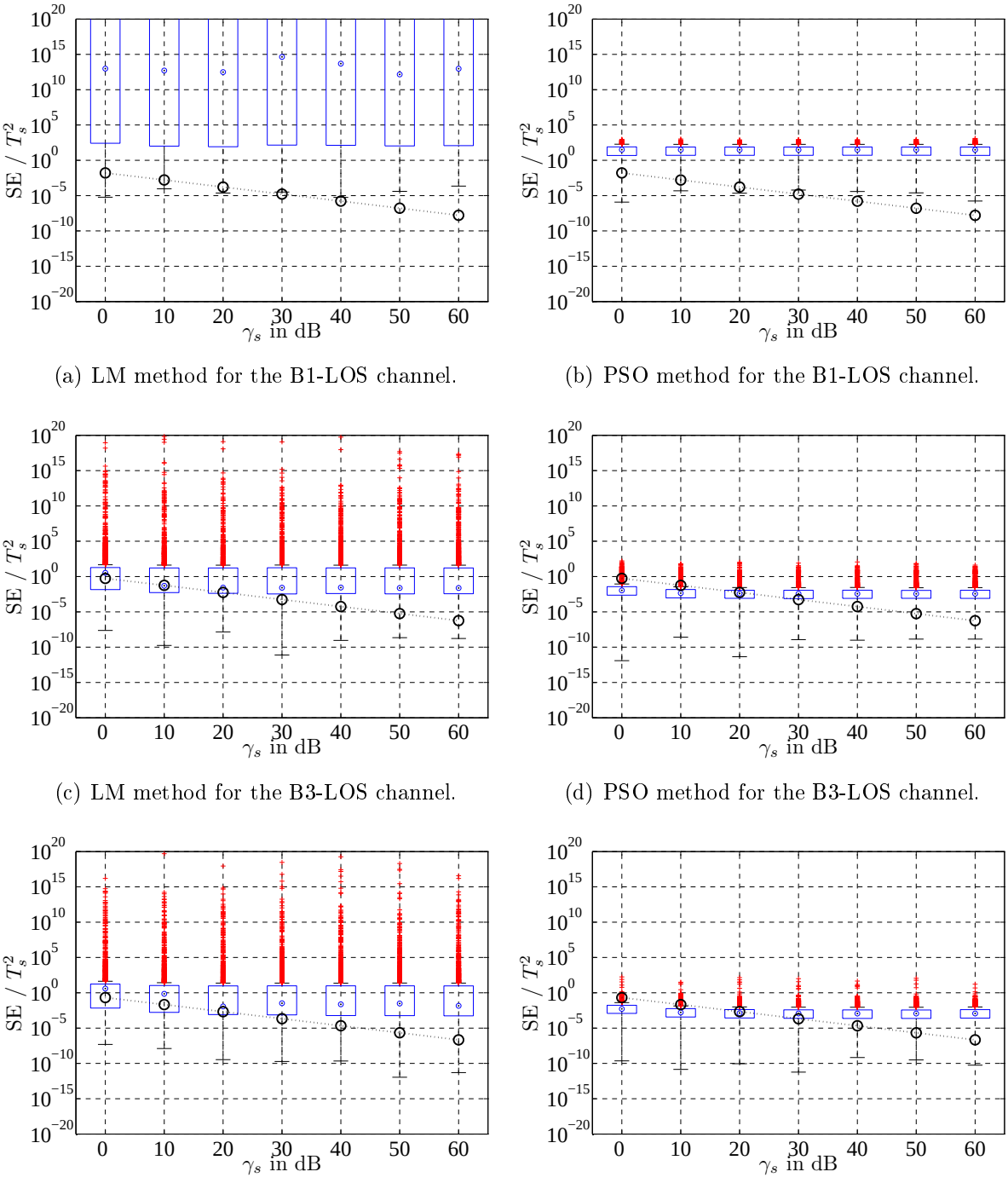
Figure 4.38: Results for the combination of the PSO and the LM method over power ratio. The curves are labelled with "S1" for the first setup and "S2" for the second setup.

the CRLBs are obtained for small excess delays. At the same time, the outage rates increase with decreasing excess delays, but, in all cases, the outage rates are below 12%. Concerning the phase offset and the power ratio, only the first setup (worst case) shows considerable deviations from the optimal performance. With respect to the phase offset, the best performance is obtained if the propagation paths are perpendicular. Otherwise, a bias occurs and the outage rate increases. The outage rate is highest for similar phases, while the MSE is below the CRLB for opposite phases. Concerning the power ratio, deviations over the whole power ratio range occur for the first setup. The outage rate is nearly constant at 8%. For small power ratios, there is a positive bias, while a negative bias is present for large power ratios. Similarly, the MSE is below the CRLB for small power ratios, while it is above the CRLB for large power ratios. Overall, the MSEs become worse the smaller the power ratio gets.

In summary, the results in Section 4.3.3 are confirmed by the results in this section: Dense multipath scenarios with similar or opposite phases and small power ratios are most challenging. Additionally, it is observed that larger SNRs facilitate channel parameter estimation in the sense that the accuracy as well as the robustness (negligible outage) are considerably improved.

WINNER Channel Models

Finally, the performance of the proposed estimator is examined for the WINNER channel models. Basically, the same simulation setup as already applied for the preceding channel models is used. However, only three particles per dimension are applied for the PSO method due to the high dimensionality of the channel parameter estimator in case of the WINNER channel models. The insights obtained for the two-path channels are exploited at this point. Nevertheless, at first, boxplots of the SE performance for the LM and the PSO method with the initial stopping criterion (LM: T=10^{-3}, $N_{\min}$=2; PSO: T=10^{-5}, $N_{\min}$=30) and without outage detection are considered as shown in Figure 4.39 and Figure 4.40 for T_s=10 ns and T_s=50 ns, respectively. A general problem is illustrated in these figures. Although it is possible to determine the CRLBs for the WINNER channel models, as already illustrated in Section 4.3.3, these CRLBs are more or less meaningless from a practical point of view: For T_s=10 ns, the CRLBs are in a reasonable range, while extreme high values are obtained for the CRLBs with T_s=50 ns because very small excess delays (ν_c/T_s<0.1) may occur, that lead to ill-conditioned Fisher information matrices, and because there may be more unknowns than observed samples. Hence, it was concluded in Section 4.3.3 that channel parameter estimation might be futile for T_s=50 ns. However, the performance of the estimator is basically the same for both symbol durations. The median values of the normalised SE are generally below 10^0 and saturate for high SNR at values between 10^{-3} and $2{\cdot}10^{-2}$. Only for the B1-LOS channel in combination with T_s=10 ns, the estimator fails completely (see Figure 4.39(a) and Figure 4.39(b)). From a theoretical point of view, the best performance is expected for this channel model because the CRLB is lowest in this case. Practically, the estimator suffers from numerical instabilities in case of the B1-LOS channel with T_s=10 ns caused



(a) LM method for the B1-LOS channel.

(b) PSO method for the B1-LOS channel.

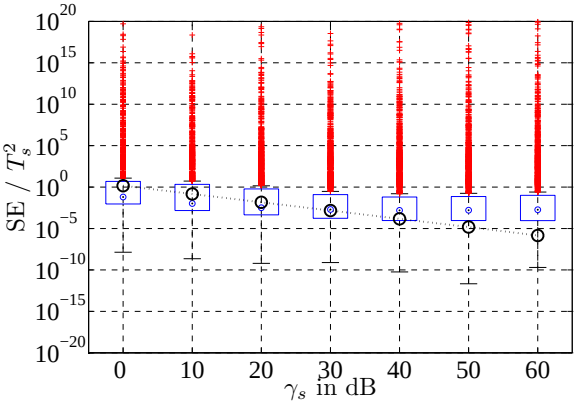
(c) LM method for the B3-LOS channel.

(d) PSO method for the B3-LOS channel.

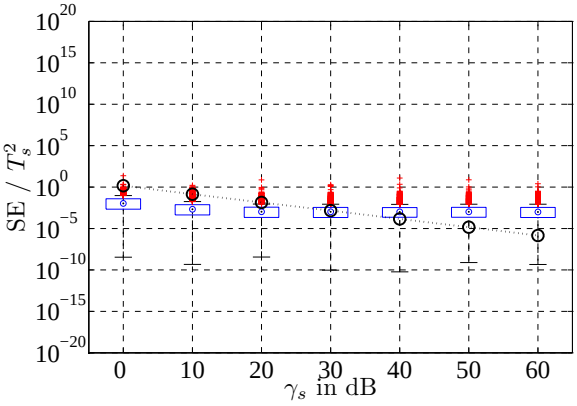
(e) LM method for the C2-LOS channel.

(f) PSO method for the C2-LOS channel.

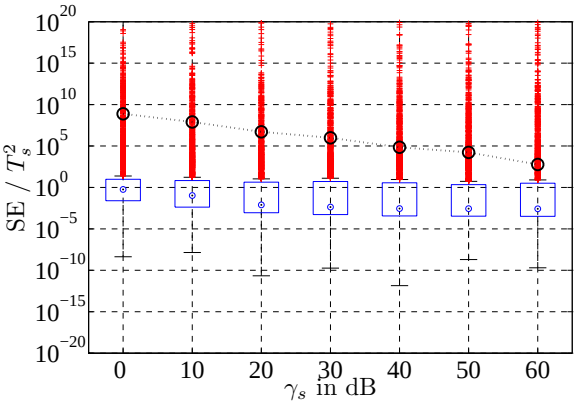
Figure 4.39: Boxplots of the SE of $\hat{\varepsilon}$ normalised to T_s^2 for the WINNER channels with T_s=10 ns. For comparison, the corresponding CRLBs are shown, which are denoted by large black circles.



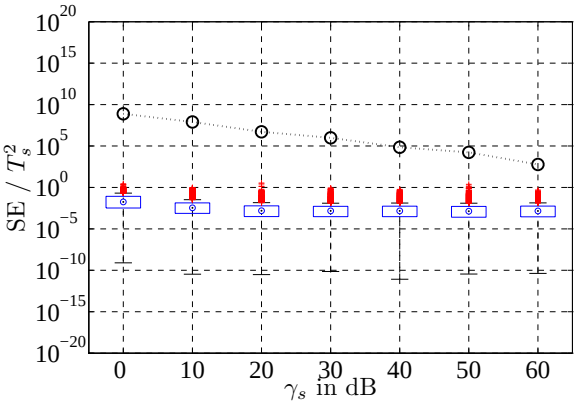
(a) LM method for the B1-LOS channel.



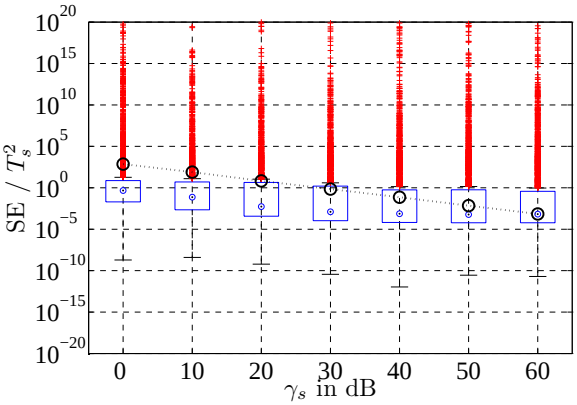
(b) PSO method for the B1-LOS channel.



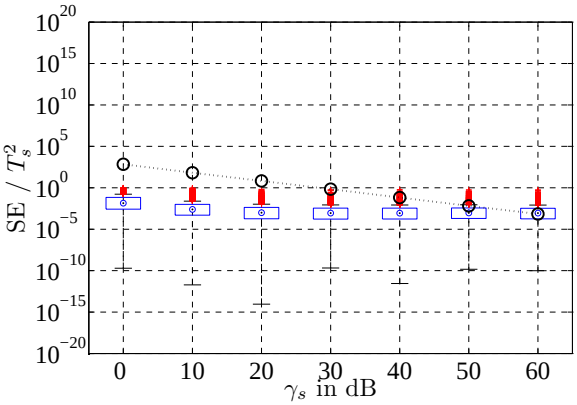
(c) LM method for the B3-LOS channel.



(d) PSO method for the B3-LOS channel.



(e) LM method for the C2-LOS channel.



(f) PSO method for the C2-LOS channel.

Figure 4.40: Boxplots of the SE of $\hat{\varepsilon}$ normalised to T_s^2 for the WINNER channels with T_s=50 ns. For comparison, the corresponding CRLBs are shown, which are denoted by large black circles.

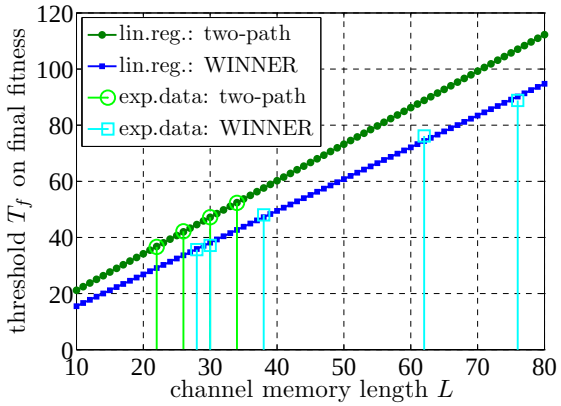


Figure 4.41: Adjusted fitness threshold for the WINNER channel models. For comparison, the fitness threshold derived for the two-path channels is plotted.

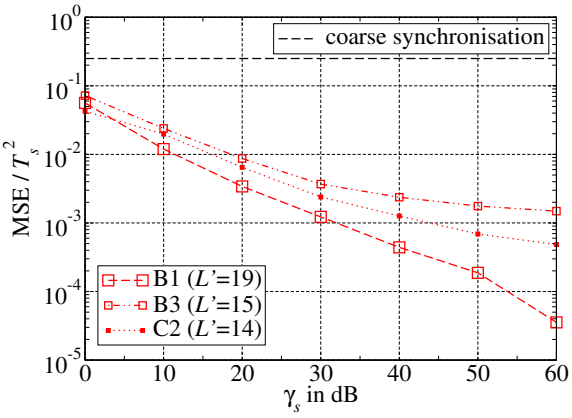
by a very large channel memory length L=110 (L'=55). From this point of view, larger symbol durations T_s are advantageous since they lead to smaller channel memory lengths L such that the estimator does not fail anymore (see Figure 4.40(a) and Figure 4.40(b)). In contrast, smaller symbol durations T_s are preferable from a theoretical point of view (based on the CRLBs) because neighbouring propagation paths become separable in this case. Hence, there is a significant discrepancy between theory and praxis. The practical point of view is preferred for the WINNER channel models, i.e. only symbol durations $T_s \geq 50$ ns are considered in the following. In this case, the CRLBs are meaningless high. Hence, they are not further considered subsequently. Instead, the normalised MSE of the assumed coarse synchronisation is plotted for comparison if the MSE of $\hat{\varepsilon}$ is considered. In this way, the improvement due to channel parameter estimation becomes clear and it can be shown that channel parameter estimation is *not* futile even for $T_s \geq 50$ ns.

For the following results, the combination of the PSO and the LM method (as developed for the two-path channels) is applied in combination with the stopping criterion and outage detection based on the fitness threshold T_f. An inspection of the fitness threshold T_f for the WINNER channel models shows that the relationship between the channel memory length L and the threshold T_f, defined in (4.100) for the two-path channels, needs to be adjusted to

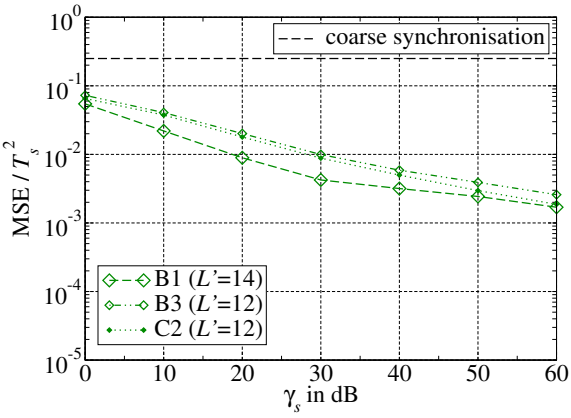
$$T_f = 1.1321 \cdot L + 4.1884. \tag{4.107}$$

The difference between the relationship for the two-path channels (4.100) and the relationship for the WINNER channel models (4.107) is illustrated in Figure 4.41. Probably, the fitness thresholds differ because fading effects[12] are considered in the WINNER channel models, that are not present in case of the two-path channels. In Figure 4.42, the results for the combination of the PSO and the LM method are shown. For the PSO method, the minimum number, for which the absolute fitness $\Omega(\tilde{\boldsymbol{\theta}})$ must be smaller than T_{PSO}=$T_f \cdot \gamma_s$+5·10^{-2}, is set to $N_{\min}$=50, while T_{LM}=$T_f \cdot \gamma_s$ and $N_{\min}$=10 are applied for

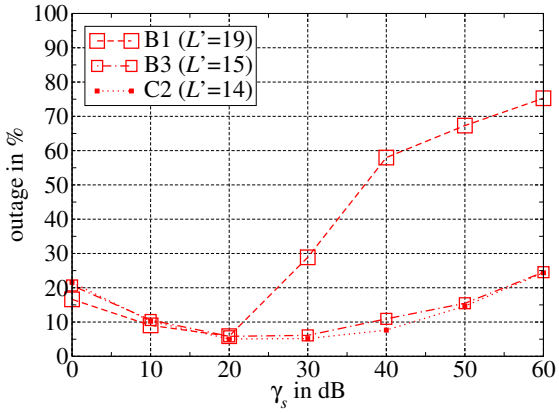
[12]For each cluster c of a WINNER channel model, R=20 rays contribute to the complex cluster amplitude f_c. Due to the superposition of the rays, fading effects occur.



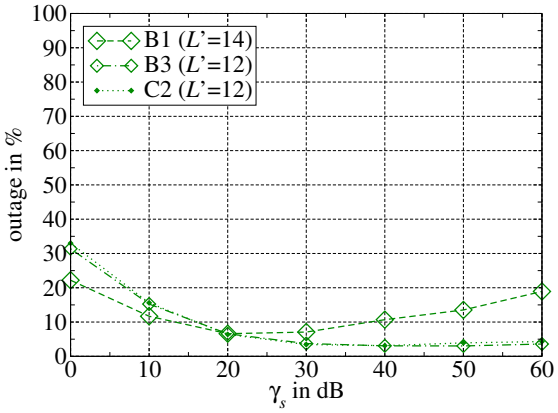
(a) MSE of $\hat{\varepsilon}$ normalised to T_s^2 for T_s=50 ns.



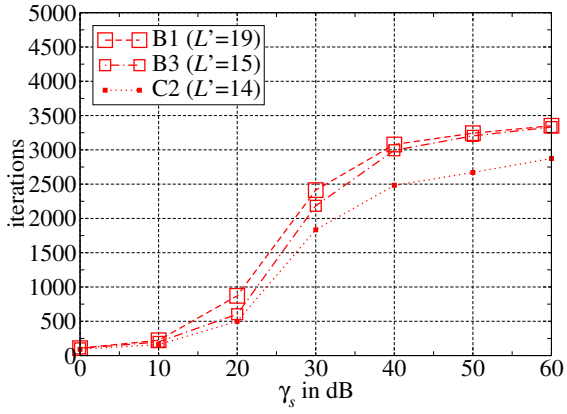
(b) MSE of $\hat{\varepsilon}$ normalised to T_s^2 for T_s=100 ns.



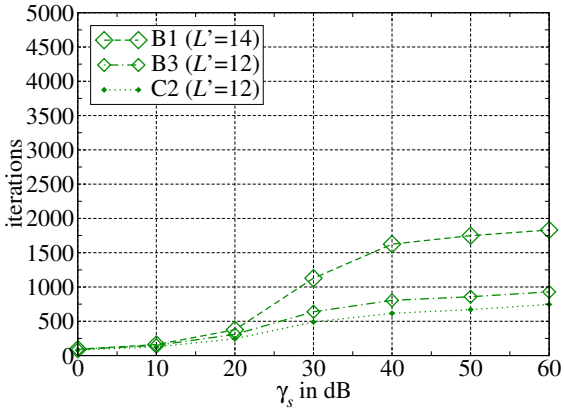
(c) Outage for T_s=50 ns.



(d) Outage for T_s=100 ns.



(e) Performed iterations for T_s=50 ns.



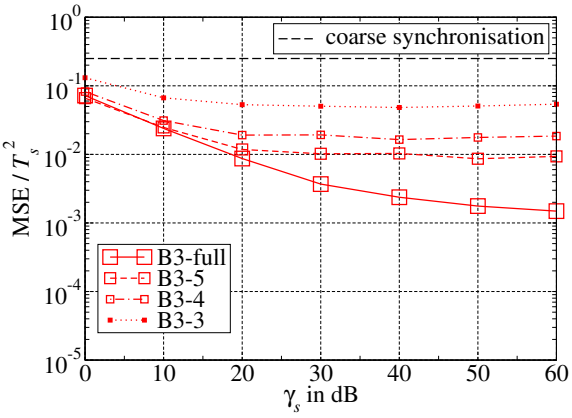
(f) Performed iterations for T_s=100 ns.

Figure 4.42: Results for the combination of the PSO and the LM method for the WINNER channel models.

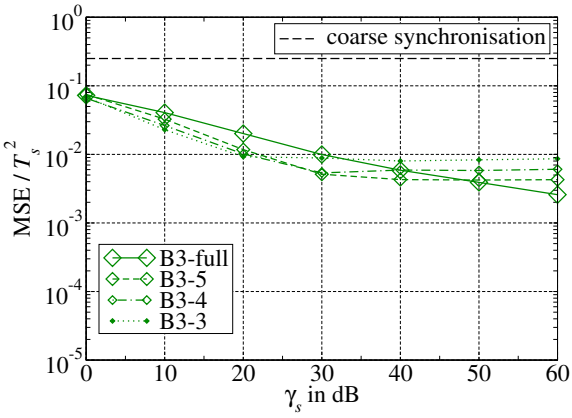
the LM method. Two different symbol durations, namely T_s=50 ns and T_s=100 ns, are considered: The results for T_s=50 ns are shown on the left hand side of Figure 4.42, while the results for T_s=100 ns are illustrated on the right hand side. In Section 4.3.3, it was shown that the B3-LOS channel is most challenging because of a small Rician factor and small excess delays. The best performance was obtained for the B1-LOS channel. Concerning the MSE performance shown in Figure 4.42(a) and Figure 4.42(b), these results are confirmed: The best and worst performance are obtained for the B1-LOS channel (urban micro-cell) and the B3-LOS channel (large indoor hall), respectively. Although the performance degrades with increasing symbol duration, even for T_s=100 ns, a significant improvement with respect to the coarse synchronisation is achieved, i.e. channel parameter estimation is generally useful. With respect to the outage rate and the number of performed iterations, the C2-LOS channel shows the best performance, while the worst performance is obtained for the B1-LOS channel. Especially for large SNRs with T_s=50 ns, the outage rate of the B1-LOS is much higher (approximately 75% at 60 dB) than the outage rate of the other channels (approximately 25% at 60 dB). For T_s=100 ns, the outage rates at high SNR are generally reduced to approximately 20% and 5% at 60 dB for the B1-LOS channel and the other two channels, respectively. Similarly, the average number of performed iterations is smaller for large symbol durations: For T_s=100 ns, between 750 and 2000 iterations are performed on average at 60 dB, while 2750 to 3500 iterations are performed at the same SNR for T_s=50 ns. Hence, there is a tradeoff between the MSE performance and the outage rate as well as the number of performed iterations.

As already mentioned in Section 4.3.3, it may be advantageous for the WINNER channels to combine multipath components with similar delays in order to reduce the number of parameters P. The estimator that considers all multipath components C is denoted as *full* estimator, while estimators that incorporate less than C components are called *reduced* estimators. Subsequently, the performance of reduced estimators with three, four and five components is compared to the performance of the full estimator. Exemplarily, the B3-LOS channel (C=10) for large indoor halls is considered. Results for the B1-LOS channel (C=8) and the C2-LOS channel (C=8) are shown in Figure C.3 and Figure C.4 in Appendix C, respectively. As before, a combination of the PSO and the LM method is used, i.e. the PSO generates an IG for the LM method, which determines the final estimate. Unfortunately, the stopping criterion and the outage detection based on the fitness threshold T_f are not applicable for the reduced estimators due to the introduction of a systematic error. Therefore, the initial stopping criterion based on the fitness improvement is utilised (LM: T=10^{-3}, $N_{\min}$=2; PSO: T=10^{-5}, $N_{\min}$=50) for the reduced estimators. Concerning outage detection, only a boundary check is performed after the overall estimation process. That way, extreme outliers outside the search space can be avoided. Corresponding results are shown in Figure 4.43. In order to distinguish between different reduced estimators, the number of considered multipath components is appended to the name of the channel model, e.g. "B3-4" denotes the estimator that incorporates four multipath components. The full estimator is denoted by "B3-full".

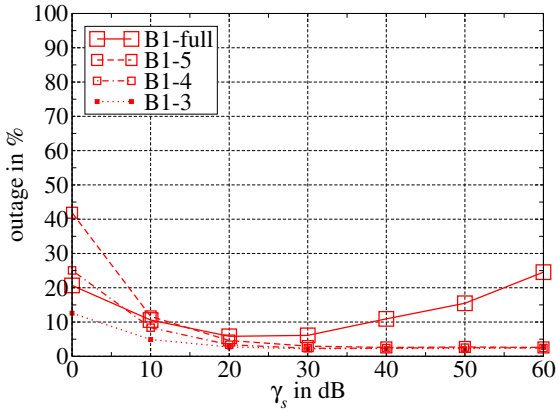
Again, there is a tradeoff between the MSE performance and the outage rate as well as the number of performed iterations. Generally, the MSEs of the reduced estimators



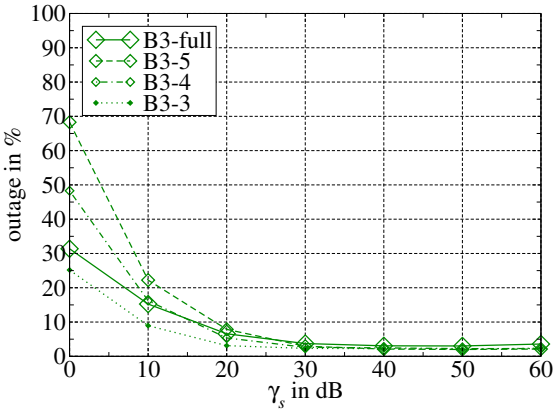
(a) MSE of $\hat{\varepsilon}$ normalised to T_s^2 for T_s=50 ns.



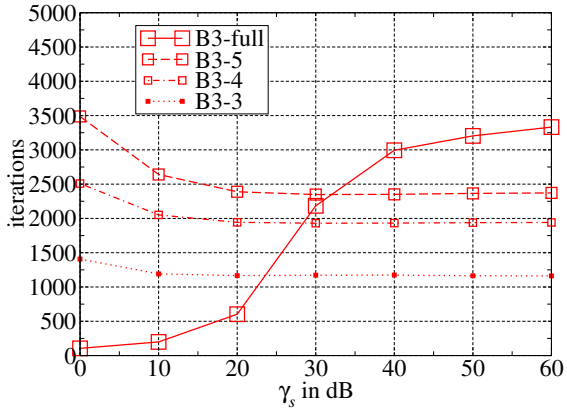
(b) MSE of $\hat{\varepsilon}$ normalised to T_s^2 for T_s=100 ns.



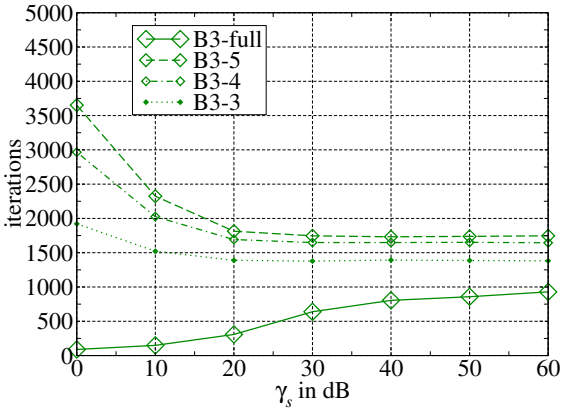
(c) Outage for T_s=50 ns.



(d) Outage for T_s=100 ns.



(e) Performed iterations for T_s=50 ns.



(f) Performed iterations for T_s=100 ns.

Figure 4.43: Performance of the full estimator and several reduced estimators in the B3-LOS channel.

decrease with increasing number of considered components. At the same time, the outage rates and the number of performed iterations increase. Consequently, reduced estimators with less components are advantageous from an outage/iteration point of view, while estimators with more components are beneficial with respect to MSEs. In comparison to the full estimator, the reduced estimators generally show a different behaviour because different stopping criteria are applied for both types. This difference is especially obvious for the average number of performed iterations: For the full estimator, the number of iterations increases with increasing SNR, while the number of iterations decreases with increasing SNR for the reduced estimators. Especially at low SNR, a large gap between the full and the reduced estimators occurs. In case of T_s=100 ns, the full estimator actually performs much less iterations over the whole SNR range. But it should be kept in mind that a *single* iteration of the full estimator is more complex due to an increased number of parameters. Interestingly, the reduced estimators exhibit better MSEs at low SNR than the full estimator for T_s=100 ns. For SNRs up to 20 dB, the reduced estimator with only three multipath components shows even the best performance. Hence, it can be concluded that the reduced estimators become more advantageous concerning the MSE with increasing symbol durations: The larger the symbol duration, the more difficult it gets to separate neighbouring components and the higher is the benefit of combining multipath components with similar delays. This effect is even more pronounced in case of the C2-LOS channel as shown in Figure C.4 in Appendix C. Since the true number of propagation paths is typically not known and symbol durations $T_s \geq 100$ ns are often applied, the reduced estimator with three multipath components is a good choice because it offers a good tradeoff between MSE and outage/iterations and may even outperform estimators that incorporate more components with respect to the MSE.

4.5 Soft Information for Channel Parameter Estimation

In the previous section, the performance of the proposed optimisation methods for channel parameter estimation has been investigated. In this context, only *hard* estimates have been considered. Often, it is desirable to gain additional information about the *reliability* of the estimates, which is also called *soft information*. The concept of soft information is already widely applied: In the area of communication, soft information is used for decoding, detection and equalisation [HROW07]. In the field of navigation, soft information is exploited for sensor fusion [HROW07]. This section aims at obtaining soft information for the parameter estimates in order to improve the positioning accuracy *before* sensor fusion is applied: For positioning, estimated TOAs from several reference objects (ROs) are needed (see Section 3.1). Some of the TOAs are more reliable, while some are less reliable. By means of soft information, the estimated TOAs can be weighted according to their reliability such that the overall positioning accuracy is increased (see Chapter 5). Parts of this section have already been published in [SAH11b].

4.5.1 Definition of Soft Information

Soft information is a measure of reliability of the (hard) estimates. The intention is to determine the a posteriori distribution of the estimates. Hence, the (hard) estimate is the mean of the distribution and soft information corresponds to the variance of the distribution. For linear estimation problems with known noise covariance matrix, the a posteriori distribution of the estimates can be determined in closed form [Kay93]. However, only few problems are linear. As already mentioned in Section 4.2, ML estimators are quite popular for more general problems because they are asymptotically (for a large number of observations or at a high SNR) unbiased and efficient. Hence, an asymptotic a posteriori distribution for the parameter estimates can be determined: For Gaussian noise with covariance matrix $\boldsymbol{C_n}=\boldsymbol{C}\cdot\sigma_n^2$, the asymptotic covariance matrix of the estimates is given by the inverse of the Fisher information matrix evaluated at the true parameters [Kay93]. Thus, the parameter estimate $\hat{\boldsymbol{\theta}}$ is asymptotically distributed according to

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}\left(\boldsymbol{\theta}, \boldsymbol{C}_{\hat{\boldsymbol{\theta}}}\right) \tag{4.108}$$

with asymptotic covariance matrix

$$\begin{aligned} \boldsymbol{C}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{I}^{-1}(\boldsymbol{\theta}) &\overset{(4.42)}{=} \left[2\,\mathrm{Re}\left\{\boldsymbol{J}^H(\boldsymbol{\theta})\boldsymbol{X}_0^H\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{X}_0\boldsymbol{J}(\boldsymbol{\theta})\right\}\right]^{-1} \\ &= \frac{\sigma_n^2}{2}\left[\mathrm{Re}\left\{\boldsymbol{J}^H(\boldsymbol{\theta})\boldsymbol{X}_0^H\boldsymbol{C}^{-1}\boldsymbol{X}_0\boldsymbol{J}(\boldsymbol{\theta})\right\}\right]^{-1}. \end{aligned} \tag{4.109}$$

In general, the true parameter vector $\boldsymbol{\theta}$ is not known. Therefore, the asymptotic covariance matrix can not be determined and an approximation has to be found. In (4.78), an approximate covariance matrix, which is often applied in the context of the LM method, has already been presented:

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} \approx \boldsymbol{H}_{\Omega}^{-1}(\hat{\boldsymbol{\theta}}) \approx \frac{\sigma_n^2}{2}\left[\mathrm{Re}\left\{\boldsymbol{J}^H(\hat{\boldsymbol{\theta}})\boldsymbol{X}_0^H\boldsymbol{C}^{-1}\boldsymbol{X}_0\boldsymbol{J}(\hat{\boldsymbol{\theta}})\right\}\right]^{-1}.$$

Obviously, the only difference between the asymptotic covariance matrix in (4.109) and the approximate covariance matrix in (4.78) is that the Jacobian is evaluated at the parameter estimate $\hat{\boldsymbol{\theta}}$ instead of the true parameter $\boldsymbol{\theta}$. Other approximate methods to obtain soft information are available in the literature as well. However, all approximate methods should be used with caution since the underlying approximation may be very poor [DS83], [DS85]. In [BW80], Bates and Watts introduce so-called *curvature measures* that can be used to judge the quality of approximate soft information. Subsequently, two different methods to obtain soft information for channel parameter estimation are proposed and analysed based on the curvature measures by Bates and Watts. First, a short description of confidence regions is included because confidence regions are closely related to soft information.

4.5.2 Confidence Regions

In [DS85], Donaldson and Schnabel investigate different methods to construct confidence regions and confidence intervals. Confidence regions and intervals are closely related to

soft information since they also indicate reliability: The estimated parameters $\hat{\boldsymbol{\theta}}$ do not coincide with the true parameters $\boldsymbol{\theta}$ because of the measurement noise. A confidence region indicates the area around the estimated parameters in which the true parameters might be with a specific probability. This probability is called *confidence level* and is often expressed as a percentage. A commonly used confidence level is 95%. For linear problems with Gaussian noise, the confidence regions are elliptical and can be determined exactly by the covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{\theta}},\text{linear}}$, which can be computed in closed form [Kay93]. The linear confidence region consists of all parameter vectors $\tilde{\boldsymbol{\theta}}$ that satisfy the following formula:

$$\left(\tilde{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right)^T \boldsymbol{C}^{-1}_{\hat{\boldsymbol{\theta}},\text{linear}} \left(\tilde{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right) \leq P\,\mathcal{F}^{1-\alpha}_{P,N-P}, \tag{4.110}$$

in which P is the number of parameters, N is the number of observations, $1-\alpha$ is the confidence level and $\mathcal{F}$ is the Fisher distribution [DS85]. In case of channel parameter estimation, $P{=}3C$ and $N{=}L{+}1$. According to [DS85], the most common method to determine a confidence region for a nonlinear problem is based on a linearisation of the problem in order to obtain an approximate covariance matrix. Subsequently, the following approximate covariance matrix is used[13]:

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}},\text{approx}} = \frac{s^2}{2}\left[\text{Re}\left\{\boldsymbol{J}^H(\hat{\boldsymbol{\theta}})\boldsymbol{X}_0^H\boldsymbol{C}^{-1}\boldsymbol{X}_0\boldsymbol{J}(\hat{\boldsymbol{\theta}})\right\}\right]^{-1}. \tag{4.111}$$

The approximate covariance matrix in (4.111) resembles the approximate covariance matrix in (4.78) with the difference that the *residual variance* $s^2{=}\Omega(\hat{\boldsymbol{\theta}})/(N{-}P)$ instead of the true variance σ_n^2 is considered. When $\boldsymbol{C}_{\hat{\boldsymbol{\theta}},\text{linear}}$ in (4.110) is replaced by $\boldsymbol{C}_{\hat{\boldsymbol{\theta}},\text{approx}}$ in (4.111), an approximate confidence region for nonlinear problems is obtained as

$$\left(\tilde{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right)^T 2\,\text{Re}\left\{\boldsymbol{J}^H(\hat{\boldsymbol{\theta}})\boldsymbol{X}_0^H\boldsymbol{C}^{-1}\boldsymbol{X}_0\boldsymbol{J}(\hat{\boldsymbol{\theta}})\right\}\left(\tilde{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right) \leq s^2\,P\,\mathcal{F}^{1-\alpha}_{P,N-P}. \tag{4.112}$$

On the one hand, the computational complexity is quite low and the results are very similar to the well known linear case. On the other hand, the approximation may be very poor and should be used with caution [DS83], [DS85]. Another (more complex) way to determine a confidence region is called *likelihood method* since it is based on the concept of likelihoods [DS85]: All parameter vectors $\tilde{\boldsymbol{\theta}}$ that satisfy

$$\Omega(\tilde{\boldsymbol{\theta}}) - \Omega(\hat{\boldsymbol{\theta}}) \leq s^2\,P\,\mathcal{F}^{1-\alpha}_{P,N-P} \tag{4.113}$$

are included in the likelihood confidence region. This region does not have to be elliptical, but can be of any form. The likelihood method is approximate for nonlinear problems as well, but more precise and robust than the method based on the approximate covariance matrix, which is called *linearisation method* in the following. The likelihood method is more robust because it does not rely on an approximation like the linearisation method, but only takes fitness values into account. There is also an exact method, which is called lack-of-fit method, that is neglected here due to its high computational complexity and because the likelihood method is already a good approximation according to [DS85].

[13] $\boldsymbol{C}_{\hat{\boldsymbol{\theta}},\text{approx}}$ corresponds to $\hat{\boldsymbol{V}}_{\boldsymbol{a}}$ in [DS85] for a complex-valued problem instead of a real-valued problem.

The accuracy of the linearisation and the likelihood method depends on the degree of nonlinearity of the problem at hand. The degree of nonlinearity depends in turn on the underlying structure of the problem and on the current parameters $\boldsymbol{\theta}$. Donaldson and Schnabel suggest in [DS85] to use the curvature measures of Bates and Watts [BW80], which are introduced in Section 4.5.4, as a diagnostic tool: With the help of the curvature measures, it can be evaluated whether the corresponding method is accurate or not.

4.5.3 Proposed Methods to Obtain Soft Information

After this excursion to confidence regions, the way of employing the above knowledge for obtaining soft information is discussed at this point. The first and straightforward idea is to use the variances of the approximate covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{\theta}},\text{approx}}$ in (4.111). This method is simple and many optimisation algorithms like the LM method already compute and output $\boldsymbol{C}_{\hat{\boldsymbol{\theta}},\text{approx}}$ or similar versions of it. But without further analysis (see Section 4.5.4 and Section 5.3.3), it is questionable if this method is precise enough. The second idea is based on likelihood confidence regions. Generally, it is quite complex to generate a likelihood confidence region since many function evaluations have to be performed in the surrounding of the parameter estimates $\hat{\boldsymbol{\theta}}$. However, probabilistic optimisation algorithms like the PSO method perform many function evaluations in the whole search space anyway. Therefore, they are well suited to determine a likelihood confidence region [SBJ+08]. Hence, the drawback of probabilistic algorithms that many function evaluations are required until convergence is transformed into an advantage with respect to likelihood confidence regions. In [SBJ+08], the following procedure is proposed: In every iteration each particle determines its fitness $\Omega(\tilde{\boldsymbol{\theta}})$, which is stored together with the corresponding parameter vector $\tilde{\boldsymbol{\theta}}$ in a table. After the optimum $\hat{\boldsymbol{\theta}}$ with fitness $\Omega(\hat{\boldsymbol{\theta}})$ is found, all parameter vectors $\tilde{\boldsymbol{\theta}}$ that fulfil

$$\Omega(\tilde{\boldsymbol{\theta}}) \leq \Omega(\hat{\boldsymbol{\theta}}) \left(1 + \frac{P}{N-P}\, \mathcal{F}_{P,N-P}^{1-\alpha}\right) \tag{4.114}$$

are selected from the table and form the likelihood confidence region. It can be observed that the density of points near the parameter estimate $\hat{\boldsymbol{\theta}}$ is higher than at the border of the likelihood confidence region. The reason is that the particles are attracted by good fitness values near the optimum and oscillate in its neighbourhood before convergence occurs. Hence, all points $\tilde{\boldsymbol{\theta}}$ form a distribution with mean and variance, where the mean coincides with the parameter estimate $\hat{\boldsymbol{\theta}}$. Therefore, the variance of this distribution can be used as soft information. Both methods proposed above will be applied to (weighted) positioning in Section 5.3.3. Prior to that, both methods are analysed with the help of the curvature measures of Bates and Watts [BW80] in the following section.

4.5.4 Accuracy Analysis of Proposed Methods based on Curvature Measures

In [BW80], Bates and Watts describe nonlinear least-squares estimation from a geometric point of view and introduce measures of nonlinearity. These measures indicate the applicability of a linearisation and its effects on inference. Hence, the accuracy of the soft information methods described above can be evaluated using these measures. In the following, the most important aspects of these so-called curvature measures are presented. First, the nonlinear least-squares problem is reviewed: A set of parameters

$$\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_P]^T$$

shall be estimated from a set of observations

$$\check{\boldsymbol{h}} = \left[\check{h}_0, \check{h}_1, ..., \check{h}_L\right]^T$$

with

$$\check{\boldsymbol{h}} = \boldsymbol{h}(\boldsymbol{\theta}) + \boldsymbol{\eta},$$

where $\boldsymbol{h}(\boldsymbol{\theta})$ is a nonlinear function of the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ is zero mean noise with covariance matrix $\boldsymbol{C}_{\boldsymbol{\eta}}$. The weighted least-squares estimate is given by the value $\hat{\boldsymbol{\theta}}$ that minimises the weighted sum of squared residuals

$$\Omega(\tilde{\boldsymbol{\theta}}) = \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right)^H \boldsymbol{C}_{\boldsymbol{\eta}}^{-1} \left(\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right), \tag{4.115}$$

which corresponds to the metric of the maximum-likelihood estimator in case of Gaussian noise. In case of white noise, the sum of squares can be rewritten as

$$\Omega(\tilde{\boldsymbol{\theta}}) = \left|\left|\check{\boldsymbol{h}} - \boldsymbol{h}(\tilde{\boldsymbol{\theta}})\right|\right|^2. \tag{4.116}$$

Geometrically, the sum of squares in (4.115) and (4.116) describe the distance between $\check{\boldsymbol{h}}$ and $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ in the $(L+1)$-dimensional observation space. Similarly, the relation between the metrics $\Omega_{\boldsymbol{y}}(\tilde{\boldsymbol{\theta}})$ and $\Omega_{\check{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}})$ has already been explained from a geometrical point of view in Section 4.2.3. Although the required geometrical fundamentals have already been introduced in Section 4.2.3, they are repeated at this point again: If a parameter vector $\tilde{\boldsymbol{\theta}}$ is changed in the P-dimensional *search space/parameter space*, the vector $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ traces a P-dimensional subspace in the $(L+1)$-dimensional *sample space/observation space*. This subspace is called *solution locus*. Hence, the function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ maps all feasible parameters in the parameter space to the solution locus in the observation space. Because of the noise $\boldsymbol{\eta}$, the observations do not lie on the solution locus, but anywhere in the observation space. The parameter estimate $\hat{\boldsymbol{\theta}}$ corresponds to the point on the solution locus $\boldsymbol{h}(\hat{\boldsymbol{\theta}})$ with the smallest distance to the point of observations $\check{\boldsymbol{h}}$. Since the function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ is nonlinear, the solution locus will be curved. For inference, the solution locus is often approximated by a tangent plane with a uniform coordinate system. The tangent plane at a specific point $\boldsymbol{h}(\tilde{\boldsymbol{\theta}}_0)$ can be described by a first order Taylor series

$$\boldsymbol{h}(\boldsymbol{\theta}) \cong \boldsymbol{h}(\tilde{\boldsymbol{\theta}}_0) + \boldsymbol{J}(\tilde{\boldsymbol{\theta}}_0) \left(\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_0\right), \tag{4.117}$$

where $\boldsymbol{J}(\tilde{\boldsymbol{\theta}}_0)$ is the Jacobian matrix of the model function $\boldsymbol{h}(\boldsymbol{\theta})$ as defined in (4.36) evaluated at $\tilde{\boldsymbol{\theta}}_0$. The informational value of inference concerning the parameter estimates highly depends on the closeness of the tangent plane to the solution locus. This closeness in turn depends on the curvature of the solution locus. Therefore, the measures of nonlinearity proposed by Bates and Watts indicate the maximum curvature of the solution locus at a specific point $\boldsymbol{h}(\tilde{\boldsymbol{\theta}}_0)$. It is important to note that there are two different kinds of curvature since two different assumptions are made concerning the tangent plane. First, it is assumed that the solution locus is planar at $\boldsymbol{h}(\tilde{\boldsymbol{\theta}}_0)$ and, hence, can be replaced by a tangent plane (*planar assumption*). Second, it is assumed that the coordinate system on the tangent plane is uniform (*uniform coordinate assumption*), i.e. the coordinate grid lines mapped from the parameter space remain equidistant and straight in the observation space. It may happen that the first assumption is fulfilled, but the second assumption is not. Then, the solution locus is planar at a specific point $\boldsymbol{h}(\tilde{\boldsymbol{\theta}}_0)$, but the coordinate grid lines are curved and not equidistant. If the planar assumption is not fulfilled, the uniform coordinate assumption is not fulfilled either.

In order to determine curvatures, Bates and Watts introduce so-called *lifted lines* [BW80]. Similar to the fact that each point $\tilde{\boldsymbol{\theta}}_0$ in the parameter space maps to a point $\boldsymbol{h}(\tilde{\boldsymbol{\theta}}_0)$ on the solution locus in the observation space, each straight line in the parameter space through $\tilde{\boldsymbol{\theta}}_0$,

$$\tilde{\boldsymbol{\theta}}(m) = \tilde{\boldsymbol{\theta}}_0 + m\boldsymbol{\vartheta}, \tag{4.118}$$

maps to a lifted line on the solution locus

$$\boldsymbol{h}_{\boldsymbol{\vartheta}}(m) = \boldsymbol{h}(\tilde{\boldsymbol{\theta}}_0 + m\boldsymbol{\vartheta}), \tag{4.119}$$

where $\boldsymbol{\vartheta}$ can be any non-zero vector in the parameter space. The tangent vector of the lifted line for m=0 at $\tilde{\boldsymbol{\theta}}_0$ is given by

$$\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}} = \left.\frac{d\boldsymbol{h}_{\boldsymbol{\vartheta}}(m)}{dm}\right|_0 = \left.\frac{d\boldsymbol{h}(\tilde{\boldsymbol{\theta}})}{d\tilde{\boldsymbol{\theta}}}\right|_{\tilde{\boldsymbol{\theta}}_0} \left.\frac{d\tilde{\boldsymbol{\theta}}(m)}{dm}\right|_0 = \boldsymbol{J}(\tilde{\boldsymbol{\theta}}_0)\,\boldsymbol{\vartheta}. \tag{4.120}$$

The set of all tangent vectors (for all possible vectors $\boldsymbol{\vartheta}$) forms the tangent plane. For measuring curvatures, second order derivatives are needed additionally. The second order derivative of the function $\boldsymbol{h}(\tilde{\boldsymbol{\theta}})$ is the Hessian given by

$$\left[\boldsymbol{H}(\tilde{\boldsymbol{\theta}})\right]_{lij} = \frac{\delta^2 h_l(\tilde{\boldsymbol{\theta}})}{\delta\tilde{\theta}_i \delta\tilde{\theta}_j}, \tag{4.121}$$

which is a three-dimensional tensor. The lth face of the Hessian is, thus, a $P\times P$ matrix:

$$\boldsymbol{H}_l(\tilde{\boldsymbol{\theta}}) = \begin{bmatrix} \frac{\delta^2 h_l(\tilde{\boldsymbol{\theta}})}{\delta\tilde{\theta}_1\delta\tilde{\theta}_1} & \cdots & \frac{\delta^2 h_l(\tilde{\boldsymbol{\theta}})}{\delta\tilde{\theta}_P\delta\tilde{\theta}_1} \\ \vdots & \ddots & \vdots \\ \frac{\delta^2 h_l(\tilde{\boldsymbol{\theta}})}{\delta\tilde{\theta}_1\delta\tilde{\theta}_P} & \cdots & \frac{\delta^2 h_l(\tilde{\boldsymbol{\theta}})}{\delta\tilde{\theta}_P\delta\tilde{\theta}_P} \end{bmatrix}. \tag{4.122}$$

The second order derivative of the lifted line is given by

$$\ddot{\boldsymbol{h}}_{\boldsymbol{\vartheta}} = \left.\frac{d^2\boldsymbol{h}_{\boldsymbol{\vartheta}}(m)}{dm^2}\right|_0 = \boldsymbol{\vartheta}^T\,\boldsymbol{H}(\tilde{\boldsymbol{\theta}}_0)\,\boldsymbol{\vartheta}, \tag{4.123}$$

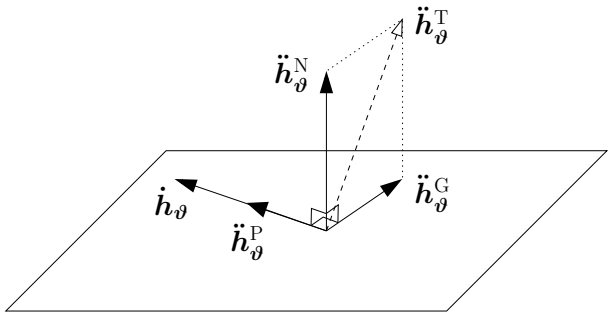


Figure 4.44: Example for the decomposition of the acceleration vector $\ddot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}$ with respect to the velocity vector $\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}$.

in which the tensor product is performed such that

$$\left[\ddot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}\right]_l = \boldsymbol{\vartheta}^T \boldsymbol{H}_l(\tilde{\boldsymbol{\theta}}_0)\, \boldsymbol{\vartheta}. \tag{4.124}$$

The derivatives of the lifted line, $\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}$ and $\ddot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}$, can be interpreted physically [BW80]: If a point moves along the lifted line $\boldsymbol{h}_{\boldsymbol{\vartheta}}(m)$ in the observation space, where m denotes the *time*, then $\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}$ and $\ddot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}$ denote the instantaneous *velocity* and instantaneous *acceleration* at time m=0, respectively. The acceleration can be decomposed into three parts

$$\ddot{\boldsymbol{h}}_{\boldsymbol{\vartheta}} = \ddot{\boldsymbol{h}}^{\mathrm{P}}_{\boldsymbol{\vartheta}} + \ddot{\boldsymbol{h}}^{\mathrm{N}}_{\boldsymbol{\vartheta}} + \ddot{\boldsymbol{h}}^{\mathrm{G}}_{\boldsymbol{\vartheta}} \tag{4.125}$$

as shown in Figure 4.44. $\ddot{\boldsymbol{h}}^{\mathrm{P}}_{\boldsymbol{\vartheta}}$ is parallel to the velocity vector $\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}$ and, thus, parallel to the tangent plane. It corresponds to the *change in velocity* of the moving point. $\ddot{\boldsymbol{h}}^{\mathrm{N}}_{\boldsymbol{\vartheta}}$ is normal to the tangent plane and describes the *change in direction* of the velocity vector $\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}$ normal to the tangent plane. $\ddot{\boldsymbol{h}}^{\mathrm{G}}_{\boldsymbol{\vartheta}}$ is parallel to the tangent plane and normal to the velocity vector $\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}$. It corresponds to the geodesic acceleration and indicates the *change in direction* of the velocity vector $\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}$ parallel to the tangent plane. Based on these acceleration components, the curvatures of the solution locus at $\tilde{\boldsymbol{\theta}}_0$ can be determined:

$$K^{\mathrm{N}}_{\boldsymbol{\vartheta}} = \frac{||\ddot{\boldsymbol{h}}^{\mathrm{N}}_{\boldsymbol{\vartheta}}||}{||\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}||^2} \tag{4.126}$$

is the normal curvature in direction of $\boldsymbol{\vartheta}$ and is called *intrinsic curvature* and

$$K^{\mathrm{T}}_{\boldsymbol{\vartheta}} = \frac{||\ddot{\boldsymbol{h}}^{\mathrm{P}}_{\boldsymbol{\vartheta}} + \ddot{\boldsymbol{h}}^{\mathrm{G}}_{\boldsymbol{\vartheta}}||}{||\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}||^2} = \frac{||\ddot{\boldsymbol{h}}^{\mathrm{T}}_{\boldsymbol{\vartheta}}||}{||\dot{\boldsymbol{h}}_{\boldsymbol{\vartheta}}||^2} \tag{4.127}$$

is the tangential[14] curvature in direction of $\boldsymbol{\vartheta}$ and is called *parameter-effects curvature*. The curvatures are divided into normal and tangential components since each component has a different influence on the accuracy of the linear approximation. On the one hand,

[14]The superscript $^{\mathrm{T}}$, which denotes *tangential*, should not be mistaken for the superscript ^{T}, which denotes the *transpose* of a matrix.

the intrinsic curvature is an intrinsic property of the solution locus. It only affects the planar assumption. On the other hand, the parameter-effects curvature only influences the uniform coordinate assumption and depends on the specific parameterisation of the problem. Hence, a reparameterisation may change the parameter-effects curvature, but not the intrinsic curvature. In order to assess the effect of the curvatures on inference, they should be normalised. A suitable scaling factor is the so-called *standard radius* $\rho{=}s\sqrt{P}$ since its square $\rho^2{=}s^2P$ appears on the right hand side in (4.112) and (4.113), which describe the confidence regions. The *relative curvatures* are given by the curvatures (4.126) and (4.127) multiplied with the standard radius:

$$\gamma_{\boldsymbol{\vartheta}}^{\mathrm{N}} = K_{\boldsymbol{\vartheta}}^{\mathrm{N}}\,\rho, \tag{4.128}$$

$$\gamma_{\boldsymbol{\vartheta}}^{\mathrm{T}} = K_{\boldsymbol{\vartheta}}^{\mathrm{T}}\,\rho. \tag{4.129}$$

If the relative curvatures are small compared to

$$\mathcal{C}_{P,N-P}^{1-\alpha} = \frac{1}{\sqrt{\mathcal{F}_{P,N-P}^{1-\alpha}}} \tag{4.130}$$

for all possible directions $\boldsymbol{\vartheta}$, then the corresponding assumptions are valid [BW80]. Hence, it is sufficient to determine the *maximum relative curvatures*[15]

$$\Gamma^{\mathrm{N}} = \max_{\boldsymbol{\vartheta}} \left\{ \gamma_{\boldsymbol{\vartheta}}^{\mathrm{N}} \right\}, \tag{4.131}$$

$$\Gamma^{\mathrm{T}} = \max_{\boldsymbol{\vartheta}} \left\{ \gamma_{\boldsymbol{\vartheta}}^{\mathrm{T}} \right\} \tag{4.132}$$

and to compare them to $\mathcal{C}_{P,N-P}^{1-\alpha}$ in order to assess the accuracy of the soft information based on different confidence regions [DS85]. If the linearisation method based on (4.112) with the approximate covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{\theta}},\mathrm{approx}}$ shall be applied, both, the planar assumption *and* the uniform coordinate assumption, have to be fulfilled. That means that the maximum relative curvatures Γ^{N} *and* Γ^{T} have to be small compared to $\mathcal{C}_{P,N-P}^{1-\alpha}$. The likelihood method based on (4.113) is more robust since only the planar assumption needs to be fulfilled and only Γ^{N} needs to be small compared to $\mathcal{C}_{P,N-P}^{1-\alpha}$.

In the following, the channel parameter estimation problem is analysed by calculating the maximum relative curvatures and by plotting the confidence regions (4.112) and (4.113) for different SNRs. The same system setup as already applied in Section 4.3.3 and Section 4.4.3 is used. For the purpose of curvature analysis, only the PSO method is applied in combination with the stopping criterion based on the fitness improvement with $T{=}10^{-7}$ and $N_{\min}{=}50$. A very tight stopping criterion is considered at this point in order to ensure that the algorithm is not stopped too early. The PSO method automatically delivers the likelihood confidence region as explained in Section 4.5.3. The approximate covariance matrix is calculated afterwards according to (4.111). A confidence level of 95%

[15] In [BW80], a simplified method to determine the maximum relative curvatures is introduced based on linear transformations of the coordinates in the parameter and the observation space. This method is neglected here because it is out of the scope of this thesis.

	LOS channel	Two-path channel *Setup 1*	Two-path channel *Setup 2*	Two-path channel *Setup 3*
Parameters	θ_1=−0.9565 θ_2= 0.2917 θ_3= 0.3162 T_s	θ_1=−0.6764 θ_2= 0.2062 θ_3= 0.3162 T_s θ_4=−0.6764 θ_5= 0.2062 θ_6= 0.8162 T_s	θ_1=−0.8732 θ_2= 0.2662 θ_3= 0.3162 T_s θ_4=−0.1191 θ_5=−0.3905 θ_6= 1.8162 T_s	θ_1=−0.9089 θ_2= 0.2771 θ_3= 0.3162 T_s θ_4=−0.2582 θ_5=−0.1744 θ_6= 1.1262 T_s
P=3C	3	6	6	6
N=L+1	19	23	23	23
$\mathcal{C}^{0.95}_{P,N-P}$	**0.55565**	**0.60873**	**0.60873**	**0.60873**
Γ^{N}@10 dB	0.53568	14.93453	2.42187	8.38223
Γ^{T}@10 dB	0.43027	69.84693	5.87543	60.80338
Γ^{N}@30 dB	0.04070	1.35313	0.20485	0.66779
Γ^{T}@30 dB	0.03155	4.02935	0.48961	4.87241
Γ^{N}@50 dB	0.00523	0.16321	0.02528	0.10790
Γ^{T}@50 dB	0.00370	0.43578	0.05963	0.80799

Table 4.3: Parameters of the investigated channel models and the corresponding maximum relative curvatures at different SNRs.

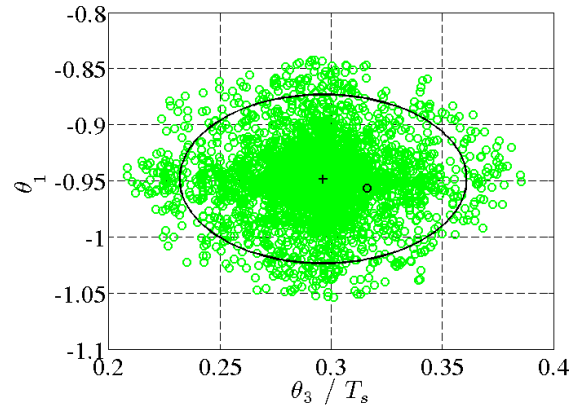
is applied (α=0.05). Since the curvature measures depend on the parameter set $\boldsymbol{\theta}$ and also on the noise samples, simulations are performed for several *fixed* channel models at different SNRs. A realisation of the LOS channel and three different realisations of the two-path channel are considered. Concerning the two-path channel, the two setups summarised in Table 4.1 are used, where the first setup corresponds to a worst case scenario and the second setup represents a best case scenario. Additionally, a third setup with ν_2/T_s=0.81, $\mathcal{P}$=9.3 and $\Delta\Phi$=51° is investigated. For each channel realisation, the sampling phase is fixed to ε=−0.3162 T_s and the phase of the LOS path is fixed to Φ_1=163.0425°. The corresponding values of the parameter vector $\boldsymbol{\theta}$ are summarised in Table 4.3. Furthermore, the number of parameters P, the number of observations N as well as the reference value $\mathcal{C}^{0.95}_{P,N-P}$ and the maximum relative curvatures Γ^{N} and Γ^{T} for different SNRs are listed in Table 4.3. Curvatures that are larger than the reference value $\mathcal{C}^{0.95}_{P,N-P}$ are marked in red colour. Generally, the maximum relative curvatures decrease with SNR. Hence, it is more likely that the soft information methods are accurate at high SNR. For the LOS channel, the maximum relative curvatures are always smaller than the reference value $\mathcal{C}^{0.95}_{P,N-P}$, i.e. both soft information methods are accurate in case of the LOS channel. In contrast, both soft information methods are inaccurate at 10 dB for the two-path channels according to Table 4.3. Especially for the first setup (worst case), the maximum relative curvatures Γ^{N} and Γ^{T} are much larger than $\mathcal{C}^{0.95}_{P,N-P}$. For the second setup (best case), the soft information methods become accurate at 30 dB, while the methods are not accurate for the remaining setups until an SNR of 50 dB. For the third setup, the linearisation method is not even accurate at 50 dB because the parameter-effects curvature Γ^{T} is larger than

the reference value $\mathcal{C}^{0.95}_{P,N-P}$. In this case, the advantage of the likelihood method, that *only* the intrinsic curvature Γ^{N} needs to smaller than $\mathcal{C}^{0.95}_{P,N-P}$, becomes clear. Generally, the intrinsic curvatures Γ^{N} are considerably smaller than the parameter-effect curvatures Γ^{T} for multipath propagation. Hence, it is more probable that the intrinsic curvature Γ^{N} is smaller than the reference value $\mathcal{C}^{0.95}_{P,N-P}$ such that the planar assumption is fulfilled. Therefore, the probability that the likelihood method is accurate is significantly higher than the probability that the linearisation method is accurate, for which the planar assumption *and* the uniform-coordinate assumption must be fulfilled (i.e. both curvatures must be smaller than $\mathcal{C}^{0.95}_{P,N-P}$). Consequently, the likelihood method is more robust. Thus, it is expected that positioning algorithms, that exploit soft information, perform better in combination with the likelihood method than in conjunction with the linearisation method. Furthermore, an improved performance is expected at high SNR because the accuracy of both methods increases with SNR. Corresponding results are presented in Section 5.3.3.

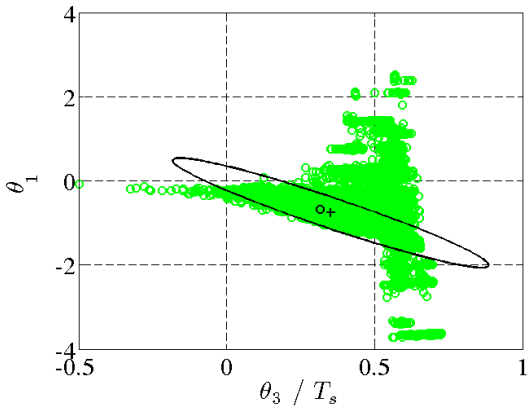
The conclusions drawn from Table 4.3 are confirmed by Figure 4.45 and Figure 4.46, in which the corresponding confidence regions based on the linearisation method (black ellipse) and the likelihood method (green circles) are plotted for all four channel realisations at different SNRs. Additionally, the estimated and the true parameters are marked in both figures: The estimated parameters $\hat{\boldsymbol{\theta}}$ are denoted by a black cross and the true parameters $\boldsymbol{\theta}$ by a black circle. In all cases, a parameter combination of the real part $\theta_1 = \mathrm{Re}\{f_1\}$ and the coarse excess delay $\theta_3 = \varrho_1 = -\varepsilon$ of the LOS path is considered. The coarse excess delay is normalised to the symbol duration T_s. According to Table 4.3, the linearisation as well as the likelihood method are generally accurate in case of the LOS channel. This is confirmed by the corresponding confidence regions shown on the left hand side of Figure 4.45: For all SNRs, the confidence region based on the linearisation method (black ellipse) and the likelihood method (green circles) are basically the same. The likelihood regions are slightly larger, but have the same elliptical shape as the confidence regions based on the linearisation method. For the two-path channels, the shape of the different confidence regions do not coincide in most cases. Especially for the first and third setup, the likelihood regions deviate severely from an elliptical shape as shown on the right hand side of Figure 4.45 and Figure 4.46, respectively. With increasing SNR or better channel conditions, the shape of the likelihood regions nearly approaches an elliptical shape. However, the linearisation ellipses do not necessarily match the likelihood regions in that case because it is less likely that the uniform coordinate assumption ($\Gamma^{\mathrm{T}} < \mathcal{C}^{0.95}_{P,N-P}$) is fulfilled in addition to the planar assumption ($\Gamma^{\mathrm{N}} < \mathcal{C}^{0.95}_{P,N-P}$). This means that the likelihood regions, that appear to be deformed with respect to the linearisation ellipses, are supposed to be more accurate.

4.6 Chapter Summary

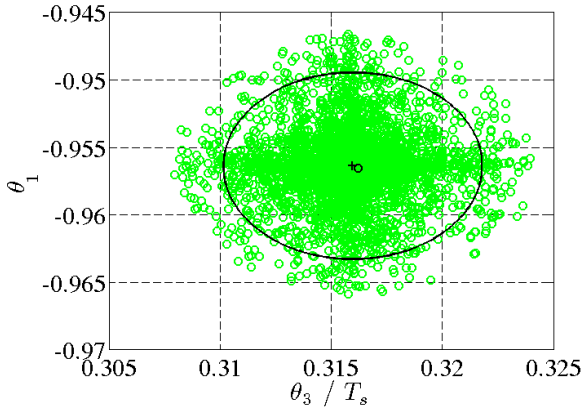
In this chapter, channel parameter estimation is investigated systematically. In the beginning, two ML estimators, one based on the received samples $\boldsymbol{y}$ and the other one based



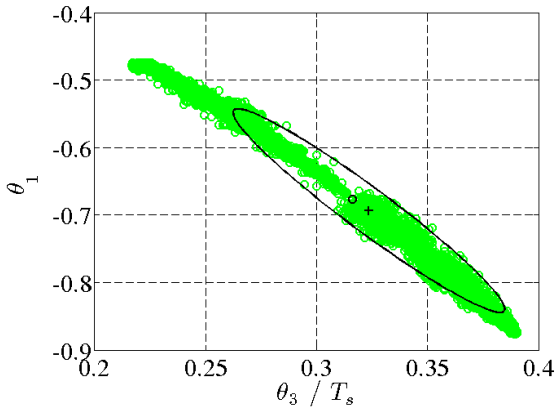
(a) LOS channel at 10 dB.



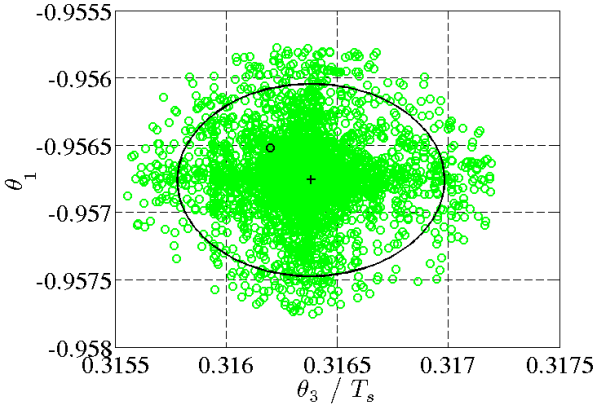
(b) Two-path channel with setup 1 at 10 dB.



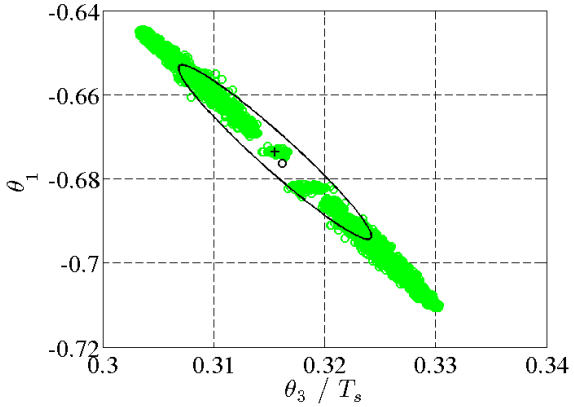
(c) LOS channel at 30 dB.



(d) Two-path channel with setup 1 at 30 dB.

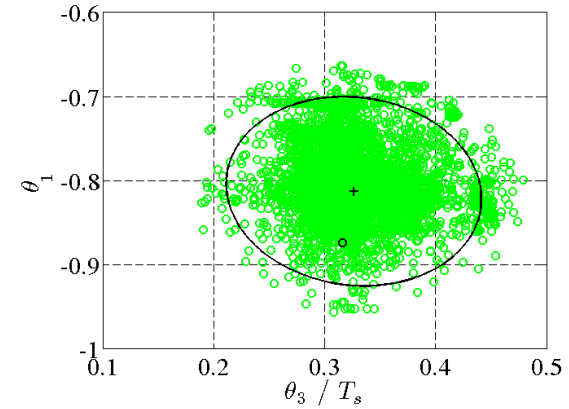


(e) LOS channel at 50 dB.

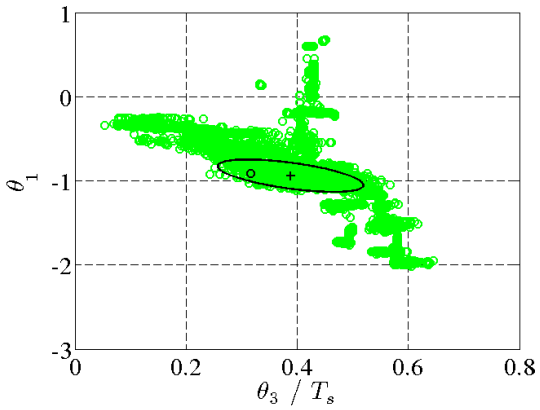


(f) Two-path channel with setup 1 at 50 dB.

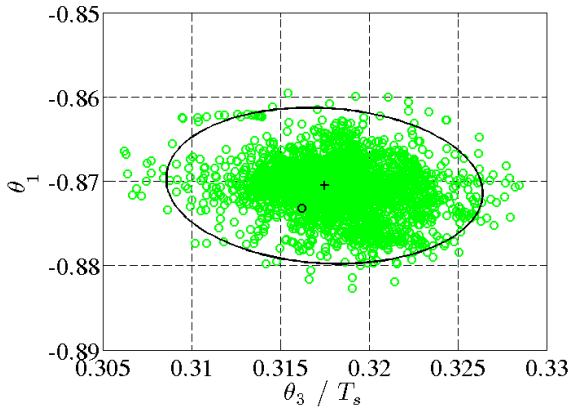
Figure 4.45: Confidence regions based on the linearisation method (black ellipse) and the likelihood method (green circles) for the LOS channel and the first setup of the two-path channel at different SNRs. The estimated parameters $\hat{\boldsymbol{\theta}}$ are denoted by a black cross and the true parameters $\boldsymbol{\theta}$ by a black circle.



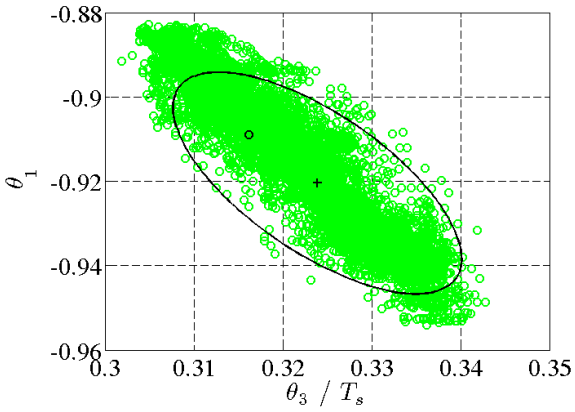
(a) Two-path channel with setup 2 at 10 dB.



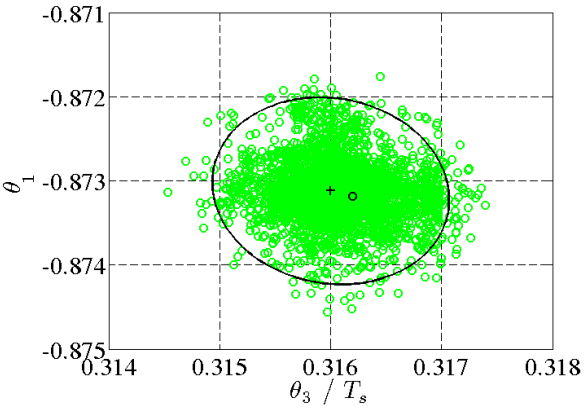
(b) Two-path channel with setup 3 at 10 dB.



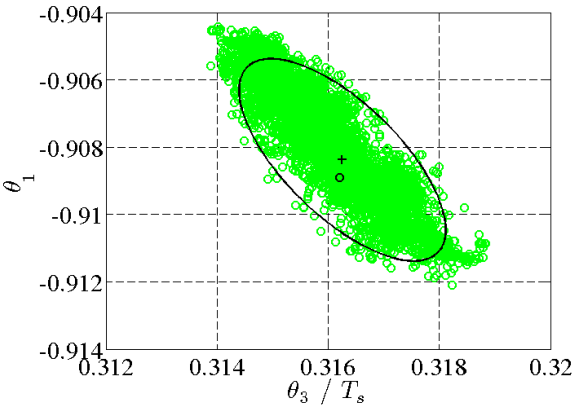
(c) Two-path channel with setup 2 at 30 dB.



(d) Two-path channel with setup 3 at 30 dB.



(e) Two-path channel with setup 2 at 50 dB.



(f) Two-path channel with setup 3 at 50 dB.

Figure 4.46: Confidence regions based on the linearisation method (black ellipse) and the likelihood method (green circles) for the second and third setup of the two-path channel at different SNRs. The estimated parameters $\hat{\boldsymbol{\theta}}$ are denoted by a black cross and the true parameters $\boldsymbol{\theta}$ by a black circle.

on pre-stage channel estimates $\check{\boldsymbol{h}}$, are derived and it is shown that both estimators are equivalent. Since the first estimator is disadvantageous from a complexity point of view, only the second estimator is considered subsequently. Performance limits for channel parameter estimation in terms of CRLBs are derived and it is observed that oversampling provides a performance gain compared to symbol-rate sampling. Especially for the WINNER channel models, oversampling proves necessary for accurate channel parameter estimation. Hence, it is suggested to apply an oversampling factor of J=2 for channel parameter estimation. Since the metric $\Omega(\tilde{\boldsymbol{\theta}})$ of the ML estimator is nonlinear, optimisation algorithms are required to find a parameter estimate $\hat{\boldsymbol{\theta}}$ that minimises this metric and, hence, maximises the likelihood function $p(\check{\boldsymbol{h}}, \tilde{\boldsymbol{\theta}})$. Several local and global optimisation methods, namely the LM, EM and SAGE method as well as the PSO and CPSO method, are introduced and their performance is investigated. Initially, the performance is analysed with the help of boxplots. Based on these results, it is concluded that the PSO method is best suited for solving the metric since it is most robust with respect to outliers even though is suffers from premature convergence at high SNR such that an error floor occurs. The remaining methods perform worse. On the one hand, the (unconstrained) local methods may converge to local optima outside the search space, that are far away from the global optimum and, thus, lead to extreme outliers. Especially, the LM method suffers from this effect. On the other hand, the CPSO methods suffer not only from premature convergence, but also from stagnation. Since a rather large outlier range is present for all methods, it is concluded that an outlier detection is generally desirable. Therefore, an outlier detection method based on the final fitness $\Omega(\hat{\boldsymbol{\theta}})$ is proposed and successfully applied to channel parameter estimation. All methods benefit from this outlier detection, but for the (unconstrained) local methods an additional boundary check is necessary to ensure a reliable outlier detection at low SNR. Very good results are obtained for all methods in conjunction with outlier detection. For the two-path channel with large excess delay ($\nu_2/T_s \in [1, 2]$), the MSEs with respect to the sampling phase $\hat{\varepsilon}$ generally correspond to the CRLBs, i.e. the estimator is optimal in this case. For the two-path channel with small excess delay ($\nu_2/T_s \in [0.05, 1]$), MSEs even below the CRLB occur at very low and very high SNRs. The deviation at low SNR is caused by a bias, while the deviation at high SNR is caused by large outage rates. With the help of fitness histories, it is shown that the large outage rates at high SNR are caused by a too early interruption of the optimisation methods. Hence, other stopping criteria are examined: Stopping criteria based on the absolute fitness (in conjunction with the fitness threshold T_f also applied for outlier detection) outperform stopping criteria based on the fitness improvement in the sense that the number of performed iterations better matches the number of required iterations derived from the fitness histories. It is shown that the best tradeoff between the MSE performance and the outage rates as well as the performed number of iterations is enabled by a sophisticated combination of the PSO and the LM method. In case of the WINNER channel models, it becomes advantageous with increasing symbol durations T_s to combine multipath components with similar delays in order to reduce the number of parameters. Generally, a reduced estimator with three multipath components is a good choice because the true number of propagations paths is typically not known and symbol durations $T_s \geq 100$ ns are often applied. Finally, soft information for channel parameter

estimation is discussed. Two different methods to obtain soft information concerning the parameter estimates, namely the linearisation and the likelihood method, are proposed. Both methods are analysed with the help of the curvature measures of Bates and Watts. According to the obtained results, the likelihood method is supposed to outperform the linearisation method because it is more robust, i.e. it is more probable that the likelihood method is accurate. Positioning results incorporating soft channel parameter estimation will be presented in Section 5.3.3.

Chapter 5

Position Estimation

As already explained in Section 3.1, positioning is typically performed in two steps, namely parameter estimation and position estimation. In the preceding chapter, parameter estimation has been examined systematically. Now, the second localisation step is investigated in more detail. Hence, it is not sufficient anymore to consider a single transmission link between a transmitter and a receiver: *Several* links between different reference objects (ROs) and a mobile station (MS) in a certain geometrical setup are required for position estimation. In the following, two-dimensional positioning based on the TOA is considered such that *at least* three ROs are needed. An extension to three-dimensional localisation is straight forward.

5.1 Fundamentals

The MS's position, that shall be estimated from the TOAs, is denoted by

$$\boldsymbol{p} = [p_1, p_2]^T = [x, y]^T, \tag{5.1}$$

while the known locations of the ROs are denoted by

$$\boldsymbol{p}_b = [p_{b,1}, p_{b,2}]^T = [x_b, y_b]^T, \qquad 1 \leq b \leq B, \tag{5.2}$$

where B is the number of reference objects. The true distance between the MS and the bth RO is a nonlinear function of the current MS's position $\boldsymbol{p}$ and can be determined according to

$$d_b(\boldsymbol{p}) = \sqrt{(x - x_b)^2 + (y - y_b)^2}. \tag{5.3}$$

These distances are estimated via the TOAs $\hat{\tau}_{b,1}$ based on

$$\hat{d}_b = \hat{\tau}_{b,1} \cdot \mathrm{c}, \tag{5.4}$$

where c is the speed of light. The estimated distances $\hat{\boldsymbol{d}}=[\hat{d}_1,\cdots,\hat{d}_B]^T$ are called *pseudo-ranges* since they consist of the true distances $\boldsymbol{d}(\boldsymbol{p})=[d_1(\boldsymbol{p}),\cdots,d_B(\boldsymbol{p})]^T$ and estimation errors $\boldsymbol{e}=[e_1,\cdots,e_B]^T$ with covariance matrix $\boldsymbol{C}_{\boldsymbol{e}}=\operatorname{diag}\left(\sigma_{e_1}^2,\cdots,\sigma_{e_B}^2\right)$:

$$\hat{\boldsymbol{d}} = \boldsymbol{d}(\boldsymbol{p}) + \boldsymbol{e}. \tag{5.5}$$

Again, a ML estimator can be applied to estimate the position of the MS:

$$\hat{\boldsymbol{p}} = \arg\min_{\tilde{\boldsymbol{p}}} \left\{ \left(\hat{\boldsymbol{d}} - \boldsymbol{d}(\tilde{\boldsymbol{p}})\right)^T \boldsymbol{C}_{\boldsymbol{e}}^{-1} \left(\hat{\boldsymbol{d}} - \boldsymbol{d}(\tilde{\boldsymbol{p}})\right) \right\} = \arg\min_{\tilde{\boldsymbol{p}}} \left\{\Omega_{\hat{\boldsymbol{d}}}(\tilde{\boldsymbol{p}})\right\} \tag{5.6}$$

Similar to the ML estimator for channel parameter estimation, the metric $\Omega_{\hat{\boldsymbol{d}}}(\tilde{\boldsymbol{p}})$ for position estimation is nonlinear due to the nonlinear distance function (5.3). Hence, all optimisation algorithms presented in Section 4.4 can be applied to localisation as well. Often, a Gauss-Newton approach is utilised for positioning. In this context, the Gauss-Newton algorithm is also known as *Taylor series (TS)* algorithm [Kap96, Chap. 2], [SZT08]. For positioning, there are additionally some approximative, non-iterative estimators like the *weighted least-squares (WLS)* algorithm available [STK05], [GCWI08]. Before discussing potential positioning algorithms, performance limits on position estimation are introduced in the following section.

5.2 Performance Limits: Cramer-Rao Lower Bound and Geometric Dilution of Precision

As in the case of channel parameter estimation, CRLBs can be determined for positioning, that correspond to the best performance that any unbiased estimator can achieve (see Section 4.3). This means that the covariance matrix of the position estimator $\boldsymbol{C}_{\hat{\boldsymbol{p}}}$ is greater than or equal to the inverse of the Fisher information matrix $\boldsymbol{I}^{-1}(\boldsymbol{p})$ [Sch91, Sec. 6.4], [Kay93, Chap. 4]:

$$\boldsymbol{C}_{\hat{\boldsymbol{p}}} - \boldsymbol{I}^{-1}(\boldsymbol{p}) \geq \boldsymbol{0} \tag{5.7}$$

i.e. the matrix $\boldsymbol{C}_{\hat{\boldsymbol{p}}}-\boldsymbol{I}^{-1}(\boldsymbol{p})$ is positive semidefinite. The CRLB for positioning is given by the trace of the inverse Fisher information matrix

$$\mathrm{CRLB}\left(\boldsymbol{p}\right) = \operatorname{tr}\left\{\boldsymbol{I}^{-1}(\boldsymbol{p})\right\} = \sum_{i=1}^{2} \left[\boldsymbol{I}^{-1}(\boldsymbol{p})\right]_{i,i}. \tag{5.8}$$

For unbiased estimators, the MSE of the position estimate is always larger than or equal to the CRLB in (5.8):

$$\mathrm{MSE}\left(\hat{\boldsymbol{p}}\right) \geq \mathrm{CRLB}\left(\boldsymbol{p}\right). \tag{5.9}$$

In order to calculate the CRLB, the Fisher information matrix $\boldsymbol{I}(\boldsymbol{p})$ needs to be determined. According to [Sch91, p. 223] and [Kay93, p. 44], a single entry of the Fisher information matrix is defined as

$$\left[\boldsymbol{I}(\boldsymbol{p})\right]_{i,j} = -\,\mathrm{E}\left\{\frac{\partial^2 \ln p(\hat{\boldsymbol{d}},\boldsymbol{p})}{\partial p_i \partial p_j}\right\} \tag{5.10}$$

Similar to the derivations in Section 4.3, the following expression is obtained for the Fisher information matrix:

$$\boldsymbol{I}(\boldsymbol{p}) = 2 \cdot \boldsymbol{J}^T(\boldsymbol{p})\, \boldsymbol{C}_{\boldsymbol{e}}^{-1} \boldsymbol{J}(\boldsymbol{p}), \tag{5.11}$$

where $\boldsymbol{J}(\boldsymbol{p})$ denotes the Jacobian matrix of the distance function $\boldsymbol{d}(\boldsymbol{p})$ given by

$$\boldsymbol{J}(\boldsymbol{p}) = \frac{\partial \boldsymbol{d}(\boldsymbol{p})}{\partial \boldsymbol{p}^T} = \left[\frac{\partial \boldsymbol{d}(\boldsymbol{p})}{\partial x}, \frac{\partial \boldsymbol{d}(\boldsymbol{p})}{\partial y}\right] = \begin{bmatrix} \frac{\partial d_1(\boldsymbol{p})}{\partial x} & \frac{\partial d_1(\boldsymbol{p})}{\partial y} \\ \vdots & \vdots \\ \frac{\partial d_B(\boldsymbol{p})}{\partial x} & \frac{\partial d_B(\boldsymbol{p})}{\partial y} \end{bmatrix}, \tag{5.12}$$

with

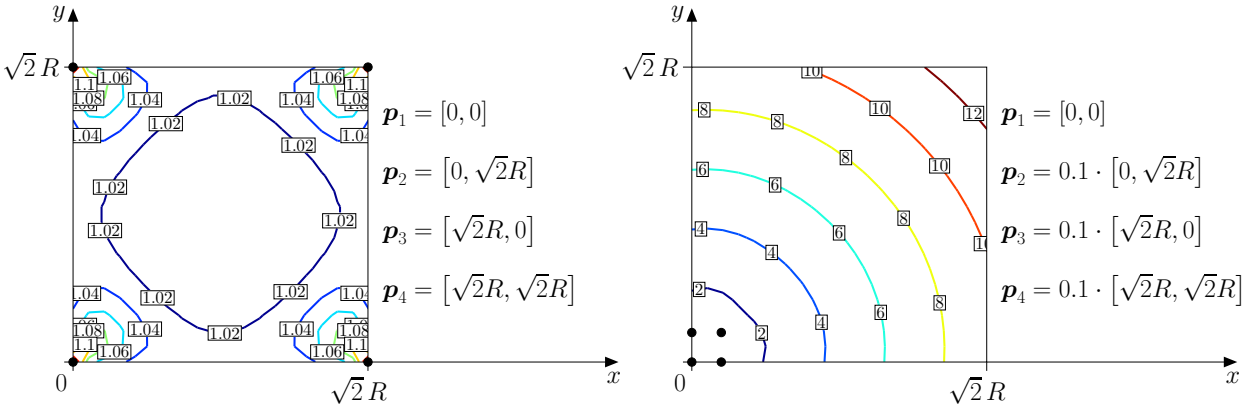
$$\frac{\partial d_b(\boldsymbol{p})}{\partial x} = \frac{x - x_b}{\sqrt{(x - x_b)^2 + (y - y_b)^2}} \tag{5.13}$$

and

$$\frac{\partial d_b(\boldsymbol{p})}{\partial y} = \frac{y - y_b}{\sqrt{(x - x_b)^2 + (y - y_b)^2}}. \tag{5.14}$$

The positions of all involved ROs are required in order to determine the Fisher information matrix. Hence, the positioning accuracy depends on the geometry between the ROs and the MS and, thus, varies with the MS's position $\boldsymbol{p}$. This effect is called *geometric dilution of precision (GDOP)* [Kap96, Chap. 7], [Lan99]. In order to separate the influence of the geometry from the influence of the estimation errors $\boldsymbol{e}$ on the positioning accuracy, it is assumed that all pseudoranges are affected by the same error variance σ_e^2=1, i.e. $\boldsymbol{C}_{\boldsymbol{e}}$=$\boldsymbol{\mathcal{I}}$. Based on this assumption, the GDOP is given as the square root of the CRLB:

$$\mathrm{GDOP}\,(\boldsymbol{p}) = \sqrt{\mathrm{CRLB}\,(\boldsymbol{p})\,\big|_{\boldsymbol{C}_{\boldsymbol{e}}=\boldsymbol{\mathcal{I}}}} = \sqrt{\frac{1}{2}\,\mathrm{tr}\left\{\left(\boldsymbol{J}^T(\boldsymbol{p})\boldsymbol{J}(\boldsymbol{p})\right)^{-1}\right\}} \quad . \tag{5.15}$$



(a) Scenario with small average GDOP.

(b) Scenario with large average GDOP.

Figure 5.1: Two-dimensional scenarios that are considered for positioning in this thesis. The ROs are denoted by black circles and the GDOP is indicated by contour lines.

In this thesis, two scenarios with different GDOP are considered for localisation. Both scenarios are shown in Figure 5.1, in which the ROs are denoted by black circles and the GDOP is indicated by contour lines. In scenario (a), the ROs are placed in the corners of a quadratic area with side length $\sqrt{2}R$, where R is the distance from every corner to the middle of the area. This geometry leads to a small GDOP on average. In contrast, the average GDOP is large for scenario (b). In this case, all ROs are placed in the lower left corner of the quadratic area. In Section 5.3.3, the performance of the proposed positioning algorithms will be investigated for both scenarios.

5.3 Positioning Algorithms

As already mentioned above, there are several positioning algorithms for localisation based on the TOA. In this thesis, two different methods are considered: On the one hand, the ML estimator in conjunction with the LM method is applied. This combination is called Taylor series (TS) algorithm subsequently. On the other hand, the suboptimum, non-iterative weighted least-squares (WLS) approach according to [STK05], [GCWI08] is taken into account. If not stated otherwise, the position estimate obtained by the WLS method is used as initial guess for the TS algorithm. First, the TS algorithm is presented in Section 5.3.1. Then, the WLS method is introduced in Section 5.3.2 and, finally, the performance of both algorithms is compared in Section 5.3.3.

5.3.1 Taylor-Series Algorithm

For the application of the TS algorithm, the Jacobian $\boldsymbol{J}_{\Omega_{\hat{\boldsymbol{d}}}}(\tilde{\boldsymbol{p}})$ and the (approximated) Hessian $\boldsymbol{H}_{\Omega_{\hat{\boldsymbol{d}}}}(\tilde{\boldsymbol{p}})$ of the metric $\Omega_{\hat{\boldsymbol{d}}}(\tilde{\boldsymbol{p}})$ need to be determined (see also (4.74) and (4.75) in Section 4.4.1):

$$\boldsymbol{J}_{\Omega_{\hat{\boldsymbol{d}}}}(\tilde{\boldsymbol{p}}) = -2 \cdot \boldsymbol{J}^T(\tilde{\boldsymbol{p}})\, \boldsymbol{C}_{\boldsymbol{e}}^{-1} \left(\hat{\boldsymbol{d}} - \boldsymbol{d}(\tilde{\boldsymbol{p}})\right), \tag{5.16}$$

$$\boldsymbol{H}_{\Omega_{\hat{\boldsymbol{d}}}}(\tilde{\boldsymbol{p}}) \approx \quad 2 \cdot \boldsymbol{J}^T(\tilde{\boldsymbol{p}})\, \boldsymbol{C}_{\boldsymbol{e}}^{-1}\, \boldsymbol{J}(\tilde{\boldsymbol{p}}). \tag{5.17}$$

Based on a current position hypothesis $\tilde{\boldsymbol{p}}_{\text{cur}}$, a new position hypothesis $\tilde{\boldsymbol{p}}_{\text{new}}$ can be determined with the help of the LM method according to (see also (4.77) in Section 4.4.1):

$$\begin{aligned} \tilde{\boldsymbol{p}}_{\text{new}} &= \tilde{\boldsymbol{p}}_{\text{cur}} - \left(\boldsymbol{H}_{\Omega_{\hat{\boldsymbol{d}}}}(\tilde{\boldsymbol{p}}_{\text{cur}}) + \omega \boldsymbol{\mathcal{I}}\right)^{-1} \cdot \boldsymbol{J}_{\Omega_{\hat{\boldsymbol{d}}}}(\tilde{\boldsymbol{p}}_{\text{cur}}) \\ &= \tilde{\boldsymbol{p}}_{\text{cur}} + \left(\boldsymbol{J}^T(\tilde{\boldsymbol{p}}_{\text{cur}})\, \boldsymbol{C}_{\boldsymbol{e}}^{-1}\, \boldsymbol{J}(\tilde{\boldsymbol{p}}_{\text{cur}}) + \omega \boldsymbol{\mathcal{I}}\right)^{-1} \cdot \boldsymbol{J}^T(\tilde{\boldsymbol{p}}_{\text{cur}})\, \boldsymbol{C}_{\boldsymbol{e}}^{-1} \left(\hat{\boldsymbol{d}} - \boldsymbol{d}(\tilde{\boldsymbol{p}}_{\text{cur}})\right), \end{aligned} \tag{5.18}$$

where ω is the tuning factor of the LM method summarised in Algorithm 4.1. Generally, the covariance matrix $\boldsymbol{C}_{\boldsymbol{e}}$ of the pseudorange errors $\boldsymbol{e}$ is not known exactly. However, it can be approximated with the help of soft information concerning the TOA estimates that can be obtained as described in Section 4.5:

$$\sigma_{e_b}^2 \approx \mathrm{c}^2 \cdot \sigma_{\hat{\tau}_{b,1}}^2 = \mathrm{c}^2 \cdot \sigma_{\hat{\varepsilon}_b}^2. \tag{5.19}$$

In order to indicate that estimated instead of true variances are used, the inverse covariance matrix in (5.18) is replaced by a weighting matrix

$$\boldsymbol{C}_{\boldsymbol{e}}^{-1} \approx \boldsymbol{W} = \operatorname{diag}\left(\frac{1}{\mathrm{c}^2\,\sigma_{\hat{\varepsilon}_1}^2}, \cdots, \frac{1}{\mathrm{c}^2\,\sigma_{\hat{\varepsilon}_B}^2}\right). \tag{5.20}$$

This leads to the following expression for the position update,

$$\tilde{\boldsymbol{p}}_{\mathrm{new}} = \tilde{\boldsymbol{p}}_{\mathrm{cur}} + \left(\boldsymbol{J}^T(\tilde{\boldsymbol{p}}_{\mathrm{cur}})\,\boldsymbol{W}\,\boldsymbol{J}(\tilde{\boldsymbol{p}}_{\mathrm{cur}}) + \omega\boldsymbol{\mathcal{I}}\right)^{-1} \cdot \boldsymbol{J}^T(\tilde{\boldsymbol{p}}_{\mathrm{cur}})\,\boldsymbol{W}\left(\hat{\boldsymbol{d}} - \boldsymbol{d}(\tilde{\boldsymbol{p}}_{\mathrm{cur}})\right), \tag{5.21}$$

that is considered for the TS algorithm in the remainder of this thesis. For the initialisation of the TS method, an initial guess of the MS's position $\tilde{\boldsymbol{p}}_0$ is needed. If the MS moves slowly, the position estimate of the preceding transmission burst can be used as initial guess. Furthermore, an initial guess for positioning can be obtained from a suboptimal, non-iterative localisation algorithm like the WLS method.

5.3.2 Weighted Least-Squares Algorithm

With the WLS method, the set of nonlinear equations described by (5.3) and (5.5) is transformed such that it can be solved in closed form [STK05], [GCWI08]. Hence, the WLS method is non-iterative and less costly than the TS algorithm. The basic idea is to transform the original set of *nonlinear* equations into a set of *linear* equations. For this purpose, squared estimated distances without errors are considered:

$$\begin{aligned}\hat{d}_b^2 &= (x - x_b)^2 + (y - y_b)^2 \\ &= x^2 + y^2 + x_b^2 + y_b^2 - 2xx_b - 2yy_b\end{aligned} \tag{5.22}$$

with $1 \leq b \leq B$. One RO is selected as reference, whose squared distance is subtracted from the squared distances of all other ROs. Without loss of generality, the first RO is chosen here:

$$\hat{d}_b^2 - \hat{d}_1^2 = \underbrace{(x_b^2 + y_b^2)}_{R_b^2} - \underbrace{(x_1^2 + y_1^2)}_{R_1^2} - 2(x_b - x_1)x - 2(y_b - y_1)y \tag{5.23}$$

with $2 \leq b \leq B$. The equation in (5.23) can be rearranged as

$$(x_b - x_1)x + (y_b - y_1)y = -\frac{1}{2}\left(\hat{d}_b^2 - \hat{d}_1^2 - R_b^2 + R_1^2\right). \tag{5.24}$$

By these transformations, a linear set of equations with respect to the MS's position $\boldsymbol{p}=[x, y]^T$ is obtained [STK05]:

$$\boldsymbol{S}\boldsymbol{p} = \boldsymbol{b} \tag{5.25}$$

with system matrix

$$\boldsymbol{S} = \begin{bmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \\ \vdots & \vdots \\ x_B - x_1 & y_B - y_1 \end{bmatrix} \tag{5.26}$$

and observation vector

$$\boldsymbol{b} = -\frac{1}{2}\begin{bmatrix} \hat{d}_2^2 - \hat{d}_1^2 - R_2^2 + R_1^2 \\ \hat{d}_3^2 - \hat{d}_1^2 - R_3^2 + R_1^2 \\ \vdots \\ \hat{d}_B^2 - \hat{d}_1^2 - R_B^2 + R_1^2 \end{bmatrix}. \tag{5.27}$$

This linear set of equations can be solved in closed form according to the least-squares principle:

$$\hat{\boldsymbol{p}} = \left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\boldsymbol{S}\boldsymbol{b}. \tag{5.28}$$

In order to improve the positioning accuracy, a weighting matrix

$$\boldsymbol{\mathcal{W}} = \operatorname{diag}\left(\frac{1}{\mathrm{c}^4\,\sigma_{\hat{\varepsilon}_2}^4}, \cdots, \frac{1}{\mathrm{c}^4\,\sigma_{\hat{\varepsilon}_B}^4}\right) \tag{5.29}$$

can be incorporated into the least-squares approach [GCWI08]:

$$\hat{\boldsymbol{p}} = \left(\boldsymbol{S}^T\boldsymbol{\mathcal{W}}\boldsymbol{S}\right)^{-1}\boldsymbol{S}\boldsymbol{\mathcal{W}}\boldsymbol{b}. \tag{5.30}$$

Based on (5.30), the initial guess for the TS algorithm is generated.

5.3.3 Performance Comparison of the Proposed Positioning Algorithms

Subsequently, the performance of the proposed positioning algorithms is examined for different channel models and different geometric scenarios. As before, the LOS channel as well as several two-path and WINNER channels are considered. For every channel model, two geometric scenarios with small and large average GDOP, respectively, according to Figure 5.1 are investigated, i.e. B=4 ROs are always assumed. For the first scenario, a better performance compared to the second scenario is expected due to the smaller average GDOP of the first scenario. In all cases, the *root mean squared error (RMSE)* of the position estimate

$$\mathrm{RMSE}\left(\hat{\boldsymbol{p}}\right) = \sqrt{\mathrm{E}\left\{|\hat{\boldsymbol{p}} - \boldsymbol{p}|^2\right\}} = \sqrt{\mathrm{E}\left\{(\hat{x}-x)^2 + (\hat{y}-y)^2\right\}} \tag{5.31}$$

is used as performance measure, where the expectation is taken with respect to different channel and position realisations. On the one hand, the positioning accuracy is examined over the quadratic area in Figure 5.1 given a certain SNR. In this case, the RMSE at a certain MS's position $\boldsymbol{p}$ is averaged over different channel realisations only. On the other hand, the RMSE of position estimation is examined over SNR. In that case, the MS's position $\boldsymbol{p}$ is determined randomly inside the quadratic area for each run at a certain SNR and the RMSE is averaged over all channel *and* position realisations. For comparison, the CRLB as derived in Section 5.2 is considered. As the RMSE instead of the MSE is

applied in conjunction with position estimation, the square root of the CRLB, which is denoted as *root CRLB (RCRLB)* in the following, needs to be calculated:

$$\mathrm{RCRLB}\left(\boldsymbol{p}\right) = \sqrt{\mathrm{CRLB}\left(\boldsymbol{p}\right)} = \sqrt{\frac{1}{2}\,\mathrm{tr}\left\{\left(\boldsymbol{J}^T(\boldsymbol{p})\;\boldsymbol{C}_{\boldsymbol{e}}^{-1}\boldsymbol{J}(\boldsymbol{p})\right)^{-1}\right\}}. \tag{5.32}$$

The covariance matrix of the estimation errors $\boldsymbol{C_e}$ is a diagonal matrix,

$$\boldsymbol{C_e} = \mathrm{diag}\left(\sigma_{e_1}^2, \cdots, \sigma_{e_B}^2\right), \tag{5.33}$$

whose entries are determined from the CRLBs of the sampling phase for channel parameter estimation:

$$\sigma_{e_b}^2 = \mathrm{c}^2 \cdot \sigma_{\tau_{b,1}}^2 = \mathrm{c}^2 \cdot \sigma_{\varepsilon_b}^2 = \mathrm{c}^2 \cdot \mathrm{CRLB}\left(\varepsilon_b\right). \tag{5.34}$$

In all figures shown below, the RMSE as well as the RCRLB are normalised with respect to the distance $d_s = \mathrm{c}\,T_s$, that corresponds to the symbol duration T_s.

For each link between a RO and the MS, the same simulation setup as already applied in Section 4.3.3 and Section 4.4.3 with $K'=100$ and $J=2$ is used. Different channel parameters are determined for every link and in each run according to the recent channel model. The distance R from a corner of the quadratic region in Figure 5.1 to its middle point is assumed to be a multiple of the distance d_s: $R=2 \cdot d_s$. In order to emphasise the effect of the GDOP, power control is considered subsequently, i.e. the same SNR is applied for all links. As before, only one-shot estimates without tracking or averaging are taken into account: For each link, channel parameter estimation is performed and soft information based on the linearisation and the likelihood method is obtained. The pseudoranges are given by

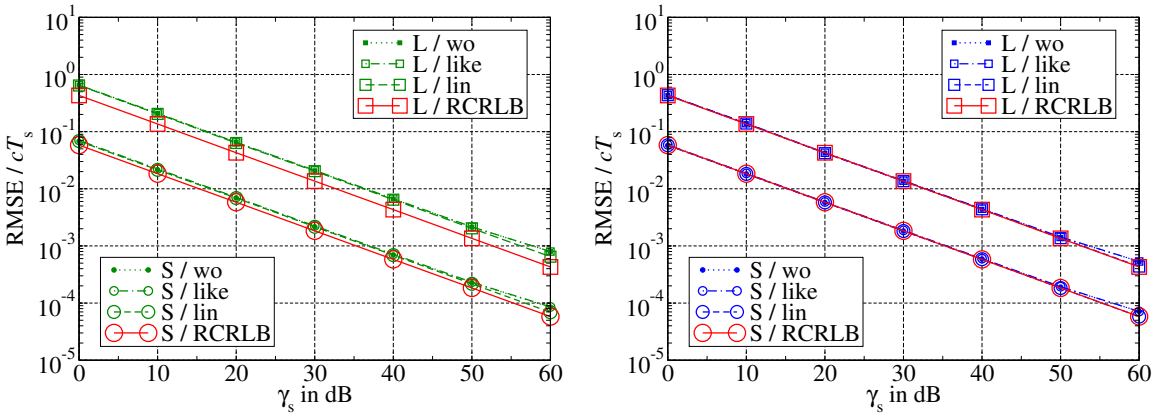
$$\hat{d}_b = d_b(\boldsymbol{p}) + \mathrm{c} \cdot (\varepsilon_b - \hat{\varepsilon}_b) \tag{5.35}$$

and the weighting matrices $\boldsymbol{W}$ and $\boldsymbol{\mathcal{W}}$ are determined according to (5.20) and (5.29) in conjunction with (5.19) applying the different soft information methods, respectively. For comparison, positioning without soft information is considered. In this case, the weighting matrices correspond to identity matrices: $\boldsymbol{W}=\boldsymbol{\mathcal{I}}$ and $\boldsymbol{\mathcal{W}}=\boldsymbol{\mathcal{I}}$. First, localisation based on the WLS method is performed, whose position estimate is used as initial guess for the TS algorithm. In case of the WLS method, the RO with the best weighting factor is used as reference and, for the TS algorithm, the stopping criterion based on the fitness improvement with $T=10^{-3}$ and $N_{\min}=2$ is applied.

LOS Channel Model

At first, the results for the LOS channel are discussed because it represents the best possible case. Channel parameter estimation is trivial in this case and it has been shown in Section 4.5.4 that both soft information methods are accurate in case of the LOS channel. Due to power control, the pseudorange errors e_b should be nearly the same for all links. Hence, positioning with and without weighting is supposed to perform equally well in a LOS channel. In Figure 5.2 and Figure 5.3, the RMSE of $\hat{\boldsymbol{p}}$ normalised to $d_s = \mathrm{c}\,T_s$ is

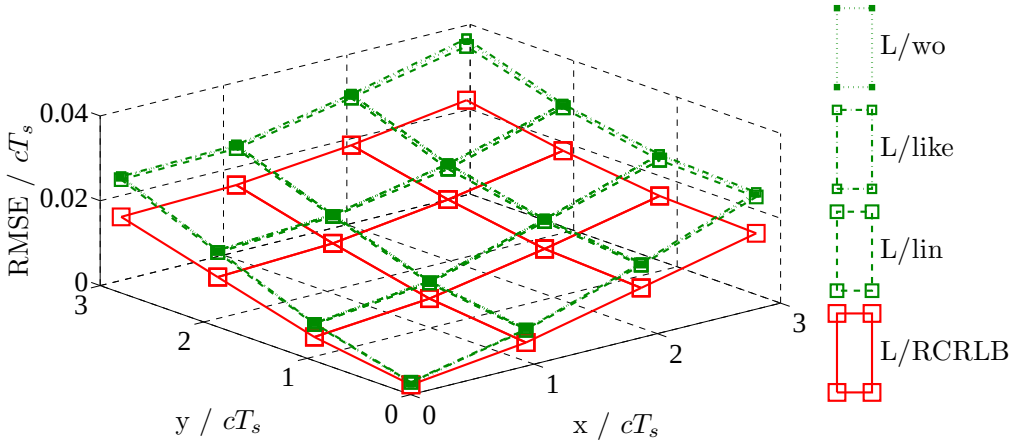
shown for the LOS channel over SNR and over space at 30 dB, respectively. As mentioned above, positioning without soft information ("wo"), with soft information from the likelihood method ("like") and with soft information from the linearisation method ("lin") is considered and the normalised RCRLB is plotted for comparison. For the likelihood method, PSO with 10 particles per dimension and T=10^{-5} as well as $N_{\min}$=30 is used, while the LM algorithm with T=10^{-3} and $N_{\min}$=2 is applied for the linearisation method. The parameter estimate of the PSO method is used as IG for the LM algorithm. In each figure, the results for the WLS method and the TS algorithm are illustrated in the subfigures (a) and (b), respectively. Curves labelled with "L" are obtained for the scenario with large average GDOP and curves labelled with "S" are obtained for the scenario with small average GDOP. Both figures show that the positioning accuracy increases with increasing SNR and decreases with increasing GDOP: As supposed previously, the performance for the scenario with small average GDOP is much better than the performance for the scenario with large average GDOP (see Figure 5.2): The gap between both RCRLBs is approximately 18 dB, i.e. significant performance gains can be achieved if the GDOP is kept low. Additionally, the assumption that positioning with and without weighting performs equally well in a LOS channel is confirmed by Figure 5.2 and Figure 5.3: The RMSEs for localisation without soft information and with soft information from the likelihood as well as the linearisation method coincide in all cases. The TS algorithm is, furthermore, an MVU estimator since the RMSEs approach the RCRLBs for all SNRs (see Figure 5.2(b)) and all positions even for very large GDOPs (see Figure 5.3(b)). The iterative TS algorithm clearly outperforms the non-iterative WLS method, for which a certain gap between the RCRLB and the RMSE occurs. In Figure 5.3(a), it can be observed that this gap depends on the position and, thus, on the GDOP: The larger the GDOP, the larger is



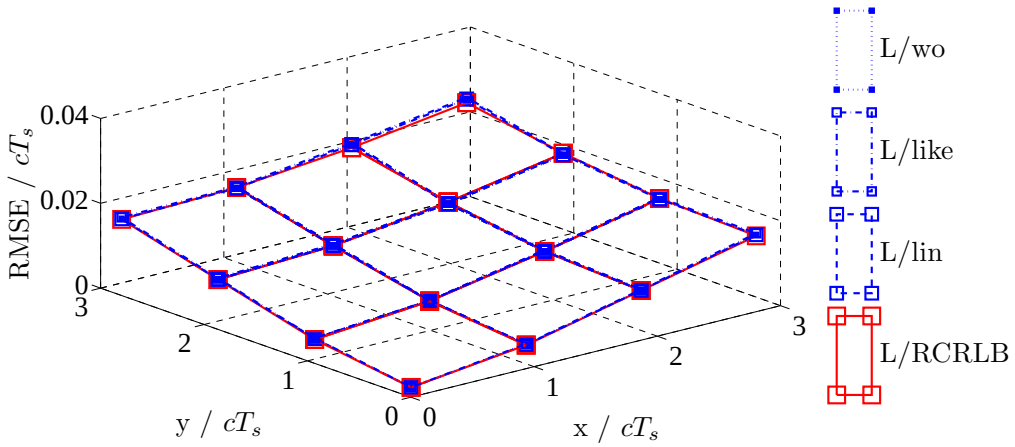
(a) WLS method. (b) TS algorithm.

Figure 5.2: RMSE of $\hat{\boldsymbol{p}}$ normalised with respect to d_s=$\mathrm{c}\,T_s$ for the LOS channel over SNR. Positioning without soft information ("wo"), with soft information from the likelihood method ("like"), and with soft information from the linearisation method ("lin") is considered. For comparison, the RCRLB normalised with respect to d_s is plotted. Curves labelled with "L" are obtained for the scenario with large average GDOP and curves labelled with "S" are obtained for the scenario with small average GDOP.

the gap. Consequently, the gap between the RCRLB and the RMSE in Figure 5.2(a) is smaller for the scenario with small average GDOP. Due to its superior performance, only the TS algorithm is considered subsequently. In general, very accurate position estimates are obtained for the LOS channel. Even for large GDOPs, the normalised RMSEs are well below 10^0, i.e. a resolution much smaller than the distance d_s corresponding to the symbol duration T_s is achieved. Since pure LOS propagation is quite unlikely to occur in terrestrial radio systems and because positioning typically suffers from multipath propagation, the above results are only of theoretical interest and provide a lower bound on the achievable localisation accuracy.



(a) WLS method.

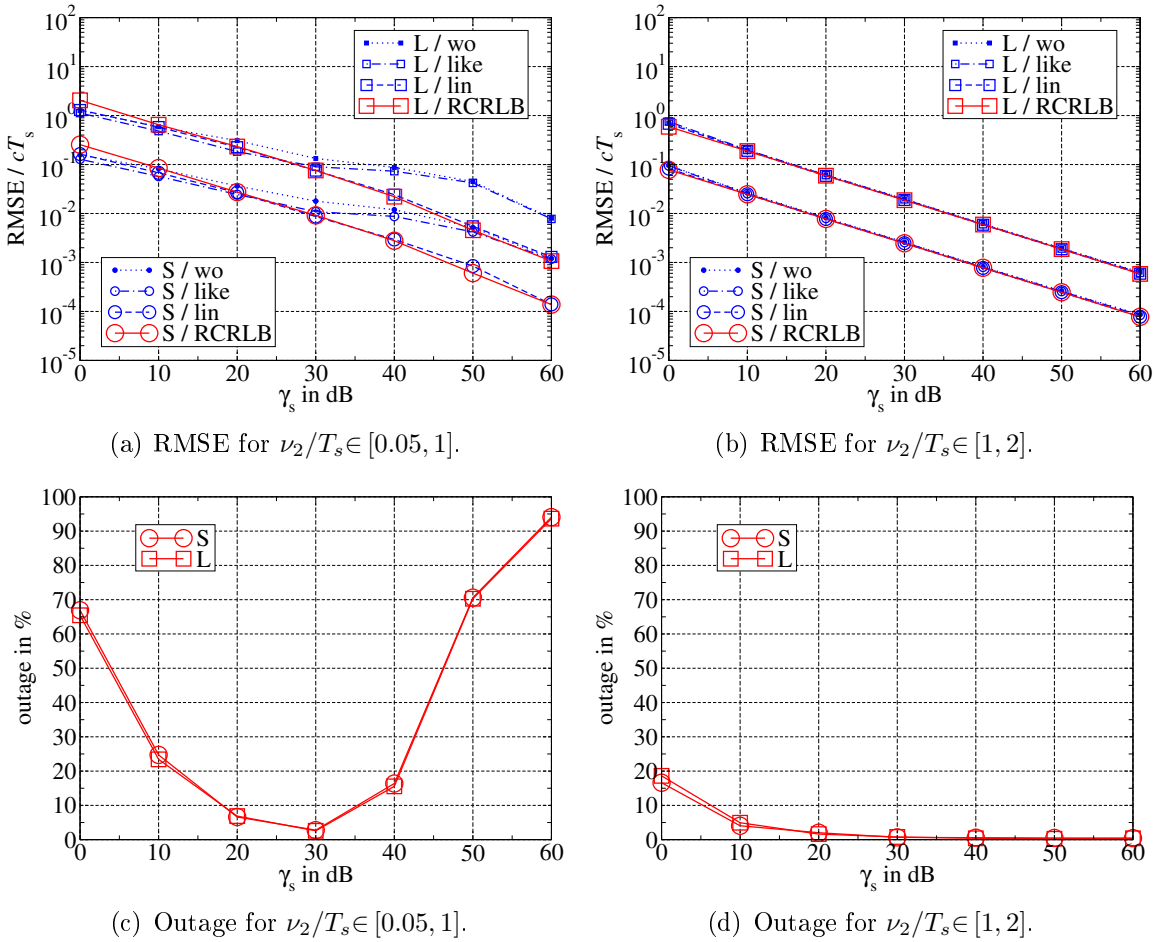


(b) TS algorithm.

Figure 5.3: RMSE of $\hat{\boldsymbol{p}}$ normalised with respect to $d_s = \mathrm{c}\,T_s$ for the LOS channel over space at 30 dB. Positioning without soft information ("wo"), with soft information from the likelihood method ("like"), and with soft information from the linearisation method ("lin") is considered. For comparison, the RCRLB normalised with respect to d_s is plotted. Only the scenario with large average GDOP ("L") is considered here.

Two-Path Channel Models

As in Section 4.3.3 and Section 4.4.3, a two-path channel with small excess delay ($\nu_2/T_s \in [0.05, 1]$) and a two-path channel with large excess delay ($\nu_2/T_s \in [1, 2]$) are considered. For each channel realisation, the power ratio $\mathcal{P}$ and the phase offset $\Delta\Phi$ are generated randomly in the intervals $[0.1, 10]$ and $[0, 2\pi[$, respectively. In contrast to the LOS channel, the pseudorange errors e_b may differ considerably for each link because the accuracy of the estimated sampling phase $\hat{\varepsilon}_b$ depends on the current channel realisation (i.e. the current combination of excess delay, phase offset and power ratio) as illustrated in Figure 4.36 to Figure 4.38. Hence, positioning with soft information may be advanta-



(a) RMSE for $\nu_2/T_s \in [0.05, 1]$.

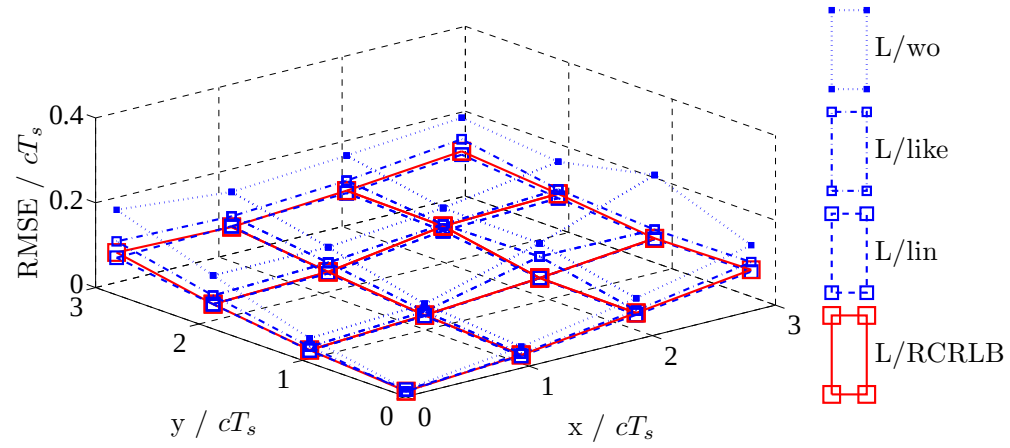
(b) RMSE for $\nu_2/T_s \in [1, 2]$.

(c) Outage for $\nu_2/T_s \in [0.05, 1]$.

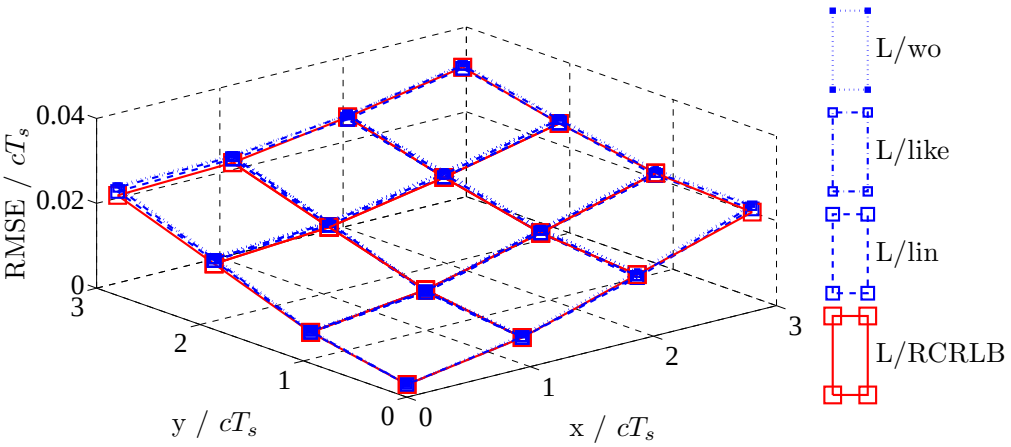
(d) Outage for $\nu_2/T_s \in [1, 2]$.

Figure 5.4: RMSE of $\hat{\boldsymbol{p}}$ normalised with respect to $d_s = \mathrm{c}\,T_s$ and outage rate in percent for the two-path channels over SNR. Positioning without soft information ("wo"), with soft information from the likelihood method ("like"), and with soft information from the linearisation method ("lin") is considered. For comparison, the RCRLB normalised with respect to d_s is plotted. Curves labelled with "L" are obtained for the scenario with large average GDOP and curves labelled with "S" are obtained for the scenario with small average GDOP.

geous in case of a two-path channel. It has been shown in Section 4.5.4 that both soft information methods are only accurate at high SNR in case of multipath propagation and that it is more likely that the likelihood method is accurate. Hence, the likelihood method is supposed to outperform the linearisation method in case of the two-path channels. Similar to the LOS channel, the LM algorithm with $T{=}10^{-3}$ and $N_{\min}{=}2$ is used for the linearisation method and the PSO method with 10 particles per dimension is applied for the likelihood method, but the stopping criterion for the PSO method is tightened to $T{=}10^{-6}$ and $N_{\min}{=}50$. Furthermore, outage detection as described in Section 4.4.3 is applied for both algorithms. In Figure 5.4 and Figure 5.5, the positioning performance for both channel models based on the TS algorithm is illustrated. The same labelling



(a) TS algorithm for $\nu_2/T_s \in [0.05, 1]$.



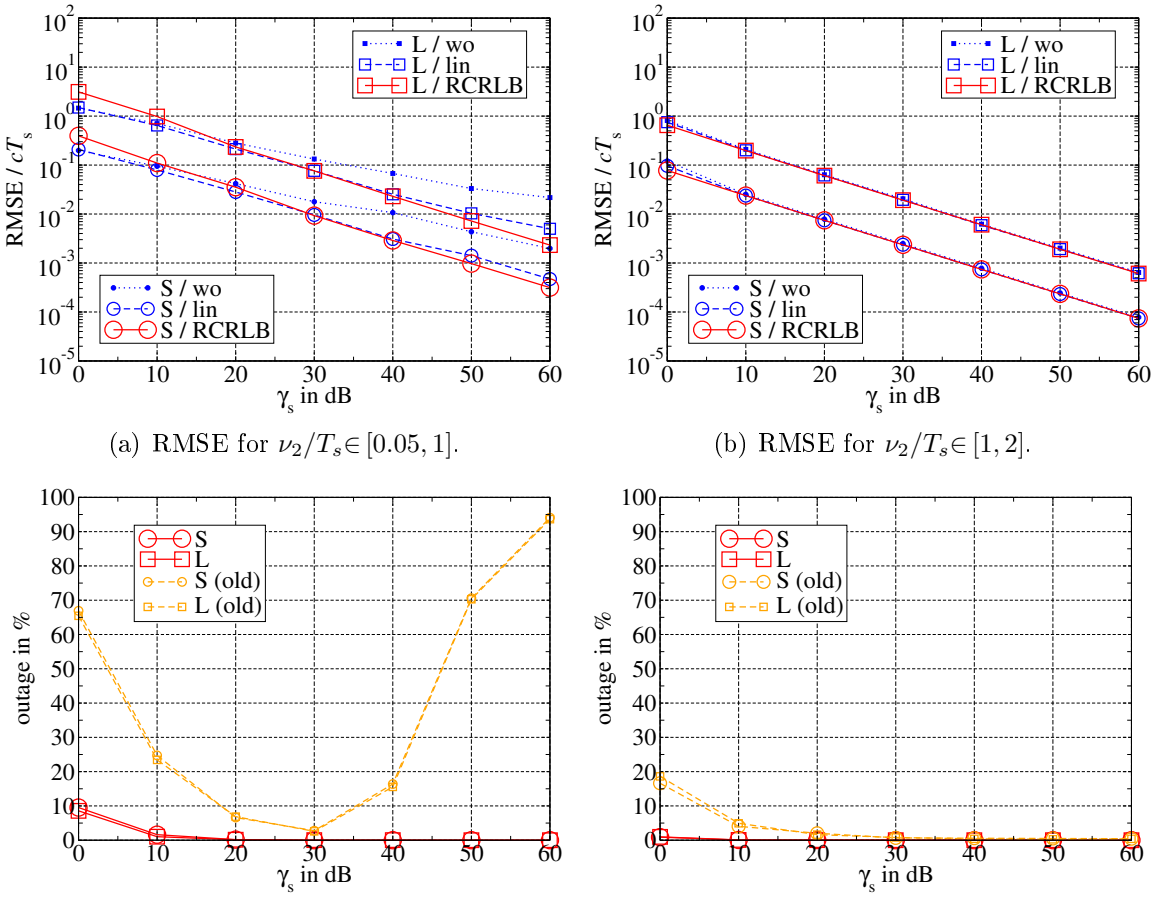
(b) TS algorithm for $\nu_2/T_s \in [1, 2]$.

Figure 5.5: RMSE of $\hat{\boldsymbol{p}}$ normalised with respect to $d_s = \mathrm{c}\,T_s$ for the two.path channels over space at 30 dB. Positioning without soft information ("wo"), with soft information from the likelihood method ("like"), and with soft information from the linearisation method ("lin") is considered. For comparison, the RCRLB normalised with respect to d_s is plotted. Only the scenario with large average GDOP ("L") is considered here.

("wo", "like", "lin", "L", "S") as for the LOS channel is applied. In Figure 5.4, not only the normalised RMSE, but also the outage rate is shown for both channel models and both geometric scenarios. In order to achieve a fair comparison between the soft information methods, a single run is only successful if *all* ROs deliver pseudoranges based on the PSO *and* the LM method. This means that a run is not successful and contributes to the overall outage rate if the PSO *or* the LM method fail for a *single* RO. As a consequence, very large outage rates are obtained for the channel with small excess delay. At low SNR, the overall outage rate is dominated by the LM algorithm, while the PSO method dominates the overall outage rate at high SNR. Only around 20 dB to 30 dB, acceptable outage rates below 10% are achieved for the channel with small excess delay. The outage rates do not depend on the geometric scenario, but only on the underlying channel model. For the channel with large excess delay, the overall outage rate is generally low and only above 10% at 0 dB.

Even though the RCRLBs and the RMSEs for the scenario with small average GDOP are much smaller than for the scenario with large average GDOP, the general behaviour of the corresponding curves is the same for both scenarios. In case of the channel with large excess delay, the RMSEs for positioning without soft information and with soft information from the linearisation as well as the likelihood method coincide with the RCRLBs over the whole SNR range. This means that the best possible performance is achieved even without weighting. Hence, soft information is not required in that case. The channel model with small excess delays is much more challenging because it is more difficult to separate closely spaced multipath components. Surprising results are obtained in that case (see Figure 5.4(a)). Although the linearisation method is supposed to be less accurate than the likelihood method, it clearly outperforms the likelihood method: At low SNR, both soft information methods show approximately the same performance. Their RMSEs are even below the RCRLB due to the bias of channel parameter estimation at low SNR. The RMSE of the likelihood method is only slightly better than the RMSE of the linearisation method. But at high SNR, the RMSE of the linearisation method is much lower than the RMSE of the likelihood method, that shows approximately the same performance as positioning without soft information. A gain of approximately 13 dB is achieved with the linearisation method in comparison to the remaining methods. In general, the RMSE of the linearisation method is very close to the RCRLB over the whole SNR range, i.e. it provides almost always the best possible performance. The above observations are confirmed by Figure 5.5, in which the RMSEs are plotted over space at 30 dB. For the channel with large excess delay, all RMSEs coincide with the RCRLB. For the channel with small excess delay, only the RMSE of the linearisation method lies on the RCRLB. For the remaining methods, a gap between the RMSE and the RCRLB occurs that increases with GDOP. At 30 dB, the gap for the likelihood method is quite small, while positioning without soft information already leads to a significant performance loss. From the above results, it can be concluded that the potential inaccuracy of the linearisation method does not influence the positioning accuracy. Actually, it seems that the absolute values of the weights in the weighting matrices $\boldsymbol{W}$ and $\boldsymbol{\mathcal{W}}$ are not crucial. Rather a correct ratio of the weights, that originate from different ROs, is relevant. Thus, rough soft information is sufficient as long as the ratio of the pseudorange variances is accurate. This condition is obviously better

fulfilled for the less complex and potentially less accurate linearisation method. Hence, it is suggested to apply the linearisation method in order to obtain soft information for the parameter estimates and, in the remainder of this thesis, only this method is considered for this purpose. This means that the sophisticated combination of the PSO and the LM method based on the fitness threshold can be applied for channel parameter estimation. The overall outage rates are automatically reduced in this way. Furthermore, only three ROs are required for two-dimensional positioning. The outage rates can be further reduced if the channel parameter estimator only needs to be successful for three instead of four ROs. Hence, positioning only fails if two or more ROs fail. Corresponding results are shown in Figure 5.6. The combination of the PSO and LM method with the thresholds T_{PSO} and T_{LM} defined in (4.105) as well as (4.106), respectively, is used. These thresh-



(a) RMSE for $\nu_2/T_s \in [0.05, 1]$.

(b) RMSE for $\nu_2/T_s \in [1, 2]$.

(c) Outage for $\nu_2/T_s \in [0.05, 1]$.

(d) Outage for $\nu_2/T_s \in [1, 2]$.

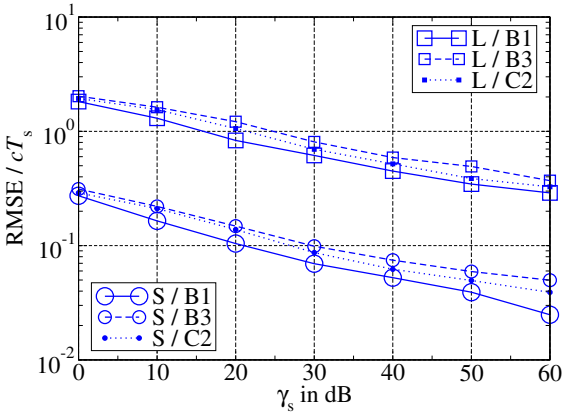
Figure 5.6: RMSE of $\hat{\boldsymbol{p}}$ normalised with respect to $d_s = c\,T_s$ and outage rate in percent for the two-path channels over SNR. Positioning without soft information ("wo") and with soft information from the linearisation method ("lin") is considered. For comparison, the RCRLB normalised with respect to d_s is plotted. Curves labelled with "L" are obtained for the scenario with large average GDOP and curves labelled with "S" are obtained for the scenario with small average GDOP.

olds are applied in conjunction with $N_{\min}$=20 and $N_{\min}$=10 as suggested in Section 4.4.3. Again, the normalised RMSEs as well as the outage rates are illustrated for both channel models and both geometric scenarios. The old outage rates are plotted for comparison. A dramatic decrease of the overall outage rate is observed. In most cases, the outage rates are negligible now. Only for the channel with small excess delay, the outage rate is around 9% at 0 dB. With respect to the RMSE, a similar behaviour as in Figure 5.4 is observed. For the channel with large excess delay, the RMSEs for positioning with and without soft information coincide with the RCRLB. In case of the channel with small excess delay, positioning with soft information clearly outperforms positioning without soft information: A significant gap between both RMSEs occurs at high SNR, where the RMSE of the linearisation method is always close to the RCRLB. Hence, severe multipath propagation can be effectively mitigated with the help soft channel parameter estimation in case of the two-path channels. Since the computation of the covariance matrix, which is needed for the linearisation method, is quite simple, the performance gain due to the exploitation of soft information comes at a very low cost.

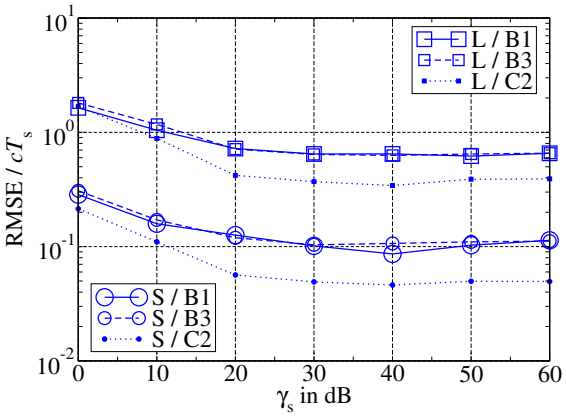
WINNER Channel Models

The WINNER channel models are the most realistic channel models applied in this thesis and are most challenging with respect to channel parameter estimation. It has been shown in Section 4.3.3 and Section 4.4.3 that the CRLBs for channel parameter estimation are often meaningless high for the WINNER channel models, while the MSEs of the sampling phase are well below the CRLB in these cases. Hence, CRLBs have not been further considered in conjunction with the WINNER channels in Section 4.4.3. Since the RCRLBs for positioning are determined based on the CRLBs for channel parameter estimation, the RCRLBs will be meaningless high for positioning as well. Therefore, RCRLBs are also neglected at this point and only RMSEs are considered for the WINNER channel models.

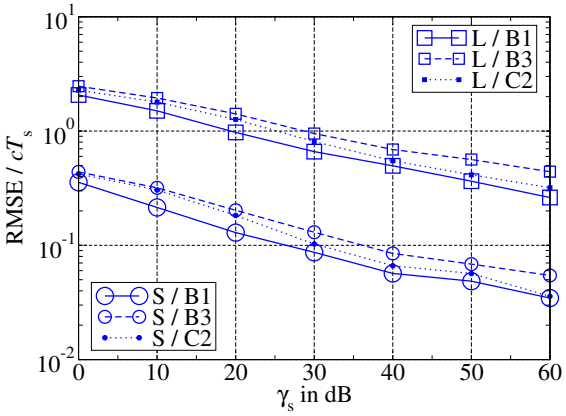
The positioning accuracy in combination with a full as well as a reduced estimator with three multipath components (as proposed in Section 4.4.3) is investigated for all WINNER channel models ("B1", "B3", "C2") subsequently. Corresponding results for T_s=100 ns are shown in Figure 5.7. On the left hand side, the results for the full estimator are illustrated, while the results for the reduced estimator are shown on the right hand side. RMSEs for positioning without soft information (first row of Figure 5.7) and with soft information from the linearisation method (second row of Figure 5.7) as well as outage rates (third row of Figure 5.7) are examined. In each subfigure, the curves for all three WINNER channel models are plotted in order to facilitate a direct comparison of these channels. At first sight, it is observed that positioning with and without soft information leads to approximately the same performance. The performance gain for positioning with soft information, that occurs for the two-path channel with small excess delay, is not observable anymore. Furthermore, the gap between the RMSEs for small average GDOP ("S") and large average GDOP ("L") has become even larger for the WINNER channel models, i.e. it is even more important to keep the GDOP as low as possible. In case of the full estimator, the best performance is obtained for the B1-LOS model (urban



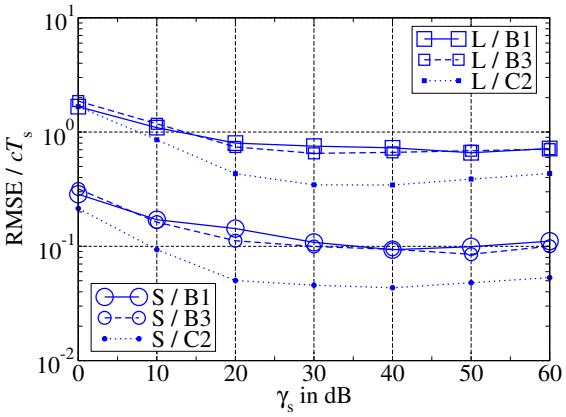
(a) RMSE for the full estimator and positioning without soft information.



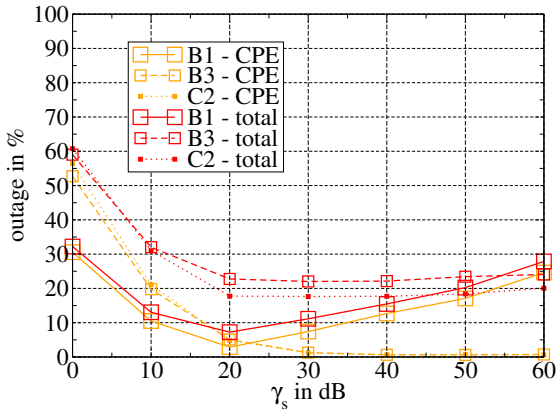
(b) RMSE for the reduced estimator and positioning without soft information.



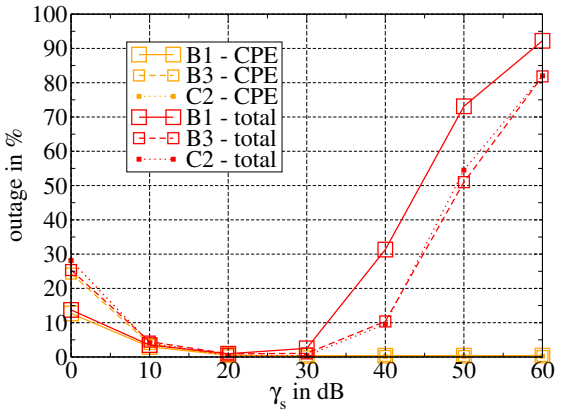
(c) RMSE for the full estimator and positioning with soft information from the linearisation method.



(d) RMSE for the reduced estimator and positioning with soft information from the linearisation method.



(e) Outage rates for the full estimator.

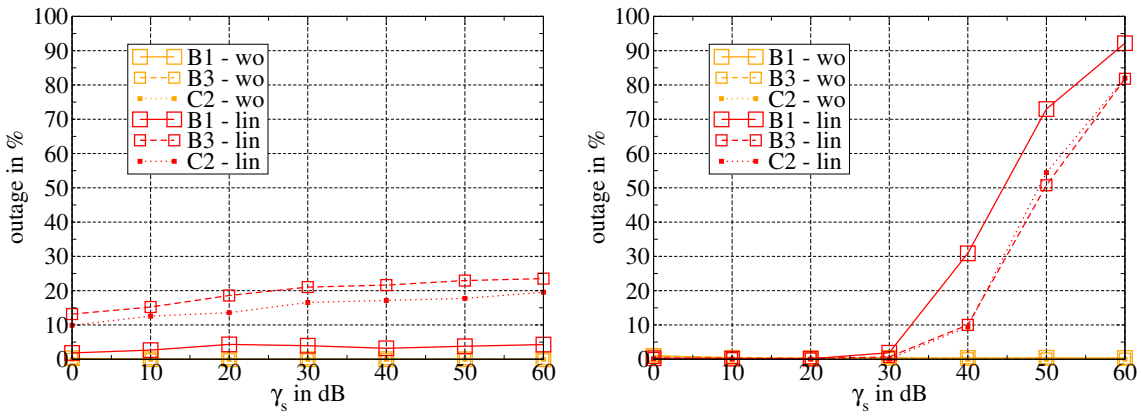


(f) Outage rates for the reduced estimator.

Figure 5.7: RMSE of $\hat{\boldsymbol{p}}$ normalised with respect to $d_s = \mathrm{c}\,T_s$ and outage rate in percent for the WINNER channels over SNR. Positioning without soft information and with soft information from the linearisation method is considered. Curves labelled with "L" are obtained for the scenario with large average GDOP and curves labelled with "S" are obtained for the scenario with small average GDOP.

micro-cell), while the B3-LOS model (large indoor hall) leads to the worst RMSEs. For the reduced estimator, a different behaviour is observed: The lowest RMSEs are obtained for the C2-LOS model (urban macro-cell), while the B1-LOS and the B3-LOS channel models show approximately the same (decreased) performance. For all channel models, an error floor occurs in case of the reduced estimator, i.e. the positioning accuracy remains approximately constant from 20 dB upwards.

With respect to the outage rates, a new aspect needs to be considered. Until now, the outage rates for positioning have only been influenced by the outage rates for channel parameter estimation, i.e. if more than two ROs fail, the MS's position can not be estimated. The positioning algorithms themselves always worked properly and did not contribute to the overall outage. In case of the WINNER channel models, the overall outage rates are influenced by the performance of the positioning algorithms as well. Unfortunately, the outage detection method applied for channel parameter estimation can not be exploited for position estimation because only estimated instead of true weighting matrices are applied. Consequently, there is no explicit distribution of the final fitness as in the case of channel parameter estimation. However, a rough outage detection can be implemented. Typically, the final fitness of the TS algorithm for positioning is below 10^2 for the examined SNR region. Therefore, a fixed threshold $T_f^{\text{pos}}=10^3$, that stays the same for all SNRs, is applied for outage detection in case of localisation, i.e. if $\Omega_{\hat{\boldsymbol{d}}}(\hat{\boldsymbol{p}})>T_f^{\text{pos}}$, the position estimate is denoted as outlier and not considered for the calculation of the RMSE. In Figure 5.7(e) and Figure 5.7(f), the corresponding outage rates are illustrated. The total outage rate as well as the contribution of channel parameter estimation ("CPE") to this total outage rate are plotted. The gap between both outage rates is caused by a failure of the TS algorithm. Obviously, a considerable part of the total outage rate is caused by a failure of the TS algorithm. Especially for the reduced estimator, the overall outage rate at high SNR is very large, while the outage rate for channel parameter estimation



(a) Outage for the full estimator.

(b) Outage for the reduced estimator.

Figure 5.8: Outage rate of the positioning algorithms in percent for the WINNER channels over SNR. Positioning without soft information ("wo") and with soft information from the linearisation method ("lin") is considered.

is close to zero and, hence, negligible. This means that channel parameter estimation is almost always successful, whereas the success or failure of position estimation is crucial. In order to further understand the high outage values for localisation, the outage rate for positioning can be split into the part caused by positioning without soft information ("wo") and the part caused by positioning with soft information from the linearisation method ("lin") as shown in Figure 5.8. Obviously, the localisation outage is mainly caused by the TS algorithm employing soft information. Hence, the weighting matrices for the WINNER channel models seem to be inaccurate, i.e. the ratio of the different entries in the weighting matrices does not reflect the true ratio of reliabilities very well.

There are two distinct reasons for this inaccuracy depending on the type of channel parameter estimator. In case of the full estimator, the covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}$ is of large dimension ($3C \times 3C$) because the number of propagation paths is quite large (C=8 for the B1-LOS and the C2-LOS channel and C=10 for the B3-LOS channel). The covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}$ is basically the inverse of the Fisher information matrix evaluated at the estimated parameters $\hat{\boldsymbol{\theta}}$. Due to closely spaced multipath components, the Fisher information matrix may become ill-conditioned or even rank-deficient. In order to avoid a complete failure of the matrix inversion needed for the calculation of the covariance matrix, singular value decomposition is applied for that purpose as described in [PTVF02, p. 62ff.]. However, due to the large size of the covariance matrix and because of the (potential) ill-conditioned Fisher information matrix, numerical instabilities occur that distort the soft information applied for localisation. As summarised in Table 4.2, very dense multipath components with a separation of 3.5 ns and 1.5 ns are present in case of the B3-LOS and the C2-LOS channel, respectively. For the first channel, this small separation occurs even twice. Therefore, the worst outage rate, in case of the full estimator, is obtained for the B3-LOS channel as illustrated in Figure 5.8(a). For all channel models, the localisation outage rates are rather constant over SNR. In contrast, the localisation outage rates for the reduced estimator are negligible at low SNR and increase rapidly for SNRs larger than 30 dB as shown in Figure 5.8(b). The worst outage rate is obtained for the B1-LOS channel. This completely different behaviour for the reduced estimator is not caused by numerical instabilities due to large and ill-conditioned matrices, but by a systematic error introduced through the reduction of the estimator. The replicas of the channel weight function incorporated in the reduced channel parameter estimator are only approximate because much less propagation paths than actually present are assumed. This simplifies the estimator, but introduces a systematic error. At low SNR, it is quite difficult to separate neighbouring paths due to large noise samples anyway and, thus, the systematic error is of little significance. At high SNR, however, neighbouring paths can be separated quite well such that the systematic error dominates the positioning performance which leads to an error floor and, apparently, to inaccurate weighting matrices.

Taking the above results into account, it can be concluded that positioning without weighting is advantageous in case of the WINNER channel models. The incorporation of soft information does not provide the desired performance gain, but, actually, complicates position estimation. Hence, only positioning without soft information is performed for the WINNER channel models in the remainder of this thesis.

5.4 Soft Information for Position Estimation

Similar to Section 4.5, soft information for positioning can be generated, i.e. the linearisation and the likelihood method are applicable for localisation as well. For the linearisation method, the output covariance of the TS algorithm is utilised and, for the likelihood method, PSO based positioning is applied.[1] As already mentioned before, soft information with respect to position estimates can be exploited for sensor fusion and/or for tracking algorithms like the Kalman filter. In contrast to channel parameter estimation, not only a single variance is of interest now, but the whole confidence region that illustrates the interdependencies between the incorporated variables. In case of two-dimensional positioning, the interdependence between the x and y component of the position estimate is revealed. If both components are independent, the confidence region based on the linearisation method is circular. Otherwise, elliptical confidence regions are obtained. As will be shown later in this section, the interdependence between the x and y component is influenced by the GDOP: The lower the GDOP, the less the components are correlated.

In the following, the accuracy of the two soft information methods for localisation is analysed with the help of the curvature measures of Bates and Watts as already done for channel parameter estimation in Section 4.5.4. For this purpose, the maximum relative curvatures Γ^{N} and Γ^{T} are calculated for positioning and the confidence regions of the linearisation as well as the likelihood method according to (4.112) and (4.113), respectively, are plotted and compared. The likelihood method is accurate if the planar assumption is fulfilled ($\Gamma^{\mathrm{N}}<\mathcal{C}^{1-\alpha}_{P,N-P}$), while the linearisation method is only accurate if the planar *and* the uniform coordinate assumption are fulfilled ($\Gamma^{\mathrm{N}}<\mathcal{C}^{1-\alpha}_{P,N-P}$, $\Gamma^{\mathrm{T}}<\mathcal{C}^{1-\alpha}_{P,N-P}$). In case of positioning, the number of parameters corresponds to the dimension of the position estimate (P=D) and the number of observations corresponds to the number of ROs (N=B). Here, two-dimensional positioning with four ROs is considered (P=2, N=4), i.e. the maximum relative curvatures must be compared to $\mathcal{C}^{0.95}_{P,N-P}$=0.22942 if α=0.05 is taken into account. As mentioned in Section 4.5.4, the curvature measures depend on the current parameters (here: on the current position) and on the current noise samples. Hence, a fixed position with $[x,y]^T/\,\mathrm{c}\,T_s$=$[1.16531, 2.18072]^T$ is considered for both geometric scenarios. This position leads to GDOP values of 1.01852 and 65.4819 in case of the scenario with small and large average GDOP, respectively. The above parameters and the corresponding maximum relative curvatures for positioning are summarised in Table 5.1. If a maximum relative curvature exceeds $\mathcal{C}^{1-\alpha}_{P,N-P}$ (i.e. the associated assumption is violated), it is marked in red colour. The corresponding confidence regions are shown in Figure 5.9 to Figure 5.11, where each figure consists of six subfigures. On the left hand side, the confidence regions for the scenario with small GDOP are shown, while the confidence regions for the scenario with large GDOP are illustrated on the right hand side. In each row, a different SNR is examined. As in Section 4.5.4, three different SNRs are considered: 10 dB, 30 dB and 50 dB. Furthermore, all three types of channel models are taken into account: the LOS channel as well as a two-path and a WINNER channel model. For each link in a certain channel model, a specific set of parameters is used and the same

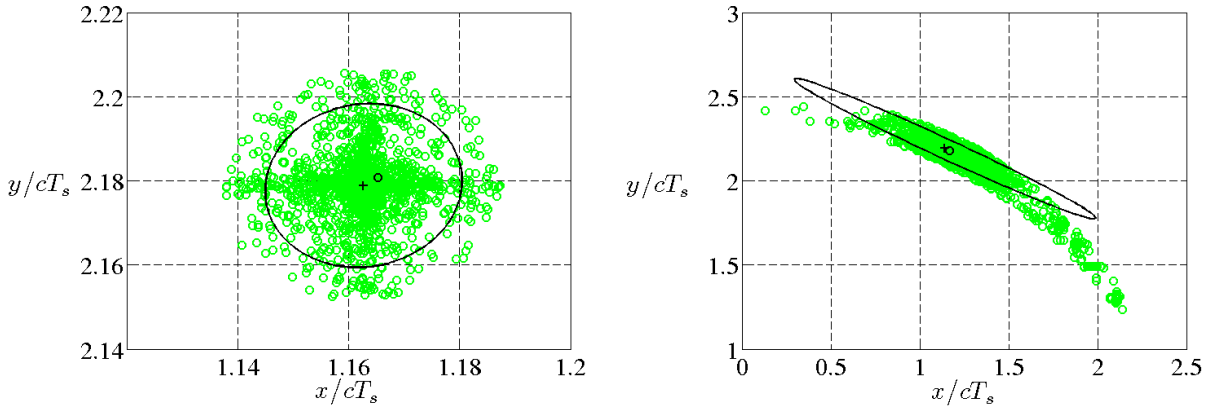
[1] In order to apply the PSO method to positioning, only the fitness function needs to be adapted.

		small GDOP	large GDOP
	Position	$x/\mathrm{c}\,T_s$=1.16531 $y/\mathrm{c}\,T_s$=2.18072	$x/\mathrm{c}\,T_s$=1.16531 $y/\mathrm{c}\,T_s$=2.18072
	GDOP	**1.01852**	**65.4819**
	P=D	2	2
	N=B	4	4
	$\mathcal{C}^{0.95}_{P,N-P}$	**0.22942**	**0.22942**
LOS channel	Γ^{N} @ 10 dB	0.00177	0.02440
	Γ^{T} @ 10 dB	0.00050	0.38899
	Γ^{N} @ 30 dB	0.00808	0.01432
	Γ^{T} @ 30 dB	0.00228	0.22867
	Γ^{N} @ 50 dB	0.00707	0.00872
	Γ^{T} @ 50 dB	0.00198	0.13925
two-path channel	Γ^{N} @ 10 dB	0.00568	0.01872
	Γ^{T} @ 10 dB	0.00188	0.30166
	Γ^{N} @ 30 dB	0.00422	0.01517
	Γ^{T} @ 30 dB	0.00117	0.24214
	Γ^{N} @ 50 dB	0.00197	0.01576
	Γ^{T} @ 50 dB	0.00055	0.25144
WINNER B3-LOS channel	Γ^{N} @ 10 dB	0.20797	0.80345
	Γ^{T} @ 10 dB	0.15029	11.68391
	Γ^{N} @ 30 dB	0.16130	0.58453
	Γ^{T} @ 30 dB	0.03242	8.11199
	Γ^{N} @ 50 dB	0.20947	0.59528
	Γ^{T} @ 50 dB	0.06372	8.24809

Table 5.1: Maximum relative curvatures for positioning for three different channel models with small and large GDOP at different SNRs.

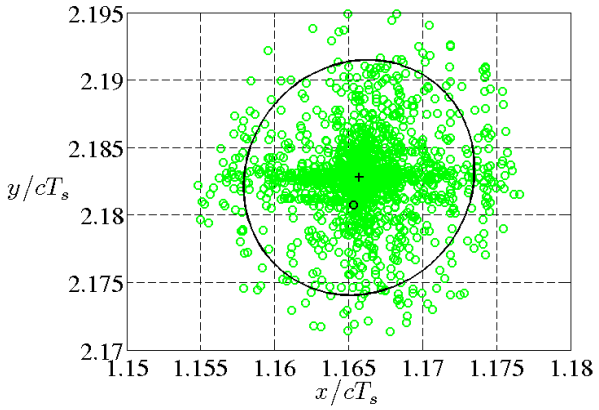
setup is applied for all SNRs and both geometric scenarios. In case of the WINNER channels, the B3-LOS model (large indoor hall) is used exemplarily at this point because it is most challenging. Results for the remaining WINNER channel models are presented in Appendix C (see Figure C.5 and Figure C.6 as well as Table C.1).

From Table 5.1, it can be observed that the GDOP has a major influence on the accuracy of the soft information methods: For small GDOP, all maximum relative curvatures are smaller than $\mathcal{C}^{1-\alpha}_{P,N-P}$, i.e. both soft information methods are accurate for *all* channel models at *all* SNRs in that case. This observation is confirmed by the confidence regions shown on the left hand side of Figure 5.9 to Figure 5.11: The confidence regions based on the linearisation method (black ellipses) and the likelihood method (green circles) are of the same shape. The likelihood regions are only slightly larger than the confidence regions based on the linearisation method. In most cases, the confidence regions are nearly circular, i.e. the x and y components of the position estimate are approximately

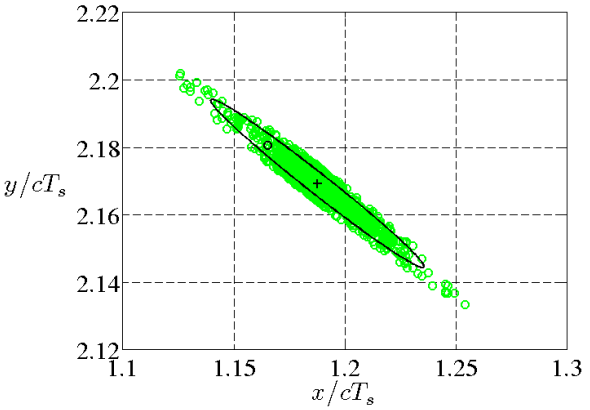


(a) Small GDOP at 10 dB.

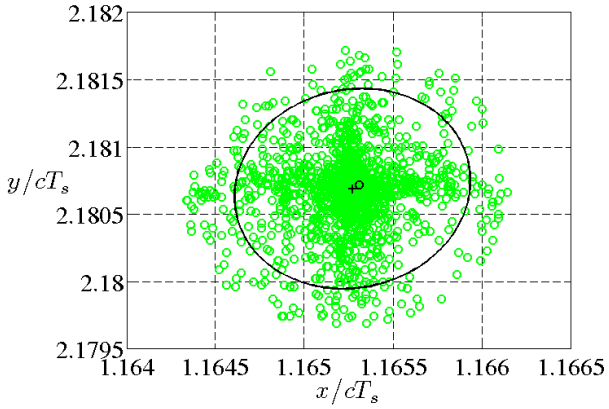
(b) Large GDOP at 10 dB.



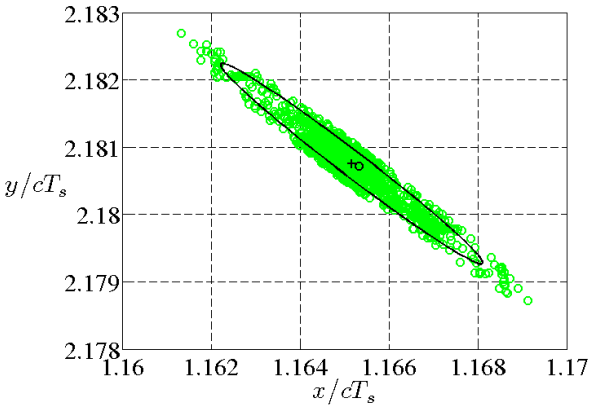
(c) Small GDOP at 30 dB.



(d) Large GDOP at 30 dB.

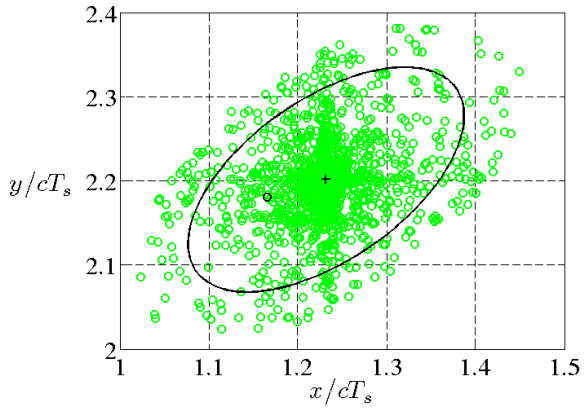


(e) Small GDOP at 50 dB.

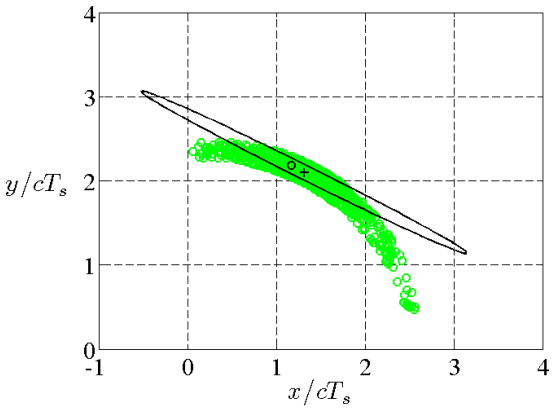


(f) Large GDOP at 50 dB.

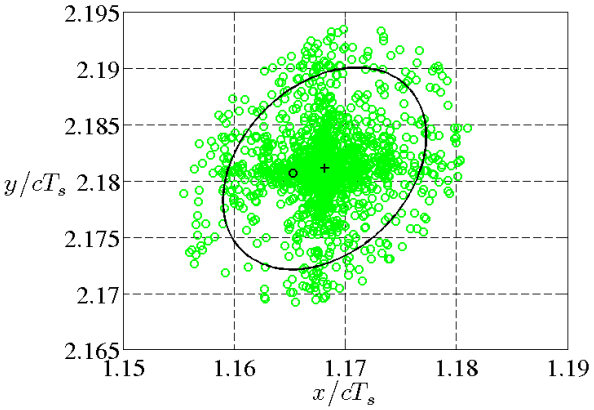
Figure 5.9: Confidence regions based on the linearisation method (black ellipse) and the likelihood method (green circles) for positioning in case of the LOS channel for different GDOPs and at different SNRs. The estimated parameters $\hat{\boldsymbol{\theta}}$ are denoted by a black cross and the true parameters $\boldsymbol{\theta}$ by a black circle.



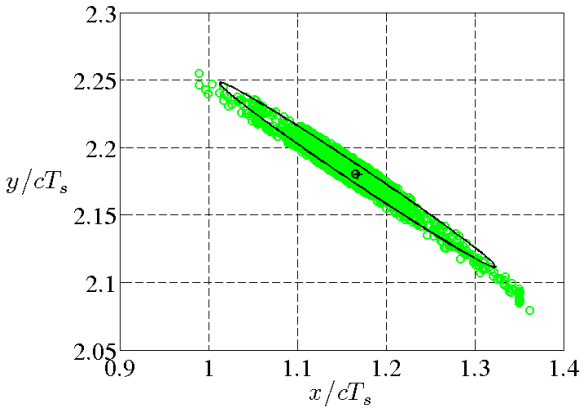
(a) Small GDOP at 10 dB.



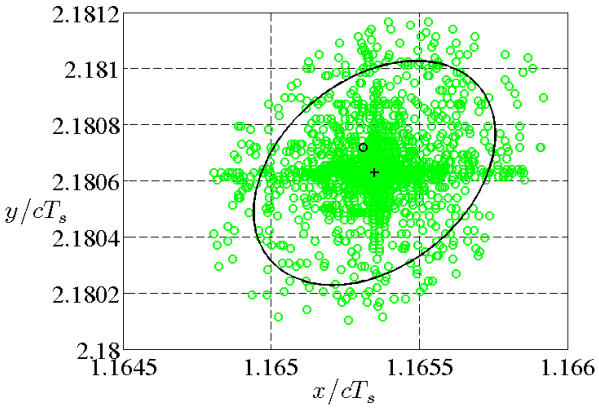
(b) Large GDOP at 10 dB.



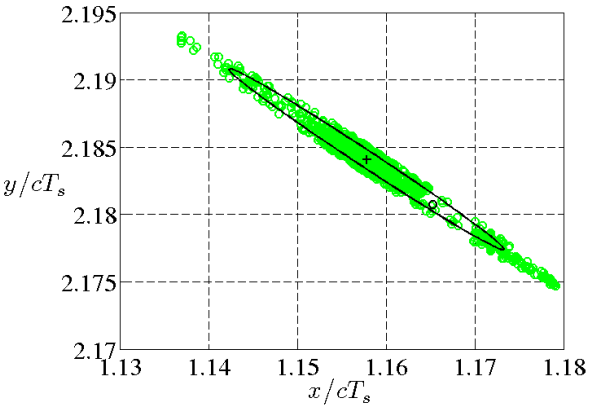
(c) Small GDOP at 30 dB.



(d) Large GDOP at 30 dB.

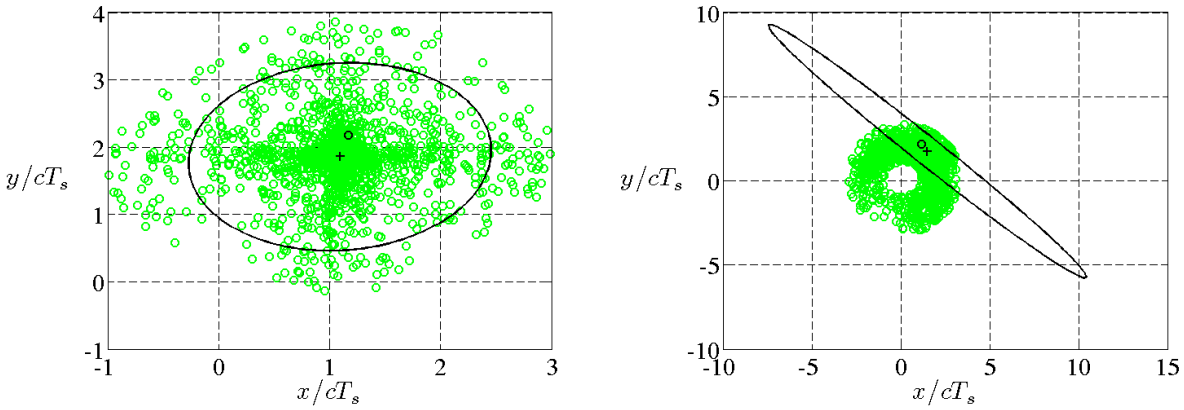


(e) Small GDOP at 50 dB.



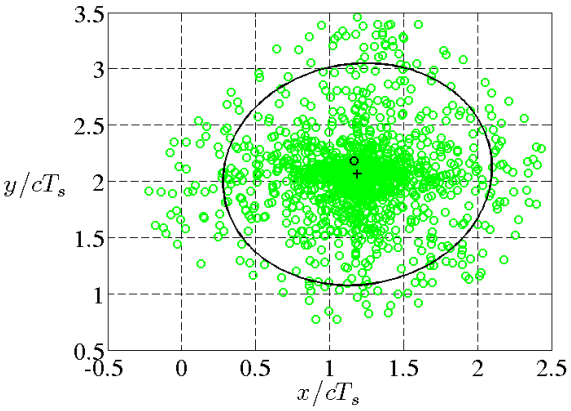
(f) Large GDOP at 50 dB.

Figure 5.10: Confidence regions based on the linearisation method (black ellipse) and the likelihood method (green circles) for positioning in case of a two-path channel for different GDOPs and at different SNRs. The estimated parameters $\hat{\boldsymbol{\theta}}$ are denoted by a black cross and the true parameters $\boldsymbol{\theta}$ by a black circle.

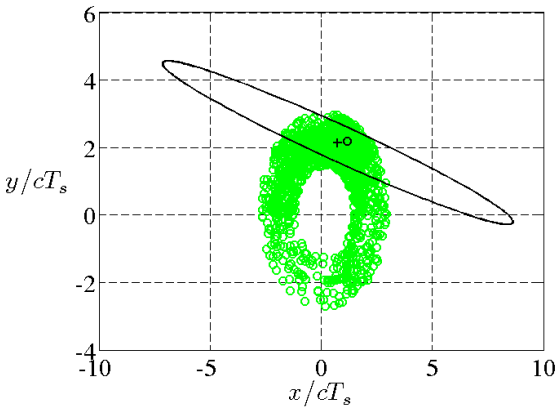


(a) Small GDOP at 10 dB.

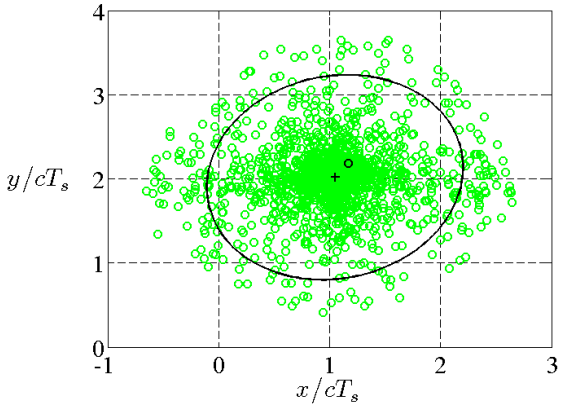
(b) Large GDOP at 10 dB.



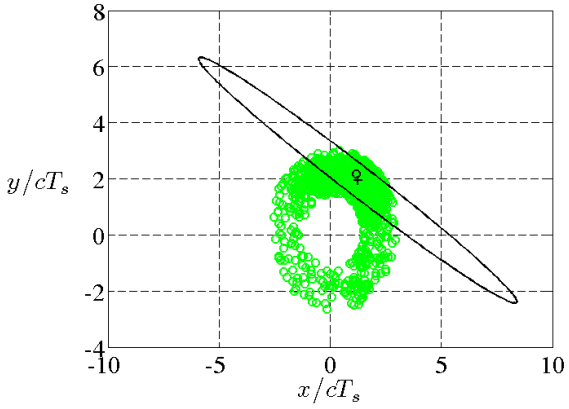
(c) Small GDOP at 30 dB.



(d) Large GDOP at 30 dB.



(e) Small GDOP at 50 dB.



(f) Large GDOP at 50 dB.

Figure 5.11: Confidence regions based on the linearisation method (black ellipse) and the likelihood method (green circles) for positioning in case of the WINNER B3-LOS channel for different GDOPs and at different SNRs. The estimated parameters $\hat{\boldsymbol{\theta}}$ are denoted by a black cross and the true parameters $\boldsymbol{\theta}$ by a black circle.

independent. In case of a large GDOP, a quite different behaviour is observed: The accuracy of the soft information methods now depends on the SNR and the current channel model: In case of the LOS channel, the uniform coordinate assumption is only violated at low SNR, while it is violated for all SNRs in case of the two-path channel model. This means that the linearisation method is *not* accurate for the LOS channel at low SNRs and for the two-path channels in general. In case of the WINNER B3-LOS channel, *both* assumptions are violated for all SNRs, i.e. neither the linearisation nor the likelihood method is accurate. These results are again confirmed by the corresponding confidence regions shown on the right hand side of Figure 5.9 to Figure 5.11. Especially at low SNR, the deviation between the linearisation and the likelihood method becomes obvious. For the LOS and the two-path channel, the likelihood regions are rather banana-shaped than elliptical. In case of the B3-LOS channel, the solution locus is extremely curved such that the likelihood region forms a closed circle. Furthermore, the confidence regions based on the linearisation method are extremely elliptical and far away from being circular. This means that there is a strong correlation between the x and y component of the position estimate.

By means of the above results, the great importance of a good GDOP is highlighted again. In case of a small GDOP, both soft information methods are accurate and, hence, the less complex linearisation method can be applied. Furthermore, the x and y component of the position estimate are nearly independent for small GDOPs. In case of a large GDOP, it is advantageous to use the likelihood method, but even the likelihood regions may be inaccurate for challenging channel models with severe multipath propagation. Consequently, the GDOP should be kept as low as possible also with respect to soft information.

5.5 Chapter Summary

In this chapter, the second step of localisation, namely position estimation based on the previously estimated channel parameters, is investigated. At first, the basic problem for localisation based on the TOA is introduced and a ML estimator as well as corresponding CRLBs are derived. The CRLB depends on the recent position of the MS and all ROs locations. Hence, the geometric configuration influences the CRLB. In order to separate the influence of the geometry from the influence of the pseudorange errors, the so-called GDOP is introduced for positioning, for which all pseudorange errors are assumed to have a unitary variance. Two different geometric scenarios with small and large average GDOP are applied throughout this chapter. Since the position estimation problem is again nonlinear, many of the insights already gained for channel parameter estimation can be exploited for positioning as well. For example, the LM algorithm can be applied to solve the nonlinear metric of the ML estimator. In the context of localisation, the LM method is denoted as TS algorithm. In comparison to channel parameter estimation, the positioning problem is less complicated. By a clever manipulation of the nonlinear set of equations, a linear set of equations can be obtained, that can be solved in closed form

with the help of a WLS approach. The performance of the proposed positioning algorithms is investigated. The iterative TS algorithm clearly outperforms the non-iterative WLS method, where the gap between both depends on the GDOP. In general, the positioning accuracy is strongly influenced by the GDOP: Significant performance gains can be obtained if the GDOP is kept low. The positioning accuracy for positioning without soft information and with soft information from the linearisation as well as the likelihood method is examined and compared. Against the prediction of Section 4.5.4, positioning based on the less accurate linearisation method outperforms positioning based on the more robust likelihood method. The potential inaccuracy of the linearisation method does not influence the positioning accuracy because a correct ratio of the weights in $\boldsymbol{W}$ and $\boldsymbol{\mathcal{W}}$ is rather important than accurate absolute values of the weights. Thus, rough soft information from the linearisation method is sufficient as long as the ratio of the pseudo-range variances is accurate. For the two-path channel with small excess, this condition is very well fulfilled: a performance gain of approximately 13 dB can be obtained with the help of soft information based on the linearisation method in comparison to unweighted positioning. For the WINNER channel models with many propagations paths and severe multipath propagation, this gain is not present anymore. Actually, soft information even complicates position estimation in case of the WINNER channel models. Hence, it is proposed to apply unweighted positioning in that case. Finally, soft information with respect to position estimates, that can be exploited for sensor fusion and/or tracking, is investigated. Again, the accuracy of the linearisation and the likelihood method are analysed with the help of the curvature measures of Bates and Watts. In contrast to channel parameter estimation, not only a single variance, but the whole confidence region, that illustrates the interdependencies between the variables, is of interest now. The accuracy of the soft information methods as well as the degree of correlation between the x and y components of the position estimates depend on the GDOP: For small GDOPs, both soft information methods are accurate for *all* SNRs and *all* channel models. Hence, the less complex linearisation method can be safely applied. For large GDOPs, even the likelihood method may fail. A main conclusion that can be drawn from this chapter is that the GDOP is of great importance: If the GDOP is kept low, not only more precise position estimates are obtained, but also the soft information of these estimates is more reliable. Consequently, the GDOP should be reduced by any means.

Chapter 6

Overall System Performance

In the preceding chapters, the positioning part of the proposed joint communication and positioning system has been investigated systematically. In order to exclude effects that may be caused by interference of the data layers, pure training transmission (ρ=1) has been assumed so far. In this chapter, the overall system performance *including* data transmission is examined. Especially, the tradeoff between communication and positioning as well as potential synergy effects are of main interest in this context. As already mentioned in the introduction part of Chapter 3, a joint communication and positioning system should provide high data rates with low bit error rate (BER) and a high localisation accuracy at the same time. Consequently, the BER as well as the RMSE of the position estimate are the main performance measures that are utilised subsequently. Since the localisation part (parameter estimation+position estimation) of the proposed system concept may fail, the outage rate is a further performance measure.

An important parameter that influences the system performance is the data rate, which is described by the bit load $\mathcal{B}$=$M \cdot R$ and can be modified by adjusting either the number of data layers M or by changing the code rate R. As already discussed in Section 3.2, a serial concatenation of an FEC and an REP code is typically applied in the context of IDM. Furthermore, the importance of the repetition part for resolving mapping ambiguities has been explained. Since code design for IDM is out of the scope of this thesis, a pure REP code in conjunction with scrambling (i.e. uncoded IDM) is applied subsequently. Thus, the code rate R is given by the inverse of the spreading factor S, which leads to $\mathcal{B}$=M/S. Another important parameter is the ratio of the pilot layer power to the total power $\rho \in [0, 1]$, that controls the tradeoff between communication and positioning. The overall SNR is divided into two parts according to the given ρ:

$$\gamma_s = \underbrace{\rho \cdot \gamma_s}_{\gamma_p} + \underbrace{(1-\rho) \cdot \gamma_s}_{\gamma_d}, \tag{6.1}$$

where γ_p is the pilot-to-noise ratio and γ_d is the data-to-noise ratio, respectively. With decreasing ρ, less power is spent for the pilot symbols and more power is spent for the data symbols. According to (3.1), each data symbol carries $\mathcal{B}$=M/S bits. In order to

achieve a fair comparison between the BERs for different bit loads $\mathcal{B}$, the ratio between the power that is spent for a single bit and the noise power (i.e. the bit-to-noise ratio: γ_b=γ_d·$\mathcal{B}$) needs to be considered. For the following simulation results, γ_b is varied from 0 dB to 18 dB and γ_s as well as γ_p are obtained according to the recent parameters ρ, M and S. The corresponding relationships in logarithmic scale given by

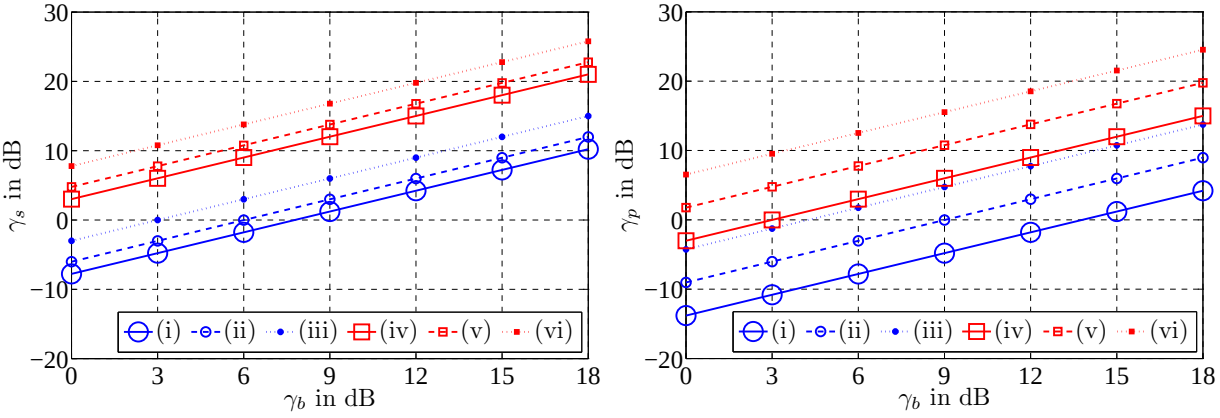
$$\gamma_s[\text{dB}] = \gamma_b[\text{dB}] + 10 \cdot \log\left(\frac{1}{1-\rho} \cdot \frac{M}{S}\right) \tag{6.2}$$

and

$$\gamma_p[\text{dB}] = \gamma_b[\text{dB}] + 10 \cdot \log\left(\frac{\rho}{1-\rho} \cdot \frac{M}{S}\right) \tag{6.3}$$

are illustrated in Figure 6.1. Of course, higher pilot-to-noise ratios γ_p are obtained for a larger ρ. Furthermore, the pilot-to-noise ratio γ_p as well as the signal-to-noise ratio γ_s increase with the bit load $\mathcal{B}$=M/S.

Before assessing the overall system performance, some considerations concerning the receiver structure of the proposed joint communication and positioning system need to be taken into account. The basic receiver structure for IDM with PLACE is illustrated for symbol-rate sampling (J=1) in Figure 3.8. The investigations in Section 4.3 have shown that oversampling is beneficial with respect to channel parameter estimation because it provides a performance gain compared to symbol-rate sampling. Hence, an oversampling factor of J=2 is generally applied for channel parameter estimation in this thesis. Now, the question is, how data detection is influenced by oversampling? It has already been explained in Section 3.2 that optimal maximum a posteriori (MAP) detection at the multi layer detector (MLD) is generally not feasible for frequency-selective channels without perfect channel knowledge. Hence, suboptimal detection strategies need to be applied



(a) Relationship between γ_b and γ_s.

(b) Relationship between γ_b and γ_p.

Figure 6.1: Relationship between γ_b and γ_s as well as γ_p. The labelling of the curves is as follows: (i) M/S=1/8 with ρ=0.25, (ii) M/S=1/8 with ρ=0.50, (iii) M/S=1/8 with ρ=0.75, (iv) M/S=12/8 with ρ=0.25, (v) M/S=12/8 with ρ=0.50, (vi) M/S=12/8 with ρ=0.75.

at the MLD. A Gaussian approximation-based detector with LLR combining (see Section 3.2) is used in this thesis because it offers a good tradeoff between complexity and performance. For LLR combining, the effective noise terms that correspond to a symbol of interest are assumed to be independent. In this case, all LLRs concerning the symbol of interest can be simply summed up. In general, the assumption of uncorrelated effective noise terms is not fulfilled due to the correlation between neighbouring received samples introduced by the frequency-selective channel (i.e. ISI). The stronger the correlation between adjacent effective noise samples, the more the performance of LLR combining degrades. Consequently, oversampling is disadvantageous in the context of LLR combining because the introduced correlation increases with the oversampling factor J. However, for moderate bit loads in conjunction with symbol-rate sampling, LLR combining performs nearly optimal. This means that the requirements for communication and positioning are in conflict: On the one hand, oversampling provides a performance gain for positioning and, on the other hand, oversampling degrades the performance for communication

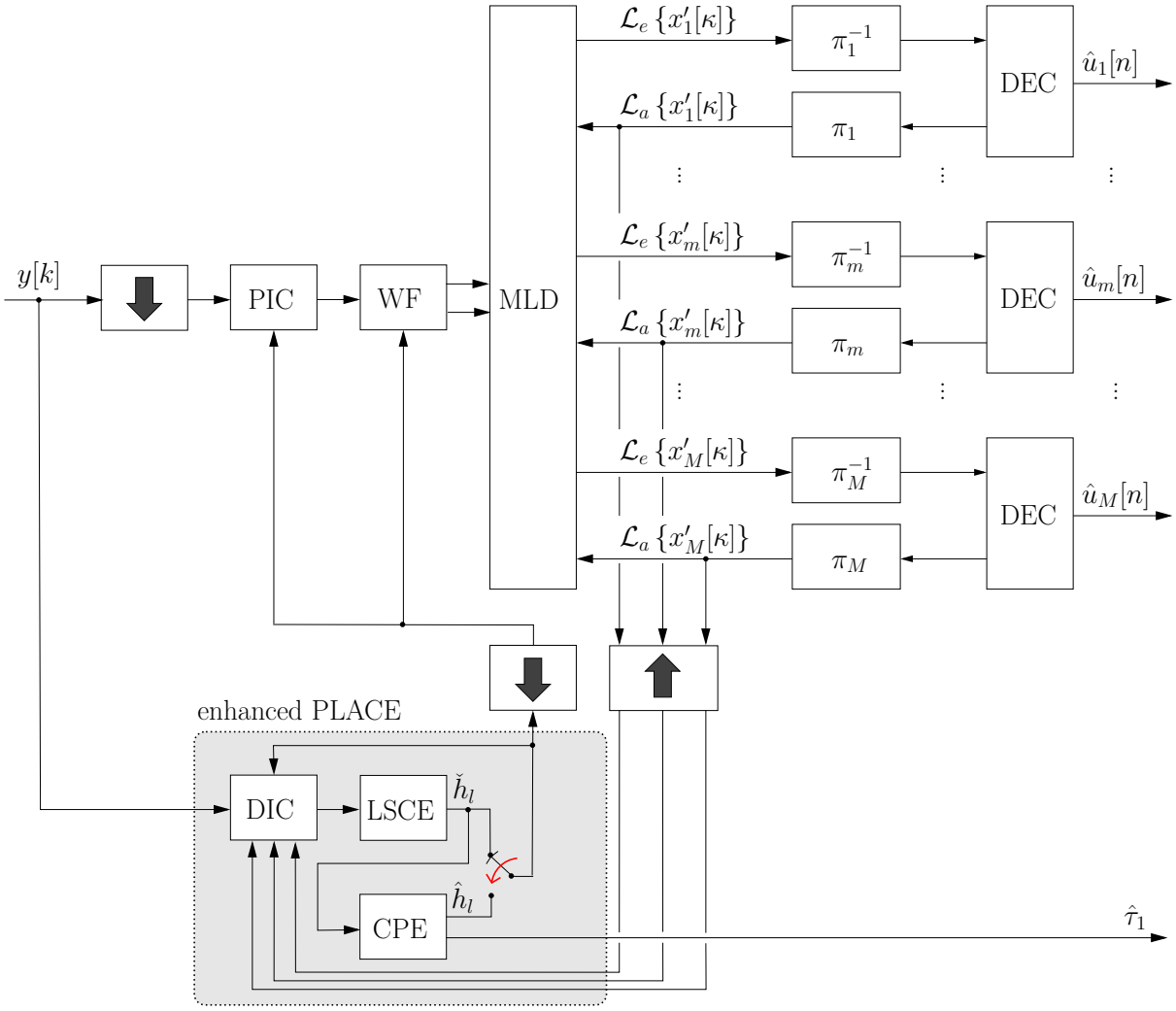


Figure 6.2: Receiver structure for the joint communication and positioning system.

purposes. This conflict can be resolved if up- and down-sampling is introduced into the receiver structure as shown in Figure 6.2: Channel parameter estimation is performed in the oversampled domain, whereas data detection is accomplished in the symbol-rate sampled domain. In this way, the benefits of oversampling for channel parameter estimation are preserved, while the complexity for data detection is kept low because the MLD with LLR combining is still applicable. In addition to up- and down-sampling, there are some further modifications of the receiver structure in Figure 6.2 compared to the basic receiver structure in Figure 3.8. Not only the enhanced PLACE unit is displayed in detail, but also a *whitening filter (WF)* is inserted before the MLD. As already mentioned above, the performance of LLR combining degrades with increasing correlation between adjacent efficient noise samples, which contain interference (MLI and ISI) as well as noise. If the noise is white[1], the correlation between adjacent efficient noise samples is only influenced by the interference part. However, in general, the noise is coloured. Hence, the correlation between adjacent efficient noise samples is influenced by the noise part as well. By introducing a WF, negative effects caused by coloured noise can be mitigated. The design of the WF is similar to the whitening approach used for least-squares channel estimation (LSCE) and channel parameter estimation (CPE) [Kay93, p. 95f.]: The received signal according to (2.43),

$$y[k] = \sum_{l=0}^{L} h_l \cdot x[k-l] + n[k], \qquad 0 \le k < K+L,$$

can be described in vector/matrix notation as:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{6.4}$$

where $\boldsymbol{y}=[y[0],\ldots,y[K+L-1]]^T$ is the observation vector containing the received samples and $\boldsymbol{H}$ is the channel matrix with Toeplitz structure:

$$\boldsymbol{H} = \begin{bmatrix} h_0 & 0 & \ldots & 0 & 0 \\ h_1 & h_0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ h_L & h_{L-1} & \ldots & 0 & 0 \\ 0 & h_L & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & h_L & h_{L-1} \\ 0 & 0 & \ldots & 0 & h_L \end{bmatrix}. \tag{6.5}$$

Furthermore, $\boldsymbol{x}=[x[0],\cdots,x[K-1]]^T$ and $\boldsymbol{n}=[n[0],\ldots,n[K+L-1]]^T$ denote the symbol vector and a zero mean Gaussian noise vector with covariance matrix $\boldsymbol{C_n}$, respectively. The covariance matrix $\boldsymbol{C_n}$ is symmetric and positive definite [Kay93, p. 95f.]. Hence, it can be factored as

$$\boldsymbol{C_n} = \boldsymbol{L_n} \cdot \boldsymbol{L_n}^T \tag{6.6}$$

[1] White noise is obtained if a square root Nyquist pulse is applied at the receiver side in combination with symbol-rate sampling.

according to the Cholesky decomposition [PTVF02, p. 99ff.]. $\boldsymbol{L_n}$ and $\boldsymbol{L_n^T}$ are a lower and an upper triangular matrix, respectively. By multiplying the received vector with the inverse of $\boldsymbol{L_n}$, a whitened signal model,

$$\underbrace{\boldsymbol{L_n^{-1}y}}_{\boldsymbol{y}_{\mathrm{WF}}} = \underbrace{\boldsymbol{L_n^{-1}H}}_{\boldsymbol{H}_{\mathrm{WF}}}\boldsymbol{x} + \underbrace{\boldsymbol{L_n^{-1}n}}_{\boldsymbol{n}_{\mathrm{WF}}}, \tag{6.7}$$

is achieved since the noise samples after the transformation, $\boldsymbol{n}_{\mathrm{WF}}$, are white:

$$\begin{aligned} \mathrm{E}\left\{\boldsymbol{n}_{\mathrm{WF}} \cdot \boldsymbol{n}_{\mathrm{WF}}^T\right\} &= \mathrm{E}\left\{(\boldsymbol{L_n^{-1}n}) \cdot (\boldsymbol{L_n^{-1}n})^T\right\} \\ &= \mathrm{E}\left\{\boldsymbol{L_n^{-1}n} \cdot \boldsymbol{n}^T(\boldsymbol{L_n^{-1}})^T\right\} \\ &= \boldsymbol{L_n^{-1}C_n}(\boldsymbol{L_n^{-1}})^T = \underbrace{\boldsymbol{L_n^{-1}L_n}}_{\boldsymbol{\mathcal{I}}} \cdot \underbrace{\boldsymbol{L_n^T}(\boldsymbol{L_n^T})^{-1}}_{\boldsymbol{\mathcal{I}}} = \boldsymbol{\mathcal{I}}. \end{aligned} \tag{6.8}$$

The whitened received samples and the whitened channel coefficients according to (6.7) are fed into the MLD (see Figure 6.2).

Another important aspect of the revised receiver structure in Figure 6.2 is that there is a switch in the enhanced PLACE unit: Either the channel coefficients for LSCE $\check{\boldsymbol{h}}$ *or* the channel coefficients for CPE $\hat{\boldsymbol{h}}$ can be used, i.e. CPE does not have to be performed in each receiver iteration. In the beginning, only LSCE is applied. Through the information exchange between channel estimation and data detection, the reliability of the pre-stage channel estimates and the soft data estimates increases over receiver iterations: With more reliable soft data, the interference of the data layers can be better cancelled before channel estimation is performed (DIC). In this way, more reliable channel estimates are obtained, which in turn lead to improved soft data estimates. Only in the very last receiver iterations, CPE can be performed to "boost" the performance because the MSE for CPE is generally much lower than for LSCE due to the exploitation of a priori information (see Section 4.4.3). Because of improved channel estimates after CPE, it is expected that the BER for data detection improves as well (synergy). At the same time, positioning is enabled. For the first receiver iteration with CPE, the combination of the PSO and the LM method proposed in Section 4.4.3 is applied. If CPE fails according to the outage detection described in Section 4.4.3, the pre-stage channel estimates $\check{\boldsymbol{h}}$ are used and, in the next receiver iteration, the combination of PSO and LM method is applied again. For successful CPE, the improved channel estimates $\hat{\boldsymbol{h}}$ are utilised and the recent parameter estimate is used as initial guess for the LM method in the next receiver iteration, i.e. the PSO method is not required anymore.

Based on the receiver structure in Figure 6.2, the overall system performance of the proposed joint communication and positioning system is examined now. The simulation setup is basically the same as in Chapter 4 and Chapter 5: The quadratic region according to Figure 5.1 with $R{=}2{\cdot}d_s$ and $B{=}4$ is assumed. In each run, the MS's position $\boldsymbol{p}$ is determined randomly inside this area. Again, both geometric scenarios with small and large average GDOP are taken into account. For each link between the MS and a RO, IDM in conjunction with PLACE as well as equal power and phase allocation for the data

layers is used:

$$a_0 = \sqrt{\rho}, \qquad \xi_0 = 0, \tag{6.9}$$

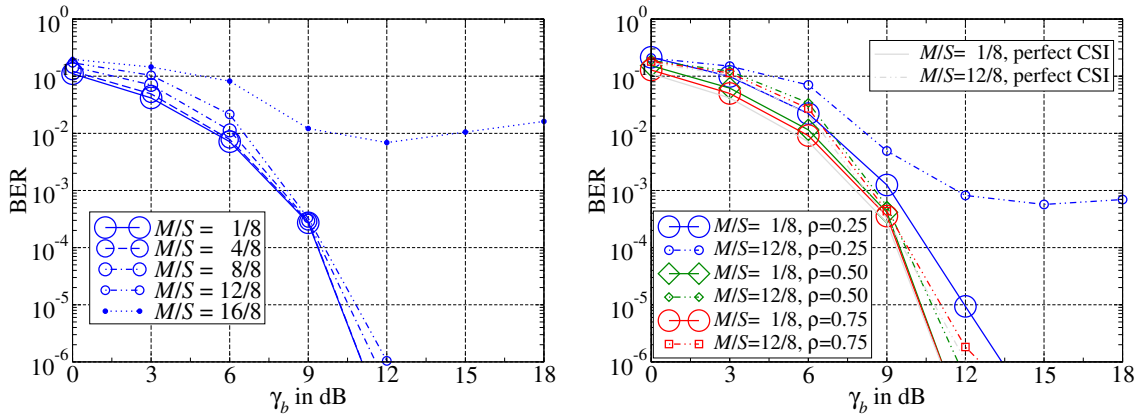
$$a_m = \sqrt{\frac{1-\rho}{M}}, \quad \xi_m = m \cdot \frac{\pi}{M+1}, \quad 1 \leq m \leq M. \tag{6.10}$$

Different bit loads $\mathcal{B}=M/S$ and pilot layer powers $\rho<1$ are considered. The spreading factor of the REP code is fixed to $S=8$ and the bit load is adjusted by changing the number of data layers M. Since the interleavers π_m, $1\leq m\leq M$, are the only means to separate the layers, they must be mutually different (see Section 3.2). Random interleavers are applied in this thesis. In contrast to Chapter 4 and Chapter 5, an increased burst length of $K'=320$ symbol durations is considered. Hence, $N=40$ bits per layer are transmitted with each burst. All bit streams are generated randomly and are mutually independent. An increase of the burst length is necessary because the initial channel estimates (when DIC can not be performed since no soft information about the data symbols is available) significantly influence the overall system performance. If the quality of the initial channel estimates is too low, the iterative receiver is not able to converge anymore. With decreasing ρ, the quality of the initial channel estimates reduces because the data layer interference increases. Hence, initial channel estimation is much more challenging compared to $\rho=1$. In order to ensure a certain quality of the initial channel estimates even for a small ρ, the burst length is increased.

For each link between the MS and a RO, different channel parameters are determined according to the recent channel model (LOS, two-path, WINNER). A Gaussian pulse shape with an effective pulse width of $T_g=8T_s$ and a shift of $s=4T_s$ is applied. At the receiver side, data detection and channel (parameter) estimation are accomplished in the symbol-rate sampled ($J=1$) and the oversampled ($J=2$) domain, respectively (see Figure 6.2). 20 receiver iterations are performed before hard estimates ($\hat{\tau}_1$ and $\hat{u}_m[n]$, $0\leq n<N$, $1\leq m\leq M$) are output. CPE is only applied in the last three receiver iterations as already described above. For positioning, power control is assumed.

6.1 LOS Channel Model

As before, the first considered channel model is the LOS channel ($L'=9$) because it represents the best possible case. In order to get a first insight into the IDM performance with LLR combining, BER curves for a single link with different bit loads $\mathcal{B}=M/S$ are shown in Figure 6.3 for IDM with perfect channel state information (CSI) as well as IDM with LSCE. Of course, the performance with channel estimation degrades in comparison to the performance with perfect CSI. Hence, the performance with perfect CSI represents a lower bound. It is observed from Figure 6.3(a) that the receiver with LLR combining is able to support bit loads up to $\mathcal{B}=1.5$. The lower the bit load, the better is the performance. For $\mathcal{B}=2$, the receiver is not able to converge anymore at high γ_b and an error floor occurs. This problem is aggravated if channel estimation is introduced as illustrated



(a) IDM with perfect CSI.

(b) IDM with LSCE.

Figure 6.3: BERs for IDM with LLR combining over γ_b in dB.

in Figure 6.3(b): For ρ=0.25, an error floor already occurs for a bit load of $\mathcal{B}$=1.5. By increasing the pilot layer power (ρ>0.25), this error floor can be eliminated (see Figure 6.3(b)). In general, the performance is improved with increasing ρ. This highlights the importance of the initial channel estimates: Due to a short block length, the quality of the initial channel estimates is quite poor and dominates the overall performance. If the quality of the initial channel estimates is too low, the iterative receiver is not able to converge anymore and an error floor occurs. For large block lengths (e.g. K'>1000), the quality of the initial channel estimates is good even for very small ρ. Hence, the influence of the pilot layer power on the BER performance reduces with the block length K': With increasing K', the BER curves for different ρ become nearly the same (see Figure C.7 in Appendix C). However, only short block lengths are of interest in this thesis because the block fading assumption needs to be fulfilled for CPE. In this case, the choice of ρ is crucial even with respect to the BER performance according to Figure 6.3(b).

In a second step, the BERs and MSEs for IDM with CPE and without CPE (i.e. LSCE only) are compared for a single link in order to test if synergy is created. It is important to note that the MSEs for LSCE and CPE,

$$\mathrm{MSE}\left(\check{\boldsymbol{h}}\right) = \mathrm{E}\left\{\frac{1}{L+1}\sum_{l=0}^{L}\left|\check{h}_l - h_l\right|^2\right\}$$

and

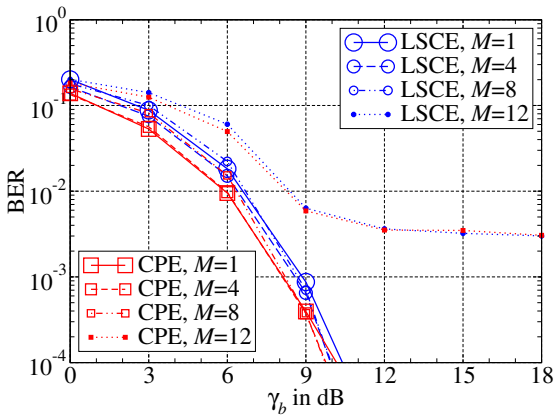
$$\mathrm{MSE}\left(\hat{\boldsymbol{h}}\right) = \mathrm{E}\left\{\frac{1}{L+1}\sum_{l=0}^{L}\left|\hat{h}_l - h_l\right|^2\right\},$$

as defined in (4.96) and (4.97), respectively, contain the factor $1/(L+1)$. This means that they correspond to the variance of a *single* channel coefficient. The overall MSE for all channel estimates is a $(L+1)$fold multiple of the MSE in (4.96) or (4.97). It is already known from Section 4.4.3 that the channel estimates $\hat{\boldsymbol{h}}$ after CPE are more reliable than the pre-stage channel estimates $\check{\boldsymbol{h}}$ obtained via LSCE. The question remains, whether this

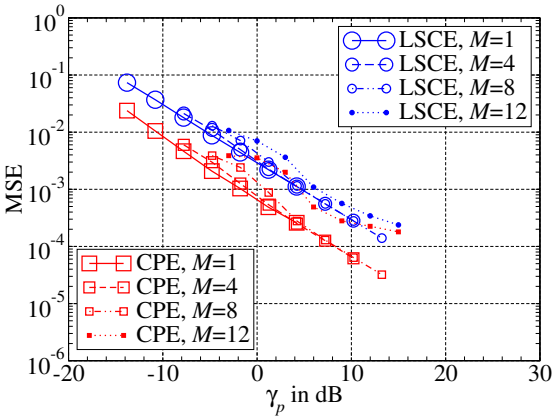
gain in MSE also improves the BER performance. Corresponding results for different pilot layer powers (ρ=0.25, 0.5, 0.75) and different bit loads (S=8, M=1, 4, 8, 12) are illustrated in Figure 6.4. On the left hand side, the BER curves are shown and, on the right hand side, the MSEs are plotted. The above observation that the overall performance improves with increasing pilot layer power is confirmed: Only for ρ=0.25 and $\mathcal{B}$=12/8=1.5, an error floor occurs. In all other cases, the receiver with LLR combining is able to converge. The MSEs decrease dramatically with increasing ρ and, hence, the BERs improve as well. The same is true for the comparison between LSCE and CPE: The MSE for CPE is much lower than the MSE for LSCE. Thus, the BERs for CPE are improved compared to the BERs for LSCE only, i.e. synergy between communication and positioning *is* created. However, the achievable gain with respect to the BER is rather small (<1 dB) and decreases with increasing ρ, i.e., especially for low pilot layer powers, CPE is able to improve the overall performance. Unfortunately, CPE is not able to eliminate an existing error floor (see Figure 6.4(a) and Figure 6.4(b)).

It has been shown in Section 4.4.3 that the gap between the MSE for LSCE and for CPE is largest in case of the LOS channel. A smaller gap is observed for all other channel models because CPE becomes more challenging with an increased number of channel parameters P. Consequently, the observed synergy effects are most pronounced in case of the LOS channel. Numerical results for the remaining channel models with respect to these synergy effects are illustrated in Figure C.8 to Figure C.12 in Appendix C. Generally, the basic behaviour remains the same, but the performance gain due to CPE is decreased in comparison to the LOS channel. In case of the WINNER B1-LOS channel model, LSCE may even outperform CPE at high γ_b because a reduced estimator, that introduces a systematic error, is applied for the WINNER channel models. In this case, the communication and positioning part should be separated, i.e. only LSCE is applied for data detection and, afterwards, CPE is performed once in order to enable positioning. However, if a full estimator is utilised, CPE generally outperforms LSCE due to the exploitation of a priori information.

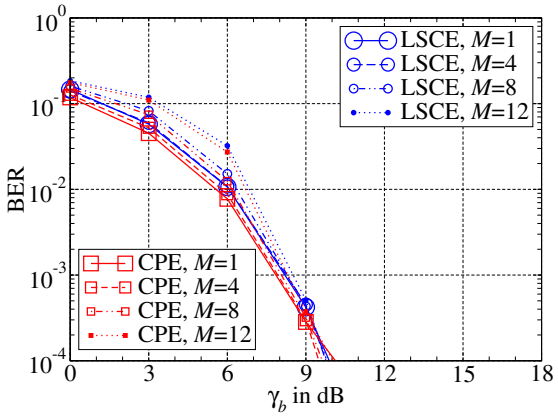
After this confirmation that synergy is created, the overall system performance including positioning is examined. Corresponding results for both geometric scenarios with small and large average GDOP are presented in Figure 6.5. BERs, RMSEs and outage rates are used as performance measures, which are compared for ρ=0.25, ρ=0.5 and ρ=0.75 with M=1 and M=8. The BERs shown in Figure 6.5(a) are averaged over all B=4 links and the outage rates illustrated in Figure 6.5(b) are determined similar to Section 5.3.3, i.e. positioning can be performed if at least three ROs are able to deliver pseudoranges $\hat{d}_b = \mathrm{c} \cdot \hat{\tau}_{b,1}$. Only if CPE for two or more ROs fails, positioning fails as well and the outage rate is increased. The BERs and the outage rates are valid for both geometric scenarios because they are not influenced by the GDOP and only depend on the underlying channel model. The RMSEs for positioning normalised to $d_s = \mathrm{c}\, T_s$ are shown for small and large average GDOP in Figure 6.5(c) and Figure 6.5(d), respectively. In contrast to Section 5.3.3, the RCRLBs for positioning are not plotted anymore due to clarity reasons. In case of the LOS channel, the RMSEs coincide with the RCRLBs over the whole pilot-to-noise ratio (PNR) range. After this general description of Figure 6.5,



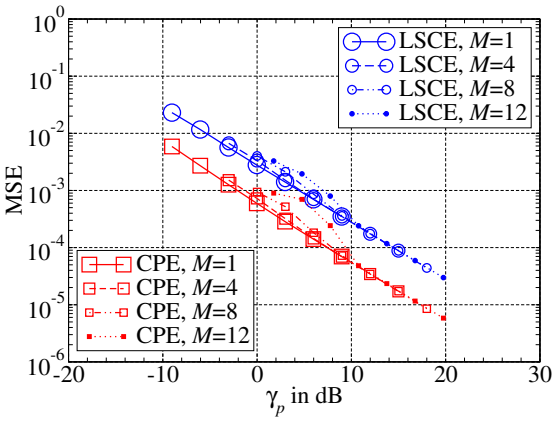
(a) BER with ρ=0.25.



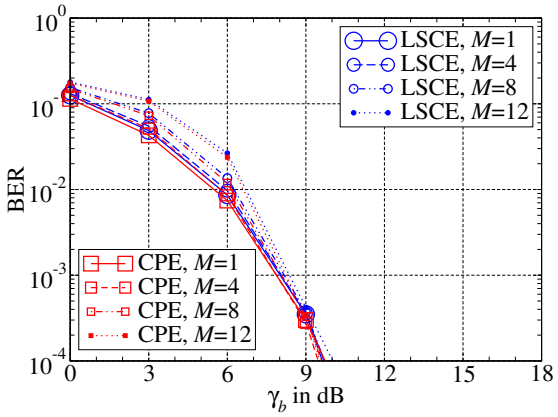
(b) MSE with ρ=0.25.



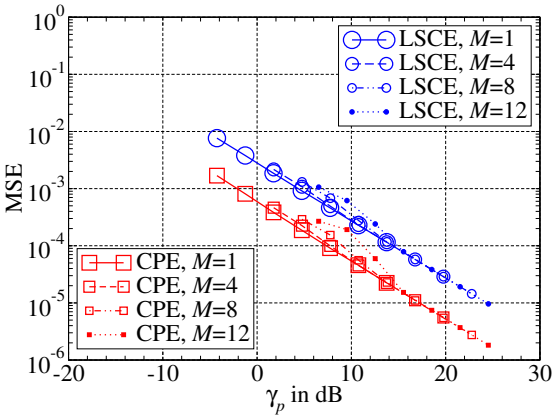
(c) BER with ρ=0.5.



(d) MSE with ρ=0.5.



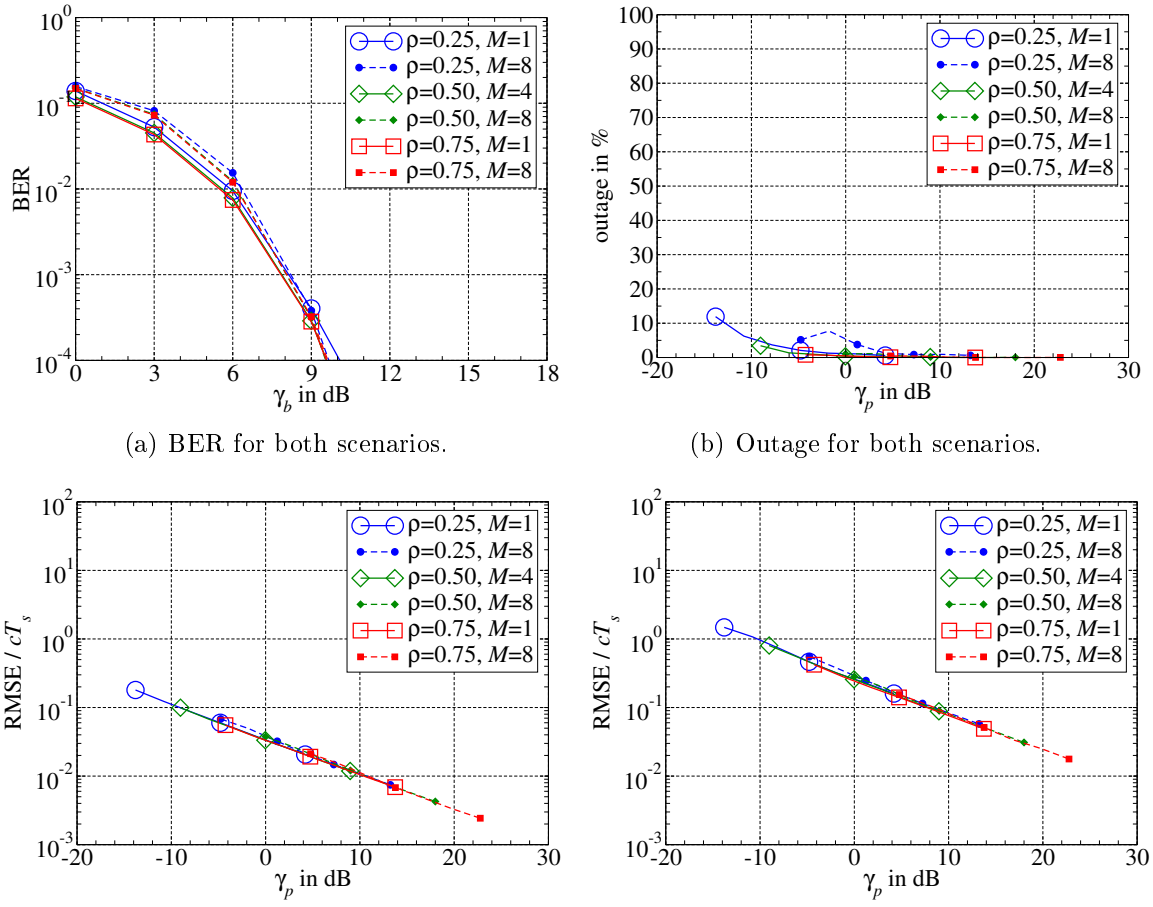
(e) BER with ρ=0.75.



(f) MSE with ρ=0.75.

Figure 6.4: Comparison of BERs as well as MSEs for LSCE and CPE with a spreading factor S=8 in case of a LOS channel.

the numerical results are now discussed in more detail: The BER curves in Figure 6.5(a) show a typical "waterfall" behaviour. For ρ=0.5 and ρ=0.75, the same BERs are obtained. Only for ρ=0.25, a slight performance degradation is observed. Furthermore, a small gap between the BERs for M=1 and M=8 is present that closes with increasing γ_b. In contrast, significant performance differences with respect to the positioning accuracy appear because the PNR strongly depends on ρ and M: Given γ_b=0 dB, a PNR of γ_p=−13.8 dB is obtained for ρ=0.25 and M=1, while the PNR is increased to γ_p=4.77 dB in case of ρ=0.75 and M=8. This means that the PNR for both cases differs by approximately 18.6 dB. Therefore, a much higher positioning accuracy is obtained for ρ=0.75 and M=8 than for ρ=0.25 and M=1: The corresponding RMSEs approximately differ by a factor of ten (e.g. a positioning error of 3 m *or* 30 m is obtained). Furthermore, the positioning accuracy is dominated by the GDOP as already concluded in Chapter 5: For a small average GDOP, the normalised RMSEs are well below 10^0 even for a very small PNR, i.e. a resolution much smaller than the distance d_s corresponding to the symbol duration T_s is achieved. For a large average GDOP, the RMSEs are only below 10^0 for γ_p>−10 dB.



(a) BER for both scenarios.

(b) Outage for both scenarios.

(c) RMSE for small average GDOP.

(d) RMSE for large average GDOP.

Figure 6.5: Results for joint communication and positioning based on IDM with a spreading factor S=8 in case of a LOS channel.

Significant performance gains can be obtained via a reduction in GDOP.

In contrast to the results in Section 4.4.3 and Section 5.3.3, outage detection is also applied for a LOS channel. For γ_p>0 dB, the outage rates are close to zero and, hence, negligible. But with decreasing PNR, positioning may even fail in case of the LOS channel: Due to large noise and data layer interference, the parameter estimate $\hat{\boldsymbol{\theta}}$ after the LM method may not be inside the predefined search space. Since there is only a single optimum for CPE in case of the LOS channel, the bounds of the search space could be expanded in order to reduce the outage rate. However, for the remaining channel models, CPE is much more challenging, i.e. several local optima exist beside the global optimum. If the search space bounds would be expanded in this case, outage detection became less reliable because premature convergence to local optima with an adequate final fitness but far away from the global optimum could not be detected anymore. Consequently, the same boundary check as applied in Section 4.4.3 and Section 5.3.3 is utilised here.

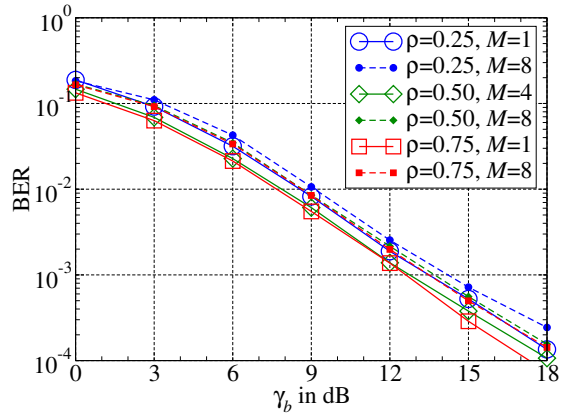
Based on the above results, it can be concluded that synergy between the communication and the positioning part is created even though the achievable performance gain with respect to BERs is quite small. For moderate bit loads, LLR combining performs well and a similar BER performance is obtained for different pilot layer powers and different bit loads. In contrast, the positioning accuracy changes considerably with different pilot layer powers and different bit loads. This means that the communication part is less susceptible to changes in the design parameters ρ and $\mathcal{B}$=M/S than the positioning part in case of a LOS channel.

6.2 Two-Path Channel Models

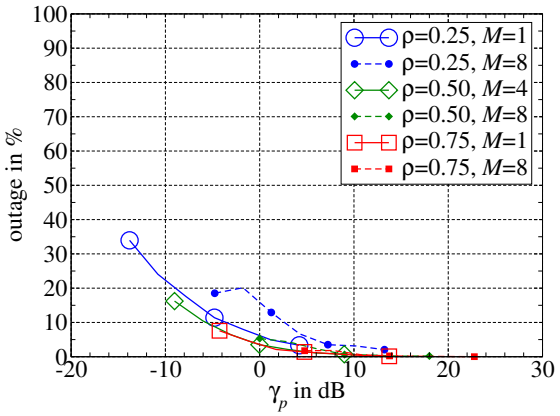
Now, it is examined whether the general behaviour observed for the LOS channel is preserved for the two-path channels. Positioning becomes more challenging due to an increased number of channel parameters P (LOS: P=3, two-path: P=6): For the two-path channels, CPE might get trapped in local optima, which lead to wrong pseudoranges. Data detection becomes more challenging as well because of an increased channel memory length L' (LOS: L'=9, two-path: L'=11), i.e. the total MSE of the initial channel estimates, that dominates the overall performance, becomes less reliable for the two-path channels. Furthermore, the correlation of neighbouring received samples introduced by the channel aggravates due to the second propagation path. This means that the independence assumption of LLR combining is less accurate than in case of the LOS channel.

As before, a two-path channel with small excess delay (ν_2/T_s∈[0.05, 1]) and a two-path channel with large excess delay (ν_2/T_s∈[1, 2]) are considered. The power ratio $\mathcal{P}$ as well as the phase offset $\Delta\Phi$ are generated randomly in the intervals [0.1,10] and [0,2π[, respectively. Soft information from the linearisation method is exploited for positioning in order to mitigate multipath propagation effects. Different pilot layer powers (ρ=0.25, 0.5, 0.75) and bit loads (S=8, M=1, 8) are considered. Corresponding results are shown in

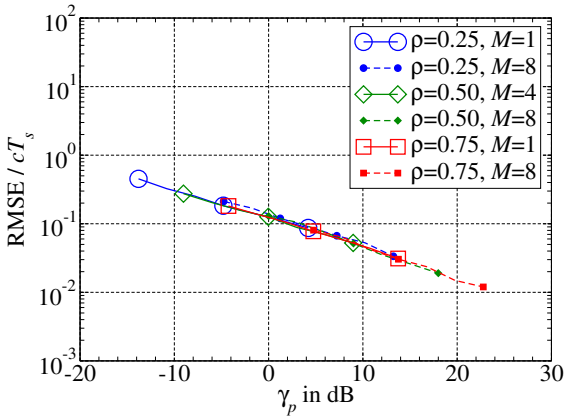
Figure 6.6 and Figure 6.7 for the channel with small and large excess delay, respectively. At first sight, it is observed that the BER performance degrades compared to the LOS channel. Especially for the channel with small excess delay, a significant performance degradation appears: The waterfall shape of the BER curves flattens such that a BER of 10^{-4} is *first* achieved around $\gamma_b \approx 18$ dB *even* for the transmission of a single layer (M=1, $\mathcal{B}$=0.125). This behaviour occurs due to two related effects caused by closely spaced multipath components. With decreasing excess delay, it becomes more difficult to separate adjacent multipath components, i.e. the components become non-resolvable. Hence, the signals of both paths overlap more and more and constructive as well as destructive superposition may appear. This means that fading effects similar to pure Rayleigh fading occur. The diversity of the channel reduces and the probability of a deep fade increases. A decrease in diversity causes an increase in correlation between adjacent channel coefficients. Hence, the correlation between adjacent received samples increases as well. Therefore, LLR combining becomes less accurate: The LLRs that are applied for the information exchange between MLD and DEC do not correspond to true LLRs.



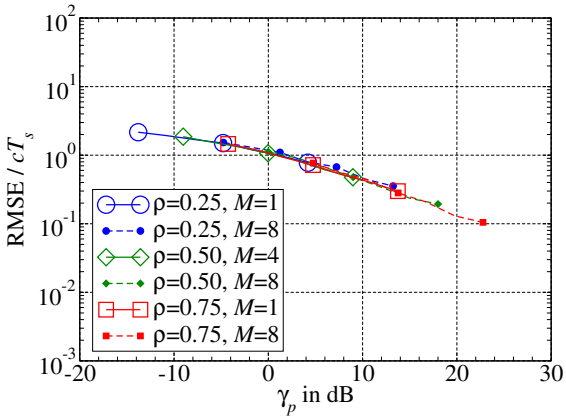
(a) BER for both scenarios.



(b) Outage for both scenarios.



(c) RMSE for small average GDOP.

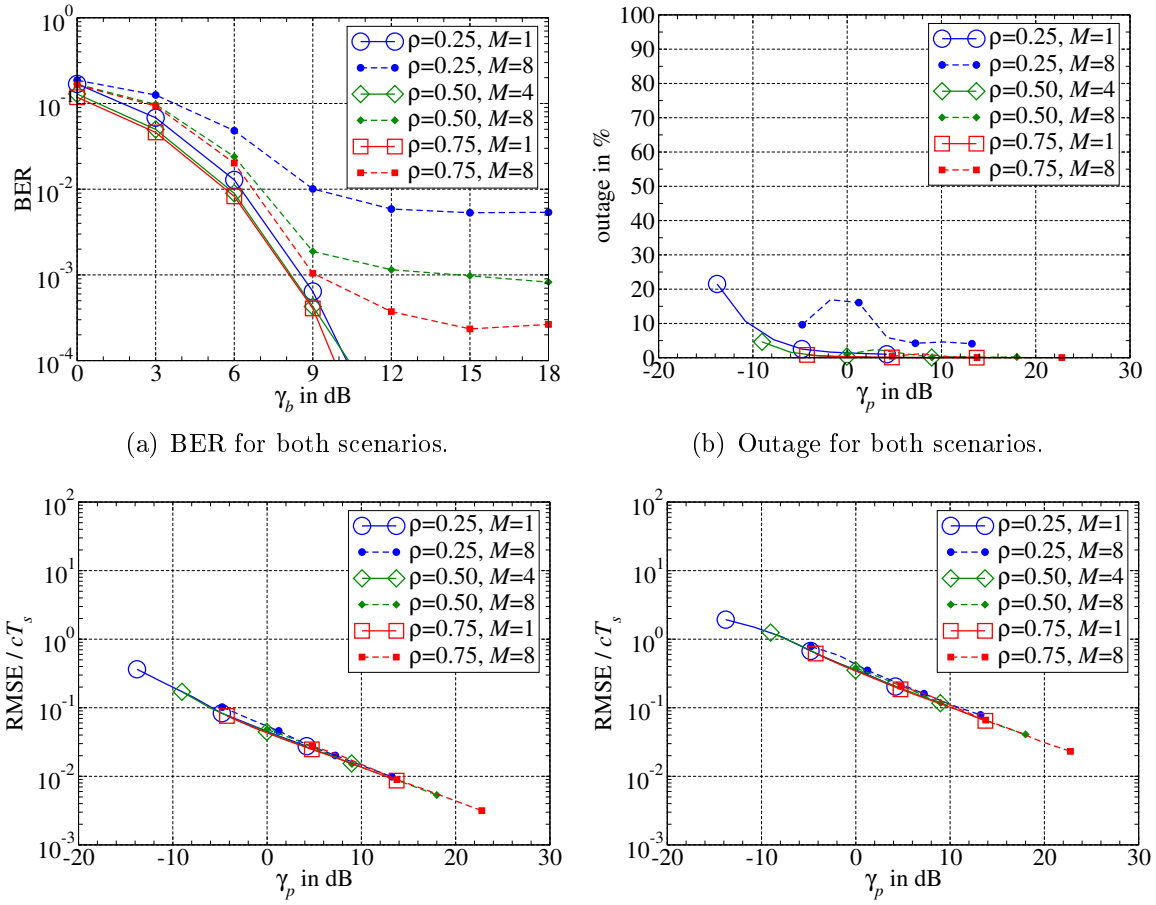


(d) RMSE for large average GDOP.

Figure 6.6: Results for joint communication and positioning based on IDM with a spreading factor S=8 in case of a two-path channel with small excess delay ($\nu_2/T_s \in [0.05, 1]$).

Typically, the LLRs are too optimistic such that wrongly estimated symbols may have an LLR with large magnitude but a wrong sign. This causes error propagation such that data detection degrades. Due to the decreased diversity and the increased correlation, the BER curves flatten even for the transmission of a single layer. In case of the channel with large excess delay, the waterfall shape of the BER curves is preserved. Hence, the decrease in diversity and the increase in correlation is limited for large excess delays. But for M=8 ($\mathcal{B}$=1), an error floor occurs for all pilot layer powers. These error floors appear due to the decreased quality of the initial channel estimates caused by a larger channel memory length L'.

With respect to positioning, the performance degradation in comparison to the LOS channel is less pronounced. Especially for the channel with large excess delay, the RMSEs are very close to the LOS performance because the propagation paths can be well separated. Furthermore, all RMSEs coincide with the RCRLB for large excess delays (not shown here). However, the outage rates increase due to the existence of local optima, in which CPE might get stuck in. For $\rho \geq 0.5$, the outage rates are negligible. For ρ=0.25,



(a) BER for both scenarios.

(b) Outage for both scenarios.

(c) RMSE for small average GDOP.

(d) RMSE for large average GDOP.

Figure 6.7: Results for joint communication and positioning based on IDM with a spreading factor S=8 in case of a two-path channel with large excess delay ($\nu_2/T_s \in [1, 2]$).

outage rates up to 20% are obtained at low PNR and the influence of data layer interference becomes obvious: For low to medium values of γ_b, the data layer interference can not be cancelled perfectly such that the MSE for channel estimation as well as the BER increase with the bit load (see also Figure C.9 in Appendix C). For large values of γ_b, the data layer interference can be mitigated almost perfectly such that the performance gap between low and high bit loads closes. This causes a "ridge" in the MSE curves for channel estimation with high bit loads that translates into a similar "ridge" for the RMSEs for positioning as well as the outage rates. For smaller pilot layer powers, this "ridge" is pronounced since the data layer interference increases. The same effect is observed for all other channel models.

For the channel with small excess delay, the positioning accuracy decreases compared to the channel with large excess delay and the outage rates at low PNR increase. However, in general, promising results are obtained even for small excess delays. For γ_p>3 dB, the outage rates are below 10% and, for a small GDOP, the normalised RMSEs are well below 10^0, i.e. a resolution much smaller than d_s is achieved. With decreasing PNR, the RMSE curves for the channel with small excess delay flatten due to a bias for CPE. Therefore, the RMSEs are below the RCRLBs at low PNR and only coincide with the RCRLBs at high PNR (not shown here).

From the above results, it can be concluded that both, the positioning *and* the communication part, depend on the underlying channel model. The first dependency was to be expected because the accuracy of CPE estimation strongly depends on the current channel parameters (excess delay, phase offset and power ratio) as demonstrated in Chapter 4. The second dependency is caused by a decrease in diversity and an increase in correlation in case of closely spaced multipath components. If only a single channel model is considered, the general behaviour already observed for the LOS channel is preserved: The communication part is less susceptible to changes in the design parameters ρ and $\mathcal{B}=M/S$ than the positioning part.

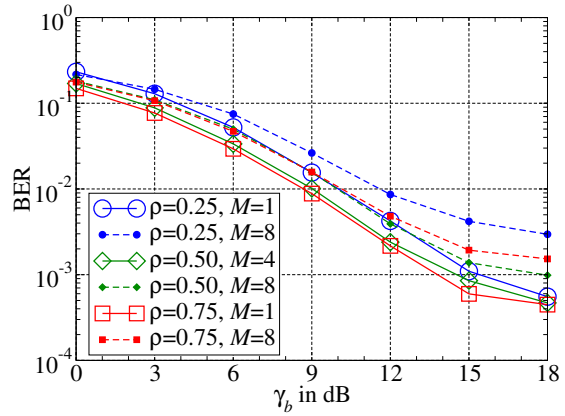
6.3 WINNER Channel Models

Finally, the performance of the proposed joint communication and positioning system is investigated for the realistic WINNER channel models. In order to keep CPE manageable, only a reduced estimator with three propagation paths as proposed in Section 4.4.3 is applied here. Positioning without soft information is performed for all three WINNER channel models (B1, B3, C2). Corresponding results for T_s=100 ns are shown in Figure 6.8 to Figure 6.10.

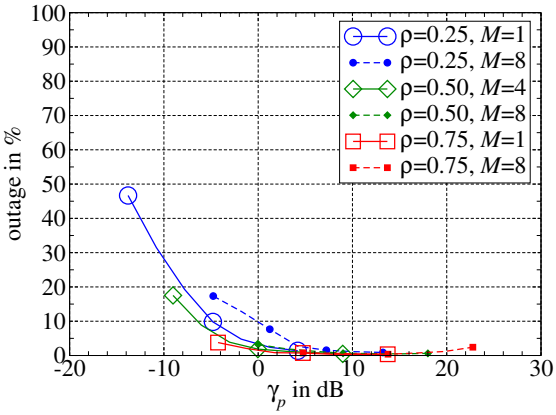
The best overall performance is obtained for the C2-LOS channel (urban macro-cell): The waterfall shape of the BER curves is preserved even though a BER of 10^{-4} is first reached around γ_b=13 dB. At the same time, the highest positioning accuracy compared to the remaining WINNER channel models is achieved. For the scenario with small

GDOP, the normalised RMSE varies between $4.1{\cdot}10^{-1}$ and $4.15{\cdot}10^{-2}$, which translates into absolute positioning errors of 12.3 m and 1.245 m in case of T_s=100 ns. In contrast, normalised RMSEs between 2.206 (66.18 m) and $3.5{\cdot}10^{-1}$ (10.5 m) are obtained for a large average GDOP, i.e. significant performance gains can be achieved by a GDOP reduction. The only drawback for the C2-LOS channel is that very large outage rates up to 90% occur for γ_p<0 dB, which is worse compared to the other two WINNER channels.

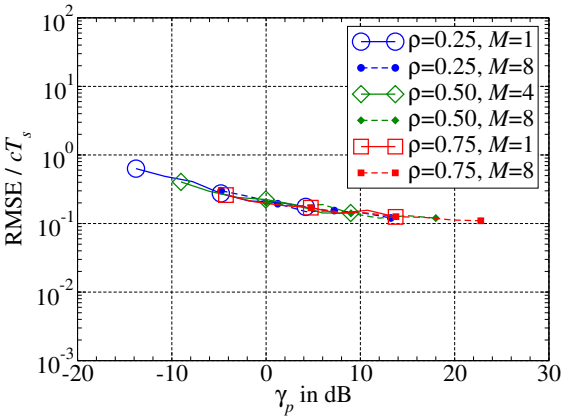
The worst overall performance is observed for the B3-LOS channel (large indoor hall), which is most challenging due to a large number of propagation paths with very small excess delays (see Table 4.2). A BER of 10^{-4} is not even reached at γ_b=18 dB and the achievable positioning accuracy varies between $8.705{\cdot}10^{-2}$ (2.61 m) for small average GDOP and 2.31 (69.3 m) for large average GDOP. The overall performance for the B1-LOS channel (urban micro-cell) lies between the two extremes of the B3-LOS channel and the C2-LOS channel. With respect to the BER, a special behaviour is observed for the B1-LOS channel: *All* BER curves show an error floor for $\gamma_b \geq 15$ dB. As already mentioned above, this performance degradation is caused by a systematic error introduced by the



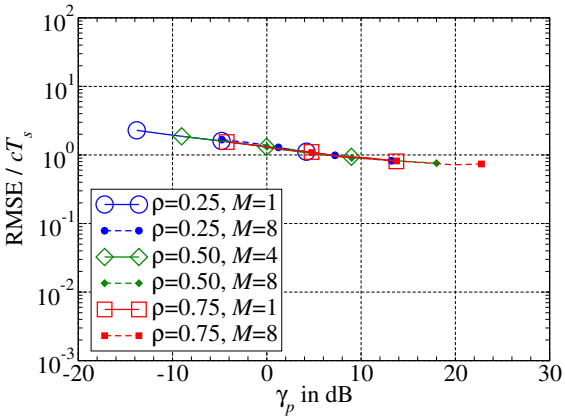
(a) BER for both scenarios.



(b) Outage for both scenarios.



(c) RMSE for small average GDOP.



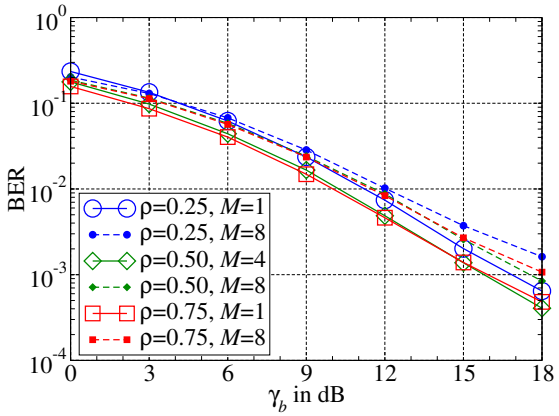
(d) RMSE for large average GDOP.

Figure 6.8: Results for joint communication and positioning based on IDM with a spreading factor S=8 in case of the WINNER B1-LOS channel.

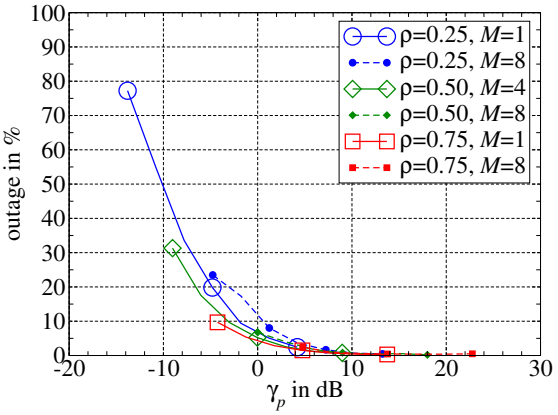
reduced estimator. If the communication part is separated from the positioning part (i.e. only LSCE is incorporated into data detection and CPE is performed at the very end), improved BER curves are obtained as shown in Figure C.10 in Appendix C. A BER of 10^{-4} is first reached at γ_b=16 dB in this case, while the positioning accuracy remains the same. In order to know whether CPE should be included into the iterative receiver or not, a (soft) block error rate (BLER) according to [HLS00], [Fri09],

$$\breve{P}_b = \sum_{m=1}^{M} \sum_{n=0}^{N-1} \frac{1}{1 + \exp\left(\left|\mathcal{L}_e\left\{u_m[n]\right\}\right|\right)}, \tag{6.11}$$

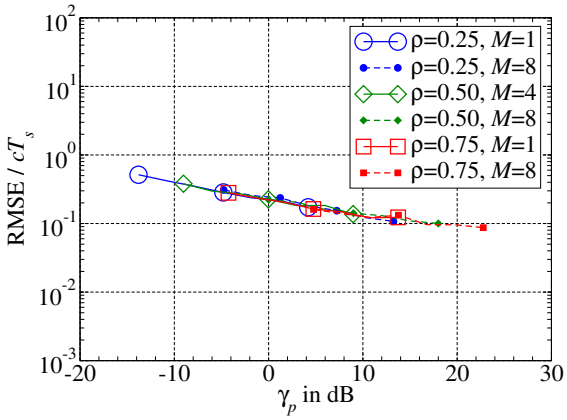
can be used even though the incorporated LLRs do not correspond to true LLRs. From the relationship between the BLER before and after the incorporation of CPE, it can be concluded whether CPE is advantageous or disadvantageous for data detection: When the BLER after the incorporation of CPE, $\breve{P}_b^{\mathrm{CPE}}$, is smaller than the BLER for the last LSCE iteration, $\breve{P}_b^{\mathrm{LSCE}}$, synergy is created and the BER is improved via CPE. If $\breve{P}_b^{\mathrm{CPE}} > \breve{P}_b^{\mathrm{LSCE}}$, communication and positioning should be separated. In this way, the best possible per-



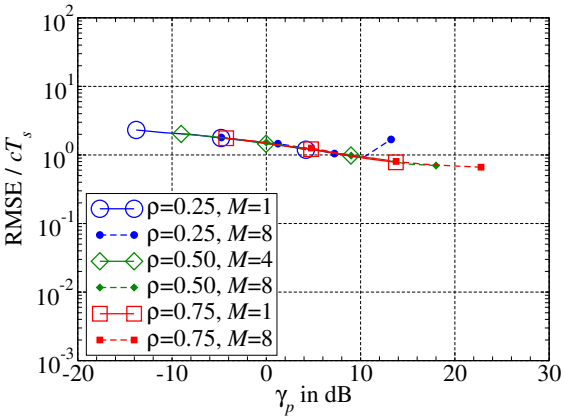
(a) BER for both scenarios.



(b) Outage for both scenarios.



(c) RMSE for small average GDOP.



(d) RMSE for large average GDOP.

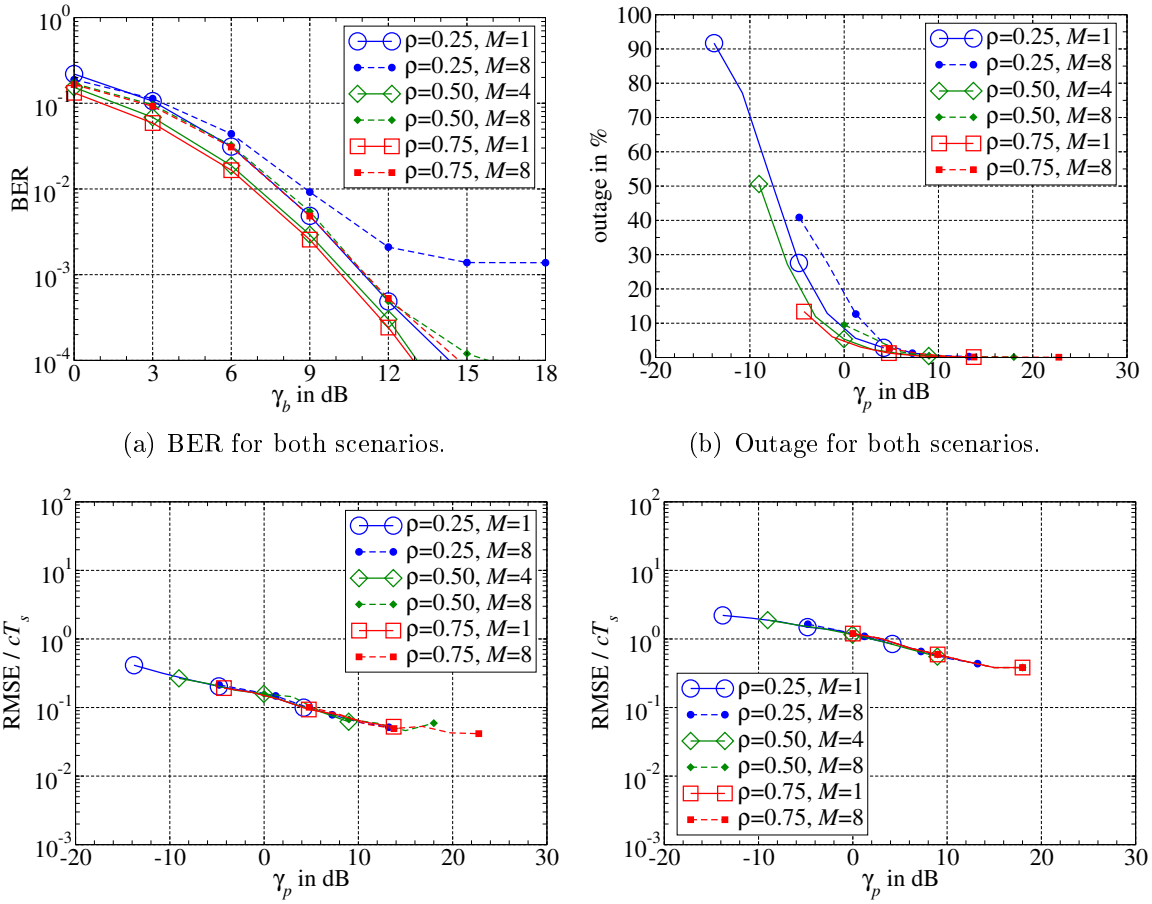
Figure 6.9: Results for joint communication and positioning based on IDM with a spreading factor S=8 in case of the WINNER B3-LOS channel.

formance can be achieved.

In summary, promising results are obtained even for the challenging WINNER channel models. Again, the importance of the GDOP is highlighted by the above results. Considering all three WINNER channel models, the positioning accuracy approximately varies between 2.31 (69.3 m) and $3.5{\cdot}10^{-1}$ (10.5 m) for large average GDOP, while it is improved to a range from $6.32{\cdot}10^{-1}$ (18.96 m) to $4.15{\cdot}10^{-2}$ (1.245 m) for small average GDOP. Hence, it is confirmed that the GDOP should be kept as low as possible.

6.4 Chapter Summary

In this chapter, the overall performance of the proposed joint communication and positioning system is investigated. At first, the relationship between γ_b, γ_p and γ_s is explained. A revised receiver structure incorporating up- and down-sampling is suggested in order to



(a) BER for both scenarios.

(b) Outage for both scenarios.

(c) RMSE for small average GDOP.

(d) RMSE for large average GDOP.

Figure 6.10: Results for joint communication and positioning based on IDM with a spreading factor S=8 in case of the WINNER C2-LOS channel.

provide the best possible performance for the communication part (J=1) as well as for the positioning part (J=2). Furthermore, a whitening filter is introduced to mitigate negative effects caused by the coloured noise. It is shown that synergy between the communication and the positioning part is generally created even though the achievable gain with respect to BERs is rather small. Promising results are obtained for all channel models. The positioning *and* the communication part depend on the underlying channel model. The more challenging a channel is for CPE, the less accurate data detection with LLR combining gets due to a decreased diversity and an increased correlation. Considering a single channel model, the communication part is less susceptible to changes in the design parameters ρ and $\mathcal{B}$=M/S than the positioning part. As in Chapter 5, the importance of a low GDOP for positioning is highlighted: For the same transmitter and receiver setup, the positioning accuracies for small and large average GDOP may differ by a factor of ten.

Chapter 7

Summary and Outlook

The combination of communication and positioning offers attractive features like improved resource allocation or tracking and guiding fire fighters on a mission. Hence, the interest in joint communication and positioning is steadily increasing. Especially in urban and indoor environments, a high localisation accuracy in the centimetre region is desirable. Unfortunately, positioning is most challenging in these environments due to severe multipath propagation. Hence, a high localisation accuracy can only be achieved by exploiting *all* possible localisation sources, whose position estimates are combined via sensor fusion. In this thesis, a system concept for joint communication and positioning is proposed whose positioning part contributes to the overall localisation accuracy via sensor fusion.

The system proposal for joint communication and positioning, that is introduced in Chapter 3, is based on interleave-division multiplexing (IDM) in combination with pilot layer aided channel estimation (PLACE) and localisation via the time of arrival (TOA): IDM offers a *flexible configuration* despite a *simple transmitter structure* and positioning based on the TOA enables *precise localisation*. A flexible configuration is especially desirable for a joint communication and positioning system because it is very difficult to provide high data rates with low bit error rate (BER) *and* accurate localisation at the same time. The connection between communication and positioning is accomplished via an enhanced PLACE unit: Not only the channel coefficients of the equivalent discrete-time channel model, that are needed for data detection, can be estimated based on the superimposed pilot layer, but also parameters of the physical channel, that enable positioning, can be extracted via channel parameter estimation. The overall system concept is investigated systematically in this thesis: At first, the localisation part, that can be further subdivided into *parameter estimation* and *position estimation*, is examined independently in Chapter 4 and Chapter 5, respectively. This means that negative effects that may be caused by data layer interference are excluded by assigning all available transmission power to the pilot layer (ρ=1, pure training transmission). In this way, phenomena that influence the positioning part can be analysed and potential countermeasures can be developed. After finding a promising setup with respect to the positioning part, the overall system performance *including* data detection (ρ<1) is examined in Chapter 6. At

all stages, the performance is examined for three different channel models that correspond to three different abstraction levels: (i) a line-of-sight (LOS) channel, that represents a best case scenario and provides a lower bound for the remaining channel models, (ii) two-path channel models, by which the influence of changing channel parameters like the excess delay or the phase offset between adjacent paths can be analysed, and (iii) wireless world initiative new radio (WINNER) channel models, that are applied to evaluate the performance in a realistic scenario. Altogether, promising results are achieved that are summarised subsequently.

Channel parameter estimation, as discussed in Chapter 4, is the core part of the proposed joint communication and positioning system. Hence, much effort is spent on the analysis of channel parameter estimation in this thesis. Two distinct maximum-likelihood estimators, that are based on received samples $\boldsymbol{y}$ and so-called pre-stage channel coefficients $\check{\boldsymbol{h}}$, respectively, are derived. It is shown that both estimators are equivalent and, hence, only the less complex estimator based on pre-stage channel coefficients $\check{\boldsymbol{h}}$ is further considered in this thesis. Before examining the performance of the proposed estimator, performance limits in terms of Cramer-Rao lower bounds (CRLBs) are derived. It is observed that oversampling provides a performance gain compared to symbol-rate sampling. Since the achievable gain is the same for all oversampling factors $J\geq 2$, moderate oversampling with $J=2$ is applied for channel parameter estimation in this thesis. Due to a nonlinear metric $\Omega(\tilde{\boldsymbol{\theta}})$, optimisation algorithms are required in order to find a parameter estimate $\hat{\boldsymbol{\theta}}$. Local as well as global optimisation algorithms are discussed. The performance of the Levenberg-Marquardt (LM), expectation-maximisation (EM) and space-alternating expectation-maximisation (SAGE) algorithm as well as the performance of the particle swarm optimisation (PSO) and cooperative particle swarm optimisation (CPSO) method are investigated with respect to channel parameter estimation. Due to many local optima of the metric $\Omega(\tilde{\boldsymbol{\theta}})$, it is very challenging to find the global optimum. First results that are analysed by means of boxplots illustrate that convergence to local optima is a basic problem for *all* optimisation methods. On the one hand, it is known that probabilistic methods like the PSO and CPSO method are incomplete and may suffer from premature convergence as well as stagnation. On the other hand, the performance of the local methods depends on the quality of the initial guess. Furthermore, the applied local methods are unconstrained such that they may converge to local optima outside the search space, which are far away from the global optimum and lead to large estimation errors. Without any modifications, the PSO method is best suited to channel parameter estimation because it is most robust with respect to outliers. However, these first results have shown that outlier detection is generally desirable for channel parameter estimation. Based on an analysis of the final fitness $\Omega(\hat{\boldsymbol{\theta}})$ for a perfect initial guess, an outlier detection method is proposed and successfully applied to channel parameter estimation. For the unconstrained local methods, an additional boundary check is required for reliable outage detection. Very good results are achieved for all methods in combination with outlier detection: In general, mean squared errors (MSEs) that lie on or near the CRLB are obtained. For small excess delays, MSEs even below the CRLB are possible at very low or very high signal-to-noise ratio (SNR). This is due to a bias at low SNR and due to large outage rates at high SNR. By comparing the performed iterations (boxplots) to

the required iterations (fitness histories), it is shown that the large outage rates at high SNR are caused by a too early interruption of the optimisation methods. Hence, the stopping criteria need to be adjusted in order to reduce the outage rates. It is shown that a stopping criterion based on the absolute fitness in conjunction with the fitness threshold, that is also applied for outlier detection, is best suited to match the performed iterations to the required iterations such that the outage rates are significantly reduced. A further performance improvement can be achieved by a sophisticated combination of the PSO and the LM method, which is applied in the remainder of this thesis. In case of the WINNER channel models, reduced estimators with less propagation paths than actually present are advantageous for symbol durations $T_s \geq 100$ ns. This means that closely spaced multipath components, that are very difficult to separate, are combined and the complexity of the estimator is reduced because less parameters are estimated compared to the full estimator. However, at high SNR, an error floor may occur due to a systematic error that is introduced by the combination of neighbouring propagation paths. Finally, soft information for channel parameter estimation is discussed in Chapter 4: Two different methods to obtain soft information are proposed and analysed with the help of the curvature measures of Bates and Watts. It is deduced that the likelihood method is more robust than the less complex linearisation method, i.e. it is more probable that the likelihood method is accurate.

In Chapter 5, position estimation based on the parameters obtained by channel parameter estimation is investigated. Again, a maximum-likelihood estimator is introduced and performance limits in terms of CRLBs are derived. In the context of position estimation, the so-called geometric dilution of precision (GDOP) is utilised in order to separate the influence of the geometric configuration from the influence of the pseudorange errors. Hence, two different scenarios with small and large average GDOP are considered. Position estimation is again a nonlinear problem. Thus, many of the insights already gained in Chapter 4 can be exploited in Chapter 5 as well. For example, the LM method, which is called TS algorithm in the context of positioning, can be used to solve the nonlinear metric of the maximum-likelihood estimator. However, position estimation is much simpler than parameter estimation such that closed form solutions can be derived for position estimation: The non-iterative WLS method performs worse than the iterative Taylor series (TS) algorithm, but it is much simpler. In this thesis, the weighted least-squares (WLS) method is used to generate an initial guess for the TS algorithm. The localisation accuracy with and without usage of soft information concerning the parameter estimates is investigated. In contrast to the prediction in Chapter 4, the linearisation method clearly outperforms the likelihood method because a correct ratio of the employed weights is rather important than accurate absolute values of the employed weights. Thus, the less complex linearisation method can be applied. In case of the two-path channel with small excess delays, a performance gain of approximately 13 dB is achieved for the linearisation method compared to unweighted positioning. Unfortunately, this gain is not preserved for the WINNER channel models. In contrast, the exploitation of soft information even complicates positioning in case of the WINNER channel models, i.e. unweighted positioning is advantageous. In general, the localisation accuracy is dominated by the GDOP. Significant performance gains can be achieved if the GDOP is kept low. As in the case

of channel parameter estimation, soft information with respect to the position estimates can be obtained. This soft information can be exploited for sensor fusion and/or tracking. This means that not only a single variance, but the whole confidence region, that illustrates the interdependencies between the variables, is of interest for position estimation. Again, the linearisation and the likelihood method are analysed with the help of curvature measures. It is shown that both methods are accurate for small average GDOP. For large GDOP, even the likelihood method may fail. Hence, the importance of a small average GDOP is again highlighted.

Based on the insights obtained from the previous results, the overall system performance is studied in Chapter 6. The basic IDM receiver structure is revised in order to provide the best possible performance for communication *and* positioning: On the one hand, up- and down-sampling is introduced such that data detection and channel (parameter) estimation can be performed in the symbol-rate sampled (J=1) and oversampled (J=2) domain, respectively. On the other hand, a whitening filter is introduced in order to mitigate negative effects caused by the coloured noise. Through the incorporation of channel parameter estimation into the iterative receiver structure, synergy effects occur because improved channel estimates are obtained via channel parameter estimation in comparison to standard least-squares channel estimation. However, the achievable synergy gain with respect to the BER is rather small ($\leq$1 dB). In general, promising results are achieved for all channel models. Both, the communication and the positioning part, depend on the underlying channel model: The more challenging a channel is for channel parameter estimation, the less accurate data detection with log-likelihood ratio (LLR) combining gets due to a decreased diversity and an increased correlation. But the communication part is less susceptible to changes in the design parameters ρ and $\mathcal{B}$=M/S than the positioning part. As before, the GDOP is of great importance: The localisation accuracies for small and large GDOP may differ by a factor of ten given the same transmitter and receiver setup. Consequently, the GDOP should be reduced in any way if possible.

In this thesis, the feasibility of the proposed joint communication and positioning system is demonstrated. In order to keep the investigations and results of this thesis manageable, some aspects have not been taken into account yet. For example, the effects of non line-of-sight (NLOS) propagation as well as adequate countermeasures need to be addressed in future work. Furthermore, the integration of the system proposal into tracking and/or sensor fusion algorithms is of interest. If a certain application is considered, the proposed system concept may be adjusted to its specific requirements, i.e. the system may be revised in order to improve its performance or in order to reduce its complexity. For example, other pulse shaping and receive filters may be implemented. Especially, square-root Nyquist filters that lead to white noise for symbol-rate sampling are of interest. Oversampling could be achieved by several parallel symbol-rate sampling filters with different phase offsets, whose outputs are adequately combined. In this way, advantages of white noise may be exploited. Furthermore, the complexity of channel parameter estimation could be reduced by the estimation of signal parameters via rotational invariance techniques (ESPRIT) [PRK85], [RK89] in conjunction with the SAGE algorithm. At the

stage of position estimation, a combination of TOA and angle of arrival (AOA) estimates may be considered: On the one hand, only a single reference object is required in this case, i.e. constraints with respect to the hearability of the positioning signals are relaxed compared to pure TOA localisation. On the other hand, multiple input multiple output (MIMO) aspects need to be taken into account at the channel parameter estimator. Last but not least, the communication part of the system proposal may be improved by a proper code design for IDM or by the application of a multi layer detector that takes the correlation introduced by the channel into account.

Appendix A

Acronyms and Abbreviations

A

AOA	angle of arrival
AOD	angle of departure
ASA	azimuth spread of arrival angles
ASD	azimuth spread of departure angles
ASK	amplitude shift keying
AWGN	additive white Gaussian noise

B

B&B	branch and bound
BER	bit error rate
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BICM	bit-interleaved coded modulation
BICM-SM	bit-interleaved coded modulation with superposition mapping
BLER	block error rate
BPSK	binary phase shift keying
BS	base station

C

CDF	cumulative distribution function
CDL	clustered delay line
CDM	code-division multiplexing
CDMA	code-division multiple access
CE	channel estimation
CPE	channel parameter estimation
CPSO	cooperative particle swarm optimisation
CRLB	Cramer-Rao lower bound
CSI	channel state information

D

DEC	decoder
DIC	data layer interference cancellation
DIRECT	divide rectangles
DS-CDMA	direct sequence code-division multiple access
DVB-T	digital video broadcasting - terrestrial

E

EM	expectation-maximisation
ENC	encoder
ESPRIT	estimation of signal parameters via rotational invariance techniques
EXIT	extrinsic information transfer

F

FEC	forward error correction
FDM	frequency-division multiplexing
fIG	fixed initial guess

G

GA	Gaussian approximation
GA	genetic algorithm
GB	global best
GDOP	geometric dilution of precision
GSM	global system for mobile communications
GPS	global positioning system

I

IB	individual best
IC	interference cancellation
IDM	interleave-division multiplexing
IDMA	interleave-division multiple access
IG	initial guess
ISI	inter-symbol interference

J

JGC	joint Gaussian combining

L

LB	local best
LLR	log-likelihood ratio
LLRC	log-likelihood ratio combining
LM	Levenberg-Marquardt
LOS	line-of-sight
LS	least-squares

	LSCE	least-squares channel estimation
	LTE	long term evolution
	LTI	linear time-invariant
	LTV	linear time-variant
M		
	MAP	maximum a posteriori probability
	MCS	multilevel coordinate search
	MIMO	multiple input multiple output
	ML	maximum-likelihood
	MLD	multi layer detector
	MLI	multi layer interference
	MMSE	minimum mean squared error
	MRC	maximum ratio combining
	MS	mobile station
	MSE	mean squared error
	MUD	multi user detector
	MVU	minimum variance unbiased
N		
	NLOS	non line-of-sight
	NLS	nonlinear least-squares
O		
	OFDM	orthogonal frequency-division multiplexing
P		
	PAPR	peak-to-average power ratio
	PDF	probability density function
	PIC	pilot layer interference cancellation
	pIG	perfect initial guess
	PLACE	pilot layer aided channel estimation
	PNR	pilot-to-noise ratio
	PSK	phase shift keying
	PSO	particle swarm optimisation
Q		
	QAM	quadrature amplitude modulation
	QOS	quality of service
R		
	RCRLB	root Cramer-Rao lower bound
	REP	repetition
	rIG	random initial guess

RMS	root mean squared
RMSE	root mean squared error
RO	reference object
RSS	received signal strength
Rx	receiver

S

SA	simulated annealing
SAGE	space-alternating expectation-maximisation
SE	squared error
SISO	single input single output
SM-EPA	superposition mapping with equal power allocation
SNR	signal-to-noise ratio

T

TDM	time-division multiplexing
TDOA	time difference of arrival
TOA	time of arrival
TOD	time of departure
TS	Taylor series
Tx	transmitter

U

UEP	unequal error protection
UMTS	universal mobile telecommunications system
US	uncorrelated scattering
UWB	ultra wideband

W

WF	whitening filter
WHERE	wireless hybrid enhanced mobile radio estimators
WHERE2	wireless hybrid enhanced mobile radio estimators (phase 2)
WINNER	wireless world initiative new radio
WLAN	wireless local area network
WLS	weighted least-squares
WSS	wide-sense stationary/wide-sense stationarity
WSSUS	wide-sense stationary uncorrelated scattering

Appendix B

Notation

Use of Fonts

x	scalar variable
$\boldsymbol{x}$	vector variable
$\boldsymbol{X}$	matrix variable
$\mathbb{X}$	set
$x(\cdot), \mathrm{x}(\cdot)$	function with continuous argument $\cdot$
$x[\cdot], \mathrm{x}[\cdot]$	function with discrete argument $\cdot$

Constants

c	speed of light ($\mathrm{c} \approx 3 \cdot 10^8$ m/s)
e	Euler's number ($\mathrm{e} \approx 2.7183$)
j	imaginary unit ($\mathrm{j} = \sqrt{-1}$)
π	circular constant ($\pi \approx 3.1416$)

Specific Sets

$\mathbb{C}$	set of all complex numbers
$\mathbb{R}$	set of all real numbers

Specific Distributions

$\mathcal{CN}(\mu, \sigma^2)$	complex Gaussian distribution with mean μ and variance σ^2

$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\mathcal{U}([a, b])$	uniform distribution in the interval $[a, b]$

Functions and Operators

$*$	convolution
$\odot$	element-wise vector multiplication
$(\cdot)^{-1}$	inverse of a scalar of a square matrix
$(\cdot)^{\dagger}$	pseudo-inverse of a matrix
$(\cdot)^{*}$	conjugate complex of a scalar, a vector or a matrix
$(\cdot)^{T}$	transpose of a vector or a matrix
$(\cdot)^{H}$	hermitian transpose of a vector of a matrix
$\arg\max\{\cdot\}$	argument of maximum
$\arg\min\{\cdot\}$	argument of minimum
$\cos(\cdot)$	cosine function
$\delta(\cdot)$	Dirac impulse
$\det(\cdot)$	determinant of a square matrix
$\mathrm{diag}\,(\cdot)$	diagonal matrix with diagonal elements given by a vector
$\mathrm{E}\{\cdot\}$, $\mu_{(\cdot)}$	expectation/mean of a random variable
$\exp(\cdot)$, $\mathrm{e}^{(\cdot)}$	exponential function
$\mathcal{F}\{\cdot\}$	Fourier transform
$\Gamma(\cdot)$	gamma function
$\mathrm{Im}\{\cdot\}$	imaginary part of a complex number
$\ln(\cdot)$	natural logarithm (to the base e)
$\log(\cdot)$	logarithm to the base 10
$\log_a(\cdot)$	logarithm to the base a
$\mathcal{L}\{\cdot\}$	log-likelihood ratio
$\mathcal{L}_a\{\cdot\}$	a priori log-likelihood ratio
$\mathcal{L}_e\{\cdot\}$	extrinsic log-likelihood ratio
$\max(\cdot)$	maximum
$\min(\cdot)$	minimum
$(\cdot) \bmod (\cdot)$	modulo operator
$p(\cdot)$	probability density function of a random variable
$P(\cdot)$	probability of an event
$\mathrm{Re}\{\cdot\}$	real part of a complex number
$\sin(\cdot)$	cosine function
$\tanh(\cdot)$	hyperbolic tangent function
$\mathrm{tr}\,(\cdot)$	trace of a matrix (sum over all diagonal elements)
$\mathrm{Var}\{\cdot\}$, $\sigma^2_{(\cdot)}$	variance of a random variable

Mathematical Accents

$\hat{(\cdot)}$	estimate
$\breve{(\cdot)}$	pre-estimate
$\tilde{(\cdot)}$	hypothesis

List of Variables

Latin Characters, Lower Case

a	shape parameter of a gamma distribution
$a_m \, \mathrm{e}^{\mathrm{j}\,\xi_m}$	complex weighting factor of mth layer (power and phase allocation)
b	scale parameter of a gamma distribution
$\boldsymbol{b}$	observation vector for the WLS method
c	constant
c	cluster index
$c\,(\tau)$	time-invariant channel weight function
$c\,(\tau, t)$	time-variant channel weight function
$c_0\,(t_E, t)$	time-variant channel impulse response
d	distance between transmitter and receiver
$d_b(\boldsymbol{p})$	distance between the mobile station and the bth reference object
d_s	distance that corresponds to the symbol duration
$\boldsymbol{d}(\boldsymbol{p})$	vector containing the distances between the mobile station and all reference objects
$\hat{\boldsymbol{d}}$	pseudorange vector
$\boldsymbol{e}$	vector of pseudorange errors
f_0	carrier frequency
f_c	fading factor of cth cluster (for block fading)
$f_c[k]$	fading factor of cth cluster at time index k
f_D	Doppler shift
$f_{D,c}$	Doppler shift of cth cluster
$f_{D,c,r}$	Doppler shift of rth ray in the cth cluster
$f_{D,\max}$	maximum Doppler shift
g_m	overall weighting factor of mth layer including power and phase allocation and a single channel coefficient for flat fading
$g_{m,l}$	overall weighting factor of mth layer including power and phase allocation and the lth channel coefficient for frequency-selective fading
$g\,(\tau)$	convolution of pulse shaping filter and receive filter
$g_{\mathrm{Tx}}\,(\tau)$	pulse shaping filter
$g_{\mathrm{Rx}}\,(\tau)$	receive filter
$h\,(\nu, t)$	normalised overall channel weight function including pulse shaping and receive filtering in addition to the physical channel

$h(\tau, t)$	overall channel weight function including pulse shaping and receive filtering in addition to the physical channel
h_l	lth channel coefficient (for block fading)
$h_l[k]$	lth channel coefficient at time index k
$\boldsymbol{h}$	vector of channel coefficients
$\boldsymbol{h}(\boldsymbol{\theta})$	vector of channel coefficients as a function of the parameter vector
$\boldsymbol{h}_{\vartheta}$	lifted line on the solution locus
$\dot{\boldsymbol{h}}_{\vartheta}$	first order derivative of the lifted line (velocity)
$\ddot{\boldsymbol{h}}_{\vartheta}$	second order derivative of the lifted line (acceleration)
k	time index (after oversampling)
l	channel coefficient index
m	layer index for interleave-division multiplexing
n	time index (before coding)
$n[k]$	noise sample at time index k
$\boldsymbol{n}$	vector of noise samples
$\boldsymbol{n}_N$	null space component of noise vector
$\boldsymbol{n}_R$	range space component of noise vector
$\boldsymbol{n}_{\mathrm{WF}}$	vector of noise samples after whitening filter
p	path loss exponent
$p(\tau)$	Gaussian pulse shape
$\boldsymbol{p}$	position vector (MS's position)
$\boldsymbol{p}_b$	position of the bth reference object
r	ray index
$r(t)$	received signal
s^2	residual variance in the context of soft information
$s(t)$	transmitted signal
t	continuous time
t_E	excitation time
$u[n]$	bit at time index n
$u_m[n]$	bit of mth layer at time index n
v	magnitude of velocity
$\boldsymbol{v}$	velocity vector
$x[k]$	transmit symbol at time index k
$x_0[k]$	pilot symbol at time index k
$x_m[k]$	coded symbol of mth layer at time index k
$y[k]$	received sample at time index k
$\boldsymbol{y}$	vector of received samples
$\boldsymbol{y}_{\mathrm{WF}}$	vector of received samples after whitening filter
$z_m[k]$	received sample of mth layer at time index k after maximum ratio combining

Latin Characters, Upper Case

A_c	normalised amplitude of the cth cluster

B	number of reference objects for positioning
B_D	Doppler spread
B_M	coherence bandwidth
B_s	signal bandwidth
C	number of clusters in the power delay profile
$C\,(f,t)$	time-variant transfer function
$\boldsymbol{C}$	covariance matrix
$\boldsymbol{C_e}$	covariance matrix of pseudorange errors
$\boldsymbol{C_n}$	covariance matrix of noise samples
$\boldsymbol{C_{\hat{p}}}$	covariance matrix of position estimates
$\boldsymbol{C_\eta}$	covariance matrix of channel estimation error (LSCE)
$\boldsymbol{C_{\hat{\theta}}}$	covariance matrix of parameter estimates
$D(f_D)$	Doppler power spectrum
G_T	transmit antenna gain
G_R	receive antenna gain
$\boldsymbol{H}$	channel matrix with Toeplitz structure
$\boldsymbol{H}_{\mathrm{WF}}$	channel matrix after whitening filter
$\boldsymbol{H}_{\Omega_{\hat{d}}}(\tilde{\boldsymbol{p}})$	Hessian of the metric for position estimation
$\boldsymbol{H}(\boldsymbol{\theta})$	Hessian of channel coefficient vector
$\boldsymbol{H}_l(\boldsymbol{\theta})$	lth face of the Hessian of channel coefficient vector
$\boldsymbol{H}_\Omega(\tilde{\boldsymbol{\theta}})$	Hessian of the metric for channel parameter estimation
I	total number of particles for the PSO method
$\boldsymbol{I}(\boldsymbol{p})$	Fisher information matrix for position estimation
$\boldsymbol{I}(\boldsymbol{\theta})$	Fisher information matrix for channel parameter estimation
J	oversampling factor
$\boldsymbol{J}(\boldsymbol{p})$	Jacobian matrix of distance vector
$\boldsymbol{J}_{\Omega_{\hat{d}}}(\tilde{\boldsymbol{p}})$	Jacobian matrix of the metric for position estimation
$\boldsymbol{J}(\boldsymbol{\theta})$	Jacobian matrix of channel coefficient vector
$\boldsymbol{J}_\Omega(\tilde{\boldsymbol{\theta}})$	Jacobian matrix of the metric for channel parameter estimation
K'	number of symbols per burst (in the symbol-rate sampled domain)
K	number of symbols per burst (in the oversampled domain)
K_R	Rician factor
$K^{\mathrm{N}}_{\boldsymbol{\vartheta}}$	intrinsic curvature (normal curvature in direction of $\boldsymbol{\vartheta}$)
$K^{\mathrm{T}}_{\boldsymbol{\vartheta}}$	parameter-effects curvature (tangential curvature in direction of $\boldsymbol{\vartheta}$)
L'	channel memory length (in the symbol-rate sampled domain)
L	channel memory length (in the oversampled domain)
L	path loss
L_{FS}	free-space path loss
L_0	reference path loss at 1 m
$\boldsymbol{L_n}$	lower triangular matrix from Cholesky decomposition of $\boldsymbol{C_n}$
M	number of data layers in interleave-division multiplexing
$M(\nu)$	normalised power delay profile
$M(\tau)$	power delay profile
N	number of bits per layer

N	number of observations
$N_{\min}$	minimum number for which the stopping criteria must be fulfilled
P	number of parameters
$\breve{P}_b$	(soft) block error rate
P_c	average power of the cth path in the power delay profile
P_S	number of dimensions per subswarm for the CPSO method
P_R	received power
P_T	transmitted power
$\boldsymbol{P}_0$	projection matrix onto the range space of $\boldsymbol{X}_0$
R	code rate
R	distance from every corner of the quadratic region in Figure 5.1 to its middle point
R_c	number of rays per cluster
S	number of subswarms for the CPSO method
S	spreading factor of a repetition code
S_σ	log-normal shadowing with variance σ^2
$S(\tau, f_D)$	delay-Doppler-function
$\boldsymbol{S}$	system matrix for the WLS method
T	threshold w.r.t. fitness improvement for the stopping criteria
T_D	coherence time
T_f	fitness threshold for outage detection
T_{f,γ_s}	normalised fitness threshold for stopping criteria/outage detection
T_g	effective width of $g(\tau)$
T_{LM}	normalised fitness threshold for the LM method
T_M	delay spread
T_{PSO}	normalised fitness threshold for the PSO method
T_s	symbol duration
$\boldsymbol{W}$	weighting matrix for the TS algorithm
$\boldsymbol{X}_0$	Toeplitz matrix of pilot layer/training sequence

Greek Characters, Lower Case

α_c	angle of departure of cth cluster
$\alpha_{c,r}$	angle of departure of rth ray in the cth cluster
$\Delta\alpha$	azimuth spread of departure angles
β_c	weighting factor of the noise components for the EM algorithm
β_c	angle of arrival of cth cluster
$\beta_{c,r}$	angle of arrival of rth ray in the cth cluster
$\Delta\beta$	azimuth spread of arrival angles
γ	scaling factor
γ	incident angle for Doppler shift
γ_c	incident angle for Doppler shift of cth cluster
$\gamma_{c,r}$	incident angle for Doppler shift of rth ray in the cth cluster
γ_b	bit-to-noise ratio

γ_d	data-to-noise ratio
γ_p	pilot-to-noise ratio
γ_s	signal-to-noise ratio
$\gamma^{\mathrm{N}}_{\boldsymbol{\vartheta}}$	relative intrinsic curvature
$\gamma^{\mathrm{T}}_{\boldsymbol{\vartheta}}$	relative parameter-effects curvature
$\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2$	random vectors of the PSO and CPSO method
ε	sampling phase
$\zeta_m[k]$	effective noise term of mth layer at time index k
$\zeta_{m,l}[k]$	effective noise term of mth layer and lth channel coefficient at time index k
$\boldsymbol{\eta}$	channel estimation error (LSCE)
$\boldsymbol{\theta}$	parameter vector for channel parameter estimation
$\tilde{\boldsymbol{\theta}}_{\mathrm{GB}}$	global best position for the PSO method
$\tilde{\boldsymbol{\theta}}_{\mathrm{IB},i}$	individual best position of the ith particle for the PSO method
$\tilde{\boldsymbol{\theta}}_{\mathrm{LB},i}$	local best position of the ith particle for the PSO method
$\tilde{\boldsymbol{\theta}}_{0,\mathrm{fIG}}$	fixed initial guess for the local methods
$\tilde{\boldsymbol{\theta}}_{0,\mathrm{pIG}}$	perfect initial guess for the local methods
$\tilde{\boldsymbol{\theta}}_{0,\mathrm{rIG}}$	random initial guess for the local methods
ϑ_r	fixed ray offset of rth ray
$\boldsymbol{\vartheta}$	any non-zero vector in the parameter space
λ	carrier wavelength
λ_1, λ_2	learning factors of the PSO and CPSO method
ν	excess delay
ν_c	excess delay of the cth cluster
$\nu_{c,r}$	excess delay of the rth ray in the cth cluster
$\nu_2^{\max}$	maximum possible excess delay for the two-path channel models
π_m	interleaver of mth layer
ρ	ratio of the pilot layer power to the total power for PLACE
ρ	standard radius in the context of curvature measures
ϱ	coarse excess delay
ϱ_c	coarse excess delay of the cth cluster
σ_n^2	variance of noise samples
τ	propagation delay
τ_1	propagation delay of the first arriving path (time of arrival)
$\pm\Delta\tau_1$	maximum sampling phase offset
τ_c	propagation delay of the cth cluster
$\tau_{c,r}$	propagation delay of the rth ray in the cth cluster
ϕ_v	direction of velocity
$\varphi_{nn}[\Delta k]$	sampled autocorrelation function of noise samples for difference Δk
$\varphi_{cc}(\tau_1, \tau_2, t_1, t_2)$	autocorrelation function of random time-variant channel weight function
$\varphi_{CC}(f_1, f_2, t_1, t_2)$	autocorrelation function of random time-variant transfer function
$\varphi_{SS}(\tau_1, \tau_2, f_{D1}, f_{D2})$	autocorrelation function of random delay-Doppler-function
χ	constriction factor of the PSO and CPSO method

$\psi_{\mathrm{Rx}}(\tau)$	autocorrelation function of the receive filter
ω	tuning factor of the Levenberg-Marquardt method

Greek Characters, Upper Case

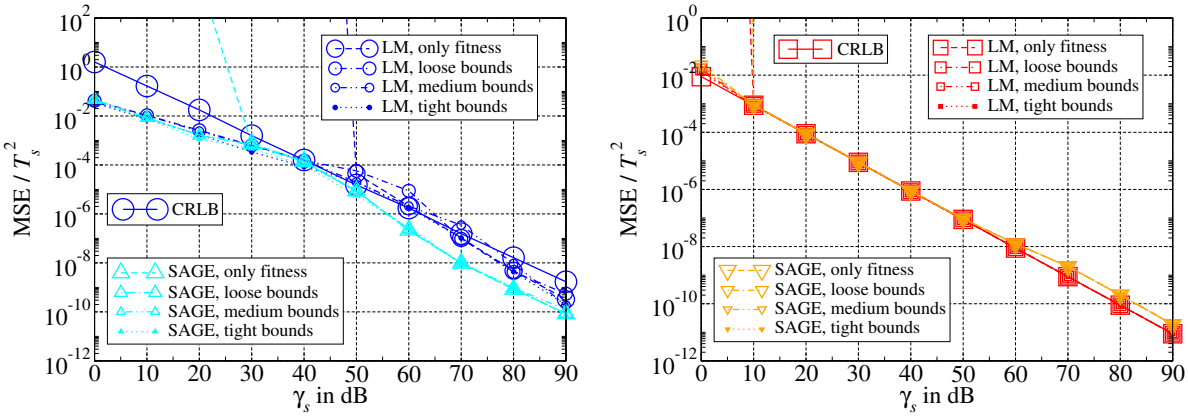
Γ^{N}	maximum relative intrinsic curvature
Γ^{T}	maximum relative parameter-effects curvature
$\Theta_{cc}(\tau, \Delta t)$	delay-time-correlation function
$\Theta_{CC}(\Delta f, \Delta t)$	time-frequency-correlation function
$\Theta_{SS}(\tau, f_D)$	scattering function/delay-Doppler-power density spectrum
$\boldsymbol{\Upsilon}$	context vector of the CPSO method
Φ	randomly generated starting phase
Φ_c	randomly generated starting phase of cth cluster
$\Phi_{c,r}$	randomly generated starting phase of rth ray in the cth cluster
$\Delta\Phi$	phase offset between adjacent clusters
$\boldsymbol{\Psi}$	auxiliary variable introduced for the SAGE algorithm
$\Omega_{\hat{\boldsymbol{d}}}(\tilde{\boldsymbol{p}})$	maximum-likelihood metric for position estimation
$\Omega(\tilde{\boldsymbol{\theta}})$	maximum-likelihood metric for channel parameter estimation
$\Omega_{\hat{\boldsymbol{h}}}(\tilde{\boldsymbol{\theta}})$	maximum-likelihood metric based on pre-stage channel estimates
$\Omega_{\boldsymbol{y}}(\tilde{\boldsymbol{\theta}})$	maximum-likelihood metric based on received samples

Calligraphic Characters

$\mathcal{B}$	bit load
$\mathcal{C}_{P,N-P}^{1-\alpha}$	reference value related to the Fisher distribution in the context of curvature measures
$\mathcal{CH}$	linear time-variant channel
$\mathcal{F}_{P,N-P}^{1-\alpha}$	Fisher distribution with confidence level $1-\alpha$
$\boldsymbol{\mathcal{H}}_{\Omega}(\tilde{\boldsymbol{\theta}})$	modified Hessian of the Levenberg-Marquardt method
$\boldsymbol{\mathcal{I}}$	identity matrix
$\mathcal{P}$	power ratio of the multipath components in the two-path channel models
$\mathcal{S}$	neighbourhood set for the PSO and CPSO method
$\mathcal{T}$	sampling period
$\boldsymbol{\mathcal{W}}$	weighting matrix for the WLS method

Appendix C

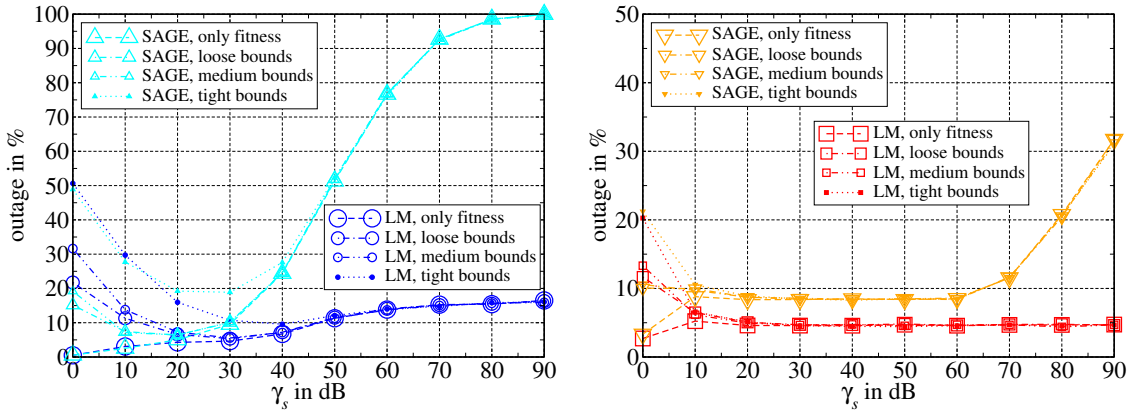
Supplementary Results



(a) Local methods with $\nu_2/T_s \in [0.05, 1]$.

(b) Local methods with $\nu_2/T_s \in [1, 2]$.

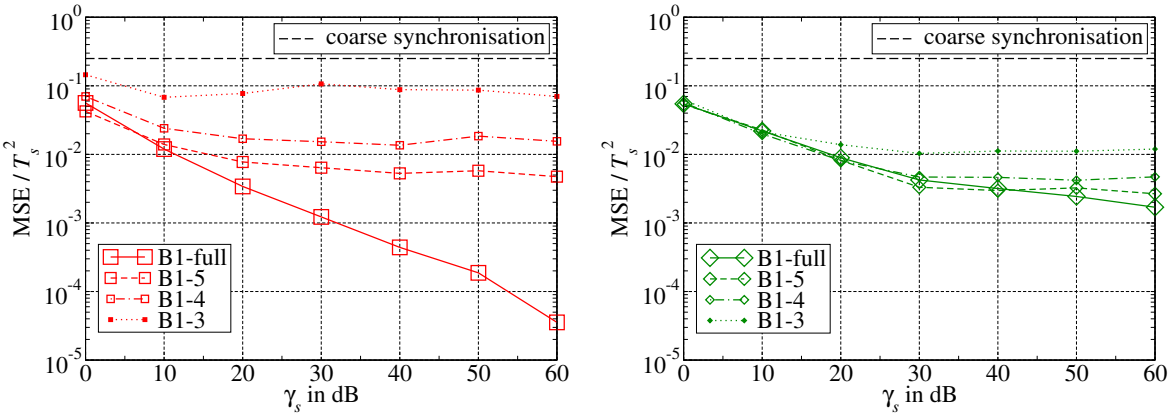
Figure C.1: MSE of $\hat{\varepsilon}$ normalised to T_s^2 for the local methods with different bound constraints in the two-path channels. In case of the tight bounds, the search space bounds also used for the global methods are applied (real and imaginary part of amplitudes in [-1,1], coarse excess delays in [-0.5T_s, $\nu_2^{\max}$]). For the medium and the loose bounds, the real and imaginary part of the amplitudes are bounded in the intervals [-2,2] and [-5,5], respectively.



(a) Local methods with $\nu_2/T_s \in [0.05, 1]$.

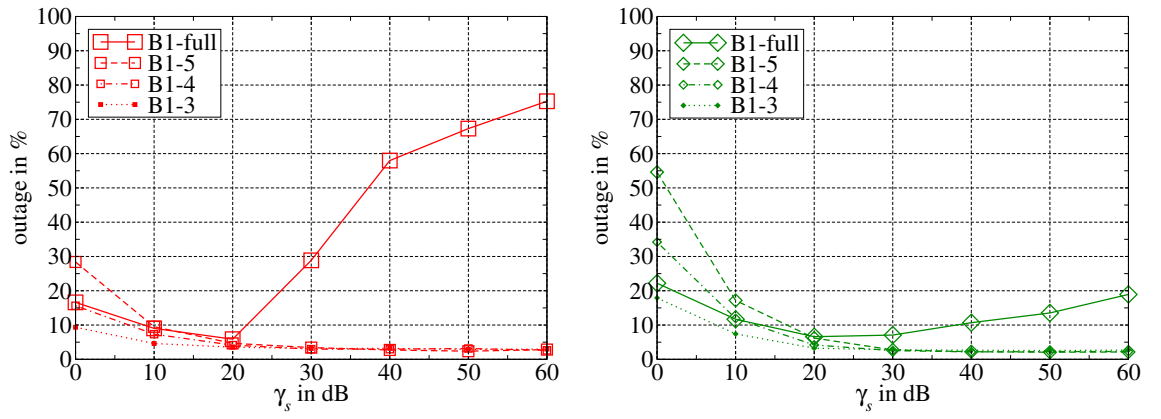
(b) Local methods with $\nu_2/T_s \in [1, 2]$.

Figure C.2: Outage for the local methods with different bound constraints in the two-path channels. In case of the tight bounds, the search space bounds also used for the global methods are applied (real and imaginary part of amplitudes in [-1,1], coarse excess delays in [-0.5T_s, $\nu_2^{\max}$]). For the medium and the loose bounds, the real and imaginary part of the amplitudes are bounded in the intervals [-2,2] and [-5,5], respectively.



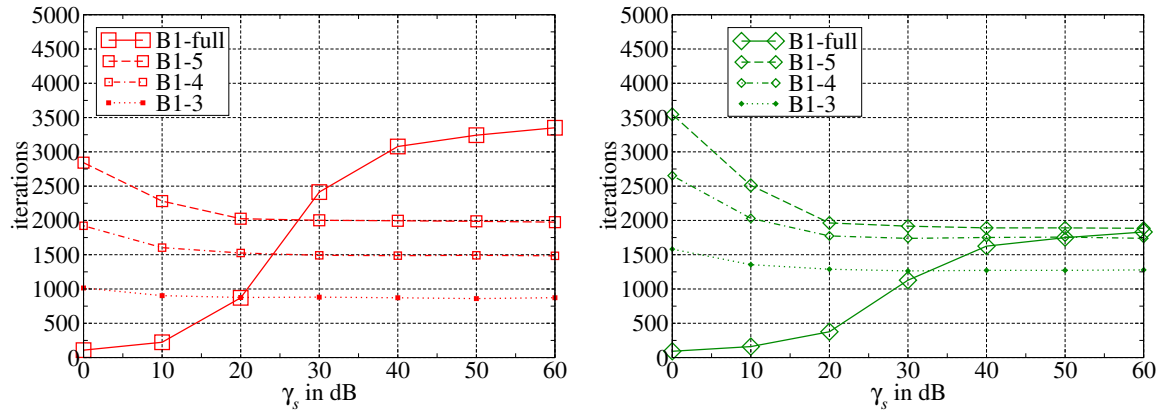
(a) MSE of $\hat{\varepsilon}$ normalised to T_s^2 for T_s=50 ns.

(b) MSE of $\hat{\varepsilon}$ normalised to T_s^2 for T_s=100 ns.



(c) Outage for T_s=50 ns.

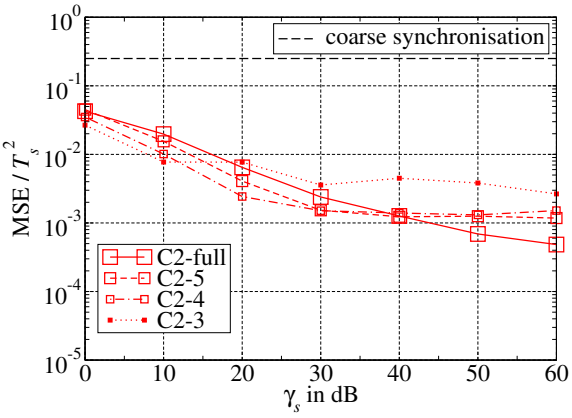
(d) Outage for T_s=100 ns.



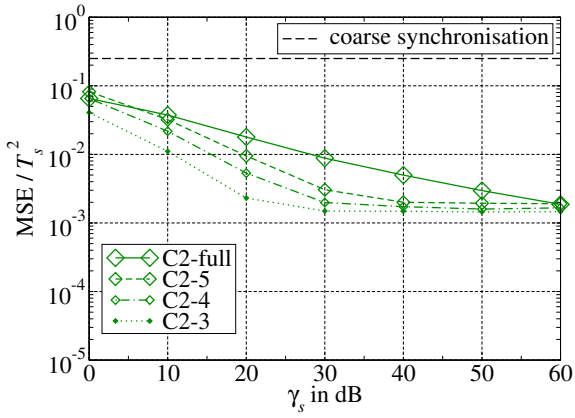
(e) Performed iterations for T_s=50 ns.

(f) Performed iterations for T_s=100 ns.

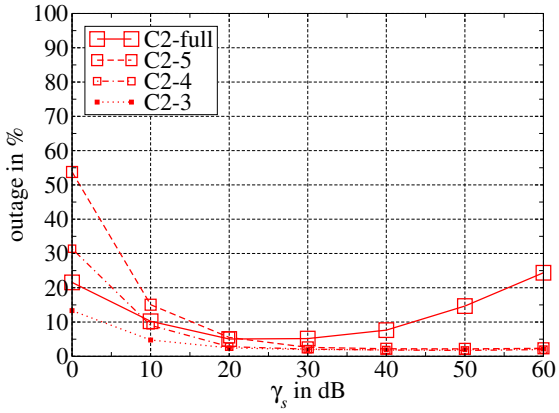
Figure C.3: Performance of the full estimator and several reduced estimators in the B1-LOS channel.



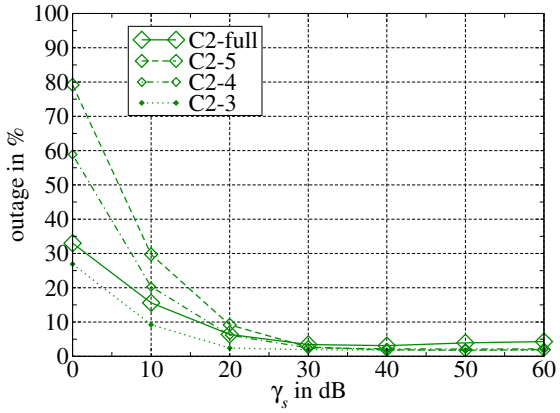
(a) MSE of $\hat{\varepsilon}$ normalised to T_s^2 for T_s=50 ns.



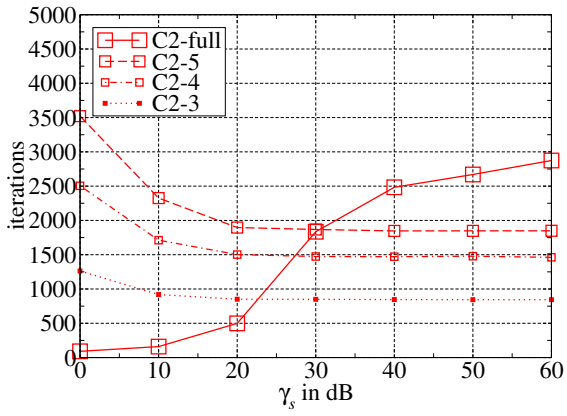
(b) MSE of $\hat{\varepsilon}$ normalised to T_s^2 for T_s=100 ns.



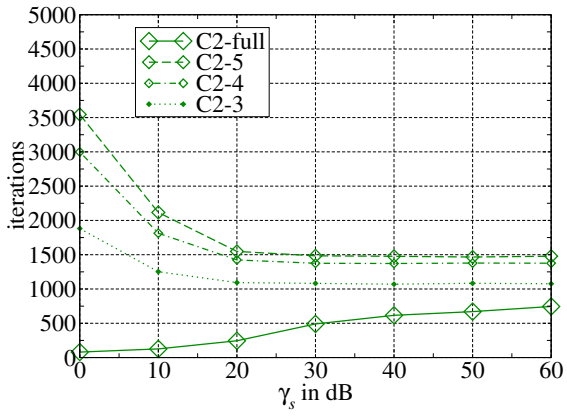
(c) Outage for T_s=50 ns.



(d) Outage for T_s=100 ns.

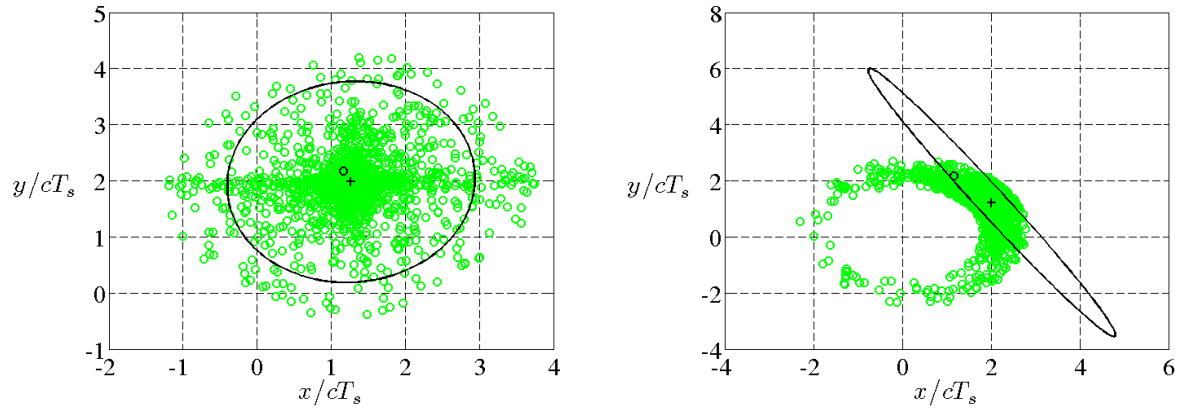


(e) Performed iterations for T_s=50 ns.



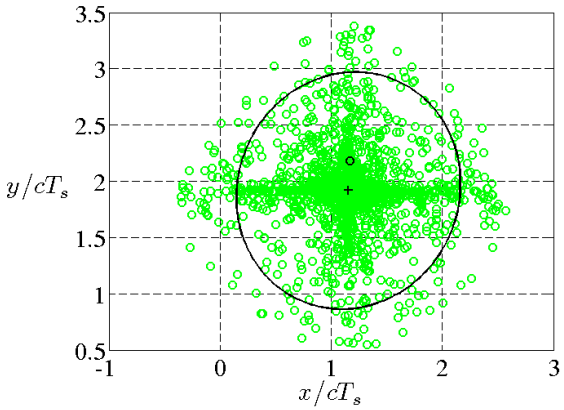
(f) Performed iterations for T_s=100 ns.

Figure C.4: Performance of the full estimator and several reduced estimators in the C2-LOS channel.

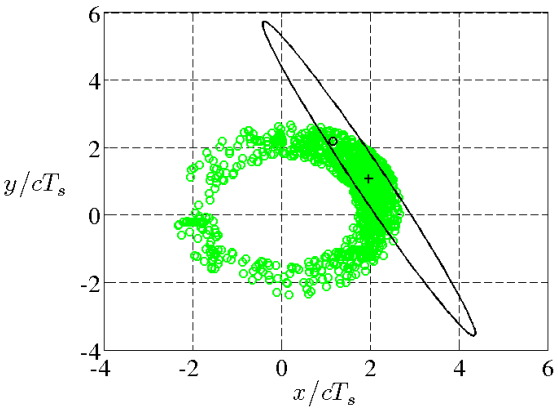


(a) Small GDOP at 10 dB.

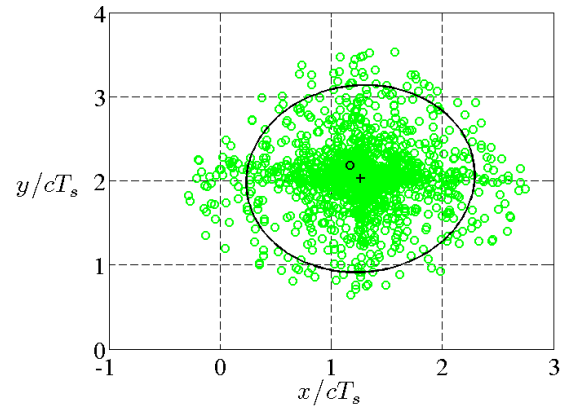
(b) Large GDOP at 10 dB.



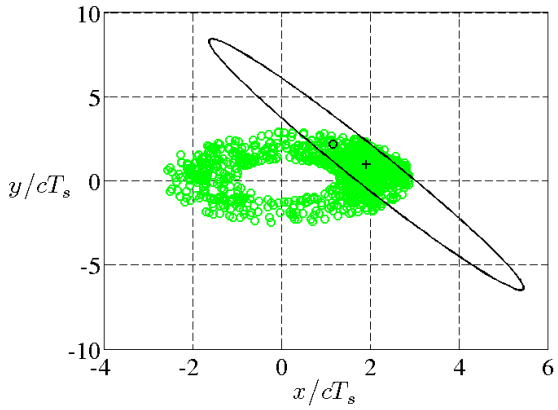
(c) Small GDOP at 30 dB.



(d) Large GDOP at 30 dB.

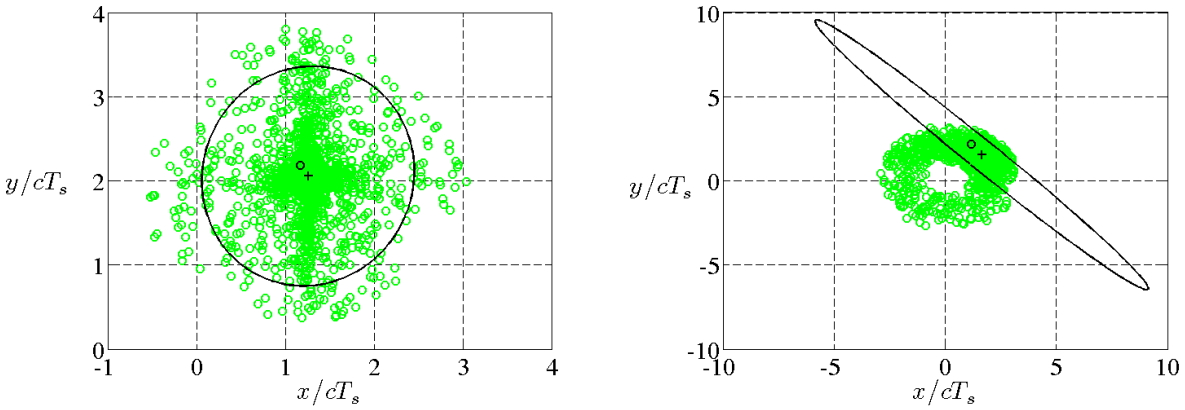


(e) Small GDOP at 50 dB.



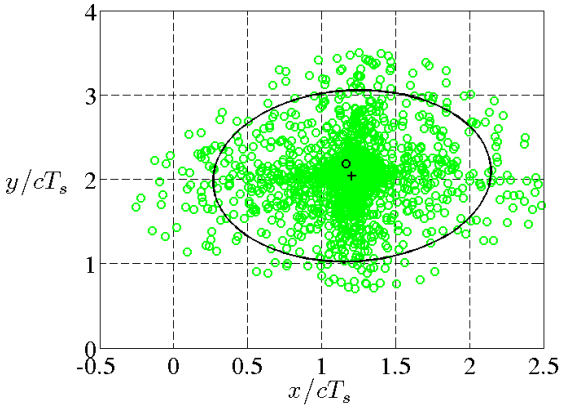
(f) Large GDOP at 50 dB.

Figure C.5: Confidence regions based on the linearisation method (black ellipse) and the likelihood method (green circles) for positioning in case of WINNER B1-LOS for different GDOPs and at different SNRs. The estimated parameters $\hat{\boldsymbol{\theta}}$ are denoted by a black cross and the true parameters $\boldsymbol{\theta}$ by a black circle.

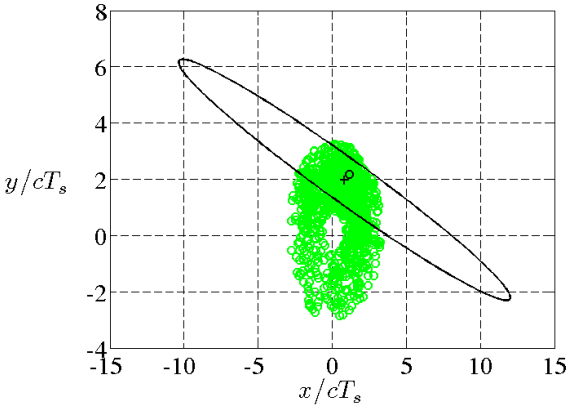


(a) Small GDOP at 10 dB.

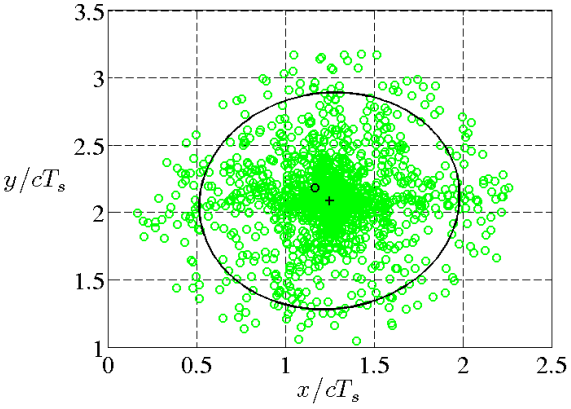
(b) Large GDOP at 10 dB.



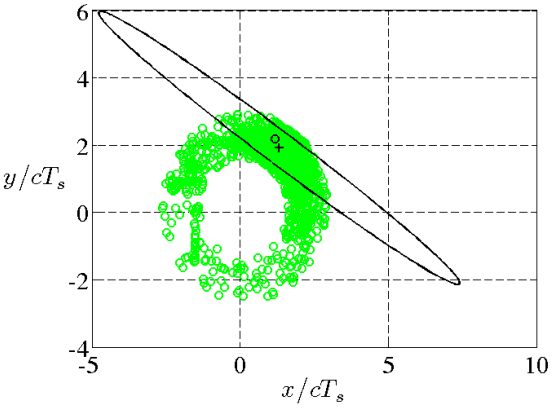
(c) Small GDOP at 30 dB.



(d) Large GDOP at 30 dB.



(e) Small GDOP at 50 dB.

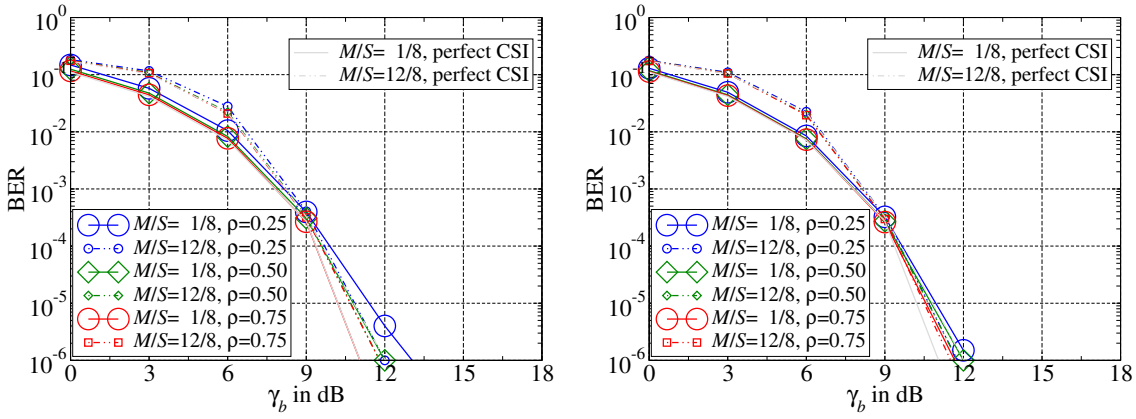


(f) Large GDOP at 50 dB.

Figure C.6: Confidence regions based on the linearisation method (black ellipse) and the likelihood method (green circles) for positioning in case of WINNER C2-LOS for different GDOPs and at different SNRs. The estimated parameters $\hat{\boldsymbol{\theta}}$ are denoted by a black cross and the true parameters $\boldsymbol{\theta}$ by a black circle.

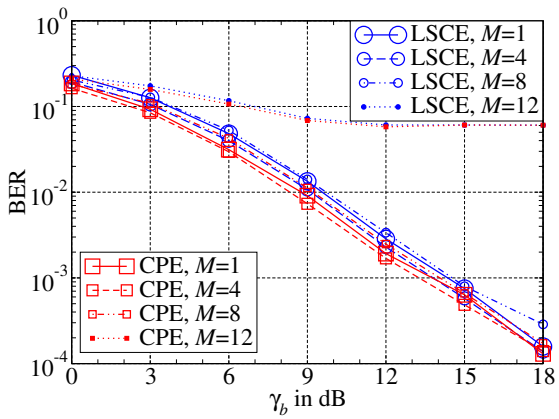
		small GDOP	large GDOP
	Position	$x/\mathrm{c}\,T_s$=1.16531 $y/\mathrm{c}\,T_s$=2.18072	$x/\mathrm{c}\,T_s$=1.16531 $y/\mathrm{c}\,T_s$=2.18072
	GDOP	**1.01852**	**65.4819**
	P=D	2	2
	N=B	4	4
	$\mathcal{C}^{0.95}_{P,N-P}$	**0.22942**	**0.22942**
WINNER B1-LOS channel	Γ^{N} @ 10 dB	0.26837	0.25188
	Γ^{T} @ 10 dB	0.07873	5.43959
	Γ^{N} @ 30 dB	0.15733	0.27301
	Γ^{T} @ 30 dB	0.09727	5.75029
	Γ^{N} @ 50 dB	0.17067	0.50717
	Γ^{T} @ 50 dB	0.05268	10.13636
WINNER C2-LOS channel	Γ^{N} @ 10 dB	0.20478	0.78810
	Γ^{T} @ 10 dB	0.06376	11.75305
	Γ^{N} @ 30 dB	0.16137	1.11493
	Γ^{T} @ 30 dB	0.03423	13.47793
	Γ^{N} @ 50 dB	0.13046	0.49991
	Γ^{T} @ 50 dB	0.04033	7.09721

Table C.1: Maximum relative curvatures for positioning for the WINNER B1-LOS and C2-LOS channel models with small and large GDOP at different SNRs.

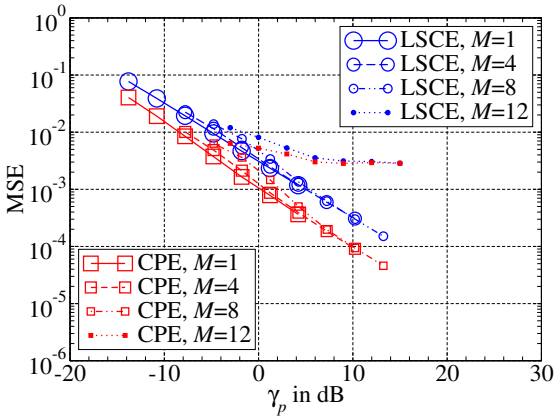


(a) K'=1000. (b) K'=2000.

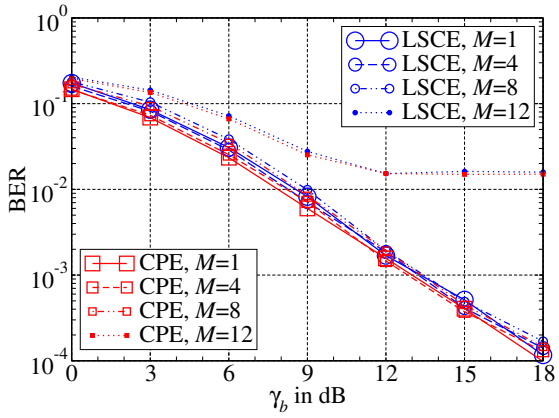
Figure C.7: BERs for IDM with LSCE, LLR combining and large burst length K' over γ_b in dB.



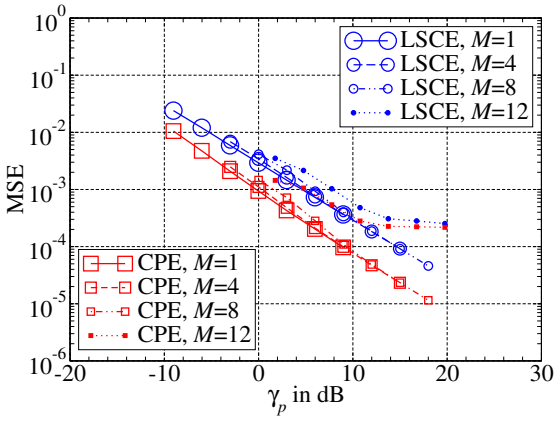
(a) BER with ρ=0.25.



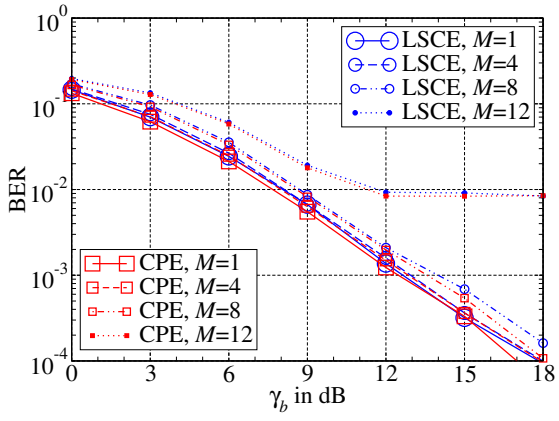
(b) MSE with ρ=0.25.



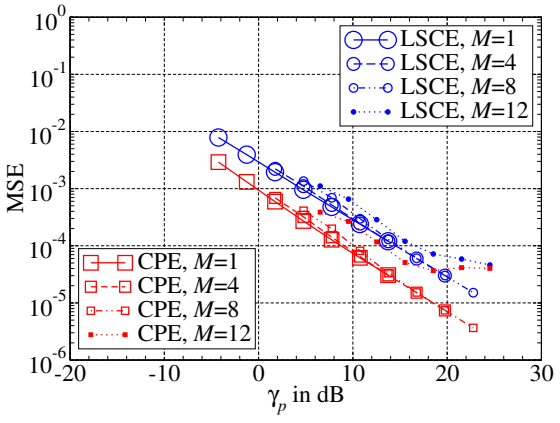
(c) BER with ρ=0.5.



(d) MSE with ρ=0.5.

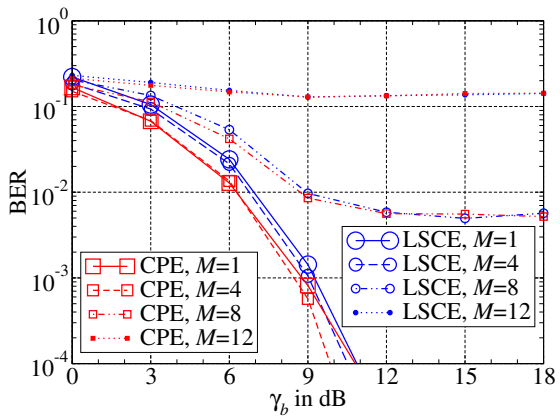


(e) BER with ρ=0.75.

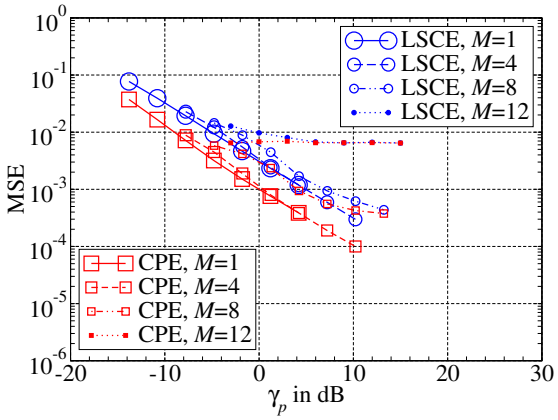


(f) MSE with ρ=0.75.

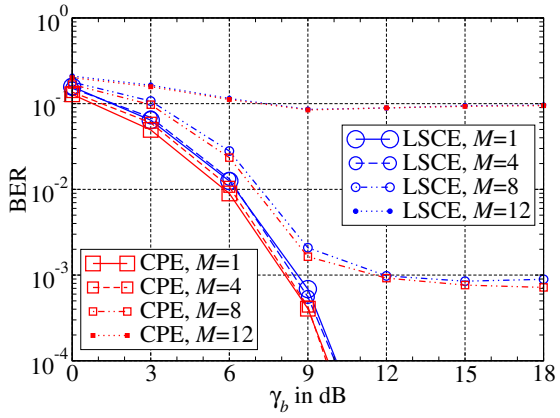
Figure C.8: Comparison of BERs as well as MSEs for LSCE and CPE with a spreading factor S=8 in case of a two-path channel with small excess delay ($\nu_2/T_s \in [0.05, 1]$).



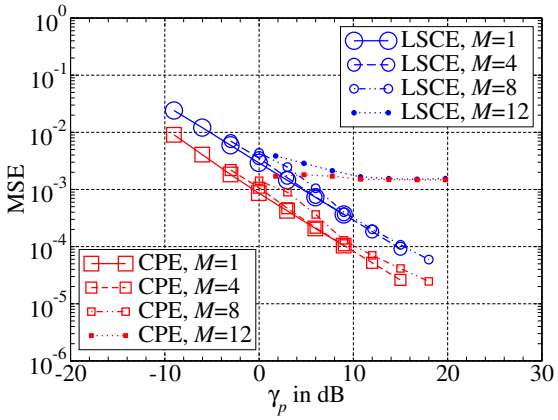
(a) BER with ρ=0.25.



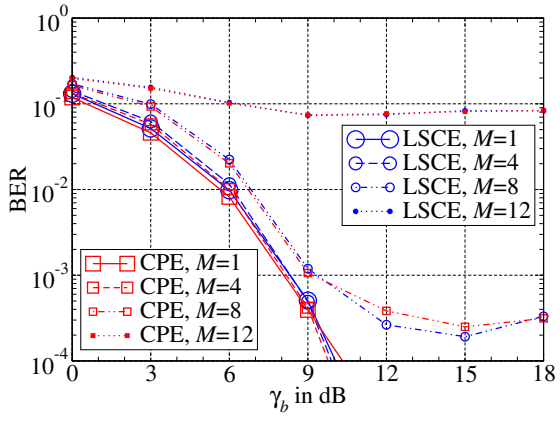
(b) MSE with ρ=0.25.



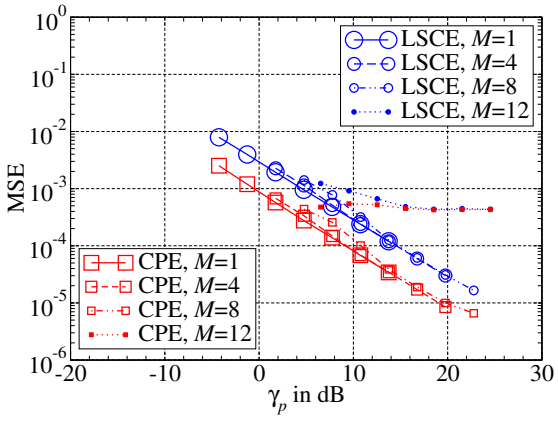
(c) BER with ρ=0.5.



(d) MSE with ρ=0.5.

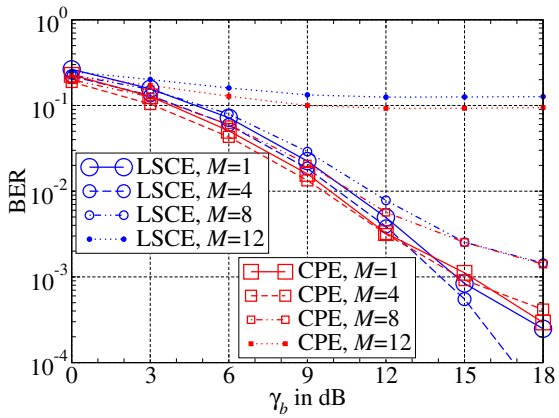


(e) BER with ρ=0.75.

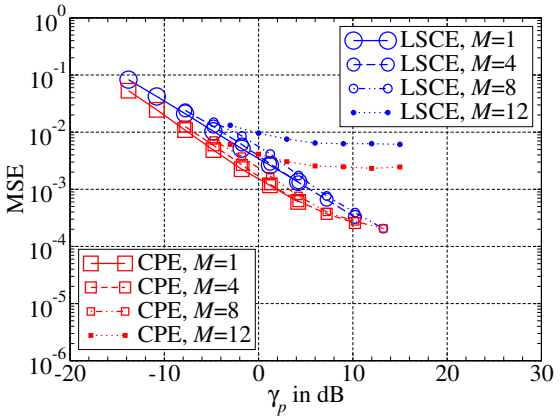


(f) MSE with ρ=0.75.

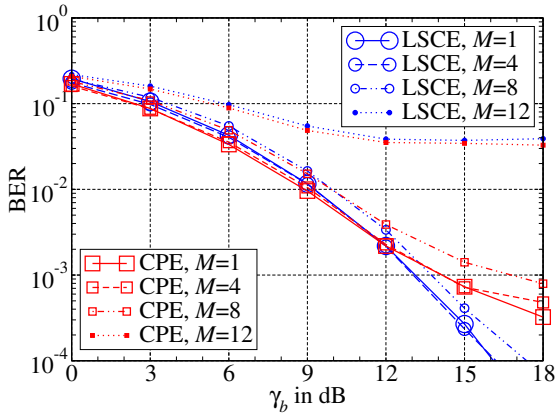
Figure C.9: Comparison of BERs as well as MSEs for LSCE and CPE with a spreading factor S=8 in case of a two-path channel with large excess delay ($\nu_2/T_s \in [1, 2]$).



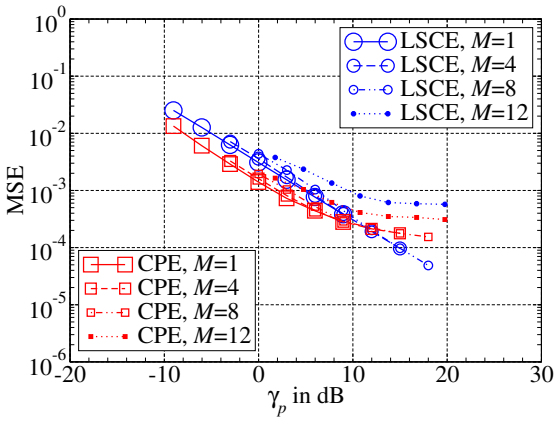
(a) BER with ρ=0.25.



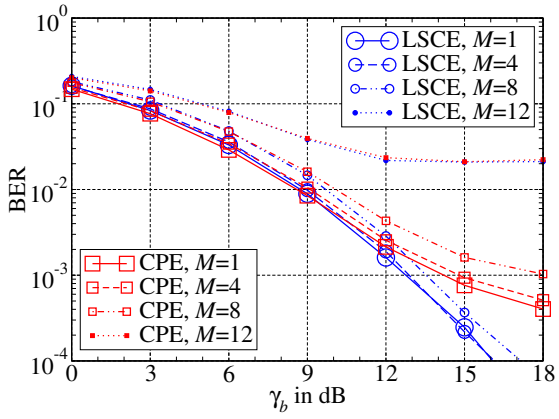
(b) MSE with ρ=0.25.



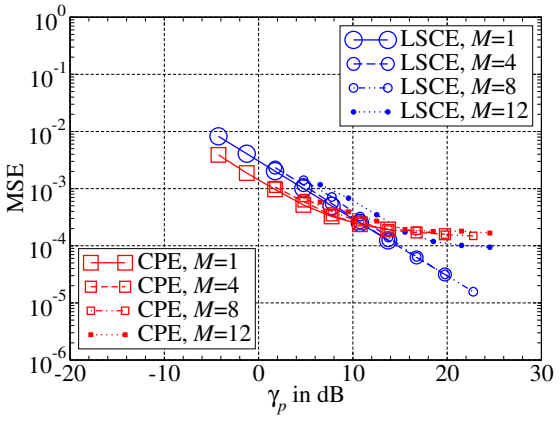
(c) BER with ρ=0.5.



(d) MSE with ρ=0.5.

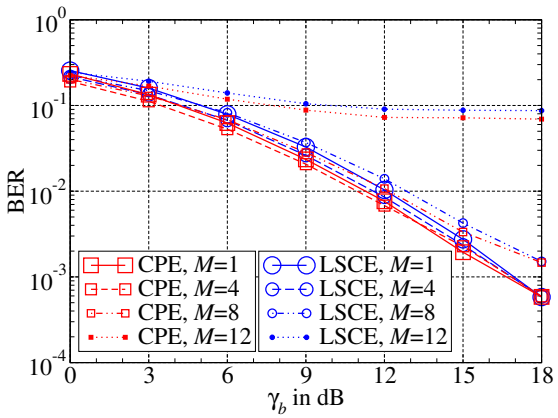


(e) BER with ρ=0.75.

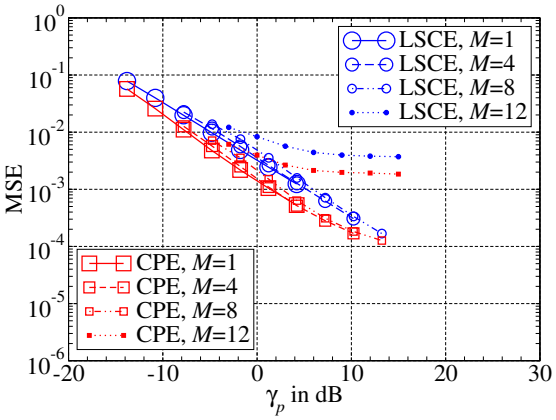


(f) MSE with ρ=0.75.

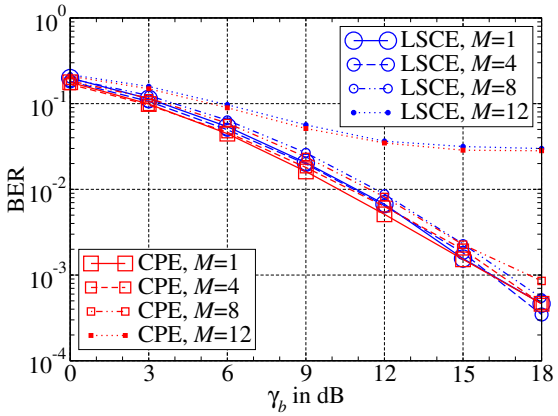
Figure C.10: Comparison of BERs as well as MSEs for LSCE and CPE with a spreading factor S=8 in case of the WINNER B1-LOS channel.



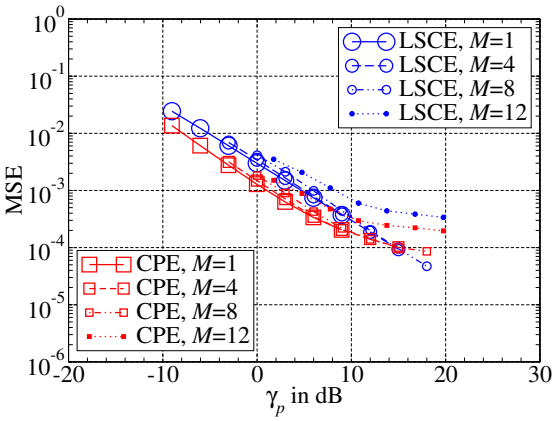
(a) BER with ρ=0.25.



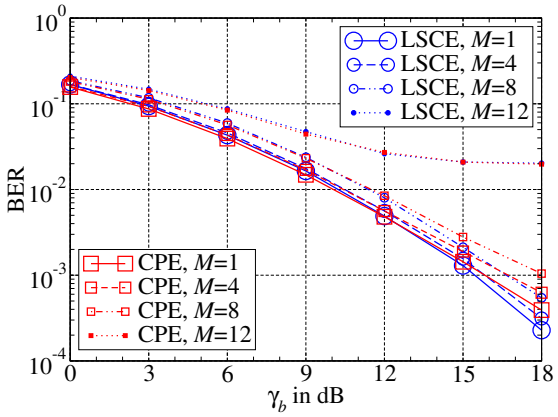
(b) MSE with ρ=0.25.



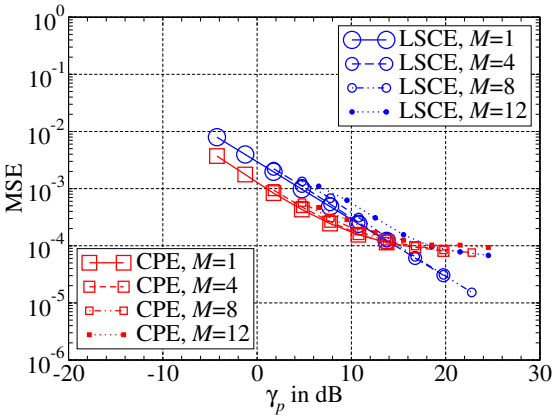
(c) BER with ρ=0.5.



(d) MSE with ρ=0.5.

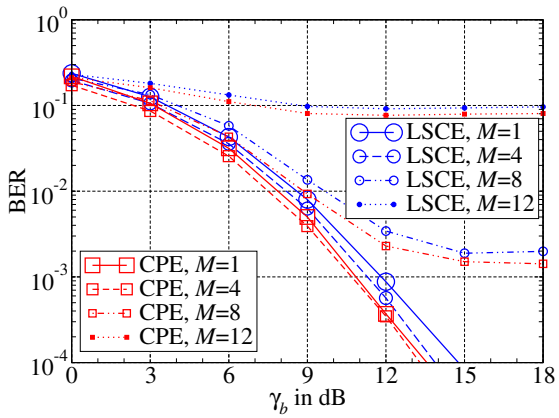


(e) BER with ρ=0.75.

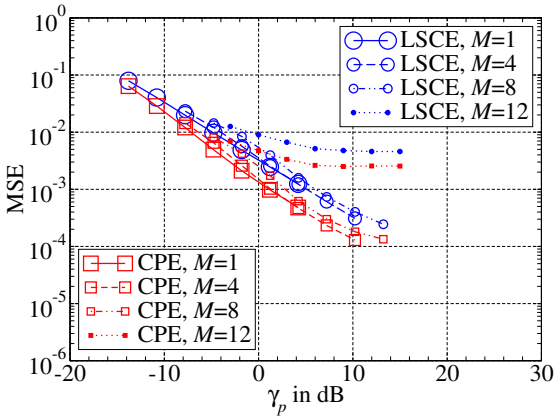


(f) MSE with ρ=0.75.

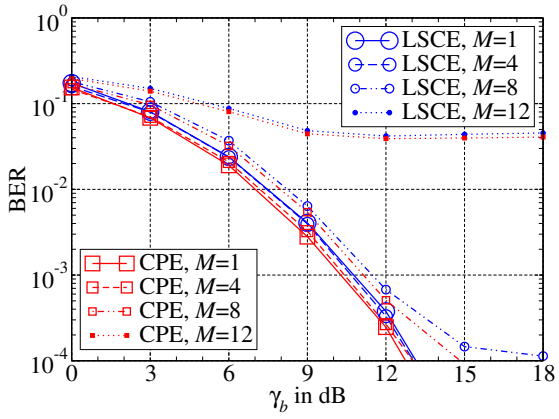
Figure C.11: Comparison of BERs as well as MSEs for LSCE and CPE with a spreading factor S=8 in case of the WINNER B3-LOS channel.



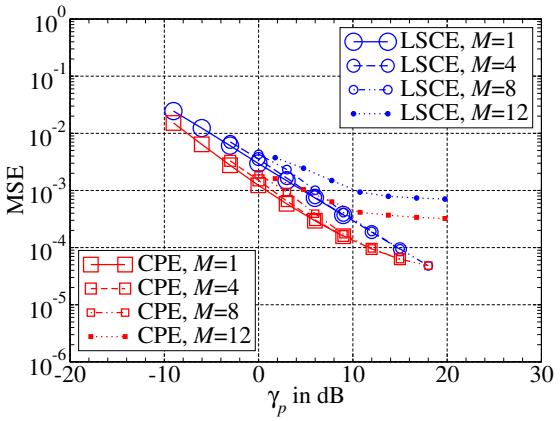
(a) BER with ρ=0.25.



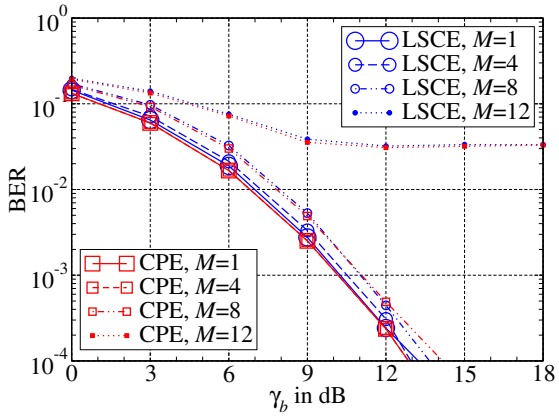
(b) MSE with ρ=0.25.



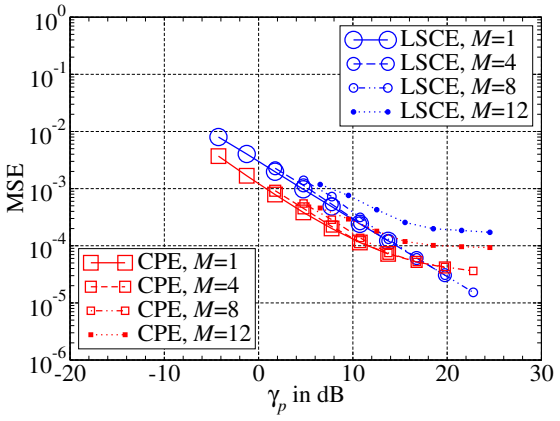
(c) BER with ρ=0.5.



(d) MSE with ρ=0.5.



(e) BER with ρ=0.75.



(f) MSE with ρ=0.75.

Figure C.12: Comparison of BERs as well as MSEs for LSCE and CPE with a spreading factor S=8 in case of the WINNER C2-LOS channel.

Bibliography

[AN07] S. Ali and P. Nobles, "A novel indoor location sensing mechanism for IEEE 802.11 b/g wireless LAN," in *Proceedings of Workshop on Positioning, Navigation and Communication (WPNC)*, Mar. 2007, pp. 9–15, Hannover, Germany.

[ANU08] F. Antreich, J. A. Nossek, and W. Utschick, "Maximum likelihood delay estimation in a navigation receiver for aeronautical applications," *Aerospace Science and Technology*, vol. 12, no. 3, pp. 256–267, 2008.

[Bel63] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Transactions on Communication Systems*, vol. 11, no. 4, pp. 360–393, 1963.

[BK07] D. Bratton and J. Kennedy, "Defining a standard for particle swarm optimization," in *Proceedings of IEEE Swarm Intelligence Symposium (SIS)*, Apr. 2007, pp. 120–127, Honolulu, Hawaii, USA.

[BL08] C. Blum and X. Li, "Swarm intelligence in optimization," in *Swarm Intelligence: Introduction and Applications*, ser. Natural Computing Series, C. Blum and D. Merkle, Eds. Springer, 2008, pp. 43–85.

[BSGS00] S. Brück, U. Sorger, S. Gligorevic, and N. Stolte, "Interleaving for outer convolutional codes in DS-CDMA systems," *IEEE Transactions on Communications*, vol. 48, no. 7, pp. 1100–1107, Jul. 2000.

[BV04] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, England: Cambridge University Press, Mar. 2004.

[BvD99] M. S. Braasch and A. J. van Dierendonck, "GPS receiver architectures and measurements," in *Proceedings of the IEEE*, vol. 87, no. 1, Jan. 1999, pp. 48–64.

[BW80] D. M. Bates and D. G. Watts, "Relative curvature measures of nonlinearity," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 42, no. 1, pp. 1–25, 1980.

[Caf00] J. J. Caffery, Jr., *Wireless location in CDMA cellular radio systems.* Boston, Massachusetts, USA: Kluwer Acadademic Publishers, 2000.

[CK02] M. Clerc and J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, Feb. 2002.

[CRE09] B. Cristea, D. Roviras, and B. Escrig, "Turbo receivers for interleave-division multiple-access systems," *IEEE Transactions on Communications*, vol. 57, no. 7, pp. 2090–2097, Jul. 2009.

[CS98] J. J. Caffery, Jr. and G. L. Stüber, "Overview of radiolocation in CDMA cellular systems," *IEEE Communications Magazine*, vol. 36, pp. 38–45, Apr. 1998.

[CT97] H. A. Cirpan and M. K. Tsatsanis, "Chip interleaving in direct sequence CDMA systems," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, Apr. 1997, pp. 3877–3880, Munich, Germany.

[CZ08] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*, 3rd ed., ser. Wiley-Interscience Series in Discrete Mathematics and Optimization. Hoboken, New Jersey, USA: Wiley-Interscience, Feb. 2008.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[DS83] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations.* Englewood Cliffs, New Jersey, USA: Prentice-Hall, 1983.

[DS85] J. R. Donaldson and R. B. Schnabel, "Computational experience with confidence regions and confidence intervals for nonlinear least squares," in *Proceedings of Symposium on the Interface of Computer Sciences and Statistics*, 1985, pp. 83–93, Lexington, Kentucky, USA.

[EM99] P. Enge and P. Misra, "Special issue on global positioning system," in *Proceedings of the IEEE*, vol. 87, no. 1, Jan. 1999, pp. 3–15.

[FDHT96] B. H. Fleury, D. Dahlhaus, R. Heddergott, and M. Tschudin, "Wideband angle of arrival estimation using the SAGE algorithm," in *Proceedings of IEEE International Symposium on Spread Spectrum Techniques and Applications*, vol. 1, Sep. 1996, pp. 79–85, Mainz, Germany.

[FH94] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.

[FOO00] P. Frenger, P. Orten, and T. Ottosson, "Code-spread CDMA using maximum free distance low-rate convolutional codes," *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 135–144, Jan. 2000.

[Fre] "FeuerWhere - tracking fire fighters," http://www.feuerwhere.de/, Freie Universität Berlin et al., [Online, accessed 05-22-2012].

[Fri09] J. C. Fricke, "Zuverlässigkeitsbasierter Cross-Layer-Entwurf digitaler Übertragungssysteme," Ph.D. dissertation, Faculty of Engineering, University of Kiel, Germany, Feb. 2009.

[FSRSM02] G. Fock, P. Schulz-Rittich, A. Schenke, and H. Meyr, "Low complexity high resolution subspace-based delay estimation for DS-CDMA," in *Proceedings of IEEE International Conference on Communications (ICC)*, Apr. 2002, New York, USA.

[FTH+99] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dalhaus, and K. I. Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 434–450, Mar. 1999.

[FW88] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 477–489, Apr. 1988.

[GCWI08] I. Guevenc, C.-C. Chong, F. Watanabe, and H. Inamura, "NLOS identification and weighted least-squares localization for UWB systems using multipath channel statistics," *EURASIP Journal on Applied Signal Processing*, vol. 2008, pp. 1–14, 2008, Article ID 271984.

[Gera] "WHERE - wireless hybrid enhanced mobile radio estimators," http://www.ict-where.eu/, German Aerospace Center (DLR) et al., [Online, accessed 05-22-2012].

[Gerb] "WHERE2 - wireless hybrid enhanced mobile radio estimators (phase 2)," http://www.ict-where2.eu/, German Aerospace Center (DLR) et al., [Online, accessed 05-22-2012].

[Gez08] S. Gezici, "A survey on wireless position estimation," *Wireless Personal Communications, Special Issue on Towards Global and Seamless Personal Navigation*, vol. 44, no. 3, pp. 263–282, Feb. 2008.

[GG05] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: Possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 41–53, Jul. 2005.

[GMW91] P. E. Gill, W. Murray, and M. H. Wright, *Numerical linear algebra and optimization*. Redwood City, California, USA: Addison-Wesley, 1991.

[GMW97] ——, *Practical optimization*. London, England: Academic Press, 1997, 11th printing.

[Gol05] A. Goldsmith, *Wireless communications*. Cambridge, England: Cambrigde University Press, 2005.

[GTG+05] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: A look at positioning aspects of future sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, Jul. 2005.

[Hen06] N. Henze, *Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls*, 6th ed. Wiesbaden, Germany: Vieweg, 2006.

[HH08] D. Hao and P. A. Höher, "Iterative estimation and cancellation of clipping noise for multi-layer IDMA systems," in *Proceedings of International ITG Conference on Source and Channel Coding (SCC)*, Jan. 2008, Ulm, Germany.

[HH10] ——, "Superposition modulation with reliability-based hybrid detection," in *Proceedings of International Symposium on Turbo Codes & Iterative Information Processing (ISTC)*, Sep. 2010, Brest, France.

[HKR97] P. Höher, S. Kaiser, and P. Robertson, "Pilot-symbol-aided channel estimation in time and frequency," in *Proceedings of Communication Theory Mini-Conference in conjunction with IEEE Global Communications Conference (GLOBECOM)*, Nov. 1997, pp. 90–96, Phoenix, Arizona, USA.

[HLS00] P. Höher, I. Land, and U. Sorger, "Log-likelihood values and Monte Carlo simulation – some fundamental results," in *Proceedings of International Symposium on Turbo Codes & Related Topics (ISTC)*, Sep. 2000, pp. 43–46, Brest, France.

[HN99] W. Huyer and A. Neumaier, "Global optimization by multilevel coordinate search," *Journal of Global Optimization*, vol. 14, no. 4, pp. 331–355, Jun. 1999.

[Höh11] P. A. Höher, *Grundlagen der digitalen Informationsübertragung: Von der Theorie zu Mobilfunkanwendungen*. Wiesbaden, Germany: Vieweg+Teubner Verlag, 2011.

[HROW07] P. A. Höher, P. Robertson, E. Offer, and T. Wörz, "The soft-output principle - reminiscences and new developments," *European Transactions on Telecommunications*, vol. 18, no. 8, pp. 829–835, 2007.

[HS07] P. A. Höher and K. Schmeink, "Joint navigation and communication based on interleave-division multiple access," in *Proceedings of International Workshop on Multi-Carrier Spread Spectrum (MCSS)*, May 2007, pp. 97–106, Herrsching, Germany.

[HSF08] P. A. Höher, H. Schöneich, and J. C. Fricke, "Multi-layer interleave-division multiple access: Theory and practice," *European Transactions on Telecommunications*, vol. 19, no. 5, pp. 523–536, Aug. 2008.

[HT99] P. Höher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, Dec. 1999, pp. 2162–2166, Rio de Janeiro, Brazil.

[Hub96] J. B. Huber, "Systemtheoretische Grundlagen der Modellierung von Mobilfunkkanälen," Kurs 5 der Ferien-Akademie der Universität Erlangen-Nürnberg und der Technischen Universität München, Tech. Rep., Sep. 1996.

[Hub02] ——, "Grundlagen der Wahrscheinlichkeitsrechnung fuer iterative Decodierverfahren," *e&i (Elektrotechnik und Informationstechnik)*, vol. 119, no. 11, p. 386ff, Nov. 2002.

[HW07] S. Helwig and R. Wanka, "Particle swarm optimization in high-dimensional bounded search spaces," in *Proceedings of IEEE Swarm Intelligence Symposium (SIS)*, Apr. 2007, pp. 198–205, Honolulu, Hawaii, USA.

[HWD+07] G. Heinrichs, J. Winkel, C. Drewes, L. Maurer, A. Springer, R. Stuhlberger, and C. Wicpalek, "System considerations for a combined UMTS/GNSS receiver," in *Proceedings of Workshop on Positioning, Navigation and Communication (WPNC)*, Mar. 2007, pp. 189–198, Hannover, Germany.

[JPS93] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the Lipschitz constant," *Journal of Optimization Theory and Application*, vol. 79, no. 1, pp. 157–181, Oct. 1993.

[Kam04] K. D. Kammeyer, *Nachrichtenübertragung*, 3rd ed. Stuttgart, Germany: Teubner Verlag, 2004.

[Kap96] E. D. Kaplan, Ed., *Understanding GPS: Principles and applications*. Boston, Massachusetts, USA: Artech House, 1996.

[Kay93] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Upper Saddle River, New Jersey, USA: Prentice-Hall, 1993.

[KDUB07] K. Kusume, G. Dietl, W. Utschick, and G. Bauch, "Performance of interleave division multiple access based on minimum mean square error detection," in *Proceedings of IEEE International Conference on Communications (ICC)*, Jun. 2007, pp. 2961–2966, Glasgow, Scotland.

[KE95] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of IEEE International Conference on Neural Networks*, vol. 4, Nov./Dec. 1995, pp. 1942–1948, Perth, Australia.

[KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.

[KHTA11] C. Knievel, P. A. Höher, A. Tyrrell, and G. Auer, "Particle swarm enhanced graph-based channel estimation for MIMO-OFDM," in *Proceedings of IEEE Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–5, Budapest, Hungary.

[KM06] J. Kennedy and R. Mendes, "Neighborhood topologies in fully informed and best-of-neighborhood particle swarms," *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, vol. 36, no. 4, pp. 515–519, Jul. 2006.

[KMH+08] P. Kyösti, J. Meinilä, L. Hentilä, X. Zhao, T. Jämsä, C. Schneider, M. Narandzić, M. Milojević, A. Hong, J. Ylitalo, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. de Yong, and T. Rautianinen, "IST-4-027756 WINNER II D1. 1.2 V1. 2 WINNER II channel models," Wireless World Initiative New Radio (WINNER), Tech. Rep., 2008. [Online]. Available: http://projects.celtic-initiative.org/winner+/WINNER2-Deliverables/D1.1.2v1.2.pdf

[KRBK97] A. S. Khayrallah, R. Ramesh, G. E. Bottomley, and D. Koilpillai, "Improved channel estimation with side information," in *Proceedings of IEEE Vehicular Technology Conference (VTC Spring)*, vol. 2, May 1997, pp. 1049–1053, Phoenix, Arizona, USA.

[KSHA10] C. Knievel, Z. Shi, P. A. Höher, and G. Auer, "2D graph-based soft channel estimation for MIMO-OFDM," in *Proceedings of IEEE International Conference on Communications (ICC)*, May 2010, pp. 1–5, Capetown, South Africa.

[Lan99] R. B. Langley, "Dilution of precision," *GPS World*, vol. 10, no. 5, pp. 52–59, 1999.

[LD60] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, no. 3, pp. 497–520, Jul. 1960.

[LLP03] L. Liu, W. K. Leung, and L. Ping, "Simple iterative chip-by-chip multiuser detection for CDMA systems," in *Proceedings of IEEE Vehicular Technology Conference (VTC Spring)*, vol. 3, Apr. 2003, pp. 2157–2161, Jeju, Korea.

[LNCP97] J.-W. Liang, B. C. Ng, J.-T. Chen, and A. Paulraj, "GMSK linearization and structured channel estimate for GSM signals," in *Proceedings of IEEE Military Communications Conference (MILCOM)*, vol. 2, Nov. 1997, pp. 817–821, Monterey, California, USA.

[LP98] H.-N. Lee and G. J. Pottie, "Fast adaptive equalization/diversity combining for time-varying dispersive channels," *IEEE Transactions on Communications*, vol. 46, no. 9, pp. 1146–1162, Sep. 1998.

[LP04] L. Liu and L. Ping, "Iterative detection of chip interleaved CDMA systems in multipath channels," *IEE Electronics Letters*, vol. 40, no. 14, pp. 884–885, Jul. 2004.

[MP02] R. H. Mahadevappa and J. G. Proakis, "Mitigating multiple access interference and intersymbol interference in uncoded CDMA systems with chip-level interleaving," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 781–792, Oct. 2002.

[MP04] X. Ma and L. Ping, "Coded modulation using superimposed binary codes," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3331–3343, Dec. 2004.

[Neu04] A. Neumaier, "Complete search in continuous global optimization and constraint satisfaction," *Acta Numerica*, vol. 13, pp. 271–369, May 2004.

[NHW11] M. Nõmm, P. A. Höher, and Y. Wang, "Phase allocation aspects of interleave-division multiplexing from PAPR prospective," in *Proceedings of IEEE Vehicular Technology Conference (VTC Spring)*, May 2011, Budapest, Hungary.

[NM65] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, no. 4, pp. 308–313, Jan. 1965.

[NW99] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, USA: Springer, 1999.

[NWH10] M. Nõmm, T. Wo, and P. A. Höher, "Multilayer APP detection for IDM," *Electronics Letters*, vol. 46, no. 1, pp. 96–97, Jan. 2010.

[Pin96] J. D. Pintér, *Global Optimization in Action: Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1996.

[PL05] K. Pahlavan and A. H. Levesque, *Wireless information networks*, 2nd ed. Hoboken, New Jersey, USA: Wiley, 2005.

[PLM02] K. Pahlavan, X. Li, and J.-P. Mäkelä, "Indoor geolocation science and technology," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 112–118, Feb. 2002.

[PLWL02] L. Ping, L. Liu, K. Wu, and W. K. Leung, "A unified approach to multiuser detection and space-time coding with low complexity and nearly optimal performance," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Oct. 2002, pp. 170–179, Monticelli, Illinois, USA.

[PLWL06] ——, "Interleave-division multiple access," *IEEE Transactions on Wireless Communications*, vol. 55, no. 4, pp. 938–947, Apr. 2006.

[PLY+00] K. Pahlavan, X. Li, M. Ylianttila, R. Chana, and M. Latvo-aho, "An overview of wireless indoor geolocation techniques and systems," in *Proceedings of International Conference on Mobile and Wireless Communications Networks (MWCN)*, ser. Lecture Notes in Computer Science, vol. 1818. Berlin/Heidelberg, Germany: Springer, May 2000, pp. 1–13.

[PRK85] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques - ESPRIT," in *Proceedings of Asilomar Conference on Circuits, Systems and Computers*, Nov. 1985, pp. 83–89, Pacific Grove, California, USA.

[PTVF02] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing*, 2nd ed. Cambridge, England: Cambridge University Press, 2002.

[PV02] K. E. Parsopoulos and M. N. Vrahatis, "Recent approaches to global optimization problems through particle swarm optimization," *Natural Computing*, vol. 1, pp. 235–306, Jun. 2002.

[Rap96] T. S. Rappaport, *Wireless communications: Principles and practice.* Upper Saddle River, New Jersey, USA: Prentice Hall, 1996.

[RK89] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[RLT03] A. Richter, M. Landmann, and R. S. Thomä, "Maximum likelihood channel parameter estimation from multidimensional channel sounding measurements," in *Proceedings of IEEE Vehicular Technology Conference (VTC Spring)*, Apr. 2003, pp. 1056–1060, Jeju, Korea.

[Ros04] S. M. Ross, *Introduction to probability and statistics for engineers and scientists*, 3rd ed. Amsterdam, Netherlands: Elsevier Academic Press, 2004.

[RRA10] P. Robertson, M. Röckl, and M. Angermann, "Advances in multi-sensor data fusion for ubiquitous positioning: novel approaches for robust localization and mapping," in *Proceedings of VDE Kongress*, Nov. 2010, Leipzig, Germany.

[SAH11a] K. Schmeink, R. Adam, and P. A. Höher, "Influence of oversampling on channel parameter estimation for joint communication and positioning," in *Proceedings of Workshop on Positioning, Navigation and Communication (WPNC)*, Apr. 2011, Dresden, Germany.

[SAH11b] ——, "Joint communication and positioning based on soft channel parameter estimation," *EURASIP Journal on Wireless Communications and Networking, Special Issue on Localization in Mobile Wireless and Sensor Networks*, 2011.

[SAH12] ——, "Performance limits of channel parameter estimation for joint communication and positioning," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, p. 178, 2012.

[SBJ+08] M. Schwaab, E. C. Biscaia, Jr., J. L. Monteiro, and J. C. Pinto, "Nonlinear parameter estimation through particle swarm optimization," *Chemical Engineering Science*, vol. 63, no. 6, pp. 1542–1552, 2008.

[SBKH10] K. Schmeink, R. Block, C. Knievel, and P. A. Höher, "Joint channel and parameter estimation for combined communication and navigation using particle swarm optimization," in *Proceedings of Workshop on Positioning, Navigation and Communication (WPNC)*, Mar. 2010, pp. 4–9, Dresden, Germany.

[SCGL05] G. Sun, J. Chen, W. Guo, and K. J. R. Liu, "Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 12–23, Jul. 2005.

[Sch91] L. L. Scharf, *Statistical signal processing: Detection, estimation, and time series analysis*, 2nd ed. Reading, Massachusetts, USA: Addison-Wesley, 1991.

[Sch08] H. Schöneich, "Adaptiver Interleave-Divison Mehrfachzugriff (IDMA) mit Anwendung in der Mobilfunkkommunikation," Ph.D. dissertation, Faculty of Engineering, University of Kiel, Germany, Jan. 2008.

[SH04] H. Schöneich and P. A. Höher, "Adaptive interleave-division multiple access - a potential air interface for 4G bearer services and wireless LANs," in *Proceedings of IEEE and IFIP International Conference on Wireless and Optical Communications and Networks (WOCN)*, Jun. 2004, pp. 179–182, Muscat, Oman.

[SH05] ——, "Semi-blind pilot-layer aided channel estimation with emphasis on interleave-division multiple access systems," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, Nov./Dec. 2005, pp. 3513–3517, St. Louis, Missouri, USA.

[SH06] ——, "Iterative pilot-layer aided channel estimation with emphasis on interleave-division multiple access systems," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–15, 2006.

[SH08] K. Schmeink and P. A. Höher, "Multi-layer interleave-division multiple access for joint communication and localization," in *Proceedings of International ITG Conference on Source and Channel Coding (SCC)*, Jan. 2008, Ulm, Germany.

[Sha48] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423,623–656, 1948.

[SHF05] H. Schöneich, P. A. Höher, and J. C. Fricke, "Adaptive 4G uplink proposal based on interleave-division multiple access," in *Proceedings of General Assembly of International Union of Radio Science (URSI GA)*, Oct. 2005, New Delhi, India.

[Sky] "SKYHOOK - the worldwide leader in location positioning, context and intelligence," http://www.skyhookwireless.com, Skyhook, [Online, accessed 05-22-2012].

[STK05] A. H. Sayed, A. Tarighat, and N. Khajenouri, "Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24–40, Jul. 2005.

[SWH10] Z. Shi, T. Wo, and P. A. Höher, "Superposition mapping with adaptive bit loading for BICM-OFDM systems," in *Proceedings of International Symposium on Turbo Codes & Iterative Information Processing (ISTC)*, Sep. 2010, Brest, France.

[SZT08] G. Shen, R. Zetik, and R. S. Thomä, "Performance comparison of TOA and TDOA based location estimation algorithms in LOS environment," in *Proceedings of Workshop on Positioning, Navigation and Communication (WPNC)*, Mar. 2008, pp. 71–78, Hannover, Germany.

[TP06] J. Tong and L. Ping, "Iterative decoding of superposition coding," in *Proceedings of International Symposium on Turbo Codes & Related Topics (ISTC)*, Apr. 2006, Munich, Germany.

[TPM06] J. Tong, L. Ping, and X. Ma, "Superposition coding with peak-power limitation," in *Proceedings of IEEE International Conference on Communications (ICC)*, Jun. 2006, Istanbul, Turkey.

[TPM09] ——, "Superposition coded modulation with peak-power limitation," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2562–2576, Jun. 2009.

[vdB01] F. van den Bergh, "An analysis of particle swarm optimizers," Ph.D. dissertation, Faculty of Natural and Agricultural Sciences, University of Pretoria, 2001.

[vdBE04] F. van den Bergh and A. P. Engelbrecht, "A cooperative approach to particle swarm optimization," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 225–239, Jun. 2004.

[vdVVP98] A.-J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Transactions on Signal Processing*, vol. 46, no. 2, pp. 405–418, Feb. 1998.

[Wei09] T. Weise, *Global Optimization Algorithms - Theory and Application*, 2nd ed. Self-Published, Jun. 2009. [Online]. Available: http://www.it-weise.de/

[WH10a] T. Wo and P. A. Höher, "Superposition mapping with application in bit-interleaved coded modulation," in *Proceedings of International ITG Conference on Source and Channel Coding (SCC)*, Jan. 2010, Siegen, Germany.

[WH10b] ——, "A universal coding approach for superposition mapping," in *Proceedings of International Symposium on Turbo Codes & Iterative Information Processing (ISTC)*, Sep. 2010, Brest, France.

[WKAR06] K. Wendlandt, M. Khider, M. Angermann, and P. Robertson, "Continuous location and direction estimation with multiple sensors using particle filtering," in *Proceedings of International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Sep. 2006, Heidelberg, Germany.

[WM97] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.

[WNHH10] T. Wo, M. Nõmm, D. Hao, and P. A. Höher, "Iterative processing for superposition mapping," *Journal of Electrical and Computer Engineering*, vol. 2010, 2010, Article ID 706464, 13 pages.

[ZL07] K. Zielinski and R. Laur, "Stopping criteria for a constrained single-objective particle swarm optimization algorithm," *Informatica (Slovenia)*, vol. 31, no. 1, pp. 51–59, 2007.

List of Figures

List of Tables

List of Algorithms