Joint Communication and Positioning based on Channel Estimation

Dissertation

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

vorgelegt von

Rebecca Carrie Adam

Kiel 2020

Tag der Einreichung: 18.06.2020
Tag der Disputation: 17.12.2020

Berichterstatter: Prof. Dr.-Ing. Peter Adam Höher
Prof. Dr.-Ing. Wolfgang Gerstacker

To Julina, Cilia and Jost

Preface

This thesis originates from my time as a research assistant at the Chair for Information and Coding Theory (ICT), Faculty of Engineering, Kiel University, Germany. I partially carried out my thesis in my free time during two maternity leaves. I finalized this work during my free time while working as a research assistant at the Intelligent Systems Group, Centre for Industrial Electronics (CIE), Department for Mechanical and Electrical Engineering, at the University of Southern Denmark (SDU).

Firstly, I want to thank my advisor, Prof. Dr-Ing. Peter Adam Höher, for his encouraging and motivating guidance during all this time, determined by the perfect balance of independent work and constructive discussions whenever required. Secondly, I am deeply thankful for the opportunity to visit international conferences and the academic experience I gained during my time at ICT. I want to thank Peter Höher for all his support in enabling and encouraging flexible, family-friendly, yet efficient scientific research opportunities. At all times, I experienced the ICT team as a scientifically excellent working environment. The complete ICT team and the DSS team, with whom we enjoyed our coffee together, were always determined by a very positive and enjoyable spirit and attitude. Therefore, I would like to thank all my former colleagues from ICT and DSS for the great time we had together. I want to express my special thanks to Ms. Sigrid Thielbörger. She was a part of this positive spirit at the faculty and will always have a special place in my heart. Furthermore, I want to thank Associate Professors Dr. Thomas Ebel and Dr. Robert Brehm for their support at SDU when I finished writing my thesis.

As a mother, I want to tell my two wonderful daughters, Cilia and Julina, how deeply thankful I am for all their patience when I was finishing this thesis. I certainly sometimes had too little time for them, and I am proud of them and their full support. As a daughter, I want to thank my mother for encouraging me to show interest in science when I was little, particularly in mathematics, so fundamental to this work and most sciences.

The most important "thank you" I would like to express to my dear husband, Jost. He patiently always supported me, strengthened me with the necessary confidence in stressful times. Without complaining, he juggled at times our double-career with kids plus me-writing-a-thesis-in-my-free-time life. He made all of this possible, accepting the extra load on his shoulders, being there for my two daughters when I could not.

Thanks again to you all!

Abstract

Mobile wireless communication systems have rapidly and globally become an integral part of everyday life and have brought forth the internet of things. With the evolution of mobile wireless communication systems, joint communication and positioning becomes increasingly important and enables a growing range of new applications. Humanity has already grown used to having access to multimedia data everywhere at every time and thereby employing all sorts of location-based services. Global navigation satellite systems can provide highly accurate positioning results whenever a line-of-sight path is available. Unfortunately, harsh physical environments are known to degrade the performance of existing systems. Therefore, ground-based systems can assist the existing position estimation gained by satellite systems.

Determining positioning-relevant information from a unified signal structure designed for a ground-based joint communication and positioning system can either complement existing systems or substitute them. Such a system framework promises to enhance the existing systems by enabling a highly accurate and reliable positioning performance and increased coverage. Furthermore, the unified signal structure yields synergetic effects.

In this thesis, I propose a channel estimation-based joint communication and positioning system that employs a virtual training matrix. This matrix consists of a relatively small training percentage, plus the detected communication data itself. Via a core semi-blind estimation approach, this iteratively includes the already detected data to accurately determine the positioning-relevant parameter, by mutually exchanging information between the communication part and the positioning part of the receiver. Synergy is created. I propose a generalized system framework, suitable to be used in conjunction with various communication system techniques.

The most critical positioning-relevant parameter, the time-of-arrival, is part of a physical multipath parameter vector. Estimating the time-of-arrival, therefore, means solving a global, non-linear, multi-dimensional optimization problem. More precisely, it means solving the so-called inverse problem. I thoroughly assess various problem formulations and variations thereof, including several different measurements and estimation algorithms.

A significant challenge, when it comes to solving the inverse problem to determine the positioning-relevant path parameters, is imposed by realistic multipath channels. Most parameter estimation algorithms have proven to perform well in moderate multipath environments. It is mathematically straightforward to optimize this performance in the sense that the number of observations has to exceed the number of parameters to be estimated. The typical parameter estimation problem, on the other hand, is based on channel estimates, and it assumes that so-called snapshot measurements are available. In the case of

realistic channel models, however, the number of observations does not necessarily exceed the number of unknowns. In this thesis, I overcome this problem, proposing a method to reduce the problem dimensionality via joint model order selection and parameter estimation. Employing the approximated and estimated parameter covariance matrix inherently constrains the estimation problem's model order selection to result in optimal parameter estimation performance and hence optimal positioning performance. To compare these results with the optimally achievable solution, I introduce a focused order-related lower bound in this thesis. Additionally, I use soft information as a weighting matrix to enhance the positioning algorithm positioning performance.

For demonstrating the feasibility and the interplay of the proposed system components, I utilize a prototype system, based on multi-layer interleave division multiple access. This proposed system framework and the investigated techniques can be employed for multiple existing systems or build the basis for future joint communication and positioning systems. The assessed estimation algorithms are transferrable to all kinds of joint communication and positioning system designs. This thesis demonstrates their capability to, in principle, successfully cope with challenging estimation problems stemming from harsh physical environments.

Keywords: Positioning, navigation, communication, estimation, optimisation, particle swarm optimisation, model order selection, information theoretic criteria

Kurzfassung

Mobile drahtlose Kommunikationssysteme sind schnell und global zu einem festen Bestandteil des Alltags geworden und haben das Internet der Dinge hervorgebracht. Mit der Entwicklung mobiler drahtloser Kommunikationssysteme wird die gemeinsame Kommunikation und Positionierung immer wichtiger. Die Menschheit hat sich bereits daran gewöhnt, jederzeit und überall auf Multimediadaten zuzugreifen und dabei alle Arten von ortsbezogenen Diensten zu nutzen. Globale Navigationssatellitensysteme sind in der Lage, hochpräzise Positionierungsergebnisse zu liefern, wenn eine direkte Sichtverbindung verfügbar ist. Leider ist bekannt, dass schwierige physikalische Umgebungen die Leistung bestehender Systeme beeinträchtigen. Hier können aber andere, bodengestützte Systeme die bestehende, durch Satellitensysteme gegebene, Positionsschätzung unterstützen.

Das Ermitteln von positionsrelevanten Informationen aus einer einheitlichen Signalstruktur, die für ein bodengestütztes gemeinsames Kommunikations- und Positionsbestimmungssystem entwickelt wurde, kann vorhandene Systeme entweder ergänzen oder ersetzen. Ein solches Systemgerüst verspricht, die vorhandenen Systeme zu verbessern, indem es eine sehr genaue und zuverlässige Positionierungsleistung und eine erhöhte Abdeckung ermöglicht. Darüber hinaus führt die einheitliche Signalstruktur zu synergetischen Effekten.

In dieser Arbeit schlage ich ein auf Kanalschätzung basierendes gemeinsames Kommunikations- und Positionierungssystem vor, das eine virtuelle Trainingsmatrix verwendet: Diese Matrix besteht aus einem relativ kleinen Trainingsprozentsatz plus den erfassten Kommunikationsdaten. Über einen zentralen semi-blinden Schätzansatz können die erfassten Daten iterativ einbezogen werden, um den positionsrelevanten Parameter genau zu bestimmen, indem Informationen zwischen dem Kommunikationsteil und dem Ortungsteil des Empfängers ausgetauscht werden, wodurch Synergie entsteht. Ich schlage ein verallgemeinertes Systemgerüst vor, das die Verwendun verschiedener Kommunikationssystemtechniken erlaubt.

Der wichtigste positionsrelevante Parameter, die Signalankunftszeit, ist Teil eines physikalischen Mehrweg-Parametervektors. Das Schätzen der Ankunftszeit bedeutet daher, ein globales nichtlineares, mehrdimensionales Optimierungsproblem zu lösen, genauer gesagt, das sogenannte inverse Problem zu lösen. Ich stelle eine gründliche Untersuchung verschiedener Problemformulierungen und deren Variationen vor, die verschiedene Messungen und Schätzalgorithmen vorstellt und bewertet.

Eine große Herausforderung bei der Lösung des inversen Problems zur Ermittlung der positionsrelevanten Pfadparameter stellen realistische Mehrwegekanäle dar. Es zeigt sich, dass die meisten Parameterschätzungsalgorithmen in moderaten Mehrwegumge-

bungen gut funktionieren. Hier ist es mathematisch verhältnismäßig einfach, die zugrundeliegende Kostenfunktion zu optimieren, wenn nur die Anzahl der Beobachtungen die Anzahl der geschätzten Parameter überschreitet. Das typische realistische Parameterschätzungsproblem dagegen basiert auf Kanalschätzungen und geht davon aus, dass sogenannte Schnappschussmessungen verfügbar sind. Bei realistischen Kanalmodellen überschreitet allerdings die Anzahl der Beobachtungen nicht immer die Anzahl der Unbekannten. In dieser Arbeit schlage ich eine Methode zur optimalen Reduzierung der Problem-Dimensionalität als Problemlösung vor: Durch die gemeinsame Auswahl der Modellordnung und Parameterschätzung, unter Verwendung der approximierten und geschätzten Parameterkovarianzmatrix, wird die Modellordnung des Schätzproblems darauf optimiert, eine optimale Parameterschätzung zu gewährleisten und damit auch optimale Positionierungsergebnisse. Um diese Ergebnisse mit der optimal erreichbaren Lösung zu vergleichen, definiere ich in dieser Arbeit eine fokussierte, modellordnungs-spezifische Untergrenze. Darüber hinaus verwende ich die Soft-Informationen als Gewichtungsmatrix, um die Positionierungsleistung des Positionierungsalgorithmus zu verbessern.

Um das Zusammenspiel und die Durchführbarkeit der vorgeschlagenen Systemkomponenten zu demonstrieren, begutachte ich in dieser Arbeit einen Systemprototypen, welcher auf Multilayer-Interleave-Division-Mehrfachzugriff basiert. Das vorgeschlagene Systemgerüst und die untersuchten Techniken können für verschiedene bestehende Systeme eingesetzt werden oder bilden die Basis für zukünftige gemeinsame Kommunikations- und Ortungssysteme. Diese Arbeit zeigt, dass die vorgestellten Schätzalgorithmen prinzipiell auf alle Arten von gemeinsamen Kommunikations- und Positionierungssystemen übertragbar sind und schwierige physikalische Umgebungen und daraus resultierende schwierige Schätzprobleme bewältigen können.

Stichwörter: Positionierung, Navigation, Kommunikation, Schätzung, Optimierung, Partikelschwarmoptimierung, Auswahl der Modellordnung, informationstheoretische Kriterien

Contents

Chapter 1

Introduction

1.1 Motivation

In the era of ubiquitous, almost unlimited access to information and connectivity via communication devices, it becomes increasingly important to be able to rely on accurate position information, always and everywhere. Rapidly emerging positioning capabilities go hand in hand with a rapid increase in location-based services. It is straightforward that joining communication and positioning yields synergy. Joint international cooperation projects like the wireless hybrid enhanced mobile radio estimators (WHERE) project [RP08] and especially its successor, the WHERE2 project [NPD+13, RSS+13] emphasize that joint communication and positioning (JCAP) is an emerging, diverse and complex research area.

Mobile communication devices nowadays are already known to provide ubiquitous positioning information [SDM14]. First approaches to equip communication devices with additional positioning capabilities have been realized by incorporating a standalone communication system as well as a standalone positioning system within a single device. Nowadays, communication systems, like the long-term evolution (LTE) and 5G standards, are designed to provide positioning capabilities of their own by utilizing so-called positioning reference signals. Furthermore, today, many different positioning technologies assist and complement each other. Hence, state-of-the-art JCAP systems already offer accurate position information in the majority of all transmission scenarios. There is, however, room for improvement in terms of accuracy, reliability, coverage, latency, complexity and synergy. In the presence of multipath and in the absence of a line of sight (LOS) connection, location estimation reliability is a challenging goal and the accuracy also suffers from harsh physical environments. Since standalone communication systems and standalone positioning systems are designed based on inherently opposing demands, any interplay between the two inevitably leads to a tradeoff.

Novel JCAP system designs potentially allow to optimize both, the communication as well as the positioning unit, equally. They promise synergetic effects: Communication algorithms can be optimized by utilizing positioning information and, in return, positioning algorithms can be improved by utilizing communication information. Regarding the communication part resource allocation, cell handover and beamforming benefit

from positioning information. On the other hand, from a positioning point of view the communication system may be used to provide positioning reference signals and further positioning relevant information, like reference object coordinates.

Typical positioning-relevant information, like the time of arrival (TOA) or the angle of arrival (AOA) can be extracted from channel measurements. Communication systems inherently provide channel measurements, called channel estimates. Therefore, the approach to employ channel estimates in order to yield accurate position information is appealing. Moreover, since channel estimation is crucial for both communication and positioning, from a JCAP perspective it can be understood as the crucial link and also as the bottleneck of the JCAP system as envisaged in this thesis. For TOA and channel estimation based JCAP, high resolution parameter estimation, model selection, the system design, modeling and the theoretical performance limits are all factors having an impact on the overall system performance. Furthermore, all these factors depend on each other. Although, all of these factors on their own are rather matured research areas that have evolved over several decades, contributions on investigating them for JCAP purposes are rare and they still are an emerging development that requires deeper insights.

Data transmission and positioning each on their own have some at first sight different system requirements. For example, considering synchronization for communication, only coarse delay estimates are required at the receiver. A TOA based positioning system, however, requires accurate delay estimates. In favor of high data rates, for communication purposes, it is beneficial to keep the overhead due to transmitting known sequences, so-called pilot sequences, as low as possible. On the other hand, for positioning purposes, increasing the pilot power is principally beneficial.

Despite such contradictory demands, channel estimation based JCAP promises to be highly synergetic and a joint design approach can target to overcome opposing demands or at least find a reasonable tradeoff.

1.2 Context, Scope and Aim

Since channel estimation can be employed as the link between communication and positioning, the fundamental equations can be formulated in a generic manner. The idea of a generic channel estimation based JCAP system framework fundamentally yields the following questions:

- How accurately can the positioning relevant information be estimated theoretically in such frameworks under the assumption of realistic physical environments (severe multipath) in conjunction with a limited amount of observations to solve the underlying estimation problem?

- Which modeling approaches, parameters, system components and algorithms lead to optimal and realistic designs?

- Which tradeoffs regarding communication and positioning requirements have an impact on such frameworks?

- Do the proposed estimation techniques asymptotically approach the associated theoretical limits and are the typical theoretical performance bounds sufficient for joint communication and positioning?
- How can soft information improve positioning?

This thesis targets the clarification of these questions, paving the way to the greater goal: To improve existing and to learn for possible alternative future JCAP designs. Therefore, the aim of this thesis is to formulate a generic system, to derive, to assess and to exploit theoretical limits thereof, to assess possible estimation and detection algorithms and finally to demonstrate a prototype system design. Note that this thesis partially builds upon the results in [Sch12] and it takes up at the for this thesis most relevant insights provided in the dissertation and related publications [SBKH10], [SAH11b], [SAH11a], [SAH12] that are briefly summarized here:

- The TOA estimation performance for a scenario with a single-input single-output (SISO) framework, snapshot channel measurements and a interleave-division multiplexing prototype communication system [Sch12] indicate: Although the position can be estimated accurately in simplified two-path channel models, the system proposal fails to achieve accurate positioning results in realistic multipath channel scenarios. The framework proposed in [Sch12] requires modification in order to tackle the challenges in more realistic channel scenarios.
- Position estimates can be improved by estimating and employing soft reliability information for the positioning-relevant estimates [SAH11b].
- Targeting high resolution positioning-relevant parameter estimation problems leads to the formulation of multi-dimensional, non-linear global optimization problems that can be solved by a wide range of available solutions from the literature. In most cases, these solutions require some kind of initial guess for the parameter estimates. Alternatively, they can be solved via heuristic global optimization strategies, like particle swarm optimization (PSO) [SBKH10], which do not require any initial guess.
- Among other parameters, the positioning performance limits depend on the sampling frequency and the number of available observations [SAH12] of the channel estimates.

Therefore, in this thesis I

- formulate and theoretically and numerically assess a channel estimation based JCAP framework and the related estimation and optimization algorithms (the framework is generalized to cover different communication signal structures and hence is applicable to different broadband cellular network technologies, 2G-5G and possibly beyond).
- extend the single snapshot channel estimation approach to a single input multiple output (SIMO) and time series measurement setup and theoretically and numerically assess the involved benefits in terms of the positioning-relevant estimation performance.

- propose and assess joint model selection and parameter estimation solutions to enable reliable high-resolution parameter estimation in more realistic channel scenarios.
- draft how positioning relevant reliability or soft information can be employed beneficially in multiple ways.
- discuss the range of different receiver modeling approaches, problem formulations and theoretical related bounds, estimation algorithms and their applicability.
- provide overall results for a prototype system.

1.3 Thesis Outline

This thesis is structured in the following manner:

Chapter 2 embeds this work into a scientific context and provides JCAP basics. It briefly summarizes positioning principles, introduces state-of-the-art JCAP systems, drafts envisaged future system concepts and relates these topics to open questions and challenges. Afterwards it motivates channel estimation based JCAP, the fundament of this thesis and outlines associated challenges.

Chapter 3 introduces the mobile radio channel modeling as used in this thesis. Note that the system performance in this thesis critically depends on the underlying channel model.

Chapter 4 introduces the proposed system design. This chapter builds the required theoretic fundament for the generalized channel estimation-based system framework in this thesis.

Chapter 5 is this thesis' main chapter. This chapter introduces and assesses the chosen estimation algorithms and the parameter estimation method developed to tackle the estimation problem specific obstacles, which are typically imposed due to the presence of realistic multipath channels. All required Cramer-Rao lower bounds are provided as the theoretically optimally achievable MSE estimation performance under the assumption that the assessed estimator belongs to the class of minimum variance unbiased estimators. Moreover, I show how to determine soft information. More specifically I show how to determine the estimate's covariance matrix and for what this is useful. I numerically assess channel estimation, joint model selection, parameter estimation and system extensions and compare these results to the theoretical lower bounds in this chapter. Finally, positioning and the related theoretical bounds are introduced.

Chapter 6 assesses a range of information theoretic criteria and their applicability to this thesis. A joint model selection and parameter estimation approach based on soft delay estimates is introduced. Due to the fact that the classical lower bounds, namely the

Cramer-Rao lower bound, fails to be an appropriate bound in case of joint model selection and parameter estimation, I propose a novel bound called the focused order-related lower bound, which I show can successfully be used to formulate the optimally achievable MSE performance for such problems.

Chapter 7 introduces a possible prototype system that, without loss of generality, is in this thesis is based on multi-layer interleave division multiple access, a non-orthogonal combination of multiple access and multiplexing. The numerical results demonstrate the channel estimation based JCAP system proposal feasibility.

Chapter 8 briefly concludes this thesis.

For the sake of brevity and as a reference to the reader, all abbreviations, symbols and mathematical definitions used througout the thesis are summarized in the Appendices B and C. Longer derivations and calculations are provided in Appendix A.
The work in this thesis was part of and furthermore builds upon a joint project on interleave-division multiple access based joint communication and navigation. Parts of the work in this thesis have been already been published in [AH13, SBKH10, SAH11b, SAH11a, SAH12] or are currently under review in [AH20].

Chapter 2

Joint Communication and Positioning (JCAP)

The purpose of this chapter is to substantiate the research of this thesis by putting it thematically into a wider context with respect to research projects and ongoing developments in the related research areas. In order to understand JCAP systems, one has to understand a few basics of communication systems, positioning principles, state-of-the-art systems, ongoing research, the special role of channel estimation, challenges and open problems.

JCAP system approaches are designed to optimize many different performance metrics for communication as well as positioning. For communication systems the bit error rate, the channel capacity, the data rate, the coverage, the energy consumption, the cost and the complexity are important performance metrics. For positioning, on the other hand, the most critical performance metrics are accuracy and coverage, followed by scalability, cost and complexity, adaptiveness and responsiveness [FNI13]. Figure 2.1 attempts to visualize the network of related topics.

2.1 Communication Systems

A typical communication system consists of a number of transmitters, equal to or larger as one, and a number of receivers, equal to or larger than one. Both ends are connected by the physical channel in-between.

As shown in the second subfigure of Figure 2.2, in this thesis we assume

- the physical channel is a wireless channel model
- either the transmitter is a base station and the receiver is a mobile device (downlink) or the transmitter is the mobile device and the receiver is a base station (uplink)
- LOS signal propagation and multipath propagation
- more than three reference objects (e.g. base stations) are available for positioning

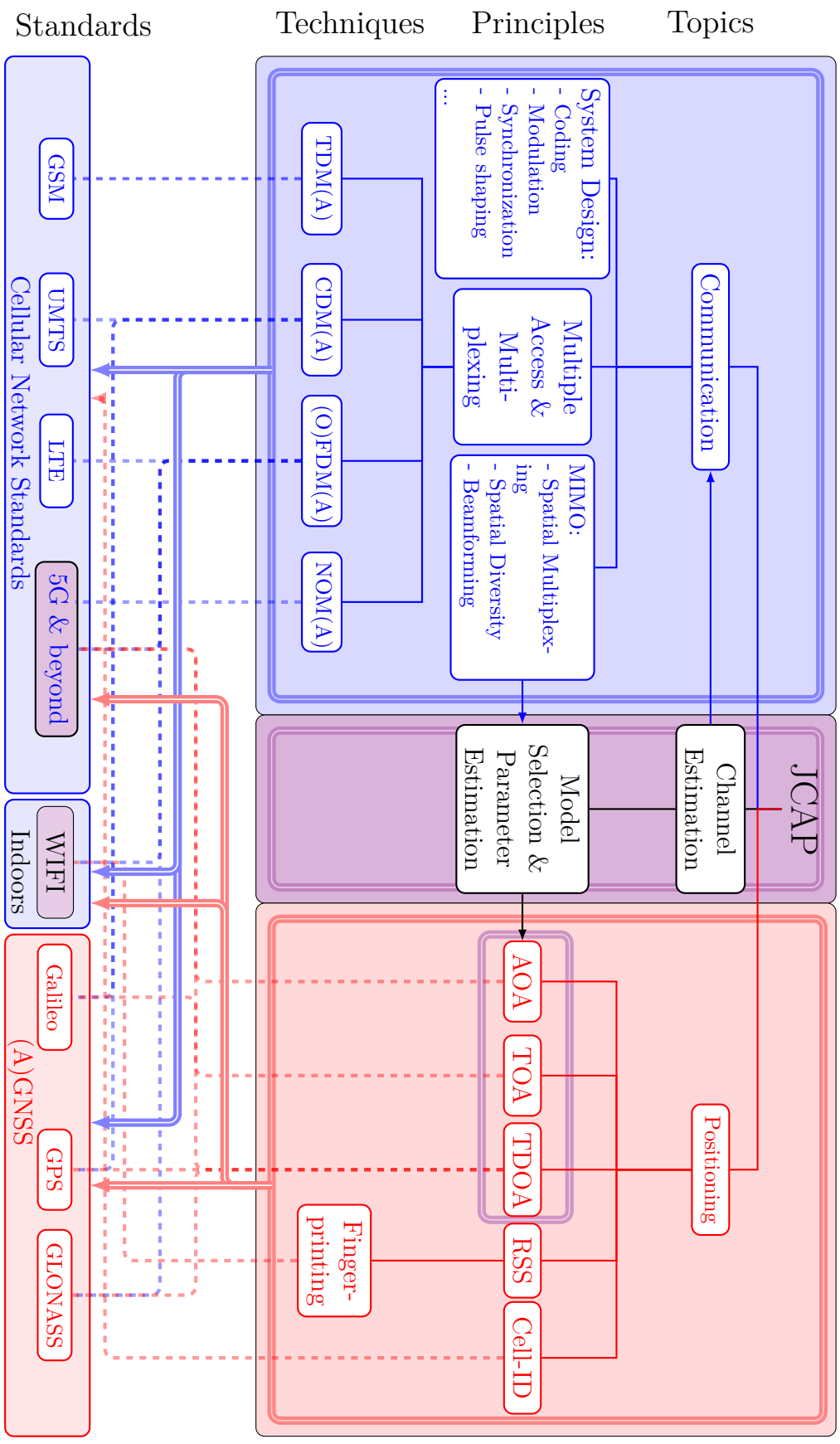
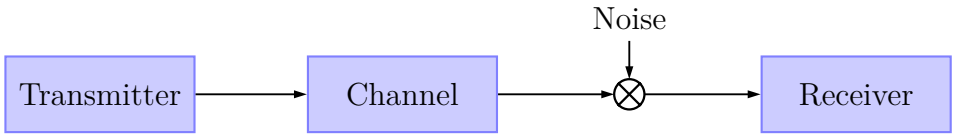
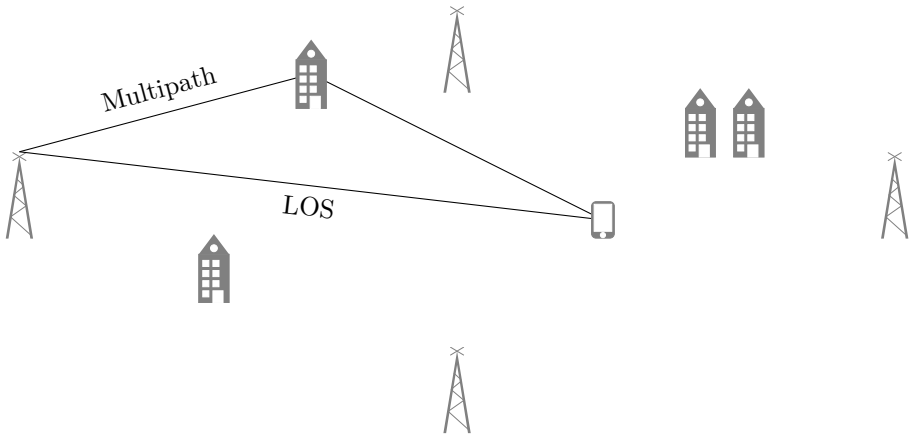


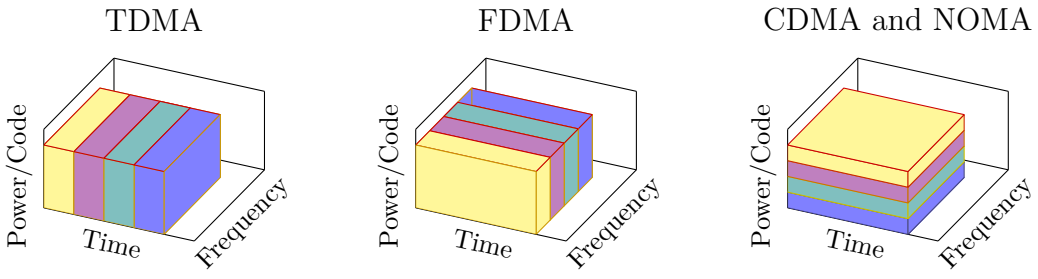
Figure 2.1: The JCAP topic overview has a tree structure with linked braches: Communication and positioning mutually depend on each other and they are unified especially for future systems like 5G.



(a) The main components of a single-user communication system setup are the transmitter, the channel and the receiver.



(b) The scenario that I assume in this thesis is a ground-based communication system setup connected by a wireless channel. The signals are assumed to propagate via multiple paths including a LOS path.



(c) The schematic visualizes the different concepts how to separate multiple users mathematically in the three domains time, frequency and power/code. Each color represents a different user.

Figure 2.2: The subfigures show a communication system in a), assumptions of this thesis in b), and multiple access schemes in c)

Multiple transmitters, or multiple users, are modelled according to the various mathematical principles of multiple access. The last two decades of mobile communication standardization have brought forth different multiple access schemes: a combination of time-division multiple access (TDMA) [FAG95] and frequency-division multiple access (FDMA) for 2G, code-division multiple access (CDMA) for 3G Universal Mobile Telecommunications System (UMTS) [SvROL99], orthogonal frequency-division multiple access (OFDMA) for 4G LTE [Erg09] and the potential setup for non-orthogonal multiple access (NOMA) for 5G [DLK+17, DWY+15]. All these different multiple access schemes operate in different mathematical domains, like the time domain, the code domain or the frequency domain, to superimpose and to allocate and separate the user data mathematically. Thereby TDMA, CDMA and OFDMA are all orthogonal schemes employing the mathematical orthogonality definition for user separation either in the time-domain, the code-domain or the frequency-domain. These schemes have the drawback that the number of users is limited by the orthogonality constraint. It is noteworthy that there is no consensus among researchers on categorizing CDMA as an orthogonal multiple access scheme [VDV19, page 168]. Here, I choose to categorize CDMA as an orthogonal multiple access scheme, if it is based on orthogonal codes.

Theoretically, the problem of a limited user number due to the orthogonality constraint is overcome by non-orthogonal schemes. This is the main reason for considering NOMA as a 5G multiple access candidate. Note that interleave-division multiple access (IDMA) [PLL03, PWW07] as well belongs to the non-orthogonal schemes. Since it is based on the idea that users can be separated by different permutations, called interleavers, IDMA can also can be classified as a special form of a code-division multiple access scheme. Like IDMA the closely related multiplexing concept interleave division multiplexing (IDM) has a few advantages in comparison with competitive multiplexing schemes. The combination of IDM and IDMA is dubbed multi-layer IDMA, as investigated in [HSF08]. To this end the preceding work to the research in this thesis was driven by the motivation to unify IDM/IDMA based communication and positioning [SH08, Sch12]. Visionary contributions in the realm of 5G listed IDMA among other NOMA approaches as a candidate for 5G multiple access [DWY+15, ARS16, VDV19]. Recently, the fifth generation decided to use both orthogonal multiple access (OMA) as well as NOMA. The fifth generation is continuing to use OFDMA, due to the ongoing usage of the cyclic prefix, frequency domain signal processing and the resulting manageable receiver complexity thereof [VDV19, page 33]. Furthermore, NOMA is considered by the 3rd generation Partnership project (3GPP) in Release 16 [VDV19, page 34]. NOMA's applicability in systems lacking centralized control or accurate channel state information is stated as it's actual attractiveness together with iterative processing and a simple implementation via IDMA offering comparably lower detection complexity than alternative solutions [VDV19, page 445]. The authors also demonstrated that IDMA offers a significant performance gain in random access as well as MIMO systems. In the document for discussion and decision the company Sony made the observation that "IDMA-based NOMA is suitable based on the agreed performance metrics and implementation related metrics" and proposed that "IDMA-based NOMA should be supported" [R1-18].

2.1.1 Transmitters

Each transmitter commonly has the purpose to transmit data error free and with a high data rate. Therefore, coding, modulation, multiplexing, pulse shaping and multiple antennas are usually employed on the transmitter Whereas coding has the purpose to prevent data detection errors by adding redundancy, higher order modulation and multiplexing is used to increase the data rate. Pulse shaping that meets the typical requirements for optimal data detection is needed as well. Furthermore, multiple antennas enable spatial multiplexing or instead the advantages of spatial diversity or beamforming can be For the purpose of channel estimation, synchronization and positioning at the receiver side pilot symbols can be transmitted together with the data. Thereby it is desirable that the overall pilot power is as small as possible in favor of a higher data rate from a communication point of view. From a positioning point of view a larger pilot power is desirable.

2.1.2 Receivers

The receiver has the purpose to receive signals and detect data. Therefore, the receiver can have multiple antennas and typically comprises functionalities like matched filtering, channel estimation, synchronization, demodulation, decoding, user and data detection. The choices of detection and estimation algorithms are crucial for the receiver design and for the performance in terms of bit error rate (BER). Hence the transmitter design has to be matched to the receiver design. Channel estimation is inherently needed for data detection in communication systems. Since channel estimates also bear positioning-relevant information like the TOA they additionally can be exploited for positioning.

Channel estimation can be employed as the link between data detection (communication) and parameter estimation (positioning) and hence this thesis concentrates on channel estimation based joint communication and positioning with a focus on positioning relevant parameter estimation. Obviously, the channel estimation performance is crucial and can be seen as a system bottleneck. Whatever specific communication system structure is applied, should therefore allow accurate channel estimation. Apart from this, obviously, various multiple access techniques and multiplexing and modulation schemes can be applied. Results of earlier contributions related to this research [AH13, Sch12, SAH12, SAH11b] substantiate the view, that the special communication signal design part is less critical than positioning-relevant estimation strategies. Note, that typical positioning systems like GPS are special forms of communication systems with the special purpose to estimate the position.

2.2 Positioning Principles

Note that there are two categories of positioning, network based and mobile based positioning. The term mobile refers to a person, an object or a device. The term network stands for multiple reference objects, objects with known coordinates and some kind of communicating processing unit. These network reference objects could refer to a satellite system, a set of base stations of a cellular wireless system or a set of connected Wi-Fi

access points or routers. In the network-based scenario a mobile device transmits a signal to the reference objects, the receivers, that are capable of positioning by utilizing the received signals. A comprehensive overview on network-based positioning is provided in [STK05]. If I refer to mobile positioning, or self-positioning, I mean the person's or the object's functionality to determine the person's or the objects own location. Then the mobile unit is the receiver and the reference objects are the transmitters. From a physical and mathematical viewpoint location-dependant information has an impact on the electromagnetic signal propagation such that a channel impulse response function can be formulated relating any received signal to the location determining parameters and the transmitted signal. Specifying this mathematical relationship the following parameters constitute the location-dependant information: The propagation delay, the propagation angle (angle of departure and angle of arrival), the propagation attenuation as well as multipath propagation delays and angles. Consequently, the most popular three different positioning principles are: Positioning based on the propagation time estimation (time-of-arrival, time-difference-of-arrival (TDOA) and Time-of-flight (TOF)), positioning based on the angle-of-arrival (AOA) and positioning based on fingerprinting, that is positioning based on the received signal strength or the power delay profile (PDP). An overview is provided in [Gez08] as in Figure 2.3.

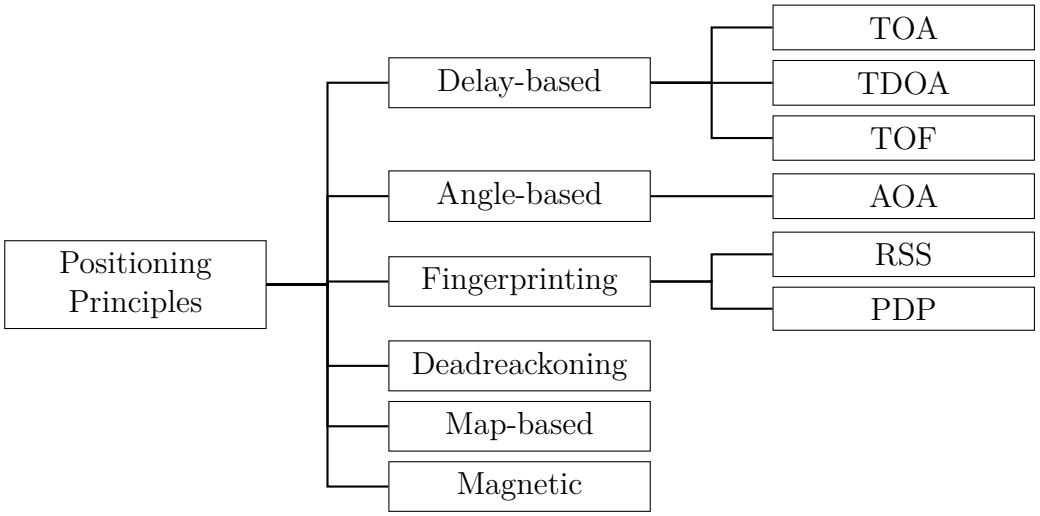
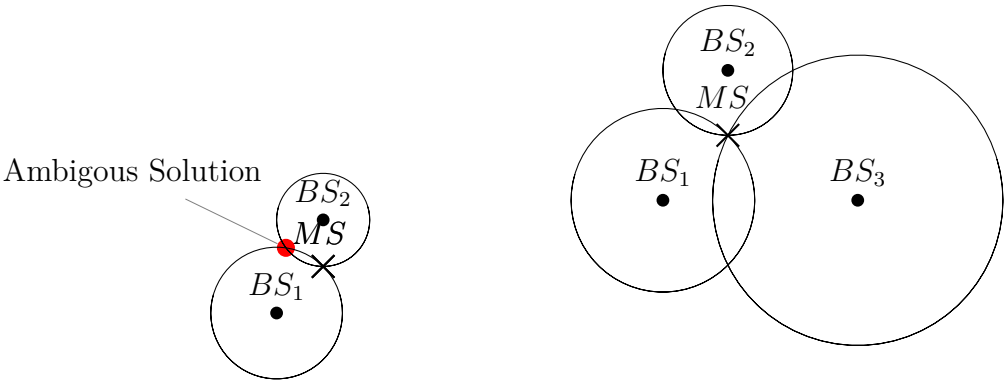


Figure 2.3: The overview shows the different categorisation of positioning principles.

2.2.1 Propagation Delay-based Positioning

In this chapter I will interchangeably use the expression reference object and basestation without loss of generality. Let (x_i, y_i, z_i) denote the ith base station Cartesian coordinates and let (x_{MS}, y_{MS}, z_{MS}) be the unknown device position. Further, let N_B be the number of base stations. For now, consider a downlink scenario, where the base stations transmit a signal and the mobile device receives the signals, if not said otherwise.



(a) Ambiguous TOA solution (b) Distinct TOA solution

Figure 2.4: The TOA positioning principle requires three distance measurements and hence ≥ 3 reference objects in order to find an unambiguous distinct solution.

2.2.1.1 Time of Arrival (TOA)

The TOA principle first of all is based on the assumption that the line-of-sight (LOS) is not obstructed. Furthermore, it is based on the assumption that via some kind of estimation step at the receiver side the estimated propagation delays of at least three transmitter-to-receiver links for 2-dimensional positioning (four links for 3-dimensional positioning) are known and hence can be used to determine the transmitter (TX)-to-receiver (Rx) distances. Let the ROs be base stations and the number of base stations be N_B. Then this is achieved by employing the linear relationship between the ith of the N_B links Tx and RX distance d_i and LOS propagation delay $\tau_{\text{toa},i}$, the TOAs, to estimate the distances also called pseudoranges by

$$\hat{d}_i = c_0 \tau_{\text{toa},i} \quad \forall i \in \{1, \ldots, N_B\}, \tag{2.1}$$

where c_0 denotes the speed of light, which is known to be $c_0 = 3 \times 10^8\,\text{m/s}$. In the 3-dimensional setup, a set of three nonlinear system equations, which later can be used to determine the MS position, can be formulated. The true distances d_i are related to the positions $\mathbf{p} = [x_{\text{MS}}, y_{\text{MS}}, z_{\text{MS}}]$ via

$$d_i = \sqrt{(x_{MS} - x_i)^2 + (y_{MS} - y_i)^2 + (z_{MS} - z_i)^2}. \tag{2.2}$$

The necessity of a third and fourth equation, hence additional reference objects, for 2-dimensional and 3-dimensional positioning, respectively, introduced to prevent ambiguity, results from the system equation's nonlinearity. Obviously, the pseudoranges d_i are gained by employing some LOS delay estimation step at the receiver side. Consequently, these estimated distances stacked in a vector depending on the coordinates will always be corrupted by an unbiased, or a zero mean, estimation error $\boldsymbol{\epsilon} = [\epsilon_1, \ldots \epsilon_{N_B}]$ with covariance matrix $\mathbf{C}_\epsilon = \text{diag}\{[\sigma^2_{\epsilon_1}, \ldots, \sigma^2_{\epsilon_{N_B}}]\}$:

$$\hat{\mathbf{d}} = \mathbf{d}(\mathbf{p}) + \boldsymbol{\epsilon}. \tag{2.3}$$

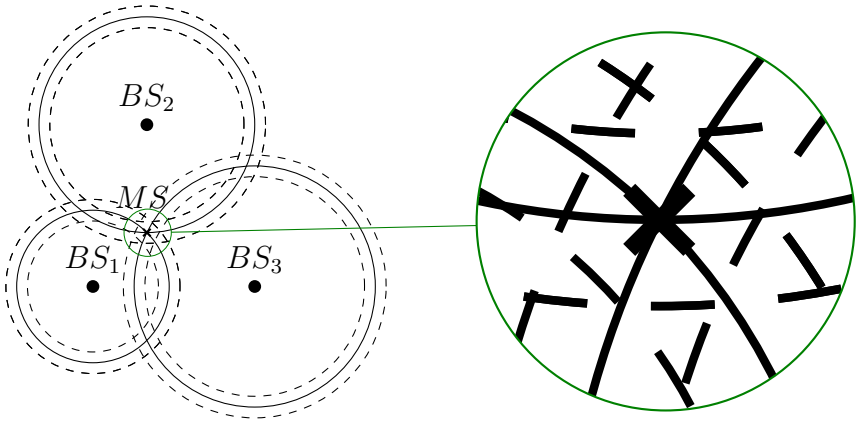


Figure 2.5: Erroneous TOA measurements lead to an intersecting area instead of a unique solution.

These errors will destroy a unique solution. More specifically the intersection point of three circles, as can be seen in Figure 2.5, instead widens to a whole intersecting area as a remaining solution area. To deal with the reality of distorted measurements, it is well known, that the error has to be included in the set of equations such that

$$c_0^2(T_i - T_0)^2 = (x_{MS} - x_i)^2 + (y_{MS} - y_i)^2 + (z_{MS} - z_i)^2 - \epsilon_i^2, \quad \forall i \in \{1, \ldots, N_B\}. \quad (2.4)$$

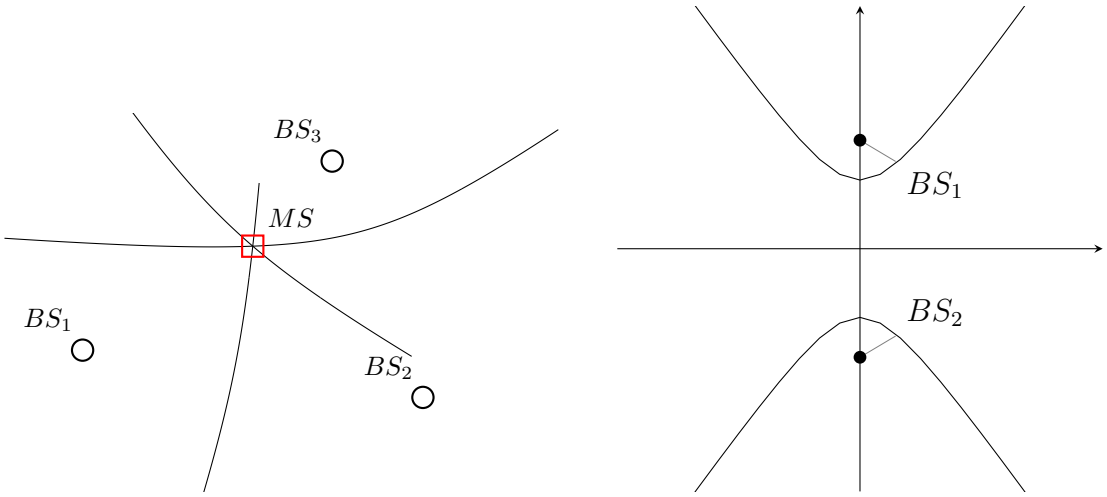
A unique optimal solution in the least squares and in the maximum likelihood sense then can be found by choosing the position that minimizes the sum of all quadratic errors ϵ_i:

$$\hat{\mathbf{p}} = \arg\min_{\tilde{\mathbf{p}}} \left\{ \left(\hat{\mathbf{d}} - \mathbf{d}(\tilde{\mathbf{p}}) \right)^T \mathbf{C}_{\boldsymbol{\epsilon}}^{-1} \left(\hat{\mathbf{d}} - \mathbf{d}(\tilde{\mathbf{p}}) \right) \right\}. \quad (2.5)$$

Another well known and practically more problematic TOA positioning error source is introduced by not perfectly synchronized clocks at the BSs and the MS, since accurate distance measurements (2.1) depend on the time difference $\tau_{1,i} = T_i - T_0$. This means all clocks have to be synchronized. This especially poses a problem for the MS. Hand-held devices are usually mass-market products and therefore their development commonly is constrained using to low cost hardware solutions. Typically, T_0 has to be treated as an additional unknown, requiring another additional equation and hence another reference object. On the network side employing highly accurate and hence costly clocks is more reasonable than in the MS. Moreover, on the network side errors due to asynchronous clocks can be measured and can be integrated into the equations.

2.2.1.2 Time Difference of Arrival (TDOA)

As already mentioned, asynchronous clocks between the BSs and the MS pose a serious problem in reality. More specifically, the value T_0 is unknown and annoying. By employing the TDOA principle instead of TOA, one cleverly circumvents actually using T_0 in the following manner: Instead of calculating the BS to MS distances directly, one



(a) TDOA principle (b) Example hyperbola

Figure 2.6: Measuring TDOAs instead of TOAs yields intersecting hyperbolas instead of intersecting circles as can be seen in a). Thereby the base stations coordinates are the hyperbola foci as can be seen in b).

builds another set of equations by using time differences and thereby cancelling out the undesirable variable T_0

$$c_0 \Delta T_{i,j} = c_0(T_i - T_0) - c_0(T_j - T_0) = c_0(T_i - T_j) \quad \forall i, j \in \{1, \ldots, N_B\}. \tag{2.6}$$

Let $N_{BS} = 3$ and assume that TDOA measurements between BS $i \in \{2, 3\}$ and are available. Then

$$c_0 \cdot \Delta T_{i,j} = \sqrt{x_{MS} - x_{BS_i})^2 + (y_{MS} - y_{BS_i})^2 + (z_{MS} - z_{BS_i})^2} - \sqrt{x_{MS} - x_{BS_1})^2 + (y_{MS} - y_{BS_1})^2 + (z_{MS} - z_{BS_1})^2}. \tag{2.7}$$

Each of the distance differences $\Delta d_{i,1} = c\Delta T_{i,1}$ to BS_1 related to a TDOA in (2.7) can geometrically be associated with a hyperbola in two dimensions and with a hyperboloid in three dimensions. Considering the two-dimensional case, every point on the hyperbola yields an equal distance difference $\Delta d_{i,1}$ to the two foci given by the locations of BS_i and BS_1. Note that in contrast to TOA where two circles intersect at two points, two hyperbolas intersect in only one distinct point intersect In Figure 2.6 the actual TDOA positioning principle as well as the hyperbolic nature of a single TDOA associated distance difference is outlined.

Obviously, apart from the advantage that TDOA even works when the MS and the BSs clocks are asynchronous, TDOA is susceptible to the same error sources as the TOA method. with the only difference that, in case of errors, the intersecting area will have a different shape than in case of TOA.

2.2.1.3 Round Trip Time of Arrival (RTTOA)

Assuming channel reciprocity, the RTTOA principle employs round-trip delay measurements between two transceivers of any kind:

$$d_{RTTOA} = c_0(T_R - T_P). \tag{2.8}$$

Here T_R denotes the round trip time and T_P denotes the processing time. The transceiver clocks do not have to be synchronized. For further details, I refer to [SDM14]. The overall setup measures time delays and the associated distances, hence the measurements geometrically specify circles similar to the ones of TOA measurements.

2.2.2 Angle-based Positioning

For the sake of simplicity and without loss of generality let us consider a two-dimensional downlink scenario, where a device provides angle of arrival (AOA) φ_i measurements of at least two, that is $i \in \{1, 2\}$ reference objects,i.e. base stations. Note that, in order to obtain these measurements, usually, multiple receive antennas are required. These angles specify straight lines, describing possible locations for the device. The unique position is the intersection of these lines. Obviously, for known base station positions (x_i, y_i) and an unknown device position (x,y) this geometrically leads to a system of equations that has to be solved. Additionally, taking into account an unknown device orientation φ, more than two equations are needed, that is more than two reference objects, in order to dissolve all ambiguities. Then the system of equations reads:

$$(y - y_i) = \tan(\varphi_i)(x - x_i + \varphi), \text{ for all } i \in \{1, N_B\}. \tag{2.9}$$

The principle is visualized in Figure 2.7.

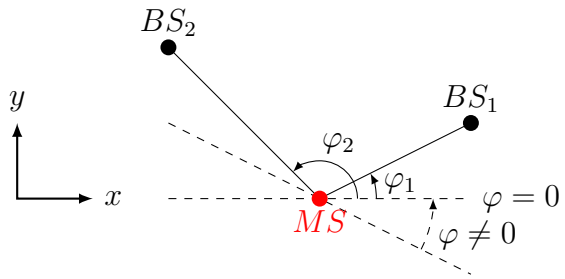


Figure 2.7: The AOA principle

2.2.3 Fingerprinting

The term fingerprinting related to positioning covers all methods that are based on the following principle: In a first step a fingerprint database for specified locations on a grid are measured and collected in an offline phase. Afterwards, in a second step, these fingerprints are then compared to online measurements. The best match is chosen via different approaches. A fingerprint measurement can be measurement that qualifies as

a fingerprint, e.g. an impulse response measurement or the power delay profile. Often it is a received signal strength. A comprehensive survey on the subject can be found in [KGA17]. A shorter mathematical introduction is provided in [HPALP09].

2.2.4 Dead Reckoning

The principle of dead reckoning is to combines apriori knowledge about the last determined position with directional information and velocity estimates, in order to estimate the new position [FNI13]. This principle has the drawback that estimation errors are cumulative over time.

2.3 State of the Art

Obviously, the most popular omnipresent JCAP system is embodied in all the mobile devices providing location based services, allowing us to navigate by determining the devices position and to be found in case of an emergency call. The ongoing evolution of positioning techniques in cellular networks, from 2G to 4G [Cam17] and also 5G, has led to a geared cooperation of different systems and hybrid techniques, working together and complementing each other to yield a high accuracy and coverage. Besides satellite based navigation, today, cell-ID based techniques, further techniques like the assisted global navigation satellite system and the capability to use specified positioning reference signals ensure a high reliability and accuracy. Next-generation techniques will more than before operate as a whole complex joint communication and positioning network unifying various varying techniques and concepts. An important novelty for the next generation will be that standalone positioning capabilities are specified directly from the beginning and not just as add-ons like before. [DRZ15]. Thereby, state-of-the-art joint communication and positioning entities can be categorized into

- parallel systems provided in a single device
- communication system extensions
- positioning system extensions
- systems based on an unified signal structure.

The usage of GPS on mobile phones embodies two parallel systems the communication system and the positioning system in a single device. Since the typical communication signal structure bears positioning relevant information, different approaches have been employed for position estimation via standalone communication systems, like for instance wireless local area network (WLAN)-based positioning. WLAN and the 1G-4G of mobile communication are examples for communication systems, which have been extended in order to additionally provide positioning information. On the other hand, the typical satellite navigation signal structure is utilized to also transmit data like for instance the reference objects' position, although it was designed for pure positioning purposes.

Although the standardized signal structures for communication and positioning show similarities typically communication and positioning systems conflict considering the system requirements and demands. Being optimally designed solely for either communication or positioning constitutes the main drawback for either a communication or a positioning system extension. Another approach is to design a joint communication system and positioning system, based on a unified signal structure.

2.3.1 Global Navigation Satellite Systems and Mobile Communication

The most popular, well known global navigation satellite system (GNSS) is the American global positioning system (GPS). Although globally the significance of GPS is unchallenged, other less famous GNSS systems coexist with GPS: The European system Galileo, the Russian Global Navigation Satellite System (GLONASS), the Chinese system called Beidou, the Japanese system Quasi-Zenith Satellite System as well as the Indian system called the Indian Regional Navigational Satellite System (IRNSS). Note that, today, at first the commercial and civil purpose of a GNSS system leaps to our mind. GPS, however, was originally developed and launched in the seventies as a military system, followed then during the cold war by the GLONASS system as a competitor to GPS also primarily for military use. Therefore, it is noteworthy that the European system Galileo has been initiated for civil use.

All the mentioned GNSS systems have in common that they employ the TOA or TDOA to further determine distances and the desired position information. Obviously, the challenges related to deploying satellites placed in orbits in space are different from than those related to ground-based positioning systems. Nonetheless, these issues are essential for GNSS systems. More specifically, note that the authors of [SDM14] address and discuss issues like satellite visibility, coverage, power, launch cost and so forth. The authors of [SDM14] also provide further insight on the GPS architecture, which can be categorized into a space segment, a user segment and a control segment.

Over the last two decades hand held mobile devices became smaller and cheaper.

Nowadays smart phones have replaced cell phones as a mass market product. Commonly, smart phones are equipped with a GPS-chip, since they also have become a mass market product, enabling permanent location awareness as well as location-based services of all kind and improved handover from one cell to another and so forth. Nonetheless, integrating a GPS-chip into a smart phone or any other communication device means employing two systems in one device, having two different signal structures and hence receiver algorithms for the communication and the positioning side. GPS is known to use 24 satellites, whereas Galileo uses 30 satellites in order to broadcast their signals. Each signal is received by the mobile station MS that has to separate the signals and that has to estimate the TOA or TDOA. The accuracy of the propagation time estimates is directly related to the positioning error. Additionally, data that comprises information about the satellites position information, has to be detected at the receiver, the MS, since it has to be known at the receiver side to be utilized as a reference object (RO). The signal separation for GPS as well as the European Galileo system is performed by employing code devision multiple access (CDMA), the Russian system uses frequency devision multiple access (FDMA). In GPS, CDMA is used in combination with pseudo random noise (PRN) sequences, that is, two code types are used, the so called coarse/acquisition (C/A) codes for civil use and the encrypted more accurate precision/encryption (P/Y) codes for military use. Thereby, the C/A codes enable a resolution of approximately 293 m, modern receivers increase this resolution to 1% thereof to approximately 2.93 m. The codes that are employed for military purposes, the P/Y codes are approximately ten times more accurate yielding a resolution of approximately 30 cm due to a ten times higher so-called chip rate, a value measuring how many CDMA-codebits, or chips, are transmitted per time unit. The C/A codes used in GPS receivers have good autocorrelation properties, which are necessary for accurate propagation time estimation as well as good cross correlation properties that are needed to guarantee orthogonality between different signals from different ROs. Being named after their inventor they are called Gold codes [Gol67], and are constructed by using linear feedback shift registers (LFSR) and especially designed generator polynomials. At the transmitter side the signals are modulated by binary phase shift keying (BPSK).

The propagation time estimates are corrupted by typically either randomly occurring errors like thermal noise or by systematic errors like clock errors, or by errors due to the speed of light variability in the ionosphere. Systematic errors can be combated by utilizing the fact that they usually are known to be highly correlated in time and space and hence can be measured in reference stations and then again be transmitted to the MS via augmentation systems like Differential GNSS (DGNSS) [SDM14] .

Depending on the required positioning accuracy, DGNSS can be utilized as a ground-based augmentation system in order to achieve higher accuracies by combating systematic errors. Further globally operating augmentation systems are listed in [SDM14].

Without doubt GNSS systems have evolved to matured, powerful and in most scenarios that come to our mind accurate positioning systems. There still are, however, NLOS and multipath scenarios, in which classical GNSS systems fail to perform as good as possibly desired. GNSS systems assume first of all that a LOS path exists between transmitter and receiver and second of all that the signal does not travel via other paths than the shortest LOS path. Consequently, non-line-of-sight (NLOS) as well as multipath propagation

deteriorates the GNSS positioning accuracy severely due to these receiver-sided modelling errors. Especially for GNSS systems, as already mentioned earlier, utilizing that the speed of light equals 3×10^8 m/s constitutes an idealized assumption, which is violated depending on the physics of the atmosphere, for instance it is violated in the ionosphere. Strong signal attenuations like for instance indoors also lead to a reduced positioning accuracy. In such environments additional systems, which assist the GNSS systems can be immensely helpful.

2.3.2 Today's Mobile Communication Standard Stand-alone Positioning Capabilities

Today's mobile communication standard has reached the fourth generation (4G) and is well known under the name long term evolution (LTE) or its extension LTE-Advanced (LTE-A). Besides traditionally employing GPS, the 4G standard specifies different standalone positioning capabilities [Cam17] that allow to determine the position of a MS, or as they are called in 4G the user equipment (UE). Partially, earlier generations, the global system for mobile communication better known as GSM and the Universal Mobile Telecommunication System better known as UMTS (2-3G), have breeded positioning techniques such as assisted GNSS (AGNSS), the Cell-ID and the Enhanced Cell-ID-based method, that are the predecessors to those used in LTE and therefore should be mentioned here. The LTE positioning techniques can be categorized into network-based, UE-based, UE-assisted and network-assisted approaches.

2.3.2.1 Assisted Global Navigation Satellite System (AGNSS)

AGNSS was specified in order to yield a better GNSS performance with respect to the weak signal reception in harsh physical environments like indoor scenarios and urban canyons, faster times to first fix and decreasing signal acquisition times and consequently saving power. The approach is based on connecting to a GNSS reference network equipped with receivers that have a clear LOS vision. Integrated in the cellular network this reference network is able to harvest all sorts of useful information like the approximate BS position, satellite visibility, ephemeris, clock correction, Doppler and so forth [SDM14]. This data is communicated back to the cellular network in case it is needed.

A GNSS equipped UE can make use of AGNSS. The UE assisted setup aims at measuring the positioning relevant information, reference signal time-difference (RSTD) for the observed TDOA (OTDOA) technique, in the UE, thereby being assisted by data transmitted from a location server, in order to hand the measurements over to the location server [Fis14] in return.Hence, although the measurements are carried out in the UE communication between the UE and a location server is required.

2.3.2.2 Cell-ID and Enhanced Cell-ID

Compared to AGNSS, the Cell-ID principle is rather simple, determining whether different nearby BS signals are or are not received. More precisely, the Cell-ID method only has to extract the transmitter ID information of commonly ≥ 3 signals from N_B different

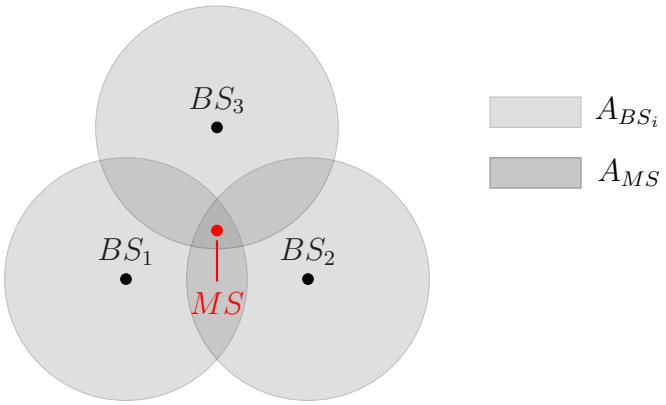


Figure 2.9: The Cell-ID positioning principle can be outlined by the intersection of BS coverage areas.

BSs, assuming that the position solution lies then in the intersecting area A_{MS} of the coverage areas A_{BS_i} for the $1 \leq i \leq N_B$ signals of these BSs. Hence, the MS position is determined by evaluating the coverage information of the nearest BSs. Naturally this approach only yields a rough positioning accuracy depending on the sizes A_{BS_i} as well as the number of Bs N_B. Nonetheless the Cell-ID and the Enhanced Cell-ID methods are known to be reliable and therefore they have been preferred for a long time over other standalone positioning capabilities. The authors of [TV04] provide an experimental study and a further discussion on how the Cell-ID can be exploited in an efficient manner.

2.3.3 WLAN-based Positioning

The well-known drawback of GNSS based positioning, is its bad performance indoors due to the high probability for NLOS propagation. WLAN signals are able to penetrate walls and they are often available on mobile devices. Fortunately, WLAN-based positioning, whenever possible, alternatively is used by some devices in order to achieve a high positioning accuracy indoors as well as in urban areas. WLAN positioning principally can be categorized by techniques that are based on angle estimation, techniques that are based on TOA or TDOA and trilateration, techniques that are based on RSS estimation and trilateration and techniques based on fingerprinting. Thereby, positioning via multilateration by employing AOAs [PS06], RSSs [KK04], TOAs [LP04] or TDOAs rquires a LOS component. Hence, these methods have the drawback that they are prone to errors due to multipath and NLOS conditions. Especially indoors this effect is very pronounced, and the accuracy is rather low. Better accuracy is achieved by fingerprinting and it can further be enhanced by additionally employing digital map data if available. In their review paper the authors of [KGA17] name two important reasons that explain the growing popularity of indoor fingerprinting with WLAN signals: The large WLAN signal deployment in public buildings and the receivers inherent ability to provide RSS measurements without extra hardware. Obstacles like walls, furniture and humans are the reason for shadowing, reflection, diffusion and diffraction and that indoors a material related signal

attenuation as well as multipath propagation has to be taken into account [STK08]. Since fingerprinting requires off-line measurements for the fingerprint database and the environment realistically often changes over time various techniques to update changes have been implemented. In large cities different self-sufficient technologies unifying have been developed and are used or at least available [Sky, awi]. Future systems will collect and unify relevant data from a homogenous network to ensure a high seamless coverage and accuracy. Another comparative study of WLAN location fingerprinting methods provides mathematical formulations [HPALP09]. In review contributions [KGA17] and others dealing with practical issues like [HJLY14] it is reported that for conventional WLAN signal fingerprinting issues like heterogenous devices, the map construction, correlated access point measurements, limited coverage and outlier detection are especially challenging. Motivated by the increasing number of mobile devices indoors, combinations of TOA and fingerprinting positioning has been proposed [KK15, KK04].

2.4 Future 5G Positioning

Note that visionary contributions located in the research realm of next generation mobile networks (NGMN) promise high speed, a high capacity, fairness, full coverage, high accuracy, scalability, high backward compatibility for the next generation 5G compared to its predecessors. Consequently, 5G currently is especially important, since it will be a seamless patchwork integrating and combining a multitude of existing and upcoming communication and positioning systems and hence it certainly will be the most widely used and important JCAP system on earth.

The next generation mobile networks alliance (NGMN) envisaged 5G to support positioning capabilities in three-dimensional space with an outdoor accuracy from 10 m to < 1 m at 80 % of the time, an indoor accuracy below 1 m and real-time high speed tracking capabilities [Nex15, KCW+17]. Thereby, the localization is supposed to cooperate and communicate with external techniques and should offer positioning information to an application programming interface. The authors of [ABC+14] also outline a prospective 5G framework and predict that it will be highly integrative with respect to LTE and WLAN for seamless user experience and high-rate coverage. They emphasize the importance of millimeter wave and massive MIMO related issues and challenges, such as path loss, architectural requirements and pilot contamination. Another issue for 5G will be a weather-related change of channel conditions. Rain is known to attenuate in the millimeter wave range [ZL06]. Additionally, the survey in [ARS16] addresses that 5G will be identified with an anywhere at anytime availability and coverage for an enormous number of connected devices, ranging in the order of thousands of connected devices, and at the same time it is proposed to enable a reduction in power consumption and energy usage. The authors of [DRZ15] illuminate the vision of 5G from a positioning perspective and point out that integrating positioning capabilities from the beginning is crucial, since this would enable a joint system with lower costs, yielding higher accuracy and coverage and further allowing the communication system to fully benefit from the available position information. Further, the authors of [DRZ15] stress that the communication system itself will be better suited than the preceding generations to combat the challenges involved with

positioning: A higher line of sight (LOS) and less harmful multipath probability will be achieved by a denser network. Furthermore, higher frequency and signal bandwidths will ease the challenges to accurately perform high resolution parameter estimation to determine positioning relevant information like the time of arrival (TOA). Another important feature named in [DRZ15] is the device-to-device (D2D) communication that enables cooperative and hence accelerated positioning [WLW09] with enhanced accuracy due to a larger amount of observations. Massive multiple input multiple output (MIMO) systems are investigated to assess localization and orientation estimation [SGD+18, SGD+15]. Proposals how to make good use of non-line-of-sight NLOS multipath components can already be found [MWBAS19]. Note that in [NLW+17] channel parameter estimation is predicted to be performed by using the space-alternating generalized expectation maximisation (SAGE) [FTH+99] or the joint iterative maximum likelihood estimation (RIMAX) [RSK06] algorithm. Note, that actually a multitude of high-resolution parameter estimation algorithms principally qualify for these kinds of problems and hence for 5G.

Concerning the 5G communication signal structure design among all users non-orthogonal multiple access (NOMA) has become a widely accepted viable candidate for dealing with the predicted large number of users in 5G. This credo is reasonable, since NOMA promises typical limitations of its orthogonal competitors: Orthogonal multiple access is limited by the resource block and the related orthogonality itself. Further, orthogonal schemes require synchronization, which becomes unpractical and difficult to deal with for an increasing number of users. NOMA can be interpreted as a framework that does not require synchronization and can be used to serve multiple users within the same orthogonal resource block regardless, which specific multiple access scheme is actually used [DLK+17].

2.5 Channel Estimation and JCAP

The channel impulse response, and consequently channel estimates, are mathematically related to positioning relevant information like the TOA τ_1 or AOA φ_1 and further parameters collected in the vector $\boldsymbol{\theta}$, which are not relevant for positioning:

$$\tau_1, \varphi_1, \boldsymbol{\theta} \mapsto h(\tau_1, \varphi_1, \boldsymbol{\theta}). \tag{2.10}$$

This relationship commonly is highly non-linear and it is determined by the instantaneous physical environment. Unfortunately, the relationship to τ_1 and φ_1 is at least partially inseparable to the relationship to the other parameters in $\boldsymbol{\theta}$. More specifically this means that extracting the TOA or AOA commonly requires at least partially extracting $\boldsymbol{\theta}$ as well. Hence, high resolution parameter estimation and multi-dimensional nonlinear global optimization is required. Consequently, it still is a challenge to find a system design to exploit this relationship between channel estimates and positioning relevant information optimally for JCAP. Nonetheless, channel estimation is inherently required in communication systems and the channel estimates can be used to determine positioning relevant information.

If a communication system design is already specified, researchers investigate, how the channel estimates can be used efficiently to provide additional positioning information.

Especially, in the realm of 4G research, due to the predefined so called positioning reference signals implemented in the 4G standard, the idea to use the communication signal structure at hand additionally for positioning became more pronounced. The authors of [DLSSG+12] proposed to estimate the channel and the positioning relevant time delays directly from the received values at the receiver side in a joint manner, and they also discussed issues regarding channel modelling for that purpose. In [dPRLSSG+14] the authors addressed the issues related to receiver sided channel modelling and joint channel and time delay estimation.

If, on the other hand, the communication system is not designed as a foregoing step, but instead is designed jointly with the positioning system this obviously is advantageous in that sense that the positioning performance is not limited by a predefined system design. Apart from that, it is desirable to have results that obviously are portable to different kinds of communication system designs with only few modifications. These are the main reasons why, in this thesis, I choose to stick to an (as far as possible) generalized and simplified system framework, whenever this is reasonable and possible to do. Focusing on aspects and most of all estimation techniques related to multiple antenna systems the entanglement of channel estimation and positioning was investigated in [Mia07] in a rather general manner and provides a good overview on this subject.

Preceding work, with partial intersections to the work and findings of this thesis, is presented in [SBKH10, SAH11b, SAH11a, SAH12, Sch12, AH13] as it belonged to the same research project. This project targeted to investigate, design and assess a ground-based joint communication and navigation system based on an interleave-division multiple acces (IDMA) and an interleave-division multiplexing (IDM) signal structure. With increasing project progress, we learned that the specific signal structure design is not as critical for the positioning performance as the limitations, obstacles and in general issues related to efficiently estimating the multipath propagation path parameters in a realistic scenario. From a positioning point of view the specific multiplexing and multiple access design turned out to be more or less interchangeable under the condition that accurate channel estimates are available. This especially is emphasized by splitting the processing steps of channel estimation and multipath parameter estimation into sequential processing units. The idea is that the parameters estimates are obtained in a second step by employing the channel estimates, which are estimated in a first step, as observations, or measurements. Therefore, by reviewing [Sch12] it becomes clear that the preciseness of channel estimation itself is critical for positioning and hence can be interpreted as the overall performance bottleneck, without relating it to a specific communication signal structure. Note that, limitations and challenges become evident when the performance limits and optimal delay estimation results of the system proposal in [Sch12] are reviewed. Algorithms that already earlier extensively had been shown to yield optimal results in simplistic low order multipath channels, like the two-path channel, were presented and investigated for the special IDM-based system structure. Especially, the results presented for the more realistic WINNER like channel models [KMH+08], indicate that not the parameter estimation algorithms pose the actual problems but the theoretically best achievable performance. As a lower variance bound for unbiased estimators the well-known Cramer-Rao lower bound embodies the best achievable performance (CRLB). In [Sch12] it can be seen that

in realistic scenarios this bound turns out to have unreasonably high values. Those values indicate that the assumption that only a single path is present would results in a lower estimation error. Hence, this bound indicates that the receiver-sided modelling is too complicated for the amount of information the receiver has to work with. Therefore, one of the goals in this thesis is to revise the receiver-sided modelling and to assess and propose less complicated estimation strategies. The dissertation [Sch12] elaborates on tuning the parameters like the oversampling factor in order to obtain lower bounds. Although the results show that this is beneficial they also show that this alone is not enough to obtain accurate parameter estimates in realistic channel conditions. This in return shows that both the choice of the specific parameter setup as well as the modelling for estimation should be revised. Only tuning the parameters responsible for the presented system, like e.g. the sampling duration T and the oversampling factor J, is clearly not enough. Consequently, this thesis addresses this problem and provides possible solutions to solve this problem. To this end I extend the parameter estimation problem to a joint model order and parameter estimation problem in this thesis.

Apart from the fact that a significant part of the project and findings presented in [Sch12] as well as in this thesis shows that estimating reliability information, more precisely meaning estimating how large the parameter estimation error is going to be, can be beneficial for the purpose of joint communication and positioning and hence it should and can be employed for different purposes.

2.6 Challenges, Design Issues and Open Questions

For JCAP systems the most important challenges are multipath propagation, NLOS conditions and limitations due do predefined systems and standards. Ground-based JCAP systems are especially promising, since, depending on the density of deployed base stations as potential ROs, they will be less prone to multipath, NLOS conditions and strong signal attenuation, as compared to their satellite-based competitor systems.
GNSS and ground-based positioning systems and system extensions, however, that employ the TOA or TDOA, have in common that both, multipath propagation and NLOS conditions, are still major challenges, although these problems are more pronounced for GNSS systems. Besides the advantages of 5G positioning systems listed earlier there are also challenges that have to be dealt with like a severe path loss and a resulting signal to noise ratio (SNR) loss and hence more sophisticated beamforming and knowledge about the physical channel due to the higher carrier frequencies [SGD+15]. For channel estimation based joint communication and parameter estimation to obtain and maintain a high resolution for the positioning relevant parameters in realistic scenarios is challenging.

2.6.1 Multipath Propagation

The GNSS system performance is degraded by multipath propagation. Under the assumption that the receiver algorithms and significant parameters like the sampling duration are similar to those in GNSS systems, ground-based systems principally experience the same problems as GNSS systems, only less severe. Furthermore, for ground-based sys-

tems like 5G it is already known that weather-specific conditions, like rain, will entail more multipath components. If high resolution parameter estimation is performed this will be challenging, since, obviously, a higher number of multipaths results in a higher number of estimates and therefore a higher complexity and higher estimation variance.

Generally, for channel-estimation-based JCAP accurate parameter estimation is required. Hence, for such JCAP systems multipath propagation is challenging, since it requires algorithms that are more complex than delay estimation via pure signal correlation. In multipath channels delay estimation via pure signal correlation is only useful for a coarse delay acquisition. Even if the receiver algorithms would be aware of the correct multipath propagation model and additionally they would principally be capable of estimating all the physical path parameters with a high resolution, the number of receiver measurements and the system design itself can yield a low positioning accuracy. Consequently, realistic channel scenarios with a huge number of multipath components require a suitable modelling with a flexible possibility to simplify the model at the receiver side. Especially the TOA CRLBs for a practical JCAP system proposal assessed in a multipath propagation scenario like in [Sch12] motivate both a revised overall optimal JCAP system proposal and parameter estimation strategies, which are tailored to work well for the instantaneous channel conditions, like for e.g. joint model selection and parameter estimation. Including multipath propagation into the estimation strategies comes with the price of a high computational cost and also often with a lack of robustness.

2.6.2 Non-line-of-sight (NLOS) Scenarios

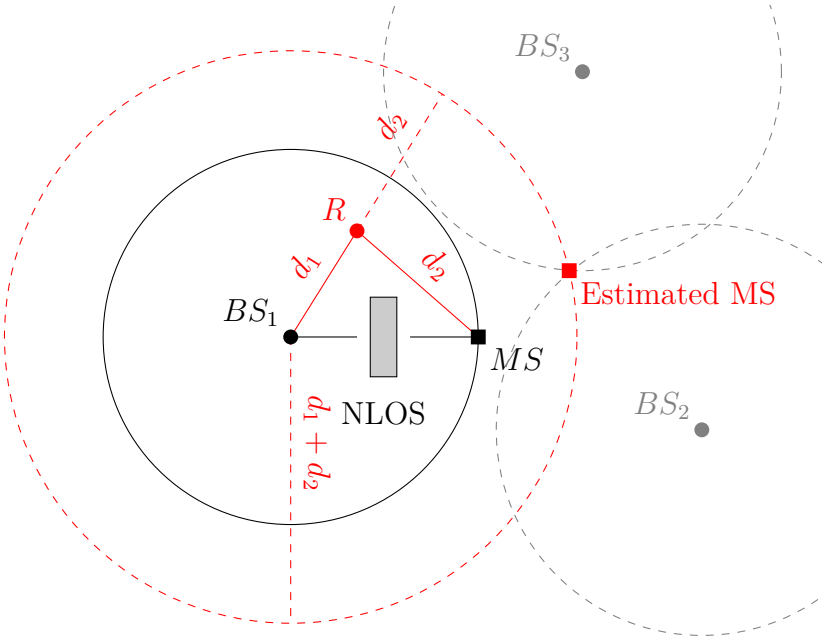


Figure 2.10: Example for NLOS as a TOA error source.

Furthermore, NLOS propagation scenarios constitute another TOA and AOA error source. Considering a TOA based system, assume that only a propagation delay larger

than the theoretical LOS delay (shortest path) is measured at a receiver. Without having additional a priori information about the physical environment the receiver wrongfully treats the measured distance as if it would be the targeted LOS distance as it is depicted in Figure 2.10. Delay-based positioning in NLOS conditions commonly leads to a positively biased NLOS estimation error, which is the distance between the first observed delay and the LOS delay. Different methods to mitigate the NLOS error are reviewed in [GC09]. Typical NLOS scenarios can be found indoors or in urban canyons, where the LOS signal is obstructed by large buildings with a high probability. At least indoors WLAN-based positioning via fingerprinting is a reasonable alternative solution to delay-based positioning. Despite the network densification envisaged for 5G, the authors of [ARS16] note that a LOS link cannot be guaranteed in dynamic outdoor environments and they therefore emphasize that it is important to explore possibilities of partially obstructed line-of-sight (LOS) and NLOS links. It is common practice to employ separate channel models as well as different estimation strategies for LOS and NLOS conditions. In this thesis we assume that a LOS component is available.

2.6.3 Why an Overall Optimal JCAP System Design from Scratch is Challenging

All the state-of-the-art JCAP systems that are used frequently are optimally designed for either communication or positioning. Positioning reference signals complement the LTE communication signal structure. GNSS signal structures on the other hand are optimized for positioning. Nevertheless, they can also be used to transmit communication data like the satellite position. Wi-Fi/WLAN signals can be employed for positioning indoors as well, however, the signal design was developed and optimized for the purpose of communication. Designing a JCAP system with an optimal unified signal structure for both communication and positioning from scratch is actually more a challenging goal than it is a completely realistic goal, due to different reasons: Apart from the limitations that will always go hand in hand with any standardization process, it would be difficult to optimally design a complete JCAP system, even without such limitations. An optimal system design for instance amongst other goals would aim at optimizing the theoretical lower performance limits like the well-known Cramer-Rao lower bound (CRLB) for the positioning relevant information. This bound is at least partially determined by the system design itself and hence only valid for this special chosen underlying setup and fixed parameters. So, to at least target an optimal system to some extent a fixed design is the prerequisite in order to define optimality. Unnecessary specifications, however, needlessly limit the performance. Consequently, in this thesis, I formulate a to a certain extent generic system framework.

Chapter 3

Mobile Radio Channel Modeling

The mobile radio channel model combines the modeling of various physical phenomena like path loss, shadowing, multipath propagation, and the Doppler effect. Therefore, to begin with, this chapter briefly reviews a few channel modeling basics. Path loss and shadowing belong to the so-called large-scale effects, describing the received signal strength variations over km ranges, while the so-called small-scale effects, like multipath propagation and the Doppler effect, cause received signal strength variations over sub-meter scales (not more than a few wavelengths). Depending on the signal parameters symbol duration and bandwidth, the small-scale effects lead to four sub-types of a phenomenon called fading, i.e., flat, frequency-selective, fast, and slow fading.

3.1 Mobile Radio Channel Modeling Basics

3.1.1 Large-Scale Effects

Large-scale channel models relate the received signal strengths to distances. Including the transmit and receiver side antenna gains G_{Tx} and G_{Rx}, a path loss model for the free-space scenario combines the distance d, the wavelength λ, the transmit power P_{Tx} and the received power P_{Rx} in the following manner, via the Friis transmission equation:

$$P_{Rx} = P_{Tx} G_{Tx} G_{Rx} \left(\frac{\lambda}{4\pi d} \right)^2 . \tag{3.1}$$

Therefore, path loss models can be used to estimate the received signal strengths roughly. This scenario is valid for GNSS. For terrestrial systems, on the other hand, these models fail, since small-scale effects have a more significant impact. Substituting the exponent 2 in (3.1) by the exponent p, the model for the path loss P_{Tx}/P_{Rx} can be generalized to other topologies. Further extending and generalizing this model by a term modeling shadowing effects (fluctuations), employing an (in the logarithmic domain) normally distributed random variable $X_\sigma \in \mathcal{N}(0, \sigma)$, with the distance-dependent variance σ further described in [Gol05, p. 51], then yields

$$\frac{P_{Tx}}{P_{Rx}}[\mathrm{dB}] = 10 \log \left(\frac{P_{Tx}}{P_{Rx}} \right) = -10 \log \left(G_{Tx} G_{Rx} \left(\frac{\lambda}{4\pi} \right)^p \right) + 10p \log(d) + X_\sigma . \tag{3.2}$$

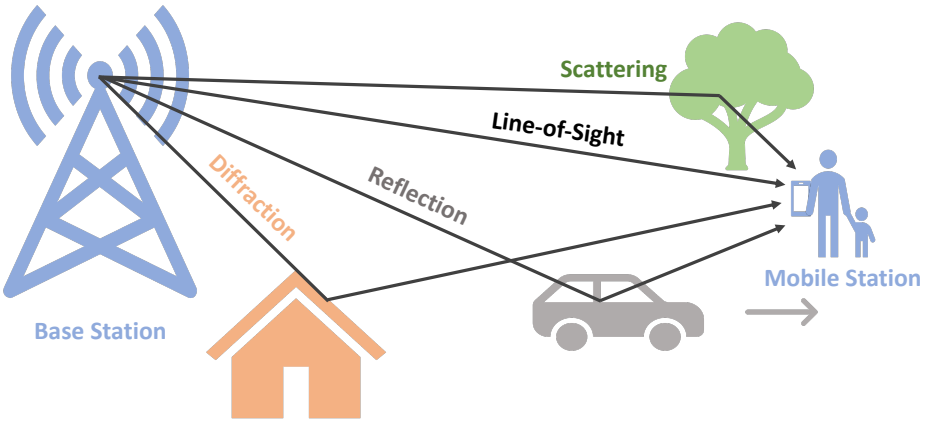


Figure 3.1: Multipath and line-of-sight signal propagation example: obstacles between transmitter and receiver reflect, diffract, and scatter.

3.2 Small-Scale Effects

3.2.1 Multipath

Signal reception via two or more paths is called multipath, caused by obstacles between transmitter and receiver. These obstacles lead to reflection, diffraction, and scattering (Figure 3.1). In a line-of-sight (LOS) multipath channel, besides the direct path, there is at least one other indirect path present. A non-line-of-sight (NLOS) channel refers to a channel, where the LOS path is blocked, and only multipath is received. We model multipath transmission via the channel weight function.

3.2.1.1 Channel Weight Function and Channel Impulse Response

The physical channel impulse response $c_0(t, t')$ expresses the time-dependent response versus time t to an excitation via a Dirac-Impulse for an excitation time t'.
For wireless channel modeling, it is a standard practice to eliminate the excitation time dependency and use the so-called resulting channel weight function instead of the channel impulse response [Höh13, p. 573].

The channel weight function $c(\tau, t)$ substitutes $\tau = t - t'$ in the channel impulse response and it is defined as

$$c(\tau, t) = c_0(t - \tau, t). \tag{3.3}$$

Considering that the channel consists of N_{MP} different transmission paths, the different path distances are related to different delays. Hence, the channel weight function $c(\tau, t)$

at time t consists of a sum of attenuated delayed Diracs with, in general, a time-varying delay $\tau_n(t) \in \mathbb{R}$ and complex amplitude $f_n(t) \in \mathbb{C}$.

$$c(\tau, t) = \sum_{n=0}^{N_{MP}} f_n(t)\delta(t - \tau_n(t)). \tag{3.4}$$

Stochastic channel modelling is needed to model the delays and complex path weights.

3.2.2 Stochastic Channel Modeling

For positioning a deterministic channel model that accurately describes the relationship to the device position is theoretically desirable. However, deterministic channel modeling is known to be unfeasible. Complex ray tracing methods are possible in the case of simplistic scenarios and geometries for channel modeling. For communication, it is common practice to statistical channel models instead. Via solving the inverse problem and applying nonlinear high-resolution parameter estimation techniques, the positioning-relevant multi-path parameters of a stochastic channel model can also be estimated. Hence, I only use stochastic channel models in this thesis. The statistical behavior, or channel characteristic, is based on random processes, commonly described via auto- and cross-correlation. The group of so-called wide-sense stationary uncorrelated scattering models covers the following usually applicable stochastic models.

3.2.2.1 Required Fundamentals

Since this thesis applies stochastic channel modeling, the following definitions are required.

The Doppler frequency shift f_d describes the frequency shift occurring due to receiver movement relative to the transmitter, such that

$$f_d = \frac{f_c v}{c_0} \cos(\phi - \theta_{vel}). \tag{3.5}$$

Thereby, f_c, c_0, v, ϕ and θ_{vel} denote the carrier frequency, the speed of light, the mobile device velocity, the angle of arrival and the travel direction.

The scatter function $S(\tau, f_d)$, also called delay Doppler power density spectrum, unifies the Doppler and the delay dependencies. It models the power density versus the delay domain as well as the power density versus the Doppler frequency.
Integrating over either the frequencies or the delays yields the power delay profile and the Doppler power spectrum, which are both used for stochastic channel modeling as applied in this thesis.

The power delay profile $S(\tau)$ is defined as

$$S(\tau) = \int_{-f_{d_{\max}}}^{f_{d_{\max}}} S(\tau, f_d)\, df_d. \tag{3.6}$$

The doppler power spectrum $S(f_d)$ is defined as

$$S(f_d) = \int_0^{f_{\tau\max}} S(\tau, f_d)\, d\tau, \tag{3.7}$$

where $\tau_{\max}$ denotes the maximum excess delay. The Doppler spread is the bandwidth for which $S(f_d) \neq 0$.

3.2.2.2 A Stochastic Channel Weight Function

Simultaneously arriving transmit signal replicas have a similar delay ($\tau_n \approx \tau_m$, $m \neq m$) if the delay difference is small compared to the symbol duration. Then these multipath components are classified as non-resolvable. They cause constructive and destructive interference.
If only non-resolvable components are present N_{MP} is replaced by N_{Rays}. If, on the contrary, resolvable path clusters ($\tau_c \neq \tau_m$, $c \neq m$), as well as non-resolvable paths, are present, the sum of N_{MP} components is replaced by a sum of C so-called multi-path clusters, comprising sum of N_{Rays} rays, in the following manner:

$$c(\tau, t) = \sum_{c=1}^{C} \underbrace{\sqrt{\frac{P_c}{\sum_{c=1}^{C} P_c} \cdot \frac{1}{N_{Rays}}} \sum_{n=0}^{N_{Rays}} f_{n,c}(t)}_{\beta_c} \delta(t - \tau_c(t)), \tag{3.8}$$

$$= \sum_{c=1}^{C} \beta_c \delta(t - \tau_c(t)). \tag{3.9}$$

Note that the powers P_c for all clusters $1 \leq c \leq C$ can be determined from the power delay profile $S(\tau)$ given by

$$S(\tau) = \sum_{c=1}^{C} P_c \delta(t - \tau_c(t)). \tag{3.10}$$

The complex path weights depend on the Doppler frequency

$$f_{n,c}(t) = e^{j(2\pi f_{D,n,c} t + \theta_{n,c})}, \tag{3.11}$$

where $\theta_{n,c}$ denotes the random starting phase drawn out of a uniform distribution in the range $[0, 2\pi]$, and $f_{D,n,c}$ denotes the Doppler frequency for cluster c and ray n.

The distinction between resolvable and non-resolvable paths is typically chosen in the literature on high-resolution parameter estimation since, for parameter estimation, it is only interesting whether a path is resolvable in a mathematical sense and not whether it physically constitutes of a cluster or a single ray.
Note that it is common practice to eliminate the LOS delay τ_1. For positioning, the LOS delay belongs to the essential parameters, and hence in this thesis τ_1 will not be eliminated.

3.2.2.3 Fading

Long-term versus short-term fading: Long-term fading models model the signal fades caused by topological variations like it is the case for shadowing. The additional term log-normal distributed term in (refeq:pathlossplusshadowing) modeling shadowing can be classified as long-term fading. Shadowing occurs due to obstacles attenuating the signal power through absorption, reflection, scattering, and diffraction between the transmitter and the receiver [Gol05, p. 27]. Typically, the variation occurs over distances that scaled by the sizes of the blocking obstacles. Short-term fading, on the other hand, models deep fades caused by constructive and destructive interference due to multipath propagation. They can be observed in a distance that maximally equals ten wavelengths [Höh13]. For example, Rayleigh and Rice fading are such short-term fading models and will be briefly introduced, since they are employed for this thesis.

Flat versus frequency-selective fading: Whether a channel can be classified as flat or frequency-selective requires us to compare the maximum channel excess delay and the symbol duration of the transmit symbol: A channel weight function can be classified as a flat fading channel if the maximum excess delay is negligible compared to the symbol period of the transmission signal. Otherwise, the channel is classified as frequency selective:

$$c(\tau, t) \text{ is } \begin{cases} \text{frequency-flat}, & \text{if } \tau_{\max} \ll T \\ \text{frequency-selective}, & \text{else} \end{cases}. \tag{3.12}$$

Equivalently, flat fading can also be characterized via the so-called coherence bandwidth $B_c = 1/\tau_{\max}$ and the signal bandwidth $B_s = 1/T$: The property $B_c \gg B_s$. Rayleigh and Rice channel models are frequency-flat channel models.

Note that for realistic terrestrial outdoor and indoor channel models, it is common practice to combine the concepts of frequency-flat fading and frequency-selective fading by employing the so-called clustered delay line model, introduced in (3.9). This model makes use of excess delay multipath clusters, for which the maximum excess delay for all clusters is $\tau_{\max} \geq T$. Hence, the model is frequency-selective since the channel is dispersive. Each cluster models Rayleigh or Rice fading like introduced in (3.13). Therefore, we gain each cluster component by modeling a sum of rays with approximately the same delay. Each cluster on its own is frequency-flat.

Rayleigh and Rice fading: Rayleigh fading is commonly used to model NLOS multipath propagation. A typical scenario is an indoor environment. On the other hand, Rice fading is used for modeling multipath propagation if an LOS component is present, which is typically the case for outdoor channels. Both channel models are closely related to each other, and they can be described via a single modeling approach, by introducing the so-called Rice factor K_R, allowing for varying between the two models smoothly. Let

θ_n be uniformly distributed in $[0, 2\pi)$. Then the model for the channel weight function is

$$c(t) = \underbrace{\sqrt{\frac{K_R}{K_R+1}}}_{\text{LOS component}} + \underbrace{\lim_{N_{Rays}\to\infty} \frac{1}{\sqrt{K_R+1}} \frac{1}{\sqrt{N_{Rays}}} \sum_{n=1}^{N_{Rays}} e^{j(\theta_n + 2\pi f_{d_n} t)}}_{\text{multipath components}}. \tag{3.13}$$

A Rice factor $K_R \neq 0$ states that a LOS component is present. In this case, a Rician channel model applies. A Rice factor of $K_R = 0$ means that an NLOS channel is modeled, and pure Rayleigh fading applies. Another noteworthy case is $K_R \to \infty$, which approaches the additive white Gaussian noise channel model [Höh13, p. 280]. The real and the imaginary part of $c(t)$ are Gaussian distributed random variables, where the mean of the real part is $\sqrt{K_R/(K_R+1)}$, and the imaginary part is zero-mean. In the Rayleigh model case, the amplitude of $c(t)$ is Rayleigh distributed, while in the Rice model case, it has a Rician distribution [Gol05, Höh13, PL05]. Rice fading and Rayleigh fading are flat fading models. They are often used to model path clusters. If more than one cluster is present, each cluster is frequency-flat, whereas the complete channel model, containing all clusters, is frequency-selective.

Fast versus slow fading: A channel varies only slowly over time (slow fading) if the maximum Doppler frequency $f_{d_{\max}}$ is significantly smaller than the signal bandwidth B_s: $f_{d_{\max}} \ll B_s$. Or, changing the perspective by taking the inverses, the so-called coherence time $T_d = 1/(2f_{d_{\max}})$ is required to be significantly larger than the symbol duration, such that

$$c(\tau, t) \text{ is } \begin{cases} \text{slow fading,} & \text{if } T_d \gg T_s \\ \text{fast fading,} & \text{else} \end{cases}. \tag{3.14}$$

3.2.2.4 Wide-Sense Stationary Uncorrelated Scattering

The Gaussian Wide-Sense Stationary Uncorrelated Scattering (WSSUS) model is a widely used model to short-term fading [Bel63, Höh13]. Consider a time-variant channel weight function and its spectrum via Fourier transform

$$c(\tau, t) \circ\!\!-\!\!- C(\tau, f_d). \tag{3.15}$$

According to the Wiener-Chintchin theorem, the associated autocorrelation function, defined as

$$r_{CC}(\tau, \tau', f_d, f_d') = \mathrm{E}\left\{F(\tau, f_d) F^*(\tau', f_d')\right\}, \tag{3.16}$$

is related to the power spectrum. A wide-sense stationary (WSS) channel has an autocorrelation function that only varies over a time difference δt. If the channel statistics fulfill the property to be wide-sense stationary (WSS) as well as fulfilling the property of uncorrelated scattering, the channel is called a WSSUS channel, and the following property holds:

$$r_{CC}(\tau, \tau', f_d, f_d') = \delta(f_d - f_d')\delta(\tau - \tau') S(\tau, f_d). \tag{3.17}$$

A common assumption [Höh13] for the GWSSUS channels is that the scatter funtion can completely be described via a factorization of the Doppler power density spectrum and the power delay profile such that

$$S(\tau, f_d) = S(\tau) \cdot S(f_d). \tag{3.18}$$

The Rayleigh and Rice channel model are only depending on $S(f_d)$, and hence are covered by this model, emphasizing the importance of this stochastic channel model category.

3.3 Channel Modeling for JCAP

Realistic channel models are commonly developed with the purpose to build and optimize wireless communication systems and not primarily to enable positioning via employing the same signal as used for communication. For instance, a typical assumption is that the sampling phase is zero due to perfect synchronization. For positioning, perfect synchronization would mean that τ_1 is already known. In this thesis, I target to estimate τ_1 by employing the channel estimates as observations and solving the inverse problem. Therefore, here $\tau_1 \neq 0$. For communication systems, it is sufficient to evaluate a single link. As explained in Chapter 2, TOA/TDOA-based positioning requires multiple reference objects. Therefore, for positioning, multiple links are necessary.

3.4 The WINNER Channel Model

For the sake of clarity, I will consider a 2-dimensional geometry. Let u, v stand for the uth and vth transmit and receive antenna, respectively. Furthermore, I define the so-called clustered delay line model according to the models provided in [KMH$^+$08]. That is, I assume that sums of multipath rays with similar excess delays build a so-called cluster like in (3.9). The distinction between rays and clusters will prove to be helpful in terms of resolvability, considering parameter estimation. Each ray is associated with index n, where a search cluster is associated with index c. Furthermore, I consider the angle of departure $\phi_{c,n}$, the angle of arrival $\varphi_{c,n}$, Doppler frequencies f_{D_n}, transmit and receive antenna spacings d_{Tx}, d_{Rx}, and starting phases $\theta_{c,n}$ uniformly distributed within $[-\pi, \pi]$. Further, assume that uniform linear arrays (ULA)s are applied. Figure 3.2 depicts the angular relationships, and Figure 3.3 shows the MIMO-specific geometry.

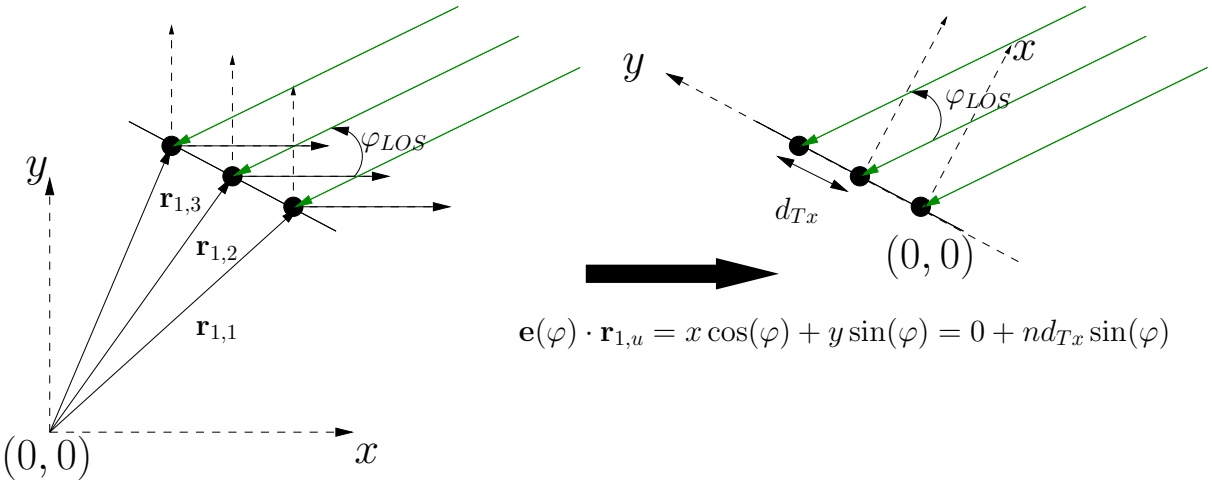


Figure 3.2: Relationship between the angle of incidence (departure or arrival) and location vector $r_{1,u}$ for antenna element u of a uniform linear array: If the coordinate system is chosen such that the x-axis equals antenna broadside, the channel angular dependencies can be described via the spacing of the antenna elements only.

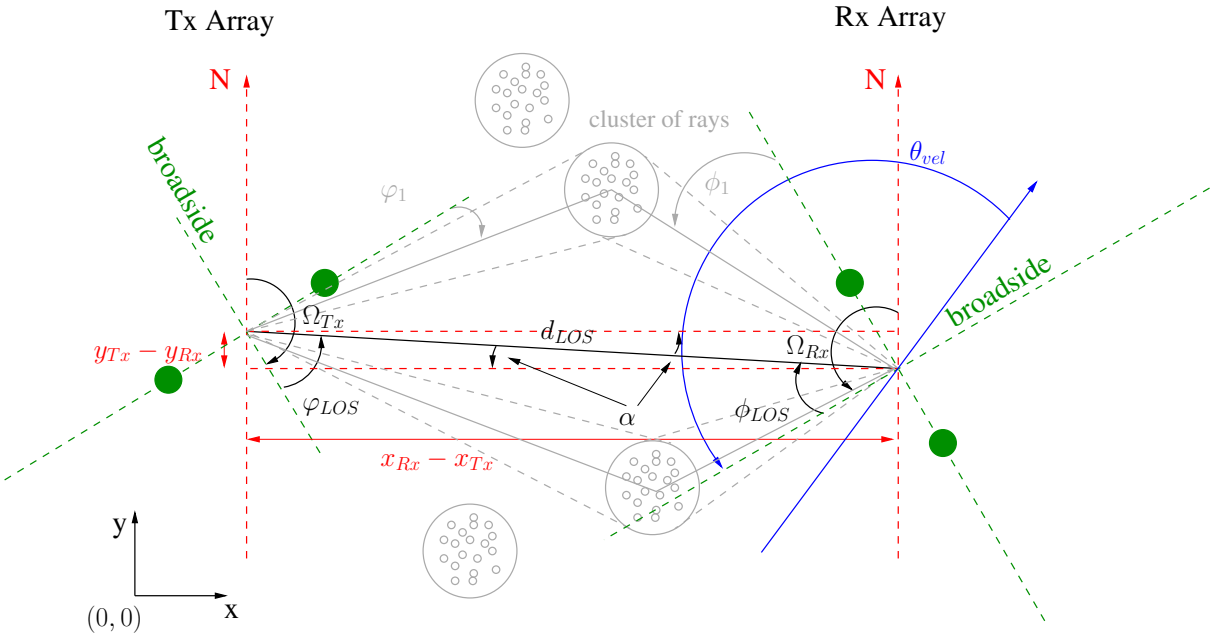


Figure 3.3: MIMO specific geometric dependencies in case two ULAs are employed: The LOS delay τ_{LOS}, angle of departure ϕ_{LOS} and angle of arrival φ_{LOS} are specified by the Cartesian coordinates of the ULA elements. It is assumed that the antenna geometry is small compared to the distance d_{LOS}. Hence, the same τ_{LOS}, ϕ_{LOS}, and φ_{LOS} are considered for each link between the antennas u, v.

Table 3.1: WINNER B1 LOS/NLOS Clustered delay line model.

		Cluster #	excess delay in [ns]	cluster power in [dB]	AOA in [°]
LOS WINNER-B1	$R = 3.3$, ASA 18°	1	0	0	0
		2 in [KMH+08]	30 35 40	-10.5 -12.7 -14.5	45
		2 here	30	-7.5	45
		3	55	-14.8	63
		4	60 65 70	-13.6 -15.8 -17.6	-69
		4	60	-10.6	-69
		5	105	-13.9	61
		6	115	-17.8	-69
		7	250	-19.6	-73
		8	460	-31.4	92

Table 3.2: Ray offset angles.

Ray number n	offset angles α_n
$1, 2$	± 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

Find the cluster's excess delays, powers, and arrival angles in Table 3.1. Then the WINNER model, as used in this thesis, can be emulated by the following steps:

1. Choose the scenario (B1,C1,...)

2. Assign reference, transmitter and receiver Cartesian coordinates

$$\mathbf{p}_0 = [0, 0], \quad \mathbf{p}_{Tx} = [x_{Tx}, y_{Tx}], \quad \mathbf{p}_{Rx} = [x_{Rx}, y_{Rx}].$$

3. Calculate distances d_{LOS} between transmitter and receiver, LOS angle of departure

and angle of arrival $\phi_{LOS}, \varphi_{LOS}$[1].

$$d_{LOS} = \sqrt{(x_{Tx} - x_{Rx})^2 + (y_{Tx} - y_{Rx})^2}, \tag{3.19}$$

$$\alpha = \arctan\left(\frac{y_{Tx} - y_{Rx}}{x_{Tx} - x_{Rx}}\right) = -\arctan\left(\frac{y_{Rx} - y_{Tx}}{x_{Tx} - x_{Rx}}\right) \tag{3.20}$$

$$\varphi_{LOS} = \begin{cases} -\arctan\left(\frac{y_{Rx} - y_{Tx}}{x_{Rx} - x_{Tx}}\right) + \frac{\pi}{2} - \Omega_{Tx}, & \text{if } x_{Rx} \geq x_{Tx} \\ -\arctan\left(\frac{y_{Rx} - y_{Tx}}{x_{Rx} - x_{Tx}}\right) - \frac{\pi}{2} - \Omega_{Tx}, & \text{else.} \end{cases} \tag{3.21}$$

$$\phi_{LOS} = \begin{cases} -\arctan\left(\frac{y_{Tx} - y_{Rx}}{x_{Tx} - x_{Rx}}\right) + \frac{\pi}{2} - \Omega_{Rx}, & \text{if } x_{Tx} \geq x_{Rx} \\ -\arctan\left(\frac{y_{Tx} - y_{Rx}}{x_{Tx} - x_{Rx}}\right) - \frac{\pi}{2} - \Omega_{Rx}, & \text{else} \end{cases}. \tag{3.22}$$

4. Assign transmitter and receiver antenna array orientations with respect to the assigned north direction.

5. Assign the speed and motion direction of the mobile station (receiver).

6. Assign the systems carrier frequency f_C.

7. Assign propagation condition (LOS/NLOS).

8. Assign Rice factor K_R and angular spreads $\Delta_\varphi, \Delta_\phi$.

9. Assign tabulated cluster powers in dB from clustered delay line model P'_c and and calculate and normalize cluster powers such that

$$P_c = \frac{10^{P'_c/10}}{\sum_{c=1}^{C} 10^{P'_c/10}}.$$

10. Assign tabulated AODs and AOAs φ_c, ϕ_c.

11. Calculate AODs and AOAs for each ray in each cluster via Table 3.2 and set

$$\varphi_{c,n} = \varphi_c + \Delta_\varphi \alpha_n.$$
$$\phi_{c,n} = \phi_c + \Delta_\phi \alpha_n.$$

12. Generate the fading components $f_{u,v,c}$ for each cluster c. Given the Doppler frequency

$$f_{d_{n,c}} = \frac{|v| \cos(\varphi_{n,c} - \theta_{vel})}{\lambda} \tag{3.23}$$

[1]Note, that the information associated with $d_{LOS}, \phi_{LOS,\varphi_{LOS}}$ is equivalent to specifying the exact locations of transmitter and receiver

and the two auxiliary arguments $\alpha_{n,c}, \alpha_{LOS}$

$$\alpha_{n,c} = \left(\theta_{c,n} + 2\pi f_{D_{n,c}} t + \frac{2\pi u d_{Tx}}{\lambda} \sin(\varphi_{c,n}) + \frac{2\pi v d_{Rx}}{\lambda} \sin(\phi_{c,n}) \right) \tag{3.24}$$

$$\alpha_{LOS} = \left(\theta_{LOS} + 2\pi f_{D_{LOS}} t + \frac{2\pi u d_{Tx}}{\lambda} \sin(\varphi_{LOS}) + \frac{2\pi v d_{Rx}}{\lambda} \sin(\phi_{LOS}) \right) \tag{3.25}$$

for each ray n and each cluster c the complex amplitude $\gamma_{u,v,c}$ is determined via

$$\gamma_{u,v,c} = \begin{cases} \frac{\sqrt{P_c}}{\sqrt{N_{Rays}+1}} \left(\sqrt{\frac{1}{K+1}} \sum_{n=1}^{N_{Rays}} \exp(j\alpha_{n,c}) + \sqrt{\frac{K}{K+1}} \exp(j\alpha_{LOS}) \right) & , \text{if } c = 1 \\ \frac{\sqrt{P_c}}{\sqrt{N_{Rays}+1}} \sum_{n=1}^{N_{Rays}} \exp(j\alpha_{n,c}) & \text{else} \end{cases} . \tag{3.26}$$

Thereby, polarisation is not considered in this thesis. Then the overall channel impulse response can be modelled by

$$c = \sum_{c=1}^{C} \gamma_{u,v,c} \delta(\tau - \tau_c). \tag{3.27}$$

13. Apply path loss and shadow fading. This step is omitted in this thesis, since the simulation results are shown for varying signal-to-noise ratios rather than varying distances. Path loss and shadow fading are proportional to a signal-to-noise ratio variation.

3.4.0.1 Delay and Path Amplitude Time Variability

Note that for positioning it is worthwhile to consider that the parameters' time variability is different for the complex path amplitudes $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_C]$ than for the actual parameters of interest, the delays collected in $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_C]$. It is essential to observe that $\boldsymbol{\tau}$ varies significantly slower than $\boldsymbol{\beta}$. More precisely, the variation $\Delta\tau_1$ in the TOA parameter τ_1 can be quantified by

$$\Delta\tau_1 = \frac{\Delta d}{c_0} = \frac{vIKT}{c_0}. \tag{3.28}$$

Here c_0 denotes the speed of light, v the relative mobile velocity, T denotes the sampling duration without oversampling, K a number of samples for which the complex path amplitudes $\boldsymbol{\beta}$ can be assumed quasi-invariant, and I shall be the number of blocks of length KT for which the delays can be considered quasi-invariant. This can easily be understood by defining an upper limit $\Delta\tau_{max}$ by inserting the maximum values $K_{max} = 1000$, $I_{max} = 100$, $T_{max} = 200\,\text{ns}$,$v_{max} = \frac{100\,\text{m}}{3.6\,\text{s}}$ as applied in (3.28) such that

$$\Delta\tau_{max} = \frac{\Delta d}{c_0} = \frac{v_{max} I_{max} K_{max} T_{max}}{c_0} \approx 2\,\text{ns} \quad (\widehat{=}\Delta d \approx 0.55\,\text{m}). \tag{3.29}$$

This means that a positioning error of 0.55 m for this setup would be unavoidable. If a higher precision is required, one of the parameters must be chosen even smaller.
Note that similarly to the block-fading assumption I am using, from now on, I will assume that $\boldsymbol{\tau}$ does not vary for a time series of I blocks.

Chapter 4

A Channel-Estimation-based JCAP System Framework

In this thesis the system framework includes the main functionalities like

- the transmission or system model,
- multipurpose channel estimation,
- data detection,
- parameter estimation,
- positioning,

including their structure and mutual relationships. Thereby, I target a formulation that can be directly applied to different multiplexing schemes or transmission concepts or needs only slight modifications. Regarding detection and estimation, I provide both, a time domain and frequency domain description system model, to later allow different estimation strategies, since the experience I gained shows that depending on the physical environment and design goal, different techniques can be applied.

4.1 System Model

The core system model consists of the composition of a so-called virtual training matrix $\mathbf{V}$, comprising data and training. It depends on the specific multiplexing and transmitter structure (coding, modulation, pulse shaping, and so forth ...). Furthermore, it consists of the physical channel and on the receiver modules: Channel estimation, data detection, parameter estimation, and positioning, as can be seen in Figure 4.1. Note that at this point it is not important whether OFDM, CDM, IDM or another multiplexing scheme is used. At this point, the idea is that via semi-blind channel estimation it is possible to employ both data and training, combined as virtual training for data detection as well as for high-resolution parameter estimation. Elements of these high-resolution parameter vector estimates are the TOA estimates, which, in return, are employed to enable high

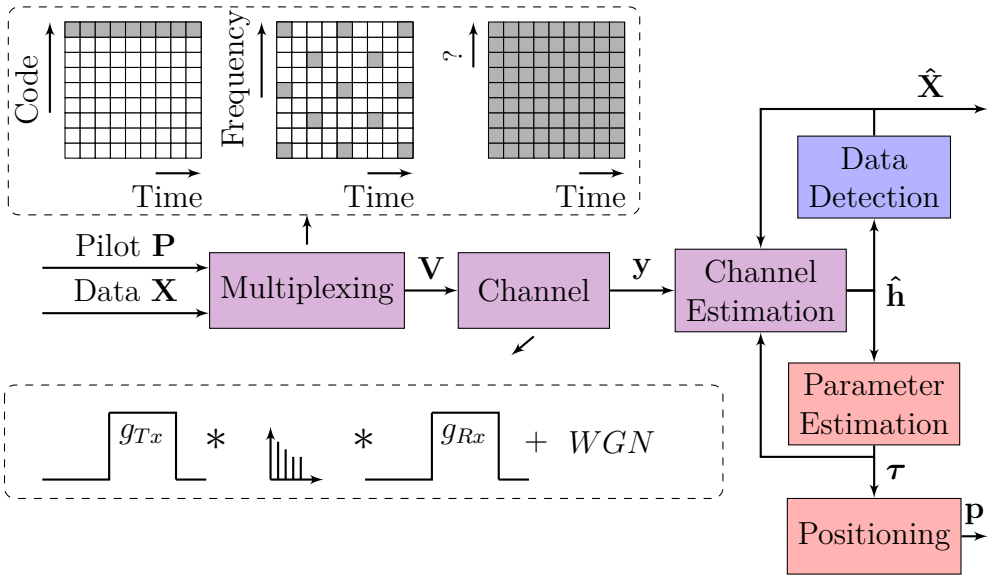


Figure 4.1: The system framework consists of a transmitter, a channel and a receiver concept. Thereby the transmitter arranges data and pilots according to the applied multiplexing and modulation scheme. The equivalent discrete time channel model comprises the physical channel as well as the pulse shaping and the receive filtering g_{Tx}, g_{Rx} and adds noise. Finally, the JCAP receiver consists of a core channel estimation module, feeding the data detection unit as well as the parameter estimation unit, which calculates positioning relevant data to be fed into the positioning algorithm.

precision positioning. Parameter estimates can be used to improve the channel estimation accuracy and hence the data detection performance. The output gained by data detection algorithms at the receiver when interpreted as virtual additional training is supposed to improve the parameter estimation accuracy in return. To obtain high precision positioning results, choosing optimal estimators for the preceding steps is an absolutely necessary prerequisite, other than choosing a specific transmission or estimation concept.

4.2 Overall Channel and Transmission Model

I will begin with the channel model formulation as it is needed. It can be employed by a parallel time-domain and frequency-domain description.

4.2.1 The Continuous Physical Channel Model

The continuous physical channel model can be described via the channel weight function $c(\tau)$, as previously introduced in Chapter 3 in (3.9) in the time domain, or as the channel transfer function in the frequency domain:

$$c(\tau) = \sum_{c=1}^{C} \beta_c \delta(\tau - \tau_c) \quad \circ\!\!-\!\!- \quad C(f) = \sum_{c=1}^{C} \beta_c e^{-j2\pi\tau_c f}. \tag{4.1}$$

Note that, for now, block-fading is assumed, meaning that the physical path parameters do not vary over the time t, that is the amplitudes and dealys time-dependency can be omitted and $c(\tau, t) = c(\tau)$.

4.2.2 Pulse Shaping and Receiver Filtering

Let the convolution of pulse shaping filter $g_{Tx}(\tau)$ and receive filter $g_{Rx}(\tau)$ be given by

$$g(\tau) = (g_{Tx} * g_{Rx})(\tau) \quad \circ\!\!-\!\!- \quad G(f) = g_{Tx}(f) g_{Rx}(f). \tag{4.2}$$

Given any delay shift τ_s, this leads to:

$$g_{\tau_s}(\tau) = g(\tau - \tau_s) \quad \circ\!\!-\!\!- \quad G_{\tau_s}(f) = G(f) e^{-j2\pi\tau_s f}, . \tag{4.3}$$

by the time-shift property of the Fourier transform.

4.2.3 The Continuous Overall Channel

The overall channel impulse response $h(t, \tau)$ is given by the convolution of the pulse shaping filter $g_{Tx}(\tau)$, the physical channel impulse response $c(\tau)$, and the receive filter $g_{Rx}(\tau)$. It can be expressed in terms of $g(\tau)$, since the convolution is commutative.

$$\begin{array}{ccccc} h(\tau) & = & (g_{Tx} * c * g_{Rx})(\tau) & = & (g * c)(\tau), \\ \circ\!\!\mid & & \circ\!\!\mid & & \circ\!\!\mid \\ H(f) & = & g_{Tx}(f) C(f) g_{Rx}(f) & = & G(f) C(f). \end{array} \tag{4.4}$$

Executing the convolution by using (4.1) leads to

$$h(\tau) = \sum_{c=1}^{C} \beta_c g(\tau - \tau_c) \quad \circ\!\!-\!\!- \quad H(f) = \sum_{c=1}^{C} G(f) \beta_c e^{-j2\pi\tau_c f}. \tag{4.5}$$

4.2.4 The Equivalent Discrete-Time Channel Model

The equivalent discrete-time channel model (EDTCM) (Figure 4.2) shows the discretization of the left hand side of (4.5), embodied in the channel coefficients $h_0, \ldots, h_L$. More precisely, the EDTCM in Figure 4.2 shows that the input-to-output behaviour for any transmit symbol $v[k]$ and any recived symbol $y[k]$ can be modeled conveniently via a discrete convolution with the channel vector, indicated by the delay elements z^{-1}. The sampling duration, in case of symbol-rate sampling, coincides with the symbol duration denoted by T. In the case of oversampling with oversampling factor J, the sampling

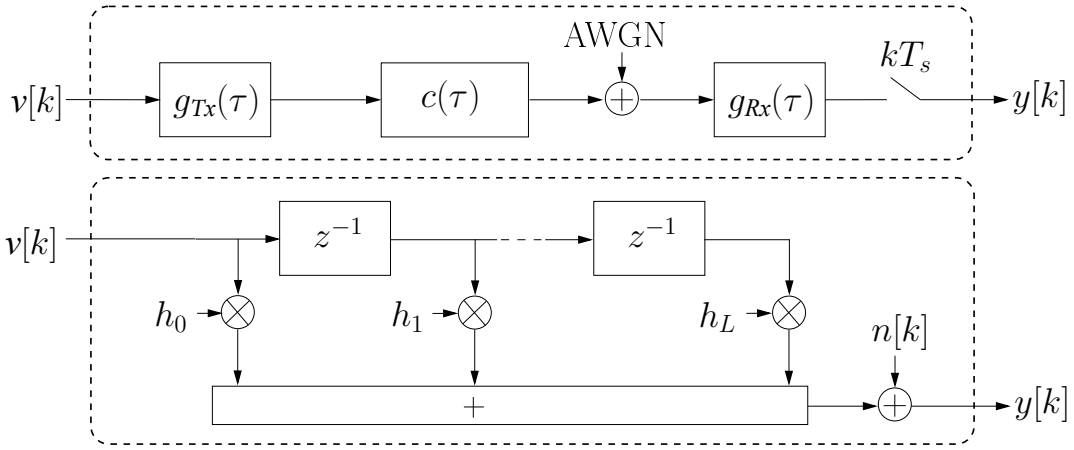


Figure 4.2: Equivalent discrete-time channel model (EDTCM) shown for $J = 1$.

duration is T/J. In this thesis, the function g, without loss of generality, is realized as a windowed delayed raised cosine with roll-off factor 0.3. Then, the window width $2W$ (W is an integer multiple of T) and the maximum channel delay τ_c specify the number of discrete samples in the channel vector, the so-called channel memory length L for symbol-rate sampling:

$$L = \left\lceil \frac{\tau_C + 2W}{T} \right\rceil. \tag{4.6}$$

Furthermore, let the the oversampled channel length be $L_J = JL + 1$. Sampling the left hand side of (4.2) at the time instants $\tau \in \{0T, T/J, 2T/J \cdots, LT\}$ and transforming via Discrete Fourier Transformation (DFT) provides the vectors

$$\begin{aligned} \mathbf{g} =& [g_0, \ldots, g_{L_J}] \quad \overset{}{\underset{L_J}{\longmapsto}} \quad \breve{\mathbf{g}} = \mathbf{F}_{L_J}\mathbf{g} = [\breve{g}_0, \ldots, \breve{g}_{L_J}], \quad \text{with} \\ g_l =& g\left(\frac{lT}{J}\right) \quad \forall l \in \{0, \ldots, L_J\} \quad \text{and} \quad \breve{g}_n = \sum_{l=0}^{L_J-1} g_n e^{\frac{-j2\pi nl}{L_J}} \quad \forall n \in \{0, \ldots, L_J\}, \end{aligned} \tag{4.7}$$

for the pulse samples without delay. Sampling the delayed pulses yields a pulse matrix with the elements given by $\forall l \in \{0, \ldots, L_J - 1\}$ $\forall c \in \{1, \ldots, C\}$:

$$[\mathbf{G}(\boldsymbol{\tau})]_{l,c} = g_{\tau_c}(l) = g\left(\frac{lT}{J} - \tau_c\right), \tag{4.8}$$

$$\begin{aligned} \mathbf{g}_{\tau_c} =& [g(0-\tau_c)], \ldots, g(LT-\tau_c)] \quad \overset{}{\underset{L_J}{\longmapsto}} \quad \breve{\mathbf{g}}_{\tau_c} = \mathbf{F}_{L_J}\mathbf{g}_{\tau_c} = [\breve{g}(0-\tau_c), \ldots, \breve{g}_{\tau_c}(LT-\tau_c)], \\ \mathbf{G}(\boldsymbol{\tau}) =& [\mathbf{g}_{\tau_1}^T, \ldots, \mathbf{g}_{\tau_C}^T] \quad \overset{}{\underset{L_J}{\longmapsto}} \quad \breve{\mathbf{G}} = \mathbf{F}_{L_J}\mathbf{G}(\boldsymbol{\tau}) = [\mathbf{F}_{L_J}\mathbf{g}_{\tau_1}^T, \ldots, \mathbf{F}_{L_J}\mathbf{g}_{\tau_C}^T]. \end{aligned}$$

Sampling at time instants $\{0T, T/J, \cdots, LT\}$ and transforming via DFT yields the vectors

$$\mathbf{h} = [h_0, \ldots, h_{L_J-1}] = \mathbf{G}(\boldsymbol{\tau})\boldsymbol{\beta} \quad \overset{}{\underset{L_J}{\vdash}} \quad \breve{\mathbf{h}} = [\breve{h}_0, \ldots, \breve{h}_{L_J-1}]^T = \mathbf{F}_{L_J}\mathbf{h} = \breve{\mathbf{G}}(\boldsymbol{\tau})\boldsymbol{\beta}, \quad \text{with}$$

$$h_l = \sum_{c=1}^{C} \beta_c g(lT - \tau_c) \quad \forall l \in \{0, \ldots, L_J\} \quad \text{and}$$

$$\breve{h}_n = \sum_{l=0}^{L_J-1} h_l e^{-\frac{j2\pi nl}{L_J}} = \sum_{l=0}^{L_J-1} \left(\sum_{c=1}^{C} \beta_c g(lT - \tau_c) \right) e^{\frac{-j2\pi nl}{L_J}} \quad \forall n \in \{0, \ldots, L_J - 1\}. \tag{4.9}$$

Sampling the physical channel weight function in the time domain doesn't make any sense. Only the transfer function can be sampled in the frequency domain. Later, we will see, that a separation of $\mathbf{g}$ and $\boldsymbol{\tau}$ is desirable for the purpose of parameter estimation. Obviously, sampling the transfer function yields a discrete expression, allowing to separate the dependencies of $\mathbf{g}$ and $\boldsymbol{\tau}$ in the frequency domain. Hence, in the following, I prepare an expression that allows to eliminate the channel vector's dependency on $\mathbf{g}$. Sampling the right hand sides of (4.1) and (4.5) at the frequencies $f_n = n/((L+1)T) \quad \forall n \in \{0, 1, \ldots, L_J\}$ yields the DFT approximations

$$\breve{c}_n \approx \sum_{c=1}^{C} \beta_c e^{-j2\pi\tau_c n/((L+1)T)}, \tag{4.10}$$

$$\breve{h}_n \approx \sum_{c=1}^{C} \breve{g}_n \beta_c e^{-j2\pi\tau_c n/((L+1)T)}. \tag{4.11}$$

Written in the more convenient matrix vector notation and defining the matrix $\mathbf{S}$ with the entries $[\mathbf{S}]_{n,c} = e^{-j2\pi\tau_c n/((L+1)T)}$ this results in

$$\breve{\mathbf{c}} \approx \mathbf{S}\boldsymbol{\beta} \quad \text{and} \quad \breve{\mathbf{h}} \approx \operatorname{diag}(\mathbf{F}_{L_J}\mathbf{g})\mathbf{S}\boldsymbol{\beta}. \tag{4.12}$$

Please note, that due to the fact that signal transmission and sampling are actually carried out in time domain, finite filter lengths lead to aliasing effects in the case of a discrete transformation to frequency domain. The resulting signals then are not bandlimited and the sampling theorem cannot be met exactly. Nonetheless, oversampling mitigates aliasing and can be helpful with respect to the sampling theorem. Therefore, if parameter estimation is carried out in the frequency-domain for a raised-cosine pulse, it will be required that $J > 1$. The definition of $\breve{h}_n$ in (4.11) has a greater theoretical than practical value. It provides a useful formulation in order to remove the pulse and to obtain a formulation that only depends on the delays. Furthermore, note that the definition of $\breve{c}_n$ has a practical value from an algorithmic point of view, since time-domain sampling with subsequent DFT would only lead to a reasonable approximation of the $\breve{c}_n$ given in (4.11), if $\frac{\tau_c}{T} \in \mathbb{Z} \quad \forall c \in \{1, \ldots, C\}$, which, in general, does not apply. Hence, sampling in the frequency domain and transforming back to the time domain can yield a better approximation.

Keeping in mind that (4.12) is only an approximation, it nevertheless allows us to formulate a time-domain representation of the channel coefficients with a separable dependency of $\boldsymbol{\tau}$ and the actual pulse $g(\tau)$

$$\breve{\mathbf{h}} \approx \operatorname{diag}(\mathbf{F}_{L_J}\mathbf{g})\mathbf{S}\boldsymbol{\beta} \quad \overset{}{\underset{L_J}{\dashv}} \quad \mathbf{h} \approx \mathbf{F}_{L_J}^{-1}\breve{\mathbf{h}} \approx \mathbf{F}_{L_J}^{-1}\operatorname{diag}(\mathbf{F}_{L_J}\mathbf{g})\mathbf{S}\boldsymbol{\beta}. \tag{4.13}$$

If it is desired to remove the dependency of the $g(\tau)$ completely, this is easier in the frequency domain than in the time domain, as can be seen later. Nonetheless, it is possible in the time domain as well, by defining an auxiliary matrix $\mathbf{F}_{L_J+1}^{-1}\mathrm{diag}(\mathbf{F}_{L_J+1}\mathbf{g})$ and multiplying the channel vector with it's pseudoinverse.

For now, let us assume no oversampling is applied ($J = 1$). For pure communication purposes employing the vector $\mathbf{h}$ of the equivalent discrete-time channel model (EDTCM) depicted in Figure 4.2 provides a sufficient tool for a linear system model. Transmitting symbols v_k for time instants $k \in \{0, \ldots, K-1\}$ over a tapped delay line channel model is performed by employing the discrete convolution, which in return is performed by utilizing a Toeplitz matrix $\mathbf{V}$ determined by the vector $\mathbf{v} = [v_0, \ldots, v_{K-1}]$

$$\mathbf{V} = \begin{pmatrix} v_L & v_{L-1} & \ldots & v_0 \\ v_{L+1} & v_L & \ldots & v_1 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & v_L \\ v_{K-L} & \vdots & \vdots & \vdots \\ v_{K-L+1} & v_{K-L} & \ldots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ v_{K-1} & v_{K-2} & \ldots & v_{K-L} \end{pmatrix}. \tag{4.14}$$

This leads to a sequence of received values stacked in the vector $\mathbf{y} = [y[0], \ldots, y[K-L]]$ such that

$$\mathbf{y} = \mathbf{VG}\boldsymbol{\beta} + \mathbf{n} = \mathbf{Vh} + \mathbf{n} \quad \underbrace{\vdash}_{K-L} \quad \breve{\mathbf{y}} = \mathbf{F}_{K-L}\mathbf{y}. \tag{4.15}$$

Thereby, $\mathbf{F}_{K-L}$ denotes the DFT matrix. Note that the matrix $\mathbf{V}$ is constructed in order to model only the transient behaviour. As visible from (4.1), 4.9), (4.5) and (4.11) the positioning relevant information, which actually corresponds to τ_1 is contained in the physical channel weight function $c(\tau)$ as a part of the vector $\boldsymbol{\tau}$.

Since the channel weight function is not observed or measured, we employ the discrete channel coefficients to determine the delay vector via parameter estimation.

4.2.5 Special Cases

Let us here treat two special cases, which allow algorithmic simplifications. The first special case arises, if the delays are integer multiples of the sampling duration, and the second one arises, if the virtual training matrix is a circulant matrix in the time domain (the convolution with the channel coefficients is circular) and processing is carried out in the frequency domain. Therefore, note that:

Remark 1. *For any square rank or period N circulant matrix $\mathbf{C}$ (defined in Appendix B), being a special kind of Toeplitz matrix, which is specified by its first column $\mathbf{c}$:*

$$\begin{aligned} \mathbf{C} &= \mathbf{F}_N^{-1}\mathrm{diag}(\mathbf{F}_N\mathbf{c})\mathbf{F}_N \\ \Leftrightarrow \quad \mathrm{diag}(\mathbf{F}_N\mathbf{c}) &= \mathbf{F}_N\mathbf{C}\mathbf{F}_N^{-1}. \end{aligned} \tag{4.16}$$

Furthermore, the definition of the circulant matrix in Appendix B.6 means that

$$\mathbf{C}\mathbf{h} = \mathbf{c} \circledast \mathbf{h} \tag{4.17}$$

Property 1 (DFT of a circulant matrix)**.** *The DFT of a circular convolution can be expressed via a pointwise multiplication in the following manner:*

$$\mathbf{F}_N(\mathbf{c} \circledast \mathbf{h}) = \mathbf{F}_N\mathbf{c} \odot \mathbf{F}_N\mathbf{h} = \breve{\mathbf{c}} \odot \breve{\mathbf{h}}. \tag{4.18}$$

4.2.5.1 Case 1: Delays are Integer Multiples of the Sampling Duration

If the delays are integer multiples of the sampling duration, firstly a straighforward implication is

$$\frac{\tau_c J}{T} \in \mathbb{Z} \quad \forall c \in \{1, \ldots, C\} \quad \Rightarrow \quad \mathbf{S} = \mathbf{F}_{L_J}. \tag{4.19}$$

Secondly, the pulse matrix $\mathbf{G}(\boldsymbol{\tau})$ becomes circulant and by using (4.18) and defining $\boldsymbol{\beta}_{zp}^T = [\boldsymbol{\beta}, 0, \ldots, 0]^T \in \mathbb{C}^{(L_J)\times 1}$ we can directly write

$$\mathbf{h} = \mathbf{G}(\boldsymbol{\tau})\boldsymbol{\beta} \quad \overset{}{\underset{L_J}{\longmapsto}} \quad \breve{\mathbf{h}} = \breve{\mathbf{g}} \odot \breve{\boldsymbol{\beta}}_{zp}. \tag{4.20}$$

4.2.5.2 Case 2: Virtual Training Matrix is Circulant

Let $\mathbf{h}_{zp} = [\mathbf{h}, 0, \ldots, 0]^T \in \mathbb{C}^{K\times 1}$ denote the zero-padded channel coefficent vector. Then the Property (4.18) also indicates that if the virtual training matrix $\mathbf{V}$ is designed to be a circular matrix by employing a cyclic prefix, as it is the case for OFDM signals, the following simplification is met:

$$\mathbf{y} = \mathbf{V}\mathbf{h} + \mathbf{n} \quad \overset{}{\underset{K}{\longmapsto}} \quad \breve{\mathbf{y}} = \breve{\mathbf{v}} \odot \breve{\mathbf{h}}_{zp} + \breve{\mathbf{n}}. \tag{4.21}$$

Note that the right hand side of (4.21) consists of a simple pointwise multiplication, which allows for simple detection and estimation techniques. Obviously, this strongly motivates techniques like OFDM. The OFDM signal is constructed in the frequency domain to begin with. In this case the length K will be determined by the numbers of subcarriers and the length of the cyclic prefix. It is implemented via fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT). If OFDM is chosen as a multiplexing scheme, we can additionally make use of the separability of the influence of $\mathbf{g}$ and $\boldsymbol{\tau}$, without any further effort.

4.3 System and Model Extensions

4.3.1 Oversampling

If oversampling is applied, a sparse length $K_J = JK$ sequence of virtual training symbols has to be defined:

$$\nu[k] = \begin{cases} v[k'] & \text{if } k/J \in \mathbb{Z} \\ 0 & \text{else} \end{cases}. \tag{4.22}$$

Up to this moment I formulated the system equations including the possibility to do oversampling ($J > 1$). Formulating both, discrete time samples and their DFTs, this was necessary. Preceding research in [Sch12] elaborately showed that oversampling is desirable for positioning purposes and rather undesirable for communication purposes. Apart from that, oversampling has the drawback to entail the necessity of using whitening filters, since oversampling leads to correlated noise samples. Furthermore, numerical results indicated that oversampling factors exceeding $J = 2$ do not suffice to yield high resolution TOA parameter estimates and do not measurably improve the performance for a reasonable SNR region and a realistic channel scenario. Consequently, assuming that oversampling is the only way to obtain the desired positioning accuracy, [Sch12] suggests to employ $J = 2$.

In order to mitigate aliasing effects transforming to the frequency domain, setting $J = 2$ suffices as well [vdVVP98]. The results also show that for communication purposes the performance degrades with increasing oversampling factor J. Employing K measurements provides a sufficient statistic for optimal detection. In [Sch12] it is suggested to always up- and downsample between communication and positioning relevant processing in order to use the benefits of oversampling for positioning and to circumvent the drawbacks for communication. This complicated measure obviously requires higher cost and energy. Therefore in this thesis other approaches to yield an increase in the number of available channel measurements than employing oversampling are preferred. The DFTs above and hence oversampling was taken into account, in order to provide a system description that allows frequency-domain parameter estimation. Now, note that the requirement ($J > 1$) for frequency domain parameter estimation solutions is a first indicator to prefer time-domain solutions over frequency-domain based solutions. From now on, if not explicitly stated otherwise, $J = 1$ and hence $L_J = L + 1$ and a time-domain solution I will present.

4.3.2 Multiple Input Multiple Output (MIMO) Systems

Apart from oversampling, a MIMO setup is an alternative approach to increase the number of available channel measurements. Given that we consider an $N_T \times N_R$ MIMO system, the system equation becomes

$$\mathbf{Y} = \mathbf{H}\mathbf{V}^{\mathbf{T}} + \mathbf{N}, \tag{4.23}$$

where $\mathbf{Y} \in \mathbb{C}^{N_R \times (K-L)}$ denotes the channel output matrix, $\mathbf{H} \in \mathbb{C}^{N_R \times N_T(L+1)}$ denotes the channel matrix itself, $\mathbf{V}^T \in \mathbb{C}^{N_T(L+1) \times (K-L)}$ denotes the virtual training matrix, and $\mathbf{N} \in \mathbb{C}^{N_R \times (K-L)}$ denotes the noise matrix. The matrices are constructed in the following manner:

$\forall k \in \{0, \ldots, K-L\}, \forall l \in \{0, \ldots, L\}, \forall u \in \{0, \ldots, N_T - 1\}, \forall v \in \{0, \ldots, N_R - 1\}$:

$$[\mathbf{Y}]_{v,k} = y_v[k+L] \tag{4.24}$$

$$[\mathbf{N}]_{v,k} = n_v[k+L] \tag{4.25}$$

$$[\mathbf{V}_u]_{k,l} = v_u[k-l+L], \qquad \mathbf{V} = [\mathbf{V}_1 \ldots \mathbf{V}_{\mathbf{N_T}}], \tag{4.26}$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{1,1}^T & \ldots & \mathbf{h}_{1,N_T}^T \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{N_R,1}^T & \ldots & \mathbf{h}_{N_R,N_T}^T. \end{bmatrix}, \qquad \mathbf{h}_{u,v} = [h_{u,v,0}, \ldots, h_{u,v,L}]^T. \tag{4.27}$$

Note that the matrices are constructed in order to model the transient behaviour only. Let $\phi_{c,n} := [\boldsymbol{\phi}]_{c,n}$ denote the angle of arrival (AOA) and let $\varphi_{c,n} := [\boldsymbol{\varphi}]_{c,n}$ denote the angle of departure (AOD), associated with multipath cluster $c \in \{1, \ldots, C\}$ and ray $n \in \{1, \ldots, N_{Rays}\}$. The vectors

$$\boldsymbol{\phi}_c = [\phi_{c,1}, \ldots, \phi_{c,N_{Rays}}], \quad \boldsymbol{\varphi}_c = [\varphi_{c,1}, \ldots, \varphi_{c,N_{Rays}}], \quad \forall c \in \{1, \ldots c\} \tag{4.28}$$

build the rows of the matrices $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$, respectively. Let $\mathbf{a_T}$, $\mathbf{a_R}$ be the so-called steering vectors, sometimes also referred to as the array response vectors, belonging to the transmit and receive antenna array, respectively. They are defined for each impinging wave. Thus, let us define these vectors for each multipath cluster $c \in \{1, \ldots, C\}$ and ray $n \in \{1, \ldots, N_{Rays}\}$ such that

$$\mathbf{a_T}_{c,n}(\phi_{c,n}) = [1, \phi, \ldots, \phi^{N_T}]^T, \quad \text{with } \phi = e^{j2\pi \frac{d_T}{\lambda} \sin(\varphi_{n,c})} \quad \text{and } \mathbf{a}_{T,c}(\phi_{c,n}) \in \mathbb{C}^{N_T}, \tag{4.29}$$

$$\mathbf{a_R}_{c,n}(\varphi_{c,n}) = [1, \phi, \ldots, \phi^{N_r}]^T, \quad \text{with } \phi = e^{j2\pi \frac{d_T}{\lambda} \sin(\phi_{n,c})} \quad \text{and } \mathbf{a}_{R,c}(\varphi_{c,n}) \in \mathbb{C}^{N_R}. \tag{4.30}$$

The channel model can be adjusted by taking into account the cluster c specific transmitter and receiver array response matrices $\mathbf{A_R}_c(\boldsymbol{\phi}_c) \in \mathbb{C}^{N_R \times N_{Rays}}$ and $\mathbf{A_T}_c(\boldsymbol{\varphi}_c) \in \mathbb{C}^{N_T \times N_{Rays}}$. The transmit and receive antenna array spacings are modelled by d_T and d_R, respectively. Then the uth and vth transmit and receive antenna array response matrix entries are

$$[\mathbf{A_T}_c(\boldsymbol{\varphi}_c)]_{u,n} = e^{(j\frac{2\pi d_T}{\lambda}(u-1)\sin(\phi_{c,n})d_T)}, \tag{4.31}$$

$$[\mathbf{A_R}_c(\boldsymbol{\phi}_c)]_{v,n} = e^{(j2\pi\frac{d_R}{\lambda}(v-1)\sin(\varphi_{c,n})d_R)}, \tag{4.32}$$

Or the matrices can be described via the steering vectors as

$$\mathbf{A_{Tc}}(\boldsymbol{\varphi}_c) = [\mathbf{a_T}_{c,1}(\varphi_{c,1}), \ldots, \mathbf{a_T}_{c,N_{Rays}}(\varphi_{c,N_{Rays}})], \tag{4.33}$$

$$\mathbf{A_{Rc}}(\boldsymbol{\phi_c}) = [\mathbf{a_R}_{c,1}(\phi_{c,1}), \ldots, \mathbf{a_R}_{c,N_{Rays}}(\phi_{c,N_{Rays}})]. \tag{4.34}$$

$$\tag{4.35}$$

Introducing an auxiliary variable $f_{c,n}$ and revisiting and reformulating the summands of (3.26)

$$\exp\left(\theta_{c,n} + 2\pi f_{d_{c,n}} t + \frac{2\pi u d_{Tx}}{\lambda}\sin(\varphi_{c,n}) + \frac{2\pi v d_{Rx}}{\lambda}\sin(\phi_{c,n})\right) \tag{4.36}$$

$$= \underbrace{\exp\left(\theta_{c,n} + 2\pi f_{d_{n,c}} t\right)}_{f_{n,c}} \exp\left(\frac{2\pi u d_{Tx}}{\lambda}\sin(\varphi_{c,n})\right) \exp\left(\frac{2\pi v d_{Rx}}{\lambda}\sin(\phi_{c,n})\right), \tag{4.37}$$

it can be directly seen that for each cluster c the spatial signature matrix $\mathbf{\Gamma}_c \in \mathbb{C}^{N_R \times N_T}$ can be defined as

$$\begin{aligned}\mathbf{\Gamma}_c(t, \boldsymbol{\phi}_c, \boldsymbol{\varphi}_c) &= \sum_{n=1}^{N_{Rays}} f_{c,n}(t)\mathbf{a}_{\mathbf{R}c,n}(\phi_{c,n})\mathbf{a}_{\mathbf{T}c,n}^T(\varphi_{c,n}) && (4.38)\\ &= \mathbf{A}_{\mathbf{R}c}(\boldsymbol{\phi}_c)\mathrm{diag}(\mathbf{f}_c)\mathbf{A}_{\mathbf{T}c}^T(\boldsymbol{\varphi}_c), \quad \text{with } \mathbf{f}_c = [f_{c,1}, \ldots, f_{c,N_{Rays}}] && (4.39)\\ &= [\boldsymbol{\gamma}_{c,1}(\boldsymbol{\phi}_c, \boldsymbol{\varphi}_c), \ldots, \boldsymbol{\gamma}_{c,N_T}(\boldsymbol{\phi}_c, \boldsymbol{\varphi}_c)]. && (4.40)\end{aligned}$$

Here $\boldsymbol{\gamma}_{c,u}(t, \boldsymbol{\phi}_c, \boldsymbol{\varphi}_c) = [\gamma_1(t, \boldsymbol{\phi}_c, \boldsymbol{\varphi}_c), \ldots, \gamma_{N_r}(t, \boldsymbol{\phi}_c, \boldsymbol{\varphi}_c)]^T$ is the uth column vector of matrix $\mathbf{\Gamma}_c(t, \boldsymbol{\phi}_c, \boldsymbol{\varphi}_c)$.
Note that the elements $\gamma_{c,u,v}(t, \boldsymbol{\phi}_c, \boldsymbol{\varphi}_c) := \beta_{c,u,v}$ if $\boldsymbol{\beta}_{u,v} := [\beta_{1,u,v}, \ldots, \beta_{C,u,v}]^T$ is defined as the complex path amplitude vector for transmit antenna u and receive antenna v for all $u \in \{1, \ldots, N_t\}$ and $v \in \{1, \ldots, N_r\}$. In the MIMO case, the complex path amplitudes are renamed to $\boldsymbol{\gamma}$. Collecting and rearranging all cluster contributions in a matrix with transmit antenna index u such that $\mathbf{\Gamma}'_u(t, \boldsymbol{\phi}, \boldsymbol{\varphi}) \in \mathbb{C}^{N_R \times C}$:

$$\mathbf{\Gamma}'_u(t, \boldsymbol{\phi}, \boldsymbol{\varphi}) = [\boldsymbol{\gamma}_{1,u}(t, \boldsymbol{\phi}_1, \boldsymbol{\varphi}_1), \ldots, \boldsymbol{\gamma}_{C,u}(t, \boldsymbol{\phi}_C, \boldsymbol{\varphi}_C)]. \quad (4.41)$$

Further I store these matrices in an overall matrix $\mathbf{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi}) \in \mathbb{C}^{N_R \times CN_T}$, defined by

$$\begin{aligned}\mathbf{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi}) &= [\mathbf{\Gamma}'_1(t, \boldsymbol{\phi}, \boldsymbol{\varphi}), \ldots, \mathbf{\Gamma}'_{N_T}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})], && (4.42)\\ \boldsymbol{\gamma} &= \mathrm{vec}\{\mathbf{\Gamma}^T\}. && (4.43)\end{aligned}$$

This leads to the cluster parameter-dependent and ray parameter-independent simplified time-continuous channel matrices $\boldsymbol{\mathcal{H}}(\tau, \boldsymbol{\tau}, \mathbf{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})) \in \mathbb{C}^{N_R \times N_T}$, specified by

$$\boldsymbol{\mathcal{H}}(\tau, \boldsymbol{\tau}, \mathbf{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})) = \sum_{c=1}^{C} \mathbf{\Gamma}_c(t, \boldsymbol{\phi}_c, \boldsymbol{\varphi}_c) g(\tau - \tau_c). \quad (4.44)$$

Let $\otimes$ denote the Kronecker product. Sampling with a sampling duration T leads to the matrix $\mathbf{H}(\boldsymbol{\tau}, \mathbf{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})) \in \mathbb{C}^{N_R \times N_T(L+1)}$:

$$\mathbf{H}(\boldsymbol{\tau}, \mathbf{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})) = \sum_{c=1}^{C} \mathbf{\Gamma}_c(t, \boldsymbol{\phi}_c, \boldsymbol{\varphi}_c) \otimes \mathbf{g}_c^T(\tau_c), \quad (4.45)$$

where $\mathbf{g}_c^T = [g(0 - \tau_c), \ldots, g(LT - \tau_c)]^T$. Let us make use of the property that the Kronecker product of an identity matrix and any other matrix yields a block-diagonal matrix such that in this case we construct an auxiliary matrix $\boldsymbol{\mathcal{G}}_{MI}(\boldsymbol{\tau})$ by

$$\boldsymbol{\mathcal{G}}_{MI}(\boldsymbol{\tau}) = \mathbf{I}_{N_T} \otimes \mathbf{G}^T(\boldsymbol{\tau}) = \begin{pmatrix} \mathbf{G}^T(\boldsymbol{\tau}) & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G}^T(\boldsymbol{\tau}) & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{G}^T(\boldsymbol{\tau}) \end{pmatrix}. \quad (4.46)$$

Keeping in mind that later, for estimation purposes, solving the equations might be easier if $\mathbf{H}(\tau, \boldsymbol{\tau}, \boldsymbol{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi}))$ is represented via a simple matrix multiplication, we employ the structure of $\boldsymbol{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})$ and reformulate (4.45) by using (4.46)

$$\mathbf{H}(\boldsymbol{\tau}, \boldsymbol{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})) = \boldsymbol{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})\boldsymbol{\mathcal{G}}_{MI}(\boldsymbol{\tau}). \tag{4.47}$$

Alternatively, vectorize this matrix expression as

$$\mathbf{h}_{MIMO}(\boldsymbol{\tau}, \boldsymbol{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})) = \boldsymbol{\mathcal{G}}_{MIMO}(\boldsymbol{\tau})\boldsymbol{\gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi}), \tag{4.48}$$

by using standard rules[1], concerning a vectorized matrix multiplication, as follows:

$$\begin{aligned}
\mathbf{h}_{MIMO}(\boldsymbol{\tau}, \boldsymbol{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})) =& \mathrm{vec}(\mathbf{H}^T(\boldsymbol{\tau}, \boldsymbol{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi}))) && (4.49)\\
=& \mathrm{vec}(\boldsymbol{\mathcal{G}}_{MI}(\boldsymbol{\tau})\boldsymbol{\Gamma}^T(t, \boldsymbol{\phi}, \boldsymbol{\varphi})) && (4.50)\\
=& (\mathbf{I}_{N_R} \otimes \boldsymbol{\mathcal{G}}_{MI}(\boldsymbol{\tau}))\mathrm{vec}(\boldsymbol{\Gamma}^T(t, \boldsymbol{\phi}, \boldsymbol{\varphi}))\\
=& (\mathbf{I}_{N_R} \otimes (\mathbf{I}_{N_T} \otimes \mathbf{G}(\boldsymbol{\tau})))\underbrace{\mathrm{vec}(\boldsymbol{\Gamma}^T(t, \boldsymbol{\phi}, \boldsymbol{\varphi}))}_{:=\boldsymbol{\gamma}(\boldsymbol{\phi}, \boldsymbol{\varphi})}\\
=& (\underbrace{\mathbf{I}_{N_R N_T} \otimes \mathbf{G}(\boldsymbol{\tau})}_{:=\boldsymbol{\mathcal{G}}_{MIMO}(\boldsymbol{\tau})})\boldsymbol{\gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})\\
=& \boldsymbol{\mathcal{G}}_{MIMO}(\boldsymbol{\tau})\boldsymbol{\gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi}). && (4.51)
\end{aligned}$$

4.3.2.1 SIMO

One can consider the single-input multiple-output (SIMO) case as a special case of the MIMO scenario given that $N_T = 1$. Let the Kathri-Rao product of two arbitrary matrices $(\bullet)\diamond(\bullet)$ be defined as the column-wise Kronecker product of these two matrices.The pulse matrix simplifies to: $\boldsymbol{\mathcal{G}}_{MI} = \boldsymbol{G}$. Then the channel matrix is determined by

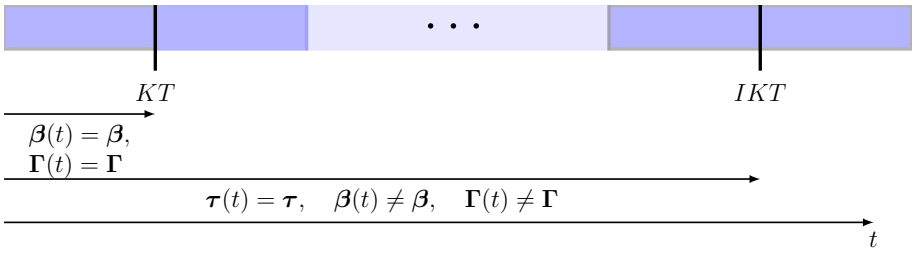
$$\mathbf{H}(\boldsymbol{\tau}, \boldsymbol{\Gamma}(\boldsymbol{\phi})) = \mathbf{A}_R(\mathbf{G} \diamond \boldsymbol{\beta}^T)^T = (\mathbf{A}_R \diamond \boldsymbol{\beta}^T)\mathbf{G}^T = \boldsymbol{\Gamma}\mathbf{G}^T(\boldsymbol{\tau}) \text{ or} \tag{4.52}$$

$$\mathbf{h}_{MO}(\boldsymbol{\tau}, \boldsymbol{\Gamma}(\boldsymbol{\phi})) = \boldsymbol{\mathcal{G}}_{MO}(\boldsymbol{\tau})\boldsymbol{\gamma}(\boldsymbol{\phi}). \tag{4.53}$$

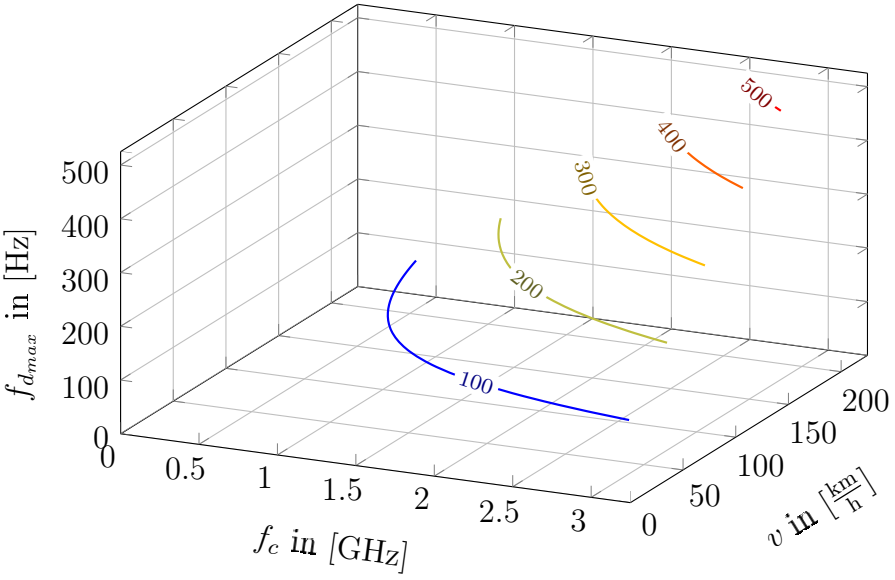
4.3.3 Time Series Measurements

Note that for the channel models applied in this thesis, $\boldsymbol{\tau}(t)$ varies slower in time than $\boldsymbol{\beta}(t)$ and $\boldsymbol{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})$, which measurably vary from block to block (block-fading assumption for the time span KT). Depending on the transmitter and receiver velocities, for the duration of a certain number of blocks, the variation of $\boldsymbol{\tau}(t)$ is negligible. We can use this property to our advantage later in the following manner: Modeling parameters in $\boldsymbol{\tau}(t) = \boldsymbol{\tau}$ as quasi-static for a time series of I bocks leads to a substantially increased number of channel measurements that can be further employed for model selection and parameter estimation. The performance of the employed model order selection as well as the TOA estimation crucially depends on the available number of channel estimations. The time variability of the parameters is shown in subfigure a) of Figure 4.3. The variability of

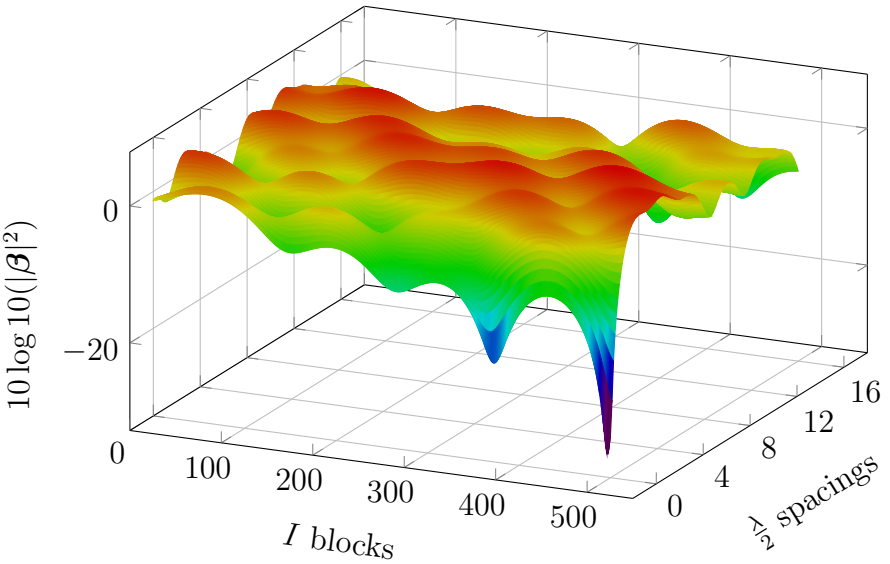
[1]Given any arbitrary matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ holds true.



(a) The physical path parameter time variability.



(b) The Doppler frequency depends on the carrier frequency and on the velocity.



(c) Channel power for a WINNER B1-LOS Rician fading channel with Rice factor $K = 3$, $v = 50\,\text{km/h}$ and $2\,\text{GHz}$ and $C = 1$.

Figure 4.3: Parameter time variability

the Doppler frequency is shown in subfigure b) and the complex path amplitude fading characteristic is shown in subfigure c). This variability indicates that it is not possible to collect an unlimited number of samples.
The continuous time-dependant channel matrix is

$$\boldsymbol{\mathcal{H}}(\boldsymbol{t}, \boldsymbol{\tau}, \boldsymbol{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi})) = \boldsymbol{\Gamma}(t, \boldsymbol{\phi}, \boldsymbol{\varphi}) \boldsymbol{\mathcal{G}}_{MI}(\boldsymbol{\tau}), \quad \in \mathbb{C}^{N_R \times N_T (L+1)}. \tag{4.54}$$

It is assumed that I blocks of estimates of $\boldsymbol{\mathcal{H}}[i]$ with $i \in \{0, \ldots, I-1\}$ are available. Therefore, the time duration between two consecutive snapshot measurements is KT and the measurement is carried out over a total time duration of IKT. We simply redefine or extend the definition of $\boldsymbol{\Gamma}$ and $\mathbf{H}$

$$\boldsymbol{\Gamma}(\boldsymbol{\phi}, \boldsymbol{\varphi})) = [\boldsymbol{\Gamma}^T(0, \boldsymbol{\phi}, \boldsymbol{\varphi}), \ldots, \boldsymbol{\Gamma}^T(I-1, \boldsymbol{\phi}, \boldsymbol{\varphi})]^T \qquad \in \mathbb{C}^{N_R I \times N_T C} \tag{4.55}$$

$$\mathbf{H}(\boldsymbol{\tau}, \boldsymbol{\Gamma}(\boldsymbol{\phi}, \boldsymbol{\varphi})) = \boldsymbol{\Gamma}(\boldsymbol{\phi}, \boldsymbol{\varphi}) \boldsymbol{\mathcal{G}}_{MI}(\boldsymbol{\tau}), \qquad \in \mathbb{C}^{N_R I \times N_T (L+1)}. \tag{4.56}$$

Similar as for the MIMO setup we can vectorize the matrix series to obtain a length $((L+1)N_R N_T I)$ complex channel vector

$$\boldsymbol{\gamma}(\boldsymbol{\phi}, \boldsymbol{\varphi})) = \text{vec}\{\boldsymbol{\Gamma}^T(\boldsymbol{\phi}, \boldsymbol{\varphi})\}, \tag{4.57}$$

$$\mathbf{h}(\boldsymbol{\tau}, \boldsymbol{\Gamma}(\boldsymbol{\phi}, \boldsymbol{\varphi})) = (\mathbf{I}_{N_R N_T I} \otimes \mathbf{G}(\boldsymbol{\tau}))\boldsymbol{\gamma} = \boldsymbol{\mathcal{G}}_{TS,MIMO}(\boldsymbol{\tau})\boldsymbol{\gamma}(\boldsymbol{\phi}, \boldsymbol{\varphi})). \tag{4.58}$$

Modelling the delays in $\boldsymbol{\tau}(t)$ as quasi-static for I symbols embodies a simplified approximated channel model. Therefore, we have to keep in mind that, in the worst case, while performing delay estimation, we actually limit the delay estimation accuracy to an accuracy of $(v_{\max}/c_0)IKT$. Another property to keep in mind regarding the time series extension is the dependency on the receiver mobility. Note that whenever the receiver is not moving (velocity $v_{\max} = 0$) or moving very slowly, the time series extension will lead to a series of constant or highly correlated measurements, which only differ in their instantaneous noise values, and therefore will not be helpful to improve the delay estimates.

4.3.4 Multiple Access

Different users $u \in \{1, \ldots, U\}$ are physically separated and hence user specific channels have to be taken into account. They are constructed via user specific complex path weights $\boldsymbol{\gamma} = [\boldsymbol{\gamma_1}, \ldots, \boldsymbol{\gamma_U}]$ and user specific delays $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_U$:

$$\mathbf{h} = \begin{bmatrix} \mathbf{I}_{N_R N_T I} \otimes \mathbf{G}(\boldsymbol{\tau}_1) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_R N_T I} \otimes \mathbf{G}(\boldsymbol{\tau}_U) \end{bmatrix} \boldsymbol{\gamma} = \boldsymbol{\mathcal{G}}_{MA,TS,MIMO}\boldsymbol{\gamma}. \tag{4.59}$$

Then the channel vector is $\mathbf{h} \in \mathbb{C}^{U \cdot I \cdot N_T \cdot N_R \cdot (L+1)}$.

4.4 Channel Estimation for JCAP

Channel estimation is inherently implemented in most available communication receiver structures, for the purpose of accurate and robust data detection. A broad range of

publications over the past two decades has been devoted to the subject, addressing different underlying communication signal architectures [vdBES+95, SH06, YLCC01, MKS04, MdCD02], time-variant versus time-invariant channel scenarios [SMH03, Lin08], different estimation strategies, including sparse and compressive techniques [BHSN10, HT08, CVM07, RS16], MIMO and training symbol design [NJU15, ZFY+16, BG06]. A relationship to positioning is established in [Mia07, AH13]. In [AH13] it is pointed out that especially for positioning using both training and data is desirable. Often, the underlying principle of linear least squares estimation based on or at least aided by training symbols is employed. Regardless, whether different multiplexing or modulation formats, or SISO, SIMO, MIMO, a time-series measurement or multiple acess systems are investigated: For the purpose of channel estimation, the underlying system equation (4.15) can often be formulated as a linear equation:

$$SISO: \quad \underbrace{\mathbf{y}}_{(K-L)\times 1} = \underbrace{\mathbf{V}}_{(K-L)\times(L+1)} \cdot \underbrace{\mathbf{h}}_{(L+1)\times 1} + \underbrace{\mathbf{n}}_{(K-L)\times 1}, \tag{4.60}$$

$$TS, MIMO, MA: \quad \underbrace{\mathbf{Y}}_{N_R\times I(K-L)} = \underbrace{\mathbf{H}}_{N_R\times UN_T(L+1)} \cdot \underbrace{\mathbf{V}^T}_{UN_T(L+1)\times I(K-L)} + \underbrace{\mathbf{N}}_{N_R\times I(K-L)}. \tag{4.61}$$

Stacking matrix columns in long vectors via vectorizations as performed similar to (4.48), (4.58) it can easily be shown that the first vector-matrix formulation suffices to model SISO, MIMO, TS and MA system equations:

$$\text{vec}\{\mathbf{Y}^T\} = \text{vec}\{\mathbf{VH}\} + \text{vec}\{\mathbf{N}^T\}, \tag{4.62}$$

$$\mathbf{y} = \underbrace{\mathbf{I}_{N_R} \otimes \mathbf{V}}_{\boldsymbol{\mathcal{V}}} \mathbf{h} + \mathbf{n}, \tag{4.63}$$

$$\mathbf{y} = \boldsymbol{\mathcal{V}}\mathbf{h} + \mathbf{n}. \tag{4.64}$$

Obviously, the representations (4.60) and (4.64) coincide. The only difference is the length and (4.60) is a special case of (4.64). Here, $\mathbf{V}$ is a matrix, which I dubbed virtual training matrix, since it either fully represents training symbols or can be decomposed into two parts, a typically and comparably small training part as well as part containing virtual training symbols, that is data symbols, which can be detected and then be used as training symbols as well. Since the exact configuration does not matter up to this point, let us assume that $\mathbf{V}$ consists of training only. The following subsection shortly introduces a linear least squares closed form solution for the channel estimates.

4.4.1 Linear Least Squares Channel Estimation

The classical and well known linear least squares (LLS) estimator here corresponds to the maximum likelihood (ML) approach and is given by multiplying the received vector with the pseudoinverse for the SISO representation in (4.60) [Kay10]:

$$\hat{\mathbf{h}} = \arg\min_{\tilde{\mathbf{h}}}\{(\mathbf{y} - \mathbf{V}\tilde{\mathbf{h}})^H(\mathbf{y} - \mathbf{V}\tilde{\mathbf{h}})\} = \underbrace{(\mathbf{V}^H\mathbf{V})^{-1}\mathbf{V}^H}_{\mathbf{V}^\dagger}\mathbf{y} = \mathbf{V}^\dagger\mathbf{y}. \tag{4.65}$$

The LLS estimator does not require any a priori information about the channel statistics. Knowledge about the physical path parameter behaviour is not something that can be assumed to be a priori available for JCAP. The goal here is to determine these parameters.

4.4.2 Weighted Linear Least Squares Channel Estimation

In case the signal is distorted by correlated noise, for instance in the case of oversampling, it is beneficial to introduce a weighting matrix $\mathbf{W}$ that weights the residual least squares error appropriately such that $\mathbf{W} = \mathbf{C}_n^{-1}$. The weighted linear least squares (WLLS) reads

$$\hat{\mathbf{h}} = (\mathbf{V}^H\mathbf{W}\mathbf{V})^{-1}\mathbf{V}^H\mathbf{W}\mathbf{y}. \tag{4.66}$$

It is straightforward to see that in the case of additive white Gaussian noise or in any case where $\mathbf{W} = \gamma\mathbf{I}$, where γ here uniquely stands for an arbitrary constant, the WLLS and the LLS estimator in (4.66) and in (4.65) are equivalent.

4.4.3 Linear Minimum Mean Squared Error Estimation

Another approach that is known to yield a better performance in terms of the mean squared error (MSE) is the so called linear minimum mean squared error (LMMSE) estimator. Since it employs additional a priori information about the channel statistics, it is here only suited to be applied after parameter estimation has been already performed, to optimize the detection and estimation performance. For the purpose of joint communication and positioning, the TOA, being part of the delay vector, is sought. Let $\mathbf{C}_h$ be the autocovariance matrix of the channel coefficients stored in $\mathbf{h}$. Further, let $\mathbf{C}_n$ be autocoveriance matrix of the noise, which is the noise covariance matrix related to the noise vector $\mathbf{n}$. Then the MMSE estimator [Kay10] is given by

$$\begin{aligned}\hat{\mathbf{h}}_{\text{MMSE}} =& \mathrm{E}\{\mathbf{h}\} + (\mathbf{V}^H\mathbf{C}_\mathbf{n}^{-1}\mathbf{V} + \mathbf{C}_\mathbf{h}^{-1})^{-1}\mathbf{V}^H\mathbf{C}_\mathbf{n}^{-1}(\mathbf{y} - \mathrm{E}\{\mathbf{h}\}) \\ =& \mathrm{E}\{\mathbf{h}\} + \mathbf{C}_\mathbf{h}\mathbf{V}^H(\mathbf{V}\mathbf{C}_\mathbf{h}\mathbf{V} + \mathbf{C}_\mathbf{n})^{-1}(\mathbf{y} - \mathrm{E}\{\mathbf{h}\}).\end{aligned} \tag{4.67}$$

Furthermore, the error covariance matrix is given by

$$\mathbf{C}_e = (\mathbf{V}^H\mathbf{C}_\mathbf{n}^{-1}\mathbf{V} + \mathbf{C}_\mathbf{h}^{-1})^{-1}. \tag{4.68}$$

Note that a LMMSE estimator (4.68) can be interpreted as a weighted least squares estimator, given that the a priori information becomes less accurate, such that $\mathbf{C}_h \approx \infty\mathbf{I}$. Given an infinite uncertainty for a zero-mean channel, it is easy to see that (4.68) and (4.66) are equal.

4.4.4 Covariance Matrix Estimation

A well known approach to gather information about the covariance matrix $\mathbf{C}_h$ is to estimate it. Channel covariance matrix estimates are later required for parameter estimation.

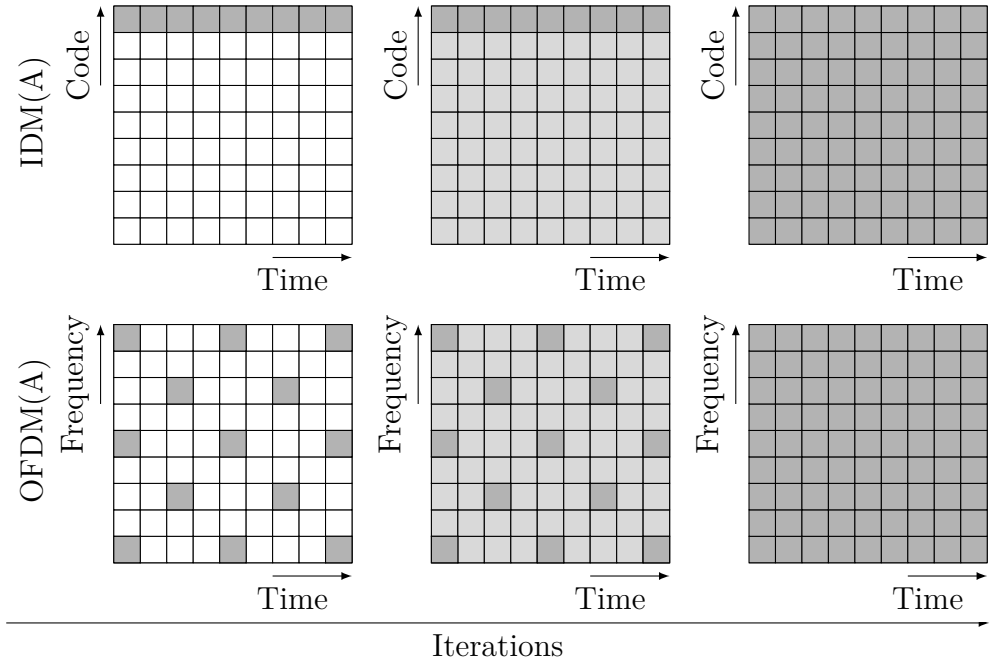


Figure 4.4: The idea is to virtualize the complete signal to training via an iterative semi-blind channel estimation approach. The different resource blocks show data (white) and training (gray). Especially JCAP profits from this approach.

This can be achieved by observing the LS channel estimates sequentially for I time instants and then calculating the sample covariance matrix by

$$\hat{\mathbf{C}}_{\hat{\mathbf{h}}} = \frac{1}{L+1} \sum_{i=1}^{I} (\hat{\mathbf{h}} - \mathrm{E}\{\hat{\mathbf{h}}\})(\hat{\mathbf{h}} - \mathrm{E}\{\hat{\mathbf{h}}\})^H. \tag{4.69}$$

4.4.5 Pulse Deconvolved Channel Estimates

The later presented frequency-domain high-resolution parameter estimation solutions require pulse deconvolved channel estimates. A deconvolution of the pulse, if desired, can easily be performed in the frequency domain. This particularly can be desirable, if frequency-domain parameter estimation strategies are applied in a further step. These pulse deconvolved channel estimates $\hat{\hat{\mathbf{h}}}_d$ can be obtained by

$$\hat{\hat{\mathbf{h}}}_d = (\mathrm{diag}(\mathbf{F}_L \mathbf{g}))^{-1} \mathbf{F}_L \mathbf{V}^\dagger \mathbf{y}. \tag{4.70}$$

4.4.6 Training-based, Blind and Semi-blind Channel Estimation

Channel estimation for JCAP as outlined in this thesis, has a central functionality serving both the communication side as well as the positioning side. Obviously, the TOA estimation accuracy directly translates into the positioning accuracy. The system proposal here assumes that parameter estimation is performed by treating the output of the

channel estimator as measurements, employed to formulate a set of equations to estimate the TOA. Hence, the TOA estimation error is strongly related to the channel estimation accuracy. Consequently, for this JCAP system framework, it is reasonable to perform accurate channel estimation.

Training symbols are symbols that are known a priori to the receiver. The more training is employed, the more accurate channel estimation can be. For the communication side it is desirable to yield a high data rate. Any form of overhead is undesirable. To this end, purely training-based channel estimators are designed to have a low ratio of training symbol energy versus overall symbol energy. This ratio is denoted by ρ. In communication systems setting $\rho = 1$ is no option due to this conflict.

A completely blind channel estimation approach assumes that no training symbols are available. Such methods are known to circumvent a training-based overhead on the one hand. On the other hand, they either require a large number of observations to make the problem identifiable, or they suffer from error propagation and hence are not reliable on the other hand.

Another approach yielding a tradeoff between training overhead, reliability and accuracy, is semi-blind channel estimation: Training symbols are complemented or aided by detected data symbols. Then as shown in Figure 4.4 with the iterations ρ iteratively approaches 1, meaning that the complete detected communication signal can be assimilated as training for accurate positioning [AH13, Mia07]. Different semi-blind channel estimation approaches for various signal models are proposed in [MM01, WH05, WSH06, MdCD02, SH06]. Here, the general virtual training matrix $\mathbf{V}_{\text{gen}} \in \mathbb{C}^{rows \times cols}$ is composed of a data part $\mathbf{D}_{\text{gen}}$ and a training part $\mathbf{T}_{\text{gen}}$. Then, the system equation has the form:

$$\mathbf{y}_{\text{gen}} = (\mathbf{D}_{\text{gen}} + \mathbf{T}_{\text{gen}})\mathbf{h} + \mathbf{w} = \mathbf{V}_{\text{gen}}\mathbf{h} + \mathbf{w}. \tag{4.71}$$

Note that this description applies for different multiplexing schemes as summarized in the following table:

	CDM/IDM/TDM	OFDM
$\mathbf{V}_{gen}$	(Toeplitz Matrix) $[\mathbf{V}_{\text{gen}}]_{k,l} = v[k-l+L]$ $\in \mathbb{C}^{K-L \times L+1}$	(stacked diagonal matrices times DFT matrix) $[\text{diag}\{\mathbf{v}\}]_{m,n} = \begin{cases} v_m & \text{if } m = n \\ 0 & \text{else} \end{cases}$ $\in \mathbb{C}^{N_s N \times N_s N}$ (N_s: subcarrier length, N symbol number)
$rows$	$K - L$	$N_s N$
$cols$	$L + 1$	$L + 1$
$\mathbf{y}_{\text{gen}}$	$\mathbf{y} = \mathbf{V}_{\text{gen}}\mathbf{h} + \mathbf{n}$ (CIR)	$\breve{\mathbf{y}} = \mathbf{V}_{\text{gen}}\mathbf{F}_{N_s}\mathbf{h} + \mathbf{n}$ (CFR)
$\hat{\mathbf{h}}$	$\mathbf{V}^{\dagger}\mathbf{y}$	$(\text{diag}\{\mathbf{v}\}\mathbf{F}_{N_s})^{\dagger}\breve{\mathbf{y}}$
$\hat{\hat{\mathbf{h}}}$	$(\mathbf{F}\mathbf{V})^{\dagger}\breve{\mathbf{y}}$	$\mathbf{V}_{\text{gen}}^{\dagger}\breve{\mathbf{y}} = (\text{diag}\{\mathbf{v}\})^{-1}\breve{\mathbf{y}}$ (pointwise division)

Obviously, for this generalised system equation the channel estimator is

$$\hat{\mathbf{h}} = \mathbf{V}_{\text{gen}}^{\dagger}\mathbf{y}_{\text{gen}}. \tag{4.72}$$

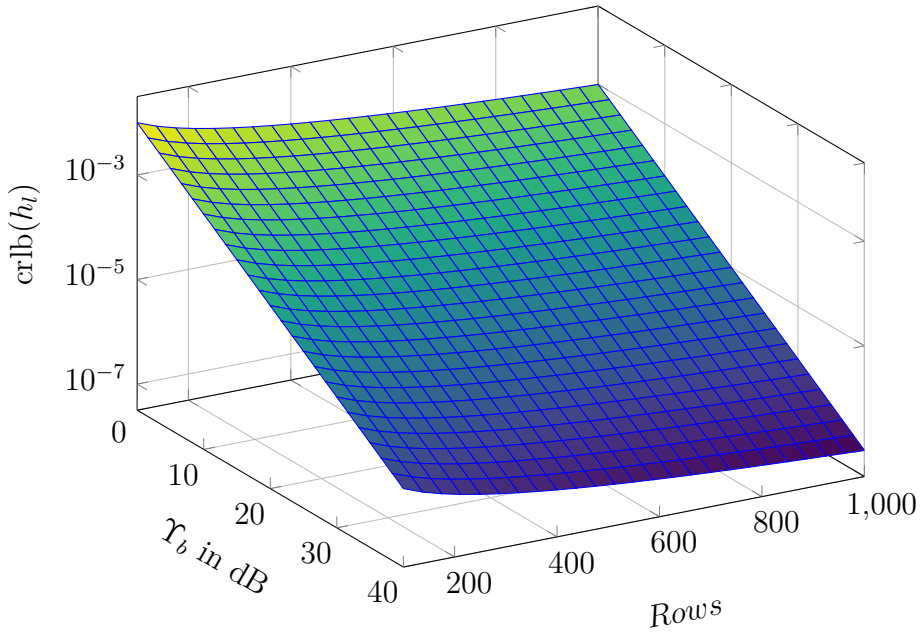


Figure 4.5: The CRLB(h_l) $\mathbf{V}$ only depends the noise variance (or signal-to-noise ratio Υ_b) and the number of rows of $\mathbf{V}_{\text{gen}}$ for $\rho \to 1$ and is the best achievable performance that can be fed to the parameter estimation unit.

Consequently, the CRLB for this estimation problem is

$$\text{CRLB}(\hat{\mathbf{h}}) = \sigma^2_{\mathbf{w}} \text{trace}\{(\mathbf{V}^H_{\text{gen}} \mathbf{V}_{\text{gen}})^{-1}\}. \tag{4.73}$$

For constant amplitude signals $\mathbf{V}^H\mathbf{V} \sim \mathbf{I}$ the CRLB can be simplified to only depending on the noise covariance $\mathbf{C}_{\mathbf{n}} = \sigma^2_{\mathbf{n}}\mathbf{I}$, $rows$ and $L+1$ for the overall vector

$$\text{CRLB}(\hat{\mathbf{h}}) = \frac{\sigma^2_{\mathbf{n}}(L+1)}{rows} \quad \text{and} \quad \text{CRLB}(h_l) = \frac{\sigma^2_{\mathbf{n}}}{rows}. \tag{4.74}$$

If we consider what the number $rows$ represents for the different multiplexing schemes, we see that it equals the number of time symbols that actually have to be transmitted. Therefore, note, that this implies the following: In case the virtual training matrix can somehow be used almost fully, ($\rho \to 1$) and it is chosen optimally with constant amplitudes, then the channel estimation performance will be similar for different mutliplexing schemes like IDM and OFDM (compare [MKS04],[vdBES+95] and [SH06]). As will be shown later, the delay estimation performance, and hence the positioning performance, depend on the channel estimation performance. All this indicates that the actual communication signal design is not as important for the positioning performance as choosing an optimal semi-blind estimation technique instead of emloying only training symbols.

4.4.7 MA, TS, MIMO Channel Estimation

In this thesis, MIMO channel estimation is desirable for two reasons: First of all, it is well known, that by employing multiple antennas the AOD and AOA can be estimated. This

is extra positioning information. Further, MIMO can be used to collect more measurements to yield a better delay estimation accuracy. In this thesis, we will employ SIMO measurements. It is straighforward to see that for a SIMO system equation the channel ML/LS channel estimator is

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{V}^H(\mathbf{V}\mathbf{V}^H)^{-1}. \tag{4.75}$$

Employing (4.64)

$$\hat{\mathbf{h}} = \boldsymbol{\mathcal{V}}^{\dagger}\mathbf{y}. \tag{4.76}$$

For the theoretical analysis presented in this thesis, lower bounds for the estimation results are an important tool to show the estimator performance and to detect room for possible improvements. Therefore, in the following sections, I assess the required error metrics, lower bounds and their derivations.

4.5 Theoretical Limits

4.5.1 Variance, Bias and MSE

For upcoming theoretical analysis I will use variance and MSE calculations as well as lower bounds. Therefore the following relationships are considered to be helpful. Let $\hat{\xi}$ be an arbitrary scalar estimate and $\hat{\boldsymbol{\xi}}$ be an arbitrary vector estimate. Then

$$\mathrm{MSE}(\hat{\xi}) = \mathrm{E}\{(\hat{\xi} - \xi)^2\} = (\mathrm{Bias}(\hat{\xi}))^2 + \mathrm{Var}\{\hat{\xi}\}.$$

In the vector case, this translates to

$$\mathrm{MSE}(\hat{\boldsymbol{\xi}}) = \mathrm{trace}(\mathbf{C}_{\hat{\boldsymbol{\xi}}}) + \mathrm{Bias}||\hat{\boldsymbol{\xi}}||^2.$$

In the scalar case, the variance of an arbitrary random variable ξ is determined by

$$\mathrm{Var}\{\xi\} = \mathrm{E}\{(\xi - \mu)^2\}, \quad \text{with} \quad \mu = \mathrm{E}\{\xi\}.$$

In the vector case, the covariance matrix, belonging to an arbitrary random column vector $\boldsymbol{\xi}$, is denoted by $\mathbf{C}_{\boldsymbol{\xi}}$ and calculated by

$$\mathbf{C}_{\boldsymbol{\xi}} = \mathrm{E}\{(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^H\}, \tag{4.77}$$

where $\boldsymbol{\mu}$ denotes the vector of component-wise expected values $\mathrm{E}\{\boldsymbol{\xi}\}$ for the vector $\boldsymbol{\xi}$. Thereby, the scalar and vector biases are defined as

$$\mathrm{Bias}(\hat{\xi}) = \mathrm{E}\{\hat{\xi}\} - \xi \quad \text{and} \quad \mathrm{Bias}(\hat{\boldsymbol{\xi}}) = \mathrm{E}\{\hat{\boldsymbol{\xi}}\} - \boldsymbol{\xi}.$$

Therefore, an unbiased estimator fulfills the property

$$\mathrm{E}\{\hat{\boldsymbol{\xi}}\} = \boldsymbol{\xi}.$$

Consequently, in the unbiased case, the MSE for each vector entry coincides with the covariance matrix, such that

$$\mathrm{MSE}(\hat{\boldsymbol{\xi}}) = \mathrm{trace}(\mathbf{C}_{\hat{\boldsymbol{\xi}}}).$$

4.5.2 Cramer-Rao Lower Bound (CRLB)

A suitable and well known lower performance bound for an arbitrary observation vector $\mathbf{s}$ and desired parameter vector $\boldsymbol{\xi}$ is the Cramer-Rao lower bound (CRLB), which, in general, is defined by

$$\operatorname{crlb}(\boldsymbol{\xi}) = \operatorname{diag}(\mathbf{F}(\boldsymbol{\xi})), \tag{4.78}$$

$$\mathbf{F}(\boldsymbol{\xi})^{-1} \leq \mathbf{C}_{\hat{\boldsymbol{\xi}}}, \quad \text{with} \tag{4.79}$$

$$[\mathbf{F}(\boldsymbol{\xi})]_{i,j} = -\operatorname{E}\left[\frac{\partial^2 \ln p(\mathbf{s};\boldsymbol{\xi})}{\partial \xi_i \partial \xi_j}\right], \tag{4.80}$$

where $\mathbf{F}$ is the Fisher information matrix (FIM). To find optimal parameters, two successive CRLBs have to be taken into account, namely the channel estimation CRLB and the delay estimation CRLB.

4.5.3 Linear Least Squares Estimation Covariance

Consider any linear system determined by an arbitrary matrix $\mathbf{X}$, a complex-valued vector $\boldsymbol{\xi}$, an additive white Gaussian noise vector $\mathbf{n} \in \mathcal{N}(0, \sigma_n^2)$, an observation vector $\mathbf{y}$, such that

$$\mathbf{y} = \mathbf{X}\boldsymbol{\xi} + \mathbf{n}. \tag{4.81}$$

Let the $\mathbf{X}$ be an $M \times N$-dimensional matrix and $\boldsymbol{\xi}$ an N-dimensional column vector. Then it is easy to show that the maximum likelihood estimate $\hat{\boldsymbol{\xi}}$ coincides with the linear least squares estimate

$$\hat{\boldsymbol{\xi}} = \mathbf{X}^{\dagger}\mathbf{y}. \tag{4.82}$$

Let $\boldsymbol{\mu}$ be the component-wise column vector for the expected values $\operatorname{E}\{\hat{\boldsymbol{\xi}}\}$. Then, the covariance matrix [Kay10] is given by

$$\mathbf{C}_{\hat{\boldsymbol{\xi}}} = \sigma_n^2 \left(\mathbf{X}^H \mathbf{X}\right)^{-1}. \tag{4.83}$$

4.5.4 Channel Estimation Cramer-Rao Lower Bound

Optimally, the whole energy of the virtual training matrix $\mathbf{V} \in \mathbb{C}^{(K-L)\times(L+1)}$ is available and $\left(\mathbf{V}^H\mathbf{V}\right)^{-1} \sim \frac{1}{K-L}\mathbf{I}_{(L+1)\times(L+1)}$ and (4.65) simplifies to

$$\hat{\mathbf{h}} = \mathbf{V}^{\dagger}\mathbf{y}. \tag{4.84}$$

Then, the lowest possible estimation error is bounded by the CRLB, such that

$$\operatorname{crlb}(\mathbf{h}) = \frac{\sigma_n^2}{K-L} \leq \operatorname{mse}_{h_l}. \tag{4.85}$$

In [AH13] it was shown that this bound is approached asymptotically with increasing signal-to-noise ratio (SNR) in a multipath scenario, which indicates, that ideally the

complete signal energy of the communication signal can be exploited for positioning as well. In the case of a purely training-based estimator, this bound is modified by dividing by the ratio of the training power to the overall transit power ρ. To preserve the quasi-time invariance during the observation, the length K should be limited such that $K \ll 1/(f_{d_{max}} T)$.

Chapter 5

Parameter Estimation

Parts of the fundamentals in this chapter have been published in [SBKH10, SAH12, SAH11b, SAH11a].

In this section, we assume that we have estimated the channel coefficient vector $\mathbf{h}$ as in (4.65). The noisy estimate consists of the channel vector $\mathbf{h}$ and an additive white Gaussian noise vector with statistical description $\mathbf{w} \sim \mathcal{N}(0, \sigma_w^2 \mathbf{I})$, such that it can be modelled as

$$\hat{\mathbf{h}} = \mathbf{h} + \mathbf{w}. \tag{5.1}$$

The channel estimates $\mathbf{h}$ potentially bears positioning relevant information like the TOA, which can be seen from revisiting the underlying channel model function for each element h_l of $\mathbf{h}$ in (4.9):

$$h_l = \sum_{c=1}^{C} \beta_c g(lT - \tau_c).$$

Since the channel estimates collected in $\hat{\mathbf{h}}$ like in (5.1) are inherently available in communication systems, they can additionally be exploited for parameter estimation, in particular, for TOA estimation, by solving the inverse problem. The accuracy of solutions to inverse problems depends on the amount of available information and on whether a priori information, like for instance statistical information such as the power delay profile, is known beforehand (or not), or whether the model function for the channel coefficients introduced in (4.9) is known (or not). To begin with, let us assume that the channel coefficient model function is known beforehand and the power delay profile isn't. This choice is reasonable, since it makes sense that the receiver knows the convolution of the pulse shape as well as the receiver filter $\mathbf{g}$ and does not know, which channel scenario momentarily applies. However, it is noteworthy that in case we assume that the model in (4.9) is known a priori, the correct model order C has to be known a priori as well. Otherwise, C could and possibly should be substituted by an model order estimate $\hat{C}$. Whether the assumption that the correct model order is known is the most reasonable choice to begin with, or whether it is not, is discussed in the later sections of this thesis. Moreover, Figure 5.1

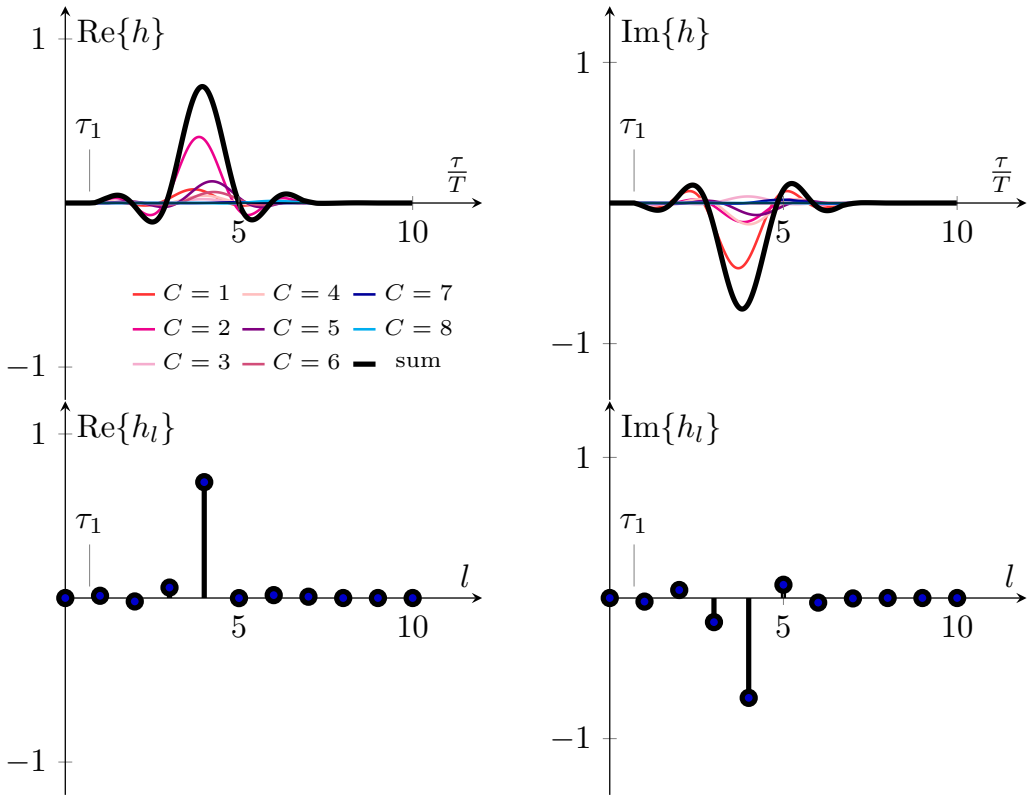


Figure 5.1: The example shows the real and imaginary parts of a typical signal superposition channel measurement snapshot: The upper pictures show the continuous channel impulse response, whereas the lower pictures show the samples $\mathbf{h} = [h_0, \ldots, h_L]$. The receiver "sees" $L+1$ estimated (noisy) samples (lower two plots: superimposed signals). The challenge is to estimate the positioning relevant information τ_1 from these noisy samples.

outlines a typically occurring channel scenario and emphasises three important difficulties assumed throughout this thesis:

- If not stated otherwise only a small number of snapshot measurements, that is channel coefficients, are available for parameter estimation.
- Typically, the excess delay to symbol duration ratio is assumed to be so small that classical correlation-based TOA estimation methods would fail.
- The roughly exponentially decaying power delay profile, varying instantaneous complex fading amplitudes combined with the difficulties mentioned above result in a high noise sensitivity.

Due to these harsh varying channel conditions, it becomes intuitively clear that the underlying estimation problem can range from being difficult to solve (considering the non-linear underlying structure and the high problem dimensionality) to being impossible to solve, without modifying the problem (considering that the problem may reveal singularities, or, at least, an ill-conditioned nature).

Note that, if a large amount of measurements is available (large sample case), the complex amplitudes β_c can be modelled by viewing them as either deterministic or stochastic variables. In the latter case, the sample covariance matrix is determined by

$$\mathrm{C}_{\boldsymbol{\beta}} = \lim_{n\to\infty} \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{\beta}(n)\boldsymbol{\beta}^H(n). \tag{5.2}$$

Two measurements taken at different times indices i and j are assumed to be independent, such that

$$\mathrm{C}_{\boldsymbol{\beta}}\delta_{i,j} = \mathrm{E}\{\boldsymbol{\beta}(i)\boldsymbol{\beta}^H(j)\}. \tag{5.3}$$

Thereby, $\delta_{i,j}$ denotes the Kronecker delta, being 1 for $i = j$ and 0 otherwise. The accurate way to model the components of $\boldsymbol{\beta}$ is to interpret them as deterministic variables. If not stated otherwise, I will proceed by interpreting $\boldsymbol{\beta}$ as a deterministic parameter vector. For now I assume that I only have one single estimated channel snapshot measurement to use for parameter estimation.

5.1 Maximum-Likelihood and Least Squares Estimation

Preferentially, we desire to estimate the parameters that can be used for the purpose of positioning. Here this is the TOA, in the single antenna scenario. In this thesis, the TOA is divided into a coarse part κ and a sampling phase τ_1

$$\tau_{\mathrm{toa}} = \kappa + \tau_1. \tag{5.4}$$

It is assumed that that κ is pre-estimated in a coarse delay estimation step during a synchronization phase. Therefore, the part, which we are actually interested in, is τ_1. However, the channel function depends on multiple different parameters. Therefore, they cannot be estimated optimally in terms of a one-dimensional parameter estimation problem. Additionally, to its dependence on the τ_1,the channel depends on the complex path amplitudes collected in $\boldsymbol{\beta} = \boldsymbol{\beta^i} + \boldsymbol{j\beta^q}$ and on the excess delays $\tau_2, \ldots, \tau_C$. Therefore, let us merge all of these inevitably desirable parameters in the overall vector

$$\boldsymbol{\theta} = [\boldsymbol{\beta^i}, \boldsymbol{\beta^q}, \boldsymbol{\tau}]. \tag{5.5}$$

Let the noise covariance matrix $\mathbf{C}_w$, belonging to the channel estimation error, be specified by

$$\mathbf{C_w} = \frac{\sigma_n^2}{K-L}\mathbf{I} = \sigma_w^2\mathbf{I}. \tag{5.6}$$

In the case that the noise samples in $\mathbf{w}$ are white and Gaussian distributed, it is known [Kay10] that the ML approach corresponds to the LS problem as is shown subsequently. The likelihood of a parameter vector $\boldsymbol{\theta}$ given the observation vector $\hat{\mathbf{h}}$ is defined by

$$\mathcal{L}(\boldsymbol{\theta}|\hat{\mathbf{h}}) = p(\hat{\mathbf{h}}|\boldsymbol{\theta}). \tag{5.7}$$

The maximum likelihood approach then aims at solving the problem

$$\hat{\boldsymbol{\theta}} = \arg\max_{\tilde{\boldsymbol{\theta}}}\{\mathcal{L}(\boldsymbol{\theta}|\hat{\mathbf{h}})\}. \tag{5.8}$$

Note that

- for any function f it holds true that $\arg\max\{f\} = -\arg\min\{f\}$ and most optimization algorithms minimize a cost function instead of maximizing it,
- exploiting the monotonicity of the logarithm is often known to lead to simpler optimization cost functions. Hence, instead of maximizing the likelihood directly, it is often preferable to maximize the logarithmic likelihood,
- constant factors and additive terms in optimization cost functions can often be omitted for the sake of simplicity, since omitting them leads to equivalent problems,
- it should be kept in mind that all optimization cost function changes, yielding equivalent optimization problems, do not necessarily maintain the exact equality to the likelihood function.

This yields the more comfortable equivalent problem

$$\hat{\boldsymbol{\theta}} = -\arg\min_{\tilde{\boldsymbol{\theta}}}\{\ln \mathcal{L}(\boldsymbol{\theta}|\hat{\mathbf{h}})\}. \tag{5.9}$$

Further, following the standard manipulations on optimization functions mentioned above the typical and well-known negative log-likelihood for this thesis can be formulated and simplified by

$$\begin{aligned}
-\ln(p(\hat{\mathbf{h}}|\boldsymbol{\theta}) &= -\ln\left(\frac{1}{|\pi^{L+1}\mathbf{C}_{\mathbf{w}}|}\exp\left((\hat{\mathbf{h}}-\tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}}))^H\mathbf{C}_{\mathbf{w}}^{-1}(\hat{\mathbf{h}}-\tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}}))\right)\right) \\
&= \underbrace{\ln\left(|\pi^{L+1}\mathbf{C}_{\mathbf{w}}|\right)}_{const} + (\hat{\mathbf{h}}(\tilde{\boldsymbol{\theta}})-\tilde{\mathbf{h}})^H\mathbf{C}_{\mathbf{w}}^{-1}(\hat{\mathbf{h}}-\tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}})) && (5.10)\\
&\overset{(5.6)}{=} \underbrace{\ln\left((\pi\sigma_w^2)^{L+1}\right)}_{const} + \frac{1}{\sigma_w^2}(\hat{\mathbf{h}}-\tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}}))^H(\hat{\mathbf{h}}-\tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}})) && (5.11)\\
&= const\cdot(\hat{\mathbf{h}}-\tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}}))^H(\hat{\mathbf{h}}-\tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}})) + const. && (5.12)
\end{aligned}$$

For the purpose of minimising the negative log-likelihood function, the constant terms in the cost function (5.12) can be omitted. For other purposes than optimization, like model selection for instance, it will become necessary to maintain the constant terms and hence to evaluate (5.11) instead of (5.12). Now, recall that the modelling function that builds the hypothesis vector $\tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}})$ is equal to

$$\tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}}) = \mathbf{G}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}}. \tag{5.13}$$

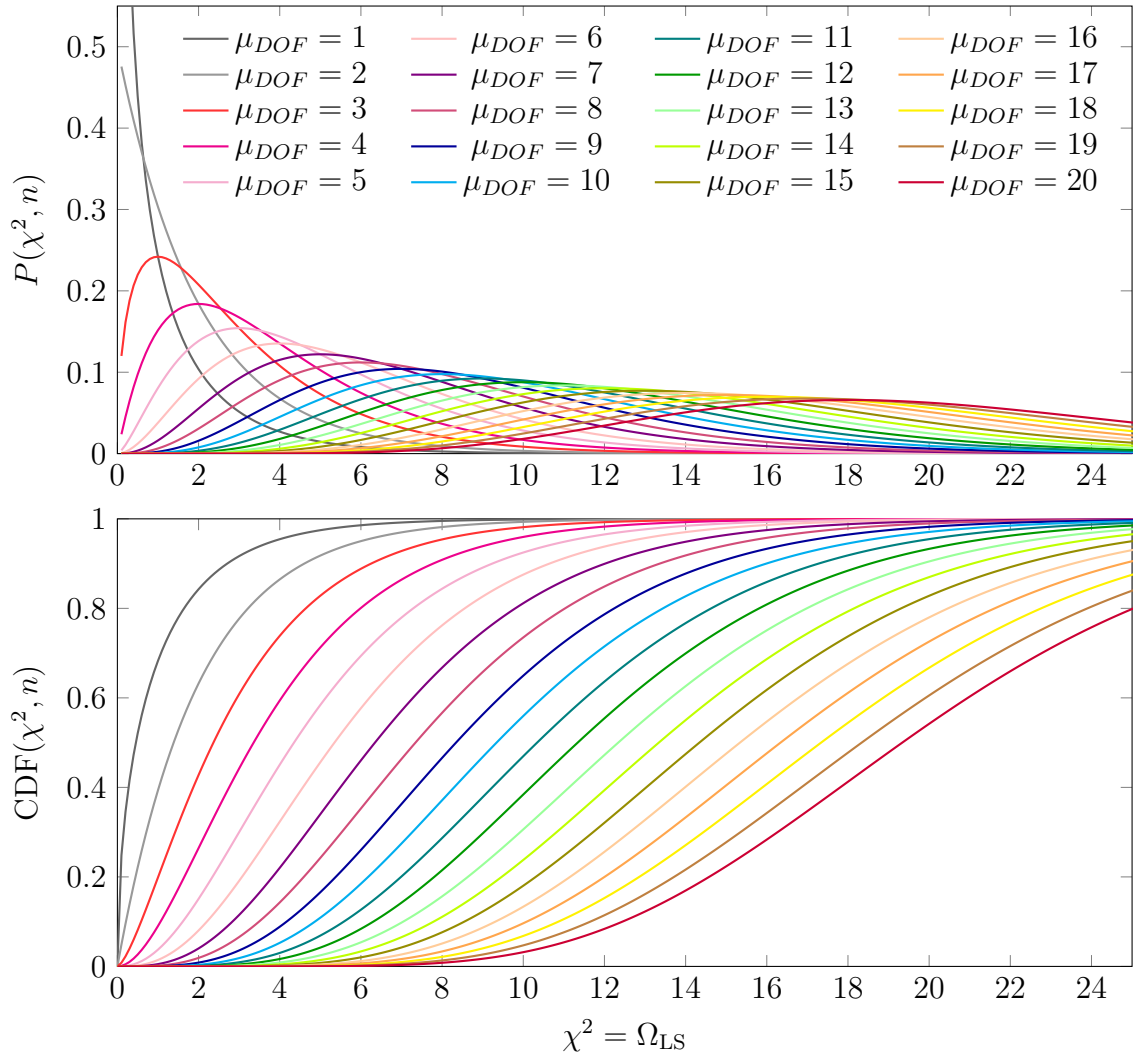


Figure 5.2: For additive white Gaussian noise the LS error's probability density function (PDF) and cumulative density function (CDF) are known functions (χ^2-distribution), uniquely specified by their degrees of freedom μ_{DOF}.

Finally, omitting the constant terms and inserting (5.12) into (5.9), the estimator may be formulated as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} \{(\hat{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}}))^H (\hat{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\boldsymbol{\theta}}))\} \tag{5.14}$$

$$= \arg\min_{\tilde{\boldsymbol{\theta}}} \{(\hat{\mathbf{h}} - \tilde{\mathbf{G}}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}})^H (\hat{\mathbf{h}} - \tilde{\mathbf{G}}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}})\}. \tag{5.15}$$

Note that in case of Gaussian distributed observations the least squares error

$$\Omega_{\mathrm{LS}} = (\hat{\mathbf{h}} - \tilde{\mathbf{G}}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}})^H (\hat{\mathbf{h}} - \tilde{\mathbf{G}}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}}) \tag{5.16}$$

is a χ^2-distributed random variable. If not specified otherwise, the channel observations $\hat{\mathbf{h}}$ are Gaussian distributed in this thesis. A χ^2-distribution can be described by the

probability density function (PDF) and cumulative density functions (CDF). The χ^2-distribution is a special case of the Γ-distribution, such that if a variable $x \sim \chi^2(k)$ then $x \sim \Gamma(\frac{k}{2}, \frac{1}{2})$. In general, the χ^2-probability density function is defined by

$$p_{\chi^2}(x, \mu_{DOF}) = \frac{x^{\frac{\mu_{DOF}}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{\mu_{DOF}}{2}} \Gamma(\frac{\mu_{DOF}}{2})}, \tag{5.17}$$

where Γ denotes the Γ-function, defined by

$$\Gamma\left(\frac{\mu_{DOF}}{2}\right) = \int_0^\infty t^{\frac{\mu_{DOF}}{2}-1} e^{-t} dt. \tag{5.18}$$

Here, μ_{DOF} denotes the number of degrees of freedom (DOF). Later, I will explain how to determine the number of degrees of freedom for the assessed LS problem. The χ^2-PDF, as well as the CDF, are shown in Figure 6.7. Figure 6.7 shows that, with increasing μ_{DOF}, the LS error distribution peak shifts more and more to the right and widens. It can be seen that for $\mu_{DOF} \to \infty$ the distribution can be approximated by a normal distribution. Furthermore, it becomes clear the least squares error exhibits known, varying PDFs and CDFs for a varying number of degrees of freedom. This special property will be exploited later for model selection and parameter estimation. Note, that later the residuals relationship to the χ^2-distribution can be exploited for different purposes. In Figure 5.8 it will be shown that experimentally the error distribution verifies the theoretically expected behaveiour.
Furthermore, note that this particular ML parameter estimation problem is in general without further specification a multi-dimensional non-linear LS problem with multiple local minima. Therefore, the optimization problem given in (5.15) is treated as a global optimization problem. If no initial guess is available, global optimization methods can be employed to solve this problem. In the case of an initial guess local optimization strategies can be considered as well. A short discussion regarding the cost function and algorithmic choice is given in Subsection 5.6.

5.2 Weighted Least Squares

The weighted least squares method proves beneficial when some entries of the observed data vector are more reliable than others. Then, a weighting matrix can be included to increase the impact of reliable observations and decrease the impact of the unreliable observations. Let a weighting matrix be defined by the matrix $\mathbf{W}$, a square positive definite matrix with a size determined by the size of the observation vector. Then, the weighted least squares problem can be formulated as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} \{(\hat{\mathbf{h}} - \tilde{\mathbf{G}}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}})^H \mathbf{W} (\hat{\mathbf{h}} - \tilde{\mathbf{G}}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}})\}. \tag{5.19}$$

For example it is straightforward to see that, if the covariance matrix for the error on the channel estimates is denoted by $\mathbf{C}_\mathbf{w}$, a reasonable choice for the weighting matrix is

$$\mathbf{W} = \mathbf{C}_\mathbf{w}^{-1}. \tag{5.20}$$

In this thesis, if not specified otherwise, the covariance matrix for the channel estimation error can be approximated by (5.6). In this case the weighted non-linear least squares estimator is the same as the unweighted estimator, since

$$\arg\min_{\tilde{\boldsymbol{\theta}}}\{\frac{1}{\sigma_w^2}(\hat{\mathbf{h}}-\tilde{\mathbf{G}}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}})^H\mathbf{I}(\hat{\mathbf{h}}-\tilde{\mathbf{G}}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}})\}=\arg\min_{\tilde{\boldsymbol{\theta}}}\{(\hat{\mathbf{h}}-\tilde{\mathbf{G}}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}})^H(\hat{\mathbf{h}}-\tilde{\mathbf{G}}(\tilde{\boldsymbol{\tau}})\tilde{\boldsymbol{\beta}})\}.$$

5.3 Separable Non-linear Least Squares

Two reasons can be named to reconsider and relax the problem given in (5.15). The first reason I mentioned earlier: Only the minor part of the estimate $\hat{\boldsymbol{\theta}}$ is actually desired in the JCAP sense. Thus it would be beneficial to be able to reduce the problem dimensionality. The second reason to relax the problem becomes obvious if we distinguish between the linear and the non-linear parameter dependencies, recalling the model function

$$\mathbf{h}=\mathbf{G}(\boldsymbol{\tau})\boldsymbol{\beta},\quad\text{with}\quad[\mathbf{G}(\boldsymbol{\tau})]_{l,c}=g(lT-\tau_c).$$

The model function depends on $\boldsymbol{\tau}$ via a nonlinear and on $\boldsymbol{\beta}$ via a linear relationship. Thus, it is well known [GP73, Kau75, RW80, OR13, Kay10] that such partially linear relationships can be exploited in the following manner: Momentarily ignoring that $\boldsymbol{\tau}$ is in fact an unknown, we are able to estimate the complex path amplitudes via the simple linear least squares estimator

$$\hat{\boldsymbol{\beta}}=\mathbf{G}(\tilde{\boldsymbol{\tau}})^{\dagger}\hat{\mathbf{h}}. \tag{5.21}$$

Thus, the complex path amplitude dependency in the optimization problem in (5.15) can be eliminated by substituting $\boldsymbol{\beta}$ with the solution (5.21), such that $\boldsymbol{\tau}$ instead of the complete $\boldsymbol{\theta}$ can be estimated by

$$\hat{\boldsymbol{\tau}}=\arg\min_{\tilde{\boldsymbol{\tau}}}\{(\hat{\mathbf{h}}-\mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger}\hat{\mathbf{h}})^H(\hat{\mathbf{h}}-\mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger}\hat{\mathbf{h}})\}. \tag{5.22}$$

Introducing the matrices $\mathbf{P}_{\mathbf{G}}=\mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger}$ and $\mathbf{P}_{\mathbf{G}}^{\perp}=\mathbf{I}-\mathbf{P}_{\mathbf{G}}$, a reduced dimensionality optimization problem, only depending on the delays in $\boldsymbol{\tau}$, can be formulated:

$$\begin{aligned}\hat{\boldsymbol{\tau}}&=\arg\min_{\tilde{\boldsymbol{\tau}}}\{-\hat{\mathbf{h}}^H\mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger}\hat{\mathbf{h}}\}\\&=\arg\min_{\tilde{\boldsymbol{\tau}}}\{-\hat{\mathbf{h}}^H\mathbf{P}_{\mathbf{G}}\hat{\mathbf{h}}\}\\&=\arg\min_{\tilde{\boldsymbol{\tau}}}\{\hat{\mathbf{h}}^H\mathbf{P}_{\mathbf{G}}^{\perp}\hat{\mathbf{h}}\}\\&=\arg\min_{\tilde{\boldsymbol{\tau}}}\{\Omega_{SNLLS}\}.\end{aligned} \tag{5.23}$$

These equalities can be understood by the stepwise estimator simplification as shown in the Appendix (A.1). Note that minimization of functions with the structure of (5.23) have been dubbed variable projection functional by Golub and Pareyra [GP73, GP03]. Minimizing Ω_{SNLLS} instead of Ω_{LS} is much simpler in terms of feasibility and computational complexity due to the dimensionality reduction, or as it is further put in [OR13]:

the beauty related to solving the separated problem instead of the full problem lies in the improved efficiency, due to less parameter dimensions and the higher probability of finding the global optimum over a local optimum, which can be understood by seeing that the cost function (5.23) has less local optima than (5.15) to begin with. Golub and Pareyra also claim in their topical review [GP03] that the separated problem is known to be less ill-conditioned than the full problem, and that it will converge faster. Nonetheless, assuming an unlimited source of computational capabilities, being equivalent, the two metrics will lead to the same performance regarding the parameter estimation MSEs. This has been seen by examining multiple simulation results carried out in the context of this thesis and has been also confirmed in [GP03]. Obviously, if $\boldsymbol{\tau}$ is estimated via (5.23), the complex path amplitudes finally can be estimated by (5.21) such that

$$\hat{\boldsymbol{\beta}} = \mathbf{G}(\hat{\boldsymbol{\tau}})^{\dagger}\hat{\mathbf{h}}. \tag{5.24}$$

5.4 Parameter Estimation-Aided Channel Estimation

If the complete vector $\boldsymbol{\theta}$ is estimated, enhanced channel estimates can be obtained via the matrix vector multiplication

$$\hat{\mathbf{h}} = \mathbf{G}(\hat{\boldsymbol{\tau}})\hat{\boldsymbol{\beta}}. \tag{5.25}$$

5.5 Detecting Estimation Success

In order to further utilize the gained parameter estimates, for instance for positioning or for improving channel estimation, it is crucial to know whether the parameter estimation has been successful or not. Here, successful means that estimated parameters should be close to the actual parameters, which, unfortunately, are not known at the receiver. More specifically, the solution should be at least equal to, or should outperform the purely correlation-based solution. Taking into account that the noisy estimation problem at hand yields a global, non-linear and multi-dimensional optimization problem, depending on the algorithmical choice it will be more or less difficult to guarantee success. Although the distance between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ is not known at the receiver side, there are indicators, which can be exploited to tell, whether the parameter estimation has been successful or not. Considering that we know that the least squares error can be modeled as the known χ^2-distribution, we can easily define a rule, which allows us to identify those parameters, which are and which aren't associated with a least squares error that, with a certain confidence, is covered by the known χ^2-distribution. This is emphasized by the results shown in Figure 5.8.

Mathematically speaking, this can be described as follows: Let Ω be the least squares error associated with the parameter estimates $\hat{\boldsymbol{\theta}}$. Further, let α specify the probability close to 1 (here I choose $\alpha = 0.999$) for which the least squares error is χ^2-distributed, and let ϵ_{χ^2} denote the associated least squares threshold value found by either table lookup [PTVF07] or calculation of the inverse CDF. Another indicator for a possible estimation failure is that one of the parameters has converged to the search space bounds, assuming

that it is a priori possible to define a bounded confined search space. Usually, that is a reasonable assumption and the upper and lower bounds are given by $\mathbf{b}_u$ and $\mathbf{b}_l$, each of the vector entries specifying an upper and lower bound for each parameter of the search space. Additionally let $\hat{s}$ denote the binary value determining success ($s = 1$) or failure ($s = 0$). Then

$$\hat{s} = \begin{cases} 1, & \text{if} \quad \dfrac{\Omega}{\sigma^2_{\hat{h}_l}} \leq \epsilon_{\chi^2}, \quad \mathbf{b}_l + \epsilon_l \leq \theta_i \leq \mathbf{b}_u - \epsilon_u, \quad \forall \theta_i, i \in \{1, \ldots, \mathcal{D}\} \\ 0, & \text{else.} \end{cases} \tag{5.26}$$

Thereby, $\mathcal{D}$ denotes the search space dimension provided by the length of $\hat{\boldsymbol{\theta}}$ and ϵ_b is a user-specified very small value. Here, it is set to $\epsilon_b = 0.01$. Whenever the χ^2-based outage detection fails, because the error cannot be appropriately modelled via a χ^2 distribution the outage detection reduces to the method of checking wether one of the parameters has converged to the bounds.

5.6 SNLLS/ML Optimization Problem Discussion

When searching for suitable optimization methods, it is important to note that unfortunately there is no such thing as a "swiss army knife" for optimization: An all-in-one for every cost function applicable and universally optimally performing optimization algorithm does not exist. This insight was thoroughly and theoretically assessed by formulating and proving the so called "No Free Lunch Theorems" in [WM97]. The theorems establish that the performance of all optimization and random search algorithms, assessed for all possible optimization functions, is equivalent. More precisely, the authors of [WM97] firstly provide information theoretic tools to relate the efficiency of optimization algorithms to the problems they are solving. Secondly they proclaim that the presented "No Free Lunch Theorems" say that, if any algorithm A yields a superior performance compared to any other algorithm B for a special problem type, there exists another subset of problems for which it is the other way around: Algorithm B will outperform algorithm A. Another noteworthy implication of the findings in [WM97] is the importance of using problem-specific a priori knowledge to adjust the optimization algorithms that are used to solve that problems. Bergh remarks that finding optimization methods that work superior for specific problems is nonetheless constructive and suggests in [Ber01] to seek for a subset of functions, for which the "No Free Lunch Theorems" do not hold.

So far we established that there is no perfect optimization method for all problems. As a straightforward consequence, we take a closer look at the optimization problems, we want to solve in this thesis, to find a suitable candidate. Fortunately, the optimization problems addressed in this thesis have a specific structure and therefore specify a small subset of all optimization functions. Obviously the problems in (5.15), (5.23) can be classified as non-linear and multi-dimensional. Both problems (5.15), (5.22) can also be classified as global optimization problems. The reasoning is not obvious, though especially not in case of the reduced problem (5.23), and therefore it is shortly discussed in the following paragraph.

(a) Global and local optima

(b) Convexity

Figure 5.3: Optimization problems showing global and local minima and convexity.

Figure 5.3 shows the difference between a global and a local optimium and it further shows a convex function.

Definition 1 (Local and global optimimum)**.** *For any function $f : I \to \mathbb{R}$, a value x_g is called the global optimum (in this case minimum), if $\forall x : f(x) \geq f(x_g)$. A value x_l is called a local optimum in case $\exists \epsilon : f(x) \geq f(x_l)$ for all x in the distance ϵ.*

The following definition and property yield a method to verify that an optimization is a local optimization problem.

Definition 2 (Convexity)**.** *Geometrically thinking, convexity means, that the line drawn between any two function values has to lie above or exactly on the function values. A method to check whether a function is convex, is to calculate its second derivative and to check whether it is always positive. In the multidimensional case that is equivalent to calculating the Hessian matrix of the cost function and showing that it is positive semi-definite.*

Property 2 (Convexity)**.** *If a function is convex, then any local minimum of this function will be equal to the global minimum $x_l = x_g$, if the function is strictly convex, it further means that the global minimum found is unique.*

Therefore, the question to raise is, whether the cost functions in this thesis are convex or not, or more specifically speaking, whether the cost function's Hessian is always positive semi-definite.

Remark 2 (Separable non-linear LS and local minima)**.** *In [OR13] emphasizes that separating the linear and non-linear parameters, such that, (5.23) is optimized instead of (5.15), definitely yields at least less local optimization function optima, which again indicates that in general we have to deal with more than one local minimum.*

The Hessians associated with the cost functions evaluated in this thesis are known matrices, as they are required for many other purposes anyway. They are required for calculating the theoretical performance via the well known Cramer-Rao Lower bound, or estimating the covariance of the estimates, or being interested in the Fisher information matrix as being a popular information theoretic measure for multiple other purposes. A

Table 5.1: Simulation setup for random sample cost function plots.

Channel		Scenario	WINNER B1-LOS
		C	2
		φ_1	$\pi/4$
		ϕ_1	$\pi/4$
		$\boldsymbol{\tau}$	$[0.6, 0.75]$
System	Signal	g_{RC} with α_{RC}	0.3
		T	200 ns
		L_g	7
	Time/Frequency	f_c	2 GHz
		J	1
		K	1000
		I	1

second interesting fact mentioned in [OR13] is that calculating the Hessian matrix for the separated problem (5.23) is actually based on calculation of the Hessian for the full problem (5.15). The Hessian derivation based on the full problem, but applicable to the separable problem, is known from [GP03, PTVF07]. It can be summarized in the following manner: Consider any non-linear least squares objective function with a residual vector $\mathbf{r} = [r_1, \ldots, r_N]$, N momentarily denoting the number of measurements in this subsection, such that the optimization problem corresponds to

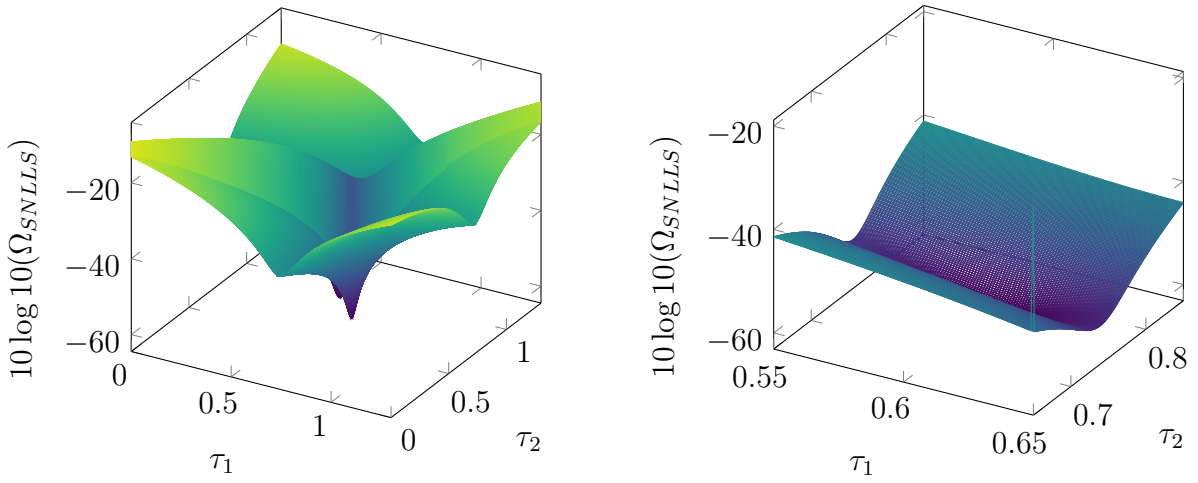
$$\hat{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} ||\mathbf{r}||_2^2 \tag{5.27}$$

similar to (5.15). Then it follows that the exact Hessian is determined by

$$\nabla^2 f(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta})^T \mathbf{J}(\boldsymbol{\theta}) + \sum_{i=1}^{N} r_i(\boldsymbol{\theta}) \nabla^2 r_i(\theta), \tag{5.28}$$

$\mathbf{J}$ denoting the Jacobian of $\mathbf{r}(\boldsymbol{\theta})$. The first dominant term only contains first derivatives, whereas the second term consist of a multiplication with the residuals them self. Assuming that the underlying modeling function is chosen correctly and further reasonably assuming that close to the actual optimum these residuals are small, the Hessian commonly is approximated by neglecting the second term in (5.28) to $\mathbf{J}^T\mathbf{J}$, in favor of stable processing. This approximated matrix is positive semi-definite by definition. Hence it is commonly employed for calculating the Fisher matrix or for instance in Gauss Newton type local optimization algorithms. Nevertheless, this matrix does not correspond to the actual exact Hessian, especially not for parameters far from the actual global optimum. It is merely an approximation, only valid close to the true minimum and therefore it should be always be used with caution. There is no way of guaranteeing that the exact Hessian is positive semi-definite due to the unpredictable nature of the residuals considering all possible parameter choices in the overall search space, including those not close to the actual optimum.

Without further insight, we should handle the problem like a global optimization problem. Having a closer visual look at the cost functions' random samples, we are dealing with, can at least provide some experience regarding the problems nature and properties. Further, seeing more than one local minimum in any cost function realization, would reinforce the decision to handle the problem as a global optimization problem.



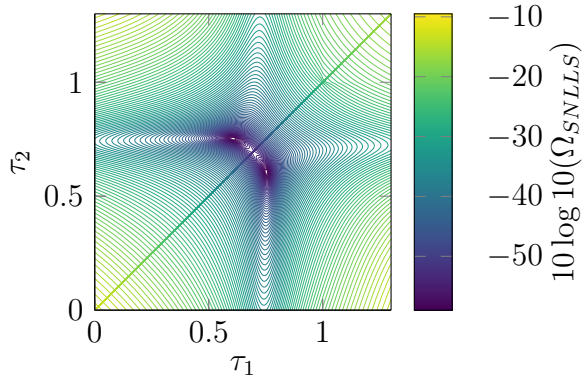
(a) 3d view

(b) 3d zoom in

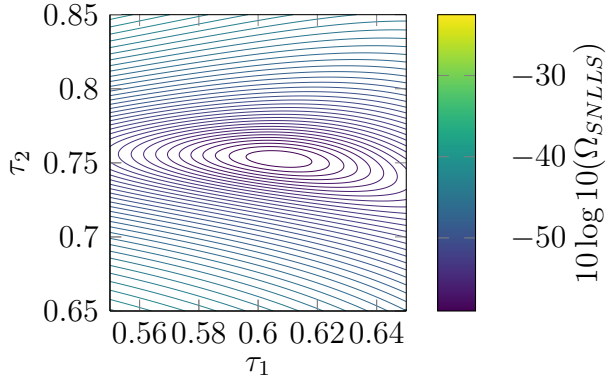
Figure 5.4: The random sample cost function surface plot and its zoom-in indicate that there are not many minima. The diagonal separation by the wall is due to the fact that the cost functions dependence is the same for all τ_c, $c \in \{1, \ldots, C\}$. Each side of the wall is therefore the mirrored version of the other side and hence a minimum on each side is to be expected.

Thanks two the dimensionality reduction via the separability of the linear and non-linear parameter components, at least for the two-path channel the cost function can be visualised. Note that the cost function depends on the channel measurements and therefore on the random channel elements. It further depends on the chosen function $g(\tau)$ and the channel estimation error. Therefore, the Figures 5.4, 5.5 and 5.6 only represent random samples, and conclusions on how to handle the optimization problem have to be drawn with caution. They are based on the simulation parameters given in Table 5.1.

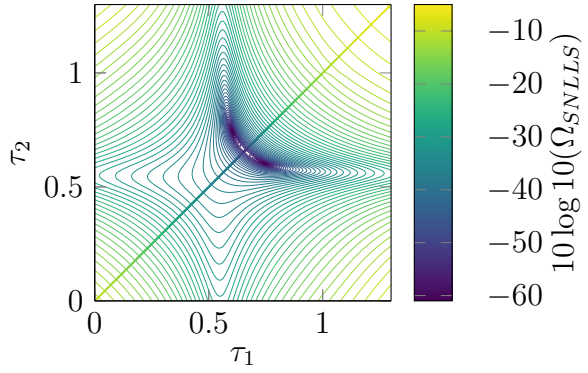
The surface plot and its zoom-in Figure 5.4 indicate twofold: Firstly, the cost function, as expected, is not convex over the entire search space. Secondly, it shows to distinct bassins divided by a diagonal wall. The diagonal wall can be interpreted as a mirror, which can be understood by seeing that the dependence of the cost function on each delay is the same. Therefore one side of the diagonal will be identical to the other side and it is not important on which side the minimum is found, since the elements of $\boldsymbol{\tau}$ are sorted in increasing order after optimization anyway. So we should only have a look at one side of the diagonal only. Comparing the contour plots of three different random cost function samples in Figure 5.5 we further see, that at least for these special two-path channel examples the optimization problem does not suffer from many local minima, and at least in these cases local optimization methods would be a good choice. Rarely it is



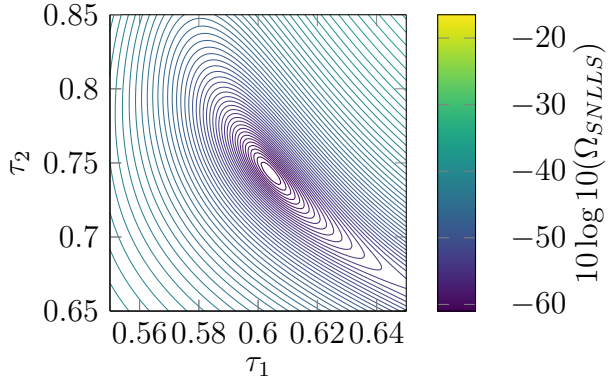
(a) Example 1): 2d contour



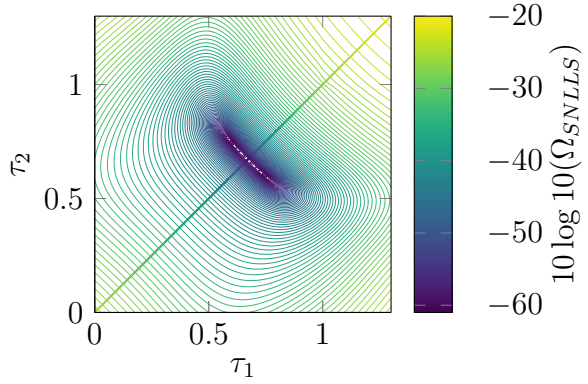
(b) Example 1): contour zoom



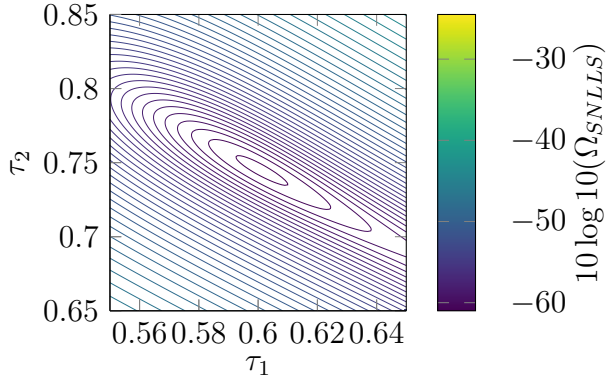
(c) Example 2): 2d contour



(d) Example 2): contour zoom



(e) Example 3): 2d contour



(f) Example 3): contour zoom

Figure 5.5: The left column shows contour plots for random cost function, enables an impression of the cost function properties and hence yields insight in how to choose the optimization algorithm. It becomes clear that although not being convex over the entire search space, the function does not bear many non-global minima. The right hand side column shows zoom ins of the left hand side plots, showing smooth bassins, a convex behaviour, near to minima.

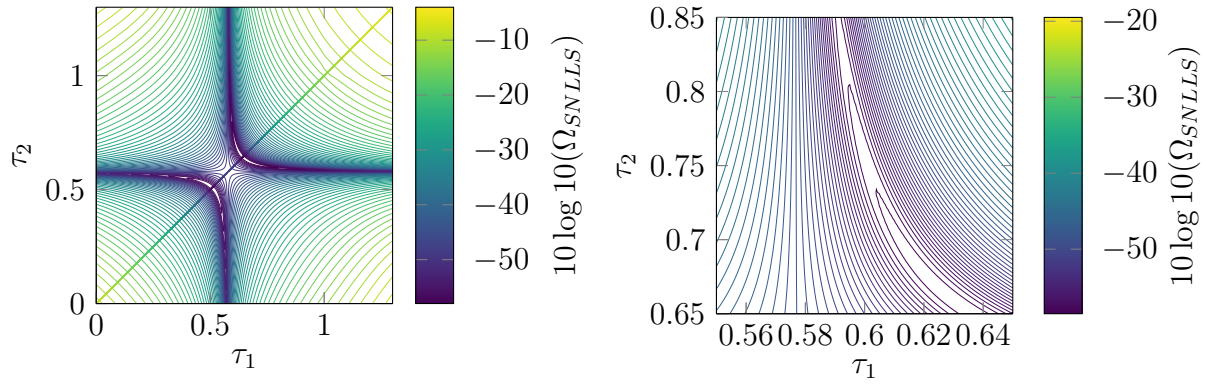


Figure 5.6: This random sample cost function plot shows that there exist rarely occurring snap shot measurements, which result in cost functions bearing more than one local minimum.

possible to find random cost function samples that have two local minima on each side of the diagonal as can be seen finally in Figure 5.6. This proves that there exist cases for which the optimization problem at hand is a global optimization. The insight yielded by visually examining the cost functions, however, indicates that a good solution might be to switch to a local optimization method after quickly exploring the search space with a global optimization method at the beginning. Consequently, it can be concluded that we are generally dealing with a global optimization problem, for which it seems as though usually there are not many local minima.

5.7 Optimising the SNLLS/ML Problem

To tackle optimization problems like (5.15), (5.22) a non-linear, multi-dimensional global optimization algorithm is required. Based on my gained experience comparing potentially competitive algorithms, I chose particle swarm optimization as the most promising and practical candidate in this thesis. This choice has further been motivated by the following considerations. Global optimization algorithms are typically categorised into deterministic, stochastic and (meta-)heuristic algorithms. Deterministic algorithms qualify whenever the possibility of failure is not an option, that is, in cases the global minimum is required to be found with certainty. Unfortunately, finding the global optimum with certainty by employing a deterministic global optimization algorithm, depending on the problem, involves indefinitely long run times, which contradict the wireless JCAP system nature. Another reason not to choose a deterministic global optimization method was that nonlinear optimization problems are still challenging to solve in this manner. More information on deterministic global optimization methods can be found in [Neu04]. A heuristic intelligent algorithm, also bearing stochastic elements regarding the search space exploration, seemed to be a reasonable choice. The most popular and suitable candidates are simulated annealing [KGV83], genetic algorithms [Mit96, SD08] and particle swarm optimization. All three algorithms have in common that they are all able to

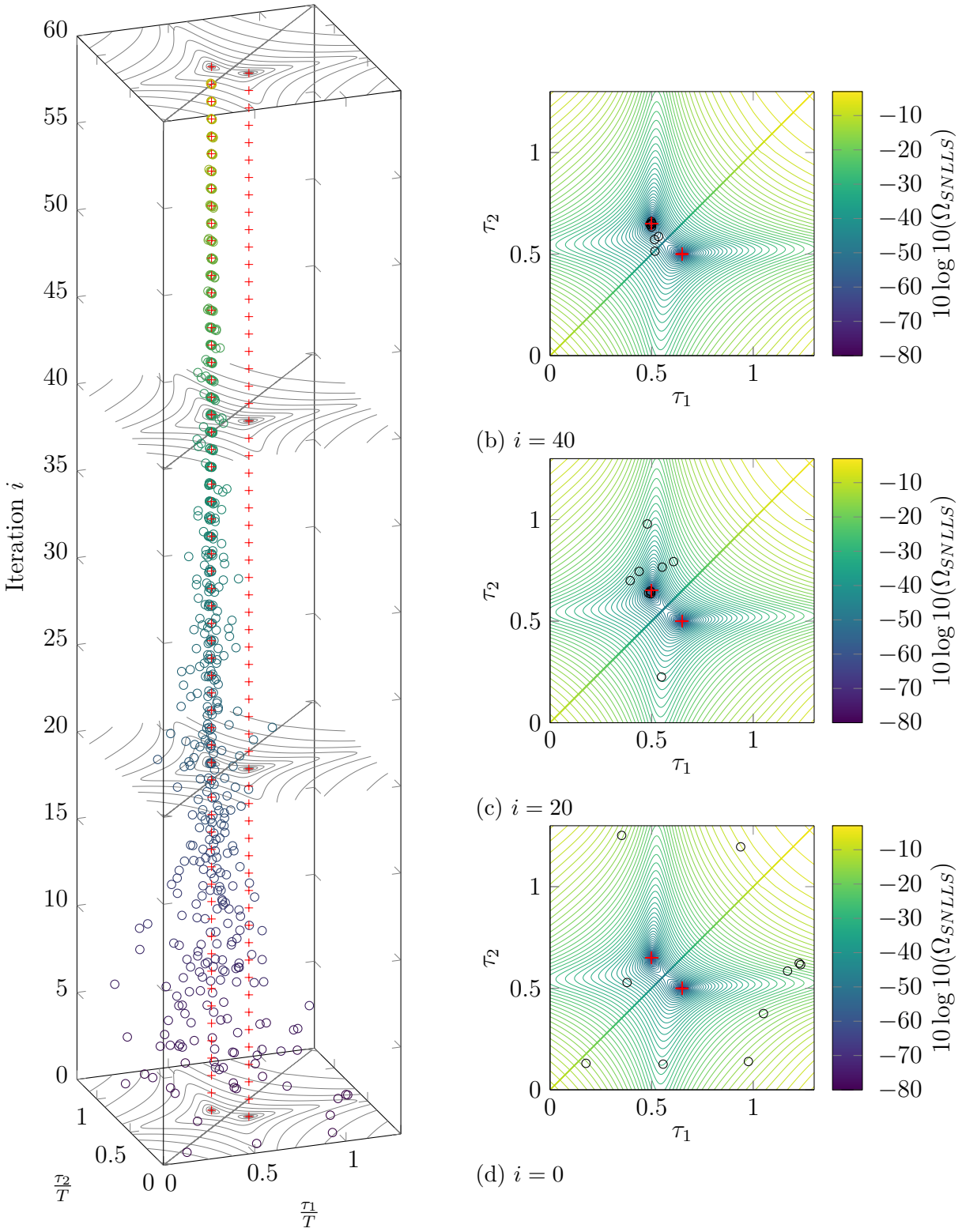
explore the search space in an intelligent manner, inspired by nature in some kind.

5.7.1 Simulated Annealing

Simulated annealing, as originally proposed in [KGV83], targets finding solutions for discrete and combinatorial optimization problems, such as the well-known traveling salesman problem [PTVF07]. It is further shown in [PTVF07] that combining it with the downhill simplex method [NMN64] it can be used as a global optimization method for continuous optimization problems. Annealing, a thermodynamic concept used in metallurgy, stands for the controlled material heating and cooling and affects the thermodynamic free energy and temperature. Slow material cooling leads to an ordered atom arrangement, yielding a low (in terms of optimization "optimal") energy state. In terms of optimization the slow material cooling represents the slow probability decrease of allowing worse cost function values than in the preceding iterations. In thermodynamics, a Boltzmann probability distribution describes the system state depending on the energy and the temperature and therefore is employed in the optimization algorithm. Contrary to local optimization algorithms all global optimization algorithms, are required to be able to not only move "downhill". To have the ability to escape local minima, they should allow to explore the search space. Therefore, from time to time, over the iterations they should allow to accept worse cost function values than the preceding ones. Simulating a high temperature at the beginning leads to such desired down- and uphill movements that, in return, yield an exploration of the overall search space. With increasing iterations, the temperature is slowly decreased and the probability of going mainly downhill, that is, performing a mainly local search, is increased.

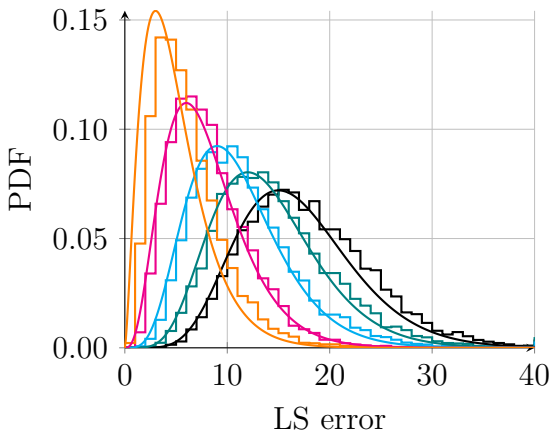
5.7.2 Genetic Algorithms

Genetic algorithms are also meta-heuristic algorithms and also copy intelligent behaviour from nature. More specifically, they copy an evolution following the principle "survival of the fittest": A population of randomly generated individuals, the candidate solutions, have a set of properties (chromosomes or genotypes) and they explore the search space iteratively as introduced in [Mit96]. The population of each iteration, representing a new generation, is determined by the laws of genetics: Evaluation and comparison of a cost function, here called fitness function, for all the members of the population allows to more often select fit individuals for recombination and random mutation to obtain the next generation. The iterations proceed until a maximum number of iterations is exceeded or a stopping criterion is met. A stopping criterion could be to terminate in case the change of fitness values lies beneath a certain tolerance value. Similar to simulated annealing due to the partially stochastic nature the algorithm allows to escape local minima and is able to find the global optimum.

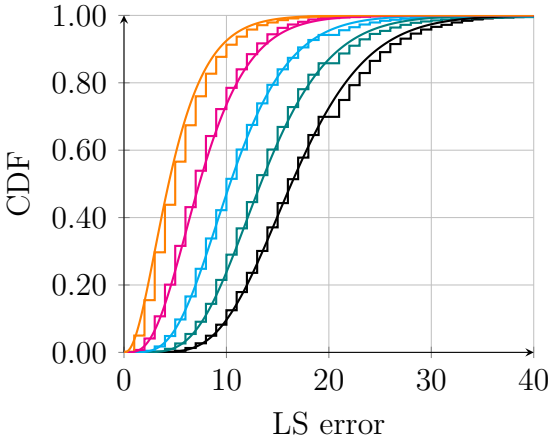


(a) Particle Swarm Optimization

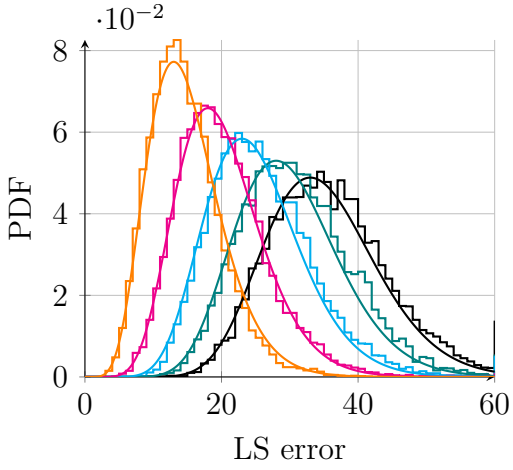
Figure 5.7: The particle swarm optimization visualization shows for a two-path scenario that the 10 particles only need a few iterations to converge to ascending order solution.



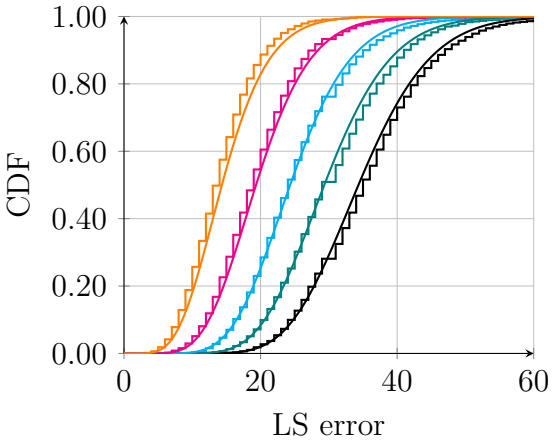
(a) PDF: $N_R = 1$, $I = 1$



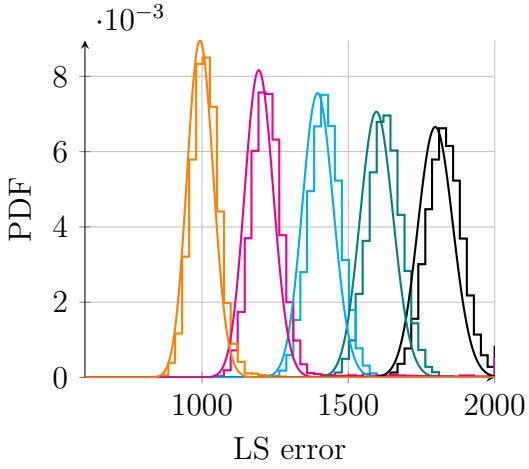
(b) CDF: $N_R = 1$, $I = 1$



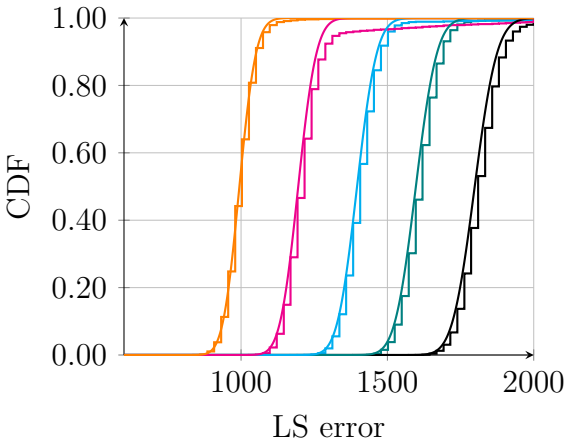
(c) PDF: $N_R = 2$, $I = 1$



(d) CDF: $N_R = 2$, $I = 1$



(e) PDF: $N_R = 1$, $I = 100$



(f) CDF: $N_R = 1$, $I = 100$

Figure 5.8: Experimental verification for $L = 9$, that the PSO estimation calculated LS error coincides with the χ^2-distribution simulated for the case $\frac{E_s}{N_0} = 60$ dB and a varying path number $C = 1, 2, 3, 4, 5$ (black, red, green, blue, purple, light blue).

Table 5.2: Simulation setup.

N_p	$10D$
I_{Max}	20000
$I_{StopMax}$	20
ε_{tol}	10^{-6}
Swarm Initialization	$\forall p \in \{1, \ldots, N_p\}$: $\forall d \in \{1, \ldots \mathcal{D}\}$: $\tau_{d,p} \in \mathcal{U}(b_l(d), b_u(d))$

5.7.3 Particle Swarm Optimization

Particle swarm optimization is yet another metaheuristic a global optimization algorithm mimicking the social behaviour of bird or fish swarms. It was originally proposed by Kennedy and Eberhart 1995 in [KE95]. Roughly ten years later Kennedy and Bratton defined a standard for PSO in [BK07], based on the most important developments in the PSO research area. The fundamental idea of PSO was that particles explore a search space by determining subsequent particle positions, by employing a partially random weighted combination

- of moving towards a so called individual best position $p_{IB,i}$, taking into account the particles' history,
- and of moving towards a so-called local[1] best position $p_{LB,i}$, the best position compared for each particle and a specified set of the particles' neighbours.

Let the number of particles be defined by N_p. Further let the optimization problem dimensionality be denoted by D. The absolute maximum number of iterations the algorithm is allowed to carry out is denoted by I_{Max}. Let further $\mathbf{v}^i$ be the velocity vector specifying the particles' movement direction.

Firstly, the algorithm has to be initialized appropriately. Hence, the particles' positions p_i, if not specified otherwise, are initially randomly uniformly distributed within the specified search space lower and upper bounds given by $\mathbf{U}_l = [b_l(1), \ldots, b_l(D)]$ and $\mathbf{U}_u = [b_u(1), \ldots, b_u(D)]$ such that

$$\mathbf{p}_i \in \mathcal{U}(\mathbf{u}_l, \mathbf{u}_u). \tag{5.29}$$

Like the particle positions, at the beginning, this vector $\mathbf{v}^1$ is initialized by random values drawn out of the uniform distribution, such that

$$\mathbf{v}^1 \in \mathcal{U}(-(\mathbf{U}_u - \mathbf{U}_l), (\mathbf{U}_u - \mathbf{U}_l)). \tag{5.30}$$

The size of the neighborhood is given by

$$N_N = \max(1, \lfloor r_N \cdot N_p \rfloor), \tag{5.31}$$

[1]Depending on the specified topology, sometimes, as it is proposed in the original algorithm instead of local best the so-called global best position is used. Using the global best leads to a full neighbourhood topology: Every particle can communicate with every other particle. For global optimization problems, due to premature convergence properties, this topology is known to be outperformed by choosing less neighbours and using the local best position.

which again is determined by the user choosing $0 < r_N \leq 1$. If not specified otherwise, in this thesis I chose $r_N = 0.25$. In favor of an improved balance between global and local exploration [BK07] the original PSO in [KE95] was modified by proposing two slightly different velocity update rules involving two slightly different parameters designed to limit the velocity. Namely, these parameters are the inertia weight ω proposed in [SE98] and the constriction factor χ [CK02]. If the inertia weight is used, it is initially set to

$$\omega = \begin{cases} \max(\mathbf{r}_I) & \text{if } \mathbf{r}_i > 0 \\ \min(\mathbf{r}_I) & \text{else,} \end{cases}$$

where $\mathbf{r}_I$ denotes the inertia range, that is, the upper and lower bounds, for an adaptive adjustment of the inertia weight over the iterations as proposed in [SE98]. If not specified otherwise it is set to $\mathbf{r}_I = [0.1, 1.1]$ in this thesis. Then these particles iteratively move through the search space by updating each of the $n = \{1, \ldots, N_p\}$ particle velocities either if the inertia weight is used by

$$\mathbf{v}_n^{i+1} = w\mathbf{v}_n^i + c_1\boldsymbol{\varepsilon}_1^i \odot (\mathbf{p}_{\mathrm{IB},n}^i - \mathbf{p_n}) + c_2\boldsymbol{\varepsilon}_2^i \odot (\mathbf{p}_{\mathrm{LB},n}^i - \mathbf{p}_n), \tag{5.32}$$

or if the constriction factor is used by

$$\mathbf{v}_n^{i+1} = \chi(\mathbf{v}_n^i + c_1\boldsymbol{\varepsilon}_1^i \odot (\mathbf{p}_{\mathrm{IB},n}^i - \mathbf{p_n}) + c_2\boldsymbol{\varepsilon}_2^i \odot (\mathbf{p}_{\mathrm{LB},n}^i - \mathbf{p}_n)). \tag{5.33}$$

In both cases, the position is then updated according to

$$\mathbf{p}_n^{i+1} = \mathbf{p}_n^i + \mathbf{v}_n^{i+1}. \tag{5.34}$$

Note, that $\boldsymbol{\varepsilon}_1^i, \boldsymbol{\varepsilon}_2^i$ are vectors of length D, with independently randomly generated elements drawn out of $\mathcal{U}(0, 1)$. Further note, that the weights c_1, c_2 correspond to a self and a social adjustment weight. To ensure convergence [BK07] proposes to use $\chi \approx 0.72984$ and $c_1 = C_2 + 2.05$. In this thesis, if not specified otherwise, I chose to use the inertia weight and $c_1 = c_2 = 1.49$. However, for a further discussion about how the choice of χ, c_1, c_2 impacts the convergence properties, I refer to [BK07].

The tradeoff between global and local search capabilities is given due to the partly random, partly deterministic updating strategies in (5.32), (5.33). The cost function values for PSO are called fitness values, as it is done in genetic algorithms. Via communication of so-called neighboring particles a local best position $\mathbf{p}_{\mathrm{LB,n}}^{i+1}$ is determined for each particle $p \in \{1, \ldots, N_p\}$. Then the globally best position is determined. The choices made in this thesis for the algorithms tuning parameters are summarized in Table 5.2. Due to PSO's heuristic nature and due to the fact that in this thesis the algorithm is employed to tackle multi-dimensional non-linear global optimization problems, conversion to the global optimum cannot be guaranteed. Therefore, a strategy to identify algorithmic failure proves to be helpful. To this end, recall that the least squares error supposedly is χ^2-distributed. In case this property is valid for the least squares error resulting from the PSO optimization procedure, this knowledge can be used to identify PSO's optimization failure. In Figure 5.7 an exemplary particle swarm optimization is shown over the iterations for a two-path channel. After a few iterations a swarm converges to one of the two global optima. An experimental verification that the LS error coincides with the analytical χ^2-distribution is shown in Figure 5.8.

Algorithm 1 Particle Swarm Optimization

1: Initialize $\tilde{\boldsymbol{\tau}}_p^0$, $\mathbf{v}_p^0$ for all particles p.
2: **while** stopping criterium is not fulfilled **do**
3: **for** all particles p **do**
4: $\mathbf{v}_p^{i+1} \leftarrow \chi(\mathbf{v}_p^i + \lambda_1 \epsilon_1^i \odot (\tilde{\boldsymbol{\tau}}_{\mathrm{IB},p}^i) + \lambda_2 \epsilon_2^i \odot (\tilde{\boldsymbol{\tau}}_{\mathrm{LB},p}^i))$.
5: $\tilde{\boldsymbol{\tau}}_p^{i+1} \leftarrow \tilde{\boldsymbol{\tau}}_p^i + \boldsymbol{v}_p^{i+1}$.
6: Evaluate fitness $\Omega(\tilde{\boldsymbol{\tau}}_p^{i+1})$.
7: Determine individual and local best positions $\tilde{\boldsymbol{\tau}}_{\mathrm{IB},p}^{i+1}$, $\tilde{\boldsymbol{\tau}}_{\mathrm{LB},p}^{i+1}$.

5.7.4 The Levenberg-Marquardt Method as a Local Search Alternative and for Hybrid use with PSO

The Levenberg-Marquardt method is a good alternative, if an initial guess is available or if it is supposed to by used in a hybrid manner together with PSO. The Levenberg-Marquardt method is an iterative local optimization method that intelligently varies between behaving like the Gauss-Newton algorithm and the steepest descent method [PTVF07]. The Gauss-Newton algorithm bases on Newton's method of finding the roots of the first derivative of a function. Thereby the function is approximated by a second-order Taylor expansion

$$\Omega(\tilde{\boldsymbol{\xi}}) \approx \Omega(\tilde{\boldsymbol{\xi}}_i) + \underbrace{(\tilde{\boldsymbol{\xi}} - \tilde{\boldsymbol{\xi}}_i)}_{\Delta\tilde{\boldsymbol{\xi}}} \nabla\Omega(\tilde{\boldsymbol{\xi}}_i) + \frac{1}{2}(\tilde{\boldsymbol{\xi}} - \tilde{\boldsymbol{\xi}}_i)^T \boldsymbol{\mathcal{H}}_\Omega(\tilde{\boldsymbol{\xi}})(\tilde{\boldsymbol{\xi}} - \tilde{\boldsymbol{\xi}}_i), \tag{5.35}$$

where $\nabla\Omega(\tilde{\boldsymbol{\xi}})$ and $\boldsymbol{\mathcal{H}}_\Omega(\tilde{\boldsymbol{\xi}})$ denote the nabla operator for the gradient and the Hessian matrix, respectively. The Jacobian of the model function $\tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}})$ is defined by

$$[\mathbf{J}(\tilde{\boldsymbol{\xi}})]_{l,m} = \frac{\partial \tilde{h}_l}{\partial \tilde{\xi}_m}. \tag{5.36}$$

The gradient is defined elementwise as

$$[\nabla\Omega(\tilde{\boldsymbol{\xi}})]_m = \frac{\partial\Omega(\tilde{\boldsymbol{\xi}})}{\partial\tilde{\xi}_m} \tag{5.37}$$

$$= -2\mathrm{Re}\left(\frac{\partial\tilde{\mathbf{h}}^H(\tilde{\boldsymbol{\xi}})}{\partial\tilde{\xi}_m}(\hat{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}}))\right), \text{ or} \tag{5.38}$$

$$\nabla\Omega(\tilde{\boldsymbol{\xi}}) = -2\mathrm{Re}\left(\mathbf{J}(\tilde{\boldsymbol{\xi}})^H(\hat{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}}))\right). \tag{5.39}$$

The derivation can be found in the Appendix A.3. By applying the product rule, it is straightforward to see that the Hessian is defined elementwise as

$$[\boldsymbol{\mathcal{H}}_\Omega(\tilde{\boldsymbol{\xi}})]_{m,n} = \frac{\partial^2 \Omega(\tilde{\boldsymbol{\xi}})}{\partial \tilde{\xi}_m \partial \tilde{\xi}_n} \tag{5.40}$$

$$= -2 \frac{\partial}{\partial \tilde{\xi}_m} \mathrm{Re}\left(\frac{\partial \tilde{\mathbf{h}}^H(\tilde{\boldsymbol{\xi}})}{\partial \tilde{\xi}_n} (\hat{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}})) \right) \tag{5.41}$$

$$= 2\mathrm{Re}\left(\frac{\partial \tilde{\mathbf{h}}^H(\tilde{\boldsymbol{\xi}})}{\partial \xi_m} \frac{\partial \tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}})}{\partial \xi_n} - \frac{\partial^2 \tilde{\mathbf{h}}^H(\tilde{\boldsymbol{\xi}})}{\partial \xi_m \partial \xi_n} (\hat{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}})) \right). \tag{5.42}$$

This exact formulation depends on the estimation error, the residual, that is small enough to be neglected, close to the exact solution. Hence, the problem is called a small-residual problem. In the large residual case the term cannot be neglected. Far away from the exact solution, the residuals can become large. Approximating the Hessian by neglecting the right hand side term of the exact solution in (5.42) such that

$$[\mathbf{H}_\Omega(\tilde{\boldsymbol{\xi}})]_{m,n} = 2\mathrm{Re}\left(\frac{\partial \tilde{\mathbf{h}}^H(\tilde{\boldsymbol{\xi}})}{\partial \xi_m} \frac{\partial \tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}})}{\partial \xi_n} \right) \text{ or} \tag{5.43}$$

$$\mathbf{H}_\Omega(\tilde{\boldsymbol{\xi}}) = 2\mathrm{Re}(\mathbf{J}^H(\tilde{\boldsymbol{\xi}})\mathbf{J}(\tilde{\boldsymbol{\xi}})) \tag{5.44}$$

$$\approx \boldsymbol{\mathcal{H}}_\Omega \tag{5.45}$$

has the advantage that the approximated Hessian terms are forced to be positive semi-definite. Positive definiteness is required when it comes to solving the constructed linearized system of equations. The Hessian corresponds to the Fisher Information matrix [Kay10].
The derivative of (5.35) is

$$\frac{\partial \Omega(\tilde{\boldsymbol{\xi}})}{\partial \tilde{\boldsymbol{\xi}}} \approx \nabla \Omega(\tilde{\boldsymbol{\xi}}_i) + \mathbf{H}_\Omega(\tilde{\boldsymbol{\xi}}_i)(\tilde{\boldsymbol{\xi}} - \tilde{\boldsymbol{\xi}}_i). \tag{5.46}$$

Setting the right hand side of (5.46) to zero and reorganising the equation yields

$$\Delta \tilde{\boldsymbol{\xi}} = -\mathbf{H}_\Omega^{-1}(\tilde{\boldsymbol{\xi}}_i) \nabla \Omega(\tilde{\boldsymbol{\xi}}_i). \tag{5.47}$$

This is the step size that can be used to find the minimum by iteratively updating the hypothetical parameter vector

$$\tilde{\boldsymbol{\xi}}_{i+1} = \tilde{\boldsymbol{\xi}}_i + \Delta \tilde{\boldsymbol{\xi}} \tag{5.48}$$

$$= \tilde{\boldsymbol{\xi}}_i - \mathbf{H}_\Omega^{-1}(\tilde{\boldsymbol{\xi}}_i) \nabla \Omega(\tilde{\boldsymbol{\xi}}_i). \tag{5.49}$$

Using the approximated Hessian in (5.49) instead of the exact Hessian leads to the Gauss-Newton setup. The Levenberg-Marquardt method yields an improvement in that (5.49) is slightly modified by using a tuning factor to vary between the actual Gauss-Newton

algorithm and the steepest descent, whenever this is reasonable to do in the following manner:

$$\tilde{\boldsymbol{\xi}}_{i+1} = \tilde{\boldsymbol{\xi}}_i - (\mathbf{H}_\Omega(\tilde{\boldsymbol{\xi}}_i) + \eta\mathbf{I})^{-1}\nabla\Omega(\tilde{\boldsymbol{\xi}}_i). \tag{5.50}$$

If η is chosen close to zero it behaves similar to the Gauss-Newton algorithm. If η is chosen large the approximated Hessian matrix is negligible compared to the terms including η and it can be seen that the method behaves like the steepest descent algorithm.

Belonging to the class of local optimization algorithms, the method requires a good initial

Algorithm 2 Levenberg-Marquardt Method

1: Initialize $\tilde{\boldsymbol{\xi}}_m$ for all parameters m and choose a moderate value for the tuning factor like $\eta = 0.001$ and set $i = 0$.
2: **while** stopping criterium is not fulfilled and $i < I_{MAX}$ **do**
3: Determine $\Omega(\tilde{\boldsymbol{\xi}}_i)$, $\mathbf{J}(\tilde{\boldsymbol{\xi}}_i)$ and $\mathbf{H}_\Omega(\tilde{\boldsymbol{\xi}}_i)$
4: $\tilde{\boldsymbol{\xi}}_{i+1} \leftarrow \tilde{\boldsymbol{\xi}}_i - (\mathbf{H}_\Omega(\tilde{\boldsymbol{\xi}}_i) + \eta\mathbf{I})^{-1}\nabla\Omega(\tilde{\boldsymbol{\xi}}_i)$.
5: **if** $\Omega(\tilde{\boldsymbol{\xi}}_{i+1}) \geq \Omega(\tilde{\boldsymbol{\xi}}_{i+1})$ **then**
6: $\tilde{\boldsymbol{\xi}}_{i+1} \leftarrow \tilde{\boldsymbol{\xi}}_i$ (go back to old hypothesis)
7: $\eta_{i+1} \leftarrow \eta_i \cdot 10$ (substantially increase tuning factor)
8: **if** $\Omega(\tilde{\boldsymbol{\xi}}_{i+1}) < \Omega(\tilde{\boldsymbol{\xi}}_{i+1})$ **then**
9: $\tilde{\boldsymbol{\xi}}_{i+1} \leftarrow \tilde{\boldsymbol{\xi}}_i - (\mathbf{H}_\Omega(\tilde{\boldsymbol{\xi}}_i) + \eta\mathbf{I})^{-1}\nabla\Omega(\tilde{\boldsymbol{\xi}}_i)$ (choose new hypothesis)
10: $\eta_{i+1} \leftarrow \eta_i \cdot 0.1$ (substantially decrease tuning factor)
11: $i \leftarrow i + 1$

guess. I define a good initial guess as a guess that is closer to the global optimum than it is close to any other local optimum. Obviously finding such good starting points is almost impossible. Using a global optimization algorithm on the same problem for the major part of the iterations and afterwards using the parameter estimates of the global optimization strategy as the initial guess for the local strategy can be a reasonable approach. Then, however, it will be an art of its own to define the number of iterations used for the global and the number of iterations used for the local strategy. An alternative approach could be to find a sub-optimal closed-form solution and to use the solution as an initial guess to the optimal local solver. Both methods would not guarantee to find the global optimum after all. That is why I, if possible, would prefer to use the global optimization algorithm to begin with. For

$$\Omega = \Omega_{SNLLS}, \quad \boldsymbol{\theta} = \boldsymbol{\tau}, \tag{5.51}$$

the nabla operator and the Hessian are

$$\nabla_{\Omega_{SNLLS}} = -2\mathrm{Re}\left\{\hat{\mathbf{h}}^H P_{\mathbf{G}}^{\perp}\mathbf{J}_{\mathbf{G}}\mathrm{diag}\left(\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right)\right\}, \tag{5.52}$$

$$\mathbf{H}_{\Omega_{SNLLS}} = 2\mathrm{Re}\left\{\mathrm{diag}\left(\hat{\mathbf{h}}^H\mathbf{G}^{\dagger^T}\right)\mathbf{J}_{\mathbf{G}}^T\mathbf{P}_{\mathbf{G}}^{\perp}\mathbf{J}_{\mathbf{G}}\mathrm{diag}\left(\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right)\right\}. \tag{5.53}$$

5.8 Derivation of the Hessian Matrix for SNLLS

To provide the derivation for (5.52) and (5.53), the following auxiliary mathematical properties are required:

Property 3. (Auxiliary Matrix Calculus Rules*:)*
Let $\mathbf{X}$ *and* $\mathbf{Y}$ *be arbitrary matrices and* $\mathbf{a}$ *be an arbitrary column vector.*
General auxiliary properties:

$$\mathbf{P}_{\mathbf{G}} \text{ and } \mathbf{P}_{\mathbf{G}}^{\perp} \text{ are idempotent and symmetric.} \tag{5.54}$$

$$\mathbf{P}_{\mathbf{G}}^{\perp}{\mathbf{G}^{\dagger}}^{T} = 0, \quad \mathbf{G}^{\dagger}\mathbf{P}_{\mathbf{G}}^{\perp} = 0 \quad \mathbf{G}^{T}\mathbf{P}_{\mathbf{G}}^{\dagger} = 0, \quad \mathbf{P}_{\mathbf{G}}\mathbf{G} = 0 \tag{5.55}$$

$$\mathbf{X}\mathbf{J}^{m,n} \text{ selects the } m\text{th column.} \tag{5.56}$$

$$\mathbf{J}^{m,n}\mathbf{X} \text{ selects the } n\text{th row.} \tag{5.57}$$

$$\mathbf{X}\mathbf{J}^{m,m}\mathbf{a} = \left[\mathbf{X}\mathrm{diag}\left(\mathbf{a}\right)\right]_{(:),m} \tag{5.58}$$

$$\mathbf{X} + \mathbf{X}^{H} = 2\mathrm{Re}\{\mathbf{A}\} \text{ if } \mathbf{X} \text{ is symmetric.} \tag{5.59}$$

General matrix derivative rules:

$$\partial(\mathbf{X}\mathbf{Y}) = (\partial\mathbf{X})\mathbf{Y} + \mathbf{X}(\partial\mathbf{Y}) \tag{5.60}$$

$$\partial(\mathbf{X}^{-1}) = -\mathbf{X}^{-1}\partial(\mathbf{X})\mathbf{X}^{-1} \tag{5.61}$$

$$\partial((\mathbf{X}^{H}\mathbf{X})^{-1}\mathbf{X}) = -\mathbf{X}^{\dagger}\partial(\mathbf{X})\mathbf{X}^{\dagger} + \mathbf{X}^{\dagger}{\mathbf{X}^{\dagger}}^{T}(\partial\mathbf{X}^{T})(\mathbf{I} - \mathbf{X}\mathbf{X}^{\dagger}) \tag{5.62}$$

Problem specific matrix derivatives:

$$\frac{\partial}{\partial\tau_m}(\mathbf{P}_{\mathbf{G}}) = \mathbf{P}_{\mathbf{G}}^{\perp}\frac{\partial\mathbf{G}}{\partial\tau_m}\mathbf{G}^{\dagger} + {\mathbf{G}^{\dagger}}^{T}\frac{\partial\mathbf{G}^{T}}{\partial\tau_m}\mathbf{P}_{\mathbf{G}}^{\perp} \tag{5.63}$$

$$\frac{\partial}{\partial\tau_m}(\mathbf{P}_{\mathbf{G}}^{\perp}) = -\frac{\partial}{\partial\tau_m}(\mathbf{P}_{\mathbf{G}}) \tag{5.64}$$

$$\frac{\partial}{\partial\tau_m}\mathbf{G} = \mathbf{J}_{\mathbf{g}}(\boldsymbol{\tau})\mathbf{J}^{m,m}, \quad \left[\frac{\partial}{\partial\tau_m}\mathbf{G}\right]_{l,n} = \frac{\partial g(lT - \tau_n)}{\partial\tau_m} \tag{5.65}$$

$$[\mathbf{J}_{\mathbf{g}}(\boldsymbol{\tau})]_{l,m} = \frac{\partial g(lT - \tau_m)}{\partial\tau_m} \tag{5.66}$$

The general matrix derivation rules (5.60), (5.61) can be found in [PP12] and [Lue96]. By employing (5.60), (5.61) it is easy to see that (5.62) is true, which can also be found in [GP73]. A compact derivation is given in the Appendix in A.4. Using (5.60), (5.62) the problem specific projection matrix derivative is straightforward to derive and also can be found in [GP73]. Given these auxiliary calculation tools, the Jacobian $\mathbf{J}_{\tilde{\mathbf{h}}}$, the SNLLS function derivative and the approximated Hessian $\mathbf{H}_{\Omega_{SNLLS}}$ can be calculated.

5.8.1 The Jacobian

$$
\begin{aligned}
[\mathbf{J}_{\tilde{\mathbf{h}}}(\boldsymbol{\tau})]_{l,m} =& \frac{\partial \tilde{h}_l}{\partial \tau_m} && (5.67)\\
=& \left[\left(\frac{\partial \mathbf{P}_{\mathbf{G}}}{\partial \tau_m}\right)\hat{\mathbf{h}}\right]_l && (5.68)\\
\overset{(5.63)}{=}& \left[\left(\mathbf{P}_{\mathbf{G}}^{\perp}\frac{\partial \mathbf{G}}{\partial \tau_m}\mathbf{G}^{\dagger} + \mathbf{G}^{\dagger^T}\frac{\partial \mathbf{G}^T}{\partial \tau_m}\mathbf{P}_{\mathbf{G}}^{\perp}\right)\hat{\mathbf{h}}\right]_l && (5.69)\\
\overset{(5.65)}{=}& \left[\left(\mathbf{P}_{\mathbf{G}}^{\perp}\mathbf{J}_{\mathbf{g}}(\boldsymbol{\tau})\mathbf{J}^{m,m}\mathbf{G}^{\dagger} + \mathbf{G}^{\dagger^T}\mathbf{J}^{m,m}{\mathbf{J}_{\mathbf{g}}}^T(\boldsymbol{\tau})\mathbf{P}_{\mathbf{G}}^{\perp}\right)\hat{\mathbf{h}}\right]_l && (5.70)\\
=& \left[\left[\mathbf{P}_{\mathbf{G}}^{\perp}\mathbf{J}_{\mathbf{g}}(\boldsymbol{\tau})\mathrm{diag}\left(\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right)\right]_{(:,m)} + \left[\mathbf{G}^{\dagger^T}\mathrm{diag}\left({\mathbf{J}_{\mathbf{g}}}^T\mathbf{P}_{\mathbf{G}}{}^{\perp}\hat{\mathbf{h}}\right)\right]_{(:,m)}\right]_l && (5.71)\\
=& \left[\mathbf{P}_{\mathbf{G}}^{\perp}\mathbf{J}_{\mathbf{g}}(\boldsymbol{\tau})\mathrm{diag}\left(\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right) + \mathbf{G}^{\dagger^T}\mathrm{diag}\left({\mathbf{J}_{\mathbf{g}}}^T\mathbf{P}_{\mathbf{G}}{}^{\perp}\hat{\mathbf{h}}\right)\right]_{l.m}. && (5.72)
\end{aligned}
$$

5.8.2 SNLLS Derivative Derivation

$$
\begin{aligned}
\frac{\partial}{\partial \tau_m}\Omega_{SNLLS} \overset{(5.39)}{=}& -2\mathrm{Re}\left\{{\mathbf{J}_{\mathbf{h}}}^T(\boldsymbol{\tau})\left(\hat{\mathbf{h}} - \tilde{\mathbf{h}}(\boldsymbol{\tau})\right)\right\} && (5.73)\\
=& -2\mathrm{Re}\{{\mathbf{J}_{\mathbf{h}}}^T(\boldsymbol{\tau})\mathbf{P}_{\mathbf{G}}{}^{\perp}\hat{\mathbf{h}}\} && (5.74)\\
\overset{(5.70,5.55)}{=}& -2\mathrm{Re}\left\{\hat{\mathbf{h}}^H\mathbf{G}^{\dagger^T}\mathbf{J}^{m,m}{\mathbf{J}_{\mathbf{g}}}^T(\boldsymbol{\tau})\mathbf{P}_{\mathbf{G}}^{\perp}\mathbf{P}_{\mathbf{G}}{}^{\perp}\hat{\mathbf{h}}\right\} && (5.75)\\
\overset{(5.54)}{=}& -2\mathrm{Re}\left\{\hat{\mathbf{h}}^H\mathbf{G}^{\dagger^T}\mathbf{J}^{m,m}{\mathbf{J}_{\mathbf{g}}}^T(\boldsymbol{\tau})\mathbf{P}_{\mathbf{G}}^{\perp}\hat{\mathbf{h}}\right\} && (5.76)\\
\overset{\mathbf{B}\in\mathbb{C}^{1\times 1}}{=}& -2\mathrm{Re}\left\{\hat{\mathbf{h}}^H\mathbf{P}_{\mathbf{G}}^{\perp}\mathbf{J}_{\mathbf{G}}\mathbf{J}^{m,m}\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right\} && (5.77)
\end{aligned}
$$

Further employing (5.58) yields (5.52).

For the further calculations let us use the auxiliaries $\mathbf{A}$ and $\mathbf{B}$ in order to work with a compact notation for the summands in (5.72).

$$
\begin{aligned}
\mathbf{A} &= \mathbf{P}_{\mathbf{G}}^{\perp}\mathbf{J}_{\mathbf{g}}(\boldsymbol{\tau})\mathrm{diag}\left(\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right) && (5.78)\\
\mathbf{B} &= \mathbf{G}^{\dagger^T}\mathrm{diag}\left({\mathbf{J}_{\mathbf{g}}}^T\mathbf{P}_{\mathbf{G}}{}^{\perp}\hat{\mathbf{h}}\right) && (5.79)
\end{aligned}
$$

5.8.3 SNLLS Approximated Hessian Derivation

In order to derive the approximated Hessian I use (5.42):

$$
\begin{aligned}
[\mathbf{H}_{\Omega_{SNLLS}}]_{m,n} &= \left[2\mathrm{Re}\left(\mathbf{J}_{\hat{\mathbf{h}}}^{H}(\boldsymbol{\tau})\mathbf{J}_{\hat{\mathbf{h}}}(\boldsymbol{\tau})\right)\right]_{m,n} && (5.80)\\
&= \left[2\mathrm{Re}\left(\mathbf{A}+\mathbf{B}\right)^{T}(\mathbf{A}+\mathbf{B})\right)\right]_{m,n} && (5.81)\\
&= \left[2\mathrm{Re}\left((\mathbf{A}^{H}\mathbf{A}+\mathbf{A}^{H}\mathbf{B}+\mathbf{B}^{H}\mathbf{A}+\mathbf{B}^{H}\mathbf{B})\right)\right]_{m,n} && (5.82)\\
&\overset{(5.55)}{=} \left[(\mathbf{A}^{H}\mathbf{A}+\mathbf{B}^{H}\mathbf{B})\right]_{m,n} && (5.83)\\
&= \Big[2\mathrm{Re}\left(\mathrm{diag}\left(\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right)\mathbf{J}_{\mathbf{g}}^{H}\mathbf{P}_{\mathbf{G}}^{\perp}\mathbf{J}_{\mathbf{g}}\mathrm{diag}\left(\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right)\right)\\
&\quad + \underbrace{\mathrm{diag}\left(\hat{\mathbf{h}}^{H}\mathbf{P}_{\mathbf{G}}{}^{\perp}\mathbf{J}_{\mathbf{g}}\right)\mathbf{G}^{\dagger}\mathbf{G}^{\dagger T}\mathrm{diag}\left(\mathbf{J}_{\mathbf{g}}{}^{T}\mathbf{P}_{\mathbf{G}}{}^{\perp}\mathbf{h}\right)}_{\approx 0}\Big]_{m,n}. && (5.84)
\end{aligned}
$$

The second under-braced summand in (5.84) contains the residual term

$$\mathbf{P}_{\mathbf{G}}{}^{\perp}\hat{\mathbf{h}} = \hat{\mathbf{h}} - \mathbf{G}\mathbf{G}^{\dagger}\hat{\mathbf{h}} = \hat{\mathbf{h}} - \mathbf{h}(\tilde{\boldsymbol{\tau}}), \tag{5.85}$$

which is zero-mean and becomes very small close to solution for $\boldsymbol{\tau}$. Hence it is easy to see that the second term is negligible, so that the approximated Hessian can further be approximated as

$$\mathbf{H}_{\Omega_{SNLLS}} = 2\mathrm{Re}\left[\mathrm{diag}\left(\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right)\mathbf{J}_{\mathbf{g}}^{H}\mathbf{P}_{\mathbf{G}}^{\perp}\mathbf{J}_{\mathbf{g}}\mathrm{diag}\left(\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right)\right]. \tag{5.86}$$

5.9 Alternative Approaches to SNLLS/ML

A large variety of approaches in the realm of multipath propagation parameter estimation has been studied over the past five decades, targeting the special needs of special problems, system requirements and hardware setups. Parameter estimation can be roughly categorized into algorithms that are employed in the frequency domain and such that can be employed in the time domain. Furthermore, they either can belong to the category acquisition or to the category of tracking like shown in Figure 5.9. Two sub-optimal approaches are discussed in the following.

5.9.1 Expectation Maximization (EM) and Space-Alternating Generalised Expectation Maximization Method (SAGE)

The SAGE algorithm [FH94] can be seen as an improved version of the EM algorithm proposed earlier by [DLR77]. The performance for JCAP was thoroughly investigated in [Sch12]. Consequently, I will only briefly summarize the underlying concept and advantages and disadvantages of the SAGE and the EM algorithm.

Like the LM algorithm, EM and SAGE yield local optimization results. Hence, as in the LM method they require an initial guess. Note that both EM and SAGE are designed to have the advantageous property that the likelihood increases, meaning that the cost

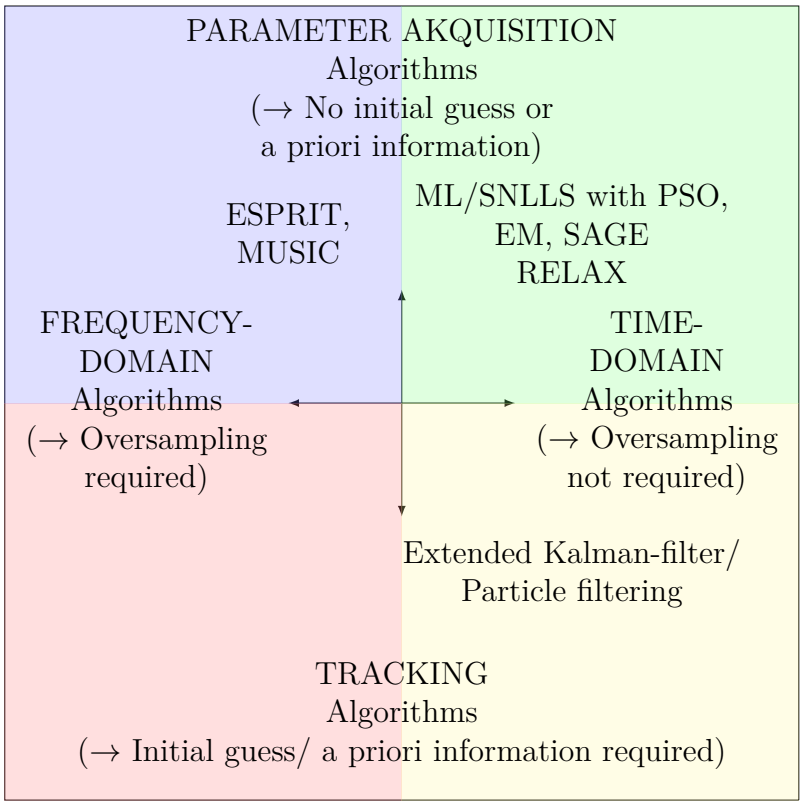


Figure 5.9: Different algorithms are suitable to solve the multipath parameter estimation problem as discussed in this thesis. ML/SNLLS with PSO does not require oversampling, or an initial guess, or a priori information, and at the same time finds the optimal solution for the specified problem.

function decreases, monotonically with each iteration, that is with each update of the parameter estimate. Consequently, it is guaranteed that the EM algorithm converges to a stationary point. The main advantage of both EM and SAGE algorithm over a local optimizer like LM can be explained by the underlying concept that allows to simplify the underlying non-linear C-dimensional optimization problem in the LM algorithm by splitting this problem into C 3-dimensional suboptimal subproblems via the model

$$\hat{\mathbf{h}} = \sum_{c=1}^{C} \hat{\mathbf{h}}_c \quad \text{with } \hat{\mathbf{h}}_c = \mathbf{h}_c + \mathbf{w}_c. \tag{5.87}$$

The components of the component-wise noise are assumed to be statistically independent, zero-mean Gaussian processes $w \in \mathcal{N}_{\mathbb{C}}(0, \iota_c \mathbf{C}_w)$. The introduced auxiliary value ι_c is constrained to $\sum_{c=1}^{C} \iota_c = 1$. In an iterative manner, each component is determined via the

expectation (E-step):

$$\hat{\mathbf{h}}_c^i = \mathbf{h}_c(\hat{\boldsymbol{\theta}}_c^i) + \mathbf{w}_c^i \tag{5.88}$$

$$= \mathbf{h}_c(\hat{\boldsymbol{\theta}}_c^i) + \iota_c^i \left(\hat{\mathbf{h}} - \sum_{c'=1}^{C} \mathbf{h}_{c'}(\hat{\boldsymbol{\theta}}_{c'}^n) \right), \tag{5.89}$$

and a new parameter update is reached via the maximization step (M-step):

$$\hat{\boldsymbol{\theta}}_c^{i+1} = \arg\min_{\tilde{\boldsymbol{\theta}}_c} \left\{ (\hat{\mathbf{h}}_c^i - \mathbf{h}_c(\tilde{\boldsymbol{\theta}}_c))^H \left(\iota_c^i \mathbf{C}_w \right)^{-1} (\hat{\mathbf{h}}_c^i - \mathbf{h}_c(\tilde{\boldsymbol{\theta}}_c)) \right\}. \tag{5.90}$$

By performing the updating step sequentially instead of parallelly in the SAGE algorithm, the convergence rate is improved compared to the EM algorithm. More specifically, each estimation result does not only depend on $\hat{\boldsymbol{\theta}}_c^n$, but also on the other estimated components. Moreover, note that the SAGE algorithm is designed to maximize the Fisher information, due to the fact that the estimation error is considered for each multipath component. Investigations carried out in [Sch12] indicate that the EM and SAGE algorithm are less sensitive to choosing an initial guess far from the optimum.

5.9.2 Estimation via Deconvolution and a Frequency-Domain Suboptimal Closed-Form Solution

Another approach to estimate the parameters is based on examining the channel coefficients in the frequency domain. This yields two advantages. The frequency domain channel coefficients can simply be divided by the DFT of sampled $g(\tau)$, to obtain a deconvolved time-domain signal. Secondly, the frequency-domain deconvolved signal exhibits an underlying rotational invariance among its subspaces. This can be exploited to find a set of linear equations, which, in return, yield a closed-form search-free solution for the delay estimates, called estimation of signal parameters via rotational invariance (ESPRIT) [RK89], [Saa97]. Hence, it is very attractive for the delay estimation in a JCAP framework.
In case $g(\tau)$ is bandlimited and sampled at or above the Nyquist rate

$$\boldsymbol{g} = [g(0) \quad g\left(\frac{T_c}{J}\right) \quad \dots \quad g\left((L+1-\frac{1}{J})T_c\right)],$$

the model function for $\bar{\boldsymbol{h}}$ can be expressed by a rank $M \times L_J$ Vandermonde matrix $\boldsymbol{S}$ as in [vdVVP98] and introduced in (4.12) as an approximation:

$$\breve{\boldsymbol{h}} = \mathrm{diag}\{\breve{\boldsymbol{g}}\}\boldsymbol{S}^T\boldsymbol{\beta}, \quad \text{with} \quad [\boldsymbol{S}]_{m,l} = e^{-2\pi j \tau_m l/((L+1)T)}. \tag{5.91}$$

Since a time domain convolution corresponds to a frequency-domain multiplication and the shifted Dirac impulses in $c(\tau)$ translate to rotations, solely $\boldsymbol{S}$ in (5.91) depends on the multipath delay information comprised in the cisoids $e^{-2\pi j\tau_m}$, whereas $\bar{\boldsymbol{g}}$ is independent

of $\boldsymbol{\tau}$. Hence, a deconvolution via Fourier transformation is applicable, and the deconvolved frequency-domain channel estimates can be written similar as introduced in 4.70 by including a selection matrix:

$$\breve{\boldsymbol{h}}^{\boldsymbol{d}} = \breve{\boldsymbol{h}}^{T} \boldsymbol{\mathcal{J}} (\mathrm{diag}\{\boldsymbol{\mathcal{J}} \breve{\boldsymbol{g}}\})^{-1} \quad \text{with} \quad \boldsymbol{\mathcal{J}} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{\lceil \frac{\mathbf{LW}}{\mathbf{2}} \rceil} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{\lfloor \frac{\mathbf{LW}}{\mathbf{2}} \rfloor} & \mathbf{0} \end{pmatrix},$$

chosen such that no division by zero occurs in the deconvolution process and the selected frequency components appear in increasing order. W denotes the selection window width and $J \geq W$. As suggested in [vdVVP98], we choose $W = 1$, since the window width is supposed to correspond to the signal main lobe. Deconvolving the channel estimates by $\boldsymbol{g}$ in (5.91) is a necessary signal preparation in order to proceed with the ESPRIT algorithm steps as will be summarized briefly.
The auxiliary construction of a rank $\varsigma \times L_{div}$ Hankel matrix $\boldsymbol{\Psi}$, by shifting and stacking sub-vectors of $\breve{\boldsymbol{h}}^{\boldsymbol{d}}$ with length ς, enables a robust channel correlation matrix estimation via a forward-backward approach [YLCC01, SM97]:

$$[\boldsymbol{\Psi}]_{n,l} = \breve{h}^{d}_{n+l-1}, \quad \hat{\boldsymbol{C}}_{\boldsymbol{\Psi}\boldsymbol{\Psi}} = \frac{1}{2L_{\mathrm{div}}} (\boldsymbol{\Psi}\boldsymbol{\Psi}^{H} + \boldsymbol{B}(\boldsymbol{\Psi}\boldsymbol{\Psi}^{H})^{*}\boldsymbol{B}^{H}).$$

Here, $L_{\mathrm{div}} = L - \varsigma - 1$ and $\boldsymbol{B}$ represents a matrix consisting of $1s$ on its anti-diagonal and $0s$ elsewhere.
Let $\varpi_i \in \{1...\varsigma\}$ be the eigenvalues of $\hat{\boldsymbol{C}}_{\boldsymbol{\Psi}\boldsymbol{\Psi}}$. Without loss of generality, we can assume $\varpi_1 \geq \varpi_2 \geq \ldots \geq \varpi_{\varsigma} = \sigma_w$. Here, σ_w denotes the channel coefficient variance. Let $\boldsymbol{U}$ be the matrix, whose columns consist of the corresponding eigenvectors. Two overlapping submatrices $\boldsymbol{U_s}$ and $\boldsymbol{U_d}$, related by a rotation matrix $\boldsymbol{\Phi}$, can be constructed by defining

$$\begin{array}{ll} \boldsymbol{U_s} & = [\boldsymbol{I}_{\nu-1} \quad \mathbf{0}_{(\nu-1)\times 1}]\boldsymbol{U} \\ \boldsymbol{U_d} & = [\mathbf{0}_{(\nu-1\times 1)} \quad \boldsymbol{I}_{\nu-1}]\boldsymbol{U} \end{array}, \qquad \boldsymbol{U_s} = \boldsymbol{U_d}\boldsymbol{\Phi}. \tag{5.92}$$

The frequency-domain physical channel inherent rotational invariance is exploited by formulating and solving the last equation in (5.92). The desired delays τ_c are related to the eigenvalues $\zeta_1, ..., \zeta_C$ of $\boldsymbol{\Phi}$ by $\mu_m = e^{-\frac{2\pi j \tau_m c}{(L+1)T}}$. Consequently, the delays can be estimated by calculating the eigenvalues $\zeta_1, ..., \zeta_C$ of the estimated least squares solution

$$\hat{\boldsymbol{\Phi}} = \boldsymbol{U_d}^{\dagger}\boldsymbol{U_s}, \tag{5.93}$$

such that

$$\hat{\tau}_c = -\frac{j(L+1)}{2\pi} \arg(\zeta_c) \quad \forall c \in \{1, \ldots, C\}. \tag{5.94}$$

A slight model mismatch has to be taken into account, since $\boldsymbol{g}$ has a finite length L_g. Consequently, the estimator is biased. This bias decreases whenever L_g or J are increased.
Proposing ESPRIT for delay estimation as an alternative to the ML solution exhibits a desirable diversity amongst the options with respect to complexity and performance in JCAP systems. All proposed algorithms do not depend on an initial guess. Yet, although

no iterative search is required for the ESPRIT-based method and a solution is guaranteed to be found, the computational cost for this closed-form solution comprises the cost for complex eigen-decomposition, calculation of a pseudoinverse, a matrix multiplication and DFT, not to forget the computational cost required by fractional sampling in order to decrease the modelling mismatch.

5.10 MIMO Channel Parameter Estimation

5.10.1 Delay Estimation

By extending the parameter definition of the SISO case to the SIMO case, such that in this chapter $\boldsymbol{\xi} = \{\boldsymbol{\varphi}, \boldsymbol{\phi}, \boldsymbol{\Gamma}^i, \boldsymbol{\Gamma}^q, \boldsymbol{\tau}\}$, it is straightforward to see that the MIMO parameter estimation problem can be extended to the joint estimation including the AOA $\boldsymbol{\varphi}$ and the AOD $\boldsymbol{\phi}$:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\min_{\tilde{\boldsymbol{\theta}}} ||\hat{\mathbf{H}} - \tilde{\mathbf{H}}(\tilde{\boldsymbol{\theta}})||_F^2, \\ &= \arg\min_{\tilde{\boldsymbol{\theta}}} ||\hat{\mathbf{H}} - \boldsymbol{\Gamma}\boldsymbol{\mathcal{G}}||_F^2. \end{aligned} \tag{5.95}$$

Whereas the reformulation of the extended parameter estimation to the SIMO case is as straightforward as it seems, let us begin with considering the case of pure delay estimation, assuming that the estimated delays provide sufficient information for an accurate position estimate. Hence, let us revise the simplified channel (4.52) model

$$\mathbf{H} = \boldsymbol{\Gamma}\boldsymbol{\mathcal{G}}.$$

Again, the concept of separability into linear and nonlinear parameters can be applied using that

$$\hat{\boldsymbol{\Gamma}} = \mathbf{H}\boldsymbol{\mathcal{G}}^T(\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}^T)^{-1}. \tag{5.96}$$

This formulation enables the formulation of a pure delay estimation problem, which does not depend on but allows to jointly estimating the angles of arrival. Hence, the delay estimation problem can be formulated via insertion of (5.96) in (5.95):

$$\hat{\boldsymbol{\tau}} = \arg\min_{\tilde{\boldsymbol{\tau}}} ||\hat{\mathbf{H}} - \hat{\mathbf{H}}\boldsymbol{\mathcal{G}}^\dagger\boldsymbol{\mathcal{G}}||_F^2. \tag{5.97}$$

Consider the SIMO case $N_T = 1$. Then, we can reformulate (5.97) in the following manner:

$$\hat{\boldsymbol{\tau}} = \arg\min_{\tilde{\boldsymbol{\tau}}} \sum_{v=1}^{N_R} ||\hat{\mathbf{h}}_v - \mathbf{G}\mathbf{G}^\dagger\hat{\mathbf{h}}_v||_F^2, \tag{5.98}$$

with $\hat{\mathbf{h}}_v = [h_{0,v}, \ldots, h_{L,v}]^T$. Further let us extend the expression to a weighted least squares problem by employing a weighting matrix $\mathbf{W}$ to emphasize the channel estimates

$$\hat{\boldsymbol{\tau}} = \arg\min_{\tilde{\boldsymbol{\tau}}} \sum_{v=1}^{N_R} \left(\hat{\mathbf{h}}_v - \hat{\mathbf{h}}_v\mathbf{G}\mathbf{G}^\dagger\right)^H \mathbf{W} \left(\hat{\mathbf{h}}_v - \hat{\mathbf{h}}_v\mathbf{G}\mathbf{G}^\dagger\right). \tag{5.99}$$

5.10.2 Spatial Signature Matrix Estimation

It is straightforward to see that the spatial signature matrix can be estimated by

$$\hat{\mathbf{\Gamma}} = \mathbf{H}(\boldsymbol{\mathcal{G}}(\hat{\boldsymbol{\tau}}))^{\dagger}. \tag{5.100}$$

5.10.3 Angle of Departure and Angle of Arrival Estimation

Already ten years ago, angle estimation was called a mature research area [Mia07]. Therefore I restrict myself to summarizing methods suitable for this framework and that have an impact on the further results in this thesis. Suitable are high resolution methods, rolling out the classical beam-forming methods [Cap69], known to be less accurate than ML estimators. Further I do not discuss the well-known and also well-suited sub-optimal subspace-based solutions based on methods like the MUSIC [Sch86], root-MUSIC [RH89], the ESPRIT [RK89] algorithm and improved variants thereof [HN95]. For more comprehensive tutorial-like introductions into the topic refer to [KV96, Mia07] or [Rib08]. Let us begin with assessing the SIMO case ($N_T = 1$). In order to do that, revisit the definition of the spatial signature matrix described by (4.40,4.41,4.43) and the dependence to the angular vector $[\boldsymbol{\varphi}_c]_n = \varphi_{c,n}$ and $[\boldsymbol{\phi}_c]_n = \phi_{c,n}$ given by

$$\begin{aligned}\boldsymbol{\gamma}_c =& \mathbf{A}_{\mathbf{R}c}(\boldsymbol{\phi}_c)\mathrm{diag}(\mathbf{f}_c)(\mathbf{A}_{\mathbf{T}c}(\boldsymbol{\varphi}_c))^T \\ =& \sum_{n=1}^{N_{Rays}} f_{c,n}\boldsymbol{a}_{\boldsymbol{R}c}(\phi_{c,n}) \quad \text{for } c = 1,\ldots,C.\end{aligned} \tag{5.101}$$

At first consider the case in which all angular spreads are assumed to be equal to zero and hence the steering matrices $\mathbf{A}_{\mathbf{T}c}(\varphi)$ and $\mathbf{A}_{\mathbf{R}c}(\phi)$ can be replaced by vectors $\mathbf{a}_{\mathbf{T}c}(\varphi)$ and $\mathbf{a}_{\mathbf{R}c}(\phi)$, not depending on a specific ray number. In case the angular spreads are assumed to be zero, that is, assuming that $\varphi_{c,n} = \varphi_c$ and $\phi_{c,n} = \phi_c$ and hence $\boldsymbol{\varphi} = [\varphi_1 \ldots \varphi_C]^T$ and $\boldsymbol{\phi} = [\phi_1 \ldots \phi_C]^T$, the following simplification applies here:

$$\boldsymbol{\gamma}_c = \mathbf{a}_{\mathbf{R}c}(\boldsymbol{\phi}) \underbrace{\sum_{n=1}^{N_{Rays}} e^{2\pi j(\theta_{c,n}+f_{D_c})}}_{\beta_c} = \mathbf{a}_{\mathbf{R}c}\beta_c \quad \text{for } c = 1,\ldots,C. \tag{5.102}$$

Given this simplified model and employing the concept of separability of linear and nonlinear estimation components as in the previous sections on delay estimation, it is straightforward to see that an ML type angle of arrival estimator for the vector $\boldsymbol{\phi}_c$ can be formulated as a one-dimensional optimization problem

$$\begin{aligned}\hat{\phi}_c =& \arg\min_{\tilde{\phi}_c} ||\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}_c(\tilde{\phi}_c)||^2 \\ =& \arg\min_{\tilde{\phi}_c} ||\hat{\boldsymbol{\gamma}} - \beta_c\tilde{\mathbf{a}}_{\mathbf{R}c}(\tilde{\phi}_c)||^2 \\ =& \arg\min_{\tilde{\phi}_c} \hat{\boldsymbol{\gamma}}_c^H(\mathbf{I}_{N_R} - \mathbf{a}_{\mathbf{R}c}(\tilde{\boldsymbol{\phi}}_c)\mathbf{a}_{\mathbf{R}c}^{\dagger}(\tilde{\phi}_c))\hat{\boldsymbol{\gamma}}_c.\end{aligned} \tag{5.103}$$

Next, consider the case that the angular spread $\sigma_\alpha = \alpha_n \Delta$ is unequal to zero. Assuming the direction of arrival angular spread is unequal to zero, the number of rays N_{Rays} has to either be known or to be estimated. In matrix vector notation, the underlying system equation for angular estimation then is given by

$$\boldsymbol{\gamma}_c = \mathbf{A}_{\mathbf{Rc}} \mathbf{f}_{\mathbf{c}}. \tag{5.104}$$

Then, it is also straightforward, assuming that a sufficient amount of observations is available, that the resulting estimator for the vector $\boldsymbol{\phi}_c$ corresponds to

$$\begin{aligned} \hat{\boldsymbol{\phi}}_c &= \arg\min_{\tilde{\boldsymbol{\phi}}_c} ||\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}_c(\tilde{\boldsymbol{\phi}}_c)||^2 \\ &= \arg\min_{\tilde{\boldsymbol{\phi}}_c} \hat{\boldsymbol{\gamma}}_c^H (\mathbf{I}_{N_R} - \mathbf{A}_{\mathbf{R}c}(\tilde{\boldsymbol{\phi}}_c) \mathbf{A}_{\mathbf{R}_c}^{\dagger}(\tilde{\boldsymbol{\phi}}_c)) \hat{\boldsymbol{\gamma}}_c. \end{aligned} \tag{5.105}$$

Obviously, (5.103) is covered by (5.105), being a special one-dimensional case thereof. These estimators have been proposed in [Lee98] as useful direction of arrival methods in case the source signals are scattered.

Nevertheless, strictly following the channel modeling construction rules provided earlier in this thesis, one can conclude that either a huge number of receive antennas is necessary to identify the parameters in question or the involved number of rays in the channel model must be comparably small to yield a higher number of equations then elements of the desired parameter vector. Rayleigh and Rician fading channel models are usually obtained by superimposing only 20 rays, since due to the central limit theorem this number, for many cases, yields a sufficient approximative channel. In reality though, an infinite number of rays is actually contributing. Taking this into account, the parametrization into a finite number of ray angles, to be estimated, is not reasonable and another parameterization, closer to reality, regarding channel modeling, should be taken into account. Further, recall that for the purpose of positioning we are only interested in estimating ϕ_{LOS}. The fact that we are not actually interested in estimating every single ray angle $\phi_{c,n}$ and we know from estimation theory, that the more parameters we estimate, the more accuracy we compromise, is another reason to find the most simple parameterization satisfying our needs. Fortunately, such alternative parametrization are provided by [TO96]. The parameterization proposed in [TO96] takes into account that local scattering leads to a non-zero angular spread and is based on the assumption that multiple measurements, that is, a whole time series of measurements, is available. Consequently, a system equation based on correlation matrices therefore can be employed. Further it is based on the assumption that the distribution of rays around the nominal direction is known and can be approximated by a normal distribution $\mathcal{N}(0, \sigma_\alpha)$. Based on that setup the authors of [TO96] suggest to estimate the parameters specified by the parametrization, given by a set containing the ϕ_{LOS}, the angular spread σ_α^2 and the signal power S. If $\sigma_\alpha = \alpha_n \Delta$ is small, approximating $\sin(\Delta\alpha_n)$ and $\cos(\Delta\alpha_n)$ via only employing the Taylor series expansion's first terms, it is easy to see that

$$\sin(\alpha_n \Delta) = \alpha_n \Delta \quad \text{and} \quad \cos(\alpha_n \Delta) = 1. \tag{5.106}$$

Consequently, this leads to the following approximation:

$$\sin(\phi_c + \Delta\alpha_n) = \sin(\phi_c)\cos(\Delta\alpha_n) + \cos(\phi_c)\sin(\Delta\alpha_n) \underbrace{=}_{\alpha_n\Delta\to 0} \sin(\phi_c) + \cos(\phi_c)\alpha_n\Delta. \tag{5.107}$$

Reviewing [TO96] it can be understood, that based on this approximation, the underlying approximate system model that is further employed for parameter estimation is given by

$$\mathbf{C}_{\boldsymbol{\gamma_1}} \approx S\mathbf{a}_{Rx}\mathbf{a}_{Rx}^H \odot \mathbf{B} + \sigma_{\boldsymbol{\gamma}_1}^2\mathbf{I}. \tag{5.108}$$

The auxiliary matrix $\mathbf{B}$ is constructed by

$$[\mathbf{B}]_{k,l} = e^{-2[pi\Delta(k-l)]^2\sigma_\alpha^2\cos^2(\phi)}. \tag{5.109}$$

Then the full ML estimator of the parameter vector stacking $\boldsymbol{\eta} = [S, \sigma_{\boldsymbol{\gamma}_1}, \phi, \sigma_\alpha]$ is obtained by minimizing the negative likelihood given by

$$\mathcal{L}(S, \sigma_{\boldsymbol{\gamma}_1}, \phi, \sigma_\alpha) = \log|\mathbf{C}_{\boldsymbol{\gamma}_1}| + \text{trace}\{\mathbf{C}_{\boldsymbol{\gamma}_1}^{-1}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}\}. \tag{5.110}$$

Thereby, $\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}$ denotes the sample correlation matrix of γ_1 yielded by the time series measurements

$$\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1} = \frac{1}{I}\sum_{i=0}^{I}\boldsymbol{\gamma}_1(t_i)\boldsymbol{\gamma}_1^*(t_i). \tag{5.111}$$

Alternatively, aiming to find less complex solutions that approximately yield the ML performance as shown in [TO95, TO96], a least squares and weighted least squares estimator can be formulated. Fortunately, since S and $\sigma_{\boldsymbol{\gamma}_1}$ are quadratic, they can be calculated in closed form as a function of the remaining two parameters in $\boldsymbol{\eta}$ and consequently the parameter estimation problem is separable in resulting in a two-dimensional instead of four-dimensional optimization problem with a weighted least squares cost function

$$\hat{S} = \frac{\text{trace}\{\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1}\}\text{trace}\{\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-2}\} - \text{trace}\{\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1}\}\text{trace}\{\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-2}\mathbf{C}\}}{\text{trace}\{\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1}\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1}\}\text{trace}\{\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-2}\} - \text{trace}\{\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-2}\}\text{trace}\{\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-2}\}}, \tag{5.112}$$

$$\hat{\sigma}_{\boldsymbol{\gamma}_1} = \frac{\text{trace}\{\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1}\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1}\}\text{trace}\{\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1}\} - \text{trace}\{\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-2}\mathbf{C}\}\text{trace}\{\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1}\}}{\text{trace}\{\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1}\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1}\}\text{trace}\{\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-2}\} - \text{trace}\{\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-2}\}\text{trace}\{\mathbf{C}\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-2}\}}, \tag{5.113}$$

$$[\hat{\phi}, \hat{\sigma}_\alpha] = \arg\min_{\tilde{\phi},\tilde{\sigma}_\alpha^2} \text{trace}\{((\hat{S}\mathbf{C}(\tilde{\phi}, \tilde{\sigma_\alpha^2}) + \hat{\sigma}_{\boldsymbol{\gamma}_1}^2\mathbf{I})\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}^{-1} - \mathbf{I})^2\}. \tag{5.114}$$

Less complex and also two-dimensional is the pure least squares estimator

$$\hat{S} = \frac{\text{trace}\{\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}\mathbf{C}\} - \text{trace}\{\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1}\}}{\text{trace}\{\mathbf{C}^2\} - N_r}, \tag{5.115}$$

$$\hat{\sigma}_{\boldsymbol{\gamma}_1} = \frac{1}{N_r}\text{trace}\{\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1} - \hat{S}\}, \tag{5.116}$$

$$[\hat{\phi}, \hat{\sigma}_\alpha] = \arg\min_{\tilde{\phi},\tilde{\sigma}_\alpha^2} \text{trace}\{(\hat{\mathbf{C}}_{\boldsymbol{\gamma}_1} - \hat{S}\mathbf{C}(\tilde{\phi}, \tilde{\sigma_\alpha^2}) - \hat{\sigma}_{\boldsymbol{\gamma}_1}^2\mathbf{I})^2\}. \tag{5.117}$$

Although both proposed least squares estimators are not efficient, applying them as an alternative to the full ML estimator is nevertheless worth consideration due to the complexity reduction [TO95].

In [TO96] the authors propose to employ a Newton-type search algorithm for optimization and indicate that the algorithm then has to be initialized carefully to converge to the global optimum. A comprehensive overview about this and other similar parameterization-based approximations is provided in [Rib08]. The authors of [AOS97] and [AO99] show that local scattering has a huge impact on the direction estimation accuracy and consequently also propose to replace the parametrization by making use of a spatial signature matrix approximation. Furthermore they investigate how local scattering effects direction estimation with MUSIC.

5.11 Time Series Extension

Employing a time series of measurements instead of a single-shot measurement $\hat{\mathbf{h}}$ for parameter estimation is suggested due to more than one reason: For starters, the number of snapshot measurements is increased by the factor I. Apart from that, time series measurements allow for a deterministic as well as a stochastic parametrization and ML (DML and SML) parameter estimation approach. Depending on the applied approach the problem dimensionality varies. Recalling the extended channel model construction for the time series measurement case, we can construct a compact matrix vector notation via defining a matrix $\mathbf{H}_{TS}$ and $\mathbf{W}$ concatenating the column vectors $\mathbf{h}(i), \quad \forall i \in \{0, \ldots, I\}$ horizontally to a size $(L+1) \times I$ matrix such that

$$\mathbf{H}_{TS} = \mathbf{G}(\boldsymbol{\tau})\boldsymbol{\Gamma} + \mathbf{W}. \tag{5.118}$$

Equation (5.118) alternatively can be vectorized[2], such that we obtain a potentially long time series extension vector similar to (4.58):

$$\begin{aligned} \mathbf{h}_{TS} =& \mathrm{vec}(\mathbf{H}_{TS}) = \underbrace{(\mathbf{I}_I \otimes \mathbf{G}(\boldsymbol{\tau}))}_{\boldsymbol{\mathcal{G}}_{TS}} \underbrace{\mathrm{vec}(\boldsymbol{\Gamma})}_{\boldsymbol{\gamma}_{TS}} + \underbrace{\mathrm{vec}(\mathbf{W})}_{\mathbf{w}_{TS}} \\ =& \boldsymbol{\mathcal{G}}_{TS}\boldsymbol{\gamma}_{TS} + \mathbf{w}_{TS}. \end{aligned} \tag{5.119}$$

For (5.119) it seems more intuitive than it might be from looking at (5.118) how to formulate the corresponding PDF $p(\hat{\mathbf{h}}|\boldsymbol{\theta})$ for parameter estimation later.

5.11.1 Deterministic Parametrization and ML Parameter Estimation

Now let us treat the N_r complex amplitudes $\boldsymbol{\gamma}$ as deterministic magnitudes. That more precisely means that we make no underlying assumptions about the complex path amplitudes possibly being some kind of stochastic variables. Then, taking into account that

[2] For vectorization of the multiplication of three arbitrary matrices $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ holds true, as before. Further, for any two arbitrary matrices $\mathrm{vec}(\mathbf{A}+\mathbf{B}) = \mathrm{vec}(\mathbf{A}) + \mathrm{vec}(\mathbf{B})$.

the measurements are fully and the unknowns are partially complex, we obtain a set of $2N_rI(L+1)$ equations with $2N_rIC + C$ unknowns. The latter value corresponds to the length of the vector

$$\boldsymbol{\theta}_{dml} = \{\boldsymbol{\tau}, \mathrm{Re}(\mathrm{vec}(\boldsymbol{\Gamma}), \mathrm{vec}(\mathrm{Im}(\boldsymbol{\Gamma}_{\boldsymbol{v}})))\} \tag{5.120}$$

associated with the underlying parametrization. By choosing this deterministic parametrization, we aim to estimate each of the I complex amplitudes. In this case, the ML approach can be extended to

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{DML} &= \arg\min_{\tilde{\boldsymbol{\theta}}_{DML}} \{-\ln(p(\hat{\mathbf{h}}(0), \ldots, \hat{\mathbf{h}}(I-1)|\tilde{\boldsymbol{\theta}}_{DML}))\} \\ &= \arg\min_{\tilde{\boldsymbol{\theta}}_{DML}} \{-\ln(p(\hat{\mathbf{h}}_{TS}|\tilde{\boldsymbol{\theta}}_{DML}))\}. \end{aligned} \tag{5.121}$$

Inserting the PDF into (5.121) and simplifying the equation as far as possible, or directly employing (B.1.13), it is straightforward to see that the ML/LS problem solution for the single measurement case can be extended by a summation over all I individual least squares errors. Further, as before, substituting the linear unknowns by their closed-form solution leads to the reduced dimensionality time series deterministic ML or LS solution

$$\hat{\boldsymbol{\tau}} = \arg\min_{\tilde{\boldsymbol{\tau}}} \left(\hat{\mathbf{h}}_{TS} - \boldsymbol{\mathcal{G}}_{TS}(\tilde{\boldsymbol{\tau}})(\boldsymbol{\mathcal{G}}_{TS}(\tilde{\boldsymbol{\tau}}))^{\dagger}\hat{\mathbf{h}}_{TS}\right)^{H} \left(\hat{\mathbf{h}}_{TS} - \boldsymbol{\mathcal{G}}(\tilde{\boldsymbol{\tau}})(\boldsymbol{\mathcal{G}}_{TS}(\tilde{\boldsymbol{\tau}}))^{\dagger}\hat{\mathbf{h}}_{TS}\right). \tag{5.122}$$

Often this kind of deterministic ML approach is also referred to as the conditional ML [OVSN93]. The cost function in (5.122) can alternatively be replaced by a concentrated form leading to

$$\hat{\boldsymbol{\tau}} = \arg\min_{\tilde{\boldsymbol{\tau}}} \{\mathrm{trace}(\mathbf{P}_{\mathbf{G}}^{\perp}\hat{\mathbf{C}}_{\hat{\mathbf{h}}})\}, \tag{5.123}$$

where the matrix $\hat{\mathbf{C}}_{\hat{\mathbf{h}}}$ is the estimated so called sample covariance matrix

$$\hat{\mathbf{C}}_{\hat{\mathbf{h}}} = \frac{1}{I}\sum_{i=0}^{I} \hat{\mathbf{h}}(i)(\hat{\mathbf{h}}(i))^{H}. \tag{5.124}$$

A derivation is provided in the Appendix A.2.

5.11.2 Stochastic Parametrization

Alternatively, exploiting the large amount of available observations, a stochastic or unconditional approach may be formulated as well, based on interpreting the complex path amplitudes as stochastic variables. Instead of indirectly employing a large number of estimates for each complex path amplitude $\hat{\boldsymbol{\beta}}_i$ for all $i \in \{0, \ldots, I-1\}$, like it is done in the deterministic approach. This enables a problem formulation based on using parts of the channel statistics, that is, the unknown amplitudes and the known or estimated channel estimation covariance matrix, $\mathbf{C}_{\boldsymbol{\beta}}$ and $\mathbf{C}_{\hat{\mathbf{h}}}$, respectively. More precisely, that means we use apriori knowledge, like knowing that the complex path amplitudes are a stochastic variable

being drawn out of a known distribution. The most popular and thoroughly documented case is, that the complex path amplitudes are zero mean complex Gaussian-distributed, and also independent and identical (IID) distributed. The latter means, that over time, the measured snapshots are distributed with the same probability distribution and further they are mutually independent. To formulate a stochastic Maximum Likelihood (SML) problem, we aim to build an equation system based on the channel covariance matrix. The covariance is defined as

$$\mathbf{C}_{\hat{\mathbf{h}}} = \mathrm{E}\{(\hat{\mathbf{h}} - \mathrm{E}\{\hat{\mathbf{h}}\})(\hat{\mathbf{h}} - \mathrm{E}\{\hat{\mathbf{h}}\})^H\} \tag{5.125}$$

in general. If the complex path amplitudes have zero mean, (5.125) simplifies to $\mathbf{C}_{\hat{\mathbf{h}}} = \mathrm{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\}$. To begin with let us assume, that the complex path amplitudes under investigations, may be modelled as circular complex Gaussian signals. If I becomes large, the statistical average represents the expected value, that is

$$\mathrm{E}\{(\hat{\mathbf{h}} - \mathrm{E}\{\hat{\mathbf{h}}\})(\hat{\mathbf{h}} - \mathrm{E}\{\hat{\mathbf{h}}\})^H\} = \lim_{I\to\infty} \frac{1}{I} \sum_{i=0}^{I} (\hat{\mathbf{h}}(i) - \mathbf{m}_{\hat{\mathbf{h}}})(\hat{\mathbf{h}}(i) - \mathbf{m}_{\hat{\mathbf{h}}})^H. \tag{5.126}$$

Thereby $\mathbf{m}_{\hat{\mathbf{h}}}$ is the mean of $\hat{\mathbf{h}}$. Hence (5.2) holds true as well. As can be found in [PP12] given any constant arbitrary matrix $\mathbf{A}$ and any other random vector $\mathbf{x}$ the following property is fulfilled:

$$\mathrm{Var}\{\mathbf{Ax}\} = \mathbf{A}\mathrm{Var}\{\mathbf{x}\}\mathbf{A}^T. \tag{5.127}$$

It is further stated in [PP12] that given any arbitrary matrices $\mathbf{A}$ and $\mathbf{B}$, any linear combination of two Gaussian-distributed random vectors $\mathbf{x} \sim \mathcal{N}(\mathbf{m}_\mathbf{x}, \mathbf{C}_\mathbf{x})$, $\mathbf{y} \sim \mathcal{N}(\mathbf{m}_\mathbf{y}, \mathbf{C}_\mathbf{y})$ with means $\mathbf{m}_\mathbf{x}$, $\mathbf{m}_\mathbf{y}$ and covariance matrices $\mathbf{C}_\mathbf{x}, \mathbf{C}_\mathbf{y}$ will again be Gaussian-distributed, that is

$$\mathbf{Ax} + \mathbf{By} + \mathbf{c} \sim \mathcal{N}(\mathbf{Am}_\mathbf{x} + \mathbf{Bm}_\mathbf{y} + \mathbf{c}, \mathbf{AC}_\mathbf{x}\mathbf{A}^T + \mathbf{BC}_\mathbf{y}\mathbf{B}^T). \tag{5.128}$$

With (5.127) and (5.128) it is easy to see that the covariance matrix of the channel estimates $\hat{\mathbf{h}} = \mathbf{G}\boldsymbol{\beta} + \mathbf{w}$ is determined by

$$\mathrm{E}\{\hat{\mathbf{h}}(i)\hat{\mathbf{h}}^H(i)\} = \mathbf{G}\mathbf{R}_{\boldsymbol{\beta}}\mathbf{G}^T + \mathbf{C}_\mathbf{w}. \tag{5.129}$$

Let us furthermore assume that the amplitudes $\boldsymbol{\beta}(t)$ are zero-mean, that is, they are temporally and spatially uncorrelated.

5.11.3 Zero-Mean Fading

Note that the zero-mean assumption, in combination with the assumption that the considered signals are Gaussian-distributed, is a typical assumption in the whole theory that has evolved around the field of sensor array processing. In our special parameter estimation problem at hand, the combination of both assumptions is reasonable, when the complex path amplitudes for $c\epsilon\{1, \ldots, C\}$ are assumed to be Rayleigh fading amplitudes.

If Rice fading applies for $c = 1$, this assumption is partially violated. The ML estimator corresponding to the scenario $\boldsymbol{\beta}^i, \boldsymbol{\beta}^q \sim \mathcal{N}(0, \mathbf{I})$ can be formulated as

$$\hat{\tau} = \arg\max_{\tilde{\boldsymbol{\tau}}} \{p(\hat{\mathbf{h}}(1), \ldots, \hat{\mathbf{h}}(I) | \mathbf{R}_{\boldsymbol{\beta}} \mathbf{C}_{\mathbf{w}}, \boldsymbol{\tau})\}. \tag{5.130}$$

In the SISO case the joint complex multivariate Gaussian PDF is determined via a product of the Gaussian PDF for the snapshot measurements

$$p(\hat{\mathbf{h}}(1), \ldots, \hat{\mathbf{h}}(I) | \mathbf{R}_{\boldsymbol{\beta}} \mathbf{C}_{\mathbf{w}}, \boldsymbol{\tau}) = \prod_{i=1}^{I} \frac{1}{|\pi^{L+1} \mathbf{C}_{\hat{\mathbf{h}}}|} e^{-\hat{\mathbf{h}}^H(i) \mathbf{C}_{\hat{\mathbf{h}}}^{-1} \hat{\mathbf{h}}(i)}. \tag{5.131}$$

In the SIMO case it reads,

$$p(\hat{\mathbf{H}}(1), \cdots, \hat{\mathbf{H}}(I) | \mathbf{R}_{\boldsymbol{\beta}} \mathbf{C}_{\mathbf{w}}, \boldsymbol{\tau}) = \prod_{v=1}^{N_r} \prod_{i=1}^{I} \frac{1}{|\pi^{L+1} \mathbf{C}_{\hat{\mathbf{h}}}|} e^{-\hat{\mathbf{h}}_v^H(i) \mathbf{C}_{\hat{\mathbf{h}}}^{-1} \hat{\mathbf{h}}_v(i)}. \tag{5.132}$$

Let the so-called sample covariance matrix be the estimate

$$\hat{\mathbf{C}}_{\hat{\mathbf{h}}} = \frac{1}{I N_r} \sum_{v=0}^{N_r} \sum_{i=1}^{I} \hat{\mathbf{h}}_v(i) \hat{\mathbf{h}}_v^H(i). \tag{5.133}$$

Since the snapshot measurements are temporally independently identically distributed, via some well known standard maximum likelihood manipulations, like taking the (monotonic) logarithm and minimizing the negative PDF, the likelihood can be defined and simplified as follows, by inserting (5.131) in (5.130):

$$\begin{aligned}
&\arg\max_{\tilde{\boldsymbol{\tau}}, \tilde{\mathbf{C}}_{\boldsymbol{\beta}}} \left\{ \prod_{v=1}^{N_r} \prod_{i=1}^{I} \frac{1}{|\pi^{L+1} \mathbf{C}_{\hat{\mathbf{h}}}|} e^{-\hat{\mathbf{h}}^H(i) \mathbf{C}_{\hat{\mathbf{h}}}^{-1} \hat{\mathbf{h}}(i)} \right\} \\
&= \arg\min_{\tilde{\boldsymbol{\tau}}, \tilde{\mathbf{C}}_{\boldsymbol{\beta}}} \left\{ -\sum_{v=1}^{N_r} \sum_{i=1}^{I} \ln \left\{ \frac{1}{|\pi^{L+1} \mathbf{C}_{\hat{\mathbf{h}}}|} e^{-\hat{\mathbf{h}}^H(i) \mathbf{C}_{\hat{\mathbf{h}}}^{-1} \hat{\mathbf{h}}(i)} \right\} \right\} \\
&= \arg\min_{\tilde{\boldsymbol{\tau}}, \tilde{\mathbf{C}}_{\boldsymbol{\beta}}} \left\{ I N_r (L+1) \ln\{|\pi\} + I N_r \ln\{|\mathbf{C}_{\hat{\mathbf{h}}}|\} + \sum_{v=1}^{N_r} \sum_{i=1}^{I} \hat{\mathbf{h}}^H(i) \mathbf{C}_{\hat{\mathbf{h}}}^{-1} \hat{\mathbf{h}}(i) \right\} \\
&= \arg\min_{\tilde{\boldsymbol{\tau}}, \tilde{\mathbf{C}}_{\boldsymbol{\beta}}} \left\{ N_r I \ln\{|\mathbf{C}_{\hat{\mathbf{h}}}|\} + \frac{N_r I}{N_r I} \sum_{v=1}^{N_r} \sum_{i=1}^{I} \hat{\mathbf{h}}^H(i) \mathbf{C}_{\hat{\mathbf{h}}}^{-1} \hat{\mathbf{h}}(i) \right\} \\
&= \arg\min_{\tilde{\boldsymbol{\tau}}, \tilde{\mathbf{C}}_{\boldsymbol{\beta}}} \left\{ N_r I \ln|\mathbf{C}_{\hat{\mathbf{h}}}| + N_r I \mathrm{trace}\{\mathbf{C}_{\hat{\mathbf{h}}}^{-1} \hat{\mathbf{C}}_{\hat{\mathbf{h}}}\} \right\}.
\end{aligned} \tag{5.134}$$

Note that the following equalities are used in this thesis:

$$\frac{1}{N_r I} \sum_{v=1}^{N_r} \sum_{i=1}^{I} \hat{\mathbf{h}}^H(i) \mathbf{C}_{\hat{\mathbf{h}}}^{-1} \hat{\mathbf{h}}(i) = \mathrm{trace}\{\mathbf{C}_{\hat{\mathbf{h}}}^{-1} \hat{\mathbf{C}}_{\hat{\mathbf{h}}}\}, \tag{5.135}$$

and in case $N_r I$ is large

$$\text{trace}\{\mathbf{C}_{\hat{\mathbf{h}}}^{-1}\hat{\mathbf{C}}_{\hat{\mathbf{h}}}\} = L + 1. \tag{5.136}$$

To maintain compact and readable equations, I omitted the dependencies on the minimizing arguments in (5.134). Further note that although the equations in (5.134) seem not to be directly related to the minimizing arguments $\tilde{\boldsymbol{\tau}}, \tilde{\mathbf{C}}_{\boldsymbol{\beta}}$, by reviewing the structure of $\mathbf{C}_{\hat{\mathbf{h}}}$ it becomes clear that indirectly the equations are related to the minimizing arguments via the relationship to $\mathbf{C}_{\hat{\mathbf{h}}}$ and the delay-dependent matrix $\mathbf{G}$:

$$\mathbf{C}_{\hat{\mathbf{h}}} = \mathbf{G}\mathbf{C}_{\boldsymbol{\beta}}\mathbf{G}^H + \sigma_{\mathbf{w}}^2\mathbf{I}. \tag{5.137}$$

Momentarily treating $\boldsymbol{\tau}$ as a known constant vector, obviously, the signal covariance matrix estimate can be obtained in closed form by

$$\hat{\mathbf{C}}_{\boldsymbol{\beta}} = \mathbf{G}^{\dagger}(\hat{\mathbf{C}}_{\hat{\mathbf{h}}} - \sigma_{\mathbf{w}}^2\mathbf{I})\mathbf{G}^{\dagger H}. \tag{5.138}$$

In this approach, $\mathbf{C}_{\boldsymbol{\beta}}$ and $\boldsymbol{\tau}$ are unknown and hence a $C^2 + C$-dimensional problem has to be solved. An alternative solution, for which the dependence on $\mathbf{C}_{\boldsymbol{\beta}}$ is eliminated and hence the problem is reduced to a C-dimensional problem, was proposed in [Jaf88] such that the resulting separable estimation probem, formulated with proof originally for the purpose of direction finding in array processing, can be translated here to the problem of delay estimation in the following manner:

$$\boldsymbol{\tau} = \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ -\frac{1}{\sigma_{\mathbf{w}}^2}\text{Trace}\{\mathbf{P}_{\mathbf{G}}\hat{\mathbf{C}}_{\hat{\mathbf{h}}}\} + \ln|\mathbf{P}_{\mathbf{G}}\hat{\mathbf{C}}_{\hat{\mathbf{h}}}\mathbf{P}_{\mathbf{G}} + \sigma_{\mathbf{w}}^2(\mathbf{I} - \mathbf{P}_{\mathbf{G}})| \right\}. \tag{5.139}$$

Thereby, $\mathbf{P}_{\mathbf{G}} = \mathbf{G}\mathbf{G}^{\dagger}$ denotes the orthogonal projection matrix. For a detailed derivation of (5.139) refer to the full proof provided in [Jaf88]. Inserting (5.137) in (5.134) it also can be understood that a so-called concentrated form of the minimization problem can be formulated as

$$\boldsymbol{\tau} = \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \ln|\mathbf{G}\hat{\mathbf{C}}_{\boldsymbol{\beta}}\mathbf{G}^H + \sigma_{\mathbf{w}}^2\mathbf{I}| \right\}. \tag{5.140}$$

It is shown below that a simplified cost function with a reduced complexity is given by

$$\boldsymbol{\tau} = \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \ln(\sigma_{\mathbf{w}}^{2((L+1)-C)}|\mathbf{G}^{\dagger}\hat{\mathbf{C}}_{\hat{\mathbf{h}}}\mathbf{G}|) \right\}. \tag{5.141}$$

Inserting (5.138) into (5.140) yields s simplfied cost function:

$$\begin{aligned}
\boldsymbol{\tau} &= \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \ln|\mathbf{G}\hat{\mathbf{C}}_{\boldsymbol{\beta}}\mathbf{G}^H + \sigma_{\mathbf{w}}^2\mathbf{I}| \right\} \\
&= \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \ln|\hat{\mathbf{C}}_{\boldsymbol{\beta}}\mathbf{G}^H\mathbf{G} + \sigma_{\mathbf{w}}^2\mathbf{I}| \right\} \\
&= \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \ln(\sigma_{\mathbf{w}}^{2(L+1)}|\sigma_{\mathbf{w}}^{-2}\hat{\mathbf{C}}_{\boldsymbol{\beta}}\mathbf{G}^H\mathbf{G} + \mathbf{I}|) \right\} \\
&= \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \ln(\sigma_{\mathbf{w}}^{2(L+1)}|\sigma_{\mathbf{w}}^{-2}\mathbf{G}^\dagger(\hat{\mathbf{C}}_{\hat{\mathbf{h}}} - \sigma_{\mathbf{w}}^2\mathbf{I})\underbrace{\mathbf{G}^{\dagger H}\mathbf{G}^H\mathbf{G}}_{\mathbf{G}} + \mathbf{I}|) \right\} \\
&= \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \ln(\sigma_{\mathbf{w}}^{2(L+1)}|\sigma_{\mathbf{w}}^{-2}\sigma_{\mathbf{w}}^2\mathbf{G}^\dagger(\sigma_{\mathbf{w}}^{-2}\hat{\mathbf{C}}_{\hat{\mathbf{h}}} - \mathbf{I})\mathbf{G} + \mathbf{I}|) \right\} \\
&= \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \ln(\sigma_{\mathbf{w}}^{2(L+1)}|\mathbf{G}^\dagger\sigma_{\mathbf{w}}^{-2}\hat{\mathbf{C}}_{\hat{\mathbf{h}}}\mathbf{G} - \underbrace{\mathbf{G}^\dagger\mathbf{G}}_{\mathbf{I}} + \mathbf{I}|) \right\} \\
&= \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \ln(\sigma_{\mathbf{w}}^{2((L+1)-C)}|\mathbf{G}^\dagger\hat{\mathbf{C}}_{\hat{\mathbf{h}}}\mathbf{G}|) \right\}. \qquad (5.142)
\end{aligned}$$

Here, I used the following two determinant properties:

Property 4. *Given any matrices* $\mathbf{A}, \mathbf{B}$ *it holds true that* $|\mathbf{AB}+\mathbf{I}| = |\mathbf{BA}+\mathbf{I}|$. *Secondly, for any matrix* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *and any scalar* b *it further holds true that* $|b\mathbf{A}| = b^n|\mathbf{A}|$.

5.12 Channel Parameter Tracking

Channel parameter tracking lies not within the focus of this thesis. Nevertheless, to keep the global optimization capabilities provided by particle swarm optimization, I hence suggest to implement a hierarchical particle swarm optimization for dynamic environments, following the proposed strategy in [JM06]. For a certain number of observations, a hierarchical concept of dividing the swarm into sub-swarms is proposed to cope with the dynamically changing cost function. Furthermore, [YOJ07] summarizes dynamical mechanisms to track dynamic environments, developed for particle swarm optimization so far.

5.13 Delay-/Parameter Estimation CRLB

For positioning, the theoretically lowest achievable TOA estimation error is especially interesting. Due to multipath propagation, the CRLB for the TOA and the associated FIM depend on a vector parametrization, which again depends on the underlying problem formulation or parametrization. The full parameterization yields a different bound, with significantly higher dimensionality and hence computational complexity, than the bound belonging to the parametrization with separated linear and nonlinear parameters. Furthermore, the SISO, SIMO and time-series-based parametrizations will yield different Fisher matrix dimensions as well.
Therefore, I assess loose TOA CRLBs $\mathrm{crlb}\,(\tau_1)$ as investigated in [SAH12] for this work.

Refer to [YB92] for a full problem description. It can be calculated via employing either the

- full SISO problem parametrization

$$\begin{aligned}
\boldsymbol{\theta} =& [\boldsymbol{\beta}^i, \boldsymbol{\beta}^q \boldsymbol{\tau}]), \quad \mathbf{F}_{\text{FULL}}(\boldsymbol{\theta}) \in \mathbb{R}^{3C \times 3C}, \text{ that is} \\
\hat{\boldsymbol{\theta}} =& \arg\min_{\tilde{\boldsymbol{\theta}}} \left\{ \frac{1}{\sigma_{w_l}} |\hat{\mathbf{h}} - \mathbf{G}(\tilde{\boldsymbol{\tau}}) \tilde{\boldsymbol{\beta}}|^2) \right\} \text{ yielding} \\
\text{CRLB}_{\text{FULL}}(\tau_1) =& \left[\text{diag}\left(\mathbf{F}(\boldsymbol{\theta})^{-1} \right) \right]_{2C+1} \text{ with} \qquad (5.143) \\
\text{CRLB}_{\text{FULL}}(\boldsymbol{\tau}) =& \frac{\sigma_{w_l}^2}{2} \left[\text{Re}\left\{ \text{diag}\left(\boldsymbol{\beta}^H \right) \mathbf{J}_{\mathbf{G}}^T \mathbf{P}_{\mathbf{G}}(\boldsymbol{\tau})^{\perp} \mathbf{J}_{\mathbf{G}} \text{diag}(\boldsymbol{\beta}) \right\} \right]^{-1} \qquad (5.144) \\
=& \left[\text{CRLB}_{\text{FULL}}(\boldsymbol{\theta}) \right]_{2C+1:3C}, \quad \text{or the} \qquad (5.145)
\end{aligned}$$

- SISO SNLLS problem parametrization

$$\begin{aligned}
\boldsymbol{\theta} =& \boldsymbol{\tau}, \mathbf{F}_{\text{SNLLS}}(\boldsymbol{\tau}) \in \mathbb{R}^{C \times C}, \text{ that is} \\
\hat{\boldsymbol{\tau}} =& \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \frac{1}{\sigma_{w_l}} \hat{\mathbf{h}}^H \mathbf{P}_{\mathbf{G}}(\tilde{\boldsymbol{\tau}})^{\perp} \hat{\mathbf{h}} \right\} \text{, yielding} \\
\text{CRLB}_{\text{SNLLS}}(\tau_1) =& [\text{diag}\left(\mathbf{F}(\boldsymbol{\tau})^{-1} \right)]_1 \qquad (5.146) \\
\text{CRLB}_{\text{SNLLS}}(\boldsymbol{\tau}) =& \frac{\sigma_{w_l}^2}{2} \left[\text{Re}\left\{ \text{diag}\left(\hat{\mathbf{h}}^H \mathbf{G}^{\dagger^T} \right) \mathbf{J}_{\mathbf{G}}^T \mathbf{P}_{\mathbf{G}}(\tilde{\boldsymbol{\tau}})^{\perp} \mathbf{J}_{\mathbf{G}} \text{diag}\left(\mathbf{G}^{\dagger} \hat{\mathbf{h}} \right) \right\} \right]^{-1} \qquad (5.147)
\end{aligned}$$

- full TS-SIMO problem parametrization

$$\begin{aligned}
\boldsymbol{\Gamma} =& [\boldsymbol{\Gamma}(0), \ldots, \boldsymbol{\Gamma}(I-1)] \text{ with } \boldsymbol{\Gamma} \in \mathbb{C}^{C \times N_R I} \qquad (5.148) \\
\boldsymbol{\theta} =& [\boldsymbol{\tau}, \text{vec}\left\{ \boldsymbol{\Gamma}^i \right\}, \text{vec}\left\{ \boldsymbol{\Gamma}^q \right\}]), \quad \mathbf{F}_{\text{FULL}}(\boldsymbol{\theta}) \in \mathbb{R}^{(2N_R IC + C) \times (2N_R IC + C)}, \text{ that is} \\
\hat{\boldsymbol{\theta}} =& \arg\min_{\tilde{\boldsymbol{\theta}}} \left\{ \sum_{v=0}^{N_R-1} \sum_{i=0}^{I-1} \frac{1}{\sigma_{w_l}} |\hat{\mathbf{h}}_{v,i} - \mathbf{G}(\tilde{\boldsymbol{\tau}}) \tilde{\boldsymbol{\beta}}_{v,i}|^2 \right\}, \quad \text{yielding} \\
\text{CRLB}_{\text{FULL}}(\tau_1) =& \left[\text{diag}\left(\mathbf{F}(\boldsymbol{\theta})^{-1} \right) \right]_1, \quad \text{with} \qquad (5.149) \\
\text{CRLB}_{\text{FULL}}(\boldsymbol{\tau}) =& \frac{\sigma_{w_l}^2}{2} \left[\text{Re}\left\{ \mathbf{J}_{\mathbf{G}}^T \mathbf{P}_{\mathbf{G}}(\boldsymbol{\tau})^{\perp} \mathbf{J}_{\mathbf{G}} \odot (\boldsymbol{\Gamma}^{\boldsymbol{H}} \boldsymbol{\Gamma})^T \right\} \right]^{-1}, \quad \text{the} \qquad (5.150)
\end{aligned}$$

- SNLLS TS-SIMO problem parametrization

$$\begin{aligned}
\boldsymbol{\theta} =& \boldsymbol{\tau}, \quad \mathbf{F}_{\text{SNLLS}}(\boldsymbol{\theta}) \in \mathbb{R}^{C \times C}, \text{ that is} \\
\hat{\boldsymbol{\tau}} =& \arg\min_{\tilde{\boldsymbol{\tau}}} \left\{ \sum_{v=0}^{N_R-1} \sum_{i=0}^{I-1} \frac{1}{\sigma_{w_l}} |\hat{\mathbf{h}}_{v,i} - \mathbf{G}(\tilde{\boldsymbol{\tau}}) \mathbf{G}(\boldsymbol{\tau})^{\dagger} \hat{\mathbf{h}}_{v,i}|^2 \right\}, \quad \text{yielding} \\
\text{CRLB}_{\text{SNLLS}}(\tau_1) =& \left[\text{diag}\left(\mathbf{F}(\boldsymbol{\theta})^{-1} \right) \right]_1, \quad \text{with} \qquad (5.151) \\
\text{CRLB}_{\text{SNLLS}}(\boldsymbol{\tau}) =& \frac{\sigma_{w_l}^2}{2} \left[\text{Re}\left\{ \mathbf{J}_{\mathbf{G}}^T \mathbf{P}_{\mathbf{G}}(\boldsymbol{\tau})^{\perp} \mathbf{J}_{\mathbf{G}} \odot ((\mathbf{G}(\boldsymbol{\tau})^{\dagger} \hat{\mathbf{H}})^H \mathbf{G}(\boldsymbol{\tau})^{\dagger} \hat{\mathbf{H}})^T \right\} \right]^{-1}. \qquad (5.152)
\end{aligned}$$

Firstly, to be able to calculate a loose bound, I approximate and substitute the channel estimation error by its lower bound

$$\sigma_{w_l}^2 \approx \text{crlb}(h_l). \tag{5.153}$$

Secondly, the SISO SNLLS and the full CRLB are special cases of the TS-SIMO SNLLS and of the full CRLB. The full and the SNLLS bounds differ in employing estimates as substitutes for the actual parameters. Additionally, the Fisher information matrix for the full problem has a higher dimensionality: a size $3C \times 3C$ (SISO) and $(2N_RIC + C) \times (2N_RIC + C)$, whereas the Fisher information for the separated parametrization for both SISO as TS-SIMO has the reduced size $C \times C$. Nevertheless, in the full problem it is also possible to find a separated formulation for the delay parameters only by dividing the complete Fisher matrix into sub-block matrices, as can be seen in the following derivations.

Since the delay estimation CRLB increases proportionally with the channel estimation CRLB, the latter has a major impact on the TOA estimation and therefore on the positioning performance. Consequently, the channel estimation error should be as low as possible. Given a typical and feasible amount of employed training, employing semi-blind channel estimation instead of a purely training based estimator ideally leads to a delay estimation error, which is smaller by a factor of $5 - 20$, as can be seen in (5.154) and (4.85).

5.13.1 Delay CRLB Derivation (SNLLS Parametrization)

It is noteworthy, that although the approximation for the resulting delay CRLB is similar apart from that the actual parameters are substituted by estimates, the derivation for both matrices is not similar: The derivation for the full problem CRLBs is based on applying a rule for inversion of partitioned matrices, whereas the SNLLS-based CRLBs are based on pseudo-inverse derivatives. The derivation for the separable parametrization is already provided as the derivation for the Hessian matrix calculation $\text{crlb}_{SNLLS}(\boldsymbol{\tau}) = [F_{SNLLS}(\boldsymbol{\tau})]^{-1} = [H_{SNLLS}(\boldsymbol{\tau})]^{-1}$ in (5.53) and Section 5.8.

5.13.2 Delay CRLB Derivation (Full Parametrization)

In [YB92] a CRLB for superimposed signals is derived, which is partially similar to the CRLB for the complete vector $\boldsymbol{\theta}$ given in [SAH12] and to the CRLB given here. Since I am mainly interested in the vector $\boldsymbol{\tau}$, however, I will provide a compact calculation for the separable delay CRLB below. The vector CRLB derivation is based on the Fisher matrix calculation and inversion:

$$\begin{aligned}
\text{crlb}\,(\tau_1) =& [\mathbf{F}(\boldsymbol{\theta})^{-1}]_{2C+1,2C+1} \\
=& \underbrace{\frac{\sigma_n^2}{K-L}}_{\text{crlb}_{h_l}} [(2\text{Re}\{\mathbf{J}(\boldsymbol{\theta})^H \mathbf{J}(\boldsymbol{\theta})\})^{-1}]_{2C+1,2C+1}, \quad \text{with} \\
[\mathbf{J}(\boldsymbol{\theta})]_{l,m} =& \frac{\partial h_l(\boldsymbol{\beta^i}, \boldsymbol{\beta^q}, \boldsymbol{\tau})}{\partial \theta_m} \text{ and } \mathbf{J}(\boldsymbol{\theta}) = [\mathbf{J}_{\boldsymbol{\beta^i}} \mathbf{J}_{\boldsymbol{\beta^q}} \mathbf{J}_{\boldsymbol{\tau}}].
\end{aligned} \tag{5.154}$$

Revisiting (4.9), the entries of the Jacobian matrix $\mathbf{J}$ can be calculated as

$$\begin{aligned}
[\mathbf{J}_{\boldsymbol{\beta}^i}]_{l,c} =& [\mathbf{G}]_{l,c} = && \frac{\partial h_l}{\partial \beta_c^i} = g(lT_s - \tau_c) \\
[\mathbf{J}_{\boldsymbol{\beta}^q}]_{l,c} =& j[\mathbf{G}]_{l,c} = && \frac{\partial h_l}{\partial \beta_c^q} = j \cdot g(lT_s - \tau_c) \\
[\mathbf{J}_{\boldsymbol{\tau}}]_{l,c} =& [\mathbf{J_g B}]_{l,c} = && \frac{\partial h_l}{\partial \tau_c} = \beta_c \frac{\partial g(lT_s - \tau_c)}{\partial \tau_c}
\end{aligned} \tag{5.155}$$

with

$$\mathbf{B} = \mathrm{diag}\{\boldsymbol{\beta}\} \quad \text{and} \quad [\mathbf{J_g}]_{l,c} = \frac{\partial g(lT_s - \tau_c)}{\partial \tau_c}. \tag{5.156}$$

Note that, despite the notational abuse, $\mathbf{F}$ corresponds to an approximate and not to the exact FIM. Calculating (5.156) finally leads to

$$\begin{aligned}
\mathbf{F} =& \frac{2}{\mathrm{crlb}_{h_l}} \mathrm{Re}\{\mathbf{J^H J}\} \\
=& \frac{2}{\mathrm{crlb}_{h_l}} \mathrm{Re} \left(\begin{array}{cc|c} \mathbf{J}_{\boldsymbol{\beta}^i}^{\mathbf{H}} \mathbf{J}_{\boldsymbol{\beta}^i} & \mathbf{J}_{\boldsymbol{\beta}^i}^{\mathbf{H}} \mathbf{J}_{\boldsymbol{\beta}^q} & \mathbf{J}_{\boldsymbol{\beta}^i}^{\mathbf{H}} \mathbf{J}_{\boldsymbol{\tau}} \\ \mathbf{J}_{\boldsymbol{\beta}^q}^{\mathbf{H}} \mathbf{J}_{\boldsymbol{\beta}^i} & \mathbf{J}_{\boldsymbol{\beta}^q}^{\mathbf{H}} \mathbf{J}_{\boldsymbol{\beta}^q} & \mathbf{J}_{\boldsymbol{\beta}^q}^{\mathbf{H}} \mathbf{J}_{\boldsymbol{\tau}} \\ \hline \mathbf{J}_{\boldsymbol{\tau}}^{\mathbf{H}} \mathbf{J}_{\boldsymbol{\beta}^i} & \mathbf{J}_{\boldsymbol{\tau}}^{\mathbf{H}} \mathbf{J}_{\boldsymbol{\beta}^q} & \mathbf{J}_{\boldsymbol{\tau}}^{\mathbf{H}} \mathbf{J}_{\boldsymbol{\tau}} \end{array} \right) \\
=& \frac{2}{\mathrm{crlb}_{h_l}} \mathrm{Re} \left(\begin{array}{cc|c} \mathbf{G^T G} & 0 & \mathbf{G^T J_g B} \\ 0 & \mathbf{G^T G} & -j\mathbf{G^T J_g B} \\ \hline \mathbf{B^H J_g^T G} & j\mathbf{B^H J_g^T G} & \mathbf{B^H J_g^T J_g B} \end{array} \right) \\
=& \frac{2}{\mathrm{crlb}_{h_l}} \left(\begin{array}{cc|c} \mathbf{G^T G} & 0 & \mathbf{G^T J_g B^i} \\ 0 & \mathbf{G^T G} & \mathbf{G^T J_g B^q} \\ \hline \mathbf{B^{iT} J_g^T G} & \mathbf{B^{qT} J_g^T G} & \mathrm{Re}\{\mathbf{B^H J_g^T J_g B}\} \end{array} \right).
\end{aligned} \tag{5.157}$$

Let

$$\mathbf{J}_{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{G} & 0 \\ 0 & j\mathbf{G} \end{pmatrix}. \tag{5.158}$$

Here the following auxiliary rule [PP12] can be applied for inversion of the Fisher matrix.

Auxiliary Rule for the Inversion of Partitioned Matrices:

Let $\mathbf{A}_{11}, \mathbf{A}_{12}, \mathbf{A}_{21}, \mathbf{A}_{22}$ be matrices of any size with $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ being square matrices and let $\mathbf{A}_{11}$ and the so called Shur complement $\mathbf{A}_{22} - \mathbf{A}_{21}{\mathbf{A}_{11}}^{-1}\mathbf{A}_{12}$ be non-singular. Then

$$\left(\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array} \right)^{-1} = \left(\begin{array}{c|c} {\mathbf{S}_1}^{-1} & -{\mathbf{A}_{11}}^{-1}\mathbf{A}_{12}{\mathbf{S}_2}^{-1} \\ \hline -{\mathbf{S}_2}^{-1}\mathbf{A}_{21}{\mathbf{A}_{11}}^{-1} & {\mathbf{S}_2}^{-1} \end{array} \right) \quad \text{with} \tag{5.159}$$

$$\mathbf{S}_1 = \mathbf{A}_{11} - \mathbf{A}_{12}{\mathbf{A}_{22}}^{-1}\mathbf{A}_{21} \tag{5.160}$$

$$\mathbf{S}_2 = \mathbf{A}_{22} - \mathbf{A}_{21}{\mathbf{A}_{11}}^{-1}\mathbf{A}_{12}. \tag{5.161}$$

The block-matrix in (5.157) has the structure

$$\left(\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right) = \left(\begin{array}{cc|c} \mathbf{G^T G} & 0 & \mathbf{G^T J_g B^i} \\ 0 & \mathbf{G^T G} & \mathbf{G^T J_g B^q} \\ \hline \mathbf{B^{iT} J_g^T G} & \mathbf{B^{qT} J_g^T G} & \mathrm{Re}\{\mathbf{B^H J_g^T J_g B}\} \end{array}\right). \tag{5.162}$$

Obviously block entries $\mathbf{A}_{11}$ as well as $\mathbf{A}_{22}$ are square matrices. Consequently, employing the auxiliary matrices $\mathbf{C}_{11}, \mathbf{C}_{12}, \mathbf{C}_{21}, \mathbf{C}_{22}$ it follows that

$$\mathbf{F}^{-1} = \frac{\mathrm{crlb}_{h_l}}{2} \mathrm{Re}\left(\begin{array}{c|c} \mathbf{C}_{11}^{-1} & \mathbf{C}_{12} \\ \hline \mathbf{C}_{21} & \mathbf{C}_{22}^{-1} \end{array}\right), \tag{5.163}$$

given the block-wise entries of the block matrix above:

$$\mathbf{C}_{22} = \mathrm{Re}\{\mathbf{J_\tau^H J_\tau}\} - \left(\mathbf{B^{iT} J_g^T G} \quad \mathbf{B^{qT} J_g^T G}\right) (\mathbf{J}_{\boldsymbol{\beta}}^H \mathbf{J}_{\boldsymbol{\beta}})^{-1} (\mathbf{G^T J_g B^i} \quad \mathbf{G^T J_g B^q})^T. \tag{5.164}$$

Note that the well known rule for the inversion of block-diagonal matrices allows to calculate the inverse of $\mathbf{J}_{\boldsymbol{\beta}}^H \mathbf{J}_{\boldsymbol{\beta}}$ as

$$\begin{pmatrix} \mathbf{G^T G} & 0 \\ 0 & \mathbf{G^T G} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{G^T G}^{-1} & 0 \\ 0 & \mathbf{G^T G}^{-1} \end{pmatrix}. \tag{5.165}$$

Further note that, considering any complex vector $\boldsymbol{\beta} \epsilon \mathbb{C}^{C \times 1}$ and any auxiliary square matrix $\mathbf{A}$, the following equality holds true:

$$(\mathrm{diag}\{\boldsymbol{\beta}\})^H \mathbf{A} \mathrm{diag}\{\boldsymbol{\beta}\} = \mathbf{R}_{\boldsymbol{\beta}\boldsymbol{\beta}}^T \odot \mathbf{A} = \mathbf{A} \odot \mathbf{R}_{\boldsymbol{\beta}\boldsymbol{\beta}}^T \quad \text{with } \mathbf{R}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \boldsymbol{\beta}\boldsymbol{\beta}^H. \tag{5.166}$$

The second part of the equality is true since the Hadamard product is commutative in general. Additionally, note that the Hadamard product is distributive in general. Finally, note that

$$\mathrm{Re}\{\mathbf{R}_{\boldsymbol{\beta}\boldsymbol{\beta}}^T\} = \mathbf{R}_{\boldsymbol{\beta}^i\boldsymbol{\beta}^i}^T + \mathbf{R}_{\boldsymbol{\beta}^q\boldsymbol{\beta}^q}^T. \tag{5.167}$$

Consequently, inserting (5.158) and (5.155) into (5.164) results in

$$\begin{aligned}
\mathbf{C}_{22} =& \mathrm{Re}\{\mathbf{B}^H\mathbf{J}_\mathbf{g}^\mathbf{T}\mathbf{J}_\mathbf{g}\mathbf{B}\} - \begin{pmatrix}\mathbf{B^{iT}J_g^TG} & \mathbf{B^{qT}J_g^TG}\end{pmatrix}\begin{pmatrix}\mathbf{G^TG} & 0\\ 0 & \mathbf{G^TG}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{G^TJ_gB^i}\\ \mathbf{G^TJ_gB^q}\end{pmatrix}\\
\overset{(5.165)}{=}& \mathrm{Re}\{\mathbf{B}^H\mathbf{J}_\mathbf{g}^\mathbf{T}\mathbf{J}_\mathbf{g}\mathbf{B}\} - \begin{pmatrix}\mathbf{B^{iT}J_g^TG(G^TG)^{-1}} & \mathbf{B^{qT}J_g^TG(G^TG)^{-1}}\end{pmatrix}\begin{pmatrix}\mathbf{G^TJ_gB^i}\\ \mathbf{G^TJ_gB^q}\end{pmatrix}\\
=& \mathrm{Re}\{\mathbf{B^H}\mathbf{J}_\mathbf{g}^\mathbf{T}\mathbf{J}_\mathbf{g}\mathbf{B}\} - \left(\mathbf{B^{iT}J_g^TG}\underbrace{(\mathbf{G^TG})^{-1}\mathbf{G}^T}_{\mathbf{G}^\dagger}\mathbf{J_gB}^i + \mathbf{B^{qT}J_g^TG}\underbrace{(\mathbf{G^TG})^{-1}\mathbf{G}^T}_{\mathbf{G}^\dagger}\mathbf{J_gB}^q\right)\\
=& \mathrm{Re}\{\mathbf{B^H}\mathbf{J}_\mathbf{g}^\mathbf{T}\mathbf{J}_\mathbf{g}\mathbf{B}\} - \left(\mathbf{B^{iT}J_g^TGG^\dagger J_gB}^i + \mathbf{B^{qT}J_g^TGG^\dagger J_gB}^q\right)\\
\overset{(5.166)}{=}& \mathrm{Re}\{\mathbf{R}_{\boldsymbol{\beta}\boldsymbol{\beta}}^T\} \odot \mathbf{J}_\mathbf{g}^\mathbf{T}\mathbf{J}_\mathbf{g} - \left(\mathbf{R}_{\boldsymbol{\beta}^i\boldsymbol{\beta}^i}^T \odot \mathbf{J}_\mathbf{g}^\mathbf{T}\mathbf{GG}^\dagger\mathbf{J}_\mathbf{g} + \mathbf{R}_{\boldsymbol{\beta}^q\boldsymbol{\beta}^q}^T \odot \mathbf{J}_\mathbf{g}^\mathbf{T}\mathbf{GG}^\dagger\mathbf{J}_\mathbf{g}\right)\\
=& \mathrm{Re}\{\mathbf{R}_{\boldsymbol{\beta}\boldsymbol{\beta}}^T\} \odot \mathbf{J}_\mathbf{g}^\mathbf{T}\mathbf{J}_\mathbf{g} - (\mathbf{R}_{\boldsymbol{\beta}^i\boldsymbol{\beta}^i}^T + \mathbf{R}_{\boldsymbol{\beta}^q\boldsymbol{\beta}^q}^T) \odot \mathbf{J}_\mathbf{g}^\mathbf{T}\mathbf{GG}^\dagger\mathbf{J}_\mathbf{g} \qquad (5.168)\\
\overset{(5.167)}{=}& \mathrm{Re}\{\mathbf{R}_{\boldsymbol{\beta}\boldsymbol{\beta}}^T\} \odot \mathbf{J}_\mathbf{g}^\mathbf{T}(\mathbf{I}_{L+1} - \mathbf{GG}^\dagger)\mathbf{J}_\mathbf{g}\\
=& \mathrm{Re}\{\mathbf{R}_{\boldsymbol{\beta}\boldsymbol{\beta}}^T \odot \mathbf{J}_\mathbf{g}^\mathbf{T}(\mathbf{I}_{L+1} - \mathbf{GG}^\dagger)\mathbf{J}_\mathbf{g}\}, \quad \text{or} \qquad (5.169)\\
=& \mathrm{Re}\{\mathbf{B}^H\mathbf{J}_\mathbf{g}^\mathbf{T}(\mathbf{I}_{L+1} - \mathbf{GG}^\dagger)\mathbf{J}_\mathbf{g}\}\mathbf{B}\}, \quad \text{or} \qquad (5.170)\\
=& \mathrm{Re}\{\mathbf{J}_{\boldsymbol{\tau}}^H(\mathbf{I} - \mathbf{J}_{\boldsymbol{\beta}}\mathbf{J}_{\boldsymbol{\beta}}^\dagger)\mathbf{J}_{\boldsymbol{\tau}}\}. \qquad (5.171)
\end{aligned}$$

Again, from a positioning point of view, we are basically interested in the TOA CRLB, which can be calculated by evaluating the single entry block $\mathbf{C}_{22}^{-1}$ in the lower right part of the partitioned matrix in (5.163). The overall CRLB separability into a CRLB part for the linear and a part for the nonlinear parameter dependencies becomes clear in (5.162) and matches well with the search space reduction addressed in (5.22). Consequently, $\mathbf{C}_{22}$ is the FIM of $\boldsymbol{\tau}$ denoted by $\mathbf{F}_{\boldsymbol{\tau}}$ and the TOA CRLB can be written compactly as

$$\begin{aligned}
\mathrm{crlb}_{\tau_1} =& \frac{\mathrm{crlb}_{h_l}}{2}[\mathbf{F}_{\boldsymbol{\tau}}^{-1}]_{1,1}, \text{ with}\\
\mathbf{F}_{\boldsymbol{\tau}} =& \mathbf{J}_{\boldsymbol{\tau}}^\mathbf{T}(\mathbf{I} - \mathbf{J}_{\boldsymbol{\beta}}\mathbf{J}_{\boldsymbol{\beta}}^\dagger)\mathbf{J}_{\boldsymbol{\tau}}\\
=& \mathrm{Re}\{\mathbf{J}_{\boldsymbol{\tau}}^\mathbf{H}(\mathbf{I} - \mathbf{J}_{\boldsymbol{\beta}}\mathbf{J}_{\boldsymbol{\beta}}^\dagger)\mathbf{J}_{\boldsymbol{\tau}}\}. \qquad (5.172)
\end{aligned}$$

5.13.3 MIMO Delay CRLB Derivation

The channel estimation CRLB is calculated like in the SIMO case via

$$\mathrm{CRLB}\{h_{u,v,l}\} = \frac{\sigma_n^2}{K-L}. \qquad (5.173)$$

Let us consider the SIMO spatial signature and delay estimation case. Then, considering some substitutions due to a slightly different signal model in [Lee98, Mia07], the delay estimation CRLB is calculated as follows:

$$\mathrm{CRLB}(\hat{\boldsymbol{\tau}}) = \frac{\mathrm{CRLB}(h_{l,u,v})}{2}\mathrm{Re}\{\mathbf{J_g}^T(\boldsymbol{\tau})\boldsymbol{P}_\mathbf{G}^\perp\mathbf{J_g}(\boldsymbol{\tau}) \odot (\boldsymbol{\Gamma}^H\boldsymbol{\Gamma})^T\}^{-1}. \qquad (5.174)$$

The derivation of (5.174) can be understood by revising the underlying signal model, such that we rearrange the channel matrix in a single long vector

$$\mathbf{vec}\{\mathbf{H^T}\} = \mathbf{vec}\{(\mathbf{\Gamma G}^T)^T\} = \mathbf{vec}\{\mathbf{G\Gamma}^T\} = (\mathbf{I}_{N_R} \otimes \mathbf{G}) \underbrace{\mathbf{vec}\{\mathbf{\Gamma}^T\}}_{\boldsymbol{\gamma}}. \tag{5.175}$$

Note that the $\mathbf{vec}\{\cdot\}$ operator concatenates column-wise. Let $\boldsymbol{\mathcal{J}}$ denote the Jacobian for multiple antennas and

$$[\boldsymbol{\mathcal{J}}]_{l,c} = \frac{\partial h_l(\boldsymbol{\gamma}^i, \boldsymbol{\gamma}^q, \boldsymbol{\tau})}{\partial \theta_c} \quad \text{and} \quad \boldsymbol{\mathcal{J}} = [\mathbf{J}_{\boldsymbol{\gamma}^i} \mathbf{J}_{\boldsymbol{\gamma}^q} \mathbf{J}_{\boldsymbol{\tau}}]. \tag{5.176}$$

Then, the sub-blocks of the Jacobian can be calculated as

$$\mathbf{J}_{\boldsymbol{\gamma}^i} = \mathbf{I}_{N_R} \otimes \mathbf{G} \qquad \text{with} \quad \frac{\partial \Omega}{\partial \gamma^i_{v,c}} = \frac{\partial h_{l,v'}}{\partial \gamma^i_{v,c}} = \begin{cases} g(lT - \tau_c) & v' = v \\ 0 & v' \neq v, \end{cases} \tag{5.177}$$

$$\mathbf{J}_{\boldsymbol{\gamma}^q} = j\mathbf{I}_{N_R} \otimes \mathbf{G} \qquad \text{with} \quad \frac{\partial \Omega}{\partial \gamma^q_{v,c}} = \frac{\partial h_{l,v'}}{\partial \gamma^q_{v,c}} = \begin{cases} jg(lT - \tau_c) & v' = v \\ 0 & v' \neq v, \end{cases}$$

$$\mathbf{J}_{\boldsymbol{\tau}} = \begin{pmatrix} \mathbf{J_g C}_1 \\ \vdots \\ \mathbf{J_g C}_{N_R} \end{pmatrix} \qquad \text{with} \quad \frac{\partial \Omega}{\partial \tau_c} = \frac{\partial h_{l,v'}}{\partial \tau_c} = \gamma_{v',c} \frac{\partial g(lT - \tau_c)}{\partial \tau_c} \quad \text{and} \tag{5.178}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \dots & \mathbf{C}_{N_R} \end{bmatrix}^T \qquad \text{with} \quad \mathbf{C}_v = \mathrm{diag}\{\boldsymbol{\gamma}_v\} \quad \forall v \epsilon \{1, \dots, N_R\}. \tag{5.179}$$

Consequently, proceeding as in the SISO case, the Fisher matrix derivation yields the SIMO Fischer matrix

$$\boldsymbol{\mathcal{F}} = \tfrac{2}{crlb_{h_l}} \mathrm{Re}\{\boldsymbol{\mathcal{J}}^{\boldsymbol{H}} \boldsymbol{\mathcal{J}}\} = \tfrac{2}{crlb_{h_l}} \mathrm{Re} \left(\begin{array}{c|c} \mathbf{J}^{\mathbf{H}}_{\boldsymbol{\gamma}} \mathbf{J}_{\boldsymbol{\gamma}} & \mathbf{J}^{\mathbf{H}}_{\boldsymbol{\gamma}} \mathbf{J}_{\boldsymbol{\tau}} \\ \hline \mathbf{J}^{\mathbf{H}}_{\boldsymbol{\tau}} \mathbf{J}_{\boldsymbol{\gamma}} & \mathbf{J}^{\mathbf{H}}_{\boldsymbol{\tau}} \mathbf{J}_{\boldsymbol{\tau}} \end{array} \right)$$

with

$$\mathbf{J}_{\boldsymbol{\gamma}} = \begin{pmatrix} \mathbf{I}_{N_R} \otimes \mathbf{G} & 0 \\ 0 & j\mathbf{I}_{N_R} \otimes \mathbf{G} \end{pmatrix}. \tag{5.180}$$

Inserting (5.179) and defining $\boldsymbol{\mathcal{G}}_{N_R} = \mathbf{I}_{N_R} \otimes \mathbf{G}$ into the Fischer matrix expression leads to

$$\boldsymbol{\mathcal{F}} = \frac{2}{\mathrm{crlb}_{h_l}} \begin{pmatrix} \boldsymbol{\mathcal{G}}^T_{N_R} \boldsymbol{\mathcal{G}}_{N_R} & 0 & \boldsymbol{\mathcal{G}}^T_{N_R} \mathbf{J}^i_{\boldsymbol{\tau}} \\ 0 & \boldsymbol{\mathcal{G}}^T_{N_R} \boldsymbol{\mathcal{G}}_{N_R} & \boldsymbol{\mathcal{G}}^T_{N_R} \mathbf{J}^q_{\boldsymbol{\tau}} \\ \mathbf{J}^{iT}_{\boldsymbol{\tau}} \boldsymbol{\mathcal{G}}_{N_R} & \mathbf{J}^{qT}_{\boldsymbol{\tau}} \boldsymbol{\mathcal{G}}_{N_R} & \mathrm{Re}\{\mathbf{J}^H_{\boldsymbol{\tau}} \mathbf{J}_{\boldsymbol{\tau}}\} \end{pmatrix}. \tag{5.181}$$

For the following derivations note that

$$\sum_{v=1}^{N_R} \mathbf{R}^T_{\boldsymbol{\gamma}_v \boldsymbol{\gamma}_v} = \left(\mathbf{\Gamma}^T \mathbf{\Gamma}^*\right)^T = \mathbf{\Gamma}^H \mathbf{\Gamma}. \tag{5.182}$$

Then, similarly as in the SIMO case, the CRLB for the delay estimation is derived as follows:

$$\begin{aligned}
\mathrm{crlb}_{\boldsymbol{\tau}} =& \mathrm{crlb}_{h_{l,v}} \boldsymbol{\mathcal{F}}_{\boldsymbol{\tau}}^{-1}, \quad \text{with} \qquad (5.183)\\
\boldsymbol{\mathcal{F}}_{\boldsymbol{\tau}} =& \mathrm{Re}\left\{ \begin{pmatrix} \mathbf{C}_1^H \mathbf{J_g}^T & \dots & \mathbf{C}_{N_R}^H \mathbf{J_g}^T \end{pmatrix} \begin{pmatrix} \mathbf{J_g C}_1 \\ \vdots \\ \mathbf{J_g C}_{N_R} \end{pmatrix} \right\} \\
& - \begin{pmatrix} \mathbf{C^{iT}} (\mathbf{I}_{N_R} \otimes \mathbf{J_g})^T \boldsymbol{\mathcal{G}} & \mathbf{C^{qT}} (\mathbf{I}_{N_R} \otimes \mathbf{J_g})^T \boldsymbol{\mathcal{G}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{G}}^{\boldsymbol{T}} \boldsymbol{\mathcal{G}} & 0 \\ 0 & \boldsymbol{\mathcal{G}}^{\boldsymbol{T}} \boldsymbol{\mathcal{G}} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\mathcal{G}}^{\boldsymbol{T}} (\mathbf{I}_{N_R} \otimes \mathbf{J_g}) \mathbf{C}^i \\ \boldsymbol{\mathcal{G}}^{\boldsymbol{T}} (\mathbf{I}_{N_R} \otimes \mathbf{J_g}) \mathbf{C}^q \end{pmatrix} \\
=& \mathrm{Re}\{ \mathbf{C}^H (\mathbf{I}_{N_R} \otimes \mathbf{J_g})^T (\mathbf{I}_{N_R} \otimes \mathbf{J_g}) \mathbf{C} \} \\
& - \begin{pmatrix} \mathbf{C^{iT}} (\mathbf{I}_{N_R} \otimes \mathbf{J_g})^T \boldsymbol{\mathcal{G}} (\boldsymbol{\mathcal{G}}^{\boldsymbol{T}} \boldsymbol{\mathcal{G}})^{-1} & \mathbf{C^{qT}} (\mathbf{I}_{N_R} \otimes \mathbf{J_g})^T \boldsymbol{\mathcal{G}} (\boldsymbol{\mathcal{G}}^{\boldsymbol{T}} \boldsymbol{\mathcal{G}})^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{G}}^{\boldsymbol{T}} (\mathbf{I}_{N_R} \otimes \mathbf{J_g}) \mathbf{C}^i \\ \boldsymbol{\mathcal{G}}^{\boldsymbol{T}} (\mathbf{I}_{N_R} \otimes \mathbf{J_g}) \mathbf{C}^q \end{pmatrix} \\
=& \mathrm{Re}\left\{ \sum_{v=1}^{N_R} \mathbf{C}_v^H \mathbf{J_g}^T \mathbf{J_g} \mathbf{C}_v \right\} \\
& - \left(\sum_{v=1}^{N_R} \mathbf{C}_v^{iT} \mathbf{J_g}^T \mathbf{G} (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{C}_v^i + \sum_{v=1}^{N_R} \mathbf{C}_v^{qT} \mathbf{J_g}^T \mathbf{G} (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{C}_v^q \right) \\
=& \mathrm{Re}\left\{ \left(\sum_{v=1}^{N_R} \mathbf{R}_{\gamma_v \gamma_v}^T \right) \odot \mathbf{J_g}^T (\mathbf{I}_{L+1} - \mathbf{G}\mathbf{G}^\dagger) \mathbf{J_g} \right\} \\
\overset{(5.182)}{=}& \mathrm{Re}\{ \boldsymbol{\Gamma}^T \boldsymbol{\Gamma}^* \odot \mathbf{J_g}^T (\mathbf{I}_{L+1} - \mathbf{G}\mathbf{G}^\dagger) \mathbf{J_g} \}. \qquad (5.184)
\end{aligned}$$

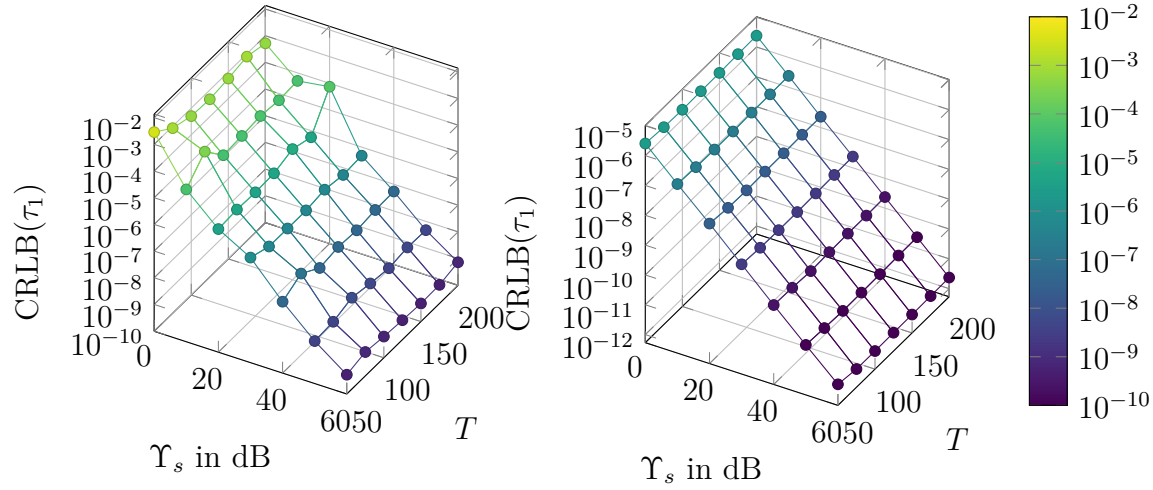
The CRLB variation over the SNR and the symbol duration T is shown in Figure 5.10. For higher cluster numbers C the Fisher information matrix becomes ill-conditioned and a meaningful CRLB cannot be calculated anymore. With increasing symbol duration T the resolution of the pulse superposition decreases and therefore the CRLB increases.

5.13.4 Time Series Extension

If the observations of a whole time series are employed, we can either formulate a CRLB based on the deterministic maximum likelihood estimator, or we can formulate a CRLB based on the stochastic maximum likelihood estimator. I begin by utilizing the conditional likelihood, thus the deterministic estimator, analogous to the SIMO case. In case multiple time series channel measurements are available, the delay CRLB is straightforward to be determined as

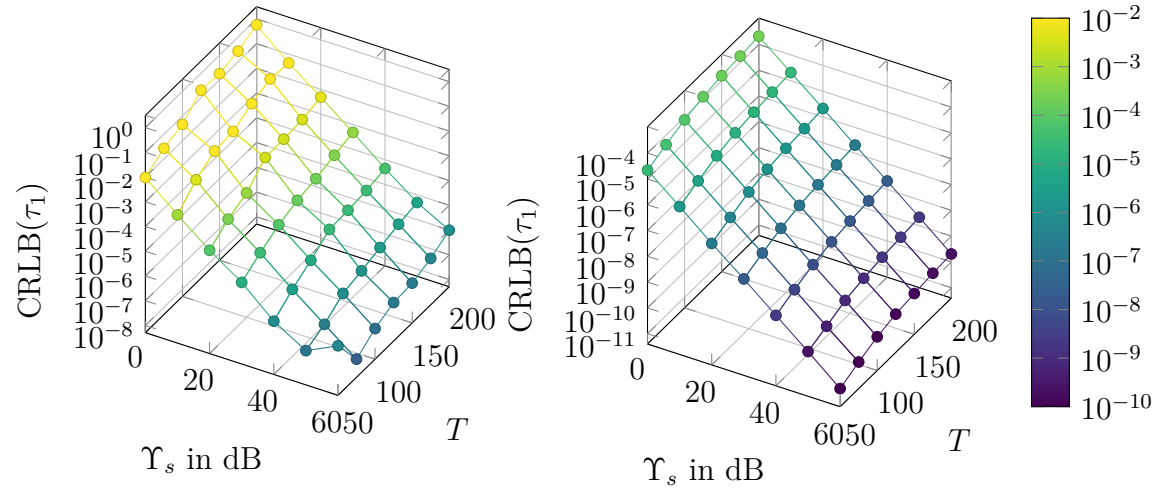
$$\begin{aligned}
\mathrm{CRLB}(\hat{\boldsymbol{\tau}})^{\text{conditional}} =& \frac{\mathrm{CRLB}(h_{l,u,v})}{2} \mathrm{Re}\{ \mathbf{J_g}^T(\boldsymbol{\tau}) \boldsymbol{P}_{\mathbf{G}}^{\perp} \mathbf{J_g}(\boldsymbol{\tau}) \odot (\boldsymbol{\Gamma}_I^H \boldsymbol{\Gamma}_I)^T \}^{-1}, \qquad (5.185)\\
\text{with } \boldsymbol{\Gamma}_I =& [\boldsymbol{\Gamma}_1^T, \dots, \boldsymbol{\Gamma}_I^T]^T. \qquad (5.186)
\end{aligned}$$

If, on the other hand, the unconditional likelihood or the stochastic maximum likelihood estimator is used to formulate a CRLB, the following formulation provided in [CY03] can



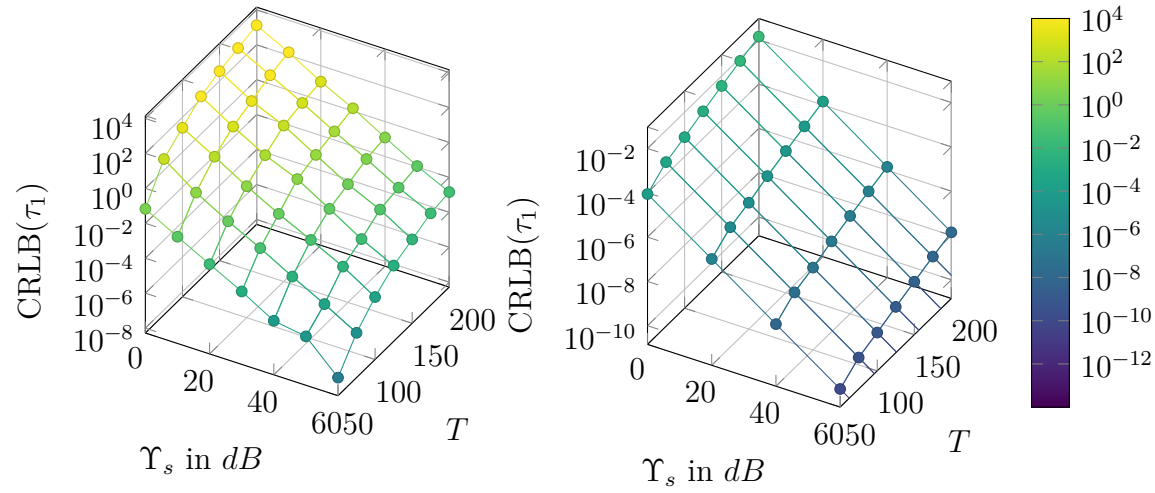
(a) $C = 1$, $I = 1$.

(b) $C = 1$, $I = 100$



(c) $C = 2$, $I = 1$.

(d) $C = 2$, $I = 100$.



(e) $C = 3$, $I = 1$.

(f) $C = 3$, $I = 100$

Figure 5.10: The delay estimation CRLB(τ_1) varies with the SNR, the symbol duration and the number of multipath clusters C.

be translated to the delay estimation problem in the following manner:

$$\mathrm{CRLB}(\hat{\boldsymbol{\tau}})^{\text{unconditional}} = \left\{ \frac{\sigma_{\mathbf{w}}^2}{2I} [\mathrm{Re}\{\mathbf{J_g}\mathbf{P_G}^{\perp}\mathbf{J_g}^T \odot (\mathbf{C}_{\boldsymbol{\beta}}\mathbf{G}^H\mathbf{R}_{\hat{\mathbf{h}}}^{-1}\mathbf{G}\mathbf{C}_{\boldsymbol{\beta}})^{\mathbf{T}}\}]^{-1} \right\}. \tag{5.187}$$

If the full Fisher matrix is needed, it needs to be derived. Although, when dealing with the stochastic maximum likelihood, the modeling function given in [Ric05] is not the same as the one here. We can adapt the derivations for the stochastic CRLB up to the point, where the required derivatives are employed. Therefore recall that the likelihood can be written as

$$L(\mathbf{H}|\boldsymbol{\theta}) = -N_R I \ln|\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})| - \mathrm{trace}(\mathbf{H}^H\mathbf{C}_{\hat{\mathbf{h}}}^{-1}(\boldsymbol{\theta})\mathbf{H}). \tag{5.188}$$

The parametrization via the parameter vector $\boldsymbol{\theta}$ summarizes the unknowns, that is, the delays τ and the physical channel covariance matrix $\mathbf{C}_{\boldsymbol{\beta}}$. Taking into account that the covariance matrices are symmetric and hence the upper triangular matrix of $\mathbf{C}_{\boldsymbol{\beta}}$ suffices to parametrize $\mathbf{C}_{\hat{\mathbf{h}}}$ properly, let $\boldsymbol{\eta}$ be the vector comprising all real and imaginary part elements of the upper triangular matrix of $\mathbf{C}_{\boldsymbol{\beta}}$. Consequently, due to splitting the matrix into an upper and lower triangular and also splitting into real and imaginary part, and noting that the main diagonal only has real values, the size of

$$\begin{aligned}\boldsymbol{\eta} =& [\underbrace{[\mathbf{C}]_{1,1}, \ldots, [\mathbf{C}]_{C,C}}_{C}, \underbrace{\mathrm{Re}\,[[\mathbf{C}_{\boldsymbol{\beta}}]_{1,2}, \ldots, [\mathbf{C}_{\boldsymbol{\beta}}]_{1,C}, [\mathbf{C}_{\boldsymbol{\beta}}]_{2,3}, \ldots, [\mathbf{C}_{\boldsymbol{\beta}}]_{2,C}, \ldots, [\mathbf{C}_{\boldsymbol{\beta}}]_{C-1,C}]}_{\frac{C^2-C}{2}}, \ldots \\ & \ldots, \underbrace{\mathrm{Im}\,[[\mathbf{C}_{\boldsymbol{\beta}}]_{1,2}, \ldots, [\mathbf{C}_{\boldsymbol{\beta}}]_{1,C}, [\mathbf{C}_{\boldsymbol{\beta}}]_{2,3}, \ldots, [\mathbf{C}_{\boldsymbol{\beta}}]_{2,C}, \ldots, [\mathbf{C}_{\boldsymbol{\beta}}]_{C-1,C}]}_{\frac{C^2-C}{2}}]\end{aligned} \tag{5.189}$$

adds up to $\frac{2}{2}(C^2 - C) + C = C^2$. Then, the parametrization of the overall parameter vector $\boldsymbol{\theta}$ has a size of $C^2 + C$ and it is given by

$$\boldsymbol{\theta} = [\boldsymbol{\tau}, \boldsymbol{\eta}]. \tag{5.190}$$

After elaborately employing some matrix derivation rules and successively simplifying the relevant terms in the equations according to [Ric05], the CRLB for $\boldsymbol{\theta}$ can be calculated in the following manner:

$$\mathrm{CRLB}_{SML}(\boldsymbol{\theta}) = \boldsymbol{\mathcal{F}}^{-1} = \frac{1}{N_R I}(\mathbf{D}^H(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\theta})) \leq \mathrm{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\}, \tag{5.191}$$

with the matrix

$$\mathbf{D}(\boldsymbol{\theta}) = \left[\mathrm{vec}\{\mathbf{L}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial}{\partial \theta_1} \mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta}) \right) \mathbf{L}^{-H}(\boldsymbol{\theta})\} \ldots \mathrm{vec}\{\mathbf{L}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial}{\partial \theta_{C^2+C}} \mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta}) \right) \mathbf{L}^{-H}(\boldsymbol{\theta})\} \right], \tag{5.192}$$

given the Cholesky decomposition $\mathbf{C}_{\hat{\mathbf{h}}} = \mathbf{L}(\boldsymbol{\theta})\mathbf{L}(\boldsymbol{\theta})^H$. Up to this point I summarized the stochastic CRLB calculations as provided in [Ric05]. The relevant derivative calculations as applicable to this thesis can be found in the Appendix A.8.

5.14 Soft Information for Estimates

Soft information is another expression for estimation reliability. If parameters are estimated, it is useful to know how reliable these estimates are. Especially in more sophisticated cases as the ones discussed here, e.g. high dimensional, non-linear, non-convex, global optimization problems, which cannot be guaranteed to find a reliable solution in every case, it is crucial to know whether the estimate is reliable or not. Furthermore, reliability information can be obviously employed to improve any further estimates. A straightforward way to quantify reliability information, or soft information, is to calculate, or, if this is not possible, to estimate the estimation vector's covariance matrix. The main diagonal covariance matrix entries represent the parameter variances and hence they can be interpreted as soft information. The inverse Fisher matrix, or inverse Hessian, of the actual estimate $\hat{\boldsymbol{\xi}}$ for an arbitrary parameter vector $\boldsymbol{\xi}$ is the estimates' covariance matrix

$$\hat{\mathbf{C}}_{\hat{\boldsymbol{\xi}}} = \mathbf{F}(\hat{\boldsymbol{\xi}})^{-1}. \tag{5.193}$$

This matrix indicates how reliable the hard estimate of $\boldsymbol{\xi}$ is. Instead of calculating the CRLB for an estimate by utilizing the correct parameter vector, here, the estimate is employed. Note, that for an unbiased estimator it holds:

$$\hat{\boldsymbol{\xi}} \sim \mathcal{N}\left(\boldsymbol{\xi}, \mathbf{C}_{\hat{\boldsymbol{\xi}}}\right). \tag{5.194}$$

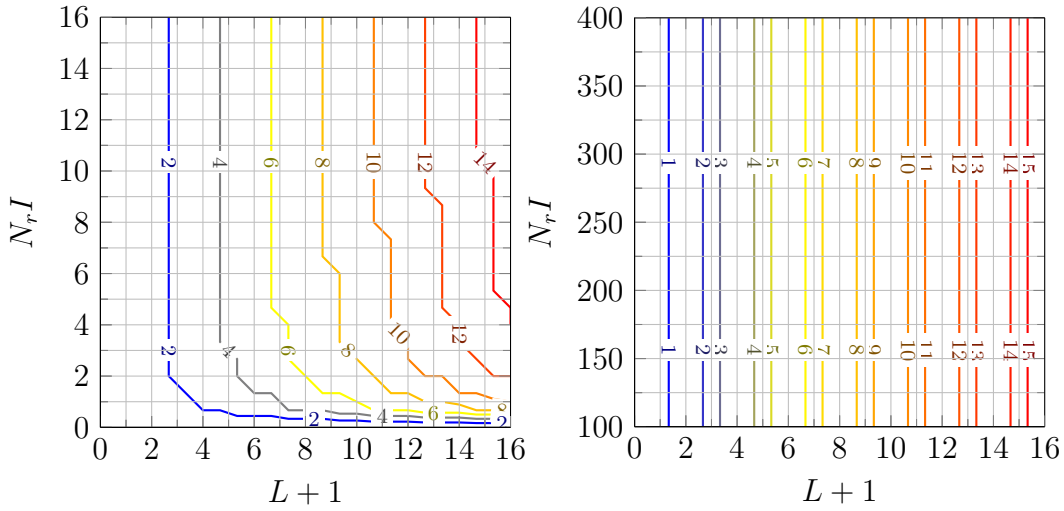
5.15 Identifiability and Overfitting

Given an infinite number of observations, an error-free mathematical model is said to be identifiable, if the underlying model parameters can be theoretically determined [JG85]. Hence, the issue of identifiability should be addressed prior to parameter estimation. A necessary condition to enable parameter identifiability is the one requiring that the underlying model provides more equations, that is, observations, than it has parameters to be estimated. Therefore, intuitively, we can formulate a necessary identifiability condition in terms of constraining the model's degrees of freedom to be larger than zero. The degrees of freedom corresponds to the number of observations minus the number of parameters that are supposed to be estimated. Keeping in mind that the estimation problem at hand consists of a real as well as a imaginary part for the observations and partially for the parameters, it is straightforward that the number of observations corresponds to $2N_R I(L+1)$ and the number of parameters is given by the dimension of $\boldsymbol{\theta} = [\boldsymbol{\tau}, \mathrm{Re}\{\mathrm{vec}\{\boldsymbol{\Gamma}\}\}, \mathrm{Im}\{\mathrm{vec}\{\boldsymbol{\Gamma}\}\}]$, being $C + 2N_R IC$. This leads to the following inequality:

$$2N_R I(L+1) - C - 2N_R IC > 0, \tag{5.195}$$

which, in return, is equivalent to the following condition:

$$C < \frac{1}{1 + \frac{1}{2N_R I}}(L+1). \tag{5.196}$$



(a) Low number of observations (b) High number of observations

Figure 5.11: Maximum identifiable multipath components C_{max}.

The maximum number of estimable multipath components consequently reads $C_{max} = \left\lfloor \frac{1}{1+\frac{1}{2N_R I}}(L+1) \right\rfloor$. From (5.196) and Figure 5.11 one can see that for a low number of observations the number of estimable multipath components theoretically are significantly smaller than the number of channel taps given by $L+1$. Reviewing the literature, two mathematically more thorough identifiability theorem formulations applicable to the model assessed in this thesis, can be found in [Lee98, WZ89a]. The issue of identifiability assuming a slightly different model than the one in this thesis is investigated in [WZ89a]. Fortunately, the author of [Lee98] clarifies that the findings presented in [WZ89a] can be adopted to a signal model corresponding to the one considered in this thesis. The main difference regarding the two contributions lies in the parameters that are supposed to be estimated. Whereas the authors of [WZ89a] refer to an angle of arrival and signal matrix estimation, [Lee98] addresses a delay and signature matrix estimation, as do we. Consider twofold: Firstly, the mathematical model structures in both contributions are similar. Secondly, as explained before, the delay estimation problem in the time domain can be represented by a phase shift estimation in the frequency domain. Therefore the applicability of [WZ89a] to [Lee98] makes sense. Consequently, recall that the underlying error-free model in this thesis is formulated such that it is similar to the one given in [Lee98]. It is described by

$$\mathbf{H}^T = \mathbf{G}(\boldsymbol{\tau})\boldsymbol{\Gamma}^T. \tag{5.197}$$

Assessing (5.197), it becomes clear that we are actually interested in estimating a set of linearly given parameters represented by the spatial signature matrix $\boldsymbol{\Gamma}^{\boldsymbol{T}}$, as well as we are interested in estimating a non-linearly dependent vector $\boldsymbol{\tau}$. To assure that the model parameters are identifiable, the following condition has to be fulfilled according to [Lee98]:

$$\mathbf{H}^T = \mathbf{G}(\boldsymbol{\tau})\boldsymbol{\Gamma}^T \neq \mathbf{G}(\boldsymbol{\tau}')\boldsymbol{\Gamma}'^T \quad \text{for all } \boldsymbol{\tau} \neq \boldsymbol{\tau}' \text{ and } \boldsymbol{\Gamma} \neq \boldsymbol{\Gamma}'. \tag{5.198}$$

In the following I will summarize the identifiability conditions as stated in [Lee98]. The proof is identical to that given in [Lee98] and hence also identical to the proof given in [WZ89a], provided that the model and the parameters are replaced appropriately, as suggested in [Lee98].

Theorem 1. *Suppose an N_R element array receives C identical waveforms with distinct delays $\boldsymbol{\tau}$. If the signal is non-ambiguous, then $\boldsymbol{\tau}$ and the spatial signature $\boldsymbol{\Gamma}$ may be uniquely determined, provided that*

$$C < \frac{L+1+m'}{2}, \tag{5.199}$$

where $m' = \operatorname{rank}(\boldsymbol{\Gamma})$. If, instead

$$C < \frac{2m'}{2m'+1}(L+1), \tag{5.200}$$

then $\boldsymbol{\tau}$ and $\boldsymbol{\Gamma}$ may be uniquely determined with probability one.

The authors of [Lee98] also point out that, assuming $\boldsymbol{\Gamma}$ has full rank, such that $m' = \min(C, N_R)$, if $C < N_R$ (that is $m' = C$), and substituting m' by C, it follows that the parameters are identifiable if $C < L+1$. On the other hand, if $N_R < C$ (that is $m' = N_R$), then the upper bound in (5.199) approaches $C < \frac{L+1}{2}$ for large numbers $L+1$. Note that, in this thesis, however, $L+1$ does not tend to become very large. Further note that, considering both the SIMO and the time series extension case instead of having simply the dimension $C \times N_R$, the matrix $\boldsymbol{\Gamma}$ has the dimension $C \times N_R I$. Finally, applying this theorem 1 to the similar and yet more complicated model proposed in this thesis, via the substitution $m' = \min(C, N_R I)$, we can formulate the following

Theorem 2. *Within the framework proposed in this thesis, the parameters of interest $\boldsymbol{\tau}$ and $\boldsymbol{\Gamma}$, if $\boldsymbol{\tau}$ is distinct and the signals are non-ambiguous, are identifiable for*

$$C < \begin{cases} L+1 & \text{if } C < N_R I \\ \frac{L+1+N_R I}{2} & \text{else,} \end{cases} \tag{5.201}$$

and else they are identifiable with probability one for

$$C < \begin{cases} L+\frac{1}{2} & \text{if } C < N_R I \\ \dfrac{1}{1+\frac{1}{2N_R I}}(L+1) & \text{else.} \end{cases} \tag{5.202}$$

Note that (5.202) is the practically relevant case in this work. Interestingly, in case $C \geq N_R I$, condition (5.202) is the same as the condition yielded by arguing that the model's degrees of freedom have to be larger than zero in (5.196). This means that not in every case, we can estimate as many multipath components as we have channel taps $L+1$.
In practice, (5.201) and (5.202) can be interpreted as follows: If the amount of observations

is limited, considering the single snapshot SISO case for any other than the single path scenario, it follows that $C > N_R I = 1$, and consequently in the ambiguous signal case:

$$C < \frac{2}{3}(L+1). \tag{5.203}$$

If we are, however, able to employ a large amount of observations evaluated by observing a whole time series and employing multiple antennas at the receiver side, then, given the channel models assessed in this thesis, we can conclude that $N_R I \geq C$. Then, at least in the idealized error-free case, the number of identifiable multipath components would approach $L + 1/2$.
More realistic, error-prone scenarios, however, might lead to estimator failure, even if the error-free model fulfills the above identifiability conditions. The matrix $\mathbf{\Gamma}$ is determined by the antenna array geometry and the complex fading amplitudes. Therefore, values in $\mathbf{\Gamma}$, which are comparably small due to instantaneous fading effects, might either cause numerical problems due to close-to-singular matrices, or they might be too small to be separable from the additive channel estimation error.
In presence of noise, even less multipath components will be identifiable. Consequently, in realistic scenarios, both $|[\mathbf{\Gamma}]_{c,i}|^2$ and $|[\mathbf{G}(\boldsymbol{\tau})]_{l,c}|^2$ should be significantly larger than the channel estimation error as can be understood from Figure 5.12.

5.16 Parameter Estimation Performance Analysis

5.16.1 The Simulation Setup

Table 5.3: Parameters for the simulations.

Category	*Subcategory*	*Parameter*	*if static*	*if variable*
Channel		Scenario	WINNER B1-LOS	
		C	2	$\in \{1, \ldots, 8\}$
		φ_1, ϕ_1	$\in \mathcal{U}[-\pi, \pi]$	
		v	50 km/h	
		τ_1	$\in \mathcal{U}[0, T]$	
System	Signal	g_{RC} with α_{RC}	0.3	
		T	200 ns	$\in \{50, \ldots, 200\}$
		L_g	7	
	Time/Frequency	f_c	2 GHz	
		J	1	$\in \{1, 2, 4, 8\}$
		I	1	$\in \{1, 100\}$
	Input/Output	N_t	1	
		N_R	1	$\in \{1, 2, 4, 8\}$
		$d_{tx}, d_{Rx}, \lambda/2$	7.5 cm	
Simulation		Runs	1000	

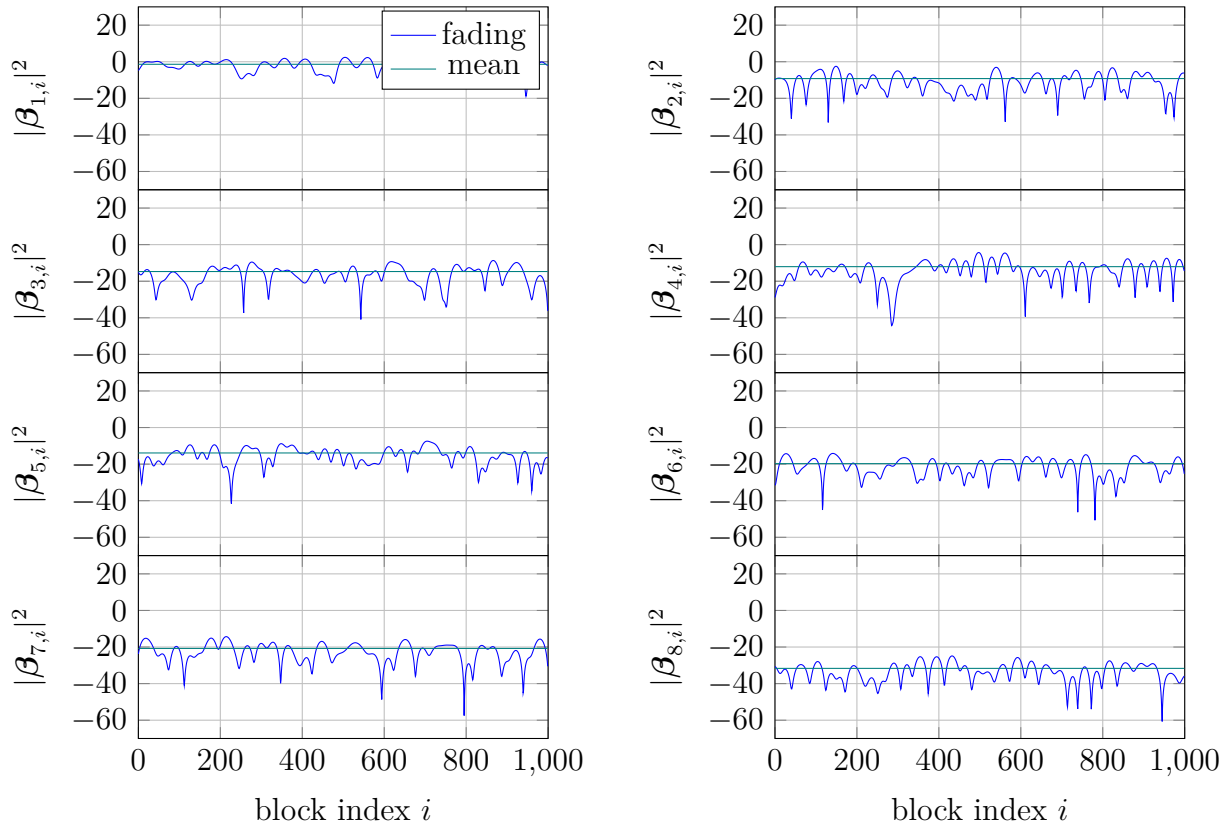


Figure 5.12: The fading plotted for each multipath component c, for the example of a WINNER B1-LOS SISO channel and $v = 50\,\mathrm{km/h}$, shows that with increasing c, due to the increasing attenuation, the sensitivity to noise becomes increasingly severe.

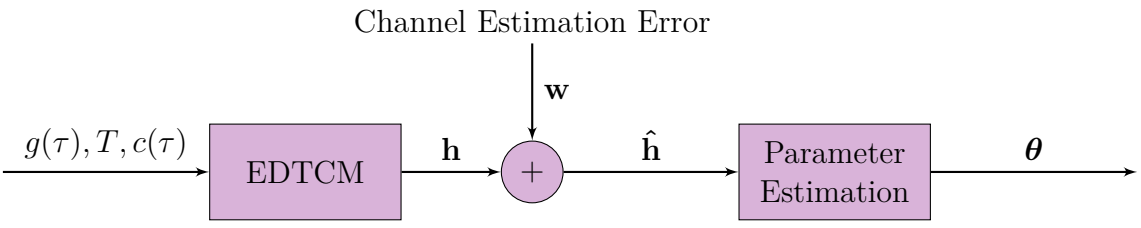


Figure 5.13: Basic simulation setup, valid for different signal designs and channel estimation methods.

To demonstrate how the parameter estimation algorithms applied in this thesis perform, I introduce a basic simulation framework. At this point I will use a generic model as shown in Figure 5.13, applicable to a wide range of communication system structures and channel estimation methods. The single shot measurement EDTCM channel coefficients are directly distorted by the uncorrelated Gaussian distributed channel estimation error, such that the system equation is

$$\hat{\mathbf{h}} = \mathbf{h} + \mathbf{w}. \tag{5.204}$$

Choosing this basic model over a specifically designed JCAP system has the advantage, that the results are valid for any kind of system. The parameter values for the different system categories are tabulated in Table 5.3. Here the "if static" column represents the simulation values for parameters that do not vary in the simulations, the "if variable" column harbors the values for the parameter being toggled in a specific simulation.

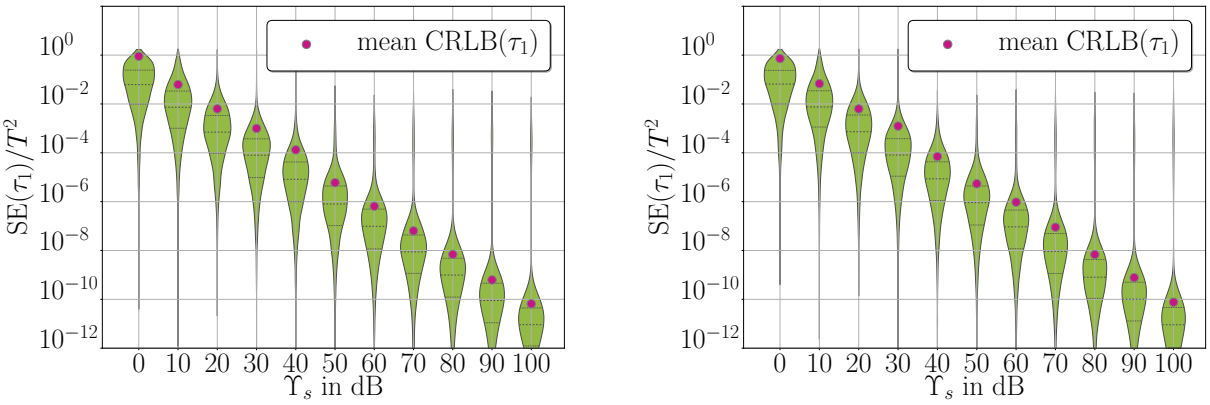
In the following subsections I show squared error distributions expressed via so-called violin-plots. Similar to a boxplot, a violinplot has the purpose to visualize data distributions. The violinplot shows rotated kernel densities to both sides, often shaped as a violin. Instead of summarizing particular statistical features of the data distributions it visually captures the complete distribution. Furthermore, some errors are shown as boxplots. The boxes are shown representing 25 to 75 percent of the data.

5.16.2 SNLLS Estimation via PSO

The results for the SISO single shot measurement case for $C \in \{1, 2, 3\}$ and $T = 200\,\text{ns}$ and for $T = 50\,\text{ns}$ are shown in Figure 5.14. It can be seen that the squared error median converges to the CRLB asymptotically, and hence it can be concluded that the SNLLS parameter estimator optimized with particle swarm optimization works properly. Furthermore, Figure 5.14 confirms that, as expected, the performance varies depending on T as well as on the present number of paths C.

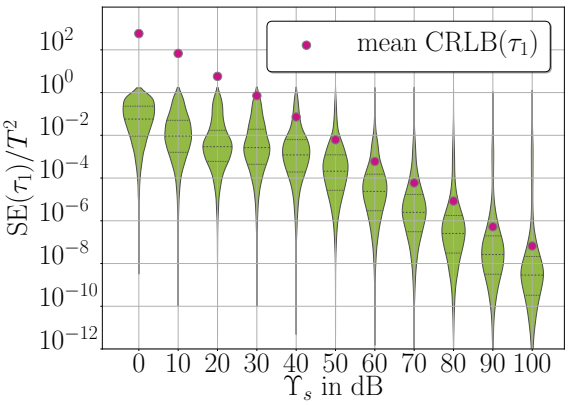
5.16.3 Estimation via ESPRIT

Compared to the PSO results for delay estimation, the ESPRIT-based solution requires oversampling. Even with oversampling it can be seen in the squared error distribution shown in the violin-plots in Figure 5.15, that the solution converges to an error floor with increasing SNR. This error floor is expected due to the finite pulse-induced model mismatch. Nevertheless, one can argue that this model mismatch, in all shown cases, is so low that it still allows a high positioning accuracy - at least for the shown two-path channel model. In the low SNR region the error has a significant amount of outliers. Both effects can be decreased by either increasing the oversampling rate J or by decreasing the sampling duration T (compare Figure 5.15).

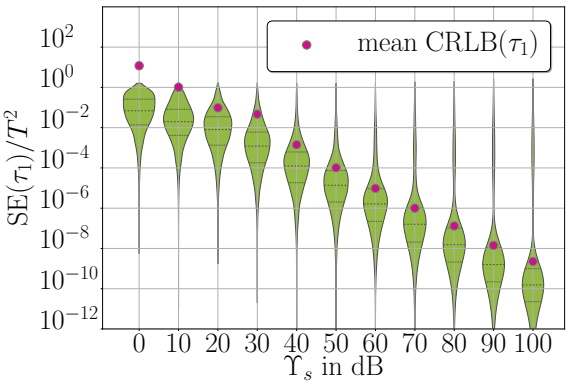


(a) $T = 200$ns, $C = 1$

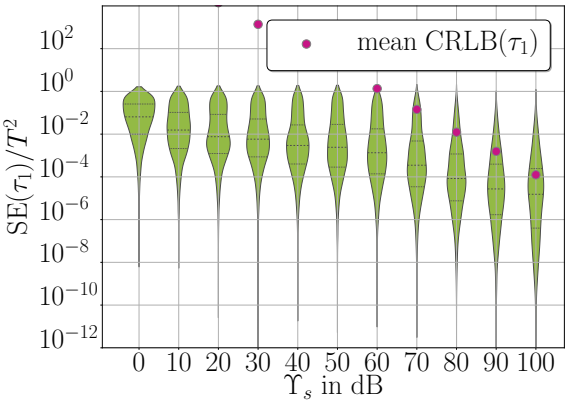
(b) $T = 50$ns, $C = 1$



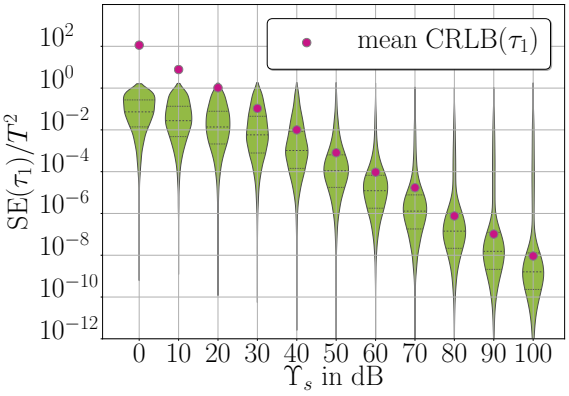
(c) $T = 200$ns, $C = 2$



(d) $T = 50$ns, $C = 2$

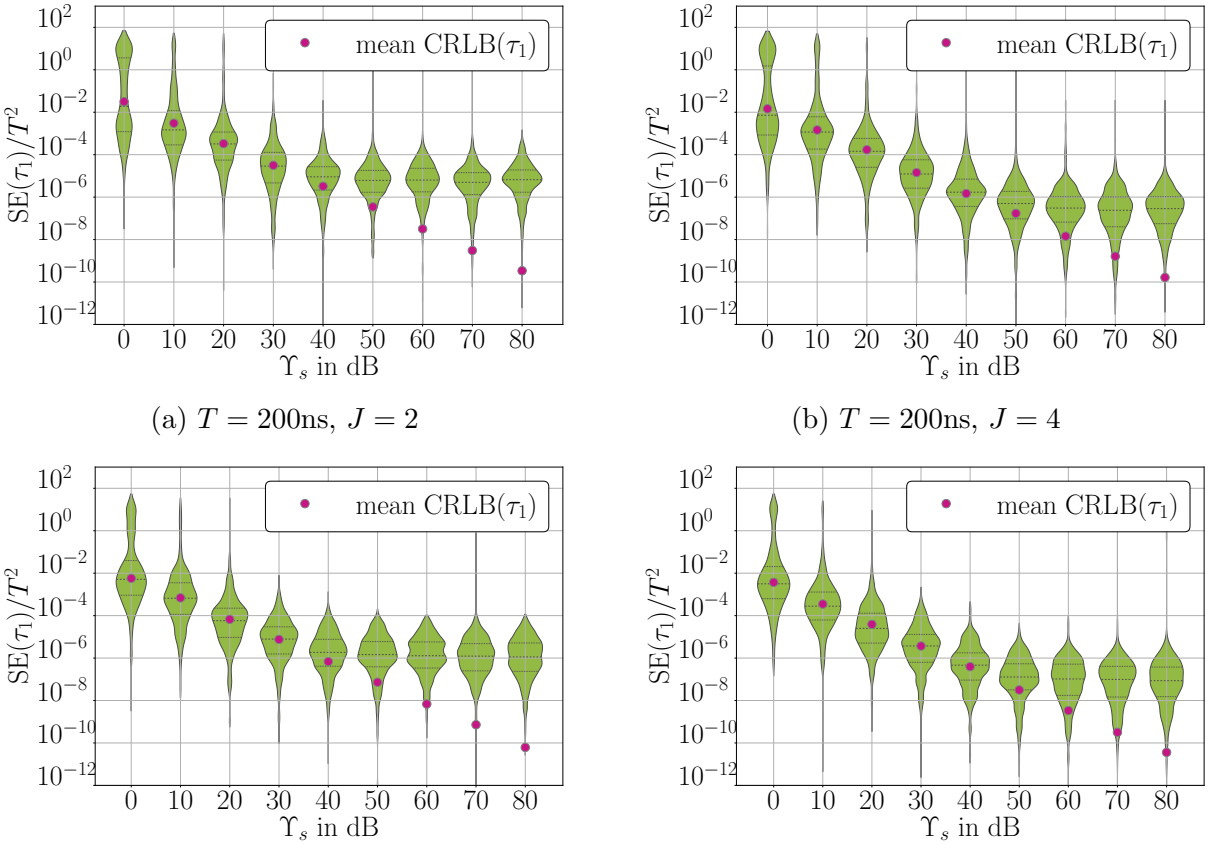


(e) $T = 200$ns, $C = 3$



(f) $T = 50$ns, $C = 3$

Figure 5.14: Violin-plots with inter-quartile ranges depicting the delay estimation squared error distributions show that, ignoring outliers, the performance converges to the CRLB for varying C and T if the PSO algorithm is applied.



(a) $T = 200$ns, $J = 2$

(b) $T = 200$ns, $J = 4$

(c) $T = 100$ns, $J = 2$

(d) $T = 100$ns, $J = 4$

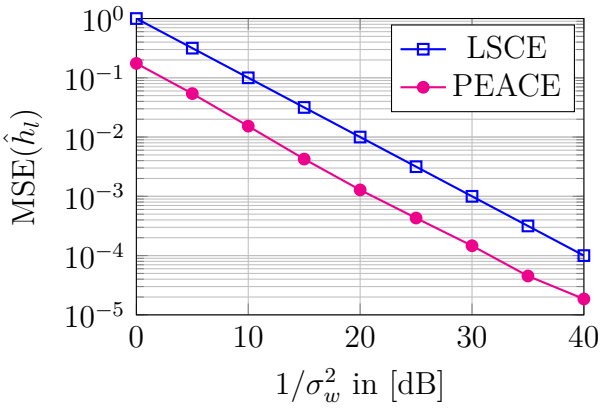
Figure 5.15: The $\mathrm{SE}(\tau_1)/T^2$ violin-plots show twofold: Firstly that the CRLB is approached for higher SNRs and secondly that due to the model mismatch caused by an finite pulse an error floor is approached in the high SNR regime, for the ESPRIT algorithm. This error floor can either be lowered by oversampling with higher frequency or by choosing a smaller sampling duration, which yields an easier-to-solve parameter estimation problem.

5.16.4 Parameter-Aided Channel Estimation

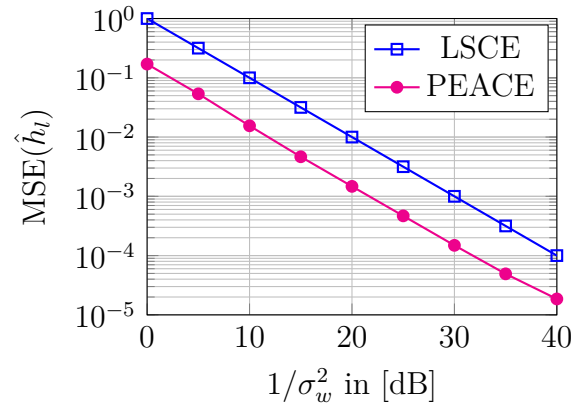
It is straightforward that, after parameter estimation has been performed, improved parameter estimation-aided channel estimates (PEACE) can be calculated via

$$\hat{\mathbf{h}}^{PEACE} = \mathbf{G}(\hat{\boldsymbol{\tau}})\hat{\boldsymbol{\beta}}. \tag{5.205}$$

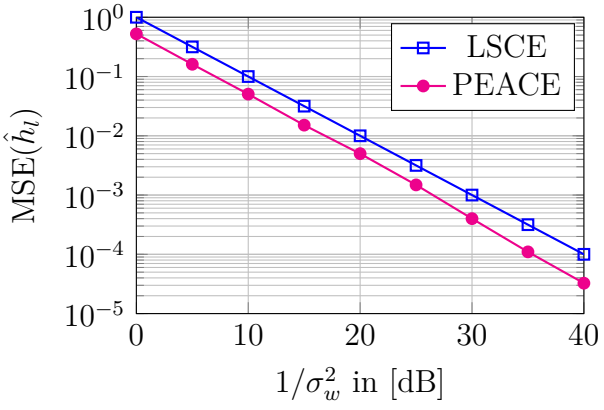
Comparative results for a sampling duration of $T = 200\,\mathrm{ns}$ and $T = 50\,\mathrm{ns}$ and for a single-path and a two-path channel are shown in Figure 5.16. The performance gain is visible. Nonetheless, it is less pronounced for the two-path channel than for the single-path channel. With an increasing number of paths the error for the PEACE estimation error will obviously increase due to the higher problem dimensionality versus the same number of observations.



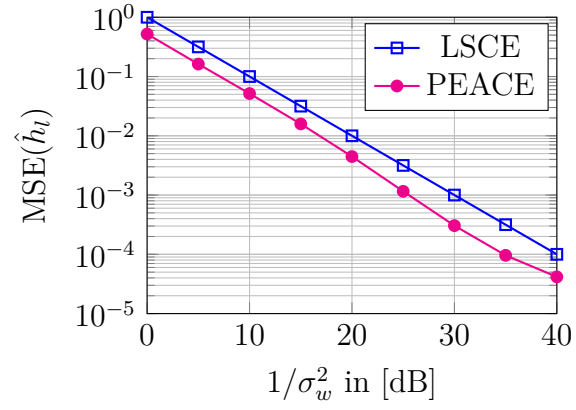
(a) $C = 1$, $T = 200\,\text{ns}$



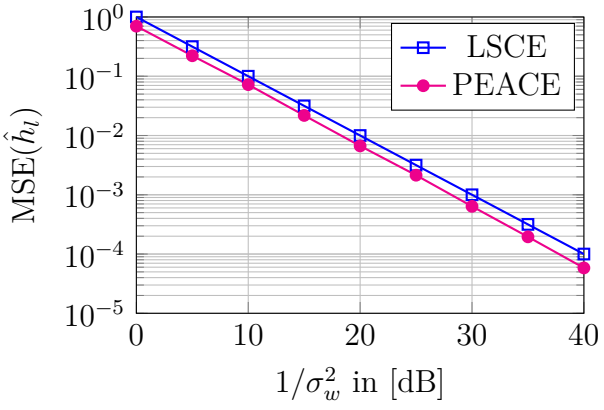
(b) $C = 1$, $T = 50\,\text{ns}$



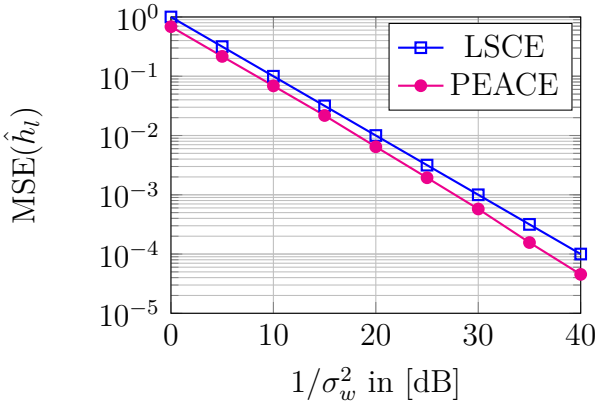
(c) $C = 2$, $T = 200\,\text{ns}$



(d) $C = 2$, $T = 50\,\text{ns}$



(e) $C = 3$, $T = 200\,\text{ns}$



(f) $C = 3$, $T = 50\,\text{ns}$

Figure 5.16: The parameter estimation aided channel estimation clearly outperforms the EDTCM channel estimation. The more multipath components added, the less pronounced is the improvement.

5.16.5 Detecting Estimation Success

In order to assess the estimation success and failure detection performance we need to define the term success precisely to use it as a reference. The delay estimation squared error should be lower than the squared error of the coarse estimation step obtained by performing purely correlation-based delay estimation. Demanding this would be one reasonable choice to classify parameter estimation performance as "succesfull". Nevertheless, taking into account that the purely correlation-based performance is known to be not more accurate than $T/2$, it is reasonable to demand a higher precision to be classified as "successful". Another "success" and "failure" classification is, to demand that, in case many estimates are observed, the squared error lies within a predefined percentile range, that is, the squared error should not be marked as an outlier when being visualized via a boxplot. Comparing two box plot visualizations of the same measurements, without and with estimation success detection in the Figure shows that the extreme outliers are found via the success estimation technique I propose here. If the method is not applied, it can be seen that the delay estimation MSE asymptotically converges to an error floor, due to the outliers. If, on the other hand, the method is applied, the MSE asymptotically converges to the CRLB. In Figure 5.17 b) it becomes clear that the χ^2-based success estimation efficiently filters in the high SNR regime, whereas the strategy to filter outliers that have converged to the search space bound works well in the low SNR regime. Both methods utilize a reliability threshold. Joining both methods determines the estimation success for all SNRs. The medium SNR regime shows a minimum, since this is the area for that both methods are still not as effective as desired.

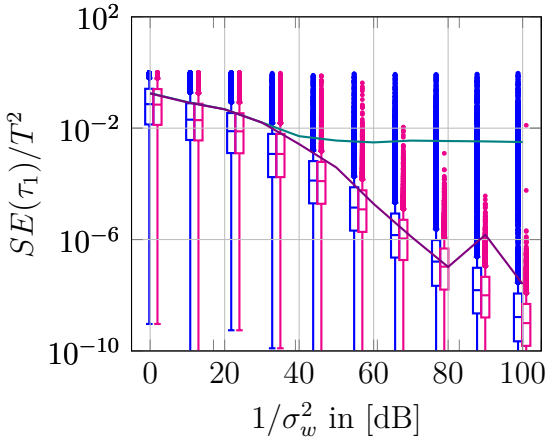
5.16.6 MIMO Channel Parameter Estimation

For delay estimation increasing the number of exploitable observations yields significant performance gains. This is either achieved by oversampling, MIMO or time series measurements. More MIMO results will be shown for joint model order selection and parameter estimation.

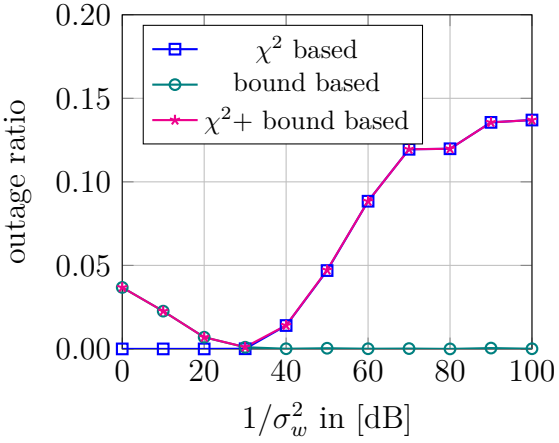
5.17 Positioning Algorithms

In Chapter 2 I discussed a range of positioning principles. Previously, the TOA method was already outlined in Subsection 2.2.1. Here, channel estimates are used to determine the TOA in a first step (together with a coarse delay estimation). Afterwards the TOA estimates for all reference objects are then used in a second step in the positioning algorithm. The distances used in the positioning algorithm are calculated by substituting the delays in (2.1) by delay estimates. The relationship between the distances and the position was given by (2.2) and the position estimation itself was already stated in (2.5). I will shortly introduce the investigated positioning algorithms that will be used in Chapter 7. In the presented results I calculate the positioning error as in (2.5):

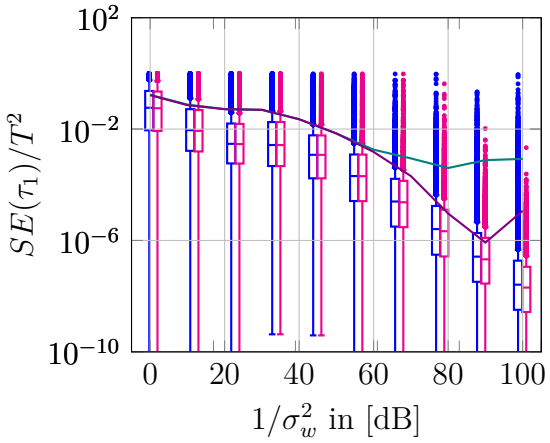
$$\hat{\mathbf{p}} = \arg\min_{\tilde{\mathbf{p}}} \left\{ \left(\hat{\mathbf{d}} - \mathbf{d}(\tilde{\mathbf{p}}) \right)^T \mathbf{C}_{\boldsymbol{\epsilon}}^{-1} \left(\hat{\mathbf{d}} - \mathbf{d}(\tilde{\mathbf{p}}) \right) \right\}.$$



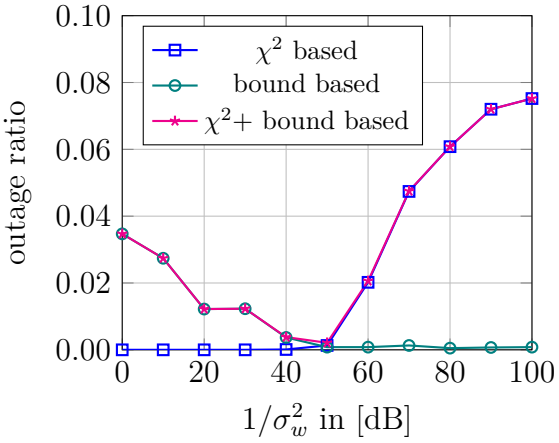
(a) SEs, $T = 50\,\text{ns}$, $C = 2$



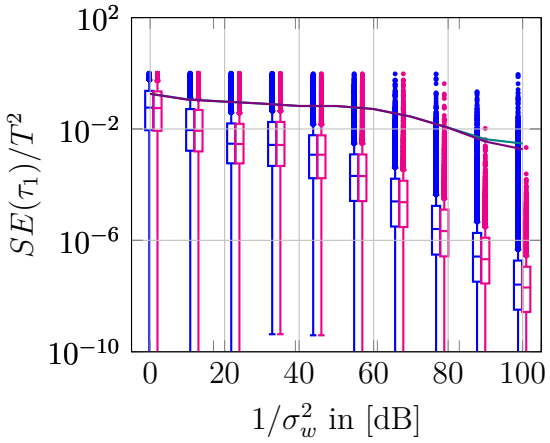
(b) Outage, $T = 50\,\text{ns}$, $C = 2$



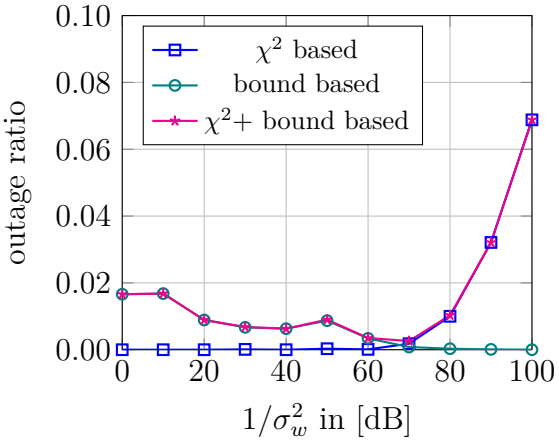
(c) SEs, $T = 200\,\text{ns}$, $C = 2$



(d) Outage, $T = 200\,\text{ns}$, $C = 2$

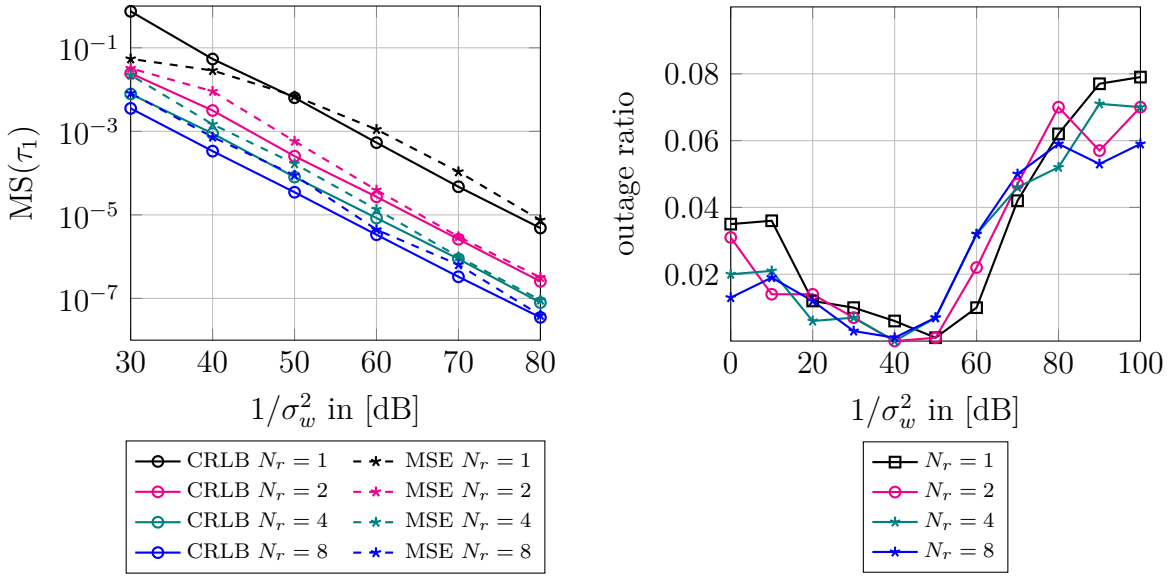


(e) SEs, $T = 200\,\text{ns}$, $C = 3$



(f) Outage, $T = 200\,\text{ns}$, $C = 3$

Figure 5.17: The magenta-colored box-plots depict the results for detecting estimation success and dismissing unsuccessful estimates, whereas the blue box plots are the reference results without detecting estimation success. The violet and teal-colored lines show the resulting MSE in case success detection + dismissing failed estimates is applied and in case it is not applied. It can be seen that critical outliers are asymptotically successfully filtered with this detection, particularly in the high SNR regime.



(a) $T = 200\,\text{ns}$, $C = 2$ (b) Outage, $T = 200\,\text{ns}$, $C = 2$

Figure 5.18: The MSE results show a decreasing MSE with increasing number of employed receive antennas N_R. For $N_R > 2$ a performance gain of 3 dB can be seen. The results do not indicate any influence on the outage rate.

An LM-algorithm-based iterative non-linear method, introduced as the Taylor series algorithm, is known to provide accurate positioning results [GZT08]. Alternatively, or as the input for the initial guess for the mentioned Taylor series algorithm, a suboptimal closed-form approach can be applied as well [STK05, GZT08].

5.17.1 Weighted Least-Squares Algorithm

The algorithm builds upon the idea to transform the non-linear equations into linear equations by ignoring the distance error. The authors of [STK05] explain in detail how this transformation is achieved. The linearized problem formulation can be outlined as follows:

$$\boldsymbol{\mathcal{A}}\mathbf{p} = \mathbf{b} \quad \text{with} \tag{5.206}$$

$$\boldsymbol{\mathcal{A}} = \begin{bmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \\ \vdots & \vdots \\ x_{N_B} - x_1 & y_{N_B} - y_1 \end{bmatrix}, \tag{5.207}$$

$$\mathbf{b} = -\frac{1}{2} \begin{bmatrix} \hat{d}_2^2 - \hat{d}_1^2 - (x_2^2 + y_2^2) + (x_1^2 + y_1^2) \\ \hat{d}_3^2 - \hat{d}_1^2 - (x_3^2 + y_3^2) + (x_1^2 + y_1^2) \\ \vdots \\ \hat{d}_{N_B}^2 - \hat{d}_1^2 - (x_{N_B}^2 + y_{N_B}^2) + (x_1^2 + y_1^2). \end{bmatrix} \tag{5.208}$$

The straightforward closed form solution to this problem then reads:

$$\hat{\mathbf{p}} = \left(\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{A}}\right)^{-1} \boldsymbol{\mathcal{A}}^T \mathbf{b}. \tag{5.209}$$

The soft information estimated here for the base station delay estimates can be used to improve the positioning accuracy via the weighting matrix

$$\mathbf{W} = \text{diag}\left(\frac{1}{c^4\sigma^4_{\epsilon_2}}, \ldots, \frac{1}{c^4\sigma^4_{\epsilon_{N_B}}}\right). \tag{5.210}$$

5.17.2 Taylor-Series Algorithm

The Taylor-series (TS) algorithm aims at solving the problem (2.5), via for instance, the Levenberg-Marquardt method that is based on a Taylor series approximation. The iterative position updates are obtained similarly to the introduction in Section 5.7.4, via

$$\mathbf{J}_{\Omega^{\text{pos}}}(\tilde{\mathbf{p}}^i) = -2 \cdot \mathbf{J}^T(\tilde{\mathbf{p}}^i)\mathbf{C}_\epsilon^{-1}(\hat{\mathbf{d}} - \mathbf{d}(\tilde{\mathbf{p}}^i)), \tag{5.211}$$

$$\mathbf{H}_{\Omega^{\text{pos}}}(\tilde{\mathbf{p}}^i) = 2 \cdot \mathbf{J}^T(\tilde{\mathbf{p}}^i)\mathbf{C}_\epsilon^{-1}\mathbf{J}(\tilde{\mathbf{p}}^i), \tag{5.212}$$

$$\tilde{\mathbf{p}}^{i+1} = \tilde{\mathbf{p}}^i - \left(\mathbf{H}_{\Omega^{\text{pos}}}(\tilde{\mathbf{p}}^i) + \eta\mathbf{I}\right)^{-1} \cdot \left(\mathbf{J}_{\Omega^{\text{pos}}}(\tilde{\mathbf{p}}^i)\right)^T. \tag{5.213}$$

5.17.3 Positioning CRLB and Geometric Dilution of Precision (GDOP)

Let us consider a 2D scenario and let $\mathbf{F}_\mathbf{p}$ denote the positioning Fisher information matrix. Further, let $\mathbf{J}(\mathbf{p})$ denote the Jacobian matrix, with

$$\mathbf{J}(\mathbf{p}) = \frac{\partial \mathbf{d}(\mathbf{p})}{\partial \mathbf{p}^T}, \quad \text{with} \tag{5.214}$$

$$\frac{\partial d_i(\mathbf{p})}{\partial x_{MS}} = \frac{x_{MS} - x_i}{\sqrt{(x_{MS} - x_i)^2 + (y_{MS} - y_i)^2}}, \tag{5.215}$$

$$\frac{\partial d_i(\mathbf{p})}{\partial y_{MS}} = \frac{y_{MS} - y_i}{\sqrt{(x_{MS} - x_i)^2 + (y_{MS} - y_i)^2}}. \tag{5.216}$$

Following the standard calculations of determining the CRLB, the positioning CRLB in this thesis reads:

$$\text{crlb}(\mathbf{p}) = \text{trace}\{\mathbf{F}_\mathbf{p}^{-1}(\mathbf{p}\}) = \text{trace}\left\{(2\mathbf{J}^T(\mathbf{p})\mathbf{J}(\mathbf{p}))^{-1}\right\}. \tag{5.217}$$

Furthermore, a measure for the geometric effect that the base station arrangement impacts the positioning accuracy, is called geometric dilution of precision (GDOP)[Lan99, Zhu92]. It is determined by assuming an identical estimation error varinace $\sigma^2_{epsilon} = 1$ on all pseudo-ranges, such that

$$\text{GDOP}(\mathbf{p}) = \sqrt{\text{CRLB}(\mathbf{p})|_{\mathbf{C}_\epsilon = \mathbf{I}}}. \tag{5.218}$$

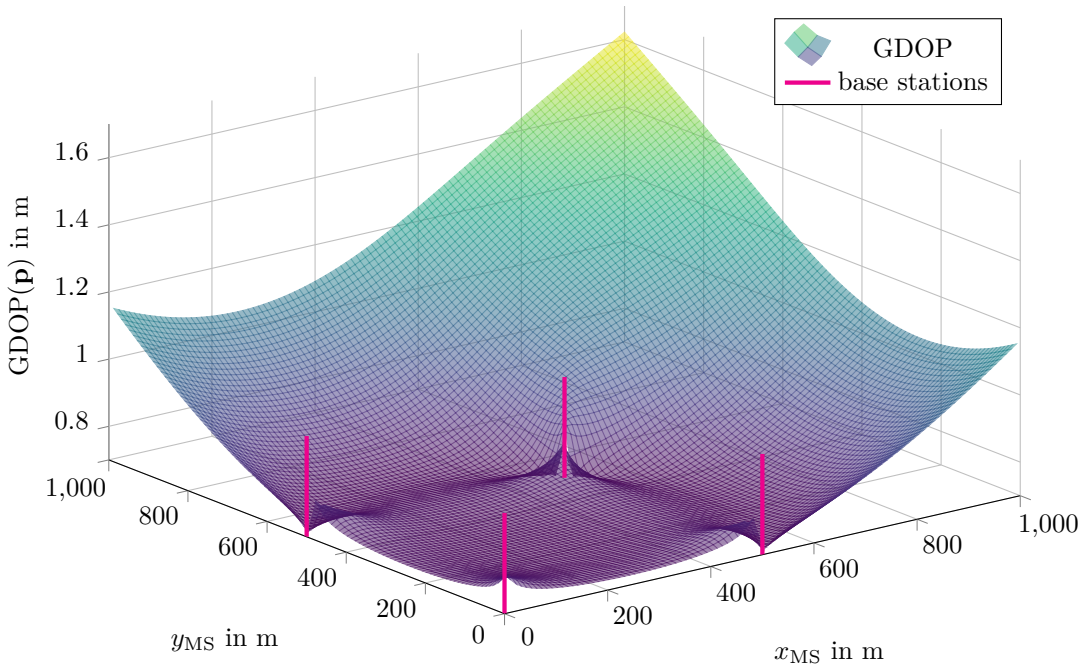


Figure 5.19: The GDOP map, exemplarily shown for a square array of four base stations, demonstrates that the GDOP is minimal at the center between all base stations, the so-called GDOP valley. Outside the square region defined by the base stations the GDOP becomes very large.

Figure 5.19 exemplarily visualizes the GDOP in a, for this thesis, typical scenario of four base stations arranged in a square array. It demonstrates that the GDOP is minimal at the center between all base stations, the so-called GDOP valley. Outside the square region defined by the base stations, the GDOP potentially becomes very large, severely affecting the positioning accuracy. The figure shows how the geometry impacts the GDOP, which is high in difficult geometric setups (MS lies outside the area spanned by the base stations) and lower in optimal geometric setups (MS lies within the area spanned by the base stations).

Chapter 6

Model Selection

Parts of this chapter are currently submitted for publication and under review in [AH20].

The extraction of positioning-relevant information, like the time or angle of arrival, from the channel estimates, requires high-resolution parameter estimation. To perform that task, knowing the number of multipath components is a mandatory prerequisite. Unfortunately, this information is not available at the receiver and therefore has to be estimated. Apart from that, the parameter estimation error and hence the positioning accuracy depends on the chosen model and model order. Choosing the correct model order does not always lead to the best estimation performance (see, for instance, pages 74 and 124 in [Sch12]). This dependence provides another reason to assess model selection for this framework. Either model order estimation has to be performed before, that is separately, or simultaneously, in a joint manner with parameter estimation. Therefore, it is interesting to study how the interplay of model selection and parameter estimation impacts the estimation performance. Model order estimation classifies as a typical model selection problem. The range of available algorithms accumulated over the last decades related to the subject of model selection is broad. Surprisingly, many of these algorithms are derived based on different theoretical approaches and therefore have slightly different interpretations. The range covers well known and hence matured strategies employing classical information-theoretic criteria [Aka73, Aka74, Sch78, Ris78, HT89, WK85, WZ89b, Saa98, BA02] and optimized extensions thereof [Boz00] as well as an approach that relates the model order to a known least-squares error distribution [QY97] and algorithms that select the model order by determining and thresholding the components relevance [Chu05, CM08, Ric05, SS15]. The latter strategies involve defining some more or less subjective, albeit efficient, threshold, based on the present amount of noise or user-specified tuning parameters. Those other criteria, based on utilizing different information-theoretic aspects, were derived by employing a robust theoretical framework and hence promise to be more objective.

Each model selection method is constrained to a set of assumptions regarding the underlying system. Hence, not every method is genuinely applicable to every model selection problem at hand. Therefore, model selection like parameter estimation should be designed problem-specifically by evaluating the signal structure and further taking into account the system-specific envisaged goals.

Engaging in the basics of information-theoretic criteria [WK85, WZ89b], one soon

learns that at least asymptotically (with a high SNR and a high number of measurements versus the number of unknowns) some estimators are proven to be consistent, meaning that the estimated model order approaches the correct model order. In our case the model order is given by the number of multipath clusters. Consistency would then mean that $\hat{C} = C$. Nevertheless, as I already mentioned earlier, $\hat{C} = C$ does not necessarily implicate the best achievable parameter estimation performance in terms of MSE. Nevertheless, any other choice than $\hat{C} = C$ inevitably leads to a modeling error and hence an asymptotically occurring parameter estimation MSE error floor. Still, depending on the setup, this modeling error can lead to a better performance than a more complex, albeit accurate, model order. Consequently, we first of all ask ourselves, whether this asymptotical behavior and hence consistency is realistic for the system setup assessed in this thesis.

This thesis aims to design a JCAP system, which is not necessarily supposed to operate under ideal asymptotical conditions (high SNR plus a relatively high number of observations versus the number of free unknowns), but instead to robustly and accurately provide parameter estimates given both high and low SNR and any number of observations. The following scenario is likely: A reasonable SNR can be assumed and a rather low number of observations versus the number of parameters, which are supposed to be estimated. In the designed JCAP system the number of complex-valued measurements, or overall observations, is determined by the number of channel coefficients $L + 1$, the oversampling factor J, the number of transmit antennas N_t, the number of receive antennas N_r as well as the number of blocks I. The number of channel coefficients $L + 1$ usually ranges between $10 - 20$. Oversampling is not desirable at all. Hence, if not specified otherwise, I assume $J = 1$. Furthermore, since it is not desired to limit the system design to a minimum number of transmit or receive antennas, I assume $N_t = N_r = 1$, if not specified differently. Additionally, I assume that the single measurement case $I = 1$ is the standard case. In this work, the different applied WINNER channel scenarios determine the number of estimated parameters. The actual number of paths C can vary between $8 - 20$. The number of parameters estimated is a multiple of C. These numbers demonstrate that, concerning the standard scenario investigated in this thesis, the ratio between the number of observations and the number of unknowns in the resulting parameter estimation problem cannot be considered a large-sample scenario. Therefore, it cannot be considered as a scenario leading to asymptotical performance when it comes to applying information-theoretic criteria. The parameter estimation MSE monotonically increases with an increasing number of unknowns. The dissertation [Sch12] shows that assuming the actual number of unknowns ($\hat{C} = C$) can lead to such high MSEs that the reader can interpret these settings as a failed receiver-sided modeling and hence as a failed estimation. It further shows that assuming a constant $\hat{C} = 2$ instead already can lead to a reasonably low parameter estimation MSE, but it appears unreasonably harsh to do so. Going further, this leads to the idea that obtaining robust high-resolution parameter estimates is most likely achievable by jointly performing model order estimation and parameter estimation. This idea, in return, leads to the question of how to perform model selection or joint model selection and parameter estimation in an optimal way regarding the presented and assessed JCAP design. In the following, I will therefore briefly

summarize the classical information-theoretic criteria for model selection.

6.1 Information-Theoretic Criteria

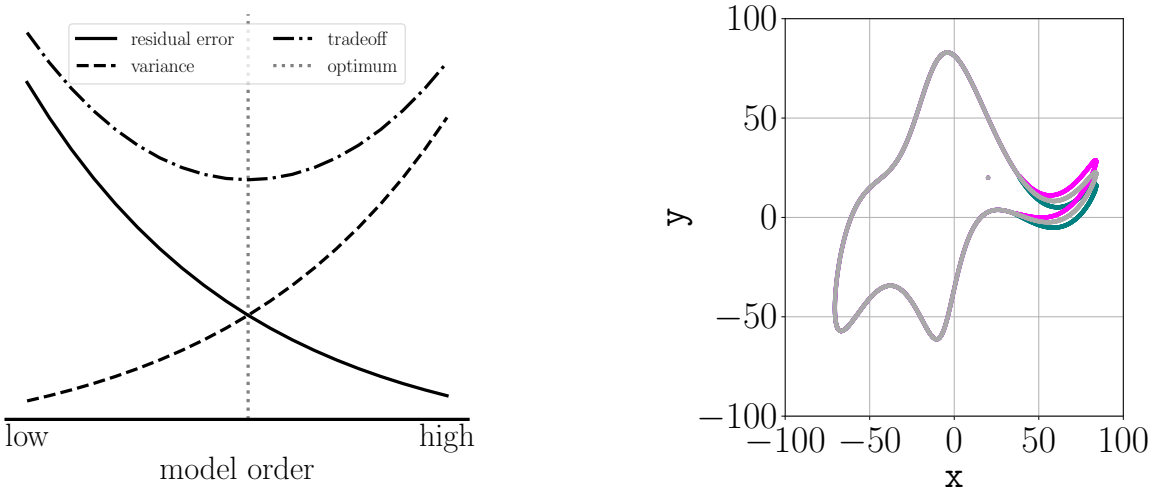


Figure 6.1: The left plot shows that the residual error decreases with increasing model order, whereas the estimates error variance increases with increasing model order. A tradeoff function can be used to find the optimum model order. The right plot shows the implementation of [MKH10] depicting an elephant constructed of 4 complex parameters and 5, in case the trunk wiggles. The modelled elephant refers to the famous quote [Dys04] and visualizes that given a certain number of parameters the possibility to model an arbitrary shape increases and therefore gives an intuitive feeling for the tradeoff between under and overfitting.

Let $\hat{\boldsymbol{\theta}}_{\tilde{C}}$ denote the estimated parameter vector for all hypothetical model orders $\tilde{C} \in \{1, \ldots, C_{max}\}$

$$\hat{\boldsymbol{\theta}}_{\tilde{C}} = \arg\min_{\tilde{\boldsymbol{\theta}}_{\tilde{C}}} \{\Omega_{\tilde{C}}(\tilde{\boldsymbol{\theta}}_{\tilde{C}})\}. \tag{6.1}$$

It is important to understand that the maximum likelihood, in case of noisy measurements, monotonically increases with $\tilde{C}$ and hence the negative likelihood monotonically decreases with $\tilde{C}$. Therefore, minimizing over all possiblehypothetical optimal parameter estimation solutions alone neither yields the true model order nor the optimal estimation performance. Firstly, minimization over all estimated solution candidates always yields the largest possible order C_{max}

$$\arg\min_{\tilde{C}} \{\Omega_{\tilde{C}}(\hat{\boldsymbol{\theta}}_{\tilde{C}})\} = C_{max} \geq C. \tag{6.2}$$

This can be explained by the fact that increasing the candidate model order leads to overfitting: When more components $\tilde{C} > C$ than actually have been used to create the

measurements are used as a candidate solution like in (6.2) the extra components, which exceed C will model noise as components and therefore virtually suggest a better model fit, that is, a higher likelihood, or equivalently, a smaller least-squares residual value. Secondly, it is important to note that, although the data is fitted tighter to the measurements by increasing $\tilde{C}$, the estimation error increases at the same time, due to the estimator's higher degree of sophistication. It is genuinely known that the theoretical parameter estimation covariances increase with increasing model complexity, that is, with increasing $\tilde{C}$. Conseqeuntly, some kind of tradeoff between goodness of fit and model order complexity is required.

Principle of Parsimony as Basis to Classical Model Selection

Proposed by [MKH10], Figure 6.1 provides an illustrative implementation to the famous quote [Dys04] of von Neumann "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk." In 1975, [Wei75] already had successfully modelled an elephant with 30 parameters. The modeled elephant visualizes the underlying problem in model selection, which is to find the optimal tradeof between under- and overfitting. Classical model order selection methods are designed to maximize the likelihood, while preventing overfitting. According to [BA02], Albert Einstein supposedly once said "everything should be made as simple as possible, but no simpler." In the context of mathematical modeling this point of view agrees with the by William Occam proposed concept of parsimony "to shave away all that is unnecessary", nowadays referred to as Occam's razor [BA02]. Thus, the underlying concept is to find a good fit and at the same time to strive for parsimony in a way that is somehow optimal. Finding the optimal tradeoff tailored to fulfill specific estimation problem requirements and contraints is the core challenge when it comes to (joint) parameter estimation and model selection. Commonly, this optimal tradeoff is achieved by adding a suitable penalty term to the likelihood.

Consider an arbitrary parameter vector $\boldsymbol{\xi}$ with the model order m, containing k free adjustable parameters and any observed data matrix $\mathbf{X}$, containing N independent measurements. Let $\mathcal{L}(\mathbf{X}|\boldsymbol{\xi})$ be the associated likelihood. Typically, an information theoretic criterion embodies the optimal tradeoff between adding the negative log-likelihood and a model-order-dependent penalty term. It can be used for joint parameter estimation and model order selection in the following manner:

$$\mathcal{C}\left(\boldsymbol{\xi}_m\right) = -\ln\left(\mathcal{L}(\mathbf{X}|\boldsymbol{\xi}_m)\right) + \mathcal{P}\left(m\right), \tag{6.3}$$

$$\hat{\boldsymbol{\xi}}_m = \arg\min_{\tilde{\boldsymbol{\xi}}} \left\{-\ln\left(\mathcal{L}(\mathbf{X}|\boldsymbol{\xi})\right)\right\}, \tag{6.4}$$

$$\hat{m} = \arg\min_{\hat{\boldsymbol{\xi}}_m} \left\{\mathcal{C}\left(\hat{\boldsymbol{\xi}}_m\right)\right\}. \tag{6.5}$$

The most popular information-theoretic criteria for model order selection known from literature were proposed by Akaike [Aka73, Aka74], who proposed the so-called Akaike

information criterion (AIC), and by Schwartz [Sch78] and Risannen [Ris78], who parallelly proposed the minimum description length criterion (MDL). These criteria aim at minimizing a criterion-dependent value with respect to the number of the problem's free adjustable parameters. This value combines the model's likelihood function and some term combatting overfitting. Both information-theoretic criteria are similar in the sense that they all were inspired by the philosophy of parsimony, while maintaining goodness of fit. Nonetheless, although they were derived basing on rather different strategies and backgrounds, they mainly differ in the choice of the penalty added to the likelihood term. These different penalty terms may, in all cases, be interpreted as terms penalizing the number of estimated parameters. Let $p(\cdot)$ denote the probability density function.

- **Akaike's Information Criterion**
  Akaike's information criterion is given by [Aka73]

$$\mathcal{C}_{-\mathrm{AIC}}(k) = -2\ln\left(\mathcal{L}\left(\mathbf{X}|\hat{\boldsymbol{\xi}}\right)\right) + 2k. \tag{6.6}$$

  This criterion origins in the idea to minimize the Kullback Leibler divergence between any hypothesis model and the true model. In [BA02], the Kullback-Leibler information or distance is explained as "the information lost" when any model g is used to approximate any other model f. Since the true model and the optimal hypothetical parameters, and therefore the true Kullback-Leibler distance are not known in the process of model selection, the hypothetical optimal parameters have to be replaced by estimates. Instead of minimizing the actual Kullback-Leibler distance the expected Kullback-Leibler distance is minimized in Akaike's information criterion. The employed maximized log-likelihood is biased and this bias can be approximated by the number of estimated parameters k [Aka74]. The criterion is valid in case of large samples and "good models", according to [BA02]. Nevertheless, estimates yielded by employing this criterion have a tendency to be overfitted. This can be explained by the fact that with increasing sample sizes, the likelihood cost term in (6.6) increases whereas the penalty term doesn't. Therefore with increasing sample sizes penalizing the number of estimated parameters looses its weight in the overall cost function.
  Another modified version, the so-called second-order version of Akaike's information criterion (c-AIC) [HT89] was designed for small sample sizes:

$$\mathcal{C}_{\mathrm{c-AIC}}(k) = -2\ln\left(\mathcal{L}\left(\mathbf{X}|\hat{\boldsymbol{\xi}}\right)\right) + 2k + \frac{2k(k+1)}{N-k-1}. \tag{6.7}$$

- **Minimum Description Length Criterion**
  The minimum description length is given by [Sch78, Ris78]

$$\mathcal{C}_{\mathrm{MDL}}(k) = -\ln\left(\mathcal{L}\left(\mathbf{X}|\hat{\boldsymbol{\xi}}\right)\right) + \frac{1}{2}k\ln I. \tag{6.8}$$

  Although similarities regarding the structure of both AIC and MDL are apparent, the derivations of both criteria are based on rather different approaches. Whereas

the idea behind the AIC was to minimize the information lost between the estimated version of a true model and the hypotheses, the minimum description length criterion aims at choosing the model providing the shortest description of measurements and has its origin in the area of coding theory. The description length in that context may be understood as the number of digits used to code binary data streams. Risannen shows in [Ris78] that describing data can be interpreted as coding. The penalty here not only depends on number of parameters that have to be estimated, but also on the sample size I. In contrast to the AIC, MDL does not tend to overfit asymptotically.

- **Information Complexity Criterion**
  All the criteria above have in common that they allow a reasonable tradeoff regarding goodness of fit and parameter complexity. Unfortunately, the criteria above do not directly relate their complexity measure with the delay estimate error variances in $\mathbf{C}_{\hat{\boldsymbol{\xi}}}$, although this would be the most reasonable choice to make. Choosing the complexity measure for the penalty based on the resulting estimation error variance and its correlation provides a natural method to decide how many components are a reasonable choice. A criterion, handling this dependency is the information complexity criterion:

$$\mathcal{C}_{\text{ICOMP}} = -2\ln\left(\mathcal{L}\left(\mathbf{X}|\boldsymbol{\xi}\right)\right) + 2\frac{k}{2}\ln\left(\frac{\text{tr}(\mathbf{C}_{\boldsymbol{\xi}})}{k}\right) - \frac{1}{2}\ln|\mathbf{C}_{\boldsymbol{\xi}}|. \tag{6.9}$$

  Let λ_i denote the ith eigenvalue of $\mathbf{C}_{\hat{\boldsymbol{\xi}}}$ for $1 \leq i \leq m$. Acknowledging that the following two properties are valid:

  Property 5. *For any square matrix $A \in \mathbb{C}^{n\times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ [PP12]:*

$$\text{tr}\left(A\right) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i, \tag{6.10}$$

$$\det\left(A\right) = \prod_{i=1}^{n} a_{ii} = \prod_{i=1}^{n} \lambda_i. \tag{6.11}$$

  Consequently, the criterion (6.9) can more compactly be written as

$$\mathcal{C}_{\text{ICOMP}} = -2\ln\left(\mathcal{L}\left(\mathbf{X}|\boldsymbol{\xi}\right)\right) + k \cdot \ln\left(\frac{\frac{1}{k}\sum_{j=0}^{k}\lambda_j}{\left(\prod_{j=1}^{k}\lambda_j\right)^{1/k}}\right). \tag{6.12}$$

The presented information theoretic criteria above have one assumption in common: They assume that the sample size I, a number of independently identically distributed time series measurements, is large. The special case, $I = 1$ is not considered at all, due to the underlying large-sample assumption. Note that for the MDL in (6.8) the penalty term

would vanish in case $I = 1$ and the criterion would not make any sense. Other model order estimation approaches do not require a large amount of independent measurements like [QY97, Ric05], and hence can be applied for a single snapshot measurement as well as in the large-sample case.

Further note that in this thesis we are interested in an optimal positioning performance and hence we seek for the information-theoretic criterion, or model selection method, yielding the best positioning performance in the end. Obviously, the best positioning performance is directly related to the best TOA estimation performance in terms of MSE.

Given this special perspective, overfitting should be prevented. For reliable positioning results, it is crucial to determine reliable, accurate TOA delay estimates. Overfitting could lead to modeling noise and mistakenly interpreting it as the TOA. Under-fitting on the other hand, will most likely lead to exclusion of the excess delays, having weaker path weights than the TOA. Hence, from an optimal positioning performance point of view, overfitting is worse than under-fitting. Overfitting, as already mentioned, either means modeling noise as system components or indicates path-splitting. Path splitting occurs, when one actual path is split into two half-energy paths with the same delay. If noise is modelled as an actual component, this can accidentally happen at the smallest estimated delay, which then is automatically and falsely interpreted as the TOA and therefore will lead to an undesirable TOA MSE. Path splitting is also not desirable, since it leads to over complex calculations, less accuracy, unnecessarily high loads and unpredictable run times.

Due to these considerations, we rule out Akeike's information criterion and variants thereof, since it is known to overfit, even asymptotically. This leaves us with the MDL criterion, the information-complexity criterion, or with choosing a threshold-based, more subjective, method to solve the problem. Theoretically, comparing the penalty terms of the MDL and the ICOMP criterion, ICOMP promisses to choose the model order most judiciously. Nevertheless, ICOMP has disadvantages in terms of practical considerations: First of all: the problem's Fisher information matrix has to be known and then it also has to be inverted for all hypotheses. Depending on the parameter vector size, this matrix can be large and hence the computational complexity could become large as well. Another point is that in this thesis I actually do not aim to solve the overall problem judiciously, but to choose that model order that results in the lowest TOA error. Although ICOMP is the only criterion that calculates a penalty related to the predicted TOA estimation accuracy, it is not directly designed to discard low energy components. It tries to judiciously decide whether low energy components, that also could be noise, are supposed to be modelled as a component or not. So overfitting still can happen. This, in return, for the time being, leaves us with the MDL criterion or an alternative subjective, based threshold method.

In the following, I will assess an approach to solve the multiple-measurement case first. Afterwards, I will present a method that also can be used in the single-measurement case.

6.2 Multiple Measurement Case

One straightforward approach of utilizing (6.6), (6.8), (6.7) would be to try using the least-squares error, as we know it corresponds to the deterministic maximum likelihood. The size of independent samples N_S would be calculated as $N_S = N_r I$, whereas the number of free adjustable parameters in $\tilde{\boldsymbol{\theta}}$ (the number of degrees of freedom in $\tilde{\boldsymbol{\theta}}$) would be determined via $k = \tilde{C}(2N_r I + 1)$. Note that, [Saa98] pointed out that this number of degrees of freedom is problematic: It violates the vital assumption that the sample size is large compared to the number of free adjustable parameters. This assumption is required for employing information-theoretic criteria like MDL and it would be violated, since

$$\lim_{N_s \to \infty} \frac{k}{\log N_s} \neq 0. \tag{6.13}$$

The problem of being confronted with an inappropriate parametrization for the information-theoretic criteria has been solved by formulating an alternative parametrization, which is related to the original parametrization, for arbitrary signals in [WZ89b] and for a stochastic ML problem formulation and circular complex Gaussian signals [WK85].
In the latter contribution, the model selection problem and the estimation problem can be separated such that the model selection problem can be determined in closed form.

Let us review the channel model vector $\mathbf{h}(t)$, which can be described by the following model equation

$$\hat{\mathbf{h}}(t) = \mathbf{G}(\boldsymbol{\theta})\boldsymbol{\beta}(t) + \mathbf{w}(t).$$

Now let us suppose, the complex path amplitudes $\boldsymbol{\beta}$ can assumed to be complex, stationary, ergodic and zero-mean Gaussian random processes with positiv-definite covariance matrices[WK85]. In general, this assumption is valid if the complex path amplitudes $\boldsymbol{\beta}(t)$ are modeled as independent Rayleigh fading multipath components like in NLOS conditions.

Based on these assumptions the separable closed form approach [WK85] is derived in the Appendix A.6 and is based on an eigenvalue-decomposition-based parametrization

$$\boldsymbol{\xi} = [\lambda_1, \ldots, \lambda_{\tilde{C}}\sigma_w^2, \mathbf{V}_1^T, \ldots, \mathbf{V}_{\tilde{C}}^T]. \tag{6.14}$$

The eigenvalues $\lambda_1, \cdots, \lambda_{\tilde{C}}$ and the eigenvectors $\mathbf{V}_1, \cdots, \mathbf{V}_{\tilde{\mathbf{C}}}$ belong to $\mathbf{C}_{\hat{\mathbf{h}}}{}^{(\tilde{C})}$. Applied to finding the number of multipath components and given that the estimates $\hat{\lambda}_i = l_i$ for all $0 \leq i \leq L$ the criterion reads

$$\mathcal{C}_{\mathrm{MDL}}(\tilde{C}) = -\ln\left(\frac{\prod_{i=\tilde{C}+1}^{L+1} l_i^{1/(L+1-\tilde{C})}}{\frac{1}{L+1-\tilde{C}}\sum_{i=\tilde{C}+1}^{L+1} l_i}\right)^{(L+1-\tilde{C})N_r I} + \frac{1}{2}\tilde{C}(2(L+1)-\tilde{C})\ln I N_r. \tag{6.15}$$

Like in [WK85], the number of free adjustable parameters of $\boldsymbol{\xi}$ is calculated in the following manner: Given that the eigenvalues associated with the covariance matrix are real and the eigenvectors are complex, $\boldsymbol{\xi}^{\tilde{C}}$ bears $\tilde{C}+2(L+1)\tilde{C}+1$ (1 is added if the error variance is also estimated) in total. Additionally, considering that the eigenvectors are constrained to have unit norm, yields a reduction of $2\tilde{C}$ parameters. A further reduction of $2\frac{1}{2}\tilde{C}(\tilde{C}-1)$ has to be considered taking into account that the eigenvectors are supposed to be mutually orthogonal. Consequently, the number of free adjustable parameters is

$$\tilde{C}+2(L+1)\tilde{C}+1-2\tilde{C}-2\frac{1}{2}\tilde{C}(\tilde{C}-1)=\tilde{C}(2(L+1)-\tilde{C})+1. \tag{6.16}$$

Note that the addition of 1 in (6.16) can be neglected when further used for minimization.

The term in the brackets in (6.15) is the ratio of the geometric to the arithmetic mean. In order to see that the likelihood term in (6.15) can be expressed in this manner, note that the estimates given by (A.44) have to be inserted into the likelihood function (A.41) and the likelihood expression is further manipulated by targeting the expression given by the ratio of the geometric to the arithmetic mean as is shown in the Appendix A.5. Note that in case $I=1, N_r=1$ the second summand is equal to zero. Finally, the estimated number of source signals is

$$\hat{C}=\arg\min_{\tilde{C}}\left\{\mathcal{C}_{\mathrm{MDL}}(\tilde{C})\right\}. \tag{6.17}$$

Further, the authors of [CY03] claim that the number of free parameters, if a forward-backward smoothed covariance matrix is used, is $\tilde{C}(L+1+\tilde{C}+1)+1$.
In [WK85] it is argued that the MDL is preferable compared to the AIC since it is consistent and the latter is not. Consistency is given if, by asymptotically increasing the sample size, the estimator approaches the real number of source signals. Furthermore, note that the sample size is finite from a practical point of view, due to the channel variability over time.

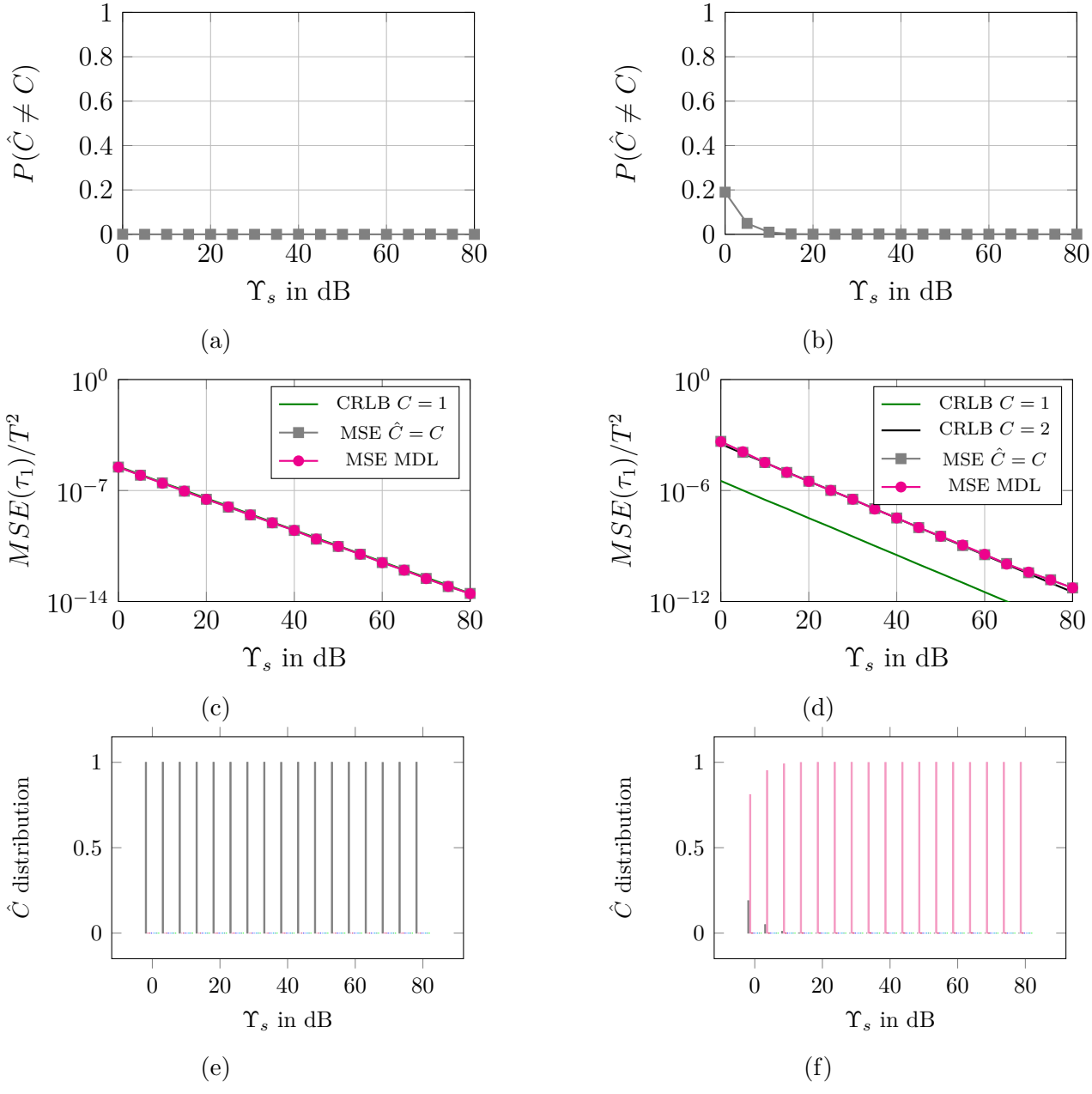
Up to this moment, the employed parametrization assumed identically distributed, zero mean, circular complex Gaussian signals for the complex path weights. Especially in the LOS scenario, which is an important scenario, when it comes to positioning, this assumption is not valid. Hence, the signals should be modeled as deterministic signals instead of modelling them as stochastic signals. Unfortunately, if we try to use the deterministic parametric model in combination with the MDL criterion, the large sample assumption is violated. Fortunately, the author of [WK85] proposed a solution [WZ89b] to anyway use the deterministic ML estimator in combination with MDL, by relating the actual parametrization to another eigendecomposition-based parametrization. Contrary to the approach proposed in [WK85], this estimator does not only estimate the model order. It also performs a joint parameter and model selection and therefore complies with (6.3)-(6.5). The drawback of this approach is that the negative likelihood optimization has to be carried out for every candidate model order. The approach was later applied to the estimation of number of multipaths in DS-CDMA in [Saa98]. The underlying system presented in [Saa98] is therefore almost equal to the system assessed here. Instead of a signal matrix depending on $\boldsymbol{\tau}$ constructed according to the rules of DS-CDMA, we have

a matrix $\mathbf{G}(\boldsymbol{\tau})$ constructed by using samples of the delayed convolved pulse. Similar to [WZ89b, Saa98] we can devide the observations into a signal and a noise subspace and can formulate a likelihood based on an eigenparametrization.

Table 6.1: Parameters of simulation setup

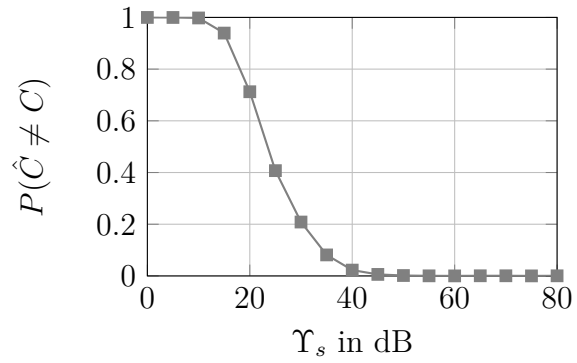
Channel		Scenario	WINNER B1-LOS
		C	$\in \{1, \ldots, 8\}$
		φ_1	$\pi/4$
		ϕ_1	$\pi/4$
		v	50 km/h
		τ_1	$\in\ \mathcal{U}[0, (L+1)T]$
System	Signal	g_{RC} with α_{RC}	0.3
		T	200 ns
		L_g	7
	Time/Frequency	f_c	2 GHz
		J	1
		K	1000
		I	100
	Input/Output	N_t	1
		N_r	1
		$d_{tx}, d_{Rx}, \lambda/2$	7.5 cm
	model selection	criterion	MDL Akaike
		parametrization	Eigen based ($\boldsymbol{\lambda}$, $\boldsymbol{V}$) SML
	TOA estimation	DML/SNLLS	with PSO
Simulation	Runs	1000	

Simulation results shown in the figures 6.2-6.5 emphasize the usefulness of joining model selection and parameter estimation. The simulations for these figures have been carried out for the simulation setup provided in Table 6.1. The results show that for a small choice of C ($C = 1$, $C = 2$) the detection algorithm almost never fails despite that a LOS channel is used although zero-mean path weights are assumed for the model order detection. The estimation results are the same as if we would have assumed $\hat{C} = C$. For larger C the model selection failure probability only converges to zero for high SNRs. Assessing the TOA estimation MSEs shown in the second rows of the Figures, it can be seen that this failure is beneficial in the sense of parameter estimation, since the MSEs for the TOAs, estimated on the basis of estimating C via MDL beforehand, are significantly lower than the MSEs for zero failure ($\hat{C} = C$). It can be seen that the results approximately approach the CRLBs associated with the estimated model order in a stepwise fashion. This property can be understood by having a look at probability $\hat{C} = C$ for all $c \in \{1, \ldots, C_{max}\}$ shown in the last rows of the Figures 6.2-6.5. Nonetheless, note that for a number of multipath components C that are close to or even larger than the number of theoretically identifiable multipath components the TOA MSEs do not monotonically

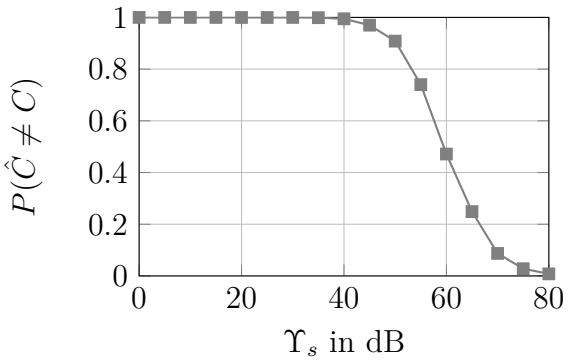


(g) Legend for all plots

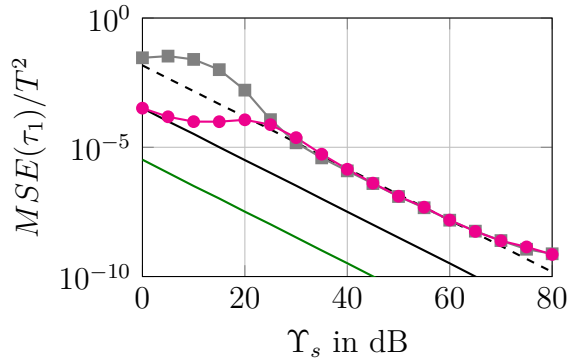
Figure 6.2: Comparing the probability of failed detection ($\hat{C} \neq C$) in the first row with the actual delay estimation MSEs for such small values of C the model order detection is almost completely correct. The subfigures a,c,e on the left side show the results for $C = 1$, the subfigures b,d,f on the right side show te results for $C = 2$.



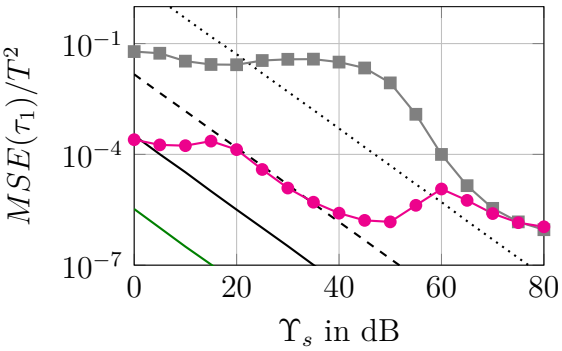
(a)



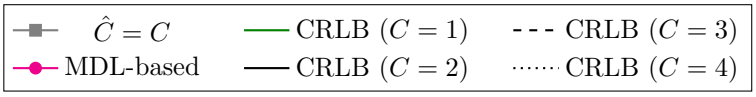
(b)



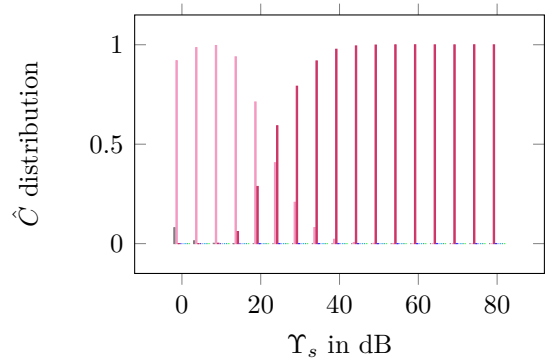
(c)



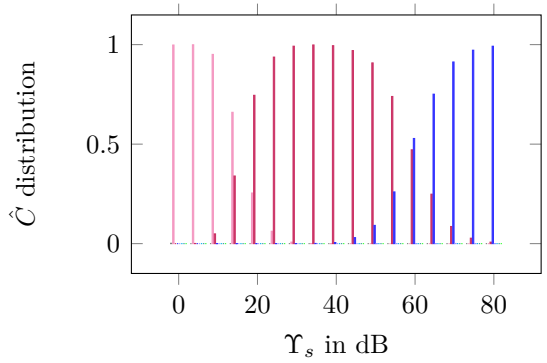
(d)



(e) Legend for subfigures (a-b)



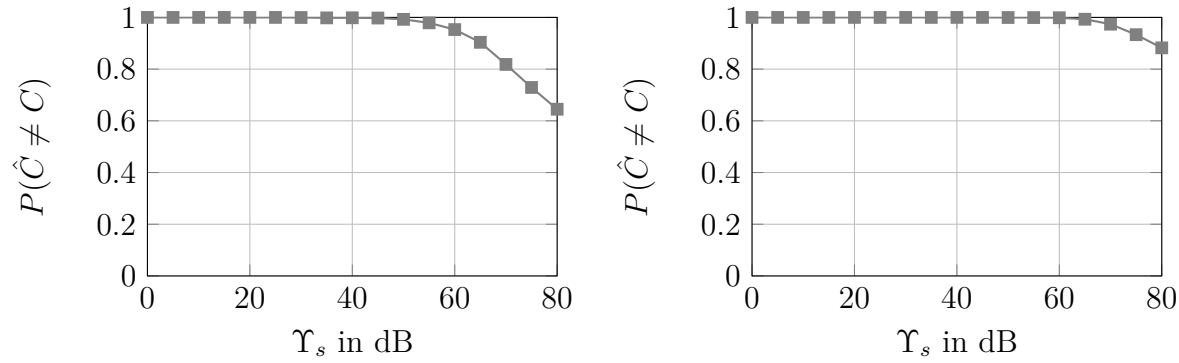
(f)



(g)

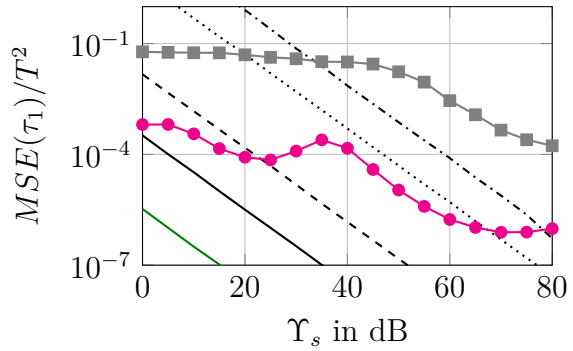
(h) Legend for all subfigure (f, g)

Figure 6.3: Comparing the probability of failed detection ($\hat{C} \neq C$) in the first row with the actual delay estimation MSEs it becomes clear that employing a simpler model than the correct one improves the estimation accuracy. The subfigures a,c,e=f on the left side show the results for $C = 3$, the subfigures b,d,g on the right side show the results for $C = 4$.

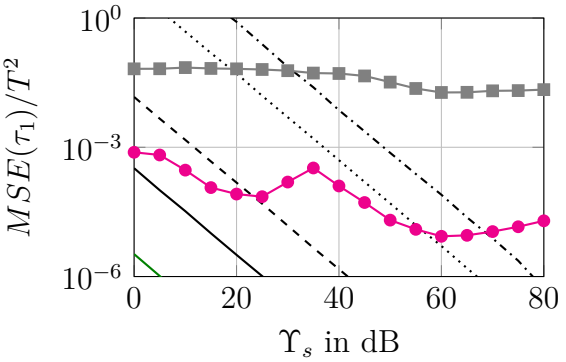


(a)

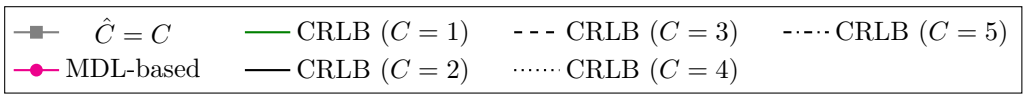
(b)



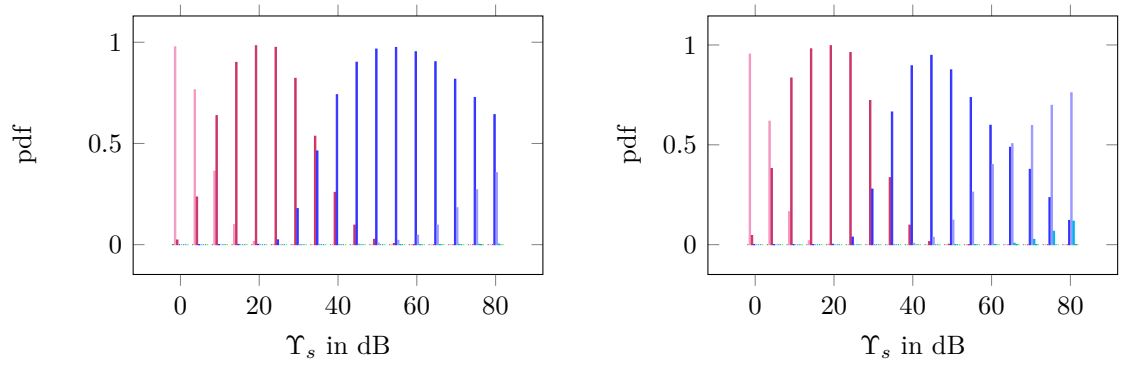
(c)



(d)



(e) Legend for subfigures (a-d)

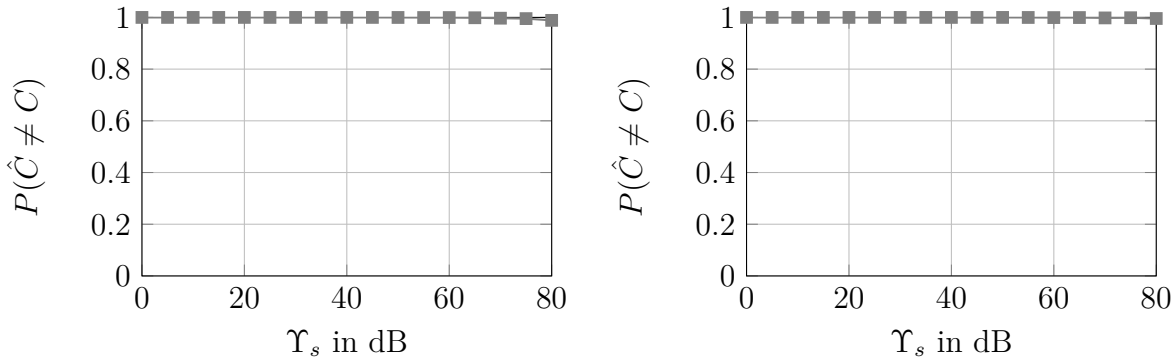


(f)

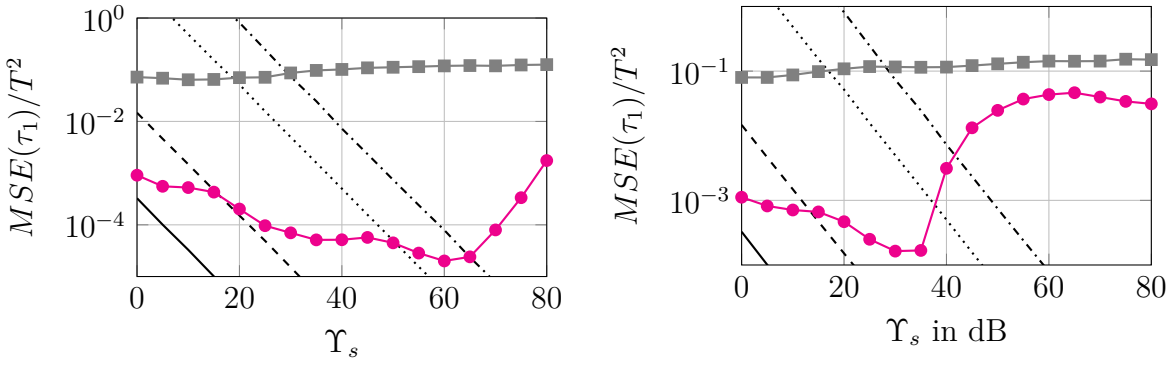
(g)

(h) Legend for all subfigures (f, g)

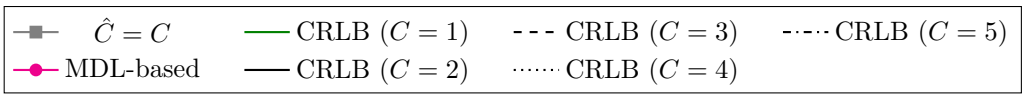
Figure 6.4: Comparing the probability of correct detection in the first row with the actual delay estimation MSEs it becomes clear that employing a simpler model than the correct one improves the estimation accuracy. The figures on the left side show the results for $C = 5$, the ones on the right side show the results for $C = 6$.



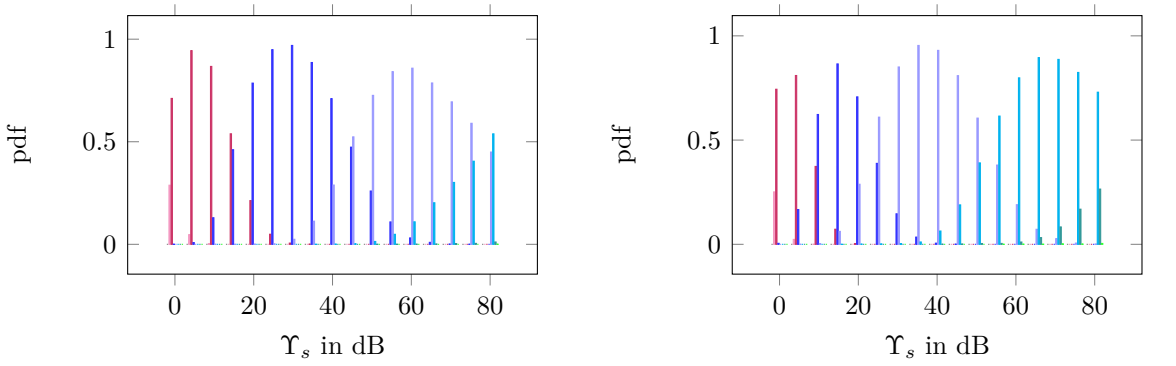
(a) (b)



(c) (d)



(e) Legend for subfigures (a-d)



(f) (g)

(h) Legend for all subfigures (f, g)

Figure 6.5: Comparing the probability of correct detection in the first row with the actual delay estimation MSEs it becomes clear that employing a simpler model than the correct one improves the estimation accuracy. The figures on the left side show the results for $C = 7$, the ones on the right side show the results for $C = 8$.

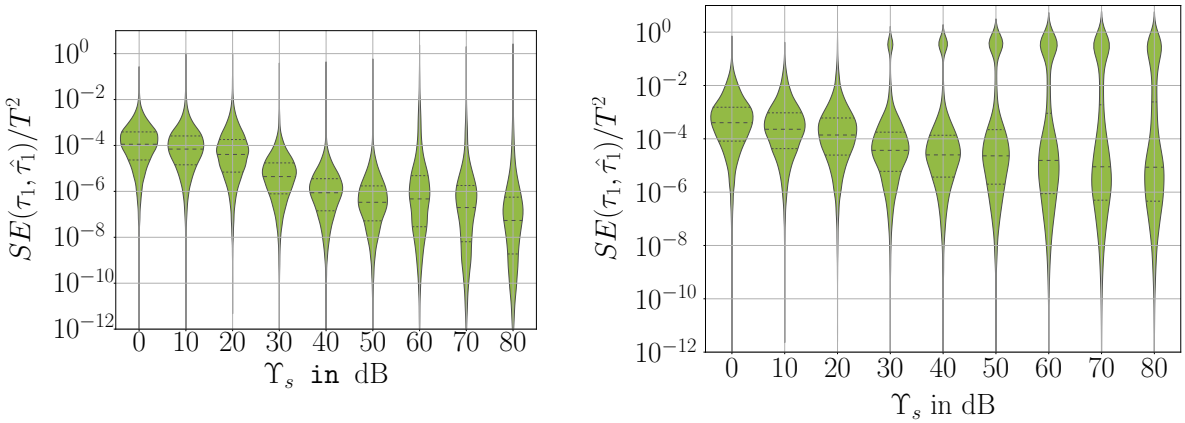


Figure 6.6: The violin-plots depicted on the left for $C = 4$ and on the right for $C = 8$ show the squared TOA estimation error distributions. The three thin lines in each violin show the median, the 25 and the 75 percentiles. It becomes clear that the major part of the error distribution lies within an error region that indicates a successful joint model order and parameter estimation.

decrease with increasing the SNR. In order to understand this behaviour a more detailed TOA error examination is required like shown in Figure 6.6. Contrary to Figure 6.5 Figure 6.6 indicates a successful model order selection and parameter estimation for the major part of the distributions. The effect, that for severe multipath like $C = 8$, the model order is overestimated and hence leads to poor TOA estimation results, can be explained by the fact that with a higher SNR a higher model order is potentially possible, which in return comes with the price of less reliability.

6.3 Single Measurement Case

6.3.0.1 Threshold-based Least Squares Approach

A threshold-based least squares joint model order selection and parameter estimation approach, working for both the single measurement case as well as the multiple measurement case, was applied to the special problem of cissoid detection in [QY97]. Note that we can easily transform the parameter estimation problem we have to deal with into one, which corresponds to a cissoid detection problem. This can be understood by acknowledging that the channel observations used for parameter estimation can be deconvolved by $G(f)$ after applying a DFT and then can be approximated by a sum of cisoids in the frequency domain. Nevertheless, the approach introduced in [QY97] is similarly applicable without any transformation and deconvolution. Hence, the detection algorithm can be adjusted to the signal model and settings used in this thesis and will be briefly introduced in the following.
The joint detection and estimation algorithm exploits the property that, when the additive noise is assumed to be Gaussian, the least-squares Ω_{SNLLS} error is χ^2-distributed

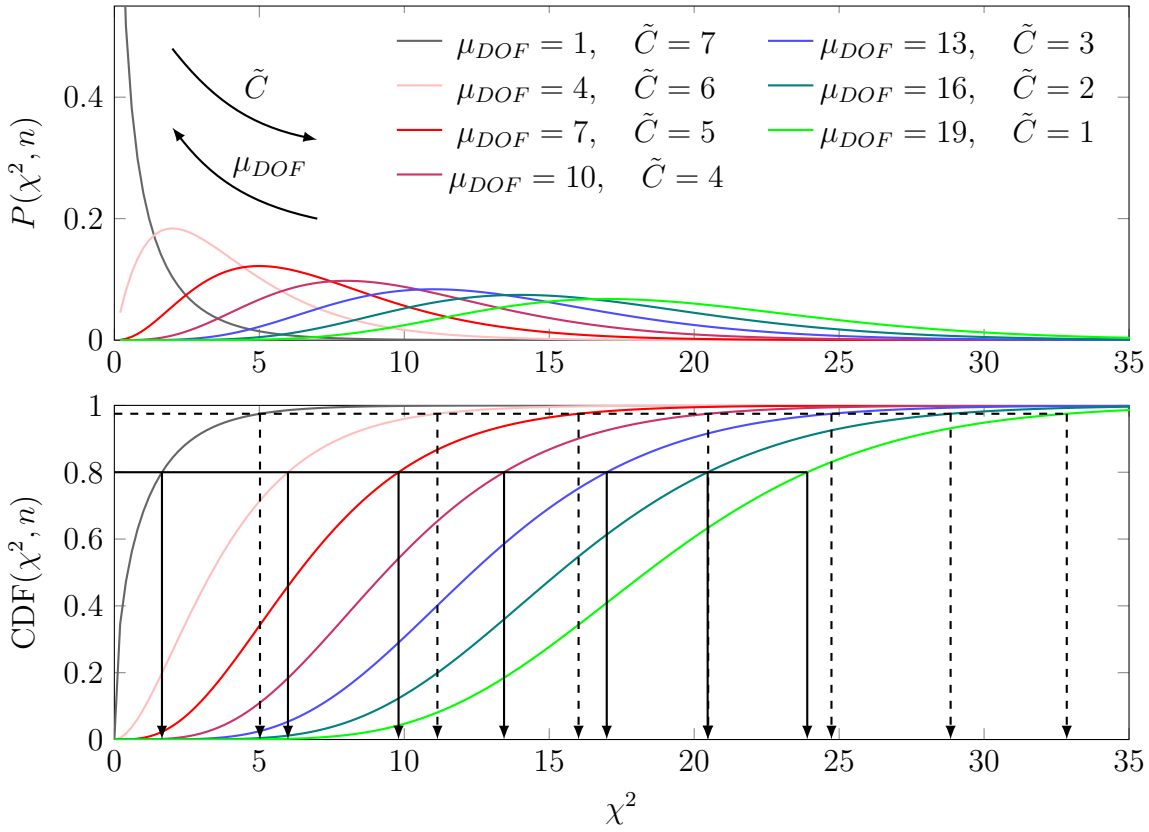


Figure 6.7: For additive white Gaussian noise the LS error's probability density function (PDF) and cumulative density function (CDF) are known functions (χ^2-distribution), uniquely specified by the number of degrees of freedom and hence the number of multipath components. The dashed and solid arrows show how decision regions are determined in order to estimate $\hat{C}$ for $\alpha = 0.8$ and $\alpha = 0.975$. It can be seen that the user-specified threshold can lead to a different decision for $\hat{C}$. Since α is specified the probability for the channel estimates having a LS error below the assigned threshold, the value should be chose close to 1.

such that

$$\Omega_{SNLLS} \approx \frac{\sigma_w^2}{2}\chi^2. \tag{6.18}$$

The distribution is depending on the number of degrees of freedom μ_{DOF} present to the assessed least-squares problem like shown in Figure 6.7. Figure 5.8 shows that this is a valid assumption for the estimator used here.The number μ_{DOF} is determined by the number of measurements, or equations, minus the number of parameters, that is, the dimension of the parameter estimation problem. The parameter estimation problem dimension is specified by the number of hypothetically assumed multipaths. Let us consider the deterministic maximum likelihood scenario. Having both, a real and imaginary part of the observations, we can count $2(L+1)N_rI$ real-valued observations. For each snapshot a compound of the complex path amplitudes and the delays are estimated jointly. Thereby

the complex path amplitudes real and imaginary part each constitute for an unknown. This is resulting in a total number of $2N_RIC$ unknowns for the complex path amplitudes plus C unknowns for the delays. Consequently, μ_{DOF} for the deterministic case is

$$\mu_{DOF} = 2(L+1)N_rI - (2N_rI\tilde{C} + \tilde{C}). \tag{6.19}$$

This huge dimensionality emphasizes the feasibility of employing a lower dimensionality stochastic ML estimator, if possible (large-sample case). In the single measurement case we have

$$\mu_{DOF} = 2(L+1) - 3\tilde{C}. \tag{6.20}$$

For example, the relation between μ_{DOF} and $\tilde{C}$ is tabulated for the case $L = 10$, $N_r = 1$, $I = 1$ here:

$\tilde{C}$	1	2	3	4	5	6	7
μ_{DOF}	19	16	13	10	7	4	1

$$\Omega_{SNLLS_{\tilde{C}}}(\tilde{\boldsymbol{\tau}}_{\tilde{C}}) = \hat{\boldsymbol{h}}^H \mathbf{P}_{\mathbf{G}\tilde{C}}^{\perp}(\tilde{\boldsymbol{\tau}})\hat{\boldsymbol{h}}^H, \quad \text{with} \tag{6.21}$$

$$\hat{\boldsymbol{\tau}}_{\tilde{C}} = [\hat{\tau}_1^{\tilde{C}} \dots \hat{\tau}_{\tilde{C}}^{\tilde{C}}], \tag{6.22}$$

$$\hat{\boldsymbol{\tau}}_{\tilde{C}} = \arg\min_{\tilde{\boldsymbol{\tau}}_{\tilde{C}}} \{\hat{\boldsymbol{h}}^H \mathbf{P}_{\mathbf{G}}{}^{\perp}(\tilde{\boldsymbol{\tau}}_{\tilde{C}})\hat{\boldsymbol{h}}^H\}. \tag{6.23}$$

Then, by choosing a confidence parameter α via table-lookup [PTVF07, PH65] or via calculation, $\forall \tilde{C} \in \{1, \dots, C_{max}\}$ thresholds $\epsilon_{\tilde{C}}$ can be defined such that

$$P\left(\frac{\sigma_{\boldsymbol{w}}^2}{2}\chi^2_{\mu_{DOF}} \leq \epsilon_{\tilde{C}}\right) = \alpha. \tag{6.24}$$

If the lookup-tables have to be calculated, that means the $\epsilon_{\tilde{C}}$ has to be calculated by considering the inverse cumulative density function such that

$$\begin{aligned} \epsilon_{\tilde{C}} =& p^{-1}_{CDF}(\alpha|\mu_{DOF}) = \{\epsilon_{\tilde{C}} : p(\epsilon_{\tilde{C}}|\mu_{DOF}) = \alpha\}, \\ \alpha =& p_{CDF}(\epsilon_{\tilde{C}}|\mu_{DOF}) = \int_0^{\epsilon_{\tilde{C}}} \frac{t^{(\mu_{DOF}-2)/2}e^{-t/2}}{2^{\mu_{DOF}/2}\Gamma(\mu_{DOF}/2)}dt. \end{aligned} \tag{6.25}$$

With $\epsilon_{\tilde{C}}$ a joint model order and parameter estimator can be formulated as

$$\hat{C} = \min\{\tilde{C}|\frac{2}{\sigma_w^2}\Omega_{SNLLS}(\tilde{C}, \hat{\boldsymbol{\tau}}_{\tilde{\boldsymbol{C}}}) \leq \epsilon_{\tilde{C}}\}, \tag{6.26}$$

$$\hat{\boldsymbol{\tau}}_{\tilde{C}} = \arg\min_{\tilde{\boldsymbol{\tau}}}\{\Omega_{SNLLS_{\tilde{C}}}(\tilde{\boldsymbol{\tau}}_{\tilde{C}})\}. \tag{6.27}$$

Note that thereby the maximum number of multipath components C_{max}, which we can reliably estimate, is limited and determined by μ_{DOF}, since $\tilde{C}_{max}$ can be obtained by setting $\mu_{DOf} \geq 0$ in (6.19). This yields the limiting number of

$$\tilde{C}_{max} = \left\lfloor \frac{L+1}{1 + \frac{1}{2N_RI}} \right\rfloor \tag{6.28}$$

estimable multipaths. More specifically, that means for the singleshot measurement case and for the large sample case:

$$\tilde{C}_{max} = \begin{cases} \left\lfloor \frac{2}{3}(L+1) \right\rfloor, & \text{if } N_r I = 1, \\ L+1 & \text{if } N_r I \to \infty. \end{cases} \tag{6.29}$$

Note that this restriction can cause problems in the single-measurement case, considering that the typically applied channel memory lengths L are rather small (typically $L \in \{10, \ldots, 20\}$). Since those $L+1$ channel taps, which significantly differ from zero, are given by the number

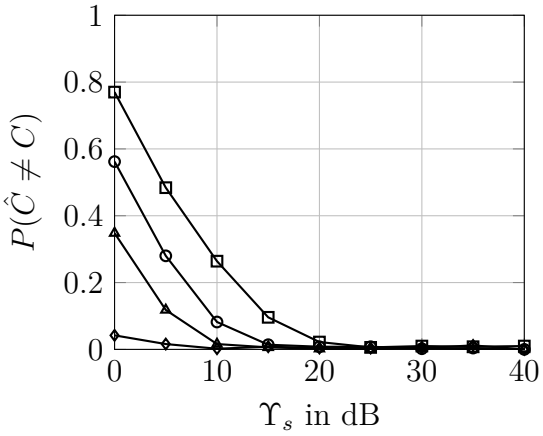
$$L = L_g - 1 + \left\lceil \frac{\tau_{max}}{T} \right\rceil, \quad L_g = 2W + 1, \tag{6.30}$$

it is clear that this number cannot be chosen arbitrarily large. Thereby, W denotes the factor for which WT represents the half window width of the window for the overall pulse impulse response $g(\tau)$. On the other hand, depending on the applied channel scenario the actual number of multipath components C typically rather ranges from 8 to 15. However, according to (6.29), it becomes clear that depending on which actual channel scenario, we are confronted with, we might not be able to reliably estimate all components ($\tilde{C}_{max} < C$). For instance, in case $C = 8$ and $L = 10$, it follows that $\tilde{C}_{max} = 7$. Therefore, increasing the number of samples either by employing multiple receive antennas or observing a time series might be necessary depending on the channel scenario. However, the algorithm is summarized in the following. Lines $1-3$ are not necessary if table-lookup is used.

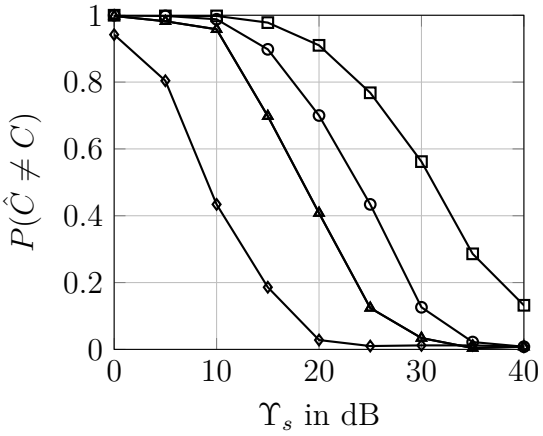
Algorithm 3 χ^2-based joint Model Selection and Parameter Estimation

1: **for** all $\tilde{C} \in \{1, \ldots, C_{max}\}$ **do**
2: Calculate $\mu_{dof}(\tilde{C})$ via (6.20).
3: Calculate $\epsilon_{\tilde{C}}$ via (6.25).
4: Calculate $\hat{\boldsymbol{\tau}}_{\tilde{C}}$ via (6.23) and the associated $\Omega_{SNLLS_{\tilde{C}}}$.
5: **if** $\Omega_{SNLLS_{\tilde{C}}} \leq \epsilon_{\tilde{C}}$ **then**
6: $\hat{C} = \tilde{C}$.
7: $\hat{\boldsymbol{\tau}} = \hat{\boldsymbol{\tau}}_{\tilde{C}}$ with $\tilde{C} = \hat{C}$.
8: **break**

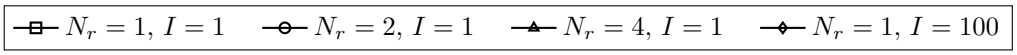
Apart from setting $I = 1$, the simulation results for the single-measurement case are shown for a similar setup to the one specified in Table 6.2. The results in the Figures 6.8 and 6.9 show that the proposed method for joint model selection and parameter estimation also works in the single measurement scenario. Similarly to the multiple-measurement case results for lower SNRs Υ_b, the algorithm prefers smaller model orders than the correct model order. Hence the delay estimation MSEs are lower than for the estimator based on the correct model order. The MSEs even are slighly lower than the CRLBs. This mismatch can be explained by the fact that the CRLB is calculated for an unbiased estimator and the applied estimator is not exactly unbiased. By using the reasonable assumption that the search space is bounded we introduce a estimation bias.



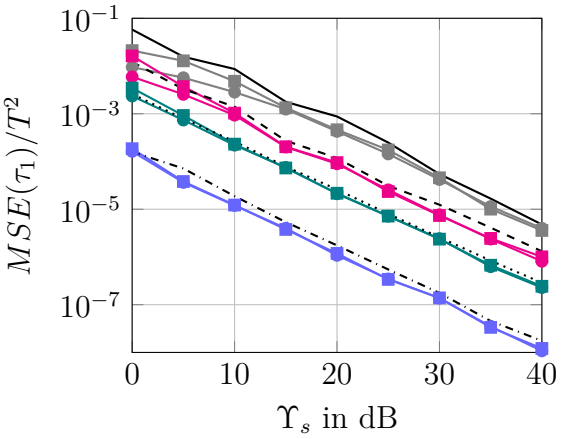
(a) $C = 2$, $P(\hat{C} \neq C)$



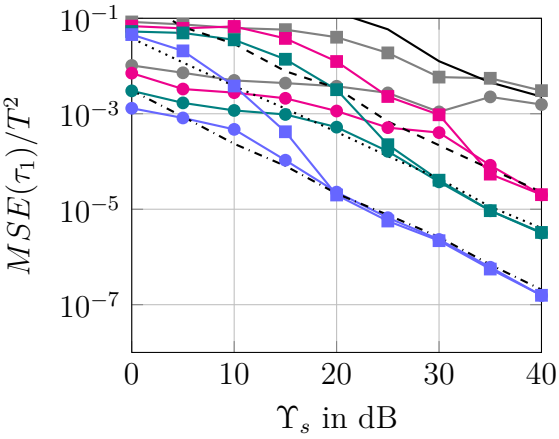
(b) $C = 3$, $P(\hat{C} \neq C)$



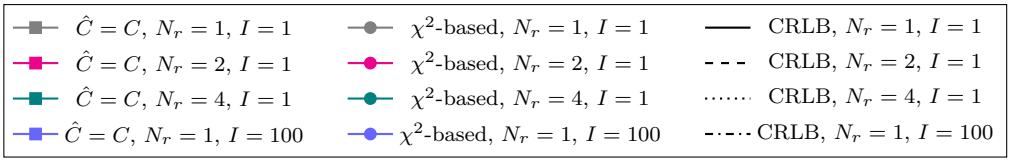
(c) Legend for subfigures (a-b)



(d) $C = 2$, MSEs

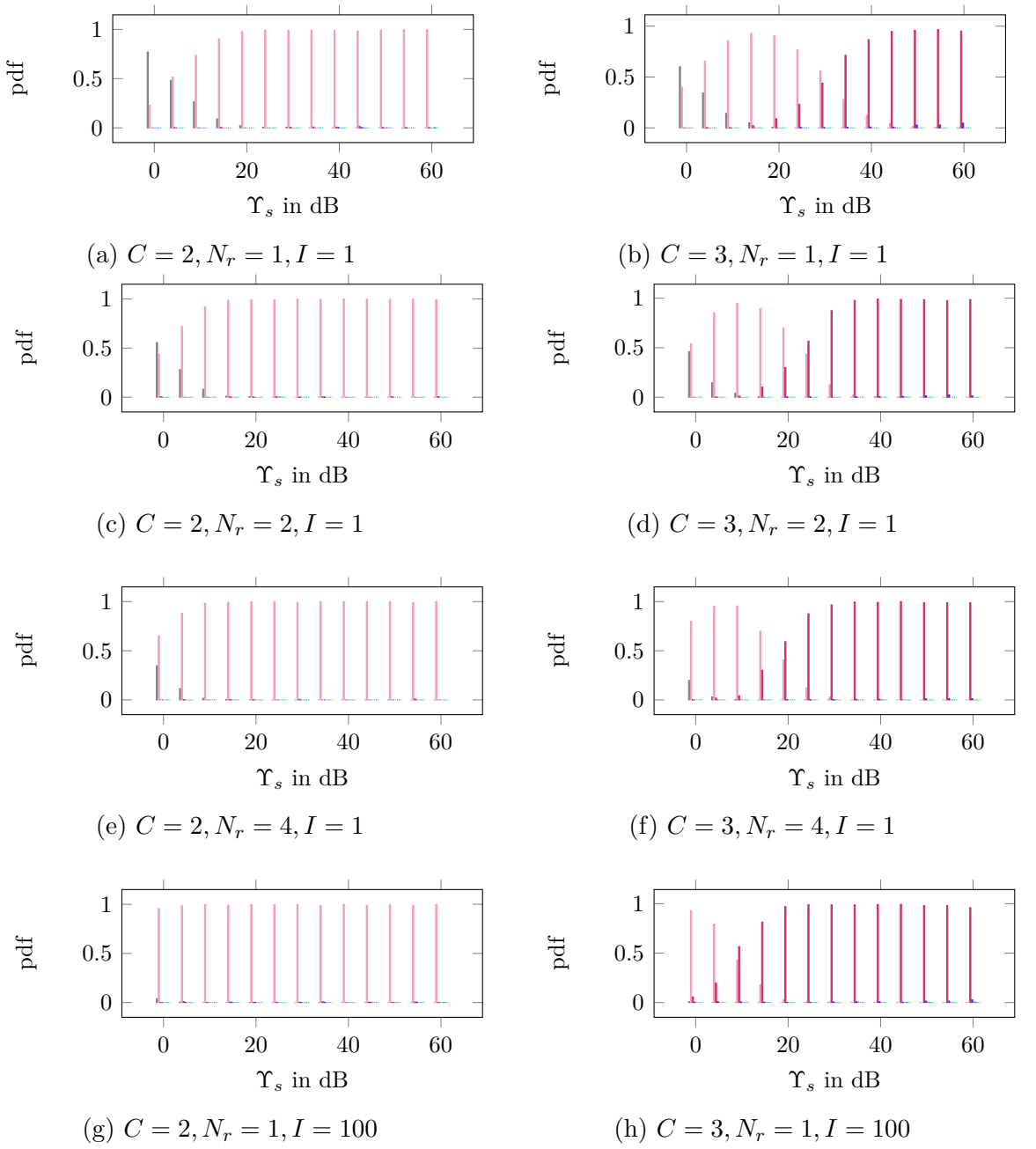


(e) $C = 3$, MSEs



(f) Legend for subfigures (d-e)

Figure 6.8: Similar to the multiple-measurement case, it can be seen that the higher C the higher the SNR is for $P(\hat{C} = C) \to 0$. Furthermore, the TOA MSEs improve if model selection and parameter estimation is applied. Increasing the number of measurements has the opposite effect as increasing the number of parameters.



(a) $C = 2, N_r = 1, I = 1$

(b) $C = 3, N_r = 1, I = 1$

(c) $C = 2, N_r = 2, I = 1$

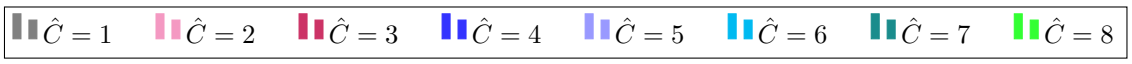
(d) $C = 3, N_r = 2, I = 1$

(e) $C = 2, N_r = 4, I = 1$

(f) $C = 3, N_r = 4, I = 1$

(g) $C = 2, N_r = 1, I = 100$

(h) $C = 3, N_r = 1, I = 100$



(i) Legend for all plots

Figure 6.9: Distributions for $P(\hat{C})$, with $\hat{C} \in \{1, \ldots, 8\}$ show similar behaviour as the distributions obtained by applying information-theoretic criteria. Increasing the number of measurements via $N_r > 1$ or $I > 1$ increases $\hat{C}$.

Table 6.2: Parameters of the simulation setup for the χ^2-based joint model selection and parameter estimation.

Channel		Scenario	WINNER B1-LOS
		C	$\in \{2, 3\}$
		φ_1	$\pi/4$
		ϕ_1	$\pi/4$
		v	50 km/h
		τ_1	$\in\ \mathcal{U}[-0.0, T]$
System	Signal	$\alpha_{RC} = 0.3$	
		T	100 ns
		L_g	7
	Time/Frequency	f_c	2 GHz
		J	1
		K	1000
		I	$\in \{1, 100\}$
	Input/Output	N_t	1
		N_r	$\in \{1, 2, 4\}$
		$d_{tx}, d_{Rx}, \lambda/2$	7.5 cm
	Model Selection & Parameter Estimation	χ^2-based	$\alpha = 0.9999$
		Parametrization	DML ($\boldsymbol{\theta} = [\boldsymbol{\tau}, \mathrm{vec}(\boldsymbol{\Gamma})]$)
		Cost	Ω_{SNLSS}
		Optimization	PSO+LM
	Runs	10000	

6.4 The Focused Order-Related Lower Bound

The CRLB depends on the model order. It is calculated based on the assumption that the underlying model order is correct. For model selection this assumption is unreasonable. Our goal is to find a model order that yields the lowest delay estimation error and not necessarily the correct model order. Therefore, whenever the model order is not correct and yet nevertheless preferable, since it results in a lower delay estimation MSE, the associated CRLB will not be correct. If the model order is underestimated the CRLB will be too optimistically low. Increasing the model order for the same SNR means increasing the CRLBs. The estimation bias introduced by the erroneous preferable model is missing in the CRLB. Conseqeuntly, for the purpose of model selection the CRLB is not the most useful tool. Therefore, I introduce a more suitable and practically motivated bound that I dub the focused order related lower bound (FORLB). Whereas the CRLB calculation requires the correct model order, the FORLB does not. Instead I tailored it to require parameter estimates, here delay estimates $\hat{\tau}_{1,\tilde{C}}$, that I trust to be optimal or at least close

to optimal, for the different hypothetical model orders $\tilde{C}$:

$$\hat{C}_{\text{opt}} = \arg\min_{\hat{\tau}_{1,\tilde{C}}} \{(\tau_1 - \hat{\tau}_{1,\tilde{C}})^2\} \tag{6.31}$$

$$\text{FORLB}(\tau_1) = (\tau_1 - \hat{\tau}_{1,\hat{C}_{opt}})^2 \tag{6.32}$$

The bound is focused, since it prefers models that result in a good TOA MSE, instead of balancing the overall delay estimation MSE. Note that focusing on a special parameter is here desirable for model selection as well, similarly as envisaged in [CH01, LPRU19] Therefore both bounds, the CRLB and the FORLB, can be seen as an aid to optimize estimation and detection algorithms on the basis of specific models and available information. They should both not be seen as the absolute lowest bound that can be achieved, since changing the estimator, or the modeling, or the apriori knowledge, always changes the bounds as well.

6.5 Model Selection based on Soft Information

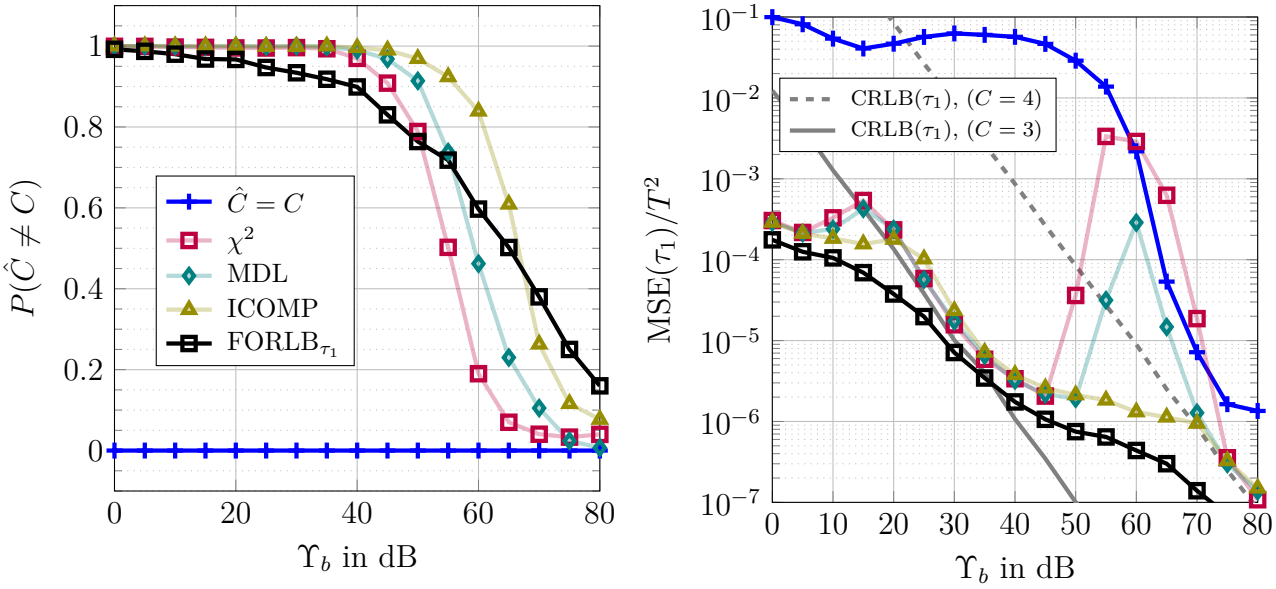
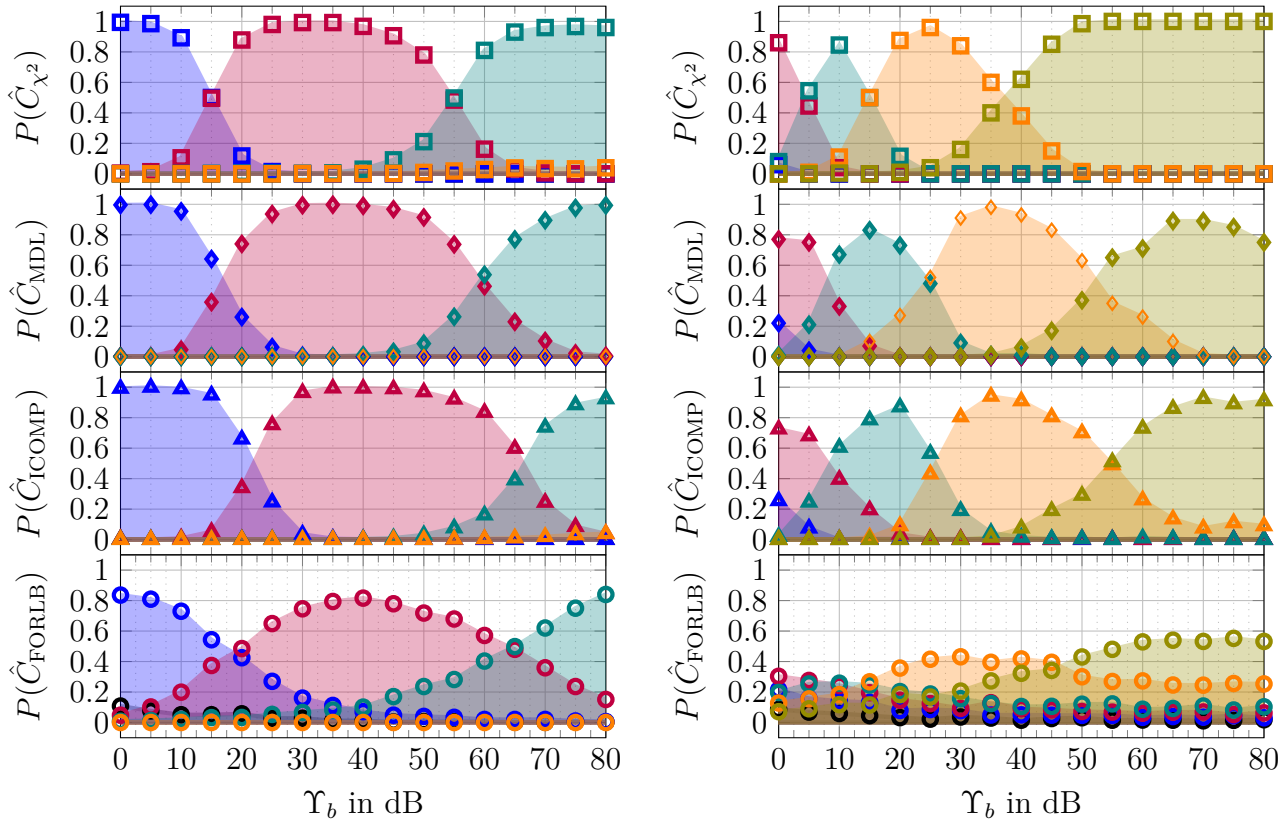


Figure 6.10: MDL, ICOMP and the χ^2-based method are compared to the case $\hat{C} = C$ and to the FORLB for $C = 4$. ICOMP is closest to the optimal FORLB.

For the joint parameter estimation and model order selection problem here, the aim is to penalize those models more that lead to a high estimation error, especially those models that lead to a high TOA estimation error. A directly related measure used to penalize estimation errors is the estimates covariance matrix $\mathbf{C}_{\hat{\boldsymbol{\theta}}}$ related to the parametrization $\boldsymbol{\theta}$,



(a) $C = 4$ (b) $C = 8$

Figure 6.11: The probability distributions for $\hat{C}$ for the different model order selection methods and the FORLB show that optimal distribution for each model order is wider than the distributions obtained via algorithms.

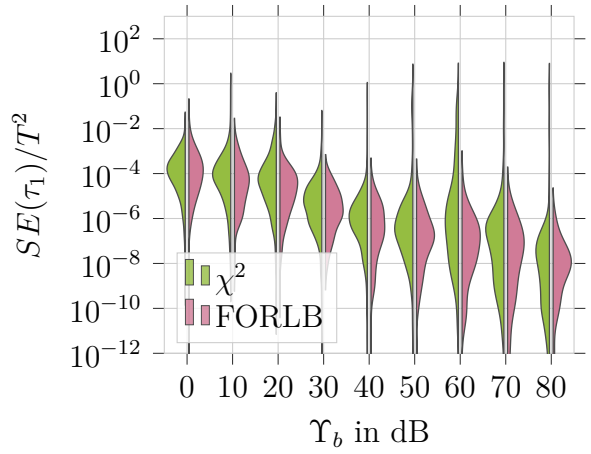
given by:

$$\text{full problem:} \quad \mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{F}(\boldsymbol{\theta})^{-1} \tag{6.33}$$

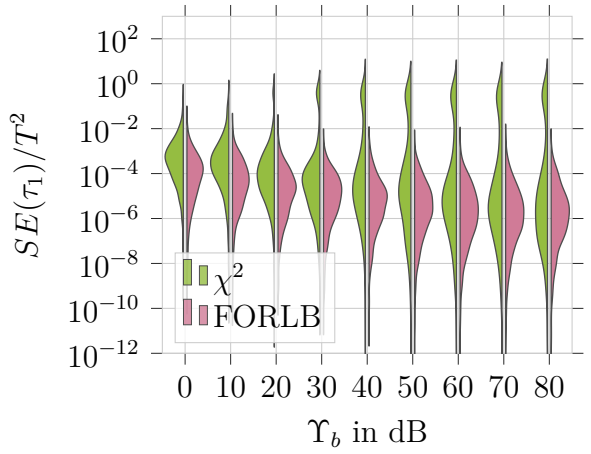
$$\text{SNLLS:} \quad \mathbf{C}_{\hat{\boldsymbol{\tau}}} = \frac{\sigma_w^2}{2}\left[\mathrm{Re}\left\{\mathbf{J_G}^T\mathbf{P_G}(\hat{\boldsymbol{\tau}})^{\perp}\mathbf{J_G} \odot \left(\left(\mathbf{G}(\hat{\boldsymbol{\tau}})^{\dagger}\hat{\mathbf{H}}\right)^H \mathbf{G}(\hat{\boldsymbol{\tau}})^{\dagger}\hat{\mathbf{H}}\right)^T\right\}\right]^{-1} \tag{6.34}$$

Thereby, the matrix $\mathbf{F}(\boldsymbol{\theta})$ is the so-called Fisher information matrix of the full problem. Note that the calculation of the Fisher matrix, and the estimates covariance matrix and the related lower bound, the CRLB, is determined and assessed in Chapter 5.

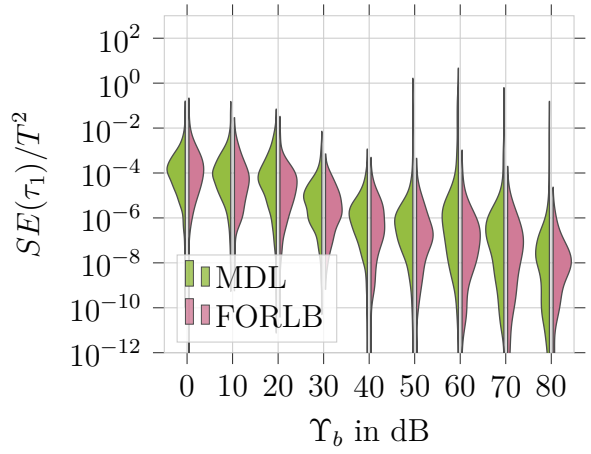
Note that this matrix has to be used with caution, like we assessed and explained in [SAH11b] due to different reasons: The matrix derivation is based on using an approximated version of the Hessian matrix instead of using the exact Hessian. The exact



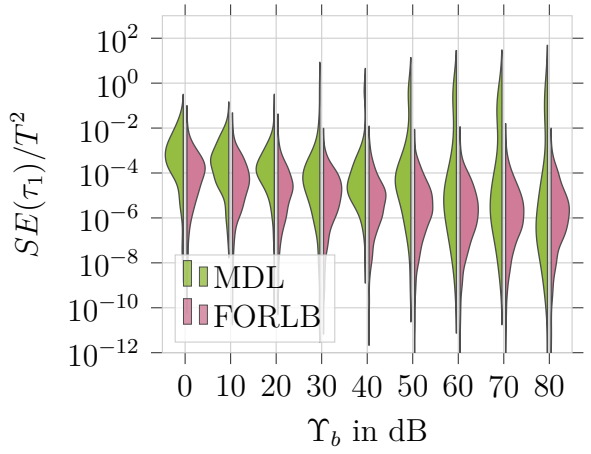
(a) $C = 4$, χ^2-based



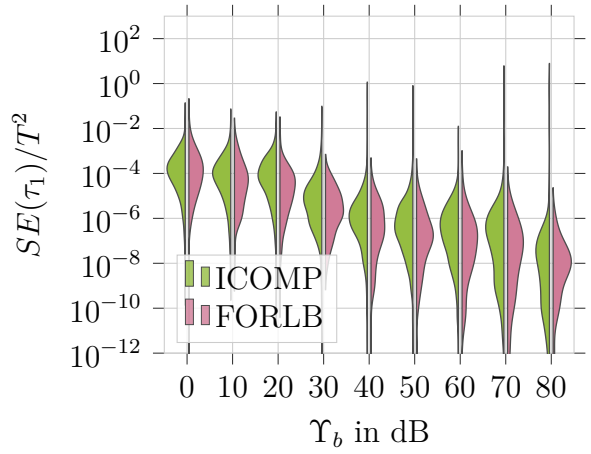
(b) $C = 8$, χ^2-based



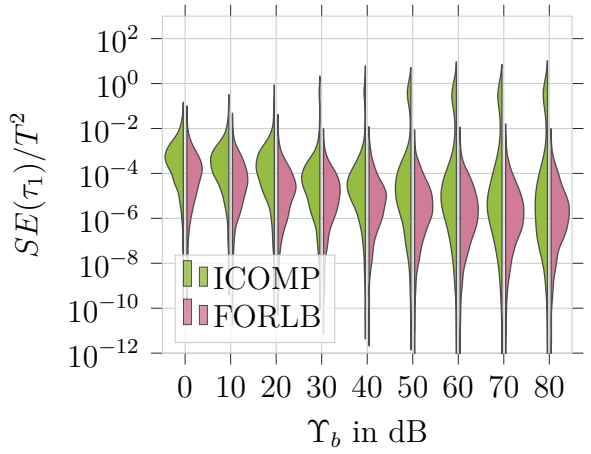
(c) $C = 4$, MDL



(d) $C = 8$, MDL



(e) $C = 4$, ICOMP



(f) $C = 8$, ICOMP

Figure 6.12: The split violin-plots show the SE distribution for a SNR range for the χ^2-based method. MDL and ICOMP compared to the optimal performance given by the FORLB.

Hessian might lead to a non-positive-definite matrix and hence disqualifies for the required Hessian inversion. Assessing the covariance estimates, related confidence regions and curvature measures, verifies that such soft information estimates are often inexact. Nevertheless, the results in [SAH11b] indicate that using them yields performance gains at least when applied to improve the positioning results by weighting the delay estimates by their estimated variances.

Recall that λ_i denotes the ith eigenvalue of the estimated parameter covariance matrix $\hat{\mathbf{C}}_{\boldsymbol{\xi}}$. Let $k = 2N_r I\tilde{C} + \tilde{C}$ denote the model order. The information complexity criterion provides a straightforward method to incorporate soft information for a joint parameter estimation and model order selection approach by utilizing the trace and the determinant of $\hat{\mathbf{C}}_{\boldsymbol{\xi}}$:

$$\mathcal{C}_{ICOMP}(\hat{\boldsymbol{\theta}}_{\tilde{C}}) = -2\ln\left(\mathcal{L}(\hat{\mathbf{H}}|\hat{\boldsymbol{\theta}}_{\tilde{C}})\right) + \tilde{k}\ln\left(\frac{\frac{1}{\tilde{k}}\sum_{j=0}^{\tilde{k}}\lambda_j}{\left(\prod_{j=1}^{\tilde{k}}\lambda_j\right)^{1/\tilde{k}}}\right), \tag{6.35}$$

$$\hat{\boldsymbol{\theta}}_{\tilde{C}} = \arg\min_{\tilde{\boldsymbol{\theta}}_{\tilde{C}}}\left\{\ln\left(\mathcal{L}(\hat{\mathbf{H}}|\boldsymbol{\theta}_{\tilde{C}})\right)\right\}, \tag{6.36}$$

$$\hat{C} = \arg\min_{\tilde{k}}\{\mathcal{C}_{\mathrm{ICOMP}}(\hat{\boldsymbol{\theta}}_{\tilde{C}})\}. \tag{6.37}$$

Comparative results for a simulation setup as in Table 6.2 (just PSO and no LM is employed) show the performance of the discussed and proposed algorithms in the Figures 6.10, 6.11 and 6.12. Firstly, note that the probability $P(\hat{C} \neq C)$ for the optimal performance is the FORLB. The FORLB is unequal to zero for the complete SNR range Υ_b. Moreover, the the probability as well as the MSE corresponding to $\hat{C} = C$ has the largest distance to the optimal performance in terms of the FORLB. Secondly, the MSE (Figure 6.10, right plot) shows that the ICOMP criterion is closest to the FORLB. This behaviour was expected, since the joint model order selection and parameter estimation via ICOMP is the only criterion that employs soft information about the delay estimates. For joint parameter estimation and model selection the violin plots in Figure 6.12 bear more information than the MSEs in Figure 6.10. The split violin plots show the SE distributions for a SNR range. Every algorithm is compared to the FORLB. The violin plot distribution show that all compared algorithms yield a TOA estimation accuracy that significantly outperforms the error achieved by correlation with $\mathbf{g}$ for the major part of the distribution, even for a realistic scenario with $C = 8$. For the realistic scenario and an increasing SNR, the violin plots exhibit widening error distributions. This can be explained by Figure 6.11, the estimated model order probability distributions. With higher SNR estimating a higher model order becomes more likely. At the same time the likeliness that either the delay estimation, the model order estimation or the soft information estimation has errors increases as well. Constraining the estimator to be even more parsimonious promises to be beneficial. This is not further investigated in this thesis, since there are different simple ways to achieve this. In principle it can be seen that joint model order

selection and parameter estimation is a useful tool in case of realistic multipath scenarios in combination with a limited amount of channel estimates.

Minimizing the remaining distance to the FORLB can be achieved by designing a criterion that focuses more on the TOA. An interesting approach called the focused information criterion is formulated by [CH01]. This approach is based on employing the estimation error bias. Whereas the estimator's variance increases with the model order, the estimation bias decreases. Therefore adding the bias to the variance yields a tradeoff and hence a criterion. That means, if we could estimate the delay estimation bias for all model orders, this criterion could improve the joint parameter estimation and model selection. Another idea also based on utilizing soft information is proposed in [Ric05]. The idea is to demand that the relative reliability associated to the estimated complex path weights is large enough. Then overestimating in every estimation step and excluding complex path weights that

$$\frac{\sigma^2_{\beta_c}}{|\beta_c|^2} \leq \varepsilon^2_{|\beta_c|} < 1 \tag{6.38}$$

provides a method to discard components via monitoring the estimates soft information. Nevertheless, for JCAP, the experience shows that overfitting entails the risk of high TOA estimation errors and hence I will not further investigate this approach.

Chapter 7

A Prototype JCAP Design

In this chapter I present a prototype system design, showcasing the overall system performance. After introducing the necessary theoretical background on interleave-division multiple access and interleave-division multiplexing, and shortly explaining the setup for the employed JCAP components, I will present the numerical results for this setup.

7.1 Multiplexing

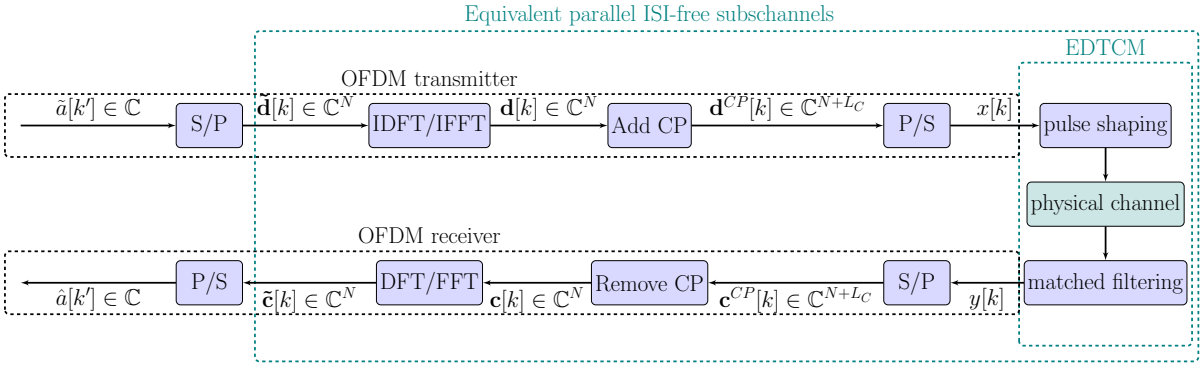


Figure 7.1: The OFDM block diagram shows that the system equations for the signal design are formulated in the frequency domain.

Multiplexing enables a user-specific flexible resource allocation, a feature especially desirable for JCAP. Multiplexing also separates different data streams of a single user. Two multiplexing schemes are especially suitable for JCAP: Firstly IDM and secondly OFDM. IDM is closely related to interleave division multiple access, a non-orthogonal multiple access scheme, which is highly desirable for future communication systems. Non-orthogonal access schemes theoretically promise to allow a higher number of simultaneous users. Furthermore, IDM offers flexible power allocation. OFDM, shown in Figure 7.1, on the other hand, offers a signal structure that is formulated in the frequency domain and hence is most suitable to be combined with techniques like ESPRIT. This can be interpreted as an advantage for JCAP, since delay estimation via ESPRIT can be carried out in a closed-form. Another advantage is that channel estimation for OFDM is simple. Due to the fact

that the cyclic convolution results in a point-wise multiplication in the frequency domain, channel estimation simplifies to parallel scalar divisions. Unfortunately, we learnt that the ESPRIT closed form solution requires oversampling. Consequently, I nevertheless prefer IDM for the prototype in this thesis and emphasize that OFDM can be used instead.

7.1.1 Interleave-Division Multiplexing (IDM)

The interleave division multiplexing (IDM) theoretical background is closely related to the principles of interleave-division multiple access (IDMA) [HSF08, SH04, Sch08]. If a single communication data sequence is transmitted, this resource is serial-to-parallel converted to N_d layers. Especially for JCAP, flexibility and adaptivity regarding the signal and training resource distribution are targeted.

7.1.1.1 IDM Transmitter

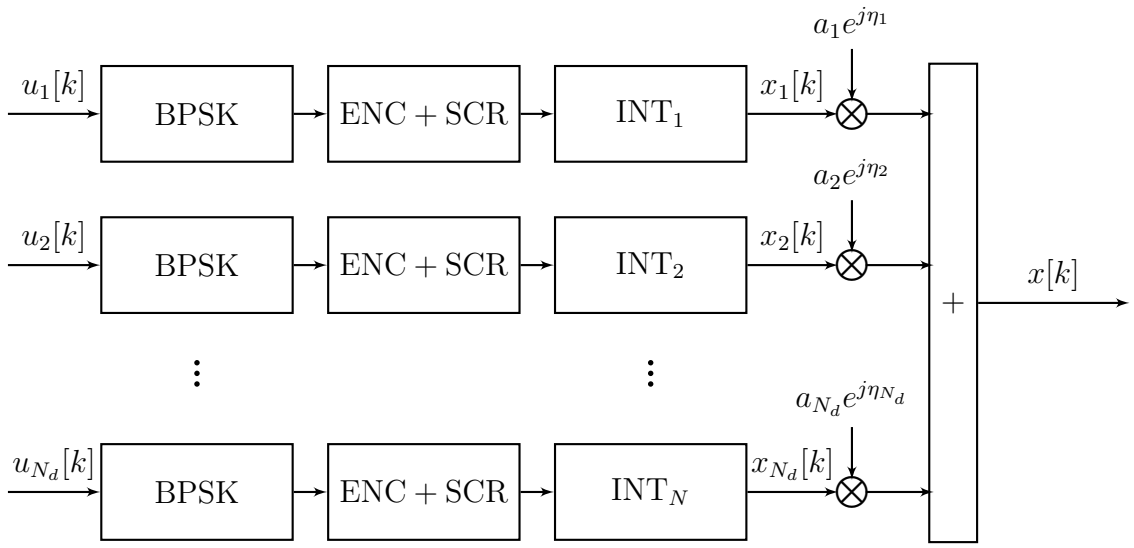


Figure 7.2: IDM transmittter

The conventional IDM transmitter signal has N_d parallel info bit sequences $u_n[k]$, with $1 \leq n \leq N_d$, parallely processed as so-called layers (Figure 7.2). First, they are modulated by binary phase-shift keying (BPSK). Then, the output is fed to an encoder (ENC), which, if not specified otherwise, is a simple repetition code with spreading factor S, followed by an scrambler (SCR) to generate a zero-mean data stream via statistically determined flipping. The scrambler output is interleaved by specific random interleavers (INT_n), leading to output $x_n[k]$. Finally, for each data layer, $x_n[k]$ is multiplied by complex weighting factors, built of layer-specific amplitudes a_n and a phase rotation $e^{j\eta_n}$,

the power allocation, before the layers' signals are superimposed:

$$x[k] = \sum_{n=1}^{N_d} a_n e^{j\eta_n} x_n[k], \tag{7.1}$$

$$a_n = \frac{1}{N_d}, \quad \eta_n = \frac{\pi n}{N_d} \quad \forall n: \quad 0 \leq n < N_d. \tag{7.2}$$

7.1.1.2 Receiver: Elementary Signal Estimator

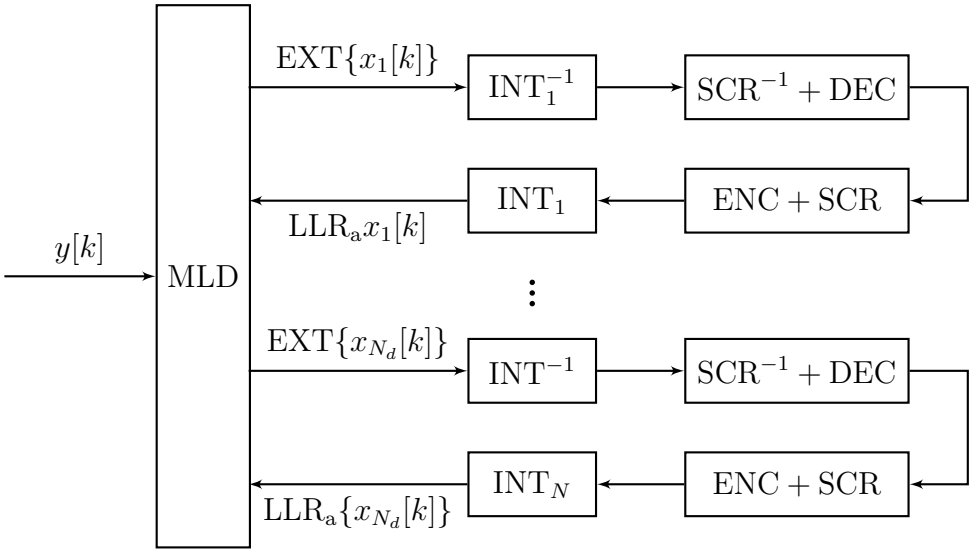


Figure 7.3: IDM receiver

The IDM receiver mainly consists of two parts: The multi-layer detector (MLD) that calculates single layer log-likelihood ratios (LLR)s and the iterative feedback cycle including the deinterleavers, the descramblers, the decoders, the encoders, the scramblers and the interleavers, for all layers, and feeding the extrinsic layer information back to the MLD (Figure 7.3). The LLR and the extrinsic information reads

$$\mathrm{LLR}\{x_n[k]\} = \underbrace{\ln \frac{p(\mathbf{y}|x_n[k]=+1)}{p(\mathbf{y}|x_n[k]=-1)}}_{\mathrm{EXT}\{x_{N_d}[k]\}} + \underbrace{\ln \frac{p(x_n[k]=+1)}{p(x_n[k]=-1)}}_{LLR_a\{x_n[k]\}} \tag{7.3}$$

The MLD consists of N_d multi-layer interference cancellations, serially concatenated with N_d single-layer detectors, to estimate the Gaussian noise mean and variance that can be calculated for each single layer.

7.2 Multiple Access

This section introduces the applied multiple access scheme in this thesis and explains the interplay with IDM.

7.2.1 IDMA and ML-IDMA

IDMA was originally proposed by the authors of [PLL03], [PLWL06] as a promising non-orthogonal code divsion multiple access (CDMA) alternative to classical orthogonal methods like the related direct sequence code division multiple access (DS-CDMA). Recently, it has gained a lot of attention as a suitable $5G$ candidate, due to its non-orthogonal nature. Orthogonality principally limits the number of users due to a limited number of code sequences, and non-orthogonal schemes are supposed to overcome these limitations. Therefore, I chose IDMA as a suitable candidate for my prototype simulations.
The concept of IDMA and IDM are related. Separating different users by different interleavers is called IDMA. Separating different data streams, called layers, by different interleavers is called IDM. A typical IDMA transmitter block diagram is shown in Figure 7.4 a). Each user's data is first coded and scrambled (for zero-mean sequences), then interleaved by random user-specific interleavers. The user-specific interleavers are random permutations of the data sequence itself. Afterwards, the chosen mapping is applied. Each user is physically separated and hence is transmitted over user-specific channels. The received signal is the summation of all the user-specific transmissions with AWGN. The combination of both IDM and IDMA is called Multi-layer IDMA (ML-IDMA). The concept is outlined in Figure 7.4 b).

7.3 Transmitter: Pilot Layer plus ML-IDMA

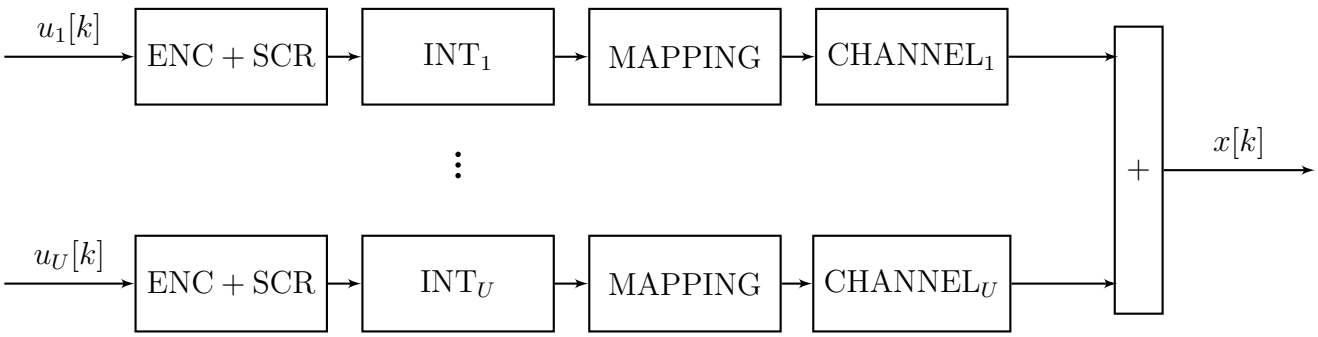
Here, in this thesis I choose ML-IDMA for multiplexing and multiple access in order to provide non-orthogonal prototype like simulation results. Figure 7.4 b) shows the complete transmit unit including a pilot layer for each user that serves both communication and positioning purposes. The pilot layer is required for channel estimation, which, in return, is required for data detection (communication) as well as for high-resolution delay estimation (positioning). Each user is physically separated and therefore each user is transmitted via distinct EDTCMs. The received signal is obtained by the summation of the different user signals transmitted over distinct channels, and adding WGN. Considering equidistant time sampling, the IDM signals can be expressed as

$$v_u[k] = \sum_{n=0}^{N_d} a_n e^{j\eta_n} x_{u,n}[k] = \underbrace{\underbrace{\sum_{n=1}^{N_d} a_n e^{j\eta_n} x_{u,n}[k]}_{\text{data}} + \underbrace{a_0 e^{j\eta_0} p_u[k]}_{\text{pilot}}}_{\text{virtual pilot}} = v_u[k]. \tag{7.4}$$

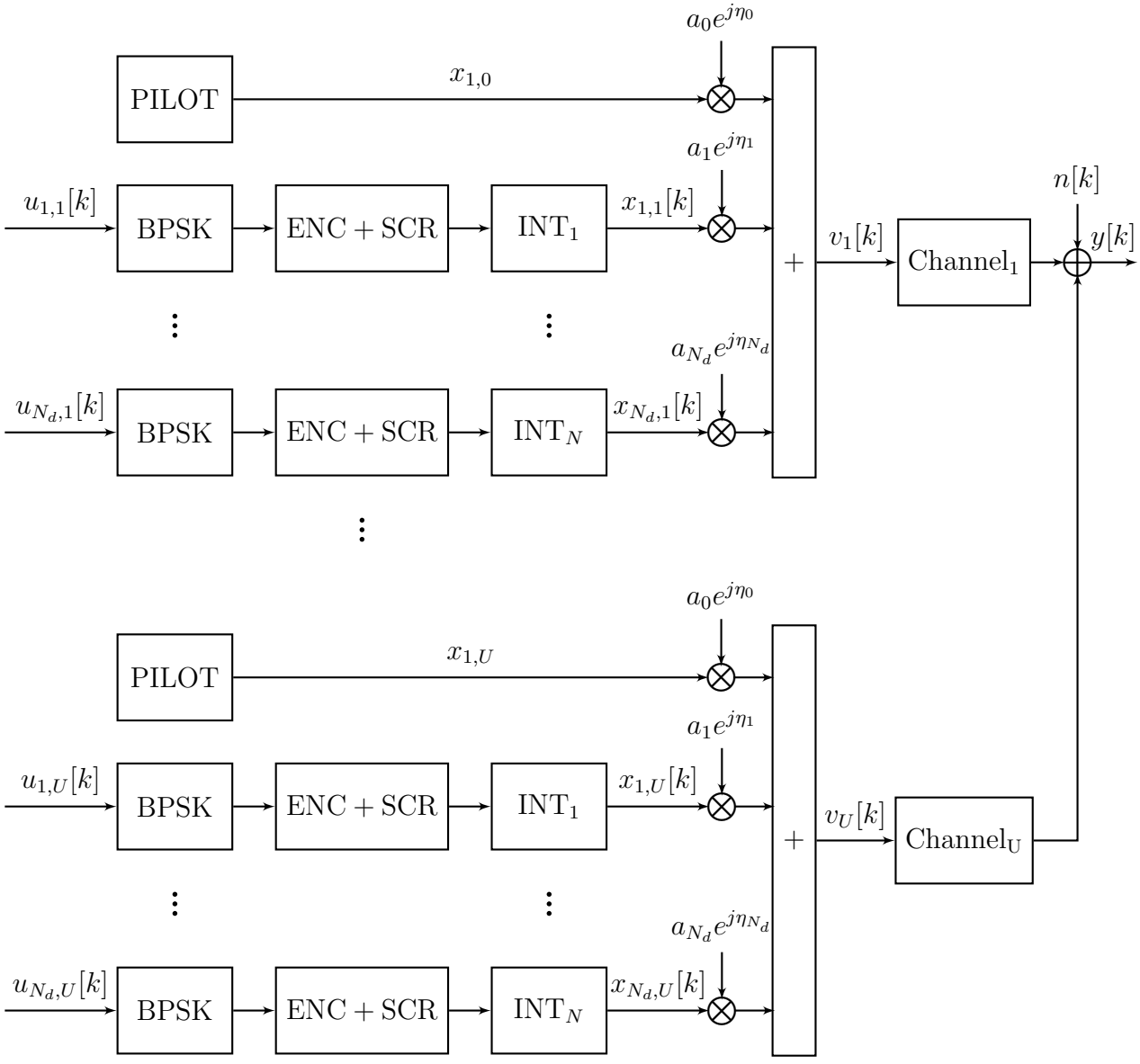
Consequently, the overall number of layers is $N = N_d + 1$. I choose the user-specific pilot sequence design in such a manner that it meets the optimal autocorrelation properties as proposed in [Chu72], as a possible realization of complex optimal sequences [NLM98]:

$$p_u[k] = p_n = \begin{cases} e^{j\pi(n^2 r/(UK)} & \text{if } UK \text{ is even} \\ e^{j\pi(n\cdot(n+1)r/(UK)} & \text{else,} \end{cases} \tag{7.5}$$

$$n = (u-1)K + k, \quad 1 \le u \le U, \quad 0 \le k \le K-1, \tag{7.6}$$



(a) IDMA transmittter



(b) ML-IDMA transmittter

Figure 7.4: a) IDMA transmitter and b) ML-IDMA transmitter.

where r is a length that is relative prime to the observation length K. Let us stack all the user-specific EDTCM channels in a redefined overall channel vector

$$\mathbf{h} = [\mathbf{h}_1^T \dots \mathbf{h}_U^T]^T. \tag{7.7}$$

Then, the ML-IDMA signal can be expressed as

$$y[k] = \sum_{u=1}^{U} \sum_{l=0}^{L} v_u[k-l] h_{(u-1)\cdot(L+1)+l} + n[k], \tag{7.8}$$

for user $1 \leq u \leq U$ and for discrete time indices $0 \leq k \leq K+L$. Let us define the following matrices in order to obtain a matrix vector notation. Let $\mathbf{V}$ be the horizontal concatenation of the users' IDM signal matrices, in Toeplitz structure, such that it can be interpreted as a virtual training matrix

$$\mathbf{V} = [\mathbf{V}_1 \dots \mathbf{V}_U] \tag{7.9}$$

$$[\mathbf{V}_u]_{k,l} = \begin{cases} v_u[k-l] & \text{if } L+1 \leq k \leq K \\ 0 & \text{else} \end{cases}, \quad \text{for } 1 \leq u \leq U. \tag{7.10}$$

Then, this can be put as

$$\mathbf{y} = \mathbf{V}\mathbf{h} + \mathbf{n} = (\mathbf{X} + \mathbf{P})\mathbf{h} + \mathbf{n}. \tag{7.11}$$

7.4 Receiver Design: Iterative JCAP Estimator

The chosen prototype receiver design (Figure 7.5) unifies communication and positioning by feeding as well an MLD algorithm for IDM(A) (for communication) as the multipath parameter delay estimation (for positioning) with semi-blind channel estimates. Therefore, the channel estimation performance is crucial for this system approach.

7.4.1 Pilot Interference Cancellation

In a first pilot interference cancellation (PIC) step, the known pilot interference is subtracted from the received signal to further pass this signal to the MLD as the input for the ith iteration:

$$\mathbf{y}^{i+1} = \mathbf{y} - \mathbf{P}\hat{\mathbf{h}}^{i+1}. \tag{7.12}$$

This interference cancellation becomes more accurate with an increasing number of iterations.

7.4.2 Iterative Elementary Signal Estimator for Multi-User and Multi-Layer Detection

Each layer's data of each user is estimated iteratively according to the elementary signal estimator principle: The MLD estimates the LLR of each layer $n \in \{1, \dots, N_d\}$ for each

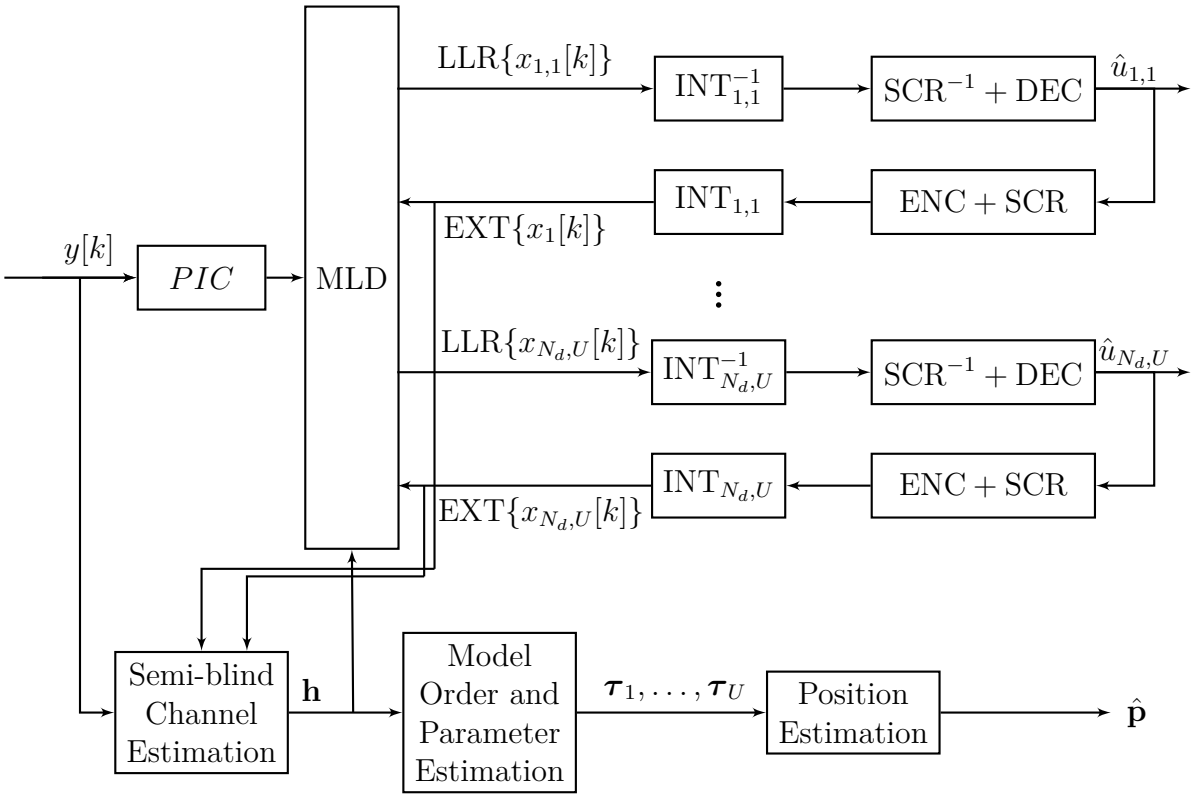


Figure 7.5: The JCAP ML-IDMA receiver uses a semi-blind core channel estimation, to simultaneously serve data detection, parameter estimation and model selection. Via iteratively exchanging mutual information, the data energy is used for high-resolution parameter estimation and therefore it enhances the positioning performance.

user $u \in \{1, \ldots, U\}$, gained by treating the superposition of other layers and users as a distortion, for which the mean and variance can be estimated. Based on these estimates, the LLR values for each layer and each user are calculated iteratively, via employing the extrinsic information due to coding and decoding. More specifically, this means that the $\mathrm{LLR}_{\mathrm{n,u}}^{\mathrm{i+1}}[k]$ is calculated according to the following principle:

$$\breve{y}_{n,u,l}^{i+1}[k] = h_{u,l} x_{u,n}[k] + s_{u,n,l}^{i+1}[k] \tag{7.13}$$

$$= y^{i+1}[k] - \left(\sum_{\substack{u'=1 \\ u' \neq u}}^{U} \sum_{\substack{l'=0 \\ l' \neq l}}^{L} h_{(u',l'} \sum_{\substack{n'=1 \\ n' \neq n}}^{N_d} a_{n'} e^{j\eta_{n'}} x_{u'.n'}[k] \right). \tag{7.14}$$

By determining the distortion's $s_{u,n,l}[k]$ mean value $\mathrm{E}\{\mathrm{s_{u,n,l}[k]}\}$ and variance $\sigma^2_{s_{u,n,l}[k]}$

$$\mathrm{E}\{s_{u,n,l}[k]\} = 0, \tag{7.15}$$

$$\sigma^2_{s_{u,n,l}[k]} = \sum_{u=1}^{U} \sum_{l=0}^{L} \sigma^2_{h_{u,l}}[k] \sum_{n=1}^{N_d} \sigma^2_{\tilde{x}_{u.n}^{i+1}}[k-l] + \sigma_n^2, \tag{7.16}$$

the log-likelihood ratios for user u, layer n, time instant k at iteration $i+1$ reads as in [SH06]

$$\mathrm{LLR}\{\tilde{x}_{n,u}^{i+1}\}[k] = \sum_{l=0}^{L} 4 \cdot \frac{\mathrm{Re}\{\hat{h}_{u,l}^{*,i+1}[k+l] \cdot \breve{y}_{u,n,l}^{i+1}[k+l]\}}{\sigma_{s_{u,n,l}}^2[k+l]}. \tag{7.17}$$

7.4.3 Core Iterative Semi-Blind Channel Estimation for Both Communication and Positioning

The channel estimation is used for both communication and positioning. Its performance is indirectly related to the positioning performance via the delay estimation error, which is affected directly. The channel estimates are calculated as

$$\hat{\mathbf{h}}^{i+1} = \mathbf{V}^{i\dagger}\mathbf{y}. \tag{7.18}$$

7.4.4 Joint Model Order Selection and Delay Estimation

7.4.4.1 Model Order is known Beforehand

For delay estimation it is reasonable to assume that a coarse delay estimation was successfully carried out beforehand, meaning that the delay relevant for positioning that we have to estimate lies within the range $[-\frac{T}{2}, \frac{T}{2}]$. If we assume that the number of multipaths is known, we can obtain the high-resolution multipath delay estimates by solving the SNLLS problem equivalent to the ML problem, introduced in (5.23), by

$$\hat{\boldsymbol{\tau}} = \arg\min_{\tilde{\boldsymbol{\tau}}} \{\hat{\mathbf{h}}^H \mathbf{P}_{\mathbf{G}}^{\perp} \hat{\mathbf{h}}\}. \tag{7.19}$$

7.4.4.2 Jointly Estimating the Model Order and the Parameters

For every hypothetical model order $\tilde{m}$, I first estimate the delays, the associated linear complex path weights, and the associated likelihoods of these events. Then I substitute

these likelihoods into the information criterion to choose the model order:

$$\hat{\boldsymbol{\tau}}_{\tilde{m}} = \arg\min_{\tilde{\boldsymbol{\tau}}_{\tilde{m}}} \{\hat{\mathbf{h}}^H \mathbf{P}_{\mathbf{G}}^{\perp} \hat{\mathbf{h}}\}, \tag{7.20}$$

$$\hat{\boldsymbol{\beta}}_m = \mathbf{G}^{\dagger} \hat{\mathbf{h}}, \tag{7.21}$$

$$\hat{\boldsymbol{\theta}}_m = [\hat{\boldsymbol{\tau}}_m, \mathrm{Re}\{\hat{\boldsymbol{\beta}}\}, \mathrm{Im}\{\hat{\boldsymbol{\beta}}\}], \tag{7.22}$$

$$\hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}_{\tilde{m}}} = \mathbf{F}_{\text{full}}^{-1} \left(\hat{\boldsymbol{\theta}}_{\tilde{m}} \right), \tag{7.23}$$

$$\text{calculate the eigenvalues } \lambda_1, \ldots \lambda_{\tilde{m}} \text{ of } \hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}_{\tilde{m}}}. \tag{7.24}$$

$$\mathcal{C}_{\text{ICOMP}}(\hat{\boldsymbol{\theta}}_m) = -2\mathcal{L}\left(\hat{\boldsymbol{\theta}}_{\tilde{m}} \right) + m \cdot \ln \left(\frac{\frac{1}{\tilde{m}} \sum_{j=0}^{\tilde{m}} \lambda_j}{\left(\prod_{j=1}^{\tilde{m}} \lambda_j \right)^{1/\tilde{m}}} \right) \tag{7.25}$$

$$\hat{m} = \arg\min_{\hat{\boldsymbol{\theta}}_{\tilde{m}}} \{\mathcal{C}_{\text{ICOMP}} \left(\hat{\boldsymbol{\theta}}_{\tilde{m}} \right)\}. \tag{7.26}$$

7.4.5 Theoretical Bounds

In the following I briefly summarize the bounds required for the prototype system design.

7.4.5.1 Channel Estimation CRLBs

In [SH06], training based (TB), TB channel estimation with interference cancellation (TB-IC), and semi-blind (SB) channel estimation methods were compared. The TB method is non-iterative and reveals the worst performance. It is used to initialize the data detection in the first iteration step, to calculate soft data estimates, which are iteratively exchanged via the TB-IC and sb estimators. With the power normalization as chosen in this thesis, these bounds can be expressed as

$$\underbrace{\frac{\sigma_n^2 + \sum_{u=1}^{U} \sigma_{h_u}^2 \sigma_{\tilde{x}_u^i}^2}{K \cdot S \cdot (\rho + (1-\rho) \sum_{u=1}^{U} |\tilde{x}_u^i|^2)}}_{crlb(h_{l,u})^{\text{sb}}} \leq \underbrace{\frac{\sigma_n^2 + \sum_{u=1}^{U} \sigma_{h_u}^2 \sigma_{\tilde{x}_u^i}^2}{K \cdot S \cdot \rho}}_{crlb(h_{l,u})^{\text{TB-IC}}} \leq \underbrace{\frac{\sigma_n^2 + \sum_{u=1}^{U} \sigma_{h_u}^2}{K \cdot S \cdot \rho}}_{crlb(h_{l,u})^{\text{TB}}}. \tag{7.27}$$

Figure 7.6 strikingly exposes the difference in the theoretically approachable mean squared error performance for the three estimators. Both the TB-IC and the sb estimator depend on the soft chip variance $\sigma_{\tilde{x}_u^i}^2 = 1 - |\tilde{x}_u^i|^2$. Depending on the iteration count and the signal-to-noise ratio, the chip variance approaches a value close to zero. For high SNRs and after a sufficient number of MLD iterations, this value becomes negligible, and the loose sb bound becomes valid. The performance difference between the semi-blind estimator, the TB-IC, and the tb estimator explains the choice to use the sb estimator for JCAP, that relies on an optimal channel estimation MSE performance. The loose lower bound can be

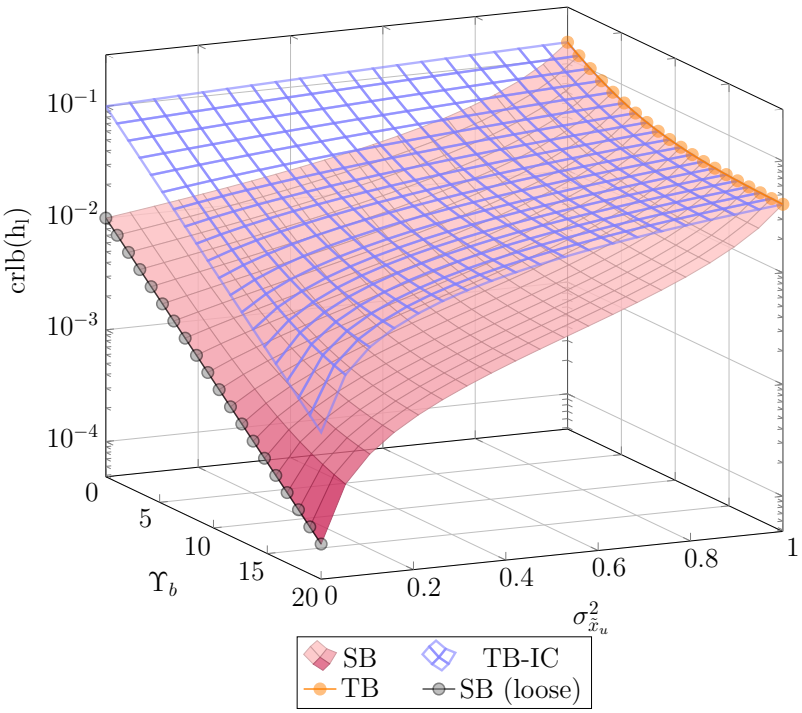


Figure 7.6: The difference of the theoretical CRLBs for channel estimation is shown for $\rho = 0.1$ and depends on the chosen method (training-based: TB, training based with interference cancellation TB-IC or semi-blind: SB). The black loose SB bound provides the lowest bound.

approached for high signal-to-noise ratios and a certain number of iterations ($\sigma^2_{\tilde{x}_u^i} \to 0$), as can be seen in the numerical results.

7.4.5.2 Delay Estimation CRLBs and FORLB

The delay CRLB is then, with (7.27) and (5.152, 5.147), determined by

$$\mathrm{crlb}^{TS,SIMO}_{\mathrm{SNLLS},\boldsymbol{\tau}}(\boldsymbol{\tau}) = \frac{\mathrm{crlb}_{\mathrm{sb}}(\mathrm{h}_{\mathrm{l,u}})}{2}\left[\mathrm{Re}\left\{\mathbf{J}_{\mathbf{G}}^T\mathbf{P}_{\mathbf{G}}(\boldsymbol{\tau})^{\perp}\mathbf{J}_{\mathbf{G}} \odot ((\mathbf{G}(\boldsymbol{\tau})^{\dagger}\hat{\mathbf{H}})^H\mathbf{G}(\boldsymbol{\tau})^{\dagger}\hat{\mathbf{H}})^T\right\}\right]^{-1} \tag{7.28}$$

$$\mathrm{crlb}_{\mathrm{SNLLS},\boldsymbol{\tau}}(\boldsymbol{\tau}) = \frac{\mathrm{crlb}_{\mathrm{sb}}(\mathrm{h}_{\mathrm{l,u}})}{2}\left[2\mathrm{Re}\left\{\mathrm{diag}\left(\hat{\mathbf{h}}^H\mathbf{G}^{\dagger^T}\right)\mathbf{J}_{\mathbf{G}}^T\mathbf{P}_{\mathbf{G}}(\tilde{\boldsymbol{\tau}})^{\perp}\mathbf{J}_{\mathbf{G}}\mathrm{diag}\left(\mathbf{G}^{\dagger}\hat{\mathbf{h}}\right)\right\}\right]^{-1}. \tag{7.29}$$

The bound that yields a lower performance measure for the joint model order selection and parameter estimation is the focused order-related bound:

$$\hat{C}_{\mathrm{opt}} = \arg\min_{\hat{\tau}_{1,\tilde{C}}}\{(\hat{\tau}_1 - \hat{\tau}_{1,\tilde{C}})^2\} \tag{7.30}$$

$$\mathrm{forlb}(\tau_1) = (\tau_1 - \hat{\tau}_{1,\hat{C}})^2. \tag{7.31}$$

7.5 Numerical Results

The numerical results I present in the following have the purpose to give an overall impression about JCAP components performance and their mutual impact.

7.5.1 The Basic Simulation Setup

As a lower performance bound and reference, I will first provide error plots for the BER for IDM and a complex-valued single-path channel. Afterwards results for a two-path channel and a WINNER B1-LOS model are assessed. The samples $g(lT - \tau_1)$ are obtained by employing a windowed sampled raised cosine pulse with rolloff factor 0.3 and half window width $W = 4T$, leading to $L_g = 2W = 8$. To cope with the later investigated multipath scenario the channel memory length is set to $L = 10 > L_g$. The information sequence with length $K_d = 100$ is spread by $S = 10$ to $K = SK_d = 1000$. The bit load $\mathcal{B} = RN_dU$ is here specified by $\mathcal{B} = N_dU/S$. In the single-path channel a zero delay is assumed, otherwise $\tau_1 \in \mathcal{U}(-T/2, /T/2)$. The MLD uses 20 receiver iterations before hard estimates for $\hat{\boldsymbol{\tau}}_u$ and $\hat{u}_{n,u}[k] \quad 0 \leq k \leq K$, $1 \leq n \leq N_d$, $1 \leq u \leq U$ are calculated. The interleavers in the MLD are random interleavers. All data sequences are generated randomly.

7.5.2 Power Normalization for ML-IDMA, SNR and PNR

The symbol-wise signal-to-noise ratio (SNR) per user Υ_s consists of the pilot-to-noise ratio (PNR) and the data-to-noise ratio by weighting it via the pilot power ratio ρ. For a single user, the weighting is

$$\Upsilon_s = \rho\Upsilon_s + (1-\rho)\Upsilon_s = \Upsilon_p + \Upsilon_d. \tag{7.32}$$

The layer-specific symbol-wise and data-wise SNR is calculated by

$$\Upsilon_{s,n,u} = \frac{P_{n,u}}{\sigma_n^2}. \tag{7.33}$$

Consequently, the overall SNR sums up to

$$\Upsilon_{\Sigma s} = \frac{E_s}{\sigma_n^2} = \sum_{u=1}^{U}\sum_{n=1}^{N}\frac{P_{n,u}}{\sigma_n^2}. \tag{7.34}$$

The user- and layer-specific power is calculated via

$$P_{n,u} = |a_{n,u}e^{j\eta_{n,u}}|^2 = \begin{cases} \rho & \text{if } n = 0 \\ \frac{(1-\rho)}{N_d} & \text{else,} \end{cases} \tag{7.35}$$

$$\Upsilon_{\Sigma d} = \sum_{u=1}^{U}\Upsilon_{d,u} = \sum_{u=1}^{U}\sum_{n=1}^{N_d}\Upsilon_{s,n,u}. \tag{7.36}$$

Here, every user has equal data power $\Upsilon_{d,u} = \Upsilon_d$. Note that I further normalize the channel power per user to 1. The bit-wise SNR or the bit-to-noise ratio is calculated by

$$\Upsilon_{b,m} = \frac{1}{\mathcal{B}}\Upsilon_d = \frac{S}{N_d \cdot U}\Upsilon_d = \frac{S}{N_d \cdot U}(1-\rho)\Upsilon_s, \tag{7.37}$$

$$\Leftrightarrow \Upsilon_s = \frac{N_d \cdot U}{S}\frac{1}{1-\rho}\Upsilon_{b,m}, \tag{7.38}$$

since $\Upsilon_d = (1-\rho)\Upsilon_s$. For the simulations in this thesis I assume that $\sum_{n=1}^{N} P_{n,u} = 1$ and hence $E_s = U$. Consequently the following equations hold:

$$\Upsilon_b = \frac{S}{N_d \cdot \sigma_n^2}, \tag{7.39}$$

$$\Upsilon_{\Sigma s} = \frac{U}{\sigma_n^2}, \quad \Upsilon_s = \frac{1}{\sigma_n^2}, \tag{7.40}$$

$$\Upsilon_{\Sigma d} = \frac{U(1-\rho)}{\sigma_n^2}, \quad \Upsilon_d = \frac{(1-\rho)}{\sigma_n^2}, \tag{7.41}$$

$$\Upsilon_{\Sigma p} = \frac{U\rho}{\sigma_n^2}, \quad \Upsilon_p = \frac{\rho}{\sigma_n^2}. \tag{7.42}$$

Therefore, we get

$$\Upsilon_s[\text{dB}] = \Upsilon_{b,m}[\text{dB}] + 10 \cdot \log 10\left(\frac{1}{1-\rho} \cdot \frac{N_d}{S}\right) \tag{7.43}$$

With (7.32) and (7.38) it is easy to see that

$$\Upsilon_p = \frac{N_d}{S} \cdot \frac{\rho}{1-\rho}\Upsilon_{b,m}, \tag{7.44}$$

$$\Upsilon_{\Sigma p} = \frac{N_d \cdot U}{S} \cdot \frac{\rho}{1-\rho}\Upsilon_{b,m}, \tag{7.45}$$

and

$$\Upsilon_{\sum p}[\text{dB}] = \Upsilon_{b,m} + 10 \cdot \log 10\left(\frac{\rho}{1-\rho} \cdot \frac{N_d \cdot U}{S}\right). \tag{7.46}$$

7.5.3 IDM and IDMA for AWGN

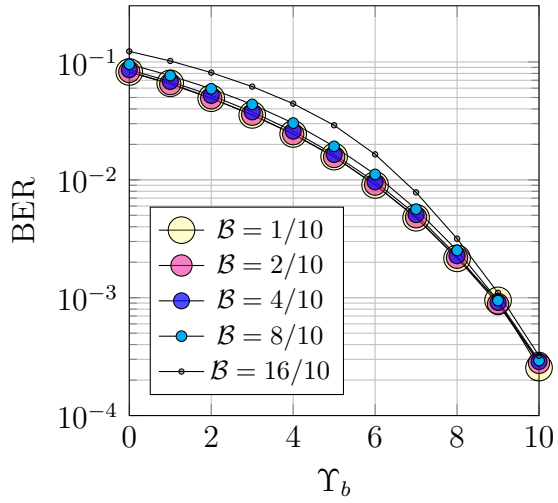
To begin with, the BER performance for a single-path AWGN channel, without delay, is investigated in Figure 7.7. For perfect channel knowledge, all BER curves converge to the single-layer performance for increasing SNR and bit loads $\mathcal{B} < 2$, as expected. For bit loads $\mathcal{B} \geq 2$, the BER curves do not converge to the single layer performance with increasing SNR [Sch12]. Furthermore, the TB-IC, and SB channel estimation BER results show a similar behavior, that is, nearly as good as the perfect channel knowledge scenario. Naturally, the TB channel estimation results in the poorest BER performance compared

to the other two approaches. Nevertheless, one could argue, that the TB results may be interpreted as satisfactory to employ it for pure communication receiver purposes, taking into account that it has the smallest complexity, due to the fact that channel estimation is not included in the iterative receiver processing. How PEACE can yield synergy here was already investigated thoroughly for the case of IDM in [Sch12]. The impact on the BER was so insignificant that it has to be decided from scenario to scenario, whether the performance gain is worth the extra computational cost.

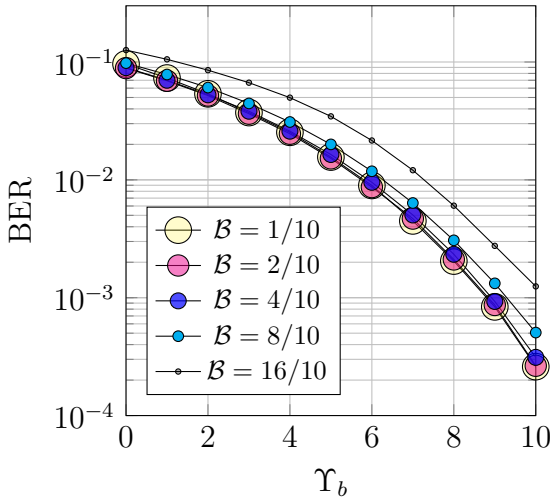
Investigating the BER plots for IDMA, we similarly see that, for perfect channel knowledge, the BER plots for multiple users converge to the single-user scenario with increasing SNR. Furthermore, to maintain this behavior, the bit load B should be kept below 2, which can be seen by comparing Figure 7.8 a) and b). Employing purely training-based channel estimation results in a comparably poor convergence behavior, which can be seen in Figure 7.8 c). Here, employing semi-blind channel estimation can obtain a convergence behavior close to the single-layer performance, for all simulated user scenarios.
The channel estimation MSE performance, plotted in Figure 7.9, indicates: Firstly, comparing the performance gains for IDM TB, IDM TB-IC channel estimation MSEs with the MSEs of IDM SB semi-blind channel estimation should clearly be applied for positioning via channel estimation. Furthermore, it becomes clear that for the channel estimation performance increasing the number of layers degrades the performance more than increasing the number of users. Finally, it can be seen that all MSEs approach the CRLB with an increasing PNR.
For the ML-IDMA single path case shown in figure 7.10 $U \cdot N_d$ is kept constant.

7.5.4 Two-path Channel

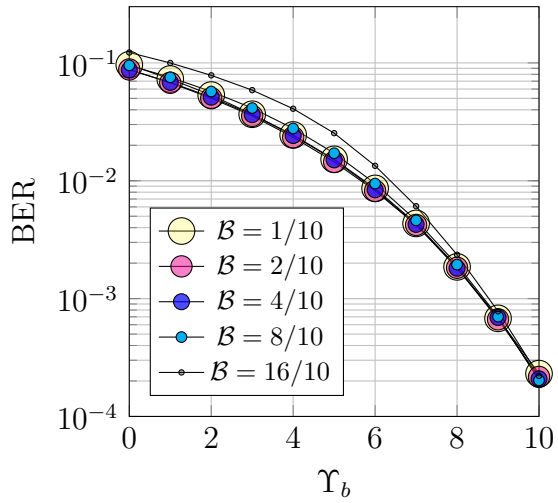
The numerical results shown for the two-path example in Figure 7.11 and in Figure 7.12 have the purpose to demonstrate the chosen algorithms' uncompromised theoretical performance. All error plots asymptotically attain the lower bounds. As expected, it can be seen that the BER is flatter compared to the single-path example, due to fading and imperfectly calculated LLRs in the multi-layer detector. Note that, due to a delay estimation bias, induced by a practically bounded search space, the delay estimation distribution plotted as violin plots lie slightly below the CRLB. For the RSE positioning results it can be seen twofold: Firstly, employing soft information enhances the RSE. Secondly, a further slight enhancement is achieved by sequentially employing the WLS and the TS positioning algorithm. Note that, due to the fact that the soft information for the positioning algorithm is gained by calculating the estimate's inverse approximate Fisher information matrix, this information is inexact for an insignificant percentage of all delay estimates. Either the estimates themself are flawed, or the Fisher information was inexact. This leads to outliers, resulting in a broad range within the violin plot distribution. Mechanisms to better detect the soft information reliability can further improve this comparison in favor of the estimation relying on soft estimates.



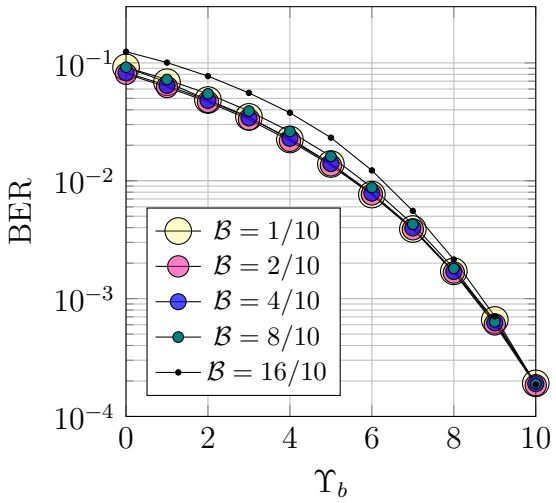
(a) Perfect CSI



(b) TB CE

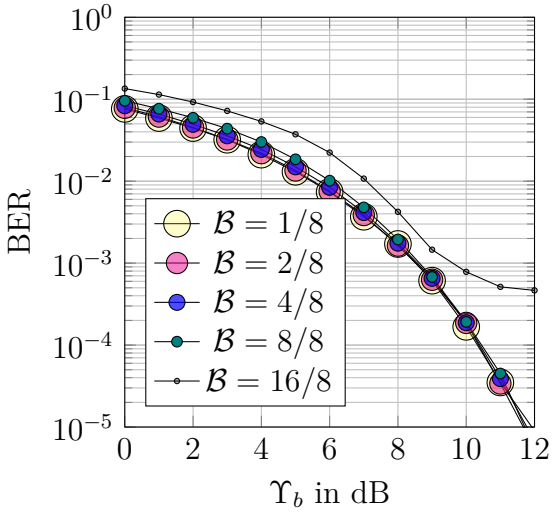


(c) TB-IC CE

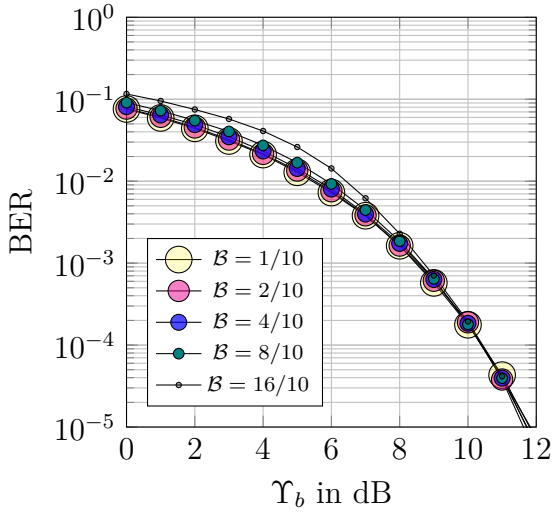


(d) SB CE

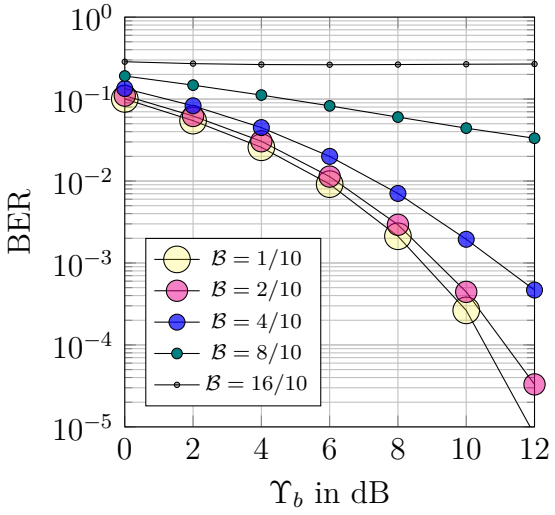
Figure 7.7: BER for IDM with complex AWGN without delay: The BER performance shows that employing SB-CE or TBIC-CE performs almost as well as having perfect channel knowledge. For communication purposes TB-CE is also able to perform well for a low number of layers.



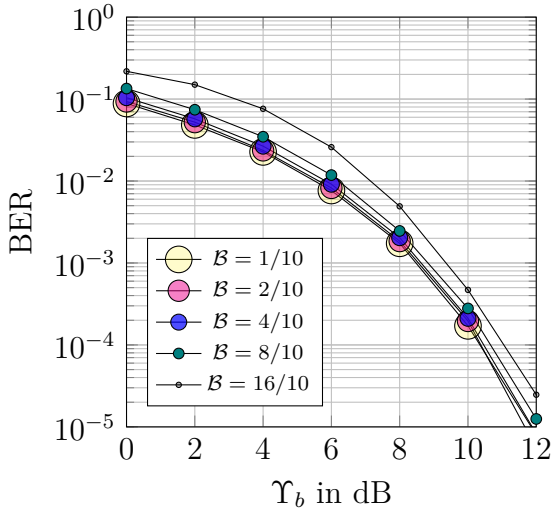
(a) Perfect CSI, ($S = 8$)



(b) Perfect CSI and $S = 10$)

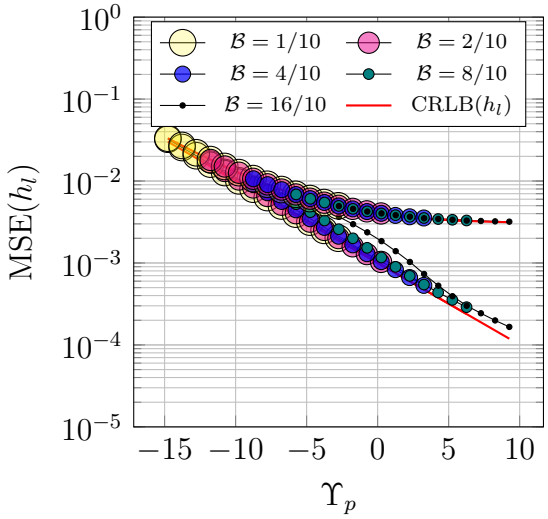


(c) TB CE, ($S = 8$)

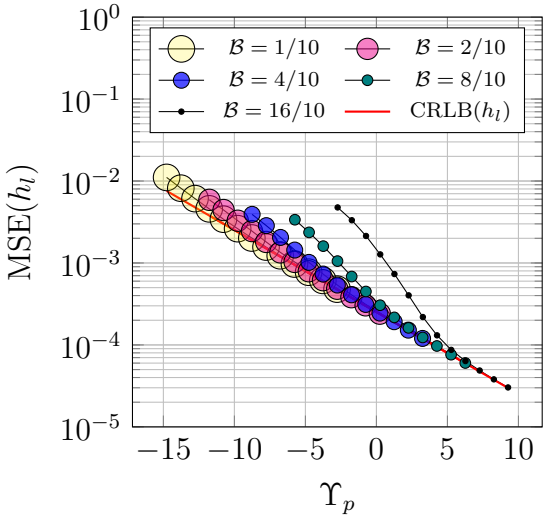


(d) SB CE and $S = 10$)

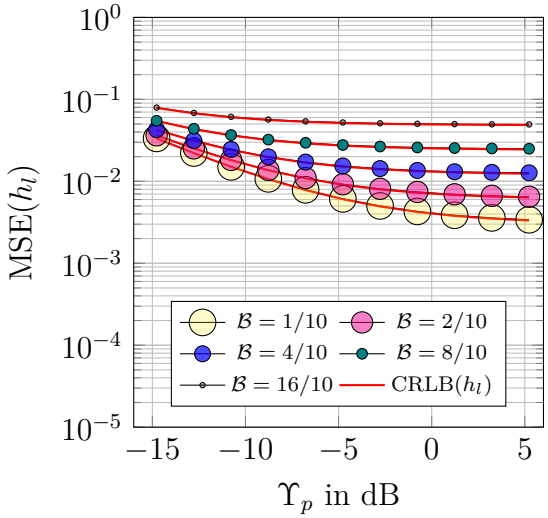
Figure 7.8: BER performance for IDMA with a complex valued AWGN channel without delay: For convergence the plots indicate that keeping the bit-load below 2 and employing SB CE is advisable.



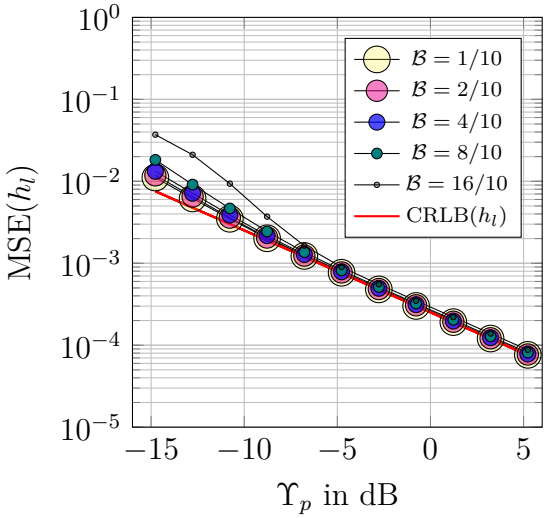
(a) IDM TB and TB-IC CE



(b) IDM SB CE

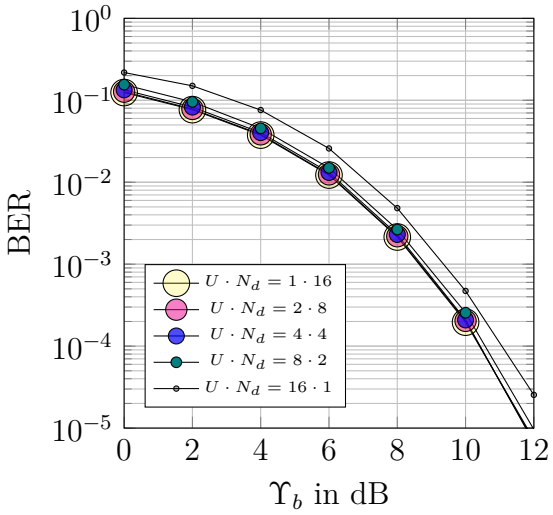


(c) IDMA TB CE

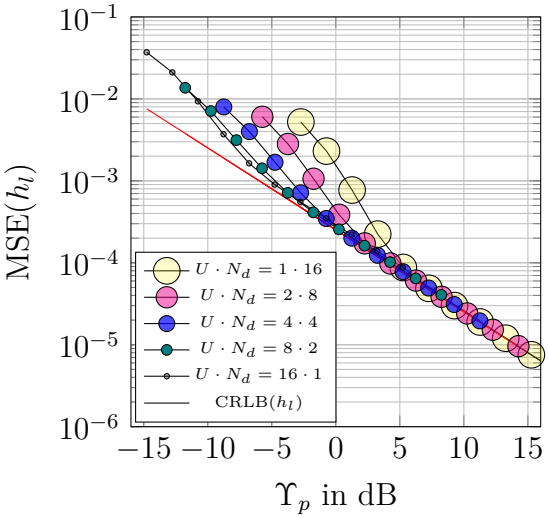


(d) IDMA SB CE

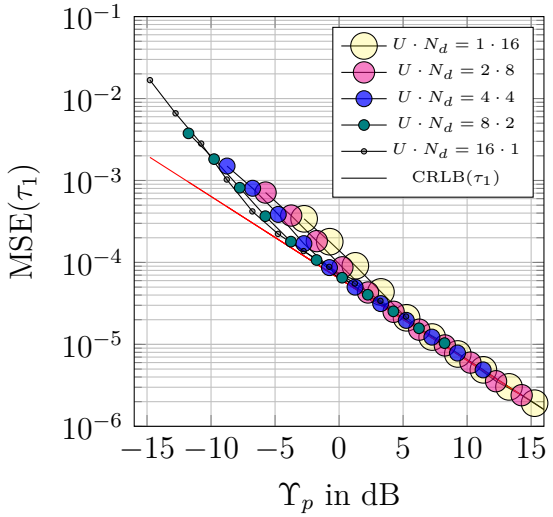
Figure 7.9: MSE TB, TB-IC and SB channel estimation performance comparison for IDM and IDMA for a complex-valued AWGN channel without delay: For both IDM and IDMA TB CE MSE plots attend an error floor for increasing SNR (a) and c)). As expected SB CE shows the best performance, although for IDM it can be seen that in order to attend the CRLBs with increasing layer number the SNR has to be increased as well.



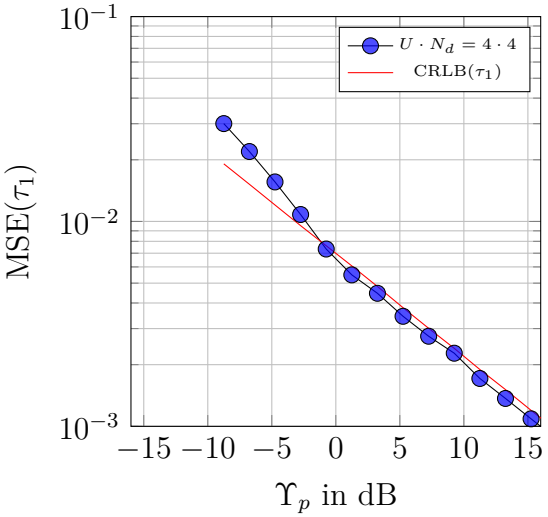
(a) BERs for communication



(b) CE MSEs for JCAP

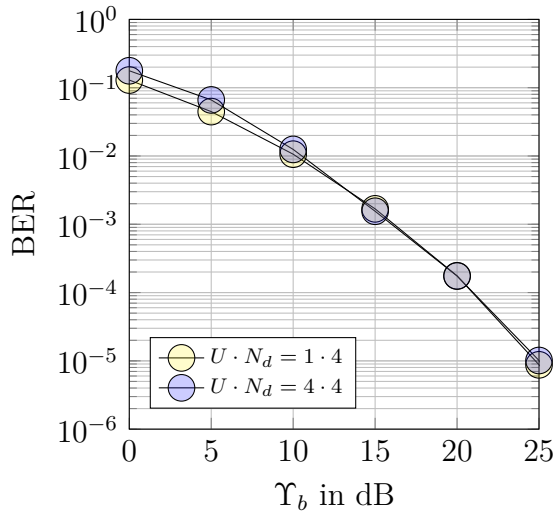


(c) DE MSEs for positioning

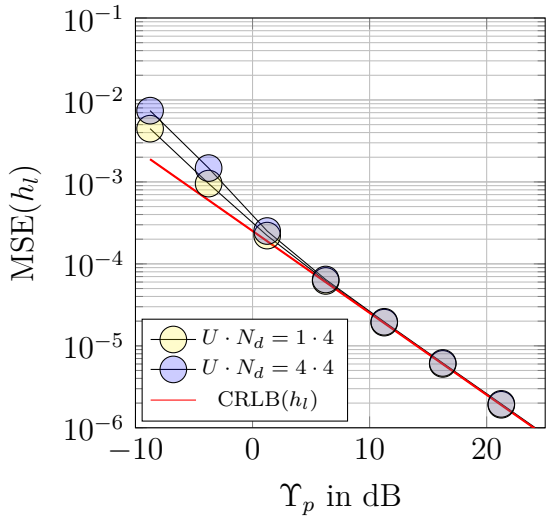


(d) DE MSEs for positioning

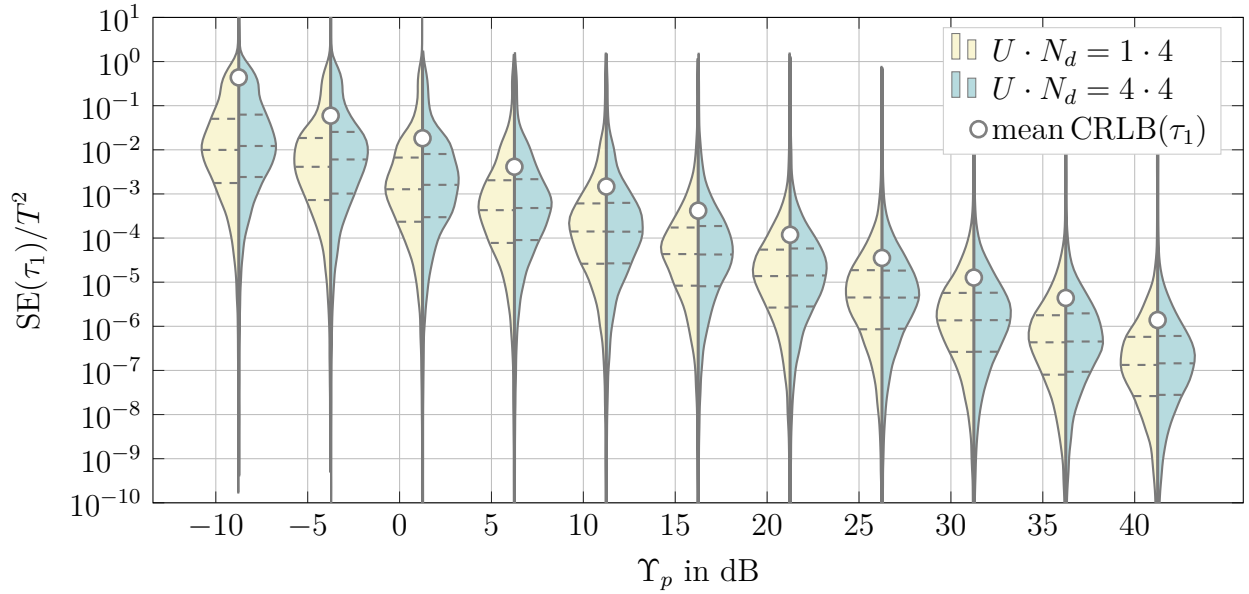
Figure 7.10: The BER and MSE performance for channel estimation and delay estimation is summarized for a complex-valued AWGN channel with zero delay: Whereas for the BER plots we see that keeping $U \cdot N_d$ constant confirms that IDMA has worse convergence properties than IDM, for the MSE performance the behavior is the other way around: IDM shows worse convergence behavior than IDMA.



(a) BERs for communication

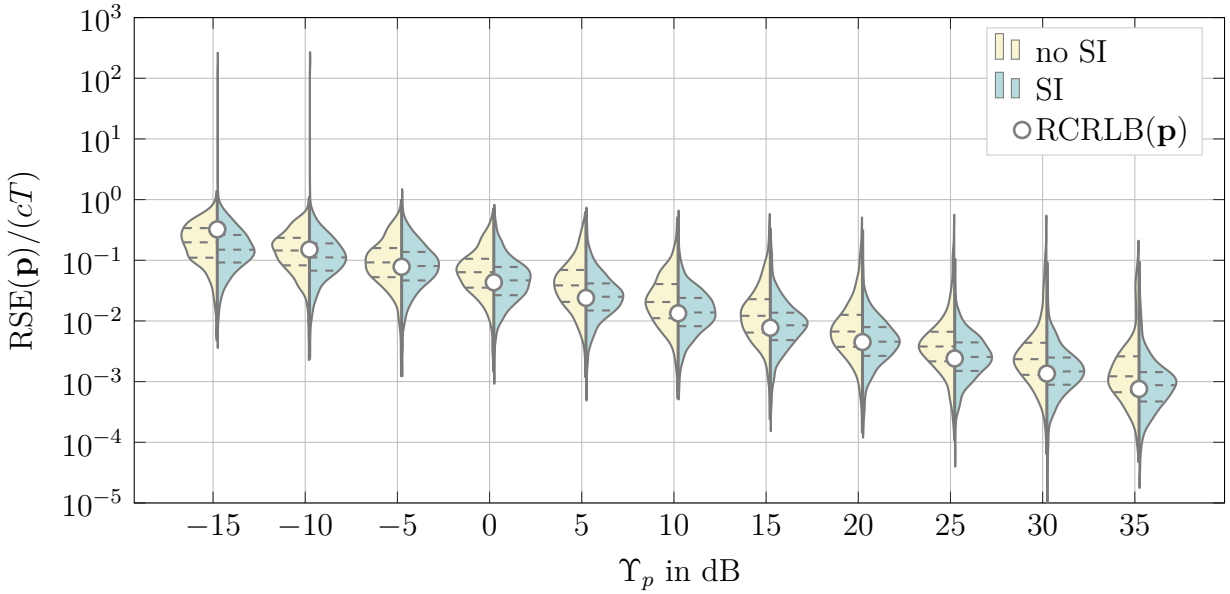


(b) CE MSEs for JCAP

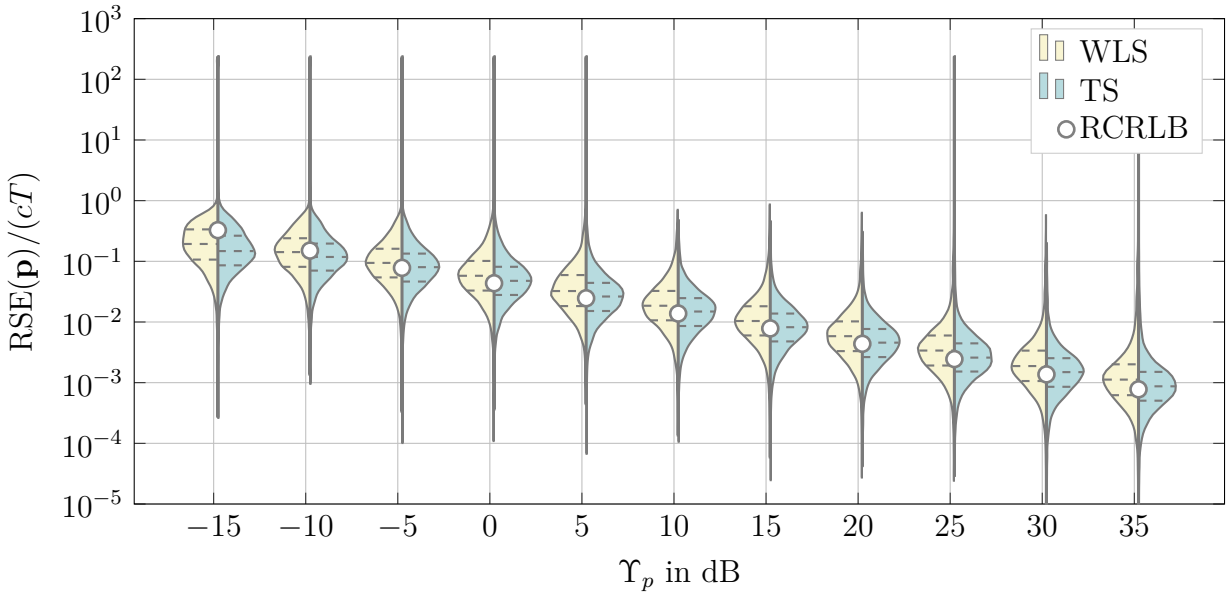


(c) TOA SEs for $U = 1$ and $U = 4$.

Figure 7.11: The BER, the channel estimation MSE and the SE distributions for the TOA estimation error all asymptotically approach the CRLBs. Due to a bounded search-space-based bias the TOA SE lies even below the CRLB.



(a) Positioning RSEs with versus without soft information (SI).

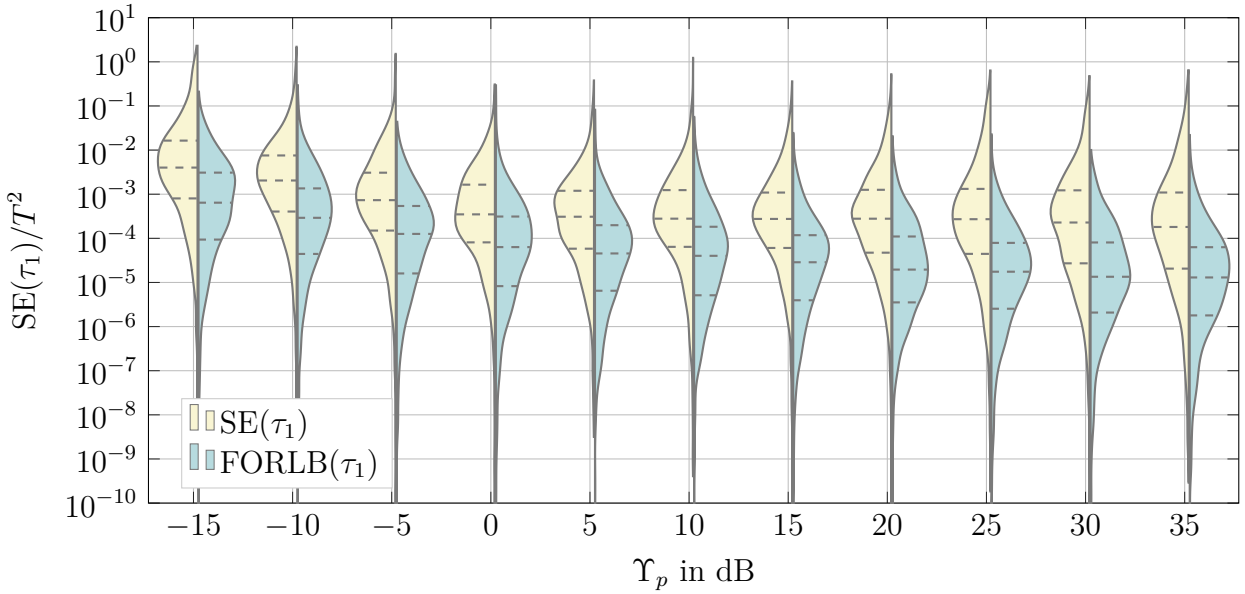


(b) Positioning RSEs with SI and determined via WLS versus TS.

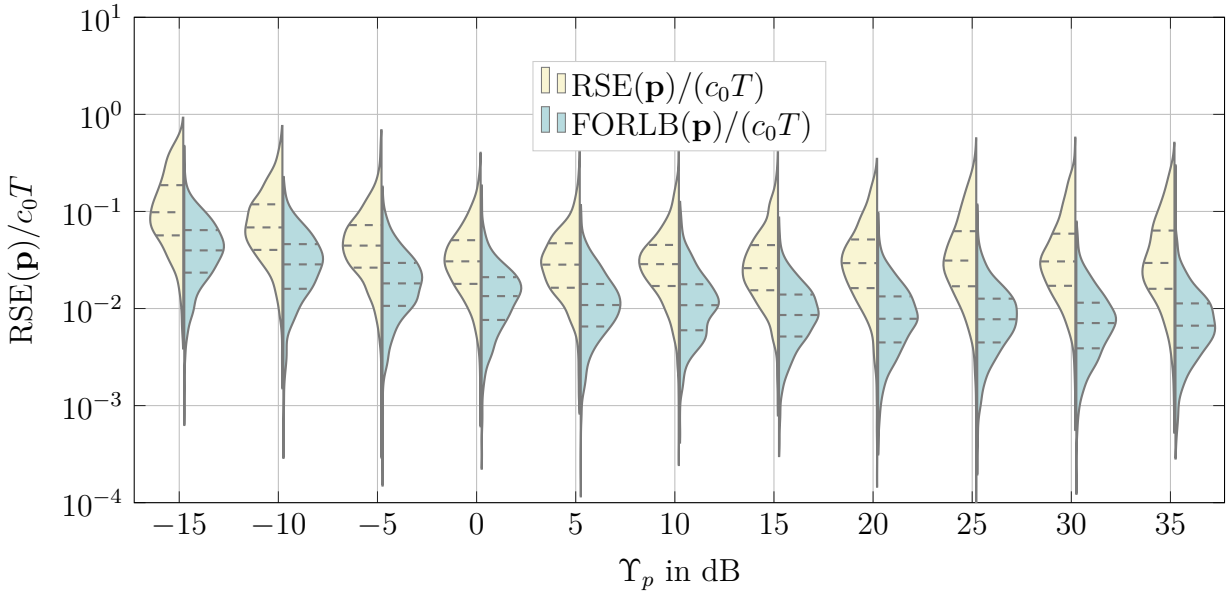
Figure 7.12: For the example of four reference objects, the violin plots show a slight improvement for the case with soft information. The chance of outliers is increased when soft information is applied. As expected, the TS positioning algorithm outperforms the WLS method.

7.5.5 WINNER Channel Model

The WINNER B1-LOS channel model scenario is chosen to demonstrate that the proposed strategy can be applied successfully to harsh physical environments, even if only a limited amount of snapshot channel estimates are available. The comparison to the FORLB demonstrates twofold: Firstly, it shows the optimally achievable performance and that the chosen methods are at least close to this bound. Secondly, it shows that for realistic scenarios and biased estimators the well known CRLB is not a well-suited performance measure. In [Sch12] [page 74] it was even shown numerically that calculating them for realistic environments leads to completely meaningless results due to ill-conditioned matrix inversion imposed by matrix singularities. The FORLB, on the other hand, is a powerful tool to determine a lower bound that provides a smooth transition from one model order's CRLB to the next higher model order's CRLB for increasing SNR. With increasing SNR or PNR, the more paths can be successfully estimated whereas, at the same time, the estimation error itself increases with increasing model order. The difference between the TOA SE and the FORLB, as well as the difference between the position RSE and the FORLB can be explained by the FORLB's complete focus on the TOA. Strategies to incorporate mechanisms that focus even more on the TOA promise to enhance the proposed method. Nevertheless, the closeness to the FORLB that is visible in Figure 7.13 is gained by employing the estimated approximated Fisher information, that yields that the model order resulting in the lowest multipath delay estimation error is chosen, which is not necessarily the correct model order or the maximum allowed model order (in case this is chosen lower that the correct model order). As expected, the BER and the channel estimates' MSE performs well (Figure 7.14).

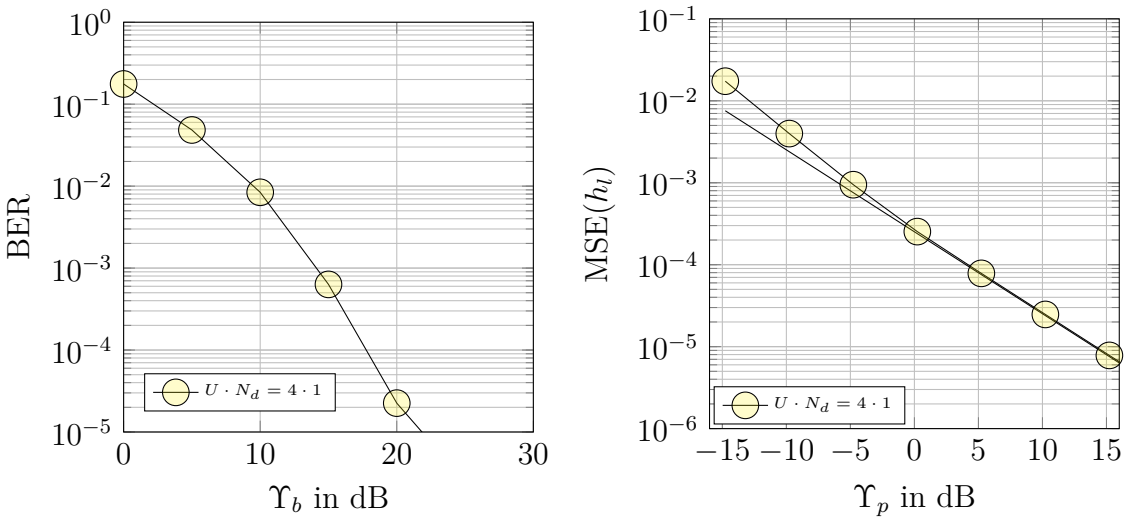


(a) Joint model order selection and parameter estimation TOA SE



(b) Positioning RSE versus the FORLB for joint model order selection and parameter estimation for the WLS algorithm without soft information.

Figure 7.13: The performance of all depicted errors shows that the proposed joint model order and parameter estimation via exploiting the approximated inverse Fisher information matrix successfully overcomes the challenge to find the tradeoff between under- and overfitting for harsh realistic environments and a comparably small amount of available measurements.



(a) BERs for communication

(b) CE MSEs for JCAP

Figure 7.14: BER and channel estimation MSE.

Chapter 8

Summary and Outlook

This chapter briefly summarizes the most relevant contributions of this thesis to a state-of-the-art JCAP system, and it gives a short outlook on where to go from here.

Summary

Especially the mobile wireless communication systems evolution during the past decade has led to the expectation of highly reliable, low latency, high precision position estimation everywhere at every time. The approach to combine different parallelly implemented positioning techniques based on different signal structures and channels targets to fulfill this expectation. Particularly investigating the overall system performance of ground-based JCAP alternatives to GNSS, based on a unified signal structure become increasingly important, due to the massive use of handheld mobile communication devices, which inherently provide the possibility of doubly exploiting the signal structure for communication and positioning. The expected growth of the internet of things, based on future wireless communication and positioning systems, underlines the necessity to investigate novel JCAP approaches.

JCAP system proposal: This thesis focuses on a terrestrial synergetic unified signal structure-based approach: It proposes and assesses a generic channel estimation-based JCAP system. Via a doubly exploitable iterative semi-blind channel estimation, the positioning side and the communication side can mutually exchange information. For positioning, I propose high-precision channel estimates, using the complete communication signal energy as virtual training. These accurate snapshot channel measurements are employed to improve the data detection accuracy iteratively and to carry out the high-resolution TOA estimation. The proposed approach applies to different multiplexing, modulation, and multiple access strategies. Since accurate channel estimation is crucial for accurate positioning, the chosen channel estimation and hence also the channel modeling itself can be interpreted as the bottleneck. At the same time, the systems turns out to be robust against changes in communication algorithms. Processing can be carried out in the time domain and the frequency domain.

Parameter Estimation and Algorithms: This thesis demonstrates that the TOA estimation is part of a multi-dimensional global non-linear optimization problem. Further, it shows how to state this problem as a separable non-linear least squares approach, either formulated in the time domain or the frequency domain. Moreover, I assessed a rotational invariance-based closed-form solution, performed in the frequency domain. The alternative approach, proposed here, operates in the time domain and relies on an iterative global optimization method, namely particle swarm optimization (PSO). The presented theory, in combination with the numerical results in this thesis, show that PSO yields a simplistic and, due to its parallelizability, practical approach that additionally results in a TOA MSE approaching the TOA CRLB with increasing SNR. PSO has outperformed alternative algorithms assessed here due to various reasons: The LM and SAGE algorithms require the availability of an initial guess close to the global optimum. PSO performs successfully without this initial guess. The ESPRIT algorithm requires over-sampling in the time domain. Unfortunately, oversampling is known not to yield any performance gain for data detection algorithms. Hence it entails an unnecessarily high processing overhead.

Joint Model Selection and Parameter Estimation: The multipath model order is a key parameter determining the underlying model complexity and hence the theoretically achievable best performance. Unfortunately, this model order is, in general, unknown and has to be estimated. It is known for decades that information-theoretic criteria can be successfully applied to find the model order for superimposed signals. The common assumption thereby is that the correct model order is sought. In this thesis, however, the ultimate goal is to find the lowest TOA MSE for realistic channel models, as opposed to finding the correct model order, which turned out not to coincide: The number of multipath components in realistic models (i.e., the correct model order) is usually close to the number of observations (channel taps). The optimal TOA MSE in such scenarios, on the other hand, turns out to be obtained by a less complicated parametrization, that is, a smaller model order. Hence, this thesis' goal was to find the model order minimizing the TOA MSE. To this end, I proposed an approach combining well suited information-theoretic criteria or threshold-based model selection methods for the single measurement case with the ML optimization problem to estimate the parameters and the model order jointly. To tailor my approach to focus on an optimized TOA accuracy instead of finding the correct model order, I introduced a soft information-based approach, called the information complexity criterion. This approach uses the approximated Fisher information matrix to find a model order that yields a low TOA MSE. To demonstrate its efficiency, I introduced a lower performance bound for the joint model selection and parameter estimation called the focused order related lower bound (FORLB). The associated numerical results show that, even for the realistic WINNER multipath channel scenario and snapshot channel estimates without oversampling, a TOA MSE performance close to the FORLB can be achieved.

Theoretic Performance Limits and Soft Information: This thesis introduces and, if necessary, derives all required lower MSE performance bounds, the so-called Cramer-Rao lower bounds (CRLBs). Particularly, the SNLLS CRLB derivation can be exploited as

well for the approximated Hessian, required in the LM algorithm and soft information determination. The thesis further assesses how the approximated Fisher information matrix can be employed to determine multipath parameter soft information for positioning as well as for model selection. The numerical results for the chosen prototype system demonstrate that employing this soft information outperforms positioning without soft information. The newly introduced FORLB for the joint model order selection and parameter estimation problem offers a suitable lower bound, where the CRLB fails for ill-conditioned problems, typical in realistic physical channel scenarios.

Prototype System: As a prototype system, I proposed an ML-IDMA based transmitter and receiver structure. At the receiver side, I employed an iterative multi-layer detector, semi-blind channel estimation, and multipath parameter estimation formulated as the SNLLS problem and solved via PSO. The results show the applicability of the chosen components. The generic theoretic framework described beforehand suggests that the communication components can be flexibly substituted by other components without severe positioning performance loss. The system assessment, however, shows that an accurate channel estimation is required for positioning. The results further show that a realistic channel model, like the B1-LOS WINNER setup investigated in this thesis, in combination with the small amount of snapshot channel estimates, requires a joint model order selection and parameter estimation tailored to find the lowest achievable TOA MSE for the assumed channel model. The results show that utilizing the approximated Fisher information in the information-theoretic criteria used to determine the optimal model order yields the best TOA performance among the investigated alternative approaches. Moreover, the numerical results show that even for the realistic B1-LOS-WINNER channel scenario high-resolution delay estimates are achieved by enforcing higher parsimony for the parameter estimation when this optimizes the TOA MSE.

Outlook

At the stage of submitting this thesis, I foresee several interesting points worthwhile to be further investigated:

Further improving the TOA MSE: At this point, it seems straightforward that determining a method of estimating the delay estimation bias would potentially lead to a further optimized version of a focused information criterion for joint model order selection and parameter estimation. I foresee that this would yield a TOA MSE even closer to the FORLB. Another seemingly promising approach to yield an enhanced TOA MSE performance in realistic scenarios would be to formulate the parameter estimation and joint model selection as a sparse estimation problem and to relate the optimal model order to the degree of sparsity.

System Considerations: All numerical investigations in this thesis and the theoretic performance bounds rely on the chosen channel model. Investigating alternative channel models will most certainly have a high impact on the investigated system concept, includ-

ing the applied algorithms, and therefore is very interesting. Finding other receiver-sided modeling attempts to optimize the TOA estimation performance is interesting as well. Another point worth investigating is how much performance can be gained by additionally employing the angle of arrival and departure estimates of all multipath components. Numerical results based on the latest mobile communication standard would be interesting as well.

Positioning: Since the positioning problem is inherently time-varying, it is customary to employ tracking algorithms to improve positioning efficiency. While this is classically implemented via extended Kalman-filtering, a promising competitive solution in the given scenario would be to employ a hierarchical particle swarm-based optimization method to track the estimated channel parameters.

Appendix A

Derivations and Calculations

A.1 Seperable Non-linear Least Squares Cost Function Derivation for a Single Snapshot Measurement

$$\hat{\boldsymbol{\tau}} = \arg\min_{\tilde{\boldsymbol{\tau}}}\{(\hat{\mathbf{h}} - \mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger}\hat{\mathbf{h}})^{H}(\hat{\mathbf{h}} - \mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger}\hat{\mathbf{h}})\} \tag{A.1}$$

$$= \arg\min_{\tilde{\boldsymbol{\tau}}}\{((\mathbf{I} - \mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger})\mathbf{h})^{H}(\mathbf{I} - \mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger})\hat{\mathbf{h}}\} \tag{A.2}$$

$$= \arg\min_{\tilde{\boldsymbol{\tau}}}\{\hat{\mathbf{h}}^{H}(\mathbf{I} - \mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger})^{H}(\mathbf{I} - \mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger})\hat{\mathbf{h}}\} \tag{A.3}$$

$$= \arg\min_{\tilde{\boldsymbol{\tau}}}\{\hat{\mathbf{h}}^{H}\underbrace{(\mathbf{I} - \mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}})^{\dagger}))}_{\mathbf{A}}{}^{2}\hat{\mathbf{h}}\} \tag{A.4}$$

$$\overbrace{=}^{\substack{\mathbf{A}\text{ idempotent}\\ \mathbf{A}^2=\mathbf{A}}} \arg\min_{\tilde{\boldsymbol{\tau}}}\{\hat{\mathbf{h}}^{H}(\mathbf{I} - \mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger})\hat{\mathbf{h}}\} \tag{A.5}$$

$$= \arg\max_{\tilde{\boldsymbol{\tau}}}\{\hat{\mathbf{h}}^{H}\mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger}\hat{\mathbf{h}}\}. \tag{A.6}$$

$$= \arg\min_{\tilde{\boldsymbol{\tau}}}\{-\hat{\mathbf{h}}^{H}\mathbf{G}(\tilde{\boldsymbol{\tau}})(\mathbf{G}(\tilde{\boldsymbol{\tau}}))^{\dagger}\hat{\mathbf{h}}\}. \tag{A.7}$$

A.2 Seperable Non-linear Least Squares Cost Function Derivation for Multiple Measurements

The deterministic maximum likelihood cost function corresponds to the least squares error. Not neglecting the constant terms as it is usually done to obtain a simple cost function version the deterministic maximum likelihood can be written as

$$\mathcal{L}_{dml}(\hat{\boldsymbol{\tau}}) = -(L+1)N_r I\ln(\sigma_w^2) - \frac{1}{\sigma_w^2}\sum_{v=1}^{N_r}\sum_{i=1}^{I}\left(\hat{\mathbf{h}}_{v,i} - \mathbf{G}\mathbf{G}^{\dagger}\hat{\mathbf{h}}_{v,i}\right)^{H}\left(\hat{\mathbf{h}}_{v,i} - \mathbf{G}\mathbf{G}^{\dagger}\hat{\mathbf{h}}_{v,i}\right). \tag{A.8}$$

and as it is pointed out in Chapter 5, assuming that σ_w^2 is a further unknown, then

$$\hat{\sigma}_w^2 = \frac{1}{(L+1)N_r I} \sum_{v=1}^{N_r} \sum_{i=1}^{I} \left(\hat{\mathbf{h}}_{v,i} - \mathbf{G}\mathbf{G}^{\dagger} \hat{\mathbf{h}}_{v,i} \right)^H \left(\hat{\mathbf{h}}_{v,i} - \mathbf{G}\mathbf{G}^{\dagger} \hat{\mathbf{h}}_{v,i} \right) \tag{A.9}$$

$$= \frac{1}{(L+1)} \text{trace}(\mathbf{P_G}^{\perp} \hat{\mathbf{C}}_{\hat{\mathbf{h}}}) \tag{A.10}$$

Inserting (A.10) into (A.8) and multiplying with -1 in order to turn the maximization into a minimization results in

$$\mathcal{L}_{dml}(\hat{\boldsymbol{\tau}}) = (L+1)N_r I \ln(\frac{1}{L+1} \text{trace}(\mathbf{P_G}^{\perp} \hat{\mathbf{C}}_{\hat{\mathbf{h}}})) \tag{A.11}$$

For the minimization taking the logarithm is not necessary. Therefore the cost function can be reduced to (5.123).

A.3 Derivation of Nabla Operator for the Levenberg-Marquard Method

The equality

$$\nabla \Omega[(\tilde{\boldsymbol{\xi}})]_m = \frac{\partial \Omega(\tilde{\boldsymbol{\xi}})}{\partial \tilde{\xi}_m} = -2\text{Re} \left\{ \frac{\partial \mathbf{h}^H(\tilde{\boldsymbol{\xi}})}{\partial \tilde{\xi}_m} (\mathbf{h} - \tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}})) \right\} \tag{A.12}$$

is derived below. Let $\boldsymbol{h}^i$, $\boldsymbol{h}^q$ denote the real and the complex part of the channel coefficient vector. Note that the least squares error $\Omega(\tilde{\boldsymbol{\xi}})$ can be formulated as

$$\Omega(\tilde{\boldsymbol{\xi}}) = (\hat{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}}))^H (\hat{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}})) = |\hat{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\boldsymbol{\xi}})|^2 \tag{A.13}$$

$$= \left(\hat{\mathbf{h}}^i - \tilde{\mathbf{h}}^{\mathbf{i}}(\tilde{\boldsymbol{\xi}}) \right)^2 + \left(\hat{\mathbf{h}}^q - \tilde{\mathbf{h}}^{\mathbf{q}}(\tilde{\boldsymbol{\xi}}) \right)^2 \tag{A.14}$$

$$= \sum_{l=0}^{L} \left(\left(\hat{h}_l^i - \tilde{h}_l^i(\tilde{\boldsymbol{\xi}}) \right)^2 + \left(\hat{h}_l^q - \tilde{h}_l^q(\tilde{\boldsymbol{\xi}}) \right)^2 \right) \tag{A.15}$$

The derivative is

$$\frac{\partial \Omega}{\partial \tilde{\xi}_m} = \sum_{l=0}^{L} \frac{\partial}{\partial \tilde{\xi}_m} \left(\left(\hat{h}_l^i - \tilde{h}_l^i(\tilde{\boldsymbol{\xi}}) \right)^2 + \left(\hat{h}_l^q - \tilde{h}_l^q(\tilde{\boldsymbol{\xi}}) \right)^2 \right) \tag{A.16}$$

Employing either chain or product rule

$$\frac{\partial \Omega}{\partial \tilde{\xi}_m} = \sum_{l=0}^{L} \left(-2 \frac{\partial \tilde{h}_l^i}{\partial \tilde{\xi}_m} (\hat{h}_l^i - \tilde{h}_l^i) - 2 \frac{\partial \tilde{h}_l^q}{\partial \xi_m} (\hat{h}_l^q - \tilde{h}_l^q) \right) - 2 \left(\tilde{\mathbf{h}} \right) \tag{A.17}$$

$$= -2 \sum_{l=0}^{L} \left(\frac{\partial \tilde{h}_l^i}{\partial \xi_m} \hat{h}_l^i - \frac{\partial \tilde{h}_l^i}{\partial \xi_m} \tilde{h}_l + \frac{\partial \tilde{h}_l^q}{\partial \xi_m} \hat{h}_l^q - \frac{\partial \tilde{h}_l^q}{\partial \xi_m} \tilde{h}_l^q \right) \tag{A.18}$$

$$= -2 \left(\frac{\partial \tilde{\mathbf{h}}^i}{\partial \xi_m}^H \hat{\mathbf{h}}_i - \frac{\partial \tilde{\mathbf{h}}^i}{\partial \xi_m}^H \tilde{\mathbf{h}} + \frac{\partial \tilde{\mathbf{h}}^q}{\partial \xi_m}^H \hat{\mathbf{h}}^q - \frac{\partial \tilde{\mathbf{h}}^q}{\partial \xi_m}^H \tilde{\mathbf{h}}^q \right) \tag{A.19}$$

$$\begin{aligned} = &-2\mathrm{Re} \left(\frac{\partial \tilde{\mathbf{h}}^i}{\partial \xi_m}^H \hat{\mathbf{h}}^i + j \frac{\partial \tilde{\mathbf{h}}^i}{\partial \xi_m}^H \hat{\mathbf{h}}^q - \frac{\partial \tilde{\mathbf{h}}^i}{\partial \xi_m}^H \tilde{\mathbf{h}}^i + j \frac{\partial \tilde{\mathbf{h}}^i}{\partial \xi_m}^H \tilde{\mathbf{h}}^q \right. \\ &\left. -j \frac{\partial \tilde{\mathbf{h}}^q}{\partial \xi_m}^H \hat{\mathbf{h}}^i + \frac{\partial \tilde{\mathbf{h}}^q}{\partial \xi_m}^H \hat{\mathbf{h}}^q + j \frac{\partial \tilde{\mathbf{h}}^q}{\partial \xi_m}^H \tilde{\mathbf{h}}^i - \frac{\partial \tilde{\mathbf{h}}^q}{\partial \xi_m}^H \tilde{\mathbf{h}}^q \right) \end{aligned} \tag{A.20}$$

$$= -2\mathrm{Re} \left(\frac{\partial \tilde{\mathbf{h}}^i}{\partial \xi_m}^T \left(\hat{\mathbf{h}}^i + j\hat{\mathbf{h}}^q - \tilde{\mathbf{h}}^i - j\tilde{\mathbf{h}}^q \right) - j \frac{\partial \tilde{\mathbf{h}}^q}{\partial \xi_m}^T (\hat{\mathbf{h}}^i + j\hat{\mathbf{h}} - \tilde{\mathbf{h}}^i - j\tilde{\mathbf{h}}^q) \right) \tag{A.21}$$

$$= -2\mathrm{Re} \left(\frac{\partial}{\xi_m} \left(\tilde{\mathbf{h}}^i - j\tilde{\mathbf{h}}^q \right)^T (\hat{\mathbf{h}}^i + j\hat{\mathbf{h}}^q - (\tilde{\mathbf{h}}^i - j\tilde{\mathbf{h}}^q)) \right) \tag{A.22}$$

$$-2\mathrm{Re} \left(\frac{\tilde{\mathbf{h}}}{\partial \xi_m} \left(\hat{\mathbf{h}} - \tilde{\mathbf{h}} \right) \right) \tag{A.23}$$

A.4 A Tall's Matrix Pseudo-Inverse Derivative Derivation

$$\partial(\mathbf{X}^\dagger) = \partial((\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H) \tag{A.24}$$

$$\overset{(5.60)}{=} \partial((\mathbf{X}^H\mathbf{X})^{-1})\mathbf{X}^H + (\mathbf{X}^H\mathbf{X})^{-1}\partial(\mathbf{X}^H) \tag{A.25}$$

$$\overset{(5.61)}{=} -(\mathbf{X}^H\mathbf{X})^{-1}\partial\left(\left(\mathbf{X}^H\mathbf{X}\right)^{-1}\right)(\mathbf{X}^H\mathbf{X})^{-1} + (\mathbf{X}^H\mathbf{X})^{-1}\partial\mathbf{X}^T \tag{A.26}$$

$$\overset{(5.60)}{=} -(\mathbf{X}^H\mathbf{X})^{-1}(\partial\mathbf{X}^T\mathbf{X} + \mathbf{X}^T\partial\mathbf{X})\mathbf{X}^\dagger + (\mathbf{X}^H\mathbf{X})^{-1}\partial\mathbf{X}^T \tag{A.27}$$

$$= (\mathbf{X}^H\mathbf{X})^{-1}\left(\partial\mathbf{X} - ((\partial\mathbf{X}^T)\mathbf{X} + \mathbf{X}^T\partial\mathbf{X})\mathbf{X}^\dagger\right) \tag{A.28}$$

$$= (\mathbf{X}^H\mathbf{X})^{-1}\left(\partial\mathbf{X}(\mathbf{I} - \mathbf{X}\mathbf{X}^\dagger) - \mathbf{X}^T(\partial\mathbf{X})\mathbf{X}^\dagger\right) \tag{A.29}$$

$$= -\mathbf{X}^\dagger\partial(\mathbf{X})\mathbf{X}^\dagger + \mathbf{X}^\dagger\mathbf{X}^{\dagger T}(\partial\mathbf{X}^T)(\mathbf{I} - \mathbf{X}\mathbf{X}^\dagger) \tag{A.30}$$

A.5 Likelihood Derivation for MDL Criterion Under Eigendecomposition-based Parametrization

The likelihood based on the eigendecomposition is formulated and simplified via

$$L(\boldsymbol{\Theta}^{(\tilde{C})}) = - N_r I \ln|\mathbf{C}_{\hat{\mathbf{h}}}^{(\tilde{C})}(\boldsymbol{\Theta}^{\tilde{C}})| + const \tag{A.31}$$

$$= - N_r I \ln \left(\left(\prod_{i=1}^{\tilde{C}} l_i \right) \left(\sigma_{\hat{h}}^{2(L+1-\tilde{C})} \right) \right) + const \tag{A.32}$$

$$= - N_r I \ln \left(\frac{\left(\prod_{i=1}^{L+1} l_i \right) \left(\sigma_{\hat{h}}^{2(L+1-\tilde{C})} \right)}{\prod_{i=\tilde{C}+1}^{L+1} l_i} \right) + const \tag{A.33}$$

$$= - N \ln \underbrace{\left(\prod_{i=1}^{L+1} l_i \right)}_{=|\hat{\mathbf{C}}_{\hat{\mathbf{h}}}|=const} - N \ln \left(\frac{\sigma_{\hat{h}}^{2(L+1-\tilde{C})}}{\prod_{i=\tilde{C}+1}^{L+1} l_i} \right) + const \tag{A.34}$$

$$= \ln \left(\frac{\prod_{i=\tilde{C}+1}^{L+1} l_i^{1/(L+1-\tilde{C})}}{\sigma_{\hat{h}}^2} \right)^{(L+1-\tilde{C})N} + const \tag{A.35}$$

$$= \ln \left(\frac{\prod_{i=\tilde{C}+1}^{L+1} l_i^{1/(L+1-\tilde{C})}}{\frac{1}{L+1-\tilde{C}} \sum_{i=\tilde{C}+1}^{L+1} l_i} \right)^{(L+1-\tilde{C})N} + const. \tag{A.36}$$

A.6 Closed-Form Solution Derivation Summary to Information Theoretic Criterion

Due to underlying assumptions the observations covariance matrix $\mathbf{C}_{\hat{\mathbf{h}}}$ can be expressed as

$$\mathbf{C}_{\hat{\mathbf{h}}} = \boldsymbol{\Psi} + \sigma^2 \mathbf{I}, \quad \text{where} \quad \boldsymbol{\Psi} = \mathbf{G} \mathbf{R}_{\boldsymbol{\beta}} \mathbf{G}^H \tag{A.37}$$

In case the matrix $\mathbf{G}$ has full column rank and $\mathbf{R}_{\boldsymbol{\beta}}$ is nonsingular, the rank of $\boldsymbol{\Psi}$ corresponds to $\tilde{C}$ and the $L+1-\tilde{C}$ smallest eigenvalues of $\boldsymbol{\Psi}$ are equal to 0 and consequently the $L+1-\tilde{C}$ eigenvalues of $\mathbf{C}_{\hat{\mathbf{h}}}$ are equal to σ^2. The eigenvalues of $\mathbf{C}_{\hat{\mathbf{h}}}$ are denoted by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{L+1}$. Namely, the smallest $L+1-q$ eigenvalues are

$$\lambda_{q+1} \geq \lambda_2 \geq \ldots \geq \lambda_{L+1} = \sigma^2. \tag{A.38}$$

Therefore, in theory the number of source signals can be determined via the number of the smallest eigenvalues of $\mathbf{C}_{\hat{\mathbf{h}}}$[Bar54, Law56]. Unfortunately, $\mathbf{C}_{\hat{\mathbf{h}}}$ is unknown and has to be estimated. In this case the calculated eigenvalues will all be different and therefore the exact number of smallest eigenvalues cannot be detected properly anymore, without introducing some subjective thresholding. Nevertheless [WK85] demonstrated how to exploit the covariance matrix and its eigendecomposition for model order estimation in a more objective manner based on information theoretic criteria. In order to understand this we have to acknowledge that in line with linear algbera's spectral theorem the channel estimate covariance matrix can also be modelled or decomposed such that

$$\mathbf{C}_{\hat{\mathbf{h}}}{}^{(\tilde{C})} = \sum_{i=1}^{\tilde{C}} (\lambda_1 - \sigma^2)\mathbf{V_i}\mathbf{V_i}^H + \sigma^2\mathbf{I}. \tag{A.39}$$

Recall that $\mathbf{H} = [\mathbf{h}(1), \ldots, \mathbf{h}(I)]$ is the set of I observations and a model given by the probability density function $f(\mathbf{H}|\mathbf{\Theta})$, where $\mathbf{\Theta}$ is a parameter vector for the new parametrization, comprising the parameters

$$\mathbf{\Theta}^{(\tilde{C})} = [\lambda_1, \ldots, \lambda_{\tilde{C}}, \sigma^2, \mathbf{V}_1^T, \ldots, \mathbf{V}_{\tilde{C}}^T]. \tag{A.40}$$

The eigenvalues $\lambda_1, \cdots, \lambda_{\tilde{C}}$ and the eigenvectors $\mathbf{V}_1, \cdots, \mathbf{V}_{\tilde{\mathbf{C}}}$ belong to $\mathbf{C}_{\hat{\mathbf{h}}}{}^{(\tilde{C})}$. Hence they can be estimated as the eigevalues and vectors obtained from an eigen decomposition of the estimated, that is the sample covariance matrix $\hat{\mathbf{C}}_{\hat{\mathbf{h}}}{}^{(\tilde{C})}$, calculated via (5.133). Then the idea is to select the model, which best fits the data that is, which minimizes the information theoretic criteria, more specifically the aim is to find the otimal model order $\tilde{C}_{opt}$. Since we have assumed that the observations are statistically independant zero mean complex Gaussian vectors the joint probabilty density is well known and is constructed according to (5.131,5.132) and lead to the same negative loglikelihood expressions as in in (5.130) and (5.134)

$$\mathrm{L}(\mathbf{\Theta}^{(\tilde{C})}) = -N_r I \ln|\mathbf{C}_{\hat{\mathbf{h}}}{}^{(\tilde{C})}| - \mathrm{trace}[\mathbf{C}_{\hat{\mathbf{h}}}{}^{(\tilde{C})}]^{-1}\mathbf{C}_{\hat{\mathbf{h}}}. \tag{A.41}$$

Note that in case $N_r I$ is large the second term in (A.41) is constant such that [CY03]

$$\mathrm{trace}[\mathbf{C}_{\hat{\mathbf{h}}}{}^{(\tilde{C})}]^{-1}\mathbf{C}_{\hat{\mathbf{h}}} = I N_r (L+1).$$

Obviously this term being independant of $\tilde{C}$ consequently does not contribute to the maximization of the likelihood and therefore subsequently may be neglected. Minimizing (A.41) leads to the maximum likelihood estimates [And63]

$$\hat{\lambda}_i = l_i, \quad i = 1, \cdots, \tilde{C} \tag{A.42}$$

$$\sigma^2 = \frac{1}{L+1-\tilde{C}} \sum_{i=\tilde{C}+1}^{L+1} l_i \tag{A.43}$$

$$\hat{\mathbf{V}}_i = \mathbf{C}_i, \quad i = 1, \cdots, \tilde{C} \tag{A.44}$$

with the eigenvalues $l_1 > l_2 > \cdots > l_{L+1}$ and eigenvectors $\mathbf{C}_1, \cdots, \mathbf{C}_{L+1}$ associated with $\mathbf{C}_{\hat{\mathbf{h}}}$. Via the simplifications in the subappendice A.5 derivation and via inserting (A.44) into (A.41) and further inserting this result together with (6.16) into (6.8) then leads to the results

$$\mathcal{C}_{\mathrm{MDL}}(\tilde{C}) = -\ln\left(\frac{\prod_{i=\tilde{C}+1}^{L+1} l_i^{1/(L+1-\tilde{C})}}{\frac{1}{L+1-\tilde{C}}\sum_{i=\tilde{C}+1}^{L+1} l_i}\right)^{(L+1-\tilde{C})N_r I} + \frac{1}{2}\tilde{C}(2(L+1)-\tilde{C})\ln I N_r. \tag{A.45}$$

A.7 Calculations Related to Theoretical Limits

A.7.1 Linear Least-Squares Covariance Matrix Estimation

The covariance matrix can be calculated as follows:

$$\mathbf{C}_{\hat{\boldsymbol{\xi}}} = \mathrm{E}\{(\hat{\boldsymbol{\xi}} - \boldsymbol{\mu})(\hat{\boldsymbol{\xi}} - \boldsymbol{\mu})^H\} \tag{A.46}$$

Since $\hat{\boldsymbol{\xi}} = \boldsymbol{\xi} + \mathbf{w}$ with $\mathbf{w} \in \mathcal{N}_C \sim (0, \sigma_w^2)$ the channel estimation error can be modelled as addtitive white gaussian noise and therefore the estimator is assumed to be unbiased. Hence, it is easy to see $\mathrm{E}\{\hat{\boldsymbol{\xi}}\} = \boldsymbol{\xi}$ and

$$\mathbf{C}_{\hat{\boldsymbol{\xi}}} = \mathrm{E}\{(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})^H\}. \tag{A.47}$$

Employing

$$\begin{aligned}
\hat{\boldsymbol{\xi}} - \boldsymbol{\xi} =& \mathbf{X}^\dagger \mathbf{y} - \boldsymbol{\xi} \\
=& \mathbf{X}^\dagger(\mathbf{X}\boldsymbol{\xi} + \mathbf{n}) - \boldsymbol{\xi} \\
=& \underbrace{(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{X}}_{\mathbf{I}}\boldsymbol{\xi} + \mathbf{X}^\dagger\mathbf{n} - \boldsymbol{\xi} \\
=& \boldsymbol{\xi} + \mathbf{X}^\dagger\mathbf{n} - \boldsymbol{\xi} \\
=& \mathbf{X}^\dagger\mathbf{n}
\end{aligned} \tag{A.48}$$

and substituting the result of (A.48) in (A.47) yields

$$\begin{aligned}
\mathbf{C}_{\hat{\boldsymbol{\xi}}} =& \mathrm{E}\{(\mathbf{X}^\dagger\mathbf{n})(\mathbf{X}^\dagger\mathbf{n})^H\} = \mathrm{E}\{\mathbf{X}^\dagger\mathbf{n}\mathbf{n}^H\mathbf{X}^{\dagger H} = \mathbf{X}^\dagger \underbrace{\mathrm{E}\{\mathbf{n}\mathbf{n}^H\}}_{\sigma_n^2\mathbf{I}_{M\times N}}\mathbf{X}^{\dagger H} = \sigma_n^2\mathbf{X}^\dagger\mathbf{X}^{\dagger H} \\
=& \sigma_n^2(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H((\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H)^H = \underbrace{(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{X}}_{\mathbf{I}}(\mathbf{X}^H\mathbf{X})^{-1H} = \sigma_n^2(\mathbf{X}^H\mathbf{X})^{-1}
\end{aligned} \tag{A.49}$$

A.8 Relevant Derivatives for the TS SML CRLB

Remember that the underlying model for $\mathbf{C}_{\hat{\mathbf{h}}}$ is determined by

$$\mathbf{C}_{\hat{\mathbf{h}}} = \mathbf{G}(\boldsymbol{\theta})\mathbf{C}_{\boldsymbol{\beta}}\mathbf{G}^T + \sigma_{\mathbf{w}}\mathbf{I}_{L+1}. \tag{A.50}$$

The derivative of any matrix $\mathbf{A}$ by any scalar x is defined as $\left[\frac{\partial \mathbf{A}}{\partial x}\right]_{i,j} = \frac{\partial}{\partial x}[A]_{i,j}$. Consequently, the relevant partial derivatives are calculated as follows

$$\left[\frac{\partial \mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})}{\partial \tau_c}\right]_{l,l'} = \frac{\partial}{\partial \tau_c}\left[\mathbf{G}(\boldsymbol{\theta})\mathbf{C}_{\boldsymbol{\theta}}\mathbf{G}^H(\boldsymbol{\theta})\right]_{l,l'}. \tag{A.51}$$

The element of a matrix multiplication of any two matrices $\mathbf{A} \in \mathbb{C}^{m\times n}$ and $\mathbf{B} \in \mathbb{C}^{n\times p}$ is known to be $[\mathbf{AB}]_{i,j} = \sum_{k=1}^{n}[\mathbf{A}]_{i,k}[\mathbf{B}]_{k,j}$. Thus, it is easy to see that applying this rule repeatedly, we get

$$\begin{aligned}\left[\mathbf{G}(\boldsymbol{\theta})\mathbf{C}_{\boldsymbol{\beta}}\mathbf{G}^T(\boldsymbol{\theta})\right]_{l,l'} &= \sum_{k=1}^{C}[\mathbf{G}(\boldsymbol{\theta})]_{l,k}\sum_{k'=1}^{C}[\mathbf{C}_{\boldsymbol{\beta}}]_{k,k'}[\mathbf{G}^T]_{k',l'} &\text{(A.52)}\\ &= \sum_{k=1}^{C} g(lT-\tau_k)\sum_{k'=1}^{C}[\mathbf{C}_{\boldsymbol{\beta}}]_{k,k'}g(l'T-\tau_{k'}) &\text{(A.53)}\\ &= \sum_{k=1}^{C}\sum_{k'=1}^{C}[\mathbf{C}_{\boldsymbol{\beta}}]_{k,k'}g(lT-\tau_k)g(l'T-\tau_{k'}). &\text{(A.54)}\end{aligned}$$

Further inserting (A.54) into (A.51) yields:

$$\left[\frac{\partial}{\partial \tau_c}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} = \frac{\partial}{\partial \tau_c}\left(\sum_{k=1}^{C}\sum_{k'=1}^{C}[\mathbf{C}_{\boldsymbol{\beta}}]_{k,k'}g(lT-\tau_k)g(l'T-\tau_{k'})\right) \tag{A.55}$$

Applying the chain rule for derivatives, this equation becomes

$$\left[\frac{\partial}{\partial \tau_c}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} = \begin{cases} 2\left[\mathbf{C}_{\boldsymbol{\beta}}\right]_{c,c} g(lT-\tau_c)\dfrac{\partial}{\partial \tau_c}(g(lT-\tau_c)) \\ +\displaystyle\sum_{\substack{k'=1\\k'\neq c}}^{C}\left[\mathbf{C}_{\boldsymbol{\beta}}\right]_{c,k'} g(lT-\tau_{k'})\frac{\partial}{\partial \tau_c}g(lT-\tau_c) \\ +\displaystyle\sum_{\substack{k=1\\k\neq c}}^{C}\left[\mathbf{C}_{\boldsymbol{\beta}}\right]_{k,c} g(lT-\tau_k)\frac{\partial}{\partial \tau_c}g(lT-\tau_c) & \text{if } l=l' \\ \\ \displaystyle\sum_{k'=1}^{C}\left[\mathbf{C}_{\boldsymbol{\beta}}\right]_{c,k'} g(l'T-\tau_{k'})\frac{\partial}{\partial \tau_c}g(lT-\tau_c) \\ +\displaystyle\sum_{k=1}^{C}\left[\mathbf{C}_{\boldsymbol{\beta}}\right]_{k,c} g(lT-\tau_k)\frac{\partial}{\partial \tau_c}g(l'T-\tau_c) & \text{else}\end{cases}.$$

With $[\mathbf{C}_{\boldsymbol{\beta}}]_{c,k'} = [\mathbf{C}_{\boldsymbol{\beta}}]_{k,c}$ and $\mathbf{c}_{\boldsymbol{\beta}_c}$ and $\mathbf{J}_{\mathbf{g}_c}$ denoting the $c$ths columns of $\mathbf{C}_{\boldsymbol{\beta}}$ and $\mathbf{J}_{\mathbf{g}}$, respectively, this equation can be simplified to

$$\begin{aligned}\left[\frac{\partial}{\partial\tau_c}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} =& \sum_{k'=1}^{C}[\mathbf{C}_{\boldsymbol{\beta}}]_{c,k'}\, g(l'T-\tau_{k'})\frac{\partial}{\partial\tau_c}g(lT-\tau_c)\\ &+\sum_{k=1}^{C}[\mathbf{C}_{\boldsymbol{\beta}}]_{k,c}\, g(lT-\tau_k)\frac{\partial}{\partial\tau_c}g(l'T-\tau_c)\\ =&[\mathbf{J}_{\mathbf{g}_c}(\mathbf{G}\mathbf{c}_{\boldsymbol{\beta}_c})^T+\mathbf{G}\mathbf{c}_{\boldsymbol{\beta}_c}\mathbf{J}_{\mathbf{G}c}^T]_{l,l'}.\end{aligned} \tag{A.56}$$

Recall that according to the parametrization in (5.189) the matrix elements of $\mathbf{C}_{\boldsymbol{\beta}}$ are related to $\boldsymbol{\eta}$ in the following manner:

$$[\mathbf{C}_{\boldsymbol{\beta}}]_{m,n} = \left[\mathbf{C}_{\boldsymbol{\beta}}^T\right]_{m,n}\begin{cases}\boldsymbol{\eta}_m & \text{if } m=n\\ \boldsymbol{\eta}_{C+n-m-1+\sum_{k=1}^{m}(C-k)}+j\boldsymbol{\eta}_{C+(C^2-C)/2+n-m-1+\sum_{k=1}^{m}(C-k)} & \text{if } n>m\\ \boldsymbol{\eta}_{C+m-n-1+\sum_{k=1}^{n}(C-k)}+j\boldsymbol{\eta}_{C+(C^2-C)/2+m-n-1+\sum_{k=1}^{n}(C-k)} & \text{else}\end{cases}. \tag{A.57}$$

This leads to the following representations of the involved partial derivatives:

$$\begin{aligned}\left[\frac{\partial}{\partial[\mathbf{C}_{\boldsymbol{\beta}}]_{m,m}}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} =& g(lT-\tau_m)g(l'T-\tau_m)\\ =&[\mathbf{g}_m^T\mathbf{g}_m]_{l,l'},\end{aligned} \tag{A.58}$$

$$\begin{aligned}\left[\frac{\partial}{\partial\,\mathrm{Re}[\mathbf{C}_{\boldsymbol{\beta}}]_{m,n}}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} =& g(lT-\tau_m)g(l'T-\tau_n)+g(lT-\tau_n)g(l'T-\tau_m)\\ =&[\mathbf{g}_m^T\mathbf{g}_n+\mathbf{g}_n^T\mathbf{g}_m]_{l,l'},\end{aligned} \tag{A.59}$$

$$\begin{aligned}\left[\frac{\partial}{\partial\,\mathrm{Im}[\mathbf{C}_{\boldsymbol{\beta}}]_{m,n}}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} =& j(g(lT-\tau_m)g(l'T-\tau_n)+g(lT-\tau_n)g(l'T-\tau_m))\\ =&j[\mathbf{g}_m^T\mathbf{g}_n+\mathbf{g}_n^T\mathbf{g}_m]_{l,l'}.\end{aligned} \tag{A.60}$$

With (A.58) to (A.60) it becomes clear that

$$\left[\frac{\partial}{\partial\theta_i}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} = \begin{cases} \left[\dfrac{\partial}{\partial\tau_i}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} & \text{if } i \leq C \\ \left[\dfrac{\partial}{\partial[\mathbf{C}_{\boldsymbol{\beta}}]_{c,c}}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} & \text{if } C < i \leq 2C, \\ & i = c + C \\ \left[\dfrac{\partial}{\partial\,\mathrm{Re}[\mathbf{C}_{\boldsymbol{\beta}}]_{m,n}}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} & \text{if } C < i \leq \dfrac{C^2-C}{2} + 2C, \\ & i = 2C + n - m - 1 + \sum\limits_{k=1}^{m}(C-k) \\ \left[\dfrac{\partial}{\partial\,\mathrm{Im}[\mathbf{C}_{\boldsymbol{\beta}}]_{m,n}}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} & \text{if } \dfrac{C^2-C}{2} + C < i \leq 2C^2, \\ & i = 2C + \dfrac{C^2-C}{2} + n - m - 1 + \sum\limits_{k=1}^{m}(C-k) \end{cases} \tag{A.61}$$

With (A.56), (A.61), $m, n \in \{1, \ldots, C\}$ and $n \geq m$ let for all $i \in \{1, \ldots, C^2 + C\}$

$$\left[\frac{\partial}{\partial\theta_i}\mathbf{C}_{\hat{\mathbf{h}}}(\boldsymbol{\theta})\right]_{l,l'} = \begin{cases} [\mathbf{J}_{\mathbf{g}_i}(\mathbf{G}\mathbf{c}_{\boldsymbol{\beta}_i})^T + \mathbf{G}\mathbf{c}_{\boldsymbol{\beta}_i}\mathbf{J}_{\mathbf{g}_i}{}^T]_{l,l'} & \text{if } i \leq C \\ [\mathbf{g}_m^T\mathbf{g}_m]_{l,l'} & \text{if } C < i \leq 2C, \\ & i = m + C \\ [\mathbf{g}_m^T\mathbf{g}_n + \mathbf{g}_n^T\mathbf{g}_m]_{l,l'} & \text{if } 2C < i \leq \dfrac{C^2-C}{2} + 2C, \\ & i = 2C + n - m - 1 + \sum\limits_{k=1}^{m}(C-k) \\ j([\mathbf{g}_m^T\mathbf{g}_n + \mathbf{g}_n^T\mathbf{g}_m]_{l,l'}) & \text{if } \dfrac{C^2-C}{2} + 2C < i \leq C^2 + C, \\ & i = 2C + \dfrac{C^2-C}{2} + n - m - 1 + \sum\limits_{k=1}^{m}(C-k) \end{cases} \tag{A.62}$$

Appendix B

Notation

B.1 Mathematical Definitions

B.1.1 DFT Matrix

The Vandermonde structured length N DFT matrix can be computed via

$$\mathbf{F}_n = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{N-1} \\ 1 & \omega^2 & \omega^{2\cdot 2} & & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \omega^{3(N-1)} \dots & \omega^{(N-1)(N-1)} \end{pmatrix}, \quad \text{with } \omega = e^{-\frac{2\pi j}{N}}. \tag{B.1}$$

B.1.2 The Jacobian Matrix

The Jacobian matrix $\mathbf{J_f}$ is defined as the matrix of all partial derivatives of an arbitrary length M vector-valued function $\mathbf{f} : \mathbb{R}^M \to \mathbb{R}^N$ that is parametrized by $\boldsymbol{\xi} \in \mathbb{R}^N$ such that

$$[\mathbf{J_f}(\boldsymbol{\xi})]_{m,n} = \frac{\partial f_m}{\partial \xi_n} \tag{B.2}$$

B.1.3 The Single-entry Matrix

The single-entry matrix is defined by:

$$[\mathbf{J}^{m,n}]_{i,j} = \begin{cases} 1 & \text{if } i = m \text{ and } j = n \\ 0 & \text{else} \end{cases} \tag{B.3}$$

B.1.4 Toeplitz Matrix

Let a Toeplitz $\mathbf{A}$ matrix be specified by two subvectors, namely $\mathbf{c}$ and $\mathbf{r}$, of the vector $\mathbf{a} = [a_{-K+1}, \dots, a_{-1}, a_0, a_1, \dots, a_{N-1}]$. Thereby, $\mathbf{c} = [c_0, \dots, c_{N-1}]$ determines the first

column and $\mathbf{r} = [r_0, \ldots, r_{K-1}]$ the first row of $\mathbf{A}$ such that

$$[a_0, \ldots, a_{-K+1}] = [\underbrace{r_0}_{=c_0}, \ldots, r_{K-1}] \quad \text{and} \quad [a_0, \ldots, a_{N-1}] = \mathbf{c} = [\underbrace{c_0}_{=r_0}, \ldots, c_{N-1}]$$

$$\mathbf{A} = \begin{pmatrix} a_0 & a_{-1} & \ldots & \ldots & a_{-K+1} \\ a_1 & a_0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & a_{-1} \\ a_{K-1} & \ldots & \ldots & a_1 & a_0 \\ \vdots & \ddots & & & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & & & \ddots & \vdots \\ a_{N-K} & \ldots & \ldots & \ldots & a_{N-2K+1} \end{pmatrix} = \begin{pmatrix} c_0 & r_1 & \ldots & \ldots & r_{K-1} \\ c_1 & c_0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & r_1 \\ c_{K-1} & \ldots & \ldots & c_1 & c_0 \\ \vdots & \ddots & & & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & & & \ddots & \vdots \\ c_{N-1} & \ldots & \ldots & \ldots & c_{N-K+1} \end{pmatrix} \quad \text{or}$$

$$[\mathbf{A}]_{n,k} = a_{n-k} = \begin{cases} c_{n-k} & \text{if } n \geq k \\ r_{k-n} & \text{else} \end{cases}, \quad \forall n \in \{0, \ldots, N-1\} \quad \text{and } \forall k \in \{0, \ldots, K-1\}. \tag{B.4}$$

B.1.5 Circulant Matrix

A square circular matrix is a special Toeplitz matrix, where the vector, which determines the complete matrix is build by concatenating a cyclic prefix such that

$$\mathbf{a} = [\underbrace{a_{-N+1}, \ldots, a_{-1}}_{\text{cyclic prefix}}, a_0, a_1, \ldots, a_{N-1}] = \mathbf{c} = [\underbrace{c_{N-1}, \ldots, c_1}_{\text{cyclic prefix}}, c_0, c_1, \ldots, c_{N-1}]. \tag{B.5}$$

Then the first column and the first row of the matrix are determined by $\mathbf{c} = [c_0, \ldots, c_{N-1}]$ and $\mathbf{r} = [c_{N-1}, \ldots, c_1]$, repectively. Consequently, a square $N \times N$ circulant matrix $\mathbf{C}$ is completely determined by its first column, corresponding to the vector $\mathbf{C}$, such that $\forall n, k \in \{0, \ldots, N-1\}$

$$\mathbf{C} = \begin{pmatrix} c_0 & c_{N-1} & \ldots & c_2 & c_1 \\ c_1 & c_0 & c_{N-1} & \ldots & c_2 \\ \vdots & c_1 & c_0 & \ddots & \vdots \\ c_{N-2} & \vdots & \ddots & \ddots & c_{N-1} \\ c_{N-1} & c_{N-2} & \ldots & c_1 & c_0 \end{pmatrix} \quad \text{or } [\mathbf{C}]_{n,k} = \begin{cases} c_{n-k} & \text{if } n \geq k \\ c_{k-n+N} & \text{else} \end{cases}. \tag{B.6}$$

B.1.6 Frobenius Norm

Let $\mathbf{A}$ be an arbitrary matrix with $[\mathbf{A}]_{m,n} = a_{m,n}$. Then the Frobenius Norm $||\mathbf{A}||_F$ is defined by

$$||\mathbf{A}||^2 = \sqrt{\sum_m \sum_n |a_{m,n}|^2}. \tag{B.7}$$

B.1.7 Kronecker Product

Let $\mathbf{A}$ be an $M \times N$ arbitrary matrix with $[\mathbf{A}]_{m,n} = a_{m,n}$ and $\mathbf{B}$ be an $U \times V$ arbitrary matrix with $[\mathbf{B}]_{u,v} = a_{u,v}$. Then the Kronecker product is a $MU \times NV$ matrix defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{1,1}\mathbf{B} & \dots & a_{1,N}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{M,1}\mathbf{B} & \dots & a_{M,N}\mathbf{B} \end{pmatrix} = (a_{m,n}\mathbf{B})_{m,n} \tag{B.8}$$

B.1.8 Kathri-Rao Product

Let $\mathbf{A}$ be an $M \times N$ arbitrary matrix with a partitioning of submatrices $\mathbf{A}_{i,j}$ with the sizes M_i, N_j. Further let $\mathbf{B}$ be an $U \times V$ arbitrary matrix with a partitioning of submatrices $\mathbf{B}_{i,j}$ with the sizes $U_i \times V_j$. The number of row partitions of $\mathbf{A}$ is the same as the number of row partitions of $\mathbf{B}$. Similarly, the number of column partitions of $\mathbf{A}$ and $\mathbf{B}$ is the same. Then the Kathri-Rao product is a $\sum M_i U_i \times \sum N_j V_j$ matrix defined as

$$\mathbf{A} \diamond \mathbf{B} = (\mathbf{A}_{i,j} \otimes \mathbf{B}_{i,j})_{i,j} \tag{B.9}$$

Note that the Kathri-Rao product is a special case of the Tracey-Singh product.

B.1.9 Hadamard Product

Let $\mathbf{A}$, $\mathbf{B}$ be $M \times N$ arbitrary matrices. Then the Hadamard product is defined as

$$[\mathbf{A} \odot \mathbf{B}]_{m,n} = a_{m,n} b_{m,n}. \tag{B.10}$$

B.1.10 Multiplication of Block Matrices

Let $\mathbf{A}$ be an arbitrary matrix with a partitioning of $M \times N$ submatrices $\mathbf{A}_{m,n}$. Let further be $\mathbf{B}$ be an arbitrary matrix with a partitioning of $N \times P$ submatrices $\mathbf{B}_{n,p}$. Then the multiplication of these block matrices is defined as a matrix with $M \times P$ submatrices such that

$$(\mathbf{AB})_{m,p} = \sum_{n=1}^{N} \mathbf{A}_{m,n}\mathbf{B}_{n,p} \tag{B.11}$$

B.1.11 Matrix Vectorization

The operator $\mathbf{vec}\{\cdot\}$ concatenates the columns of a matrix.

B.1.12 Square Root of a Matrix

The square root of an arbitrary matrix square matrix $\mathbf{A}$ is defined as

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{V_A}\text{diag}\{\mathbf{d}^{\frac{1}{2}}\}{\mathbf{V_A}}^{-1}. \tag{B.12}$$

Here $\mathbf{V_A}$ is a matrix composed of a columnwise concatenation of the eigen vectors of $\mathbf{A}$ and $\mathbf{d}$ denotes a vector containing all the eigen values of $\mathbf{A}$. Note that in case $\mathbf{A}$ is a Hermitian matrix ${\mathbf{V_A}}^{-1} = {\mathbf{V_A}}^{H}$ is valid.

B.1.13 On Multivariate Complex Gaussian Random Variables and their Probability Density Functions

A multivariate complex Gaussian can be thought of as a more dimensional normal distribution. Let and a multivariate Gaussian zero mean random vector $\mathbf{n} \sim \mathcal{N}(0, \boldsymbol{\Sigma_n})$ of size $K \times 1$, where $\boldsymbol{\Sigma}$ is the covariance matrix. Then the K-variate complex Gaussian random vector probability density function is defined as [Goo63]

$$p(n) = p(n_1, \ldots, n_K) = \frac{1}{\pi^K |\boldsymbol{\Sigma}_n|} \exp(-\mathbf{n}^H \boldsymbol{\Sigma}_n^{-1} \mathbf{n}). \tag{B.13}$$

Now, let $\mathbf{x}$ denote any deterministic arbitrary vector. Further, consider a vector $\mathbf{y}$ that is constructed in the following manner

$$\mathbf{y} = \mathbf{x} + \mathbf{n}. \tag{B.14}$$

Then $\mathbf{y}$ is a multivariate Gaussian distributed vector. This equation type typically occurs in communication systems. For Maximum Likelihood problems a typical PDF that has to be formulated is the conditional pdf of some observations stacked in $\mathbf{y}$ given a set of parameters of any parametrization stacked in the vector $\boldsymbol{\theta}$ that $\mathbf{y}$ depends on. This conditional pdf is the same as $p(\mathbf{n})$

$$p(\mathbf{y}|\boldsymbol{\theta}) = p(y_1, \ldots, y_K|\boldsymbol{\theta}) = p(n) \tag{B.15}$$

$$= \frac{1}{\pi^K |\boldsymbol{\Sigma}_n|} \exp(-(\mathbf{y} - \mathbf{x})^H \boldsymbol{\Sigma}_n^{-1} (\mathbf{y} - \mathbf{x})). \tag{B.16}$$

Another often needed term is the negativ loglikelihood function belonging to this pdf and parametrization, which is given by

$$-\ln(p(\mathbf{y}|\boldsymbol{\theta})) = \underbrace{K \ln\{\pi\}}_{\text{constant}} + \ln(|\boldsymbol{\Sigma}_n|) - (\mathbf{y} - \mathbf{x}) \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{x}) \tag{B.17}$$

For optimization purposes the constant is expendable and hence often is omitted.

B.2 Notational Conventions

Symbol	Short Description
$\forall$	for all
$\exists$	It exists
$\in$	Element of
$(\bullet)^*$	Conjugate transpose
$(\bullet)^H$	Hermitian transpose
$(\bullet)^\dagger$	Pseudo-inverse
$(\bullet)^\perp$	Orthogonal matrix
$\mathbf{P}_\bullet$	Projection matrix $(\bullet)(\bullet)^\dagger$
$\mathbf{P}^\perp_\bullet$	Orthogonal projection matrix $\mathbf{I} - (\bullet)(\bullet)^\dagger$
$(\bullet)^i$	in-phase part (real part)
$(\bullet)^q$	quadrature-phase part (complex part)
$*$	Convolution
$\circledast$	Circular convolution
$\odot$	Pointwise multiplication
$\otimes$	Kronecker product
$\diamond$	Kathri-Rao product
diag$\{\mathbf{X}\}$	selects main diagonal of matrix $\mathbf{X}$ as row vector
diag$\{\mathbf{x}\}$	builds diagonal matrix from vector $\mathbf{x}$
$\overset{N}{\vdash\!\!-}$	Length N discrete Fourier transformation
$\overset{N}{-\!\!\dashv}$	Length N Discrete inverse Fourier transformation
$\circ\!\!-$	Fourier transformation
$-\!\!\circ$	Inverse fourier transformation
$\mathbf{I}_N$	size $N \times N$ identity matrix
tr$(\bullet)$	Trace
x	Scalar variable
$\mathbf{x}$	Bold lower case letters denote vectors
$\mathbf{X}$	Bold upper case letters denote matrices
$\mathbf{F}_N$	size N discrete Fourier transform matrix
$\mathbb{X}$	This font denotes sets
vec $\{\bullet\}$	concatenates matrix column-wise
$\mathcal{U}[a, b)$	Uniform distribution within the one sided open interval $[a, b)$

B.3 Arabic Lower Case Letter Notation

Symbol	$\in$	**Short Description**
$\mathbf{a}_{Tc,n}$	$\mathbb{R}^{1\times N_T}$	array steering vector for cluster c and ray n
$\mathbf{a}_{Rc,n}$	$\mathbb{R}^{1\times N_R}$	array steering vector for cluster c and ray n
$\mathbf{b}_l$	$\mathbb{R}^\bullet$	lower bound vector for parameter estimation cost function

$\mathbf{b}_u$	$\mathbb{R}^{\bullet}$	upper bound vector for parameter estimation cost function
c_0	$\mathbb{R}^{+}$	speed of light
$c(\tau, t)$	$\mathbb{C}$	physical channel weight function
c_1	$\mathbb{R}^{+}$	Self adjustment weight for PSO
c_2	$\mathbb{R}^{+}$	Social adjustment weight for PSO
d_{LOS}	$\mathbb{R}^{+}$	distance between transmitter and receiver
d_{T_x}	$\mathbb{R}^{+}$	transmit antenna spacing for ULA
d_{R_x}	$\mathbb{R}^{+}$	receive antenna spacing for ULA
$\hat{\mathbf{d}}$	$\mathbb{R}^{1\times\bullet}$	pseudoranges
$f_n(t)$	$\mathbb{C}$	complex fading amplitude
f_{d_n}	$\mathbb{R}^{+}$	Doppler frequency
g_{Tx}	$\mathbb{R}^{+}$	pulse shaping filter
g_{Rx}	$\mathbb{R}^{+}$	receive filter
$\mathbf{h}$	$\mathbb{C}^{(L+1)\times 1}$	SISO EDTCM channel coefficients
$\mathbf{h}_{MO}$	$\mathbb{C}^{N_R(L+1)\times 1}$	Stacked SIMO measurements of the EDTCM channel
$\mathbf{h}_{TS}$	$\mathbb{C}^{I(L+1)\times 1}$	Stacked time series snapshot measurements of the EDTCM channel coefficients ($\mathbf{h}_{TS} = \text{vec}(\mathbf{H}_{TS})$)
$\hat{\tilde{\mathbf{h}}}_d$, $\hat{\tilde{\mathbf{h}}}^{\mathbf{d}}$	$\mathbb{C}^{(L+1\times 1}$	pulse deconvoluted channel estimate
k	$\mathbb{N}^{+}$	time index (if used for sytem description)
k	$\mathbb{N}^{+}$	number of free adjustable parameters (if used for information heoretic criteria)
m	$\mathbb{N}^{+}$	model order
$\mathbf{p}_n^i$	$\mathbb{R}^{\bullet}$	PSO specific particle position in iteration i and dimension n
$\mathbf{p}_{IB,n}^i$	$\mathbb{R}^{\bullet}$	PSO specific individual best particle position in iteration i and dimension n
$\mathbf{p}_{LB,n}^i$	$\mathbb{R}^{\bullet}$	PSO specific local best particle position in iteration i and dimension n
$\mathbf{p}_0 = [0, 0]$	$\mathbb{R}^{1\times 2}$	reference point in cartesian coordinates
$\mathbf{p}_{Tx} = [x_{Tx}, y_{Tx}]$	$\mathbb{R}^{1\times 2}$	cartesian transmitter coordinates
$\mathbf{p}_{Rx} = [x_{Rx}, y_{Rx}]$	$\mathbb{R}^{1\times 2}$	cartesian receiver coordinates
p	$\mathbb{R}^{+}$	topology dependent exponent in path loss model
$\mathbf{p}$	$\mathbb{R}^{1\times 2}$	cartesian coordinate
s	$\{0, 1\}$	binary value determining parameter estimation success
$\mathbf{r}$	residual vector	
$\mathbf{r_I}$		vector specifying the Inertia range for PSO
r_N	$\mathbb{N}^{+}$	user specified value to choose the neighborhood size for PSO particles

$s_{u,n,l}[k]$	$\mathbb{C}$	distortion value for layer n, user u and tap l for MLD
t	$\mathbb{R}^+$	time
$\mathbf{v}_n^i$	$\mathbb{R}^\bullet$	PSO specific particle velocity vector in iteration i for particle n
v	$\mathbb{R}^+$	Mobile velocity
$v_{\max}$	$\mathbb{R}^+$	Maximum mobile velocity
$\mathbf{w}$	$\mathbb{C}^{L+1}$	Channel estiation error
x_{MS}	$\mathbb{R}^+$	x Device position component in a carthesian coordinate system
x_i	$\mathbb{R}^+$	x Reference object position component in a carthesian coordinate system for the ith base station
x_{Rx}	$\mathbb{R}^+$	x position component in a carthesian coordinate system for a receiver
x_{Tx}	$\mathbb{R}^+$	x position component in a carthesian coordinate system for a transmitter
y_{MS}	$\mathbb{R}^+$	Second device position component in a carthesian coordinate system
y_i	$\mathbb{R}^+$	Second reference object position component in a carthesian coordinate system for the ith base station
$y[k]$	$\mathbb{C}$	received value at time intant k
$y^{i+1}[k]$	$\mathbb{C}$	received value without interference of pilot symbols in iteration i
$\breve{y}_{u,n,l}^{i+1}$	$\mathbb{C}$	layer, user and tap specific modelled received value for MLD

B.4 Arabic Upper Case Letter Notation

Symbol	$\in$	**Short Description**
A_{MS}	$\mathbb{R}$	Intersectiong coverage areas of N_{BS} BSs (Cell-ID and Enhanced Cell-ID)
A_{BS_i}	$\mathbb{R}$	ith BS coverage area (CellID- and Enhanced Cell-ID)
$\mathbf{A}_T$	$\mathbb{R}^{N_t \times N_{Rays}}$	transmitter array response matrix
$\mathbf{A}_R$	$\mathbb{R}^{N_{Rays} \times N_R}$	receiver array response matrix
$\mathbf{B}$		auxiliary matrix $[\mathbf{B}]_{k,l} = e^{-2[pi\Delta(k-l)]^2\sigma_\alpha^2\cos^2(\phi)}$
C	$\mathbb{N}^+$	number of cluster in clustered delay line model
$\tilde{C}_{max}$	$\mathbb{N}^+$	maximum hypothetical number of theoretically assessable multipath components
$\mathbf{C}_{\boldsymbol{\epsilon}}$	$\mathbb{R}^{N_b,N_b}$	covariance matrix for pseudoranges estimation error
$\mathbf{C}_w$	$\mathbb{R}^{(L+1)\times(L+1)}$	channel estimation error covariance matrix

$\mathbf{C}_{\boldsymbol{\xi}}$	$\mathbb{R}^{(L+1)\times(L+1)}$	Soft estimation matrix used in ICOMP
D	$\mathbb{N}^{+}$	Number of optimization problem dimensions
$\mathbf{D}_{\text{gen}}$	$\mathbb{C}^{\bullet\times\bullet}$	generalized data matrix
$\mathbf{G}(\boldsymbol{\tau})$	$\mathbb{R}^{(L+1)\times C}$	Pulse matrix: $[\mathbf{G}(\boldsymbol{\tau})]_{l,c} = g(lT - \tau_c)$
$G_{rx}(f)$	$\mathbb{C}^{+}$	pulse shaping filter transfer function
$G_{rx}(f)$	$\mathbb{C}^{+}$	receive filter transfer function
G_{Rx}	$\mathbb{R}^{+}$	receive antenna gain
G_{Tx}	$\mathbb{R}^{+}$	transmit antenna gain
$\mathbf{H}$	$\mathbb{C}^{N_r\times N_T(L+1)}$	MIMO EDTCM channel coefficient matrix
$\mathbf{H}_{TS}$	$\mathbb{C}^{I\times(L+1)}$	Stacked time series snapshot measurements of the EDTCM channel
I	$\mathbb{N}^{+}$	number of blocks
I_{Max}	$\mathbb{N}^{+}$	Maximum Number of PSO Iterations
$I_{StopMax}$	$\mathbb{N}^{+}$	Maximum Number of iterations PSO stopping criterium has to be fulfilled
K	$\mathbb{N}^{+}$	number of samples in one block
K_R	$\mathbb{R}^{+}$	Rice factor
L	$\mathbb{N}^{+}$	overall channel memory length
L_g	$\mathbb{N}^{+}$	length of windowed convoluted transmit and receive filter
N_B	$\mathbb{N}^{+}$	number of base stations
N_{MP}	$\mathbb{N}^{+}$	number of resolvable and non-resolvable multipathes
N_{Rays}	$\mathbb{N}^{+}$	number of multipath rays with similar
N_R	$\mathbb{N}^{+}$	number of receiva antennas
N_T	$\mathbb{N}^{+}$	number of transmit antennas
N_p	$\mathbb{N}^{+}$	number of particles for PSO
N_N	$\mathbb{N}^{+}$	Neighborhood size for PSO
$P(\bullet)$	$\mathbb{R}+, \leq 1$	probability
P_c	$\mathbb{R}^{+}$	mean power of cluster c
$\mathbf{P}_{\mathbf{G}}$	$\mathbb{R}^{(L+1)\times(L+1)}$	$\mathbf{G}(\boldsymbol{\tau})(\mathbf{G}(\boldsymbol{\tau})^{\dagger}$ (see $\mathbf{G}$) orthogonal projection matrix
$\mathbf{P}_{\mathbf{G}}^{\perp}$	$\mathbb{R}^{(L+1)\times(L+1)}$	$\mathbf{I} - \mathbf{P}_{\mathbf{G}}$ (see $\mathbf{P}_{\mathbf{G}}$)
P_{Rx}	$\mathbb{R}^{+}$	received power
P_{Tx}	$\mathbb{R}^{+}$	transmit power
R	$\mathbb{R}^{+}$	coderate
S	$\mathbb{N}^{+}$	spreading factor for repetiton code
$\mathbf{S}$	$\mathbb{C}^{L_J\times C}$	phase rotation matrix for delays in frequency domain
T	$\mathbb{R}^{+}$	sampling time
$\mathbf{T}_{\text{gen}}$	$\mathbb{C}^{\bullet\times\bullet}$	generalized training matrix
$\mathbf{u}_l$	$\mathbb{R}^{1\times\mathcal{D}}$	lower search space bound for PSO
$\mathbf{u}_u$	$\mathbb{R}^{1\times\mathcal{D}}$	upper search space bound for PSO
$\mathbf{U}$	$\mathbb{N}^{+}$	number of users

$\mathbf{U}$		eigenvector matrix used in ESPRIT
$\mathbf{U_s}$		used for ESPRIT (see $\mathbf{U}$)
$\mathbf{U_d}$		used for ESPRIT (see $\mathbf{U}$)
$\mathbf{v}_n^{i+1}$	$\mathbb{R}^{1\times N_p}$	velocity vector for PSO
V	$\mathbb{R}^{(K-L)\times(L+1)}$	virtual training matrix
W	$\mathbb{N}^+$	integer factor times T specifying half pulse window width
X_σ	$\mathbb{R}$	log-normal distributed variable accounting for shadowing.

B.5 Greek Lower Case Letter Notation

Symbol	$\in$	Short Description
α	$\mathbb{R}^+$	confidence parameter ($0 < \alpha < 1$), typically close to 1
$\boldsymbol{\beta}$	$\mathbb{C}^{C\times 1}$	complex cluster path weights.
$\boldsymbol{\gamma}_{R,c}$	$\mathbb{C}^{N_r\times 1}$	receive antenna array spatial signature vector for cluster c
$\delta_{i,j}$	$\{0,1\}$	Kronecker delta $\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & else \end{cases}$
ϵ_{χ^2}	$\mathbb{R}^+$	threshold for determining parameter estimation success/failure
$\epsilon_{\tilde{C}}$	$\mathbb{R}^+$	threshold for algorithm (3)
ϵ_b	$\mathbb{R}^{+,<<1}$	small value near zero for lower and upper bound values for PSO
ε_1^i, ε_2^i	$\mathcal{U}^D(0,1)$	
ε_{tol}	$\mathbb{R}^{+,<<1}$	PSO stopping criterium tolerance
ε_i	$\mathbb{R}^+$	random weighting factors in PSO
ζ_c	$\mathbb{R}$	cth rotation matrix eigenvalue for ESPRIT
η	$\mathbb{R}^+$	tuning parameter for Levenberg-Marquardt method
$\boldsymbol{\eta}$	$\mathbb{R}^+$	parameter vector for upper triangular covariance matrix elements in TS SML CRLB
η_n	$\mathbb{C}^N$	phase for equal power allocation phase rotation of layer n
$\boldsymbol{\theta}$	$\mathbb{R}^{1\times 2\cdot N_R\cdot I\cdot C+C}$	full ML parameter vector
θ_{vel}	$\mathbb{R}^+$	mobile device's direction of travel
$\vartheta_{Tx,Rx}$, $\vartheta_{Rx,Tx}$	$\mathbb{R}^+$	Tx and Rx angles between LOS direction and antenna broadside
ι_c^n	$\mathbb{R}^+$	auxiliary value for appropriately weighting for EM and SAGE
κ	$\mathbb{R}$	coarse delay estimate part
λ	$\mathbb{R}^+$	wavelength
λ_i	$\mathbb{R}^+$	ith eigenvalue of $\mathbf{C}_{\boldsymbol{\xi}}$ in ICOMP
μ_{dof}	$\mathbb{N}^+$	degrees of freedom

$\nu[k']$	$\mathbb{C}$	oversampled virtual data sequence at index k'
$\boldsymbol{\xi}$	$\mathbb{R}^{\bullet}$	arbitrary parameter vector
ϖ		Eigenvalues of $\hat{\mathbf{C}}_{\boldsymbol{\Psi}\boldsymbol{\Psi}}$
ρ	$\mathbb{R}^{+}$	ratio of training to signal power
σ_n^2	$\mathbb{R}^{+}$	noise variance on received values
σ_w^2	$\mathbb{R}^{+}$	error variance on channel estimates
ς	$\mathbb{N}^{+}$	value $< L+1$ for FB and ESPRIT
τ_{TOA}	$\mathbb{R}^{+}$	TOA
$\boldsymbol{\tau}$	$\mathbb{R}^{1\times C}$	multipath delay vector
$\boldsymbol{\tau}_{\mathrm{toa},b}$	$\mathbb{R}^{1\times C}$	TOA for base station b
τ_{max}	$\mathbb{R}^{+}$	maximum excess delay
φ_i	$\mathbb{R}^{+}$	Angle of arrival associated with the ith reference object
φ_c	$\mathbb{R}^{+}$	Angle of departure associated with cluster c
ϕ_c	$\mathbb{R}^{+}$	Angle of arrival associated with cluster c
$\varphi_{c,n}$	$\mathbb{R}^{+}$	Angle of departure associated with cluster c and ray n
$\phi_{c,n}$	$\mathbb{R}^{+}$	Angle of arrival associated with cluster c and ray n
χ	$\mathbb{R}^{+}$	Constriction factor for PSO
χ^2	$\mathbb{R}^{+}$	Symbol for χ^2 distribution
ψ		
ω	$\mathbb{R}^{+}$	Inertia weight for PSO

B.6 Greek Upper Case Letter Notation

Symbol	$\in$	**Short Description**
$\Gamma(\cdot)$		Gamma function
$\boldsymbol{\Gamma}(t,\boldsymbol{\phi},\boldsymbol{\varphi})$	$\mathbb{C}^{N_R\times CN_T}$	overall spatial signature matrix
$\boldsymbol{\Gamma}_c(t,\boldsymbol{\phi}_c,\boldsymbol{\varphi}_c)$	$\mathbb{C}^{N_R\times N_T}$	spatial signature matrix for cluster c
$\boldsymbol{\Gamma'}_u(t,\boldsymbol{\phi},\boldsymbol{\varphi})$	$\mathbb{C}^{N_R\times C}$	spatial signature matrix for transmit antenna u
Λ	$\mathbb{R}$	wavelength
$\boldsymbol{\Phi}$	$\mathbb{C}^{\bullet\times\bullet}$	Rotation matrix for ESPRIT
$\boldsymbol{\Psi}$	$\mathbb{C}^{\varsigma\times(L+1)}$	Hankelmatrix for FB in ESPRIT
υ_b	$\mathbb{R}$	bit-to-noise ratio
Υ_b	$\mathbb{R}$	symbol-to-noise ratio
Υ_d	$\mathbb{R}$	data-to-noise ratio
Υ_p	$\mathbb{R}$	pilot-to-noise ratio
Ω	$\mathbb{R}$	cost function for LS and ML

B.7 Calligraphic Upper Case Letter Notation

Symbol	$\in$	Short Description
$\boldsymbol{\mathcal{A}}$		matrix used in TS-positioning algorithm
$\mathcal{B}$		Bit-load
$\boldsymbol{\mathcal{B}}$	$\{0,1\}^{L_{\text{div}}\times(L+1)}$	auxiliary matrix for ESPRIT and FB
$\mathcal{C}_{(\bullet)}$	$\mathbb{R}$	information theoretic criteria
$\mathcal{C}_{ICOMP}$	$\mathbb{R}$	Information complexity criterion
$\hat{\boldsymbol{\mathcal{C}}}_{\hat{\mathbf{h}}}$	$\mathbb{R}^{(L+1)\times(L+1)}$	Channel estimates sample
$\mathcal{D}$		Search space dimension
$\mathcal{G}_{MI}$		Multiple input pulse matrix ($\mathbf{I}_{N_T}\otimes\mathbf{G}$)
$\boldsymbol{\mathcal{J}}$	$\mathbb{R}^{L_J\times(L+1)W}$	Selection matrix for ESPRIT
$\mathcal{G}_{MI}$		MIMO pulse matrix ($\mathbf{I}_{N_RN_T}\otimes\mathbf{G}$)
$\mathcal{L}$		Likelihood
$\boldsymbol{\mathcal{G}}$	$\mathbb{R}^{N_RI(L+1)\times N_RIC}$	SIMO TS extended pulse matrix ($=\mathbf{I}_{N_RI}\otimes\mathbf{G}$)
$\boldsymbol{\mathcal{G}}_{MI}$	$\mathbb{R}^{N_T(L+1)\times N_TC}$	Multiple input extended pulse matrix ($=\mathbf{I}_{N_T}\otimes\mathbf{G}$)
$\boldsymbol{\mathcal{G}}_{TS}$	$\mathbb{R}^{I(L+1)\times IC}$	TS extended pulse matrix ($=\mathbf{I}_I\otimes\mathbf{G}$)
$\boldsymbol{\mathcal{G}}_{MO}$	$\mathbb{R}^{N_R(L+1)\times N_RC}$	SIMO extended pulse matrix ($=\mathbf{I}_{N_R}\otimes\mathbf{G}$)
$\boldsymbol{\mathcal{H}}(\tau,\boldsymbol{\tau},\boldsymbol{\Gamma})$	$\mathbb{C}^{N_R\times N_T}$	Continuous MIMO channel matrix
$\mathcal{P}(\bullet)$	$\mathbb{R}$	penalty term (for instance for ITC)
$\boldsymbol{\mathcal{V}}$	$\mathbb{C}^{IN_R(K-L)\times IN_RN_TU(L+1)}$	MA,TS, MIMO virtual training matrix

Appendix C

List of Abbreviations

Numbers:		
	3GGP	Third generation partnership project
A:		
	AIC	Akaike's information criterion
	AGNSS	Assisted GNSS
	AWGN	Additive white Gaussian noise
B:		
	BS	Base station
	BPSK	Binary phase shift keying
C:		
	CDF	Cumulative density function
	CDM	Code-division multiplexing
	CDMA	Code-division multiple access
	CRLB	Cramer-Rao lower bound
D:		
	DML	Deterministic maximum likelihood
	DFT	Discrete Fourier transfomation
	DGNSS	Differential GNSS
	DOF	Degrees of freedom
E:		
	ESPRIT	Estimation of parameters via rotational invariance techniques
F:		
	FDM	Frequency-division multiplexing
	FDMA	Frequency-divison multiple access
	FFT	Fast Fourier transform
G:		
	GALILEO	Europian GNSS
	GDOP	Geometric delution of Precision
	GLONASS	Russian GNSS
	GNSS	Global navigation satellite systems
	GPS	Global positioning system

H:

I:

ICOMP	Information complexity criterion
IDM	Interleave-division multiplexing
IDMA	Interleave-division multiple access
IFFT	Inverse fast Fourier transform
IRNSS	Indian regional navigational satallite system

J:

K:

LFSR	Linear feedback shift registers
LLR	Log-likelihood Ratio
LLS	Linear least squares
LOS	Line of sight
LS	Least Squares
LTE	Long term evolution
LTE-A	LTE-Advanced

M:

MDL	Minimum description length
ML	Maximum-likelihood
MS	Mobile station

N:

NGMN	Next generation mobile networks
NLLS	Non-linear least squares
NLOS	Non line of sight

O:

OFDM	Orthogonal frequency-division multiplexing
OFDMA	Orthogonal frequency-division multiple access
OMA	Orthogonal multiple access

P:

PDF	Probability density function
PDP	Power delay profile
PRN	Pseudo random noise sequence

Q:

R:

RO	Reference object

S:

SML	Stochastic maximum likelihood
SNLLS	Separable Non-linear least squares

T:

TDM	Time-division multiplexing
TDMA	Time-division multiple access
TDOA	Time-difference of arrival
TOA	Time of arrival
TOF	Time-of-flight

U:

UE	User equipment
UMTS	Universal mobile telecommunications system

Bibliography

[ABC+14] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, and A. C. K. Soong, "What will 5G be ?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[AH13] R. Adam and P. A. Hoeher, "Semi-blind channel estimation for joint communication and positioning," in *Proc. 10th IEEE Workshop on Positioning, Navigation and Communication (WPNC)*, Dresden, Germany, Mar. 2013, pp. 1–5.

[AH20] R. Adam and P. A. Hoeher, "Simultaneous model and parameter estimation for joint communication and positioning," *IEEE Access*, vol. 9, pp. 2934–2949, Dec. 2020.

[Aka73] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *2nd International Symposium on Information Theory*, Tsahkadsor, Sep. 1973, pp. 267–281.

[Aka74] ——, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[And63] T. W. Anderson, "Asymptotic theory for principal component analysis," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 122–148, Mar. 1963.

[AO99] D. Astely and B. Ottersten, "The effects of local scattering on direction of arrival estimation with music," *IEEE Transactions on Signal Processing*, vol. 47, no. 12, pp. 3220–3234, Dec. 1999.

[AOS97] D. Asztely, B. Ottersten, and A. Swindlehurst, "A generalized array manifold model for local scattering in wireless communications," in *Proc. EEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, Munich, Germany, 1997, pp. 4021–4024.

[ARS16] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, Feb. 2016.

[awi] "awiloc® Indoor Lokalisierung, Tracking und Navigation." [Online]. Available: https://www.iis.fraunhofer.de/de/ff/lv/lok/tech/awiloc.html

[BA02] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York, NY: Springer, 2002.

[Bar54] M. S. Bartlett, "A note on the multiplying factors for various $\chi 2$ approximations," *Journal of the Royal Statistical Society Series B*, vol. 16, no. 2, pp. 296–298, Jul. 1954.

[Bel63] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Transactions on Communications*, vol. 11, no. 4, pp. 360–393, Dec. 1963.

[Ber01] F. V. D. Bergh, "An analysis of particle swarm optimizers," Ph.D. dissertation, University of Pretoria, 2001.

[BG06] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[BHSN10] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[BK07] D. Bratton and J. Kennedy, "Defining a standard for particle swarm optimization," in *Proc. IEEE Swarm Intelligence Symposium*, Honolulu, Hawai, USA, Apr. 2007, pp. 120–127.

[Boz00] H. Bozdogan, "Akaike's information criterion and recent developments in information complexity," *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 62–91, Mar. 2000.

[Cam17] R. S. Campos, "Evolution of positioning techniques in cellular networks, from 2G to 4G," *Wireless Communications and Mobile Computing*, vol. 2017, p. 17, 2017.

[Cap69] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[CH01] G. Claeskens and N. L. Hjort, "The focussed information criterion," in *Model Selection and Model Averaging*. Cambridge University Press, Jan. 2001, pp. 145–191.

[Chu72] D. Chu, "Polyphase codes with good periodic correlation properties (Corresp.)," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 531–532, Jul. 1972.

[Chu05] P.-J. Chung, "ML estimation under misspecification of signals," in *Proc. of the 39th Asilomar Conference on Signals, Systems*, vol. 1, Pacific Grove, CA, USA, 2005, pp. 1698–1701.

[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, 2002.

[CM08] P. J. Chung and C. Mecklenbraeuker, "Deterministic ML estimation for unknown numbers of signals," in *Proc. 16th European Signal Processing Conference (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008, pp. 1–5.

[CVM07] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1743–1763, May 2007.

[CY03] Q. Cheng and H. Yingbo, "A review of parametric high-resolution methods," in *High-Resolution and Robust Signal Processing*, H. Yingbo, A. Gershman, and Q. Cheng, Eds. CRC Press, 2003, ch. 1, pp. 1–62.

[DLK+17] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[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 (Methodological)*, vol. 39, no. 1, pp. 1–22, Sep. 1977.

[DLSSG+12] J. A. Del Peral-Rosado, J. A. López-Salcedo, G. Seco-Granados, F. Zanier, and M. Crisci, "Joint channel and time delay estimation for LTE positioning reference signals," in *Proc. 6th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing*, Noordwijk, Netherlands, Dec. 2012.

[dPRLSSG+14] J. A. del Peral-Rosado, J. A. López-Salcedo, G. Seco-Granados, F. Zanier, and M. Crisci, "Joint maximum likelihood time-delay estimation for LTE positioning in multipath channels," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 33, Dec. 2014.

[DRZ15] A. Dammann, R. Raulefs, and S. Zhang, "On prospects of positioning in 5G," in *Proc. IEEE International Conference on Communication Workshop (ICCW)*, London, UK, Jun. 2015, pp. 1207–1213.

[DWY+15] L. Dai, B. Wang, Y. Yuan, S. Han, C.-l. I, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[Dys04] F. Dyson, "A meeting with Enrico Fermi," *Nature*, vol. 427, no. 6972, pp. 297–297, Jan. 2004.

[Erg09] M. Ergen, *Mobile Broadband including WiMAX and LTE*. New York, NY: Springer Science+Business Media, 2009.

[FAG95] D. Falconer, F. Adachi, and B. Gudmundson, "Time division multiple access methods for wireless personal communications," *IEEE Communications Magazine*, vol. 33, no. 1, pp. 50–57, Jan. 1995.

[FH94] J. Fessler and A. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.

[Fis14] S. Fischer, "Observed time difference of arrival (OTDOA) positioning in 3GPP LTE," 2014. [Online]. Available: https://www.qualcomm.com/media/documents/files/otdoa-positioning-in-3gpp-lte.pdf

[FNI13] Z. Farid, R. Nordin, and M. Ismail, "Recent advances in wireless indoor localization techniques and system," *Journal of Computer Networks and Communications*, vol. 2013, p. 12, Aug. 2013.

[FTH+99] B. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. Ingeman 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.

[GC09] I. Güvenç and C. C. Chong, "A survey on TOA based wireless localization and NLOS mitigation techniques," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 3, pp. 107–124, 2009.

[Gez08] S. Gezici, "A survey on wireless position estimation," *Wireless Personal Communications*, vol. 44, no. 3, pp. 263–282, Feb. 2008.

[Gol67] R. Gold, "Optimal binary sequences for spread spectrum multiplexing (Corresp.)," *IEEE Transactions on Information Theory*, vol. 13, no. 4, pp. 619–621, Oct. 1967.

[Gol05] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, Aug. 2005.

[Goo63] N. R. Goodman, "Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction)," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 152–177, Mar. 1963.

[GP73] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM Journal on Numerical Analysis*, vol. 10, no. 2, pp. 413–432, Apr. 1973.

[GP03] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. R1–R26, Apr. 2003.

[GZT08] Guowei Shen, R. Zetik, and R. S. Thoma, "Performance comparison of TOA and TDOA based location estimation algorithms in LOS environment," in *Proc. 5th Workshop on Positioning, Navigation and Communication*. Hannover, Germany: IEEE, Mar. 2008, pp. 71–78.

[HJLY14] D. Han, S. Jung, M. Lee, and G. Yoon, "Building a practical Wi-Fi-based indoor navigation system," *IEEE Pervasive Computing*, vol. 13, pp. 72–79, May 2014.

[HN95] M. Haardt and J. Nossek, "Unitary ESPRIT: how to obtain increased estimation accuracy with a reduced computational burden," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1232–1242, May 1995.

[Höh13] P. A. Höher, *Grundlagen der digitalen Informationsübertragung*. Wiesbaden: Springer Fachmedien, 2013.

[HPALP09] V. Honkavirta, T. Perala, S. Ali-Lyotte, and R. Piche, "A comparative survey of WLAN location fingerprinting methods," in *Proc. 6th Workshop on Positioning, Navigation and Communication (WPNC)*, Hannover, Germany, Mar. 2009, pp. 243–251.

[HSF08] P. A. Hoeher, H. Schoeneich, and J. C. Fricke, "Multi-layer interleave-division multiple access: Theory and practice," *European Transactions on Telecommunications (ETT)*, vol. 19, no. 5, pp. 523–536, Jan. 2008.

[HT89] C. M. Hurvich and C.-L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, no. 2, pp. 297–307, Jun. 1989.

[HT08] F. Hlawatsch and G. Tauböck, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, Mar. 2008, pp. 2885–2888.

[Jaf88] A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources: a separable solution," in *Proc. International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, New York, NY, USA, Apr. 1988.

[JG85] J. A. Jacquez and P. Greif, "Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design," *Mathematical Biosciences*, vol. 77, no. 1-2, pp. 201–227, Dec. 1985.

[JM06] S. Janson and M. Middendorf, "A hierarchical particle swarm optimizer for noisy and dynamic environments," *Genetic Programming and Evolvable Machines*, vol. 7, no. 4, pp. 329–354, Sep. 2006.

[Kau75] L. Kaufman, "A variable projection method for solving separable nonlinear least squares problems," *BIT Numerical Mathematics*, vol. 15, no. 1, pp. 49–57, Mar. 1975.

[Kay10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, 1st ed. Upper Saddle River, New Jersey, USA: Prentice Hall, 2010.

[KCW+17] M. Koivisto, M. Costa, J. Werner, K. Heiska, J. Talvitie, K. Leppänen, V. Koivunen, M. Valkama, and S. Member, "Joint device positioning and clock synchronization in 5G ultra-dense networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2866–2881, Mar. 2017.

[KE95] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. International Conference on Neural Networks (ICNN)*, Perth, Australia, Nov. 1995, pp. 1942–1948.

[KGA17] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern WLAN fingerprinting indoor positioning methods and deployment challenges," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1974–2002, Mar. 2017.

[KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.

[KK04] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for WLAN location fingerprinting," in *Proc. the 1st Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MOBIQUITOUS)*, Boston, MA, (USA), Aug. 2004, pp. 14–23.

[KK15] J. J. Khalifeh and Z. M. Kassas, "Indoor localization based on floor plans and power maps: Non-line of sight to virtual line of sight," in *Proc. 28th Int. Tech. Meeting Satellite Div. Inst. Navig. (ION GNSS+)*, Tampa, Florida (USA), Sep. 2015, pp. 2291–2300.

[KMH+08] P. Kyösti, J. Meinila, L. Hentilae, X. Zhao, T. Jämsa, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. de Yong, and T. Rautianin, "IST-4-027756 WINNER II D1. 1.2 V1. 2 WINNER II channel models," 2008. [Online]. Available: https://www5.tu-ilmenau.de/nt/generic/paper_pdfs/PartIIofD1.1.2.pdf

[KV96] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[Lan99] R. B. Langley, "Dilution of precision," *GPS World*, vol. 10, pp. 52–59, May 1999.

[Law56] D. N. Lawley, "Tests of significance for the latent roots of covariance and correlation matrices," *Biometrika*, vol. 43, no. 1-2, pp. 128–136, Jun. 1956.

[Lee98] A. Lee Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Transactions on Signal Processing*, vol. 46, no. 2, pp. 449–462, Feb. 1998.

[Lin08] J. Lin, "Least-squares channel estimation for mobile OFDM communication on time-varying frequency-selective fading channels," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 6, pp. 3538–3550, Nov. 2008.

[LP04] X. Li and K. Pahlavan, "Super-resolution TOA estimation with diversity for indoor geolocation," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 224–234, Jan. 2004.

[LPRU19] J. Lohmeyer, F. Palm, H. Reuvers, and J.-P. Urbain, "Focused information criterion for locally misspecified vector autoregressive models," *Econometric Reviews*, vol. 38, no. 7, pp. 763–792, Aug. 2019.

[Lue96] H. Luetkepohl, *Handbook of matrices*. New York: John Wiley, 1996.

[MdCD02] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.

[Mia07] H. Miao, "Channel estimation and positioning for multiple antenna systems," Ph.D. dissertation, University of Oulu, 2007.

[Mit96] M. Mitchell, *An introduction to genetic algorithms*. Cambridge, Massachussetts - London, England: MIT Press, Sep. 1996.

[MKH10] J. Mayer, K. Khairy, and J. Howard, "Drawing an elephant with four complex parameters," *American Journal of Physics*, vol. 78, no. 6, pp. 648–649, May 2010.

[MKS04] X. Ma, H. Kobayashi, and S. C. Schwartz, "EM-based channel estimation algorithms for OFDM," *EURASIP Journal on Advances in Signal Processing*, no. 10, pp. 1460–1477, Sep. 2004.

[MM01] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3065–3073, Dec. 2001.

[MWBAS19] R. Mendrzik, H. Wymeersch, G. Bauch, and Z. Abu-Shaban, "Harnessing NLOS components for position and orientation estimation in 5G millimeter wave MIMO," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 93–107, Jan. 2019.

[Neu04] A. Neumaier, "Complete search in continuous global optimization and constraint satisfaction," *Acta Numerica*, vol. 13, pp. 271–369, Jun. 2004.

[Nex15] Next Generation Mobile Networks, "Next generation mobile networks 5G white paper," 2015. [Online]. Available: https://www.ngmn.org/wp-content/uploads/NGMN_5G_White_Paper_V1_0.pdf

[NJU15] D. Neumann, M. Joham, and W. Utschick, "Channel estimation in massive MIMO systems," 2015. [Online]. Available: https://arxiv.org/abs/1503.08691

[NLM98] J. C. Ng, K. B. Letaief, and R. D. Murch, "Complex optimal sequences with constant magnitude for fast channel estimation initialization," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 305–308, Mar. 1998.

[NLW+17] J. Nurmi, E. S. Lohan, H. Wymeersch, G. Seco-Granados, and O. Nykänen, Eds., *Multi-Technology Positioning*. Springer International Publishing, 2017.

[NMN64] J. Nelder, R. Mead, and A. Neumaier, "A simplex method for function minimization," *Computer Journal*, vol. 7, no. 1, pp. 308–313, Jan. 1964.

[NPD+13] J. J. Nielsen, T. Pedersen, B. Denis, A. Dammann, R. Raulefs, D. Slock, J. Rodriguez, N. Yi, Y. Ma, I. Aramba, M. Raspopoulos, and S. Stavrou, "Where2 system description," ICT-248894 WHERE2 Deliverable D1.9, Tech. Rep., 2013.

[OR13] D. P. O'Leary and B. W. Rust, "Variable projection for nonlinear least squares problems," *Computational Optimization and Applications*, vol. 54, no. 3, pp. 579–593, Apr. 2013.

[OVSN93] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in *Radar Array Processing*, S. Haykin, J. Litva, and T. J. Shephard, Eds. Berlin: Springer-Verlag, 1993, pp. 99–151.

[PH65] E. S. Pearson and H. O. Hartley, "Biometrika tables for statisticians," *Biometrische Zeitschrift*, vol. 7, no. 2, pp. 132–132, 1965.

[PL05] K. Pahlavan and A. H. Levesque, "Wireless information networks." Hoboken, New Jersey, USA: John Wiley & Sons, 2005, ch. 13: RF loc.

[PLL03] L. Ping, L. Liu, and W. K. Leung, "Interleave division multiple access (IDMA) communication systems," in *Proc. 3rd International Symposium on Turbo Codes and Related Topics*, Brest, France, Sep. 2003, pp. 173–180.

[PLWL06] L. Ping, L. Liu, K. Wu, and W. Leung, "Interleave division multiple-access," *IEEE Transactions on Wireless Communications*, vol. 5, no. 4, pp. 938–947, Apr. 2006.

[PP12] K. B. Petersen and M. S. Pedersen, "The Matrix Cookbook," Tech. Rep., 2012.

[PS06] R. Peng and M. L. Sichitiu, "Angle of arrival localization for wireless sensor networks," in *Proc. 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, Reston, VA, USA, Sep. 2006, pp. 374–382.

[PTVF07] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. New York, NY, USA: Cambridge University Press, 2007.

[PWW07] L. Ping, P. Wang, and X. Wang, "Recent progress in interleave-division multiple-access (IDMA)," in *Proc. IEEE Military Communications Conference (MILCOM)*, Orlando, USA, Oct. 2007, pp. 1–7.

[QY97] Qi Cheng and Yingbo Hua, "Detection of cisoids using least square error function," *IEEE Transactions on Signal Processing*, vol. 45, no. 6, pp. 1584–1590, Jun. 1997.

[R1-18] R1-1808331 - 3rd Generation Partnership Project, "Consideration on the transmitter side signal processing for NOMA, Sony, RAN1#94," Gothenburg, 2018. [Online]. Available: https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_94/Docs/R1-1808331.zip

[RH89] B. Rao and K. Hari, "Performance analysis of Root-Music," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.

[Rib08] C. B. Riberio, "Propagation parameter estimation in MIMO systems," PhD Thesis, Helsinki University of Technology, 2008.

[Ric05] A. Richter, "Estimation of radio channel parameters: Models and algorithms," Ph.D. dissertation, Illmenau, Germany, 2005.

[Ris78] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, Sep. 1978.

[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.

[RP08] R. Raulefs and S. Plass, "Combining wireless communications and navigation -the WHERE project," in *Proc. IEEE 68th Vehicular Technology Conference*, Ottawa, Canada, Sep. 2008, pp. 1–5.

[RS16] N. Rashmi and M. Sarvagya, "Sparse channel estimation using orthogonal matching pursuit algorithm for SCM-OFDM system," in *Proc. International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. Los Angeles: IEEE, Nov. 2016, pp. 1224–1227.

[RSK06] A. Richter, J. Salmi, and V. Koivunen, "An algorithm for estimation and tracking of distributed diffuse scattering in mobile radio channels," in *Proc. IEEE 7th Workshop on Signal Processing Advances in Wireless Communications*. IEEE, Jul. 2006, pp. 1–5.

[RSS+13] M. Raspopoulos, S. Stavrou, J. Stéphan, B. Uguen, J. Domínguez, L. D. Celis, B. Denis, D. Slock, and G. Aga, "Final: Location information extraction," ICT–248894 WHERE2 D2.5 Final:, Tech. Rep., 2013.

[RW80] A. Ruhe and P. Å. Wedin, "Algorithms for separable nonlinear least squares problems," *SIAM Review*, vol. 22, no. 3, pp. 318–337, Jul. 1980.

[Saa97] H. Saarnisaari, "TLS-ESPRIT in a time delay estimation," in *Proc. 47th Vehicular Technology Conference. Technology in Motion*, Pheonix, USA, May 1997, pp. 1619–1623.

[Saa98] ——, "Estimation of numbers of multipaths in DS-CDMA by the MDL principle," in *Proc. 1988 IEEE 5th International Symposium on Spread Spectrum Techniques and Applications - Proceedings. Spread Technology to Africa*, Sun City, South Africa, Sep. 1998, pp. 258–261.

[SAH11a] K. Schmeink, R. Adam, and P. A. Hoeher, "Influence of oversampling on channel parameter estimation for joint communication and positioning," in *Proc. 8th Workshop on Positioning, Navigation and Communication*, Dresden, Germany, Apr. 2011, pp. 161–165.

[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*, no. 1, p. 185, Dec. 2011.

[SAH12] ——, "Performance limits of channel parameter estimation for joint communication and positioning," *EURASIP Journal on Advances in Signal Processing*, no. 1, p. 178, 2012.

[SBKH10] K. Schmeink, R. Block, C. Knievel, and P. A. Hoeher, "Joint channel and parameter estimation for combined communication and navigation using particle swarm optimization," in *Proc. 7th Workshop on Positioning, Navigation and Communication (WPNC)*, Dresden, Germany, Mar. 2010, pp. 4–9.

[Sch78] G. Schwartz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[Sch86] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[Sch08] H. Schoeneich, "Adaptiver Interleave-Division Mehrfachzugriff (IDMA) mit Anwendung in der Mobilfunkkommunikation," Ph.D. dissertation, Kiel University, 2008.

[Sch12] K. Schmeink, "Joint communication and positioning based on interleave-division multiplexing," Ph.D. dissertation, Kiel, Germany, 2012.

[SD08] S. N. Sivanandam and S. N. Deepa, *Introduction to Genetic Algorithms*. Berlin, Heidelberg, New York: Springer, 2008.

[SDM14] S. Sand, A. Dammann, and C. Mensing, *Positioning in Wireless Communications Systems*. Chichester, UK: John Wiley & Sons, Ltd, Mar. 2014.

[SE98] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *Proc. IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence*. Achorage, AK, USA: IEEE, May 1998, pp. 69–73.

[SGD+15] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "5G position and orientation estimation through millimeter wave MIMO," in *2015 IEEE Globecom Workshops*. San Diego, CA, USA: IEEE, 12 2015, pp. 1–6.

[SGD+18] ——, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1822–1835, Mar. 2018.

[SH04] H. Schoeneich and P. A. Hoeher, "Adaptive interleave-division multiple access - a potential air interface for 4G bearer services and wireless LANs," in *Proc. 1st IEEE and IFIP Int. Conf. on Wireless and Optical Communications and Networks (WOCN)*, Muscat, Oman, Jun. 2004, pp. 179–182.

[SH06] ——, "Iterative pilot-layer aided channel estimation with emphasis on interleave-division multiple access systems," *EURASIP Journal on Advances in Signal Processing*, pp. 1–16, Dec. 2006.

[SH08] K. Schmeink and P. A. Hoeher, "Multi-layer interleave-division multiple access for joint communication and navigation," in *Proc. of International ITG Conference on Source and Channel Coding*, Ulm, Germany, Jan. 2008.

[Sky] "Skyhook." [Online]. Available: https://www.skyhook.com

[SM97] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, 1st ed. Englewood Cliffs, NJ: Prentice Hall, 1997.

[SMH03] D. Schafhubes, G. Matz, and F. Hlawatsch, "Kalman tracking of time-varying channels in wireless MIMO-OFDM systems," in *Proc. the Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, Pacific Grove, Nov. 2003, pp. 1261–1265.

[SS15] D. Shutin and N. Schneckenburger, "Joint detection and super-resolution estimation of multipath signal parameters using incremental automatic relevance determination," Mar. 2015. [Online]. Available: https://arxiv.org/pdf/1503.01898.pdf

[STK05] H. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location," *IEEE Signal Processing Magazine*, vol. 22, no. July 2005, pp. 24–40, Jun. 2005.

[STK08] Shih-Hau Fang, Tsung-Nan Lin, and Kun-Chou Lee, "A novel algorithm for multipath fingerprinting in indoor WLAN environments," *IEEE Transactions on Wireless Communications*, vol. 7, no. 9, pp. 3579–3588, Sep. 2008.

[SvROL99] F. Swarts, P. van Rooyen, I. Oppermann, and M. P. Loetter, Eds., *CDMA techniques for third generation mobile systems*, 1st ed. New York, NY: Springer Science+Business, 1999.

[TO95] T. Trump and B. Ottersten, "Maximum likelihood estimation of nominal direction of arrival and angular spread using an array of sensors," in *Proc. of COST 229, Adaptive Systems, Intelligent Approaches, Massively Parrallel Computing and Emergent Techniques in Signal Processing and Communications*, Vigo, Spain, Oct. 1995.

[TO96] ——, "Estimation of nominal direction of arrival and angular spread using an array of sensors," *Signal Processing*, vol. 50, no. 1-2, pp. 57–69, Apr. 1996.

[TV04] E. Trevisani and A. Vitaletti, "Cell-ID location technique, limits and benefits: An experimental study," in *Proc. Sixth IEEE Workshop on Mobile Computing Systems and Applications*, Windermere, Cumbria, UK, Dec. 2004, pp. 51–60.

[vdBES+95] J.-J. van de Beek, O. Edfors, M. Sandell, S. Wilson, and P. Borjesson, "On channel estimation in OFDM systems," in *Proc. IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century*, Chicago, USA, Jul. 1995, pp. 815–819.

[VDV19] M. Vaezi, Z. Ding, and H. Vincent Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond*, M. Vaezi, Z. Ding, and H. V. Poor, Eds. Cham: Springer International Publishing, 2019.

[vdVVP98] A.-J. van der Veen, M. 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.

[Wei75] J. Wei, "Least squares fitting of an elephant," *Chemtech*, vol. 5, pp. 128–129, 1975.

[WH05] T. Wo and P. A. Hoeher, "Semi-blind channel estimation for frequency-selective MIMO systems," in *Proc. 14th IST Mobile & Wireless Commun. Summit*, vol. paper no., Dresden, Jun. 2005.

[WK85] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[WLW09] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 427–250, Mar. 2009.

[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, Apr. 1997.

[WSH06] T. Wo, A. Scherb, and P. A. Hoeher, "Performance analysis of maximum-likelihood semiblind estimation of MIMO channels," in *Proc. EEE 63rd Vehicular Technology Conference*, Melbourne, Australia, May 2006, pp. 1738–1742.

[WZ89a] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, 1989.

[WZ89b] ——, "Detection of the number of coherent signals by the MDL principle," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 8, pp. 1190–1196, Aug. 1989.

[YB92] S. Yau and Y. Bresler, "A compact Cramer-Rao bound expression for parametric estimation of superimposed signals," *IEEE Transactions on Signal Processing*, vol. 40, no. 5, pp. 1226–1230, May 1992.

[YLCC01] B. Yang, K. Letaief, R. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Transactions on Communications*, vol. 49, no. 3, pp. 467–479, Mar. 2001.

[YOJ07] S. Yang, Y. S. Ong, and Y. Jin, Eds., *Evolutionary Computation in Dynamic and Uncertain Environments*. Berlin, Heidelberg, New York: Springer, 2007.

[ZFY+16] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and S. Li, "Channel estimation for millimeter wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.

[Zhu92] J. Zhu, "Calculation of geometric dilution of precision," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 893–895, Jul. 1992.

[ZL06] Q. Zao and J. Li, "Rain attenuation in millimeter wave ranges," in *Proc. of IEEE International Syposium Antennas, Propagation and EM Theory*, Guilin, China, Oct. 2006, pp. 1–4.

List of Figures

List of Tables