Multi-Dimensional Channel Estimation for MIMO-OFDM

Dissertation

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

vorgelegt von

Christopher Knievel

Kiel 2014

Tag der Einreichung:	28.10.2013
Tag der Disputation:	02.07.2014
Berichterstatter:	Prof. Dr.-Ing. Peter Adam Höher
	Prof. Dr. Sc. Techn. Bernard H. Fleury
	Prof. Dr.-Ing. Armin Dekorsy

Für Imke, Joona und Levi

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter in der Arbeitsgruppe für Informations- und Codierungstheorie der Christian-Albrechts-Universität zu Kiel.

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. Peter Adam Höher, der diese Arbeit betreut hat und mir die Freiheit in der Ausgestaltung der Forschungsschwerpunkte gegeben hat. Für das Interesse an dieser Arbeit und die Übernahme des Korreferats danke ich Herrn Prof. Dr. Sc. Techn. Bernard H. Fleury und Herrn Prof. Dr.-Ing. Armin Dekorsy. Ebenso danke ich dem Vorsitzenden der Promotionskommission Prof. Dr.-Ing. habil Thomas Meurer sowie Herrn Prof. Dr.-Ing. Werner Rosenkranz.

Insbesondere bedanke ich mich bei Dr. Gunther Auer der mir in vielen Diskussionen wertvolle Anregungen gab, die diese Arbeit nachhaltig beeinflußt haben. Ferner danke ich Dr. Alexander Tyrrell, Dr. Petra Weitkemper und Dr. Hidekazu Taoka für die fachlichen Diskussionen, die zum Gelingen dieser Arbeit beigetragen und zu vielen fruchtbaren Ideen geführt haben.

Weiterhin möchte ich mich bei allen Kollegen der Arbeitsgruppe für Informations- und Codierungstheorie und der Arbeitsgruppe für Digital Signal Processing and System Theory bedanken, die zu einer angenehmen Arbeitsatmosphäre beigetragen haben. Besonders hervorzuheben sind Meelis Noemm, Rebecca Adam, Dr. Kathrin Schmeink, Gilbert Forkel und Christian Lüke, die sich bei dieser Arbeit und Veröffentlichungen die Zeit genommen haben, diese zu lesen und Korrekturen einzubringen.

Schließlich danke ich in besonderer Weise meinen Eltern, die mir das Studium der Elektrotechnik ermöglicht haben und mich immer nach besten Kräften unterstützt haben, meinem Bruder Lennart für das Lesen der Arbeit, sowie meiner Frau Imke für ihre stete Unterstützung.

Kiel, im August 2014
Christopher Knievel

Kurzfassung

DIE DIGITALE drahtlose Kommunikation begann in den 1990er Jahren mit der zunehmenden Verbreitung von GSM. Seitdem haben sich Mobilfunksysteme drastisch weiterentwickelt. Aktuelle Mobilfunkstandards nähern sich dem Ziel eines omnipräsenten Kommunikationssystems an und erfüllen damit den Wunsch mit jedem Menschen zu jeder Zeit an jedem Ort kommunizieren zu können. Heutzutage ist die Akzeptanz von Smartphones und Tablets immens und das mobile Internet ist die zentrale Anwendung. Ausgehend von dem momentanen Wachstum wird das Datenaufkommen in Mobilfunk-Netzwerken im Jahr 2020, im Vergleich zum Jahr 2010, um den Faktor 1000 gestiegen sein und 100 Exabyte überschreiten.

Unglücklicherweise ist die verfügbare Bandbreite beschränkt und muss daher effizient genutzt werden. Schlüsseltechnologien, wie z.B. Mehrantennensysteme (multiple-input multiple-output, MIMO), orthogonale Frequenzmultiplexverfahren (orthogonal frequency-division multiplexing, OFDM) sowie weitere MIMO Codierverfahren, vergrößern die theoretisch erreichbare Kanalkapazität und kommen bereits in der Mehrheit der Mobilfunkstandards zum Einsatz. Auf der einen Seite verspricht MIMO-OFDM erhebliche Diversitäts- und/oder Kapazitätsgewinne. Auf der anderen Seite steigt die Komplexität der optimalen Maximum-Likelihood Detektion exponientiell und ist infolgedessen nicht haltbar. Zusätzlich wächst der benötigte Mehraufwand für die Kanalschätzung mit der Anzahl der verwendeten Antennen und reduziert dadurch die Bandbreiteneffizienz. Iterative Empfänger, die Datendetektion und Kanalschätzung im Verbund ausführen, sind potentielle Wegbereiter um den Mehraufwand des Trainings zu reduzieren und sich gleichzeitig der maximalen Kapazität mit geringerem Aufwand anzunähern.

Im Rahmen dieser Arbeit wird ein graphenbasierter Empfänger für iterative Datendetektion und Kanalschätzung entwickelt. Der vorgeschlagene multidimensionale Faktor Graph führt sogenannte Transferknoten ein, die die Korrelation benachbarter Kanalkoeffizienten in beliebigen Dimensionen, z.B. Zeit, Frequenz und Raum, ausnutzen. Hierdurch wird eine einfache und flexible Empfängerstruktur realisiert mit deren Hilfe weiche Kanalschätzung und Datendetektion in mehrdimensionalen, dispersiven Kanälen mit beliebiger Modulation und Codierung durchgeführt werden kann. Allerdings weist der Faktorgraph suboptimale Schleifen auf. Um die maximale Performance zu erreichen, wurde neben dem Ablauf des Nachrichtenaustausches und des Vorgangs zur Kombination von Nachrichten auch die Initialisierung speziell angepasst. Im Gegensatz zu herkömmlichen Methoden, bei denen mehrere Knoten zur Vermeidung von Schleifen zusammengefasst werden, verringern die vorgeschlagenen Methoden die leistungsmindernde Effekte von Schleifen, erhalten aber zugleich die geringe Komplexität des Empfängers. Zusätzlich wird ein neuartiger Detektionsalgorithmus vorgestellt, der baumbasierte Detektionsalgorithmen mit dem sogenannten Gauss-Detektor verknüpft. Der resultierende baumbasierte Gauss-Detektor (Gaussian tree search detector) lässt sich ideal in das graphenbasierte Framework einbinden und verringert weiter die Gesamtkomplexität des Empfängers. Zusätzlich wird Particle Swarm Optimization (PSO) zum Zweck der initialen Kanalschätzung untersucht. Der biologisch inspirierte Algorithmus ist insbesonders wegen seiner schnellen Konvergenz zu einem akzeptablen MSE und seiner vielseitigen Abstimmungsmöglichkeiten auf eine Vielzahl von Optimierungsproblemen interessant. Da PSO keine a priori Informationen benötigt, ist er speziell für die Initialisierung geeignet. Sowohl ein kooperativer Ansatz für PSO für Antennensysteme mit extrem vielen Antennen als auch ein multiobjective PSO für Kanäle, die in Zeit und Frequenz dispersiv sind, werden evaluiert.

Die Leistungsfähigkeit des multidimensionalen graphenbasierten iterativen Empfängers wird mit Hilfe von Monte Carlo Simulationen untersucht. Die Simulationsergebnisse werden mit denen eines dem Stand der Technik entsprechenden Empfängers verglichen. Es wird gezeigt, dass ähnliche oder bessere Ergebnisse mit geringerem Aufwand erreicht werden.

Eine weitere ansprechende Eigenschaft von iterativen semi-blinden Kanalschätzern ist, dass der mögliche Abstand von Trainingssymbolen die Grenzen des Nyquist-Shannon Abtasttheorem überschreiten kann. Im Rahmen dieser Arbeit wird eine Beziehung zwischen dem Trainingsabstand und dem Kanalcode formuliert. In Abhängigkeit des gewählten Kanalcodes und der Coderate folgt der maximale Trainingsabstand der vorgeschlagenen "coded sampling bound".

Stichwörter: Drahtlose Kommunikation, MIMO, OFDM, Graphentheorie, Belief propgation, baumbasierte Detektion, Kanalschätzung, Evolutionäre Algorithmen, Particle Swarm Optimization, Abtasttheorem

Abstract

DIGITAL wireless communication started in the 1990s with the wide-spread deployment of GSM. Since then, wireless systems evolved dramatically. Current wireless standards approach the goal of an omnipresent communication system, which fulfils the wish to communicate with anyone, anywhere at anytime. Nowadays, the acceptance of smartphones and/or tablets is huge and the mobile internet is the core application. Given the current growth, the estimated data traffic in wireless networks in 2020 might be 1000 times higher than that of 2010, exceeding 127 exabyte.

Unfortunately, the available radio spectrum is scarce and hence, needs to be utilized efficiently. Key technologies, such as multiple-input multiple-output (MIMO), orthogonal frequency-division multiplexing (OFDM) as well as various MIMO precoding techniques increase the theoretically achievable channel capacity considerably and are used in the majority of wireless standards. On the one hand, MIMO-OFDM promises substantial diversity and/or capacity gains. On the other hand, the complexity of optimum maximum-likelihood detection grows exponentially and is thus, not sustainable. Additionally, the required signaling overhead increases with the number of antennas and thereby reduces the bandwidth efficiency. Iterative receivers which jointly carry out channel estimation and data detection are a potential enabler to reduce the pilot overhead and approach optimum capacity at often reduced complexity.

In this thesis, a graph-based receiver is developed, which iteratively performs joint data detection and channel estimation. The proposed multi-dimensional factor graph introduces transfer nodes that exploit correlation of adjacent channel coefficients in an arbitrary number of dimensions (e.g. time, frequency, and space). This establishes a simple and flexible receiver structure that facilitates soft channel estimation and data detection in multi-dimensional dispersive channels, and supports arbitrary modulation and channel coding schemes. However, the factor graph exhibits suboptimal cycles. In order to reach the maximum performance, the message exchange schedule, the process of combining messages, and the initialization are adapted. Unlike conventional approaches, which merge nodes of the factor graph to avoid cycles, the proposed message combining methods mitigate the impairing effects of short cycles and retain a low computational complexity. Furthermore, a novel detection algorithm is presented, which combines tree-based MIMO detection with a Gaussian detector. The resulting detector, termed Gaussian tree search detection, integrates well within the factor graph framework and reduces further the overall complexity of the receiver. Additionally, particle swarm optimization (PSO) is investigated for the purpose of initial channel estimation. The bio-inspired algorithm is particularly interesting because of its fast convergence to a reasonable MSE and its versatile adaptation to a variety of optimization problems. It is especially suited for initialization since no a priori information is required. A cooperative approach to PSO is proposed for large-scale antenna implementations as well as a multi-objective PSO for time-varying frequency-selective channels.

The performance of the multi-dimensional graph-based soft iterative receiver is evaluated by means of Monte Carlo simulations. The achieved results are compared to the performance of an iterative state-of-the-art receiver. It is shown that a similar or better performance is achieved at a lower complexity.

An appealing feature of iterative semi-blind channel estimation is that the supported pilot spacings may exceed the limits given the by Nyquist-Shannon sampling theorem. In this thesis, a relation between pilot spacing and channel code is formulated. Depending on the chosen channel code and code rate, the maximum spacing approaches the proposed "coded sampling bound".

Keywords: Wireless communications, MIMO, OFDM, graph theory, belief propagation, tree-based detection, channel estimation, evolutionary algorithms, particle swarm optimization, sampling theorem

Contents

1

Introduction

1.1 Motivation

DURING the last decades, communication systems evolved continuously at an impressive speed. While the first (1G) and second generation (2G) were restricted to voice and text messages, the third generation (3G) embraced the mobile internet and enabled location-aware services. The fourth generation (4G) aims at mobile video conferencing and high-quality 3D graphics. With the advancing deployment of the mobile ecosystem, the desire to communicate with anyone, anywhere at anytime becomes stronger. Already, the demand of high-rate wireless communication systems is increasing exponentially [Eri13]. Given the current growth, the estimated data traffic in wireless networks in 2020 might be 1000 times higher than that in 2010 [HHI⁺12]. In order to meet these challenging demands several concepts have to be combined. Broadband communication is such a concept and represents an appealing strategy to increase the achievable capacity of a wireless channel [SBM⁺04]. In combination with orthogonal frequency-division multiplexing (OFDM), the assigned bandwidth is used efficiently. Unfortunately, the available radio spectrum is limited and an efficient use is therefore of utmost importance. The utilization of multiple antennas at both the transmitter side and the receiver side constitute a so-called multiple-input multiple-output (MIMO) system, which increases the achievable capacity linearly as a function of the number of transmit and receive antennas. In current research, asymmetric MIMO systems attract considerable interest of research [Mar10, LTEM13, RPL⁺13], since the available physical space is typically too constrained to ensure uncorrelated transmit and receive antennas. This is especially true at the mobile station. Hence, so-called massive MIMO systems target this problem by deploying a large number of antennas at the base station and assume single-antenna mobile stations. While this concept is appealing in multiple ways, i.e. very precise beamforming in the downlink and significantly reduced transmit power in the uplink, accurate channel state information at the transmitter side (CSIT) is crucial. In the absence of CSIT, versatile low-complexity multiple-input single-output detection and estimation algorithms are required to achieve a desirable throughput. Here, the proposed concepts for channel estimation as well as data detection offer viable solutions towards massive MIMO systems.

Moreover, advanced forward error correction codes, such as turbo codes and/or LDPC codes, have made a substantial progress and are able to approach the Shannon limit. The equivalent concept has been applied to equalization, which is dubbed “turbo equalization” [DJB$^+$95, KST04], and yields equally impressive gains. To facilitate coherent detection, knowledge of the channel response is required at the receiver end. The most common method to provide the receiver with channel state information (CSI) is to embed pilots, known at the receiver, within the transmitted signal stream [Cav91]. To reconstruct the OFDM channel impulse response at the positions of the unknown data symbols, two-dimensional interpolation and filtering over time and frequency is often adopted [Hoe91, HKR97b, Aue03b]. With the growing popularity of MIMO transmission, channel estimators operating not only over time and/or frequency, but also the spatial domain emerged [SZF02, MJ05, CL07, Aue12].

While MIMO-OFDM promises substantial diversity and/or capacity gains [FG98, Tel99], the required pilot overhead grows in proportion to the number of transmit antennas [HH03]. Iterative receivers which jointly carry out channel estimation and data detection are a potential enabler to reduce the required pilot overhead. These iterative receivers refine channel estimates by generation pseudo-pilots by previously detected data symbols [LWL01, VW01, CH03, AB07].

Unfortunately, the complexity of the performance-optimal maximum-likelihood (ML) estimator grows exponentially with the modulation order and the number of transmit antennas [SJS03]. The class of suboptimal iterative receivers based on the expectation maximization (EM) algorithm significantly reduce the computational cost and have attracted considerable interest recently [LWL01, XG03, KB06, YJ09]. However, the conventional structure of the EM based receiver prohibits the use of reliability information in terms of log-likelihood ratios (LLR) for iterative channel estimation, but relies on hard decisions instead. Furthermore, initialization of an EM based iterative receiver is susceptible to estimation errors. For the initialization, EM therefore requires either a preamble or a computationally complex algorithm, such as a linear MMSE estimator.

Graph-based algorithms pose a viable alternative for iterative receivers [WS01]. Factor graphs [KFL01, LDH$^+$07] constitute a versatile framework that has been applied to a variety of signal processing problems. Several graph-based receivers for iterative joint detection and channel estimation in combination with the sum-product algorithm have been published in recent years [NSRL05, NMH09, ZGH09, WHS12]. Nevertheless, all of these receivers are either designed for single-antenna systems and/or a single dimension, that is the estimation of the channel impulse response is done in either time domain or frequency domain. While the generalization to multiple antennas and/or channel estimation in multiple dimensions is often conceptually straight-forward, it is in general non-trivial.

The aim of this thesis is the development of a graph-based soft iterative receiver (GSIR), which facilitates joint data detection and channel estimation at a low complexity. The underlying graph structure and the message exchange algorithms are derived under the premise of a unified integration of multiple antennas as well as multiple dimensions. The corresponding receiver is termed multi-dimensional GSIR (MD-GSIR). In order to reach the best performance, the message exchange schedule, the process of combining messages, and the initialization are adapted. Since optimum ML detection is considered to be computationally infeasible, an alternative sub-optimum approach is desirable. However, popular tree-based detection methods based on QR decomposition are difficult to be implemented due to the proposed graph-structure. An alternative approach is given by the so-called Gaussian detector [PLL03], which approximates the multi-antenna interference by a Gaussian random variable and thereby reduces the complexity. Unfortunately, the achievable performance is rather poor for higher-order modulation. A novel detection method is presented, which combines the Gaussian detector with tree-based detection. The resulting Gaussian tree-search detector is ideally suited for the implementation in the MD-GSIR and offers a flexible trade-off between performance and complexity.

For the purpose of providing initial CSI, particle swarm optimization (PSO) has been thoroughly investigated. The concept of cooperative PSO has been applied to MIMO channel estimation and shows to be especially appealing for modern massive MIMO systems. Nevertheless, in previous work it is restricted to time-invariant flat-fading channels. Therefore, an extension to time-varying frequency-selective channels based on multi-objective PSO is proposed. The versatility of the PSO algorithm and the quick convergence to a reasonable mean squared error performance are beneficial to maintain an overall low-complexity solution. In combination with an initialization based on multi-objective PSO, the MD-GSIR reaches its optimum performance under various channel conditions. In extensive simulations, the achievable performance of the MD-GSIR is highlighted and compared to a state-of-the-art iterative receiver. It is shown that a similar performance is reached at a lower complexity. Additionally, the requirements w.r.t. pilot overhead and a priori information are lower. Moreover, the versatility of the MD-GSIR is illustrated by the adaptation of codebook-based beamforming.

One of the virtues of iterative joint channel estimation and data detection is the reduced pilot overhead. The pilots can be separated to such an extent that the Nyquist-Shannon sampling theorem is violated. This effect has been observed for the MD-GSIR and has been individually reported in [tBSS00, SJS03, XM11]. In this thesis, a relation between the spacing of pilots and the channel code as well as code rate is formulated for iterative semi-blind channel estimation. Depending on the channel code and code rate the resulting maximum pilot spacing is upper bounded, by the so-called "coded sampling bound".

The major contributions of this thesis can be summarized as follows:

- Design and evaluation of a novel Gaussian tree search (GTS) detector, which offers an efficient approach for the detection of rank-deficient MIMO systems.

- The combination of correlated random variables is thoroughly investigated and applied within the MD-GSIR.

- Particle Swarm Optimization is examined for the use of MIMO channel estimation in combination with a flat-fading time-invariant channel as well as a frequency-selective time-varying channel.

- A graph-based framework for multi-dimensional joint channel estimation and data detection is developed. The convergence behavior is studied in detail. Furthermore, initialization, message combining, and message scheduling is adapted to achieve the performance of a state-of-the-art receiver at lower complexity.
- The Nyquist-Shannon sampling theorem is evaluated for iterative semi-blind channel estimation and a corresponding coded sampling bound is derived.

Parts of this thesis were published as journal papers or refereed conference papers in [KSHA10], [KHAT11a], [KNH11], [KHTA12a], [KH12], [KHTA12b], [KHA12], and [KHH$^+$13].

1.2 Thesis Outline

This thesis is organized as follows: In **Chapter 2**, the mobile radio channel is introduced. A description of a linear channel model and the underlying first- and second-order characterization of the selectivity in time, frequency, and space is presented. Subsequently, more practically oriented channel scenarios given by the WINNER channel models are reviewed, which are used throughout this thesis to assess the performance of the proposed algorithms. Furthermore, the concept of MIMO is briefly discussed and an introduction to codebook-based beamforming is given, followed by an outline of the advantages and challenges of massive MIMO systems. To conclude this chapter, the basic principle as well as implementation aspects of OFDM are investigated.

Chapter 3 reviews data detection methods and channel estimation algorithms suitable for iterative processing. An overview of MIMO detection methods is given, which comprises tree-based detection based on QR decomposition, a Gaussian detector, and a novel Gaussian tree search (GTS) detection. The latter combines tree-based detection with the Gaussian approximation and thereby supersedes the QR decomposition. The complexity of the three algorithms is discussed and simulation results illustrating the BER performance are shown. Subsequently, pilot-based channel estimation (PACE) is studied. To facilitate coherent detection, pilots are multiplexed into the data stream, which are used at the receiver side to estimate the channel response. Two common methods for PACE are introduced, namely least-squares and minimum mean squared error channel estimation. The achievable performance w.r.t. MSE and spectral efficiency depends on the pilot grid. Hence, strategies for the design of a pilot grid are summarized and common pilot grids are studied. The exchange and the combining of reliability information is a key component of turbo decoding and turbo equalization. A common property is that the random values which are combined have to be statistically independent. Under certain circumstances, these random measures may be correlated. The combination of correlated random measures is explained in detail.

The applicability of particle swarm optimization (PSO) to MIMO channel estimation is evaluated in **Chapter 4**. Beginning with the evolution of PSO, several general modifications, which increase the performance, are summed up. Three different variants of

PSO are presented, which are able to solve single-objective optimization problems. This means, there is only one single/multi-dimensional solution which yields the optimum result. Hence, given a time-invariant flat-fading channel, PSO can be directly applied. It is illustrated that with increasing dimensions the number of iterations is increasing exponentially. With the proposed cooperative approach, the number of necessary iterations can be significantly reduced. The initialization of a massive MIMO system is identified as an ideal application for cooperative PSO. Moreover, as common wireless channels are typically time-varying and frequency-selective, a multi-objective PSO (MOPSO) is derived. Its performance is evaluated with four channel scenarios, which are part of the WINNER channel model. By using linear prediction, the MSE performance is increased compared to a LS channel estimator. Similarly to the conventional single-objective PSO, the major advantage of MOPSO lies in its fast convergence to a reasonable MSE. Thus, it is ideally suited to provide initial channel state information.

In **Chapter 5**, the graph-based soft iterative receiver (GSIR), which facilitates joint soft data detection and channel estimation, is derived. The factorization of the objective function, the development of the underlying graph structure, and the message generation are studied in detail. The core part of the GSIR is the message exchange in arbitrary channel dimensions, which is established by so-called transfer nodes. The complete factor graph follows a unified design to implement multiple antennas and multiple dimensions while simultaneously maintaining a very low computational complexity. However, the focus of minimal computational complexity has to be balanced with a trade-off between complexity and achievable performance. Due to the premise of low complexity, the scheduling, and the message combining are adapted to significantly increase the performance. It is shown that the performance of the GSIR can be further increased by a suitable initialization. The MOPSO algorithm presented in Chapter 4, improves the overall performance of the GSIR and even outperforms a Wiener-filter-initialized GSIR with QPSK modulation. The achievable performance in four WINNER channel scenarios is thoroughly investigated for a wide range of modulation formats and code rates. It is observed that the achievable performance is similar to a state-of-the-art receiver. However, the proposed GSIR is more robust w.r.t. a priori information, requires less pilots, and is of lower complexity. The versatility of the receiver is highlighted with the implementation of a codebook-based beamforming transmission. The transfer nodes are adapted to take precoding into account and reach impressive performance gains.

The pilot overhead induced by coherent detection in MIMO systems can be significant. Joint channel estimation and data detection is able to reduce this overhead and thereby increases the achievable bandwidth efficiency. Interestingly, the Nyquist-Shannon sampling theorem is not strictly limiting the pilot spacing any longer. An upper limit of the pilot spacing for semi-blind iterative channel estimation is derived in **Chapter 6**. The dependency of the channel code as well as code rate on the initial pilot spacing is analyzed. It is shown that depending on the channel code and code rate, the pilot spacing can be arbitrarily large. Furthermore, the implications for a practical receiver design are drawn.

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

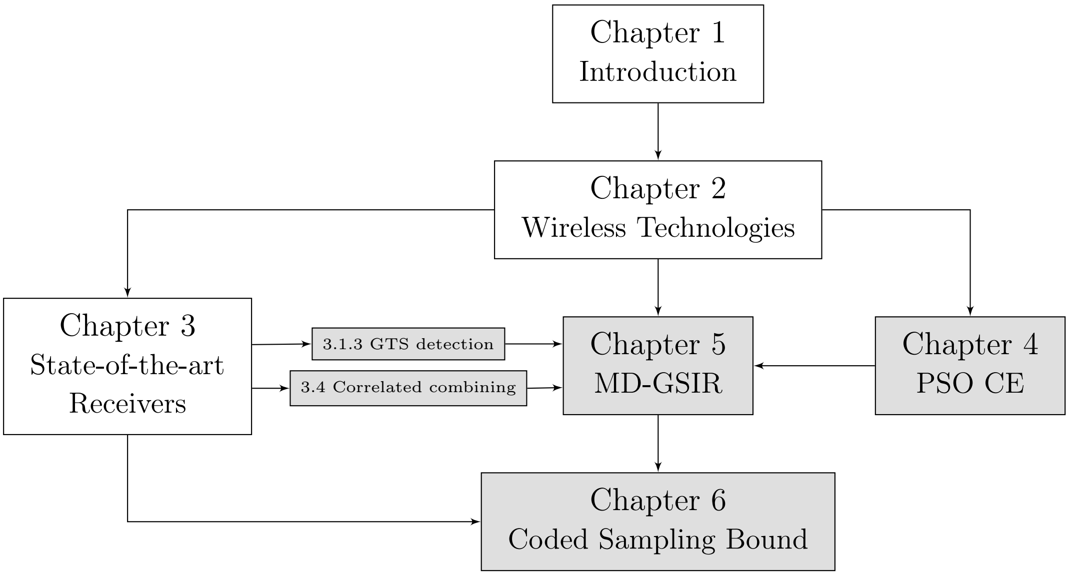


Figure 1.1: Graphical outline of the thesis.

A graphical overview of the thesis structure is given in Figure 1.1. Hereby, gray colored boxed consist of review as well as introductory parts. The white colored boxes represent original contributions developed in the course of this thesis.

2

Wireless Multi-Antenna Multi-Carrier Technologies

INTENSIVE research on multiple antenna systems started a little more than a decade ago in the area of wireless communications. Initiated by the realization that the use of multiple antennas substantially improves the performance of a wireless system in terms of reliability, capacity, quality of service, etc., antenna arrays are deployed at the transmitter side as well as at the receiver side. In addition to antenna technologies, broadband transmission is often used to further increase the peak data rate. For the development, assessment, and optimization of such systems suitable channel models are required. The aim of this chapter is to provide an introduction of the wireless technologies which have been addressed in this thesis as well as an general overview of the area.

A comprehensive study of linear time variant systems was introduced by Bello in his seminal papers [Bel63, Bel64]. This concept was later extended, among others, by Fleury to the spatial domain [Fle00]. In the beginning of this chapter, a subset of their ideas, relevant for this thesis, are presented. Based on this theoretical foundation, the WINNER channel model is presented subsequently, which has been derived based on extensive measurement campaigns. Furthermore, the potential gains of multiple-input multiple-output (MIMO) systems are highlighted. MIMO techniques, such as beamforming, are briefly examined, followed by a discussion of modern large-scale implementations of MIMO systems. The resulting challenges are briefly described. Orthogonal frequency-division multiplexing (OFDM) is often considered as one of the key enablers to facilitate high data rate transmission. Subsequently, the advantages and design trade-offs are presented.

2.1 Time-Varying Linear Channels

A linear system $\mathcal{S}$ transforms a continuous signal $x(t)$ into the output signal $y(t)$ as follows:

$$x(t) \overset{\mathcal{S}}{\mapsto} y(t). \tag{2.1}$$

Such a linear system is completely characterized by its response signal $h(\tau, f_{\mathrm{D}})$ given an impulse-shaped input signal [Bel63, Hub96]. The resulting transformation is represented by

$$\delta(t - \tau_{\mathrm{A}}) \overset{\mathcal{S}}{\mapsto} h(\tau, f_{\mathrm{D}}), \tag{2.2}$$

where t denotes the time and τ_{A} the excitation time. The response signal $h(\tau, f_{\mathrm{D}})$ is, thus, the time-variant impulse response of a linear time-variant channel.

Typically, an input signal of a mobile radio channel experiences multipath propagation due to reflection, diffraction and scattering of the electromagnetic wave at various objects. Thus, it is likely that rays emanating from a transmitter Tx reach a receiver Rx via different multipaths and consequently with different delays. A common approach is to group a set of rays with roughly the same delay into one cluster, which leads to M_{c} resolvable clusters. Furthermore, a linear time-variant channel, in particular the mobile radio channel, is assumed to be wide-sense stationary uncorrelated scattering (WSSUS) [Pro00], i.e. the fading statistics remain constant over short periods of time and the multipaths are independent and identically distributed (i.i.d.). Although this approximation is typically not fulfilled for real channels [Mat05], the virtue of this assumption lies in the significant reduction of parameters. This explains that the vast majority of channel models rely on the WSSUS assumption [Hub96].

A detailed description of the selectivity in time and frequency domain is given in Bello's original model. An analogous extension to the spatial domain is presented in [Fle00], which additionally considers the incidence direction Ω. Without loss of generality, only the receiver side is considered. The corresponding channel impulse response is given by

$$h(\Omega, \tau, f_{\mathrm{D}}) = \sum_{i=1}^{\mathrm{M_c}} \alpha_i(\Omega_i, t)\delta(\tau - \tau_i), \tag{2.3}$$

with

$$\alpha_i(\Omega_i, t) = a_i \cdot c_n(\Omega_i) \cdot \exp\left(2j\pi f_{\mathrm{D,i}} t\right) \exp\left(-2j\pi f_{\mathrm{c}} \tau_i\right) \exp\left(-j2\pi\lambda^{-1}(\Omega \cdot x)\right), \tag{2.4}$$

where a_i, Ω_i denote the amplitude and incidence direction of the ith impinging wave and $f_{\mathrm{D,i}}$, f_{c} refer to the Doppler frequency and carrier frequency. The vector/point Ω is determined by its spherical coordinates $(\phi, \varphi) \in [-\pi, \pi) \times [0, \pi]$ as illustrated in Fig 2.1. The angles ϕ and φ denote the azimuth and coelevation of Ω. Furthermore, $c_n(\Omega)$ is the complex field pattern of the nth array element, and λ is the wavelength [Fle00]. In this thesis, the coelevation angle φ is assumed to be zero, hence, the magnetic wave is only propagating horizontally. As a result, the incidence direction is completely described by the azimuth angle ϕ.

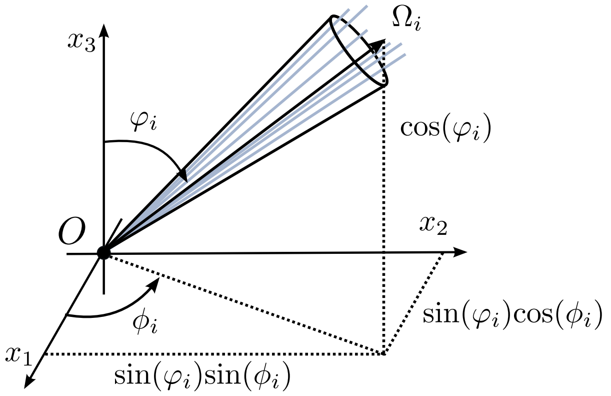


Figure 2.1: Multipath propagation and characterization of an incidence direction Ω_i.

2.1.1 Channel Correlation Functions and Power Spectra

As mentioned in the previous section, the mobile radio channel is characterized by the time-varying channel impulse response $h(\Omega, \tau, f_\mathrm{D})$ and/or by the time-varying channel transfer function $H(x, f, t) \doteq \mathcal{F}\left(h(\Omega, \tau, f_\mathrm{D})\right)$, which is the Fourier transform of $h(\Omega, \tau, f_\mathrm{D})$. Under the presumption of WSSUS, the space-frequency-time and the direction-delay-Doppler correlation function respectively, are defined as

$$\mathrm{E}\left\{H^*(x, f, t)H(x{+}\Delta x, f{+}\Delta f, t{+}\Delta t)\right\} \doteq \theta_{HH}\left(\Delta x, \Delta f, \Delta t\right) \tag{2.5}$$
$$\mathrm{E}\left\{h^*(\Omega, \tau, f_\mathrm{D})h(\Omega{+}\Delta\Omega, \tau{+}\Delta\tau, f_\mathrm{D}{+}\Delta f_\mathrm{D})\right\} \doteq \Theta_{SS}\left(\Omega, \tau, f_\mathrm{D}\right)\delta(\Delta\Omega)\delta(\Delta\tau)\delta(\Delta f_\mathrm{D}) \tag{2.6}$$

and provide the second-order characterization of selectivity, jointly in space, frequency and time as well as of dispersion in direction, delay, and Doppler frequency [Fle00]. The direction-delay-Doppler power spectrum

$$\Theta_{SS}\left(\Omega, \tau, f_\mathrm{D}\right) \doteq \mathrm{E}\left\{|h(\Omega, \tau, f_\mathrm{D})|^2\right\} \tag{2.7}$$

describes the distribution of the average output power as a function of direction, delay, and Doppler frequency.

The duality of time and frequency domain initially described by Bello in [Bel64] and extended by Fleury in [Fle00] connects the space-frequency-time correlation function $\theta_{HH}\left(\Delta x, \Delta f, \Delta t\right)$ with the direction-delay-Doppler power spectrum $\Theta_{SS}\left(\Omega, \tau, f_\mathrm{D}\right)$ via corresponding Fourier transformations [Hub96, Fle00]:

$$\begin{aligned}\theta_{HH}\left(\Delta x, \Delta f, \Delta t\right) =& \mathcal{F}^{-1}_{\Delta x}\left\{\mathcal{F}^{-1}_{\Delta t}\left\{\mathcal{F}_{\Delta f}\left\{\Theta_{SS}\left(\Omega, \tau, f_\mathrm{D}\right)\right\}\right\}\right\} \\ =& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Theta_{SS}\left(\Omega, \tau, f_\mathrm{D}\right) \\ & \cdot \exp\left(-j2\pi\left[(\Omega\cdot\Delta x)\lambda^{-1} - \Delta f\tau + \Delta t f_\mathrm{D}\right]\right)d\Omega d\tau df_\mathrm{D}.\end{aligned} \tag{2.8}$$

Hereby, $\mathcal{F}_{\Delta\cdot}\{\cdot\}$ and $\mathcal{F}^{-1}_{\Delta\cdot}\{\cdot\}$ correspond to the Fourier and inverse Fourier transform, respectively, as a function of the domain-specific dispersion parameter. Given a wide-sense stationary channel, the correlation function of a specific domain is obtained by setting the parameters of the remaining domains to zero. Obviously, the degree of correlation depends on the domain-specific spreading function. By integrating the direction-delay-Doppler power spectrum $\Theta_{SS}\left(\Omega, \tau, f_{\mathrm{D}}\right)$ over the Doppler frequency f_{D} and the incidence direction Ω, the delay spread function or power delay profile (PDP) is obtained:

$$\Theta_{SS}\left(\tau\right) \doteq \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \Theta_{SS}\left(\Omega, \tau, f_{\mathrm{D}}\right) d\Omega df_{\mathrm{D}}. \tag{2.9}$$

Correspondingly, the frequency correlation function is the Fourier transform of the power delay profile $\Theta_{SS}\left(\tau\right)$:

$$\theta_{HH}\left(\Delta f\right) \doteq \mathcal{F}_{\Delta f}\left\{\Theta_{SS}\left(\tau\right)\right\} = \int\limits_{-\infty}^{\infty} \Theta_{SS}\left(\tau\right) \exp\left(-j2\pi\tau\Delta f\right) d\tau. \tag{2.10}$$

The power delay profile describes the average received power as a function of the time delay τ. A common radio channel consists of multiple resolvable propagation paths, such that a transmitted signal arrives with different propagation delays at a receiver. The time delay τ describes the propagation delay which exceeds the delay of the direct path. The time $\tau_{\max}$ denotes the *maximum excess delay* and is given by the time between the first and last received component. If the signal duration T_{s} is smaller than the maximum excess delay $\tau_{\max}$ ($T_{\mathrm{s}}<\tau_{\max}$), the channel is said to be *frequency-selective* and generates intersymbol interference (ISI). The channel is *frequency-nonselective* or *flat fading* if the signal duration is larger than the maximum excess delay ($T_{\mathrm{s}}>\tau_{\max}$). Hence, all multipath components arrive within the symbol duration and are, thus, not resolvable [Skl97]. Due to the duality of time and frequency, a similar classification of the multipath fading can be done in the frequency domain on the basis of the frequency correlation function. Hereby, the *coherence bandwidth* $(\Delta f)_c$ refers to the range of frequencies which are mutually correlated with a certain threshold $c_f \in [0, 1)$ [Fle00, Skl97]:

$$(\Delta f)_c \doteq \min\left\{\Delta f > 0 : \left|\theta_{HH}\left(\Delta f\right)\right| = c_f\right\}. \tag{2.11}$$

Typically, the threshold c_f is chosen relatively large, e.g.: $c_f \geq 0.9$. In addition, the channel is said to be frequency-selective if the signal bandwidth B is larger than the coherence bandwidth $(\Delta f)_c$. Furthermore, the maximum excess delay is reciprocally related to the coherence bandwidth via its inverse within a multiplicative constant [Skl97]. Besides the maximum excess delay, other characteristic parameters can be extracted from the power delay profile, such as the root mean squared (rms) delay spread τ_{rms} as well as the mean excess delay $\overline{\tau}$. While the maximum excess delay is one important parameter for the classification of the fading, a more useful measure is offered by the rms delay spread, since it also considers the actual shape of the power delay profile, which has a significant influence on the fading characteristic. The rms delay spread is defined as follows

$$\tau_{\mathrm{rms}} \doteq \sqrt{\frac{\int_{0}^{\infty} (\tau - \overline{\tau})^2 \cdot \Theta_{SS}\left(\tau\right) d\tau}{\int_{0}^{\infty} \Theta_{SS}\left(\tau\right) d\tau}}, \tag{2.12}$$

with the mean excess delay given by

$$\overline{\tau} \doteq \frac{\int\limits_{0}^{\infty} \tau \cdot \Theta_{SS}\left(\tau\right) d\tau}{\int\limits_{0}^{\infty} \Theta_{SS}\left(\tau\right) d\tau}. \tag{2.13}$$

By integrating the direction-delay-Doppler power spectrum over the incidence direction Ω and the time delay τ, the Doppler power density spectrum is obtained:

$$\Theta_{SS}\left(f_{\mathrm{D}}\right) \doteq \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \Theta_{SS}\left(\Omega, \tau, f_{\mathrm{D}}\right) d\Omega d\tau. \tag{2.14}$$

The corresponding time correlation function is given by the inverse Fourier transform of the Doppler power density spectrum:

$$\theta_{HH}\left(\Delta t\right) \doteq \mathcal{F}_{\Delta t}^{-1}\left\{\Theta_{SS}\left(f_{\mathrm{D}}\right)\right\} = \int\limits_{-\infty}^{\infty} \Theta_{SS}\left(f_{\mathrm{D}}\right) exp\left(j2\pi f_{\mathrm{D}}\Delta t\right) df_{\mathrm{D}}. \tag{2.15}$$

Moreover, the Doppler power density spectrum describes the average received power as a function of the Doppler frequency f_{D}, whereas the maximum occurring Doppler frequency

$$f_{\mathrm{D,max}} \doteq \frac{v \cdot f_{\mathrm{c}}}{c} \tag{2.16}$$

depends on the velocity of the mobile station v, the carrier frequency f_{c}, and the speed of light c. A moving transmitter and/or receiver or a change in the environment may cause a change in the multipath propagation such that the amplitude and phase of the transmitted signal change over time. Accordingly, the channel is said to be time-varying. The coherence time $(\Delta t)_c$ denotes the duration over which the channel is essentially time-invariant, i.e. the time correlation function is above a threshold c_t:

$$\left(\Delta t\right)_c \doteq \min\left\{\Delta t > 0 : \left|\theta_{HH}\left(\Delta t\right)\right| = c_t\right\}. \tag{2.17}$$

Hence, the channel is *time-selective* if the coherence time $(\Delta t)_c$ is smaller than the symbol duration T_{s}. The term *fast fading* is often used to classify the situation where the coherence time is much smaller than the symbol duration. On the other hand, if the coherence time is larger than the symbol duration, the channel is said to be non-selective in time or *time-invariant*. Correspondingly, a situation for which the channel statistics remain quasi-constant is termed *slow fading*.

Analogous to the delay and Doppler domain, the direction spread function is obtained by integrating the direction-delay-Doppler power spectrum over the time delay τ and the Doppler frequency f_{D}:

$$\Theta_{SS}\left(\Omega\right) \doteq \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \Theta_{SS}\left(\Omega, \tau, f_{\mathrm{D}}\right) d\tau df_{\mathrm{D}}. \tag{2.18}$$

The corresponding space correlation function is given by

$$\theta_{HH}\left(\Delta x\right) \doteq \mathcal{F}_{\Delta x}^{-1}\left\{\Theta_{SS}\left(\Omega\right)\right\} = \int_{-\infty}^{\infty} \Theta_{SS}\left(\Omega\right)\exp\left(j2\pi(\Omega{\cdot}\Delta x)\lambda^{-1}\right) d\Omega. \tag{2.19}$$

Furthermore, a so-called $\tilde{\Omega}$-constrained distance correlation function can be defined as follows [Fle00]

$$\theta_{HH}^{\tilde{\Omega}}\left(\Delta d\right) \doteq \theta_{HH}\left(\Delta x\right)|_{\Delta x=\Delta d\tilde{\Omega}}, \tag{2.20}$$

which is of interest when the output of a linear antenna array is investigated. The $\tilde{\Omega}$-constrained space correlation function is uniquely determined by its sample at $\lambda/2$-equidistant points along the line $x = d\tilde{\Omega}$ [Fle00]. Similarly to the coherence time and coherence frequency, a coherence distance can be defined. The $\tilde{\Omega}$-constrained coherence distance of the radio channel is given by

$$(\Delta d)_{\tilde{\Omega},c} \doteq \min\left\{\Delta d > 0 : |\theta_{HH}^{\tilde{\Omega}}\left(\Delta d\right)| = c_d\right\}. \tag{2.21}$$

The meaning of the $\tilde{\Omega}$-constrained coherence distance depends on the selection of c_d. When chosen relatively large (e.g. $c_d \approx 0.8 - 0.9$), the coherence distance can be interpreted as the maximum length of displacement along the $\tilde{\Omega}$-direction for which the variation of the channel is constant. While this interpretation can be formulated generally within the concept of a WSS process, a second interpretation is possible assuming that only one resolvable multipath component arrives per incidence direction. Under this condition, a small value of c_d (e.g. $c_d \leq 0.5$) can be interpreted as the least separation, beyond which samples of the $\tilde{\Omega}$-constrained channel transfer function are uncorrelated [Fle00].

2.1.2 Example Correlation Functions

The first- and second-order characterization of a LTV channel presented in the previous section—in terms of multi-dimensional autocorrelation function and power spectrum—is often used for channel estimation algorithms, cf. Section 5.1.4 and Section 6.1. Hence, a few selected examples are presented in the following. It is obvious from (2.10), (2.15), and (2.19), that the distribution of the time delay, Doppler frequency, and incidence direction has a paramount influence on the selectivity in frequency, time, and spatial domain, respectively. Although the distribution of each parameter depends highly on the environment, i.e. the fading of an indoor channel exhibits a different behavior than a rural channel model, only few distributions for each domain are commonly applied to approximate the true fading behavior.

Simplified distributions for the direction, delay, and Doppler power spectrum are assumed to illustrate the dependency of the autocorrelation functions on the distribution of the dispersion parameters. Without loss of generality, it is assumed that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Theta_{SS}\left(\Omega, \tau, f_{\mathrm{D}}\right) d\Omega d\tau df_{\mathrm{D}} \doteq 1. \tag{2.22}$$

(a) Angular spread ϑ around azimuth angle ϕ=0°

(b) Uniform distribution of the azimuth angle ϕ

(c) Uniform distribution of the time delay τ

(d) Exponential decaying distribution of the time delay τ

(e) Uniform distribution of the Doppler frequency f_{D}

(f) Doppler frequency f_{D} distributed with Jakes Doppler power spectrum

Figure 2.2: Exemplary distribution functions for the incidence direction Ω, the time delay τ, and the Doppler frequency f_{D}.

A uniform distribution for the time delay between $\tau_i \in [-\tau_{\max}/2, \tau_{\max}/2]$ and for the Doppler frequency between $f_{\mathrm{D,i}} \in [-f_{\mathrm{D,max}}, +f_{\mathrm{D,max}}]$ is chosen. Furthermore, the impinging waves are uniformly distributed within the angular spread $\vartheta \in [-\vartheta_{\max}/2, +\vartheta_{\max}/2]$ around an azimuth angle of $\phi = 0°$. Sample distributions are shown in Figures 2.2c, 2.2e, and 2.2a, respectively. Corresponding autocorrelation functions in space, frequency, and time are given by

$$\theta_{HH}(\Delta x) = \operatorname{sinc}(\vartheta \Delta x), \tag{2.23}$$

$$\theta_{HH}(\Delta f) = \operatorname{sinc}(\tau_{\max} \Delta f), \tag{2.24}$$

$$\theta_{HH}(\Delta t) = \operatorname{sinc}(2 f_{\mathrm{D,max}} \Delta t), \tag{2.25}$$

where $\operatorname{sinc}(x) = \sin(\pi x)/(\pi x)$. As expected, the function describing the domain-specific autocorrelation function is equal in this case, depending only on the domain-specific dispersion parameters. Obviously, the simplified distributions in this constructed example are not suitable to accurately model a realistic multipath propagation. Nevertheless, they may be used as approximations when detailed information of the fading statistics is not available, i.e. only the maximum value of the dispersion parameter is given.

Numerous measurement campaigns have been made in order to develop more accurate distributions which resemble the reality closer. Among several popular distributions an exponential decaying distribution for the power delay profile (cf. Figure 2.2d), the so-called Jakes' model for the Doppler power spectrum (cf. Figure 2.2f), and a uniform distribution of the azimuth angle ϕ are often applied [Jak75, SW94, Pro00].

In case the coelevation angle is zero and the azimuth of the impinging waves is uniformly distributed within 360°, we refer to *2D isotropic scattering*, while an additional uniform distribution of the coelevation angle between 0 and 180° is termed *3D isotropic scattering* [Hoe13]. The impact of non-isotropic scattering shows that the effect of different incidence directions on the delay power profile as well as the Doppler power spectrum is not negligible [IST07, SG08, Aue12]. This is motivated by the fact that dominant spatially separated reflectors with different azimuth angles lead to distinct delays [Aue12].

The direction spread function for a uniform distribution of the azimuth angle ϕ

$$\Theta_{SS}\left(\phi\right) = \begin{cases} \frac{1}{2\pi} & \text{for } -\pi < \phi \leq \pi \\ 0 & \text{else.} \end{cases} \tag{2.26}$$

can be observed in densely buildup areas without a line-of-sight [Fle00]. This distribution describes the above mentioned 2D isotropic scattering. The corresponding direction autocorrelation function is given by [SW94]:

$$\theta_{HH}\left(\Delta x\right) = J_0\left(2\pi\frac{\Delta x}{\lambda}\right), \tag{2.27}$$

where $J_0(\cdot)$ is the Bessel function of the first kind and order zero. The power delay profile for an exponential decaying power delay profile is given by

$$\Theta_{SS}\left(\tau\right) = \begin{cases} \frac{1}{\tau_{\mathrm{rms}}}\exp\left(-\tau/\tau_{\mathrm{rms}}\right) & \text{for } 0 \leq \tau \leq \tau_{\mathrm{max}} \\ 0 & \text{else,} \end{cases} \tag{2.28}$$

which results after a Fourier transformation into the frequency correlation function

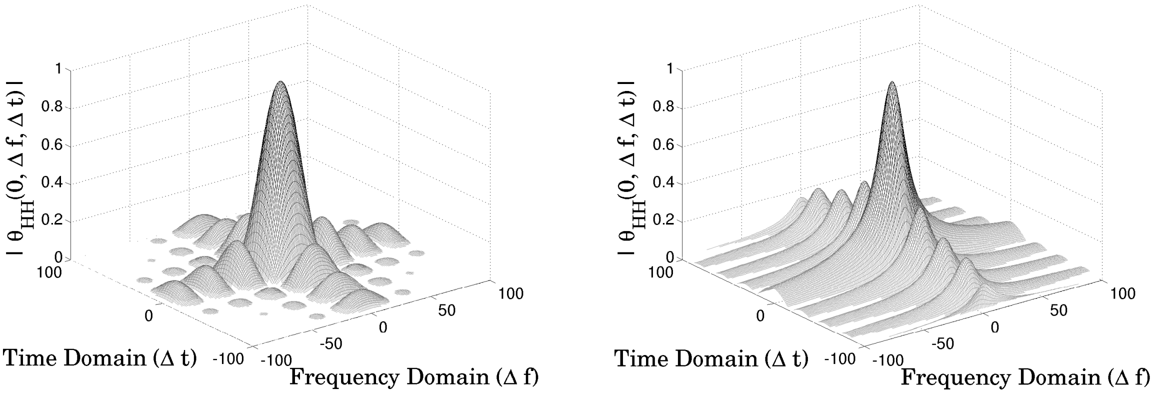
$$\theta_{HH}\left(\Delta f\right) = \frac{1}{1 + j2\pi\tau_{\mathrm{rms}}\Delta f}. \tag{2.29}$$

In case of non-isotropic scattering, different incidence directions lead to distinct delays. Hence, a joint frequency-space correlation function $\theta_{HH}(\Delta x, \Delta f)$ is required to model the fading characteristics accurately. However, the calculation of the corresponding azimuth-delay spectrum can be decomposed as [PMF00]

$$\Theta_{SS}\left(\phi, \tau\right) \propto \Theta_{SS}\left(\phi\right)\Theta_{SS}\left(\tau\right). \tag{2.30}$$

Motivated from the results of several measurement campaigns, the decomposition has been applied for various channel conditions [PMF00, Aue12]. Moreover, if the multipath rays can be grouped into M_{c} clusters with similar delays and incidence directions, the azimuth-delay spectrum is calculated as [PMF00]

$$\Theta_{SS}\left(\phi, \tau\right) = \sum_{c=1}^{M_{\mathrm{c}}} \Theta_{SS,c}\left(\phi, \tau\right). \tag{2.31}$$



(a) Time-frequency correlation with a uniform distribution in delay and Doppler domain

(b) Time-frequency correlation with a Jakes spectrum in the Doppler domain and exponential decaying function in the delay domain

Figure 2.3: Absolute value of the time-frequency autocorrelation function as a function of power delay profile and Doppler power spectrum.

The assumption of 2D isotropic scattering for the azimuth angle ϕ results in a Doppler power spectrum, which is often referred to as "Jakes spectrum". A mathematical description of the Jakes spectrum is given by [Jak75, Pro00]:

$$\Theta_{SS}\left(f_{\mathrm{D}}\right) = \begin{cases} \frac{1}{\pi f_{\mathrm{D,max}}} \frac{1}{\sqrt{1-\left(f_{\mathrm{D}}/f_{\mathrm{D,max}}\right)^2}} & \text{for } |f_{\mathrm{D}}| \leq f_{\mathrm{D,max}} \\ 0 & \text{else.} \end{cases} \tag{2.32}$$

In addition, the inverse Fourier transform of the Jakes' spectrum yields the well-known time autocorrelation function

$$\theta_{HH}\left(\Delta t\right) = J_0\left(2\pi f_{\mathrm{D,max}} \Delta t\right). \tag{2.33}$$

Given a dominant incidence direction, the obtained Doppler power spectrum corresponds to partial sections of the Jakes spectrum [Hoe13].

Without loss of generality, the two-dimensional correlation function in time and frequency domain with a uniform distribution for the power delay profile and Doppler power spectrum is given exemplary in Figure 2.3a. The dispersion parameters for time and frequency are chosen as $\tau_{\mathrm{max}}\Delta f = 2 f_{\mathrm{D,max}}\Delta t = 0.01$. The resulting symmetry is apparent. Additionally, the two-dimensional correlation function with an exponential decaying power delay profile and a Jakes Doppler power spectrum is shown in Figure 2.3b. The periodic sharp declines in the amplitude of the correlation function along the time domain are characteristic. The exponential decay of the amplitude in the frequency domain is also clearly visible. Although the two shapes vary significantly for larger values of Δt and Δf, they are rather similar for smaller values. Hence, for a limited range, the uniform distributions represent suitable approximations.

2.2 WINNER Channel Model

In order to evaluate the performance of the proposed graph-based iterative receiver in a realistic environment, different scenarios of the wireless world initiative new radio (WINNER) channel models, derived by measurement campaigns, are used [IST07]. The trade-off between computational complexity of the channel model and the degree of accuracy has been addressed by implementing two different concepts. Namely, a generic model suitable for system level simulations, and a clustered delay model (CDL) with reduced complexity, which is used for link level simulations. The parameters of the different scenarios within the CDL model are based on the expectation values of the generic model [IST07]. To assess the performance of the proposed graph-based receiver in various channel conditions, the scenarios considered in this thesis comprise an indoor office (A1 NLOS), a typical urban micro-cell (B1 NLOS), a suburban (C1 NLOS), and finally a typical urban macro-cell (C2 NLOS) channel model. The chosen scenarios differ significantly with respect to power delay profile and direction spread. All scenarios are applicable in the frequency range from 2 GHz to 6 GHz.

Taking the assumptions of the WINNER channel model into account, a continuous channel transfer function in space, frequency, and time, corresponding to (2.3), is given by

$$h\left(\Omega_{\mathrm{AoA}}, \Omega_{\mathrm{AoD}}, f, t\right) = \frac{1}{\sqrt{M_{\mathrm{c}}}} \sum_{c=1}^{M_{\mathrm{c}}} \sqrt{\frac{P_c}{M_{\mathrm{r}}}} \sum_{r=1}^{M_{\mathrm{r}}} \left\{ \exp\left(j\left(\Phi_{c,r} + 2\pi\left(f_{\mathrm{D,c,r}} t - \tau_c f + \Omega_{\mathrm{AoD}} + \Omega_{\mathrm{AoA}}\right)\right)\right) \right\}. \tag{2.34}$$

Hereby, M multipaths are separated into M_{c} clusters each with M_{r} rays. The number of clusters depend on the scenario and can as many as 20, whereas each cluster is composed of $M_{\mathrm{r}} = 20$ rays. A single link, as modeled by the WINNER channel, is illustrated in Figure 2.4. Furthermore, the WINNER system is based on the orthogonal frequency-division multiplexing (OFDM) scheme; see also Chapter 2.4 for further information on OFDM. Due to the orthogonality introduced by OFDM, intersymbol interference (ISI) and intercarrier interference (ICI) can be neglected. Hence, a discrete representation of the channel transfer function corresponding to (2.3) is given by

$$\begin{aligned} h_{n,m}[l,k] = \frac{1}{\sqrt{M_{\mathrm{c}}}} \sum_{c=1}^{M_{\mathrm{c}}} \sqrt{\frac{P_c}{M_{\mathrm{r}}}} \sum_{r=1}^{M_{\mathrm{r}}} \Bigg\{ & \exp\left(j\left(\Phi_{c,r} + 2\pi f_{\mathrm{D,c,r}} k T_{\mathrm{s}} - 2\pi \tau_c l F_{\mathrm{s}}\right)\right) + \\ & \exp\left(j\left(2\pi m \frac{d_{\mathrm{Tx}}}{\lambda} \sin(\varphi_{c,r}) + 2\pi n \frac{d_{\mathrm{Rx}}}{\lambda} \sin(\phi_{c,r})\right)\right) \Bigg\}. \end{aligned} \tag{2.35}$$

The channel coefficient $h_{n,m}[l,k]$ relates transmit (Tx) antenna m with receive (Rx) antenna n at OFDM subcarrier l and OFDM symbol k and is obtained by sampling $h(\Omega_{\mathrm{AoA}}, \Omega_{\mathrm{AoD}}, f, t)$ at frequency $f = lF_{\mathrm{s}}$, time $t = kT_{\mathrm{s}}$, and direction $\Omega_{\mathrm{AoD}} = m\frac{d_{\mathrm{Tx}}}{\lambda}$, $\Omega_{\mathrm{AoA}} = n\frac{d_{\mathrm{Rx}}}{\lambda}$, where T_{s} and F_{s} denote the OFDM symbol duration and subcarrier spacing, respectively. Moreover, a linear antenna array is assumed, with a spacing of d_{Tx} at the transmitter side and d_{Rx} at the receiver side with a wavelength λ. Each scat-

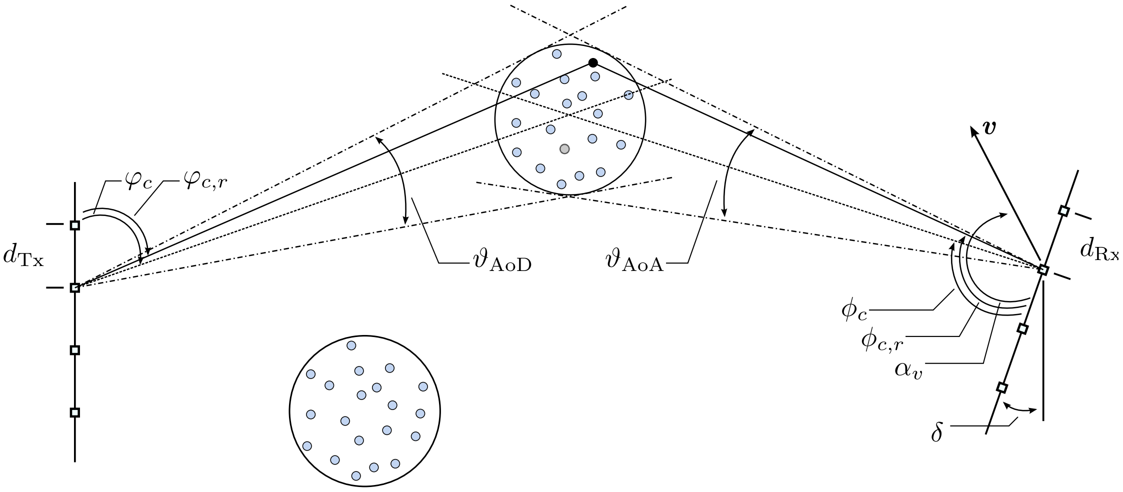


Figure 2.4: Single link of the WINNER channel models [IST07].

terer is characterized by an individual random-phase $\Phi_{c,r} \in [0, 2\pi)$, a propagation delay $\tau_c \in [0, \tau_{\max}]$, and a Doppler frequency $f_{\mathrm{D,c,r}} \in [-f_{\mathrm{D,max}}, f_{\mathrm{D,max}}]$. The maximum Doppler frequency depends on the angle of the velocity vector α_v and the angle of arrival $\phi_{c,r}$ as follows

$$f_{\mathrm{D,c,r}} = \frac{|v|}{\lambda} \cos\left(\phi_{c,r} - \alpha_v\right). \tag{2.36}$$

A multipath component departs with an angle $\varphi_{c,r}$ and is received with $\phi_{c,r}$. Characteristic parameters are tabulated for each scenario, among others, a mean angle of arrival (AoA) ϕ_c and a mean angle of departure (AoD) φ_c for each cluster c. In combination with the azimuth spread of departing angles ϑ_{AoD} and the azimuth spread of arriving angles ϑ_{AoA}, the ray specific angle of arrival and departure, respectively, are given by

$$\varphi_{c,r} = \varphi_c + \vartheta_{\mathrm{AoD}} \cdot \gamma_r, \tag{2.37}$$

$$\phi_{c,r} = \phi_c + \vartheta_{\mathrm{AoA}} \cdot \gamma_r, \tag{2.38}$$

where γ_r refers to the offset angles of each ray (cf. Table B.1). Furthermore, the composite angular spread at the transmitter and the receiver are defined as follows [Aue12]:

$$\Phi_{\mathrm{AoD}} = \left| \arg\max_{c,r} \left(\varphi_{c,r}\right) - \arg\min_{c,r} \left(\varphi_{c,r}\right) \right|, \tag{2.39}$$

$$\Phi_{\mathrm{AoA}} = \left| \arg\max_{c,r} \left(\phi_{c,r}\right) - \arg\min_{c,r} \left(\phi_{c,r}\right) \right|. \tag{2.40}$$

Additionally, the mean composite AoD and AoA, respectively, are given by

$$\Psi_{\mathrm{AoD}} = \frac{1}{2} \left(\arg\max_{c,r} \left(\varphi_{c,r}\right) + \arg\min_{c,r} \left(\varphi_{c,r}\right) \right), \tag{2.41}$$

$$\Psi_{\mathrm{AoA}} = \frac{1}{2} \left(\arg\max_{c,r} \left(\phi_{c,r}\right) + \arg\min_{c,r} \left(\phi_{c,r}\right) \right). \tag{2.42}$$

Table 2.1: Characteristic values for the WINNER channel scenarios

Channel scenario	Φ_{AoD}	Φ_{AoA}	ϑ_{AoD}	ϑ_{AoA}	$\tau_{\max}$	τ_{rms}	$\lvert v\rvert_{\max}$
A1 NLOS	248°	244°	5°	5°	175 ns	25.01 μs	5 km/h
B1 NLOS	113°	238°	12°	10°	615 ns	77.32 μs	70 km/h
C1 NLOS	43°	240°	2°	10°	750 ns	77.38 μs	120 km/h
C2 NLOS	39°	222°	2°	15°	1845 ns	245.05 μs	120 km/h

A detailed overview of the tabulated values for the scenarios under investigation are given in Appendix B. Characteristic parameters of each scenario are derived and the corresponding results are given in Table 2.1. As can be seen from Table 2.1, the scenarios differ significantly with respect to the maximum excess delay with the exception of B1 and C1. Although the scenarios B1 and C1 have a similar value for the maximum excess delay, the composite spread of the departing rays is much smaller for the C1 scenario, which results in a higher spatial correlation. Furthermore, with an exception for the A1 NLOS, a high spatial correlation is more likely to occur at the transmitter side, since the composite azimuth spread at the transmitter Φ_{AoD} is significantly smaller compared to the composite azimuth spread at the receiver Φ_{AoA}. The absolute value of the channel transfer function as a function of time and frequency is shown in Figure 2.5. As can be seen, a signal is exposed to fading since the multiple propagation paths are superimposed in a constructive and destructive fashion. This leads to the characteristic deep fades of the amplitude with an attenuation of close to 60 dB.

2.3 MIMO Wireless Communications

Research focused intensively on the achievable limits of multiple-input multiple-output (MIMO) systems, which was mainly sparked by the promising research results of [FG98, Tel99], predicting that MIMO systems are able to facilitate high data rate as well as high reliability communication links. Nowadays, MIMO systems are an integral core part of most wireless communications standards, such as LTE [3GP08], IEEE 802.16 (WiMAX) [IEE06], and IEEE 802.11n (WiFi-n) [IEE09]. Previously the pitfall of reliable communication, random multipath fading is now exploited to increase the capacity C, quality of service (QoS), and/or signal-to-noise ratio (SNR) ξ.

In general, a wireless communication link is characterized by three fundamental components: (1) noise, (2) multipath fading, and (3) interference. Nowadays, noise and multipath fading are no longer the limiting factors for advanced MIMO receivers, which are rather interference limited. Nevertheless, a rich scattering environment is a prerequisite for the majority of MIMO techniques which exploit the spatial diversity that is obtained by spatially separated antennas [SBM+04, Gol05].

Multiple antennas at the transmitter and receiver side can be used to increase the data rate through multiplexing or to improve the reliability of the system through diversity [BCC+07]. In order to increase the capacity, independent data streams are transmitted from each of the transmit antennas. This variant is known as *spatial multiplexing.*

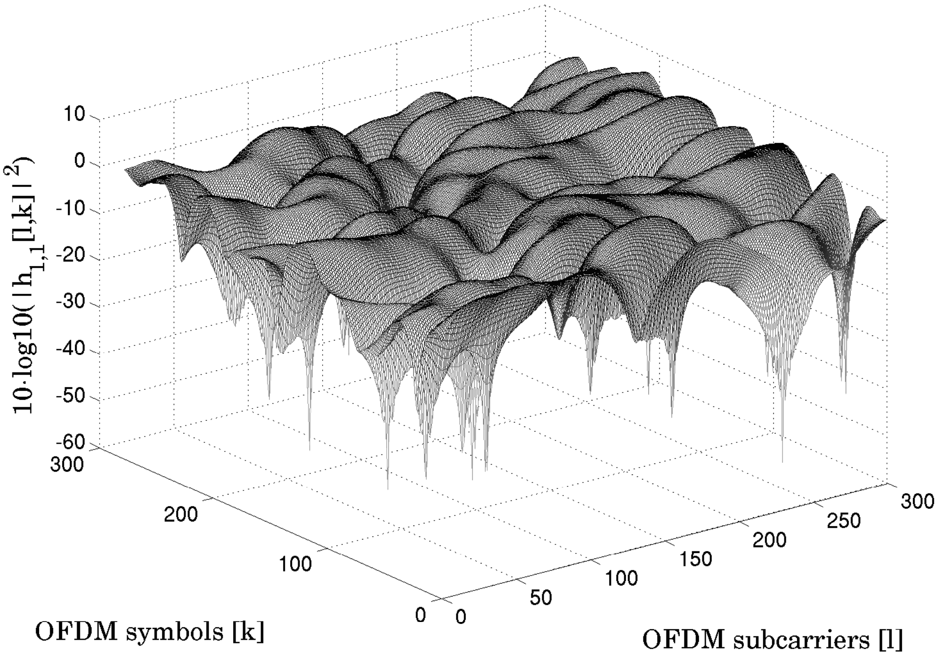


Figure 2.5: Magnitude of the channel transfer function in time and frequency domain with the WINNER C2 NLOS scenario.

Under the assumption that the transmit antennas are sufficiently separated in combination with a rich scattering environment, the transmitted signals will experience a unique spatial signature, which will allow the receiver to separate the different superimposed streams. Under optimum conditions, the capacity increases by $\min(N_\mathrm{T}, N_\mathrm{R})$, where N_T and N_R denote the number of transmit and receive antennas, respectively [Gol05, BCC$^+$07]. Moreover, most receiver implementations require that $N_\mathrm{R} \geq N_\mathrm{T}$ [Hoe13]. It is worth to mention that the channel estimation algorithms evaluated in Chapter 4 as well as the graph-based receiver proposed in Chapter 5 are not limited by this condition.

An alternative to spatial multiplexing, which increases the throughput of a user, is given by *spatial diversity*, which aims to increase the signal-to-noise ratio. Hereby, the same information is transmitted from all transmit antennas. The probability that all links between the transmit and the receive antennas are simultaneously in a deep fade is relatively low, i.e. it is the product of their individual fading probabilities, correspondingly the transmission of the signal is more reliable [Hoe13]. Spatial multiplexing and spatial diversity are the extreme cases of the diversity-multiplexing trade-off for MIMO channels. It has been shown, that a flexible transition between these two strategies is possible [ZT03].

Unlike conventional point-to-point communications, in a MIMO wireless system the throughput of a system is interference limited. Obviously, increasing the SNR by boosting the transmission power can not mitigate the effect that signals emanating from the transmit antennas interfere with each other. Consequently, advanced MIMO techniques have a strong impact on the interference and correspondingly on the reusability of a resource. They can be divided into four categories to either reject, avoid, coordinate or exploit interference [Ges12]. The various strategies can be applied for single as well as multi-user systems. This thesis, however, is restricted to the single-user case. An overview of MIMO techniques applicable to 4G mobile networks, such as LTE-A, is given in [BPG+09]. One of the more mature techniques is known as *beamforming*. A directed transmission is achieved by controlling the phase and amplitude of the signal emanated from each antenna element, such that constructive and destructive interference at desired angles is achieved. A receiver might adjust its beam pattern in order to achieve a high directivity towards the dominant angles of reception [MSL+09]. An SNR gain achieved by means of beamforming is often termed *array gain*. Beamforming is explained in more detail in Section 2.3.2.

2.3.1 Narrowband MIMO Model

For a MIMO system with N_T transmit and N_R receive antennas, the corresponding MIMO channel for a given OFDM subcarrier l and OFDM symbol k, respectively, can be represented by the channel matrix $\boldsymbol{H} \in \mathbb{C}^{N_\mathrm{R} \times N_\mathrm{T}}$:

$$\boldsymbol{H}[l,k] = \begin{bmatrix} h_{1,1}[l,k] & h_{1,2}[l,k] & \cdots & h_{1,N_\mathrm{T}}[l,k] \\ h_{2,1}[l,k] & h_{2,2}[l,k] & \cdots & h_{2,N_\mathrm{T}}[l,k] \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_\mathrm{R},1}[l,k] & h_{N_\mathrm{R},2}[l,k] & \cdots & h_{N_\mathrm{R},N_\mathrm{T}}[l,k] \end{bmatrix}, \tag{2.43}$$

where $h_{n,m}[l,k]$ may be generated as described in Section 2.2.

Given perfect knowledge of the channel matrix $\boldsymbol{H}$, the capacity of the MIMO system is given by [Tel99, BCC+07]:

$$C = \mathrm{E}\left\{\log\det\left(\mathbf{I}_{N_\mathrm{R}} + \frac{\xi}{N_\mathrm{T}}\boldsymbol{H}\boldsymbol{H}^\mathrm{H}\right)\right\}. \tag{2.44}$$

Assuming that the channel matrix $\boldsymbol{H}$ has full rank and the overall transmit power is kept constant, the capacity increases linearly with $\min(N_\mathrm{T}, N_\mathrm{R})$. The linear increase in capacity is only achieved, if both the transmitter and the receiver side increase the number of antennas simultaneously. If the number of antennas is bounded at one side, i.e. the number of receive antennas is fixed while the number of transmit antennas is increased to infinity, the capacity is bounded in N_T and converges to $N_\mathrm{R}\log(1+\xi)$. On the other hand, if the number of transmit antennas is fixed while increasing the number of receive antennas, the capacity increases with $\log(N_\mathrm{R})$. The main difference between the two latter approaches is that increasing the number of receive antennas results in an increase of received power. However, since the overall transmit power is fixed, the available power is shared between all available transmit antennas, hence, the overall power is not increased by increasing the number of transmit antennas [BCC+07]. Figure 2.6 shows the channel capacity as a function of the number of transmit and receive antennas. It can be seen, that the increase from one transmit and receive antenna to four yields a significant capacity gain. However, a further increase of the antennas on either the transmitter or receiver side follows the above mentioned bounds.

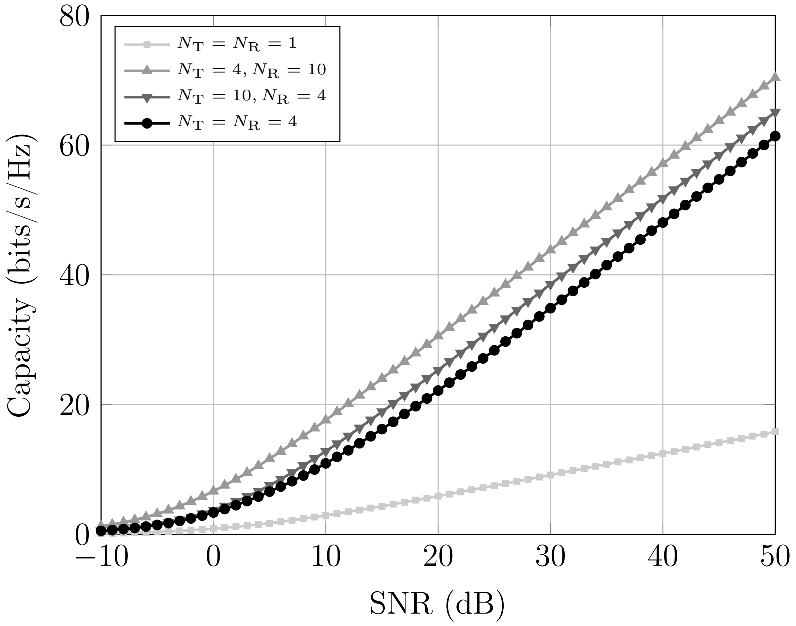


Figure 2.6: Channel capacity as a function of the number of transmit and receive antennas.

The potential capacity gain has lead to the widespread application of MIMO in wireless standards.

2.3.2 Codebook-based Beamforming

Beamforming is a MIMO spatial processing technology, which allows directed transmission and reception of signals. Hereby, the spatial domain is exploited to increase the signal-to-interference plus noise (SINR) ratio and/or increase the data rate. Although beamforming is typically applied to achieve a diversity and/or array gain, multiple streams can be processed in parallel, increasing the data rate for one user. Under ideal conditions, the maximum number of independent streams is equal to the number of antennas in the antenna array [Cox12]. Multiple antennas at the transmitter and/or the receiver side are mandatory in order to apply beamforming since the principle is based on the controlled superposition of the transmitted signals, thus, creating constructive and destructive interferences in the desired directions. However, beamforming requires the knowledge of channel state information at the transmitter (CSIT). Typically, this information needs to be transmitted from the receiver to the transmitter, as uplink and downlink channel are often not reciprocal, which renders the estimation of CSI by the transmitter infeasible [BCC$^+$07].

In order to reduce the amount of feedback information, a common codebook is maintained at both the transmitter and the receiver side, which comprises a finite set of beamforming vectors $\mathbf{w}^{(i)}$, termed codewords. The length of the codebook, i.e. the number of codewords, is a trade-off between the number of bits required to address a codeword and the suitability of the resulting beam pattern [STB09]. Typically, a codebook is designed to facilitate efficient codebook storage as well as codeword search [IH09]. An overview

Table 2.2: DFT-Codebook for two transmit antennas.

i	$\mathbf{w}^{(i)}$
1	$1/\sqrt{2}\,[+1,+1]^{\mathrm{T}}$
2	$1/\sqrt{2}\,[+1,-1]^{\mathrm{T}}$
3	$1/\sqrt{2}\,[+1,+\jmath]^{\mathrm{T}}$
4	$1/\sqrt{2}\,[+1,-\jmath]^{\mathrm{T}}$

Table 2.3: Grassmannian-Codebook for two transmit antennas.

i	$\mathbf{w}^{(i)}$
1	$[-0.1612-0.7348\jmath, -0.5135-0.4128\jmath]^{\mathrm{T}}$
2	$[-0.0787-0.3192\jmath, -0.2506+0.9106\jmath]^{\mathrm{T}}$
3	$[-0.2399+0.5985\jmath, -0.7641-0.0212\jmath]^{\mathrm{T}}$
4	$[-0.9541+0.0000\jmath, +0.2996+0.0000\jmath]^{\mathrm{T}}$

of available limited feedback precoding methods is provided in [LHL$^+$08]. Two widely applied codebooks are under investigation, namely the DFT and the Grassmannian codebook [LHS03]. The corresponding codebooks for two transmit antennas are given in Table 2.2 and Table 2.3, respectively. The possible antenna patterns which can be generated by the DFT and the Grassmannian codebook are exemplary visualized in Appendix C. Besides a simple codebook generation, the DFT codebook has several advantages and has been chosen for the use in LTE systems [STB09]. The appealing properties of the DFT codebook comprises [STB09]: (1) Constant modulus: As an be seen in Table 2.2, the beam-weights perform only a change in the phase without affecting the amplitude. This ensures a constant load for the power amplifiers. (2) Nested property: That means that lower rank codebooks represent a subset of higher rank codebooks. Thereby, the complexity of generating feedback is reduced for the mobile station. (3) Reduced complex multiplications: For two transmit antennas, the beam-weights consists of QPSK symbols only, hence, a sign operation replaces a complex multiplication, which results in a reduced complexity.

A different approach is pursued for the Grassmannian codebook. Here, the design of the beamforming codebooks can be interpreted as the discretization of the Grassmannian manifold [LHS03]. The entries of the Grassmannian codebook are generated such that the chordal distance

$$d_{\text{chord}}(\mathbf{w}^{(i)}, \mathbf{w}^{(j)}) \doteq \frac{1}{\sqrt{2}} \|\mathbf{w}^{(i)}\mathbf{w}^{(i)\mathrm{H}} - \mathbf{w}^{(j)}\mathbf{w}^{(j)\mathrm{H}}\|_F \tag{2.45}$$

is maximized, whereby $\|\cdot\|_F$ corresponds to the Frobenius norm. The resulting codebook entries differ maximally from each other [Hoe13].

Revisiting (2.44), the MIMO channel capacity taking codebook-based beamforming into account is given by

$$C(\mathbf{w}^{(i)}) = \mathrm{E}\left\{\log\det\left(\mathbf{I}_{N_{\mathrm{R}}} + \frac{\xi}{N_{\mathrm{T}}}\boldsymbol{H}\mathbf{w}^{(i)}\mathbf{w}^{(i)\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)\right\}. \tag{2.46}$$

The optimum codeword $\mathbf{w}_{\mathrm{opt}}^{(i)}$ is chosen such that it maximizes the given metric:

$$\mathbf{w}_{\mathrm{opt}}^{(i)} = \arg\max_{\mathbf{w}^{(i)}} \left\{ C(\mathbf{w}^{(i)}) \right\}. \tag{2.47}$$

In this thesis, the optimum codeword maximizes the channel capacity of a given channel matrix $\boldsymbol{H}$. Depending on the selectivity of the channel in time and/or frequency domain, the selected codewords need to be updated, which requires the calculation of the metric (2.47). Given a frequency-selective channel, each OFDM subcarrier may have an individual codeword in order to maximize the channel capacity. On the other hand, given a frequency-flat channel, the same codeword for all subcarriers is optimum. The update interval depends on the Doppler frequency and, thus, the selectivity in time domain. Chosen codewords need to be updated frequently, i.e. each OFDM symbol for a fast fading channel whereas the update interval can be extended for slow fading channels. Additionally, feedback delays deteriorate the performance as already outdated beamforming weights are applied. Hence, the application of codebook-based beamforming for fast fading channels is only limited. Section 5.6 highlights the impact of codebook-based beamforming on the performance of graph-based channel estimation as well as the selection of the codebook on the achievable bit error performance.

2.3.3 Large-Scale MIMO Implementations

A conclusion that may be drawn from the above mentioned fact—that the capacity is increasing linearly with $\min(N_{\mathrm{T}}, N_{\mathrm{R}})$—is, that a system just needs a sufficient number of antennas in order to fulfill arbitrary spectral efficiency demands. This idea has led to the investigation of large-scale MIMO implementations with number of transmit and receive antennas of the order of tens to hundreds, termed *large-MIMO* [VMCR08, MZCR09]. The challenges with large-MIMO are on the one hand, the lack of computationally efficient algorithms for data detection and channel estimation. An heuristic approach based on likelihood ascent search for data detection is proposed in [VMCR08]. The applicability of particle swarm optimization (PSO) for large-MIMO channel estimation is proposed and evaluated in [KNH11] and [KH12], respectively. The results are additionally presented in Chapter 4. On the other hand, the physical placement of hundreds of antennas while maintaining individual spatial signatures is especially difficult for applications with constrained dimensions, such as handheld devices. That is, not to mention, the limited power supply.

An alternative approach, which takes the dimension and power constraints into account, is given by *massive MIMO* [Mar10, RPL+13]. The base station, which is typically less restricted in terms of space and energy, is equipped with hundreds of antennas while the mobile stations are limited to a single antenna. This approach has several advantages, e.g. channel estimation for the uplink, i.e. the transmission from the mobile station to the base station, can be significantly simplified by using series expansion techniques [RPL+13]. In the downlink, due to the large number of antennas, beamforming is optimal [RPL+13].

However, a major problem of massive MIMO is pilot contamination due to interfering symbols. Typically, training sequences are designed to be orthogonal, however, the number of orthogonal training sequences is limited and, in case of a multi-cell environment, have to be shared across adjacent cells. Pilot contamination occurs when pilots are reused in neighboring cells. Although pilots may be arranged to interfere with data symbols, which would mitigate the interference as the power of pilots could be larger than that of data symbols, the probability of interfering pilots would remain high for a multi-cell environment [RPL+13]. A similar effect was studied in [KHAT11a] and described in Section 4.2.2, where orthogonal as well as random training sequences are superimposed. The application of random training sequences is particularly interesting since random sequences are not designed to be orthogonal and, thus, their separation is typically challenging. The channel estimation algorithm based on multi-objective particle swarm optimization (MOPSO) is able to separate even random training sequences and is a possible solution to mitigate the pilot contamination problem observed in massive MIMO.

2.4 Orthogonal Frequency-Division Multiplexing

Future wireless systems will have to deal with an increasing amount of traffic, which requires high spectral efficiency as well as flexibility of the air interface. Partly, these challenges are addressed with MIMO technologies (cf. Section 2.3), smaller cell sizes (e.g. femtocell networks [CA08]), and broadband transmission [SBM+04, SSO+07]. An increased bandwidth which exceeds the coherence bandwidth introduces intersymbol interference (ISI). Thus, due to multipath propagation, replicas of a transmitted symbol arrive with a delay at the receiver and will be superimposed with successive symbols. With increasing bandwidth the effect of ISI is increased, hence, more symbols will interfere with each other. The effect of ISI is twofold at the receiver side: ISI will degrade the performance when ignored and if taken into account, will significantly increase the complexity of the receiver [CS00].

Orthogonal frequency-division multiplexing (OFDM) is seen as a possible solution to facilitate high efficiency broadband transmission since, in conjunction with a cyclic prefix, it mitigates ISI, provides orthogonal multiple access within one cell, and enables flexible allocation of radio resources [SSO+07]. The main idea of OFDM is to partition the wide bandwidth channel into multiple orthogonal subchannels, each with a smaller bandwidth such that the reduced bandwidth is smaller than the coherence bandwidth of the channel. As a result, every subcarrier is invariant in the frequency domain. Since the subchannels are orthogonal, detecting the transmitted signal is comparably easier.

Figure 2.7 illustrates a block diagram of a discrete-time baseband OFDM system. At the transmitter side on the left, L data symbols are mapped on parallel subcarriers. Typically, the data symbols comprise multiple information bits, i.e. they are modulated by an arbitrary digital modulation format, e.g. quadrature amplitude modulation (QAM). Each subcarrier can be modulated individually, known as *bit loading* [Hoe13]. In order to modulate the L data symbols, given by $\boldsymbol{x} = (x[0], \dots, x[L-1])^{\mathrm{T}}$, on the OFDM subcarriers, an inverse discrete Fourier transform (IDFT) is employed according to

$$s[\kappa] = \frac{1}{\sqrt{L}} \sum_{l=0}^{L-1} x[l] \exp\left(j2\pi l\kappa/L\right), \qquad \kappa = 0, \dots, L-1 \tag{2.48}$$

where $1/\sqrt{L}$ is a scaling factor. After a parallel-to-serial (P/S) conversion, the time domain signal yields [Pro00, Hoe13]

$$s(t) = \frac{1}{\sqrt{L}} \sum_{l=0}^{L-1} x[l] \exp\left(j2\pi lt/T_{\mathrm{s}}\right), \qquad 0 \le t \le T_{\mathrm{s}} \tag{2.49}$$

where T_{s} is the OFDM symbol duration. Sampling the continuous time signal at times $t = \kappa T_{\mathrm{s}}/L$ where $\kappa = 0, \dots, L-1$, results in the discrete time signal given by (2.48). After the addition of a cyclic prefix (CP), the signal is convolved with the channel transfer function $\boldsymbol{g} = \mathrm{IDFT}(\boldsymbol{h})$, where $\boldsymbol{h}$ is the channel transfer function of length L. A white Gaussian noise term $\boldsymbol{n}$ is added. At the receiver side, the cyclic prefix (CP^{-1}) is removed and the stream is converted from serial to parallel (S/P). A subsequent discrete Fourier transformation (DFT) demodulates the received sequence to the frequency domain, represented by the receive vector $\boldsymbol{y} = (y[0], \dots, y[L-1])^{\mathrm{T}}$.

The orthogonality of an OFDM signal is apparent in the frequency domain, as can be seen in Figure 2.8. At the maximum amplitude of each OFDM subcarrier, all remaining subcarriers have an amplitude of zero, thus, they do not interfere. The subcarriers are separated by a subcarrier spacing $F_{\mathrm{s}} = 1/T_{\mathrm{s}}$, whereas the subcarrier spacing F_{s} and, inherently, the OFDM symbol duration T_{s} are design parameters, which are chosen under the following constraints [Cox12]:

$$F_{\mathrm{s}} \gg f_{\mathrm{D}}, \tag{2.50}$$

$$T_{\mathrm{s}} \gg \tau_{\max}. \tag{2.51}$$

The first constraint is introduced by the Doppler frequency f_{D}. Similarly to rays in a multipath environment, which are shifted to higher or lower frequencies due to the Doppler frequency (cf. Section 2.1.1), OFDM subcarriers are equally shifted to higher or lower frequencies. While a common Doppler shift corresponds to a frequency offset and, thus, may be corrected [Moo94, RK99], it has a more severe impact in rich scattering environments, as different subcarrier frequencies experience different Doppler shifts. Accordingly,

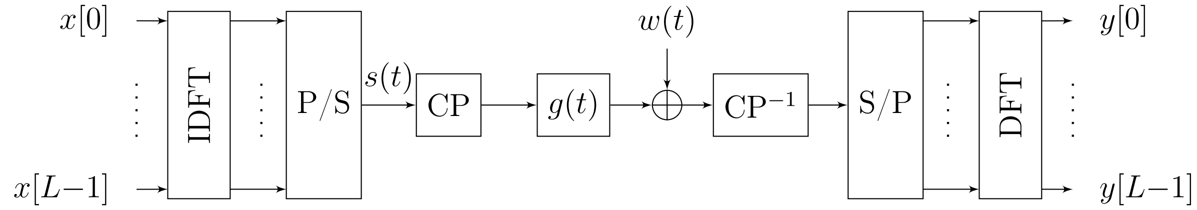


Figure 2.7: Baseband OFDM system model.

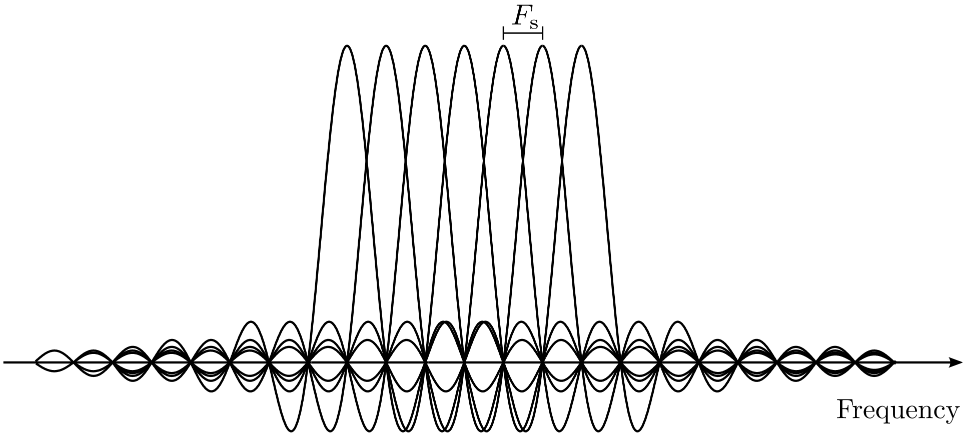


Figure 2.8: OFDM signal representation in frequency domain.

at the maximum response of an OFDM subcarrier, adjacent OFDM subcarriers will now interfere with each other. An effect which is commonly referred to as intercarrier interference (ICI). The effect is more pronounced with increasing velocity [RK99]. Given a sufficiently long subcarrier spacing, which fulfills (2.50), the amount of interference is still acceptable [Cox12].

The second constraint, given by (2.51), aims to minimize the impact of intersymbol interference (ISI), i.e. due to multipath propagation, successive OFDM symbols may interfere with each other. A common solution to avoid ISI is to introduce a guard interval, often realized by a so called cyclic prefix. Hereby, L_CP samples are taken from the end of each OFDM symbol $(x[L-L_\mathrm{CP}], x[L-L_\mathrm{CP}+1], \dots, x[L-1])$ and are appended to the beginning of each block. The duration of the CP is designed such that it incorporates the maximum delay τ_max. Accordingly, the total OFDM symbol duration is extended to $T_\mathrm{tot} = T_\mathrm{s} + T_\mathrm{CP}$. Since the CP is removed at the receiver, the delayed part of the previous OFDM symbol is no longer interfering. By selecting T_s according to (2.51), the rate loss due to the CP is negligible [Pro00, Cox12].

An alternative system-theoretical interpretation for the CP is that it converts the linear convolution $(*)$ of the channel transfer function $\boldsymbol{g}$ and the transmitted signal $\boldsymbol{x}$ into a circular convolution:

$$\boldsymbol{y} = \mathrm{DFT}\left(\mathrm{IDFT}\left(\boldsymbol{x}\right) \circledast \frac{\boldsymbol{g}}{\sqrt{L}} + \boldsymbol{n}\right) = \boldsymbol{x} \odot \boldsymbol{h} + \boldsymbol{n}, \tag{2.52}$$

where $\circledast$ and $\odot$ denote the circular convolution and element-wise multiplication, respectively. Due to the convolution theorem, the transmitted vector $\boldsymbol{x}$ is simply multiplied with the channel transfer function $\boldsymbol{h}$. Accordingly, every OFDM subcarrier has a single individual scalar weighting factor. A compensation of this factor is, thus, significantly simplified [Hoe13].

3

State-of-the-Art Receivers for Wireless Systems

IN HIS 1948 LANDMARK paper [Sha48], Shannon quantified the maximum achievable capacity of communication channels. Since then, several key technologies have been devised to approach the capacity. In the area of wireless communications, MIMO and OFDM, as well as corresponding techniques, are recognized as one of the most significant breakthroughs in modern communications [HEHA11], cf. Chapter 2. Despite the promising goals and achievements, real-world applications still have a considerable gap to the theoretical bound [HHI$^+$12]. This is partly reasoned in the rising computational complexity of optimum detectors, absence of perfect channel knowledge as well as required signaling overhead for practical implementations. This chapter aims to give an overview on current state-of-the-art receivers comprising data detection and channel estimation algorithms. Moreover, a novel low-complexity MIMO detector is introduced.

It is well known that the capacity can only be reached in combination with an outer channel code. However, the optimum joint detector and decoder is computationally complex and thus, infeasible. An efficient approximate solution is given by the so-called "turbo principle", first proposed by Douillard et al. in [DJB$^+$95]. Thereby, reliability information [HOP96] is exchanged iteratively between a detector and a channel decoder based on, for example, the maximum a posteriori probability decoder [BCJR74]. Section 3.1 introduces MIMO detectors suitable for turbo processing. A trade-off between the achievable performance and the required computational complexity is identified. Besides the a posteriori probability detector, which yields the optimum performance at a high complexity, several sub-optimum detectors are presented. The so-called Gaussian detector offers good BER performance under certain conditions while maintaining a very low complexity. Additionally, a novel suboptimal approach has been devised within this thesis: The Gaussian tree search detector outperforms the classical Gaussian detector and simultaneously offers a flexible trade-off between performance and complexity.

An overview of MIMO-OFDM channel estimation techniques is given in Section 3.2. Since pilot-based techniques are the common method to acquire channel state information [Aue12], Section 3.3 focuses on the design of pilot grids, i.e. which pilot grid yields the lowest mean squared error with the least amount of pilots.

While the focus of the previous sections has been on the presentation of optimum and sub-optimum data detection and channel estimation methods, Section 3.4 presents general methods for the combining of soft messages. The exchange of soft messages and accordingly their combination has been identified as a key tool to facilitate low complexity data detection and channel estimation. Unfortunately, with the use of suboptimal algorithms, several ideal assumptions are not fulfilled and the achievable performance is deteriorated. Section 3.4 examines how correlated reliability information can be combined optimally. By considering the correlation during the combining step, a more accurate result is obtained.

3.1 Iterative MIMO Detection

The MIMO-OFDM system under consideration consists of N_T transmit and N_R receive antennas, and can be represented after OFDM demodulation as

$$\boldsymbol{y}[l,k] = \boldsymbol{H}[l,k]\boldsymbol{x}[l,k] + \boldsymbol{n}[l,k], \tag{3.1}$$

where $l \in 0,1,\ldots,L-1$ and $k \in 0,1,\ldots,K-1$ represent the OFDM subcarrier and OFDM symbol index, respectively. The received symbol vector is given by $\boldsymbol{y}[l,k] \in \mathbb{C}^{N_\mathrm{R}\times 1}$. Additionally, the channel matrix $\boldsymbol{H}[l,k] \in \mathbb{C}^{N_\mathrm{R}\times N_\mathrm{T}}$ contains the channel coefficients, which are assumed to be WSS, complex Gaussian variables with zero mean. Moreover, the channel coefficients are assumed to be known at the receiver side in this section. Typically, channel state information (CSI) is obtained beforehand by means of channel estimation. A transmitted signal vector is denoted by $\boldsymbol{x}[l,k] \in \mathbb{C}^{N_\mathrm{T}\times 1}$, whereas each element of $\boldsymbol{x}[l,k]$ belongs to a finite constellation set $\mathcal{S}$ of size $|\mathcal{S}|$. More specifically, x consists of a set of $N_\mathrm{b} \doteq \log_2(|\mathcal{S}|)$ coded bits $\boldsymbol{c}' = [c'_1,\ldots,c'_{N_\mathrm{b}}]$, which are mapped to a symbol. The code bit vector $\boldsymbol{c}$ aggregates all elements of the transmit vector $\boldsymbol{x}$, i.e. $\mathbf{c} = [c_1,\ldots,c_i,\ldots,c_{N_\mathrm{T}N_\mathrm{b}}]^\mathrm{T}$. Furthermore, N_b and $N_\mathrm{m} \doteq 2^{N_\mathrm{b}} = |\mathcal{S}|$ refer to the number of bits per symbol and the cardinality of the constellation set, respectively. Finally, $\boldsymbol{n}[l,k] \in \mathbb{C}^{N_\mathrm{R}\times 1}$ is the zero-mean complex Gaussian noise vector with variance σ_n^2 per term. Without loss of generality, the time and frequency index l and k are omitted in the following to improve readability.

It can be seen from (3.1) that the receiver observes a linear superposition of N_T individually transmitted data symbols, which results in an effective constellation set $\mathcal{X}$ of size $|\mathcal{X}| = 2^{N_\mathrm{T}\cdot N_\mathrm{b}}$. The receivers task is to separate the transmitted symbols. The symbol-wise maximum-likelihood (ML) detector tests all possible hypotheses and selects the transmitted symbols which minimize the Euclidean distance according to

$$\hat{\boldsymbol{x}}^\mathrm{ML} = \arg\min_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|. \tag{3.2}$$

Only for equi-probable symbols, the ML detector yields the optimum solution with respect to (w.r.t.) the minimum symbol error probability [Pro00].

In a practical system, the transmitted sequence is typically encoded by means of a channel code. For an optimum detection, the constraints of the channel code have to be considered jointly with the complete received sequence [Ht03].

Since joint detection and decoding is often computationally too complex, the detection and decoding steps are executed separately, whereas both components exchange reliability (soft) information. This separation of detection and decoding by means of iterative processing, based on the so-called "turbo principle" is, strictly speaking, not optimal, but has shown to be very effective and computationally efficient [Ht03]. However, it requires soft-input soft-output (SiSo) detection algorithms, i.e. reliability information is utilized in form of a priori information at the input and refined reliabilities are subsequently given to the output. Unlike the ML detector, the maximum a posteriori (MAP) detector is able to take a priori probabilities of transmitted symbols (different to equal probability) into account and minimizes the probability of a symbol error [Pro00, Hoe13]:

$$\hat{\boldsymbol{x}}^{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}\in\mathcal{X}} p\left(\boldsymbol{x}|\boldsymbol{y}\right), \tag{3.3}$$

where $p(\boldsymbol{x}|\boldsymbol{y})$ is the conditional probability density function (PDF) of the candidate solution $\boldsymbol{x}$, given the observed vector $\boldsymbol{y}$. In case the maximization in (3.3) is omitted, i.e. the decoder feeds back $p(\boldsymbol{x}|\boldsymbol{y})$ for all possible hypotheses, the corresponding detector is termed a posteriori probability (APP) detector [Hoe13]. The APP is typically given in form of an a posteriori log-likelihood ratio (LLR) defined as follows:

$$\begin{aligned}
\mathrm{L}_{\mathrm{P}}(c_i|\boldsymbol{y}) &\doteq \ln \frac{P\left(c_i=+1|\boldsymbol{y}\right)}{P\left(c_i=-1|\boldsymbol{y}\right)} \qquad (3.4)\\
&= \ln \frac{\sum_{\boldsymbol{x}\in\mathcal{X}^i_{+1}} p(\boldsymbol{y}|\boldsymbol{x})\cdot P(\boldsymbol{x})}{\sum_{\boldsymbol{x}\in\mathcal{X}^i_{-1}} p(\boldsymbol{y}|\boldsymbol{x})\cdot P(\boldsymbol{x})} \qquad (3.5)\\
&= \underbrace{\ln \frac{P(c_i=1)}{P(c_i=-1)}}_{\mathrm{L}_{\mathrm{A}}(c_i)} + \underbrace{\ln \frac{\sum_{\boldsymbol{x}\in\mathcal{X}^i_{+1}} p(\boldsymbol{y}|\boldsymbol{x})\cdot \exp\left(\frac{1}{2}\boldsymbol{c}^{\mathrm{T}}_{\sim i}\mathbf{L}_{\mathrm{A},\sim\mathbf{i}}\right)}{\sum_{\boldsymbol{x}\in\mathcal{X}^i_{-1}} p(\boldsymbol{y}|\boldsymbol{x})\cdot \exp\left(\frac{1}{2}\boldsymbol{c}^{\mathrm{T}}_{\sim i}\mathbf{L}_{\mathrm{A},\sim\mathbf{i}}\right)},}_{\mathrm{L}_{\mathrm{E}}(c_i|\boldsymbol{y})} \qquad (3.6)
\end{aligned}$$

where $\mathcal{X}^i_{+1}$ denotes the subset of symbols $\boldsymbol{x}\in\mathcal{X}$ whose bit patterns have the value $+1$ at the position i, that is

$$\mathcal{X}^i_{+1} = \{\boldsymbol{x}|c_i=+1\}, \qquad \mathcal{X}^i_{-1} = \{\boldsymbol{x}|c_i=-1\}. \tag{3.7}$$

Moreover, $\boldsymbol{c}^{\mathrm{T}}_{\sim i}$ denotes the column vector of $\mathbf{c}$ omitting its ith element c_i, and $\mathbf{L}_{\mathrm{A},\sim\mathbf{i}}$ is the vector of a priori LLR values omitting the ith element as well [Ht03]. As can be seen from (3.6), the a posteriori LLR is separated into two components, L_{A} and L_{E}, referring to the a priori LLR and extrinsic LLR, respectively. An essential part of calculating the a posteriori LLR lies in the calculation of the conditional probability

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\left(2\pi\sigma_n^2\right)^{N_{\mathrm{R}}/2}}\cdot\exp\left(-\frac{1}{2\sigma_n^2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2\right). \tag{3.8}$$

It is of importance to mention that (3.6) as well as (3.8) assume statistically independent code bits and inherently symbols. This assumption is typically justified with the use of a sufficiently long interleaver, which distributes the bits such that they can be assumed to be uncorrelated. While this assumption holds for an optimum detector, correlation might be inadvertently introduced nevertheless due to the use of a suboptimal detection algorithm and/or a factor graph structure with cycles [PRV96, KHA12]. Section 3.4 describes in detail how to handle correlated random measures.

One of the main problems of solving (3.4) is, that the search set $\mathcal{X}$ grows exponentially with the number of transmit antennas and the number of bits per symbol. In order to reduce the computational complexity, the so-called max-log approximation [RHV97]

$$\ln\left(\exp(\delta_1) + \ldots + \exp(\delta_n)\right) \approx \max_{i \in \{1,\ldots,n\}} \delta_i \tag{3.9}$$

is commonly applied for the calculation of (3.4). Hereby, the logarithm of the sum of exponential terms is approximated by its largest term. Inserting (3.8) and

$$P(\boldsymbol{x}) = \prod_{i=1}^{N_{\mathrm{T}} \cdot N_{\mathrm{b}}} \frac{\exp(c_i \cdot \mathrm{L}_{\mathrm{A}}(c_i)/2)}{\exp(-\mathrm{L}_{\mathrm{A}}(c_i)/2) + \exp(\mathrm{L}_{\mathrm{A}}(c_i)/2)} \tag{3.10}$$

into (3.5) [Hoe13] and by applying the max-log approximation, the calculation of the a posteriori LLR simplifies to

$$\mathrm{L}_{\mathrm{P}}(c_i|\boldsymbol{y}) \approx \max_{\boldsymbol{x} \in \mathcal{X}^i_{+1}} \left\{\Lambda^{\mathrm{APP}}(\boldsymbol{x})\right\} - \max_{\boldsymbol{x} \in \mathcal{X}^i_{-1}} \left\{\Lambda^{\mathrm{APP}}(\boldsymbol{x})\right\}, \tag{3.11}$$

with

$$\Lambda^{\mathrm{APP}}(\boldsymbol{x}) = -\frac{1}{2\sigma_n^2}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \mathbf{c}^{\mathrm{T}} \cdot \frac{1}{2}\mathbf{L}_{\mathrm{A}}. \tag{3.12}$$

Nevertheless, the max-log approximation is not able to reduce the computational complexity significantly, since the size of the search set $\mathcal{X}$ is not reduced. A selection of sub-optimum approaches is presented in the following, which are able to solve (3.11) efficiently and obtain a close-to-optimum performance.

3.1.1 QR-Based Detection

A large variety of detection algorithms use a QR-factorization as a precondition to reduce the computational complexity of calculating (3.11) [MHC$^+$05, VB99, AEVZ02, KYIG05]. Hereby, the channel matrix $\boldsymbol{H}$ is factorized according to $\boldsymbol{H} = \mathbf{QR}$, where $\mathbf{Q} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ is orthogonal ($\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}$) and $\mathbf{R}$ is upper triangular. By applying the QR-factorization, the vector norm in (3.12) can be rewritten as [Lar09]

$$\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 = \|\mathbf{Q}^{\mathrm{T}}\boldsymbol{y} - \mathbf{R}\boldsymbol{x}\|^2 + \|(\mathbf{I} - \mathbf{Q}\mathbf{Q}^{\mathrm{T}})\boldsymbol{y}\|^2, \tag{3.13}$$

where the last term does not depend on $\boldsymbol{x}$ and, thus, can be neglected in the maximization. The corresponding metric is given by

$$\begin{aligned}\Lambda^{\mathrm{QR}}(\boldsymbol{x}) &= \frac{1}{2\sigma_n^2}\|\tilde{\mathbf{y}} - \mathbf{R}\boldsymbol{x}\|^2 - \mathbf{c}^{\mathrm{T}} \cdot \frac{1}{2}\mathbf{L}_{\mathrm{A}} \\ &= \frac{1}{2\sigma_n^2}\sum_{i=1}^{N_{\mathrm{T}}}\Big|\tilde{y}_i - \sum_{j=i}^{N_{\mathrm{T}}} r_{ij}x_j\Big|^2 - \mathbf{c}^{\mathrm{T}} \cdot \frac{1}{2}\mathbf{L}_{\mathrm{A}} \qquad (3.14)\end{aligned}$$

with $\tilde{\mathbf{y}} = \mathbf{Q}^{\mathrm{T}}\boldsymbol{y}$. The complexity of the QR decomposition is of the order $\mathcal{O}(2/3 \cdot N_{\mathrm{T}}^3)$ for a channel matrix with full rank and $N_{\mathrm{T}} \approx N_{\mathrm{R}}$. The complexity approaches $\mathcal{O}(N_{\mathrm{R}} \cdot N_{\mathrm{T}}^2)$ for a MIMO setup with $N_{\mathrm{R}} \gg N_{\mathrm{T}}$ [Dah08]. Unfortunately, the performance degrades catastrophically when $N_{\mathrm{T}} > N_{\mathrm{R}}$ [JH07], since the multi-antenna interference is not longer completely resolved within the triangular matrix $\mathbf{R}$.

Due to the upper triangular structure of $\mathbf{R}$, the metric Λ^{QR} can be visualized as a decision tree with $(N_{\mathrm{T}} + 1)$ layers. Exemplary for the case of binary phase-shift keying (BPSK) ($N_{\mathrm{b}} = 1$) and four transmit antennas ($N_{\mathrm{T}} = 4$), the resulting decision tree is shown in Figure 3.1. At each layer, $|\mathcal{S}|$ branches are emanating from each node, resulting in $|\mathcal{S}|^{N_{\mathrm{T}}} = |\mathcal{X}|$ leaf nodes at the bottom of the tree. At each node, the partial path metric

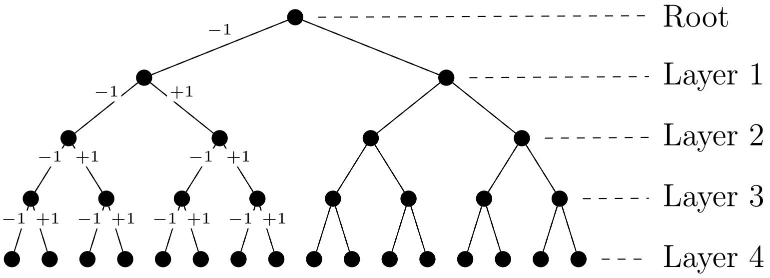


Figure 3.1: Illustration of the MIMO detection problem as a decision tree.

$$\lambda_{\mathrm{par}}^{\mathrm{QR}}(\boldsymbol{x}_m) = \frac{1}{2\sigma_n^2}\sum_{i=m}^{N_{\mathrm{T}}}\Big|\tilde{y}_i - \sum_{j=i}^{N_{\mathrm{T}}} r_{ij}x_j\Big|^2 - \frac{1}{2}\sum_{i=1}^{m\cdot N_{\mathrm{b}}} c_i \mathrm{L}_{\mathrm{A}}(c_i) \qquad (3.15)$$

is evaluated, where $\boldsymbol{x}_m = [x_m, \ldots, x_{N_{\mathrm{T}}}]^{\mathrm{T}}$ with $m = N_{\mathrm{T}}, N_{\mathrm{T}} - 1, \ldots, 2, 1$. Because of the upper triangular structure, the index m is reversed in its order. The complete path from the root node to a leaf node represents a hypothesis $\tilde{\boldsymbol{x}} = (x_1, \ldots, x_{N_{\mathrm{T}}})$. Evaluating all $|\mathcal{X}|$ hypotheses results again in an exponential complexity and, hence, there are several approaches to reduce the number of leaves that are evaluated.

Example 3.1 The following example illustrates the principle of the QR decomposition for a noiseless 3×3 MIMO system. After factorizing $\boldsymbol{H}$, (3.13) is given by

$$\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\\ \tilde{y}_3\end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13}\\ 0 & r_{22} & r_{23}\\ 0 & 0 & r_{33}\end{bmatrix} \cdot \begin{bmatrix} x_1\\ x_2\\ x_3\end{bmatrix}. \qquad (3.16)$$

The algorithm starts at the bottom, solving $x_3 = \tilde{y}_3/r_{33}$ and uses this result subsequently to solve the second equation. By successively solving the given equations from bottom to top, the estimates of the unknown transmitted symbols $\boldsymbol{x}$ are obtained [Dah08]:

$$\hat{x}_m = \left(\tilde{y}_m - \sum_{k=m+1}^{N_\mathrm{T}} r_{m,k}\hat{x}_k \right) / r_{m,m}, \tag{3.17}$$

with $m = N_\mathrm{T}, N_\mathrm{T} - 1, \ldots, 2, 1$.

The so-called *sphere decoder* [VB99, AEVZ02] is a prominent solution to reduce the overall complexity of the MIMO detection problem. The algorithm follows all paths which have a cumulative metric smaller than R. Thus, all paths with $\lambda_\mathrm{par}^\mathrm{QR}(\boldsymbol{x}_m) > R$ are excluded from the tree search [Lar09]. The performance as well as complexity of the sphere decoder depends on the chosen sphere radius R. A trade-off between performance and complexity has to be found, i.e. choosing a small value for R improves the complexity but results in a higher bit error probability and vice versa, choosing a large value results in a high complexity with close-to-optimum performance. It has been shown in [JO05], that although the sphere decoder can be efficient under certain conditions, the complexity is still increasing exponentially.

An alternative approach, termed *QRD-M detection*, was proposed in [KYIG05], which offers a fixed computational complexity depending on the parameter M_sl. The main difference to the sphere decoder lies in the way with which paths within the tree are pruned, i.e. removed from the decision tree. The QRD-M detector selects the M_sl most significant leaves based on the path metric given in (3.15). Thus, at most $M_\mathrm{sl}{\cdot}N_\mathrm{m}$ nodes have to be evaluated at each layer. The computational complexity of a metric calculation with QRD-M detection is of the order $\mathcal{O}(1/2 \cdot N_\mathrm{T}^2)$ [Dah08]. Overall, the number of visited nodes is given by

$$M_\mathrm{vn} = \sum_{i=0}^{N_\mathrm{T}-1} \min\left(M_\mathrm{sl}, (N_\mathrm{m})^i\right) \cdot N_\mathrm{m}. \tag{3.18}$$

The exemplary decision tree with pruned nodes is shown in Figure 3.2. The dashed lines

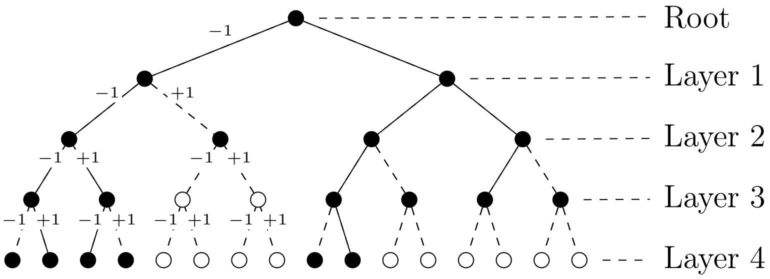


Figure 3.2: Illustration of pruned MIMO decision tree search with $M_\mathrm{sl} = 3$.

indicate that the remaining parts of the branches are pruned. A partial or full metric calculation needs to be done within the solid black nodes/leaves, respectively. Due to pruning, the costs of the overall metric calculations can be reduced, here visualized by the white nodes. Obviously, the potential to reduce the computational cost is largest when pruning is applied for a tree with higher-order modulations and/or a large number of transmit antennas.

The QR decomposition enables the representation of the detection problem as a tree-search; a prerequisite for a large variety of low complexity MIMO detection algorithms. However, it requires that the number of receive antennas is equal to or larger than the number of transmit antennas, i.e. $N_\mathrm{R} \geq N_\mathrm{T}$. In scenarios where $N_\mathrm{T} > N_\mathrm{R}$, which is often referred to as *rank-deficient* scenario, the performance is significantly deteriorated [JH07]. In light of the proposed graph-based receiver concept, the QR decomposition is not suitable (cf. Section 5.1.6 for a more detailed explanation). Therefore, in the subsequent sections two alternative low-complexity approaches for MIMO detection are presented.

3.1.2 Gaussian Detection

An aggressive approach to reduce the computational complexity is given by the so-called elementary signal estimator (ESE) initially proposed for interleave-division multiple access (IDMA) in [PLL03]. Applied to the case of MIMO detection, the estimator approximates the interference for the mth transmit antenna of $(N_\mathrm{T} - 1)$ antennas at the nth receive antenna by a Gaussian random variable

$$\zeta_{n,m} \sim \mathcal{CN}\left\{\mu_{\zeta_{n,m}}, \sigma^2_{\zeta_{n,m}}\right\}, \tag{3.19}$$

with mean and variance

$$\mu_{\zeta_{n,m}} = \sum_{\substack{i=1 \\ i\neq m}}^{N_\mathrm{T}} h_{n,i} \cdot \mu_{x_i}, \qquad \sigma^2_{\zeta_{n,m}} = \sum_{\substack{i=1 \\ i\neq m}}^{N_\mathrm{T}} |h_{n,i}|^2 \cdot \sigma^2_{x_i} + \sigma^2_n. \tag{3.20}$$

The mean and variance of a symbol x is defined as $\mu_x = \mathrm{E}\{x\}$ and $\sigma^2_x = \mathrm{Var}\{x\}$, respectively. The Gaussian approximation of the interfering terms is motivated by the central limit theorem, which states that the superposition of a sufficiently large number of independent and identically distributed (i.i.d.) random variables results in an approximate Gaussian distribution [Hoe13]. Despite the constraint of "sufficiently large" number of superimposed variables—which in this case depends on the number of transmit antennas—the Gaussian approximation is also applied for MIMO systems with only a few number of transmit antennas. As the numerical results provided in Section 3.1.4 will show, the Gaussian approximation is sufficiently accurate in combination with lower-order modulation and few transmit antennas. The corresponding detector is termed "Gaussian detector" (GA) in the following. By applying the Gaussian approximation for the nth receive and mth transmit antenna, (3.1) can be rewritten according to

$$\begin{aligned} y_n &= h_{n,m}x_m + \sum_{\substack{i=1 \\ i \neq m}}^{N_\mathrm{T}} h_{n,i}x_i + n_n \\ &\approx h_{n,m}x_m + \zeta_{n,m}. \end{aligned} \tag{3.21}$$

Effectively, the Gaussian detector ignores a large fraction of the original constellation set $\mathcal{X}$ and focuses instead on the constellation set $\mathcal{S}$ of a single desired antenna with size N_m. The calculation of the a posteriori LLR with the Gaussian detector is given by

$$\mathrm{L_P}(c_n|\boldsymbol{y}) \approx \max_{x_m \in \mathcal{S}^i_{+1}} \left\{ \Lambda^{\mathrm{GA}}(x_m) \right\} - \max_{x_m \in \mathcal{S}^i_{-1}} \left\{ \Lambda^{\mathrm{GA}}(x_m) \right\}, \tag{3.22}$$

with

$$\Lambda^{\mathrm{GA}}(x_m) = -\sum_{n=1}^{N_\mathrm{R}} \frac{\left| y_n - h_{n,m}x_m - \mu_{\zeta_{n,m}} \right|^2}{2(\sigma_n^2 + \sigma^2_{\zeta_{n,m}})} + \frac{1}{2}\sum_{m=1}^{N_\mathrm{b}} c'_m \mathrm{L_A}(c'_m). \tag{3.23}$$

The complexity of the Gaussian detection is significantly reduced compared to the APP detector, at the cost of a decreased performance, especially with few transmit antennas and higher-order modulation (cf. Section 3.1.4).

3.1.3 Gaussian Tree Search Detection

A combined approach of the Gaussian approximation and tree-search based detection methods is given by the so-called *Gaussian tree search (GTS) detector*. First presented for a single-antenna system and superposition modulation (SM) in [HH12], it is applied here for MIMO detection with QAM modulation. The central idea of Gaussian tree search is to facilitate MIMO detection within a decision tree without the use of a QR decomposition.

The challenge of the Gaussian tree search is to calculate the metric of an incomplete path whose symbol pattern is only partially known. Unlike QR decomposition, where the unknown symbols are successively determined, the ambiguities introduced by the channel matrix $\boldsymbol{H}$ are still present and, hence, for each received observation y, N_T variables exist. By applying the Gaussian approximation for the calculation of the metric of a partial path, the unknown layers are represented by a Gaussian random variable. That means, at the mth layer, $(N_\mathrm{T} - m)$ interfering terms are summarized by a corresponding Gaussian variable. Thus, the metric at the mth layer is defined according to

$$\Lambda^{\mathrm{GTS}}_{\mathrm{par}}(\boldsymbol{x}_m) \approx -\frac{1}{2}\sum_{n=1}^{N_\mathrm{R}} \frac{\left| y_n - \sum_{j=1}^{m} h_{n,j}x_j - \mu_{\zeta_{n,m}} \right|^2}{\sigma_n^2 + \sigma^2_{\zeta_{n,m}}} + \frac{1}{2}\sum_{n=1}^{m \cdot N_\mathrm{b}} c_n \mathrm{L_A}(c_n), \tag{3.24}$$

where $\boldsymbol{x}_m = [x_1, x_2, \ldots, x_m]^\mathrm{T}$ is the partial transmitted symbol vector and $\zeta_{n,m} \sim \mathcal{CN}(\mu_{\zeta_{n,m}}, \sigma^2_{\zeta_{n,m}})$ the soft estimate of the unknown layers—the equivalent term to (3.19)—with mean and variance given by

$$\mu_{\zeta_{n,m}} = \sum_{j=m+1}^{N_\mathrm{T}} h_{n,j}\mu_{x_j}, \qquad \sigma^2_{\zeta_{n,m}} = \sum_{j=m+1}^{N_\mathrm{T}} |h_{n,j}|^2 \sigma^2_{x_j}. \tag{3.25}$$

It can be seen that at the first layer, $(N_\mathrm{T} - 1)$ interfering terms are approximated by a single Gaussian variable. With increasing depth, less interfering terms are approximated and, hence, the accuracy of the metric is improved. At the last layer, the a posteriori LLR can be calculated by

$$\mathrm{L_P}(c_n|\boldsymbol{y}) \approx \max_{x_m \in \mathcal{L}^i_{+1}} \left\{\Lambda^{\mathrm{GTS}}(x_m)\right\} - \max_{x_m \in \mathcal{L}^i_{-1}} \left\{\Lambda^{\mathrm{GTS}}(x_m)\right\}, \tag{3.26}$$

where the size of the search set $\mathcal{L}$ is determined by $|\mathcal{L}| = M_\mathrm{sl} \cdot N_\mathrm{m}$. The parameter M_sl is equivalently used as with the QRD-M detection, i.e. it determines the number of branches that are kept alive during the traversal of the decision tree. Due to the rough Gaussian approximation, especially at the upper layers, M_sl has to be chosen comparably large until—with increasing iterations—the a priori information is sufficiently good in order to avoid the pruning of the correct branch. Particularly, pruning at the top layers is very critical, since a pruned branch can not be recovered at lower layers. In order to avoid this problem, it is beneficial to sort the layers w.r.t. their instantaneous received signal power, which can be determined by the estimated channel coefficients. Furthermore, the problem of a missing counter-hypothesis may arise caused by the pruning of nodes within a decision tree. The partial sets $\mathcal{L}^i_{+1}$ and $\mathcal{L}^i_{-1}$, used in (3.26), need to contain at least one hypothesis with $c_i = \pm 1$ at the ith position. Under certain circumstances, one of the two (either $+1$ or -1) might be pruned, which leads to infinitely large or small LLR values, respectively. In case of a missing counter-hypothesis, the missing counterpart is generated by inverting the available hypothesis.

By means of an improved initialization, e.g. with an MMSE detector [Kay09], M_sl can be set to a lower value already at the beginning of the GTS detection without degrading the performance, hence, reducing the overall complexity. Reliable a priori information, obtained by iterative turbo processing, contribute significantly in order to keep the correct branch alive during the traversal of the tree. As a result, M_sl is typically smaller during iterations as compared to its initial value.

Example 3.2 A simple example is given in the following to illustrate the principal idea of GTS detection and to highlight the difference to QRD-M detection. A noiseless 3 × 3 MIMO system is assumed again:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & h_{1,3} \\ h_{2,1} & h_{2,2} & h_{2,3} \\ h_{3,1} & h_{3,2} & h_{3,3} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{3.27}$$

The calculation of a transmit symbol x_m by means of a GTS detection is given by

$$\hat{x}_m = \sum_{n=1}^{N_\mathrm{R}} \left(y_n - \sum_{k=1}^{m-1} h_{n,k}\hat{x}_k - \underbrace{\sum_{l=m+1}^{N_\mathrm{T}} h_{n,l}\mu_{x_l}}_{\mu_{\zeta_{n,m}}} \right) / h_{n,m}, \tag{3.28}$$

with $m = N_\mathrm{T}, N_\mathrm{T} - 1, \ldots, 1$. For the cases of $m = 1$ and $m = N_\mathrm{T}$, the first and/or second sum within the bracket, respectively, are treated to be zero.

For each of the N_T transmit symbols $N_\mathrm{R}(N_\mathrm{T} - 1)$ multiplications and N_R divisions are required. Thus, in total $N_\mathrm{T}(N_\mathrm{R}(N_\mathrm{T}-1))$ multiplications and $N_\mathrm{T}N_\mathrm{R}$ divisions are necessary for a metric calculation with the Gaussian tree search. For a MIMO system with $N_\mathrm{T} \approx N_\mathrm{R}$, the complexity of a metric calculation is of the order $\mathcal{O}(N_\mathrm{T}^3)$ and with $N_\mathrm{T} \gg N_\mathrm{R}$: $\mathcal{O}(N_\mathrm{T}^2)$. A direct comparison to the costs of a QRD-M metric calculation shows that the GTS metric calculation is computationally more expensive. However, the costs of the QR factorization are not included. Moreover, the fact that QRD-M is limited to the case of $N_\mathrm{R} \leq N_\mathrm{T}$ is an advantage for the GTS and is, accordingly, a favorable detection algorithm in combination with the proposed graph-based receiver concept. The overall complexity of all detection algorithms depend on the number of hypotheses that are tested. In the case of QRD-M and GTS, this number is adaptive and a trade-off between complexity and performance has to be found. The subsequent section elaborates on this trade-off.

3.1.4 Performance/Complexity Trade-Off

The principle as well as the complexity of current state-of-the-art detection algorithms have been addressed in the previous section. Optimum performance (w.r.t. lowest bit-error-rate (BER)) is achieved with the APP detector requiring the highest computational complexity. The performance of the Gaussian detector, with the lowest complexity, depends significantly on the modulation order as well as the number of transmit antennas. The complexity of GTS detection depends on the parameter M_sl and thus, inherently, on the modulation order and number of transmit antennas as well. In the following, the Gaussian tree search is compared to both, an APP detector, providing optimum performance, and a Gaussian detector, which exhibits lowest computational complexity. It is shown that GTS detection offers a flexible trade-off between performance and complexity by appropriately choosing M_sl. The three selected detection algorithms are evaluated in a MIMO-OFDM system with $L = 300$ OFDM subcarriers and $K = 14$ OFDM symbols, respectively. A rate-1/2 turbo code is applied with a total of 5 global iterations where a global iteration comprises one iteration for the turbo code and one iteration for the data detection algorithm. The receiver setup is shown in Figure 3.3. The exchange of LLR values between the MIMO detector and the channel decoder is visualized by L_P and L_E, respectively. The subscripts of the LLR values indicate which component they belong to, i.e. subscript 1 is related to the MIMO detector and subscript 2 to the channel decoder. An AWGN channel is assumed. Table 3.1 contains the parameters under evaluation, e.g.

the number of transmit and receive antennas, modulation order, and the parameter M_{sl} for GTS detection. The parameter M_{sl} is determined empirically, i.e. the lowest value which achieves a close-to-optimum performance and reduces the complexity compared to the APP is chosen. Based on the *law of diminishing returns*, an additional constraint on the selection of M_{sl} is introduced, i.e. the increase of computational complexity has to yield a reasonable improvement w.r.t. BER performance. The number of metric calculations is used as a measure to compare the computational complexity between the algorithms. Despite the different computational costs per metric calculation, the number of metric calculations represents a critical parameter, since an exponentially increasing number of calculations eventually renders an arbitrary detection algorithm infeasible. For GTS, the number of metric calculations is given by the number of visited nodes, M_{vn}, defined in (3.18). Furthermore, the ratio of required metric calculations, R_c, of the detection algorithms with respect to the optimum APP detector is included as well. The results for a 2×2 MIMO system with QPSK ($N_{\mathrm{b}} = 2$), 16-QAM ($N_{\mathrm{b}} = 4$), and 64-QAM ($N_{\mathrm{b}} = 6$)

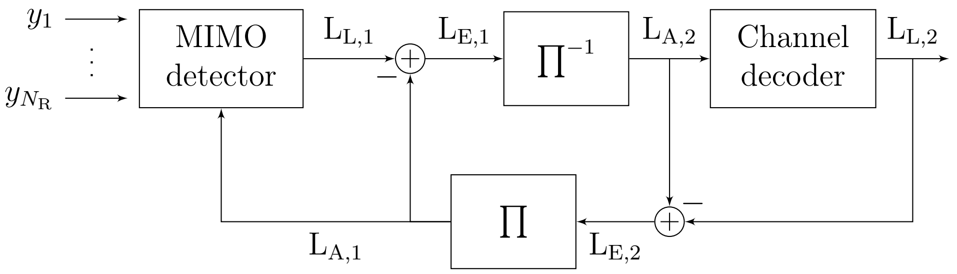
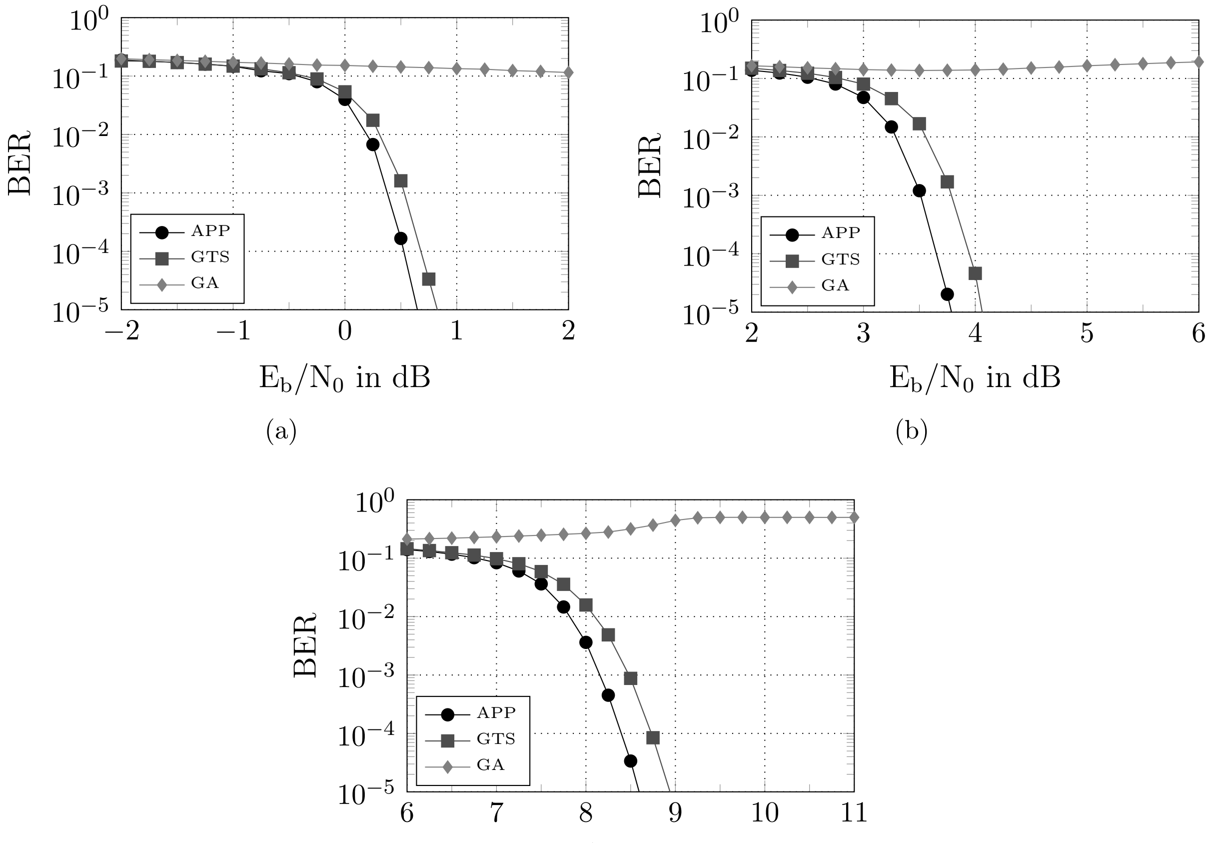


Figure 3.3: Receiver with iterative detection and decoding.

Table 3.1: Exemplary overview of computational complexity as a function of detection algorithm.

	$N_{\mathrm{T}} = N_{\mathrm{R}}$	N_{b}	M_{sl}	# of metric calculations	R_{c}
APP	2	2	x	16	100%
	2	4	x	256	100%
	2	6	x	4096	100%
	4	2	x	256	100%
	4	4	x	65536	100%
GA	2	2	x	4	25%
	2	4	x	16	1/16%
	2	6	x	64	1/64%
	4	2	x	4	1/64%
	4	4	x	16	1/4096%
GTS	2	2	2	12	75%
	2	4	8	176	68.75%
	2	6	30	1984	48.44%
	4	2	10	100	39.06%
	4	4	650	14768	22.60%



(a)

(b)

(c)

Figure 3.4: BER performance of different detection algorithms with a 2 × 2 antenna configuration and perfect channel knowledge. (a) with QPSK modulation, (b) with 16-QAM, and (c) with 64-QAM.

are given in Figure 3.4. As expected, the APP detector achieves the best performance independent of the modulation format. The BER performance of the Gaussian detector is nearly identical to the APP performance for QPSK modulation, and surprisingly better than the Gaussian tree search. This effect is caused by the chosen value of M_{sl} for which it is likely that the branch with the correct transmitted vector is pruned. Once a branch is removed from the decision tree, it can not be recovered at a later stage. Thus, the performance of the GTS with QPSK modulation and two transmit antennas, is slightly worse compared to the GA detector. However, the impact of the Gaussian approximation of the interfering term can be seen from the poor performance of GA in case of higher-order modulations. On the one hand, the central limit theorem states that the accuracy of the Gaussian approximation improves with increasing number of transmit antennas. On the other hand, the accuracy of the Gaussian approximation becomes worse with increasing modulation order [FSMH05]. It is thus not surprising that the GA is not converging for 16-QAM and 64-QAM. The performance of GTS for 16-QAM and 64-QAM is roughly 0.4 dB worse compared to the APP performance at a BER of 10^{-5}, but the computational

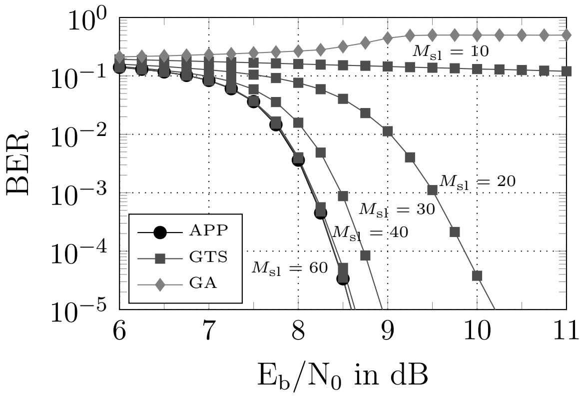


Figure 3.5: BER performance of Gaussian tree search as a function of M_{sl} with a 2×2 MIMO system and 64-QAM.

complexity is significantly reduced, with $R_c = 68.75\%$ and $R_c = 48.44\%$ of metric calculations for the GTS. As a matter of fact, the number of metric calculations does not directly translate into the number of hypotheses for GTS since partial path metric calculations have to be performed during the evaluation of the decision tree. Interestingly, the GTS suffers as well from the poor Gaussian approximation in combination with 64-QAM, as can be seen from a relatively large value of $M_{sl} = 30$. However, the overall complexity of the GTS detection can be reduced by providing initial a priori information. This information can be generated by means of an MMSE detection, for example. The effect of an initialization on the overall receiver complexity is studied in more detail in Section 5.4. Figure 3.5 shows the BER performance of the Gaussian tree search as a function of M_{sl} for a 2×2 MIMO system and 64-QAM. For comparison, the BER performance of GA and APP are included as well. For $M_{sl} = 10$, the performance of GTS is similar compared to the Gaussian detector for the SNR range of interest. With $M_{sl} = 10$, it is likely that the branch including the correct decision is pruned and, hence, the resulting performance is poor. With increasing M_{sl}, the BER performance of GTS detection approaches that of the APP detector. A value of $M_{sl} = 40$ approaches the APP performance up to 0.04 dB at a BER of 10^{-5}. Interestingly, optimal performance is achieved with $M_{sl} = 60$ and a further increase in complexity does not yield a gain w.r.t. BER performance. The chosen value of $M_{sl} = 30$ offers the best compromise between performance and complexity. The results for a 4×4 MIMO system for QPSK and 16-QAM, respectively, are shown in Figure 3.6. The performance of GTS detection is again about 0.2 dB and 0.4 dB worse compared to the APP performance. However, the gain w.r.t. the computational complexity compared to the APP detector could be further increased. As can be seen in Table 3.1, the relative ratio of metric calculations, R_c, is reduced by a factor of two for QPSK modulation with four transmit and receive antennas compared to the 2×2 scenario. Roughly, a total of 39% of metric calculation is required to reach a close-to-optimum solution. A similar gain is observed for 16-QAM modulation. About 22.6% ($M_{sl} = 650$) of metric calculations are needed to approach the APP performance up to 0.4 dB, whereas 26.27% ($M_{sl} = 800$) are needed for the optimum APP performance. It can be concluded from the values of M_{sl} given in Table 3.1 and obtained from additional numerical simulations,

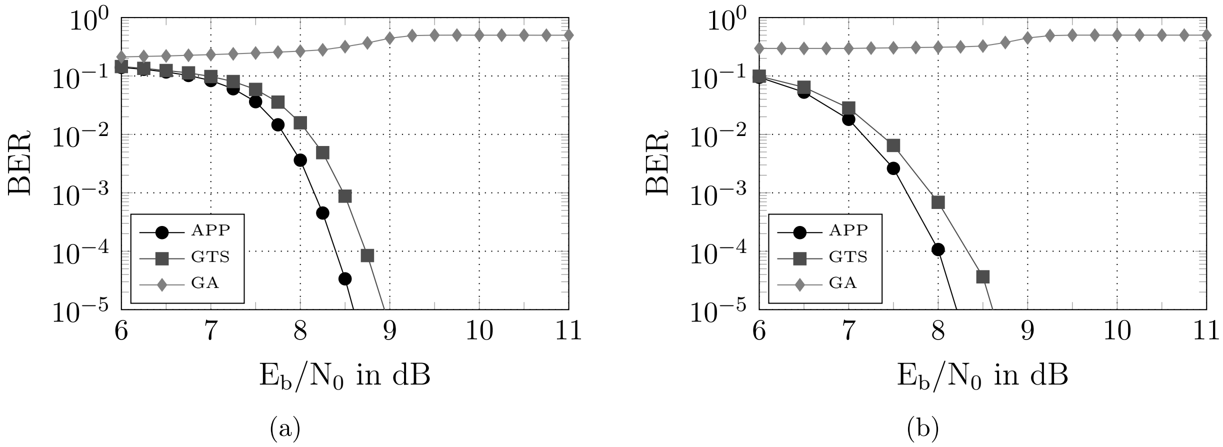


Figure 3.6: BER performance of different detection algorithms with a 4 × 4 antenna configuration and perfect channel knowledge. (a) with QPSK modulation, (b) with 16-QAM.

that the initial value of M_{sl}, which is required for GTS detection to converge, increases exponentially. However, for the parameters under investigation, GTS detection offers a good performance at an acceptable complexity. The initial complexity of GTS detection can be reduced by means of a suitable initialization.

The results for 64-QAM are omitted due to complexity reasons; for the APP, $16.78 \cdot 10^6$ hypotheses need to be evaluated for each OFDM subcarrier and OFDM symbol, which exceeds the available resources w.r.t. processing time and memory capacity.

3.2 Channel Estimation

The promising gains of MIMO-OFDM systems in terms of increased capacity, improved SNR, and high quality of service, can only be achieved with accurate channel state information [GJJV03]. In order to facilitate coherent detection, the channel has to be estimated at the receiver side. In general, channel estimation algorithms can be divided into three categories [Hoe13]: (1) pilot-based channel estimation, (2) data-aided (semi-blind) channel estimation, and (3) blind channel estimation. The literature on MIMO-OFDM channel estimation is extensive and a detailed overview of blind and non-blind channel estimation methods is given in [OA07] and [HAW11], respectively.

The most common method to acquire channel state information at the receiver side is to use pilots known to the transmitter and receiver [Aue12]. The corresponding estimation is commonly known as pilot-aided channel estimation (PACE). Pilots can be arranged depending on their purpose and provide the receiver sampled version of the channel response [Cav91]. Typically, for synchronization a pilot block occupying the first OFDM symbol is used, whereas for the tracking of time-varying, frequency-selective channels, pilots are multiplexed with data symbols. However, special care has to be taken for the allocation of multiplexed pilots, as will be explained in Section 3.3. In order to reconstruct the channel coefficients at the unknown data positions, multi-dimensional interpolation/filtering in time (OFDM symbols), frequency (OFDM subcarriers), and space (transmit/receive antennas) is often adopted [Hoe91, HKR97b, Aue12]. Although the spacing of pilots is upper bounded by the sampling theorem (cf. Section 3.3.1), an oversampling ratio of two for SISO systems [HKR97b] and even higher for MIMO systems [Aue03a] is recommended to achieve a good performance.

An alternative to interpolation/filtering in the frequency domain is given by so-called transform domain techniques. Instead of estimating the channel in the frequency domain, it is transformed, e.g. by means of an IFFT, to the time domain where the amount of coefficients is significantly reduced. With knowledge of the channel memory length, i.e. the number of non-zero coefficients in time domain, all coefficients exceeding this number can be set to zero (clipping), which results in a noise reduction when the coefficients are transformed back to the frequency domain [OA07].

The idea of semi-blind channel estimation (SBCE) is appealing, as it promises to improve the estimation accuracy by using reliably detected data symbols as so-called pseudo pilots. In case of error-free data detection, the complete transmitted sequence can be used, which significantly reduces the mean squared error (MSE) of the estimates as well as relaxes the constraints on the pilot density (cf. Chapter 6). Iterative receivers, which jointly carry out channel estimation and data detection, are seen as potential enablers to reduce the pilot overhead for MIMO-OFDM. Unfortunately, the complexity of performance-optimal ML and minimum mean squared error (MMSE) estimators grow exponentially with the modulation order and the number of transmit antennas [VW01, SJS03, CH03, AB07]. The class of suboptimal iterative receivers based on the expectation maximization (EM) algorithm significantly reduce computational costs and have attracted considerable interest [LWL01, XG03, KB06, YJ09]. Graph-based algorithms pose a viable alternative for iterative receivers and are subject of Chapter 5. However, especially at low SNR and/or fast fading channels, semi-blind approaches are sensitive to error propagation.

Since pilots do not carry any useful information, the effective throughput is limited. This problem is exacerbated, in case of MIMO systems, as the required pilot overhead grows proportionally with the number of transmit antennas [HH03, CA07]. Blind channel estimation techniques use the statistical behavior of the received signals [OA07]. The drawback of blind estimation techniques is their long acquisition time and their comparably high computational complexity [Hoe13].

This thesis is restricted to pilot-based and semi-blind channel estimation methods. Therefore, two widely applied methods used for pilot based channel estimation, namely least-squares (LS) and minimum mean squared error (MMSE), are presented in the following. Semi-blind channel estimation methods are the focus of Chapter 5 and Chapter 6.

Least-Squares Channel Estimation

Typically, training sequences are used in wireless communications systems to obtain initial information of the channel state information. The common approach for MIMO systems is to design the pilot grid orthogonal in time, frequency, and space. Nevertheless, other approaches are feasible as well [Li02]. OFDM subcarriers are allocated to pilots ensuring that they do not interfere with data symbols as well as with pilots of other antennas (see also Section 3.3.2). On the basis of an orthogonal pilot grid, the system model of (3.1), for PACE, is changed into an effective single antenna system in which only the positions of pilots are considered:

$$\boldsymbol{y}_{\mathrm{P}} = \boldsymbol{X}_{\mathrm{P}}\boldsymbol{h}_{\mathrm{P}} + \boldsymbol{n}_{\mathrm{P}}, \tag{3.29}$$

with

$$\boldsymbol{y}_{\mathrm{P}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_{\mathrm{P}}} \end{bmatrix}, \quad \boldsymbol{X}_{\mathrm{P}} = \begin{bmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x_{N_{\mathrm{P}}} \end{bmatrix}, \quad \boldsymbol{h}_{\mathrm{P}} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_{N_{\mathrm{P}}} \end{bmatrix}, \quad \boldsymbol{n}_{\mathrm{P}} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{N_{\mathrm{P}}} \end{bmatrix}, \tag{3.30}$$

where N_{P} denotes the number of pilots. A simple method to obtain initial channel estimates at pilot positions is given by the least-squares (LS) estimator:

$$\hat{\boldsymbol{h}}_{\mathrm{LS}} = \left(\boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}}\right)^{-1}\boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}} \cdot \boldsymbol{y}_{\mathrm{P}}. \tag{3.31}$$

For OFDM and the chosen orthogonal pilot grid, the LS estimator can be further simplified to $\hat{\boldsymbol{h}}_{\mathrm{LS}} = \boldsymbol{X}_{\mathrm{P}}^{-1}\boldsymbol{y}_{\mathrm{P}}$, since intersymbol interference is prevented and the symmetry of the corresponding diagonal matrix is fulfilled. The LS estimator is often applied for systems where the statistical properties of the channel are unknown and/or the optimal estimator is too complicated to apply [Kay09]. The trade-off between performance and complexity, as discussed for data detection algorithms in Section 3.1.4, is also valid for channel estimation. Hereby, the LS approach offers low complexity at the cost of a reduced performance.

Minimum Mean Squared Error Channel Estimation

The performance of the LS estimator can be further improved by exploiting information of the channel correlation as well as information of the noise process into account. By assuming that $\boldsymbol{h}_{\mathrm{P}}$ and $\boldsymbol{X}_{\mathrm{P}}$ are zero mean, the MMSE estimator is linear. Hence, the resulting estimate is given by [Kay09]:

$$\hat{\boldsymbol{h}}_{\mathrm{MMSE}} = \boldsymbol{\theta}_{\boldsymbol{hy}}\boldsymbol{\theta}_{\boldsymbol{yy}}^{-1}\boldsymbol{y}_{\mathrm{P}}, \tag{3.32}$$

where

$$\boldsymbol{\theta}_{\boldsymbol{hy}} = \mathrm{E}\left\{\boldsymbol{h}_{\mathrm{P}}\boldsymbol{y}_{\mathrm{P}}^{\mathrm{H}}\right\} = \boldsymbol{\theta}_{\boldsymbol{hh}}\boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}} \tag{3.33a}$$

$$\boldsymbol{\theta}_{\boldsymbol{yy}} = \mathrm{E}\left\{\boldsymbol{y}_{\mathrm{P}}\boldsymbol{y}_{\mathrm{P}}^{\mathrm{H}}\right\} = \boldsymbol{X}_{\mathrm{P}}\boldsymbol{\theta}_{\boldsymbol{hh}}\boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}} + \sigma_n^2\mathbf{I}_{\mathrm{P}}. \tag{3.33b}$$

Hereby, $\boldsymbol{\theta}_{\boldsymbol{hh}}$ corresponds to the autocorrelation function as defined in (2.8) depending on the position of pilots. Inserting (3.33) into (3.32) results in

$$\hat{\boldsymbol{h}}_{\mathrm{MMSE}} = \boldsymbol{\theta}_{\boldsymbol{hh}}\left(\boldsymbol{\theta}_{\boldsymbol{hh}} + \sigma_n^2(\boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}})^{-1}\right)^{-1}\boldsymbol{X}_{\mathrm{P}}^{-1}\boldsymbol{y}_{\mathrm{P}}. \tag{3.34}$$

The MMSE estimator can be interpreted as a post-processing filter of the LS estimate, as can be seen from (3.34), where the term $\boldsymbol{X}_{\mathrm{P}}^{-1}\boldsymbol{y}_{\mathrm{P}}$ corresponds to the LS estimate.

Due to the calculation of the multi-dimensional autocorrelation function $\boldsymbol{\theta}_{\boldsymbol{hh}}$ and the subsequent matrix inversion, the complexity of the MMSE is significantly larger compared to the LS estimator. Several approaches try to reduce the complexity: First, by applying the estimator independently for each domain, instead of using a multi-dimensional estimator. It has been shown in [HKR97a, HKR97b], that the separation of an MD-estimator into M-1D estimators has only a minor influence on the achievable performance. Furthermore, a generalization of an MMSE estimator to arbitrary pilot grids is presented in [HKR97a, HKR97b] as well. Secondly, by assuming a uniform distributed spectrum (cf. Section 2.1.2), the autocorrelation function can be pre-calculated and stored in tables.

3.3 Pilot Allocation for MIMO-OFDM Systems

The optimum pilot allocation is a trade-off between the accuracy of the channel estimation and the spectral efficiency of a system [OA07]. In general, pilot design can be differentiated into the categories: pilot grid, orthogonal pilot set, and pilot type [AC09a]. In order to sample the channel response sufficiently, the sampling theorem has to be obeyed, however, in case of a MIMO system, the amount of pilots dictated by the sampling theorem might not be sufficient to resolve the interference of adjacent antennas. Hence, more pilots are required to improve the channel estimation accuracy. Optimum pilot patterns for MIMO-OFDM have been studied extensively, providing a minimum of pilot overhead while minimizing the mean squared error of the estimates. The Nyquist-Shannon sampling theorem, which determines the maximum spacing of pilots for which the channel response can be reconstructed error-free, is presented briefly in Section 3.3.1. It will be shown in Chapter 6, that these bounds are valid for pilot-based channel estimation but can be relaxed for semi-blind channel estimation methods. Subsequently, common pilot grids used for MIMO-OFDM systems are presented in Section 3.3.2. Additionally, orthogonal pilot sets are briefly discussed as well as pilot types suitable for beamforming.

3.3.1 Multi-Dimensional Sampling Theorem

PACE is typically applied for the initialization of subsequent detection and estimation algorithms, which provides sampled versions of the noisy channel response. In case of OFDM, the channel response is often correlated in time and frequency. Thus, to provide estimates at also the unknown data positions interpolation/filtering is applied. As said before, sampling the channel response more often yields more accurate estimates at the cost of a reduced bandwidth efficiency. Accordingly, it is beneficial to sample the response as low as possible. Lower bounds for the perfect reconstruction of a sampled channel response given a minimum pilot density are provided by the multi-dimensional sampling theorem [DM84]. Moreover, the sampling theorem states that any signal with a bandlimited spectrum can be perfectly reconstructed by an ideal lowpass filter with equi-spaced pilots [Hoe13].

For MIMO-OFDM, three to four dimensions—i.e. time, frequency, and space at the transmitter and/or receiver—are typically exploited for multi-dimensional channel estimation. These dimensions exhibit a sufficiently strong correlation in order to improve the estimation accuracy by increasing the number of observations. Typically, a correlation coefficient of at least 0.8 is required to yield a gain as will be shown in Section 5.3 and is independently reported in [OA07]. Considering the scenarios of the WINNER channel models described in Section 2.2, a sufficient correlation is more likely for time, frequency, and the spatial domain at the transmitter side. The three-dimensional spectrum, as defined in (2.9), is confined in the ranges $[-f_{\mathrm{D,max}}, f_{\mathrm{D,max}}]$, $[0, \tau_{\mathrm{max}}]$, and $[\Psi_{\mathrm{AoD}} - \Phi_{\mathrm{AoD}}/2, \Psi_{\mathrm{AoD}} + \Phi_{\mathrm{AoD}}/2]$. Accordingly, the unknown channel impulse response at data positions can be reconstructed by an ideal lowpass filter whose passband matches the above mentioned ranges. Multiplexed, equi-spaced pilots provide initial channel estimates for a corresponding lowpass filter. Due to the periodic sampling of the channel response, replicas of its autocorrelation spectrum, known as aliases, are created [Aue12]. Pilots must be located sufficiently close in such a way that a spectral overlap between the filter passband and its aliases is prevented.

Let D_{s}, D_{f}, and D_{t} denote the spacing of pilots in space, frequency, and time, respectively [Aue12, Hoe13]. According to the Nyquist-Shannon sampling theorem, the spacing of pilots for the three domains is as follows:

$$D_{\mathrm{s}} < \frac{\lambda}{d_{\mathrm{Tx}}} \cdot \frac{1}{2\sin\left(\frac{\Phi_{\mathrm{AoD}}}{2}\right) \cdot \cos\left(\Psi_{\mathrm{AoD}}\right)}, \tag{3.35a}$$

$$D_{\mathrm{f}} < \frac{1}{\tau_{\mathrm{max}} F_{\mathrm{s}}}, \tag{3.35b}$$

$$D_{\mathrm{t}} < \frac{1}{2 f_{\mathrm{D,max}} T_{\mathrm{s}}}. \tag{3.35c}$$

It can be seen that the spacing in space depends on the composite angular spread Φ_{AoD} and the mean composite AoD Ψ_{AoD}. A mean composite AoD of $\cos(\Psi_{\mathrm{AoD}}) = 1$ represents the worst case scenario, given a composite angular spread. Obviously, the spacing D_{s} is further reduced with a composite angular spread of $\sin(\Phi_{\mathrm{AoD}}/2) = 1$. However, assuming a fixed angular spread is considered to be unrealistic and therefore omitted. The spacing of D_{f} and D_{t} depends on the normalized maximum fading $\tau_{\mathrm{max}} F_{\mathrm{s}}$ and $f_{\mathrm{D,max}} T_{\mathrm{s}}$, respectively. The minimum pilot density for 3D-PACE, assuming the worst case scenario, results in [Aue12]

$$\frac{1}{D_{\mathrm{s}} D_{\mathrm{f}} D_{\mathrm{t}}} > \delta_{\mathrm{min}} = 4\frac{d_{\mathrm{Tx}}}{\lambda} \sin\left(\frac{\Phi_{\mathrm{AoD}}}{2}\right) \cdot \tau_{\mathrm{max}} F_{\mathrm{s}} \cdot f_{\mathrm{D,max}} T_{\mathrm{s}}. \tag{3.36}$$

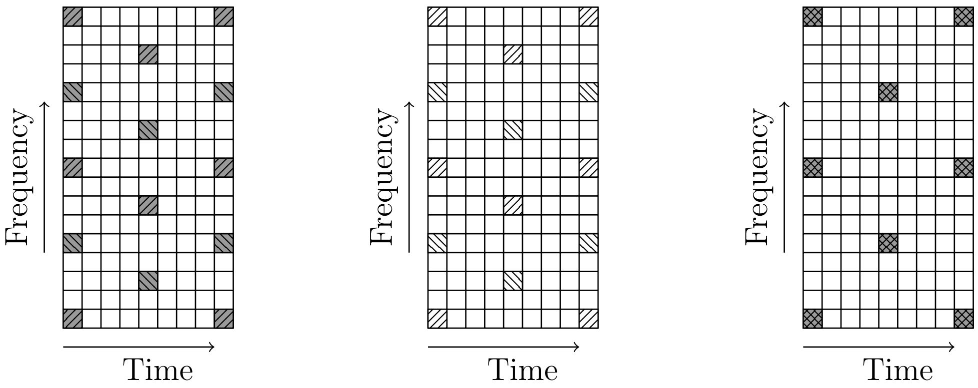
Since the received sequence is often impaired by noise and/or interference, oversampling has to be performed in practice. As mentioned before, an oversampling ratio of two for SISO systems [HKR97b] and even higher for MIMO systems [Aue03a] is recommended to achieve a good performance.

While the sampling theorem defines the maximum spacing of pilots for PACE, the strict bound can be relaxed when semi-blind channel estimation is performed. Data symbols are iteratively detected and with increasing reliability, can be used as pseudo-pilots. The spacing of pilots has to be sufficiently narrow as to provide a "good enough" estimate to enable convergence of the iterative receiver. More detailed information about the sampling theorem in conjunction with iterative semi-blind channel estimation is given in Chapter 6.

3.3.2 Pilot Grids for MIMO-OFDM

The primary goal of a pilot grid is to minimize the mean squared error of an estimate with a minimum of pilots. This can be achieved with different grids depending on the channel estimation method. While arbitrary grids are possible, a few selected grids dominate the literature and will be presented in the following.

Figure 3.7 shows three popular pilot grids used for MIMO-OFDM with two transmit antennas. Each square represents a resource element, i.e. the smallest assignable unit which consists of one OFDM subcarrier during one OFDM symbol and can be used for data symbols and/or pilots. The ascending hatching denotes pilots of the first antenna, while the descending refers to pilots of the second antenna. A gray background, as shown in Figures 3.7a and 3.7c, indicates, that the specific resource element is used for pilots exclusively. A white background, shown in Figure 3.7b, symbolizes pilots superimposed with data symbols. Thus, for the pilot grid of Figure 3.7a, pilots are multiplexed orthogonally in time, frequency, and space, arranged in a so-called diamond shaped grid. Extensions to other regular pilot grids is straightforward. The initial MIMO channel estimation is effectively converted into a SISO channel estimation problem. Although other possibilities exist to orthogonally separate training symbols, they lead to higher complexity and/or at least the same pilot overhead [Aue04]. Moreover, the pilots are typically equidistantly spaced in time and frequency [Hoe91]. The resulting pilot grid minimizes the harmonic mean of the MSE of channel estimates over all OFDM subcarriers [CL05] and additionally maximizes the capacity [ATV02, CA07]. Although this pilot grid simplifies the initial PACE, it inherently introduces drawbacks to the system. On the one hand, the bandwidth efficiency is reduced since many subcarriers are reserved for the transmission of pilots or are required to be silent. And on the other hand, the silent subcarriers increase the peak



(a) Orthogonal multiplexing of pilots in time, frequency, and space

(b) Pilots superimposed with data symbols

(c) Pilots superimposed with pilots

Figure 3.7: Pilot Grids for MIMO-OFDM.

to average power ratio (PAPR) [OA07], which is already a major drawback of OFDM systems. In order to mitigate this problem, pilots can be superimposed with data symbols, as illustrated in Figure 3.7b. In this case, data symbols are interfering with pilots, which deteriorates the MSE of PACE. Moreover, $N_{\mathrm{T}} \cdot N_{\mathrm{P}}$ unknowns have to be estimated with N_{P} equations being available.

Figure 3.7c illustrates the superposition of pilots of all transmit antennas. Several advantages are introduced with this grid: Since all transmit antennas can position pilots at both the beginning and end of a burst, so-called edge effects are reduced. As can be seen for the first two grids, either the first or second transmit antenna transmits a pilot. Due to the effectively widened pilot spacing, the MSE deteriorates especially at the corresponding edges. A simplified channel estimation is additionally provided, given a special pilot sequence designed according to [Li00, Li02]:

$$x_{i,m} = \exp\left(-j2\pi \cdot i \cdot m / N_{\mathrm{T}}\right), \qquad 1 \leq i \leq N'_{\mathrm{P}}, \quad 1 \leq m \leq N_{\mathrm{T}}, \tag{3.37}$$

where N'_{P} denotes the number of pilots per transmit antenna.

The pilot type distinguishes between *dedicated pilots* that are subject to the same user-specific precoding scheme as data symbols, and *common pilots* which are transmitted without any precoding. Common pilots are distributed throughout the assigned bandwidth and are used to acquire channel quality information (CQI). An improved channel estimation quality is achieved by means of interpolation/filtering. The major drawback of common pilots is the high overhead since at least N_{T} pilots need to be transmitted. Even more pilots are required if interpolation is applied. However, they can be flexibly deployed and provide the necessary CQI measurements over the entire band [AC09a].

Dedicated pilots are user-specific precoded pilots that change according to the applied precoding scheme. Typically, interpolation is only feasible between pilots that are subject to the same precoding. For spatially separated beams, a pilot reuse is possible which may reduce the pilot overhead to that of a single antenna system. Unfortunately, dedicated pilots cannot be used to obtain CQI over the entire band, which is needed for link adaptation at the transmitter side [AC09a].

It is shown in Section 5.6 that the proposed graph-based iterative receiver is easily adapted to channel estimation based on dedicated pilots. Given a priori information about the applied precoding, channel estimation across pilots of different precoding is enabled.

3.4 Combining of Correlated Random Measures

Message combining is an essential component in most digital communication systems. Correlation between random measures has a significant impact on the combining process. In order to provide the best estimate after combining, correlation must be considered. Typically, two extreme cases can be identified: (1) multiple observations are combined to represent a single variable ($p(y_1, \ldots, y_{N_{\mathrm{RM}}}|x)$) and (2) multiple variables are combined, which are instances of the same observation ($p(y|x_1, \ldots, x_{N_{\mathrm{RM}}})$). The first case appears in numerous applications, such as sensor networks with one transmit antenna and multiple receive antennas (single-input multiple-output, SIMO) providing multiple observations of, for example, a transmitted signal. The latter is typical for systems with multiple transmit antennas and a single receive antenna (multiple-input multiple-output, MISO). In the following, the combination of correlated variables as well as correlated observations is studied.

3.4.1 Combination of Multiple Correlated Observations

In digital communications, multiple observations can efficiently be combined by utilizing log-likelihood values (L-values). In numerous applications, such as sensor networks and/or systems with one transmit antenna and multiple receive antennas, multiple observations of one variable are provided. The corresponding single-input multiple-output (SIMO) system is given by

$$\boldsymbol{y} = \boldsymbol{i} \cdot x + \boldsymbol{n}, \tag{3.38}$$

where $\boldsymbol{y}$ contains N_{RM} observations, x is the transmit symbol multiplied with the unit vector $\boldsymbol{i}$ of appropriate length, and $\boldsymbol{n}$ contains the noise components with variance σ^2. Without loss of generality, the transmit symbol is assumed to be BPSK-modulated. Revisiting (3.6), the extrinsic LLR can be written as

$$\mathrm{L_E}\left(\boldsymbol{y}|x\right) \doteq \ln \frac{p\left(\boldsymbol{y}|x=+1\right)}{p\left(\boldsymbol{y}|x=-1\right)}. \tag{3.39}$$

The conditional probability density function (pdf) $p(\boldsymbol{y}|x)$ is defined as

$$p\left(\mathbf{y}|x\right) \doteq \frac{1}{(2\pi)^{N_{\mathrm{RM}}/2}|\mathbf{\Sigma}|^{1/2}} \cdot \exp\left(-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{i}\cdot\mu_x\right)^{\mathrm{T}}\mathbf{\Sigma}^{-1}\left(\mathbf{y}-\boldsymbol{i}\cdot\mu_x\right)\right), \tag{3.40}$$

where $\mathbf{\Sigma}$ refers to the covariance matrix [HOP96] and μ_x to the mean value of the variable x. Since x is not to be estimated in this case, the mean value μ_x corresponds to the hypothesis of $\tilde{x} = \pm 1$. Commonly, uncorrelated noise samples are assumed. Hence, the covariance matrix has non-zero entries only along its main diagonal and the conditional pdf simplifies to (3.8). In this case, the reliability information of all N_{RM} received symbols can be combined according to

$$\mathrm{L_E}\left(\boldsymbol{y}|x\right) = \sum_{i=1}^{N_{\mathrm{RM}}} \mathrm{L_E}\left(y_i|x\right), \tag{3.41}$$

where [PRV96]

$$\begin{aligned}\mathrm{L_E}\left(y_i|x\right) &= \ln\left(\frac{p(y_i|x=+1)}{p(y_i|x=-1)}\right)\\ &= \ln\left(\exp\left(-\frac{1}{2\sigma_i^2}\cdot\left((y_i-1)^2-(y_i+1)^2\right)\right)\right)\\ &= \frac{2}{\sigma_i^2}y_i.\end{aligned} \tag{3.42}$$

However, this is only true for uncorrelated observations. Given correlated L-values $\mathrm{L_E}(y_i|x)$ *equal gain combining*, according to (3.41), is too optimistic (cf. Example 3.3). The problem can be solved by a weighted superposition of L-values. The weighted superposition of L-values results in *unequal gain combining* and can be written as follows:

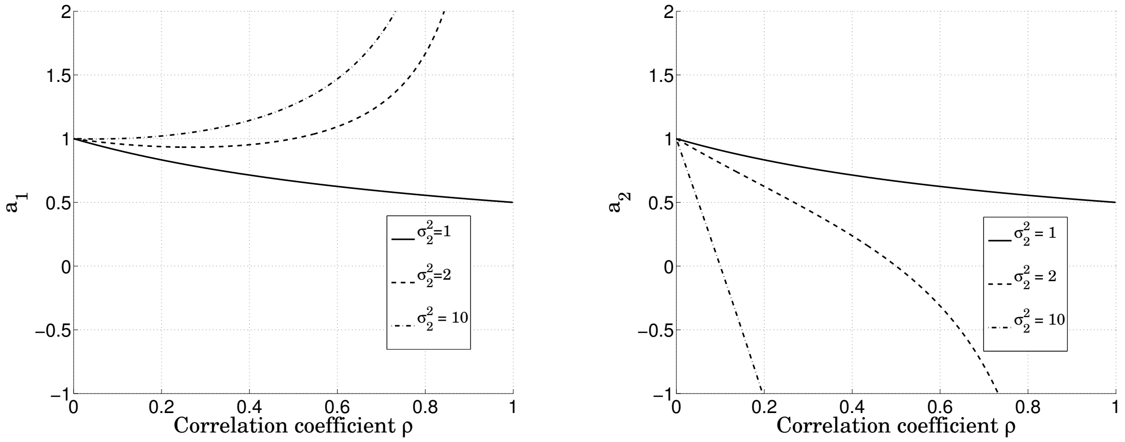
$$\mathrm{L_E}\left(\mathbf{y}|x\right) = \sum_{i=1}^{N_{\mathrm{RM}}} a_i \mathrm{L_E}\left(y_i|x\right). \tag{3.43}$$

Inserting (3.40) into (3.39) yields

$$\begin{aligned}\mathrm{L_E}\left(\mathbf{y}|x\right) =& \frac{1}{2}\left[-(\mathbf{y}-1)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-1)+(\mathbf{y}+1)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}+1)\right]\\ =& \sum_{i=1}^{N_{\mathrm{RM}}} y_i\left[2\boldsymbol{\Sigma}_{ii}^{-1}+\sum_{j=1,j\neq i}^{N_{\mathrm{RM}}}\left(\boldsymbol{\Sigma}_{ij}^{-1}+\boldsymbol{\Sigma}_{ji}^{-1}\right)\right]\\ =& \sum_{i=1}^{N_{\mathrm{RM}}} 2y_i\underbrace{\left[\boldsymbol{\Sigma}_{ii}^{-1}+\sum_{j=1,j\neq i}^{N_{\mathrm{RM}}}\boldsymbol{\Sigma}_{ij}^{-1}\right]}_{\lambda_i}\\ =& \sum_{i=1}^{N_{\mathrm{RM}}} \mathrm{L_E}(y_i|x)\cdot\underbrace{\sigma_i^2\lambda_i.}_{a_i}\end{aligned} \tag{3.44}$$

The covariance matrix $\boldsymbol{\Sigma}$ can either be calculated online or is pre-calculated and stored in tables, which would reduce the computational complexity. Furthermore, in case the variances differ significantly in their value, a mismatched correlation may lead to a very large magnitude of the weighting factors which deteriorates the estimation accuracy [LGC88].

> **Example 3.3** The combining of two correlated L-values is used as an example to further illustrate the principle of correlated combining. Hereby, a transmit symbol x is received by two antennas. The correlated observations are given by y_1 and y_2. The conditional pdf $p(y_1, y_2|x)$ is thus given by
>
> $$\begin{aligned}p\left(y_1, y_2|x\right) =& \frac{1}{2\pi\hat{\sigma}_1\hat{\sigma}_2\sqrt{1-\rho^2}}\cdot\\ \exp&\left\{-\frac{\hat{\sigma}_2^2(y_1-x)^2-2\rho\hat{\sigma}_1\hat{\sigma}_2(y_1-x)(y_2-x)+\hat{\sigma}_1^2(y_2-x)^2}{2\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)}\right\},\end{aligned} \tag{3.45}$$
>
> where
>
> $$\rho = \frac{\mathrm{E}\left\{(y_1-x)\cdot(y_2-x)\right\}}{\hat{\sigma}_1\hat{\sigma}_2}. \tag{3.46}$$
>
> Inserting (3.45) into (3.39) yields
>
> $$\mathrm{L_E}\left(y_1, y_2|x\right) = \ln\left(\frac{\exp\left\{-\frac{\hat{\sigma}_2^2(y_1-1)^2-2\rho\hat{\sigma}_1\hat{\sigma}_2(y_1-1)(y_2-1)+\hat{\sigma}_1(y_2-1)^2}{2\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)}\right\}}{\exp\left\{-\frac{\hat{\sigma}_2^2(y_1+1)^2-2\rho\hat{\sigma}_1\hat{\sigma}_2(y_1+1)(y_2+1)+\hat{\sigma}_1(y_2+1)^2}{2\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)}\right\}}\right). \tag{3.47}$$



(a) Weighting factor a1

(b) Weighting factor a2

Figure 3.8: Weighting factors for two L-values as a function of correlation and variance.

After straightforward derivation, we obtain:

$$
\begin{aligned}
\mathrm{L_E}\left(y_1, y_2|x\right) &= \frac{1}{1-\rho^2}\left[\frac{2y_1}{\hat{\sigma}_1^2} - \frac{2\rho(y_1+y_2)}{\hat{\sigma}_1\hat{\sigma}_2} + \frac{2y_2}{\hat{\sigma}_2^2}\right] \\
&= \frac{1}{1-\rho^2}\left[L(y_1) - \rho\left(\frac{\hat{\sigma}_1}{\hat{\sigma}_2}L(y_1) + \frac{\hat{\sigma}_2}{\hat{\sigma}_1}L(y_2)\right) + \mathrm{L_E}(y_2)\right] \\
&\doteq a_1\mathrm{L_E}(y_1) + a_2\mathrm{L_E}(y_2),
\end{aligned}
\tag{3.48}
$$

with

$$
a_1 = \frac{1-\rho\frac{\hat{\sigma}_1}{\hat{\sigma}_2}}{1-\rho^2}, \qquad a_2 = \frac{1-\rho\frac{\hat{\sigma}_2}{\hat{\sigma}_1}}{1-\rho^2}. \tag{3.49}
$$

The weighting factors a_1 and a_2 with a correlation of $\rho = 0$ result in $a_1 = a_2 = 1$. Whereas for a correlation of $\rho \to 1$ and equal variances $\hat{\sigma}_1/\hat{\sigma}_2 = 1$, the weighting factors are $a_1 = a_2 = 1/2$. Figure 3.8 depicts the weighting factors as a function of the correlation coefficient ρ, with $\sigma_1^2 = 1$ and different variances of σ_2^2. The improvement due to the combining is gradually reduced with increasing correlation. In general, no improvement with the combining of variables is achieved when $\rho\hat{\sigma}_1 > \hat{\sigma}_2$ and/or $\rho\hat{\sigma}_2 > \hat{\sigma}_1$. The weighting factors for $\hat{\sigma}_1^2 = 1$ and $\hat{\sigma}_2^2 = 10$ illustrate the importance of a correct correlation coefficient. It can be seen that for a correlation above $\rho = 0.2$, the influence of the second L-value dominates and finally determines the combined L-value.

3.4.2 Combination of Multiple Correlated Variables

The application of combining multiple random variables of one observation (MISO) appears frequently in various kinds of situations. A common example for instance is given by the following experiment: A physical entity (e.g. temperature, location, etc.) is measured several times by $N_{\mathrm{RM}} = 2$ independent persons. Due to measurement inaccuracies, e.g. human error, the observed values contain errors. Expressed in mathematical terms, the estimates can be represented by two Gaussian distributions defined by a mean and variance: $p \sim \mathcal{N}(\hat{\mu}_i, \hat{\sigma}_i^2)$, $i \in \{1, 2\}$ where the mean value relates to the actual observation/estimate and the variance refers to its reliability. Because of a specific training, the second person is able to measure more accurately, i.e. the resulting variance is smaller. The combination of the two estimated values will result in an even smaller variance and, thus, in a more accurate estimate [May79, LGC88]. Accordingly, the aim of MISO is to improve the quality of the estimated variable, opposed to the previous application of SIMO, where the hypotheses of a discrete variable were evaluated.

Prominent technical applications that combine random variables to improve the estimation accuracy are the Kalman filter [May79] as well as factor graphs and the Sum-Product algorithm [KFL01, Loe04]. Typically within a factor graph, messages are exchanged, which refer to the estimates of a random variable. Quite often the true pdf of a random variable is approximated by a Gaussian pdf in order to reduce the computational complexity. Multiple Gaussian pdfs, which represent individual estimates of a random variable, are combined within the factor graph to increase the estimation accuracy.

The best linear unbiased estimator (BLUE) combines the N_{RM} individual uncorrelated variables $p(y_i|x_i) \sim \mathcal{CN}(\hat{\mu}_i, \hat{\sigma}_i^2)$ as follows:

$$p(y|x) = \prod_{i=1}^{N_{\mathrm{RM}}} p(y_i|x_i) \tag{3.50}$$

with mean and variance

$$\hat{\mu} = \frac{\sum\limits_{i=1}^{N_{\mathrm{RM}}} \frac{\hat{\mu}_i}{\hat{\sigma}_i^2}}{\sum\limits_{i=1}^{N_{\mathrm{RM}}} \frac{1}{\hat{\sigma}_i^2}}, \qquad \hat{\sigma}^2 = \frac{1}{\sum\limits_{i=1}^{N_{\mathrm{RM}}} \frac{1}{\hat{\sigma}_i^2}}. \tag{3.51}$$

If the individual messages are correlated, the message combining process has to be changed accordingly. In order to obtain the lowest possible mean squared error of an estimate $\hat{\mu}$ in terms of BLUE, it is required that

1. $\hat{\mu}$ is a linear combination of individual estimates $\hat{\mu}_i$;
2. $\hat{\mu}$ provides an unbiased estimate of the true value μ;
3. $\hat{\mu}$ has the lowest possible variance.

A general way to express Condition 1 is given by

$$\hat{\mu} = \sum_{i=1}^{N_{\mathrm{RM}}} \alpha_i \hat{\mu}_i. \tag{3.52}$$

Condition 2 requires $\hat{\mu}$ to be unbiased, i.e. $\mathrm{E}\{\hat{\mu}\}=\mu$. Assuming that the individual estimates $\hat{\mu}_i$ are unbiased, a normalization constraint is required, thus Condition 2 is fulfilled:

$$\sum_{i=1}^{N_{\mathrm{RM}}} \alpha_i \doteq 1. \tag{3.53}$$

The weighting factors α_i need to be determined to such an extent that the variance of the estimate is minimized. In [LGC88], Lagrangian multipliers are applied to determine the weighting factors

$$\boldsymbol{\alpha} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{i}_{N_{\mathrm{RM}}}}{\boldsymbol{i}_{N_{\mathrm{RM}}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{i}_{N_{\mathrm{RM}}}}, \tag{3.54}$$

where $\boldsymbol{\Sigma}$ represents the covariance matrix as before and $\boldsymbol{i}_{N_{\mathrm{RM}}}$ denotes a unit vector of length N_{RM}. Based on (3.52), the variance $\hat{\sigma}^2$ can be calculated as follows [LGC88]:

$$\hat{\sigma}^2 = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\alpha}, \tag{3.55}$$

where $\boldsymbol{\alpha}$ is the column vector of the weighting factors α_i.

Example 3.4 ($N_{\mathrm{RM}} = 2$) The following example illustrates the combination of correlated Gaussian distributions in more detail. Given $p(y_1|x_1) \sim \mathcal{CN}(\hat{\mu}_1, \hat{\sigma}_1^2)$ and $p(y_2|x_2) \sim \mathcal{CN}(\hat{\mu}_2, \hat{\sigma}_2^2)$, the covariance matrix results in

$$\boldsymbol{\Sigma} = \begin{pmatrix} \hat{\sigma}_1^2 & \rho\hat{\sigma}_1\hat{\sigma}_2 \\ \rho\hat{\sigma}_1\hat{\sigma}_2 & \hat{\sigma}_2^2 \end{pmatrix}. \tag{3.56}$$

The new mean value of the combined Gaussian function according to (3.52) is calculated as

$$\hat{\mu} = \alpha_1 \cdot \hat{\mu}_1 + \alpha_2 \cdot \hat{\mu}_2. \tag{3.57}$$

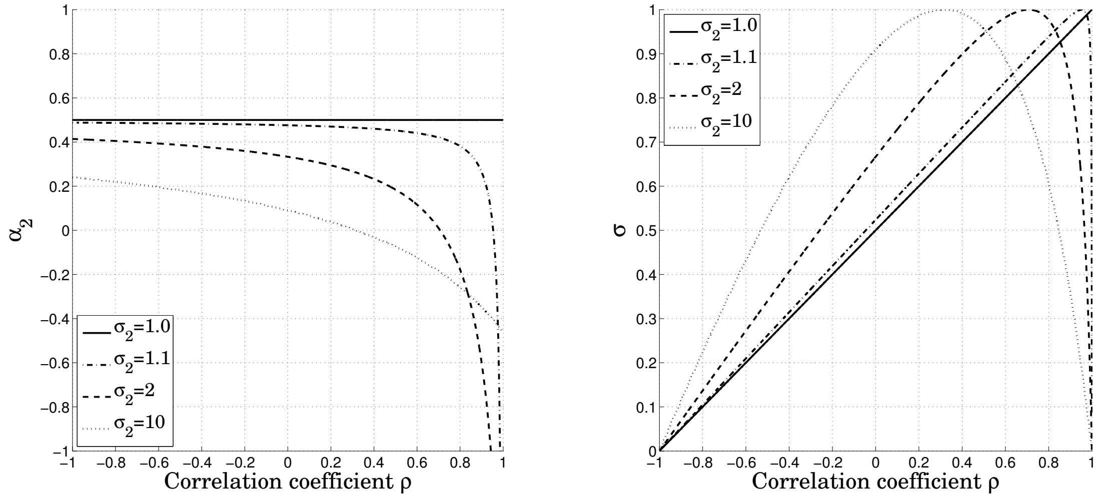
According to (3.54), the weighting factors are

$$\alpha_1 = \frac{\hat{\sigma}_2\left(\hat{\sigma}_2 - \rho\hat{\sigma}_1\right)}{\hat{\sigma}_1^2 - 2\rho\hat{\sigma}_1\hat{\sigma}_2 + \hat{\sigma}_2^2}, \qquad \alpha_2 = \frac{\hat{\sigma}_1\left(\hat{\sigma}_1 - \rho\hat{\sigma}_2\right)}{\hat{\sigma}_1^2 - 2\rho\hat{\sigma}_1\hat{\sigma}_2 + \hat{\sigma}_2^2}. \tag{3.58}$$

The variance of the estimated parameter is given by

$$\hat{\sigma}^2 = \left(1 - \rho^2\right)\frac{\hat{\sigma}_1^2\hat{\sigma}_2^2}{\hat{\sigma}_1^2 - 2\rho\hat{\sigma}_1\hat{\sigma}_2 + \hat{\sigma}_2^2}. \tag{3.59}$$

It is obvious, that (3.51) and (3.57), (3.59) are equivalent if $\rho = 0$. The weighting factor α_2 and the resulting variance after combining are shown in Figure 3.9. The results are intuitive for equal variance $\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = 1$, where the weighting factors are $\alpha_1 = \alpha_2 = 0.5$ independent of the correlation. However, an improvement is only achieved if $\hat{\sigma}_1 \neq \rho\hat{\sigma}_2$ and/or $\hat{\sigma}_2 \neq \rho\hat{\sigma}_1$. Moreover, it can be seen that for negative correlations the variance even reduces, which means that the estimate gets more reliable. For a correlation of $\rho = -1$, the estimate can actually be determined perfectly ($\sigma = 0$). Similarly to the combining of correlated observations, in case of $\hat{\sigma}_1/\hat{\sigma}_2 \neq 1$, the estimation accuracy depends on the accuracy of the correlation. A mismatch deteriorates the achievable accuracy.



(a) Weighting factor α_2 as a function of correlation, for different variance values $\hat{\sigma}_2^2$

(b) Resulting variance after combining as a function of correlation, for different values $\hat{\sigma}_2^2$

Figure 3.9: Properties of the BLUE solution as a function of correlation, for varying values of $\hat{\sigma}_2^2$ and fixed $\hat{\sigma}_1^2 = 1.0$.

Example 3.5 The purpose of the following example is to explain the relation between the SIMO and the MISO case in more detail. In the previous chapter, the SIMO case explained the combining of correlated LLR values and depends hereby on a discrete variable x. Under this condition, every hypotheses can be tested, i.e. x does not need to be estimated. Furthermore, it is assumed that the multiple observations are obtained by a single variable.

The MISO case, on the other hand, is used to combine correlated variables in order to improve the accuracy of the combined variable. Consequently, the hypotheses of the transmit variable x may be first combined and subsequently used to calculate an LLR value. It is shown in the following that the two methods obtain equivalent results.

The corresponding system model is given by

$$\begin{aligned} y_1 &= x + n_1 \\ y_2 &= x + n_2, \end{aligned} \tag{3.60}$$

where n is a Gaussian white noise term with variance σ_n^2. Without loss of generality, it is assumed that the observations are uncorrelated, i.e. $\rho = 0$. Thus, (3.44) for the SIMO case yields

$$\mathrm{L}_{\mathrm{E,SIMO}} = \frac{2}{\hat{\sigma}_n^2} y_1 + \frac{2}{\hat{\sigma}_n^2} y_2. \tag{3.61}$$

The MISO case is used to obtain an estimate of x. Given the two observations y_1 and y_2, two corresponding hypotheses of x are obtained: $\hat{\mu}_1 = y_1$ and $\hat{\mu}_2 = y_2$.

Combining the two estimates according to (3.51) results in

$$\hat{\mu} = \frac{\hat{\mu}_1/\hat{\sigma}_n^2 + \hat{\mu}_2/\hat{\sigma}_n^2}{1/\hat{\sigma}_n^2 + 1/\hat{\sigma}_n^2}, \qquad \hat{\sigma} = \frac{1}{1/\hat{\sigma}_n^2 + 1/\hat{\sigma}_n^2}. \tag{3.62}$$

Hereby, the LLR value is given by

$$\mathrm{L}_{\mathrm{E,MISO}} = \frac{2}{\hat{\sigma}^2}\hat{\mu}. \tag{3.63}$$

Inserting (3.62) into (3.63) yields

$$\begin{aligned} \mathrm{L}_{\mathrm{E,MISO}} &= \frac{2}{\frac{1}{1/\hat{\sigma}_n^2+1/\hat{\sigma}_n^2}} \cdot \frac{\hat{\mu}_1/\hat{\sigma}_n^2 + \hat{\mu}_2/\hat{\sigma}_n^2}{1/\hat{\sigma}_n^2 + 1/\hat{\sigma}_n^2} \\ &= \frac{4}{\hat{\sigma}_n^2} \cdot \frac{\hat{\mu}_1 + \hat{\mu}_2}{2} \\ &= \frac{2}{\hat{\sigma}_n^2} \cdot (\hat{\mu}_1 + \hat{\mu}_2) = \mathrm{L}_{\mathrm{E,SIMO}}. \end{aligned} \tag{3.64}$$

3.5 Chapter Summary

In this chapter, concepts suitable for iterative MIMO detection methods are presented. Besides the optimum APP detector, sub-optimal MIMO detectors are discussed and their performance is evaluated for different modulation orders and MIMO constellations within an AWGN channel. A novel MIMO detector, termed Gaussian tree search, is introduced, which offers a flexible trade-off between computational complexity and performance. Additionally, the Gaussian tree search integrates well in the proposed graph-based receiver concept, in contrast to classical MIMO tree search detection methods based on QR decomposition. Subsequently, pilot-based channel estimation methods and popular pilot grids are briefly described. The multi-dimensional sampling theorem is discussed, which states the maximum spacing of pilots up to which a bandlimited signal can be perfectly reconstructed. Finally, general methods for the combining of correlated random measures are introduced. By taking the correlation into account, the reliability of soft information is calculated more precisely. As a result, a receiver achieves a higher estimation accuracy and improves its overall performance. Often the convergence speed can be improved since more reliable estimates are obtained, which yields an additional gain in terms of complexity.

4

Particle Swarm Optimization

ADVANCED iterative receiver concepts offer a performance close to the optimum performance, often at reduced complexity. Precise channel state information is required to obtain this performance. Providing accurate initial channel state information is of importance since it has a major influence on the attainable performance and the complexity of the following receiver algorithm. At the same time, the complexity of the initial estimator should be sufficiently low as to keep the overall receiver complexity at a minimum.

Generally, channel estimation can be interpreted as an optimization problem, e.g. to minimize the Euclidean distance between the estimated and the true channel coefficients. The straightforward solution to this problem incorporates matrix inversion and leads to the well-known LS and/or MMSE estimator, as discussed in Section 3.2.

Evolutionary algorithms (EA), in particular particle swarm optimization (PSO), have recently received much attention. PSO is famous for its simple implementation and good performance in solving global optimization problems. Being a heuristic approach, it does not require any additional information of the optimization problem [BR03], which is particularly interesting for initial channel estimation. PSO has been successfully applied to a wide range of technical optimization problems, including channel estimation. However, most publications in the area of digital communications ignore the improvements developed by the EA community. Therefore, in Section 4.1, a general overview of PSO is given as well as improvements, which provide good performances without problem-specific fine-tuning. A cooperative approach to PSO is given, which mitigates the problem of early convergence in high-dimensional problems. Although PSO only has a few parameters that need to be adjusted for a specific optimization problem, an alternative approach with only one parameter, termed bare bone PSO, is presented. Additionally, a generalization to multimodal objective functions is examined with the multi-objective PSO. Cooperative approaches are proposed for MIMO channel estimation in Section 4.1.2 as well as the multi-objective PSO for a MIMO system with a time-varying frequency-selective channel in Section 4.2.2. Instead of determining the maximum number of iterations needed for convergence empirically, a semi-analytical method is developed, which enables the evaluation of the complexity for a wide range of parameters. Knowledge of the required number of iterations is essential for a practical receiver design since iterations have a significant impact on the overall complexity. A detailed discussion about the complexity of PSO is given in Section 4.3. Additionally, possible applications w.r.t. performance and complexity are outlined. The conclusions of this chapter are drawn in Section 4.4.

4.1 General Overview

Heuristic, nature-inspired algorithms, such as particle swarm optimization (PSO) [KE95, KE01] or genetic algorithms (GAO) [Gol89, ACH05], are attractive solutions to facilitate low-complexity MIMO channel estimation. According to the no-free-lunch theorem [WM97], an arbitrary optimizer can be tuned to precisely solve a specific optimization problem. However, averaged over a wide range of optimization problems, different strategies come to similar results. It is thus not a matter of which optimizer is chosen, but instead how easily it can be adapted to the underlying optimization problem. Out of the vast amount of global optimizers, evolutionary algorithms (EAs) stand out due to their simple and elegant approach of mimicking nature to solve technical optimization problems. A few algorithms dominate the class of evolutionary algorithms. Amongst others, particle swarm optimization and genetic algorithms are particularly popular.

PSO is a population-based heuristic global optimization algorithm, which originated in modeling the social behavior of bird flocks and fish schools. It has been applied to a variety of technical optimization problems, including channel and parameter estimation [GLHL07, TJK07, BTA^{+}07, DLL08, SBKH10, CHH^{+}10] as well as data detection [PCYH09] and multiuser detection [SSC^{+}07]. Unfortunately, a fair evaluation of PSO is rather difficult due to the wide range of available modifications and the fact that the algorithm is often tuned to optimum performance for a specific optimization problem by empirical measures.

Genetic algorithms are inspired by natural evolution. Accordingly, population members are termed chromosomes. Based on an optimization metric, a subset of chromosomes is selected to breed a new generation, which are subsequently used to generate a new generation by means of crossover and/or mutation. PSO and GAO share many similarities as both start with a randomly initialized population; both use a fitness value to evaluate their population members. The main difference lies within the selection of leaders (in terms of PSO) or parents (in terms of GAO) as well as the update of position and/or generation of new members, respectively. Population members within PSO are updated iteratively and influence themselves directly by their personal best position. On the contrary, population members in GAO pass characteristic information to their children. It is difficult to compare the performance of PSO and GAO in general as both depend on the specific optimization problem. Additionally, a similar variety of possible implementations exists also for GAO. However, several publications in the field of digital communications come to the conclusion that PSO is advantageous compared to GAO in terms of computa-

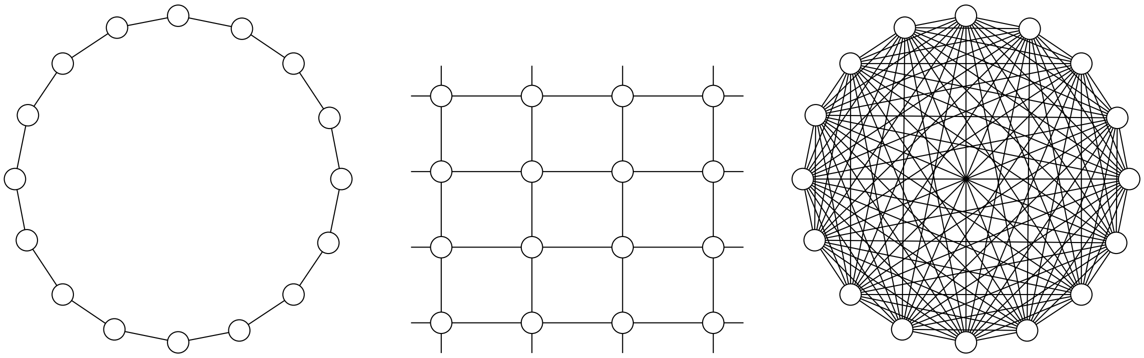

(a) Local neighborhood (lbest) (b) Von Neumann topology (c) Global neighborhood (gbest)

Figure 4.1: PSO neighborhood topologies.

tional complexity, convergence speed, and accuracy [D'O08, LS08, QJJ08]. Additionally, fewer parameters need to be set for the PSO algorithm.

4.1.1 PSO

The standard PSO is described by Algorithm 1. Initially, all N_p particles of a swarm are randomly positioned throughout the feasible search region $[\boldsymbol{S}_\mathrm{min}, \boldsymbol{S}_\mathrm{max}]$, where $\boldsymbol{S} \in \mathbb{R}^D$. Hence, the particles of a swarm "fly" through a D-dimensional search space, which is gradually explored by adjusting the trajectory of each particle at each iteration. Within each iteration, the current position of a particle $\boldsymbol{p}_i = [p_1, \ldots, p_D]$ is used as a candidate solution for the optimization metric termed fitness function. The fitness value of a particle is distributed to all particles within a so-called *neighborhood.* Different neighborhood topologies have been evaluated in [MKN03, HM04]. Among a large variety, three popular topologies are shown in Figure 4.1, namely the local neighborhood or *lbest* topology (Fig. 4.1a), the von Neumann topology (Fig. 4.1b), and the global neighborhood or *gbest* topology (Fig. 4.1c). From left to right, the number of neighbors is increasing, starting from two neighbors (lbest), to four neighbors (von Neumann), and finally to all particles within a swarm (gbest). The number of neighbors defines the social interaction between particles, i.e. the more particles are associated to a certain neighborhood, the faster this neighborhood will converge to a local/global optimum. Accordingly, the topology offers a trade-off between *exploration* and *exploitation.* Hereby exploration refers to the situation in which particles cover a wide search region, whereas exploitation means that particles are searching within a very narrow region, thus, converging to the global optimum. The performance of PSO with a given topology depends on the specific optimization problem. In general, the larger the neighborhood the faster the swarm will converge. Hence, in some optimization problems, the global neighborhood suffers from early convergence. Under these circumstances, a trade-off between performance and complexity (similar to Chapter 3.1.4) can be identified, since more iterations result in a larger computational complexity. Fortunately, in the case of MIMO channel estimation, the global neighborhood topology offers good performance and converges fast, as will be shown in Section 4.2.

Accordingly, a fully connected swarm is used in the following. As a result, the previous best position of a particle is termed personal best $\boldsymbol{p}_i^{\mathrm{IB}}$, whereas the previous best position of the swarm is called global best $\boldsymbol{p}^{\mathrm{GB}}$. The velocity vector of a particle i is updated according to [SE98, KE01]:

$$\boldsymbol{v}_i' = \omega \boldsymbol{v}_i + c_1 \boldsymbol{\varepsilon}_1 \circ \left(\boldsymbol{p}_i^{\mathrm{IB}} - \boldsymbol{p}_i\right) + c_2 \boldsymbol{\varepsilon}_2 \circ \left(\boldsymbol{p}^{\mathrm{GB}} - \boldsymbol{p}_i\right), \tag{4.1}$$

where $\circ$ denotes the entrywise product. The variables $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ denote random numbers in the range of [0,1]. The inertia weight ω typically decreases from 0.9 to 0.4 over the course of iterations. The social and cognitive parameters c_1 and c_2 are acceleration coefficients towards the personal and/or global best position, respectively. The velocity vector of a particle is, similar to the search space, restricted within certain boundaries $[\boldsymbol{v}_{\min}, \boldsymbol{v}_{\max}]$. Particles which are beyond the boundaries of the search space and velocity are reset to the corresponding boundary limits.

The update function (4.1) was published in 1998 in [SE98] as part of an already revised version of the PSO algorithm. The original update function of PSO published in 1995 [KE95] did not include the inertia weight or the cognitive and social parameters. Since then, an overwhelming amount of variations have been proposed. However, no standard algorithm or set of parameters has yet emerged, which delivers optimum performance independent of the optimization problem. Hence, parameters are tuned for each specific problem and settings determined by means of empirical measures are often applied. For example, the authors of [BE07] propose a so-called standardized version of PSO, which incorporates several general applicable improvements, i.e. bound handling, swarm size, and an update equation replacing the inertia weight with a constriction factor. The standardized version improves the performance for most optimization problems compared to the original version. In this thesis, only general applicable optimizations for PSO are considered. Although adaptive versions [ZZLC09] are also able to improve the performance of the standard PSO, their parameters typically need to be optimized for each optimization problem.

The update rule of the standardized version based on the constriction factor χ is given

Algorithm 1 Standard PSO algorithm.

```
Initialize swarm
Locate leader
$i = 1$
while $i < i_{\max}$ or convergence do
  for each particle do
    Update velocity and position using (4.1) / (4.2), (4.4)
    Evaluation using (4.5)
    Update pBest
    Update leader
  end for
  $i$++
end while
```

by

$$\boldsymbol{v}_i' = \chi \cdot \left\{ \boldsymbol{v}_i + c_1 \boldsymbol{\varepsilon}_1 \circ \left(\boldsymbol{p}_i^{\mathrm{IB}} - \boldsymbol{p}_i \right) + c_2 \boldsymbol{\varepsilon}_2 \circ \left(\boldsymbol{p}^{\mathrm{GB}} - \boldsymbol{p}_i \right) \right\}, \tag{4.2}$$

with

$$\chi = \frac{2}{|2 - \varphi - \sqrt{\varphi^2 - 4 \cdot \varphi}|}, \tag{4.3}$$

where $\varphi = c_1 + c_2$, $\varphi > 4$. The factors c_1 and c_2 are constraints on the velocity towards the global and the personal best position. According to [CD01], suitable values for a wide range of test functions are as follows: $c_1 = 2.8$ and $c_2 = 1.3$, which results in $\chi \approx 0.7298$. The standardized update function (4.2) as well as the above mentioned parameters are applied throughout all simulations. The position of a particle is updated subsequently according to

$$\boldsymbol{p}_i' = \boldsymbol{p}_i + \boldsymbol{v}_i'. \tag{4.4}$$

The updated velocity vector $\boldsymbol{v}_i'$ is added to the current position $\boldsymbol{p}_i$ of a particle. The new position $\boldsymbol{p}_i'$ is used as a candidate solution for the optimization metric. The optimization performed by PSO is described by

$$p^{\mathrm{OPT}} = \arg \min_{\boldsymbol{p}_i} f(\boldsymbol{p}_i). \tag{4.5}$$

The fitness function $f(\cdot)$ denotes the problem-specific optimization metric and is discussed in more detail in Section 4.2. The so far emerged personal and/or global best $\boldsymbol{p}_i^{\mathrm{IB}}$ and $\boldsymbol{p}^{\mathrm{GB}}$, respectively, are replaced by the updated position $\boldsymbol{p}_i'$, if the fitness value p^{OPT} is improved compared to the values of the personal and the global best position. This procedure is repeated until PSO has converged or the maximum number of iterations $i_{\max}$ is reached. The parameter $i_{\max}$ is chosen to be sufficiently large to prevent that the algorithm is stopped before the global optimum could be found. Frequently, the optimum solution is found within a fraction of $i_{\max}$. Therefore, a stopping criterion is necessary to reduce the average number of iterations needed for convergence. The probability of a rapid improvement of the fitness value during the first few iterations is comparably large. However, a further improvement of the fitness value is not directly related to the iterations but happens spontaneously, i.e. the fitness value can remain constant for a certain number of iterations and then continue to improve with further iterations. In general, the fitness improvement over iterations is a suitable stopping criterion [ZL07]. However, due to the nonlinear behavior of the fitness value w.r.t. iterations, the fitness improvement should reach a threshold t_{h} and maintain this value for γ iterations. An overview of alternative suitable stopping criteria for PSO is given in [ZL07].

In case PSO converges, all particles $\boldsymbol{p}$ of the swarm are located at the same position, which minimizes (4.5). Without loss of generality, only minimization problems are considered.

4.1.2 Cooperative PSO

In general, population-based optimization algorithms are searching for a small, specified volume in a D-dimensional search space, surrounding the global optimum. In order to converge to the global optimum, an optimization algorithm needs to create a sample within this region. The probability of generating a sample within the region is the volume of the region divided by the volume of the search space [vdBE04]. This probability decreases exponentially with increasing dimensionality of the search space. This effect is often termed "curse of dimensionality". PSO is known to perform rather poor for high-dimensional problems. A large variety of solutions is proposed to solve this problem. In [Hen09], the update function (4.1) is changed to take adaptive parameters into account. These parameters are changed over the course of iterations and improve the converge behavior of the PSO algorithm. However, the optimum set of parameters remains problem dependent. An alternative solution to improve the performance of the original PSO algorithm is given by a so-called cooperative approach to particle swarm optimization (CPSO) presented in [vdBE04]. The CPSO approach relies on the original update equation and is described in the following. The pseudocode describing CPSO is given by Algorithm 2. The N_p particles of the PSO swarm are now separated into N_s swarms with N'_p particles. The number of particles for both PSO and CPSO should be chosen within a certain range. Too few particles ($N_\mathrm{p}, N'_\mathrm{p} < 5$) lead to a deteriorated performance, while too many are not able to improve the performance ($N_\mathrm{p}, N'_\mathrm{p} > 100$). About 30 particles is a good trade-off between complexity and performance for a majority of optimization problems [CD01]. For CPSO the overall number of particles is important. Only a limited amount of particles per subswarm is required in combination with many subswarms and vice versa.

Separating the high-dimensional search space into sets of smaller dimension improves the performance significantly, given that the dimensions are mutually independent. Accordingly, the D-dimensional problem is split into $N_\mathrm{s} = D/\delta$ subsets and optimized sepa-

Algorithm 2 Cooperative PSO algorithm.

```
Initialize $N_s$ swarms with $N'_\mathrm{p}$ particles
Locate leader
$i = 1$
while $i < i_\mathrm{max}$ or convergence do
  for each swarm do
    for each particle do
      Update position using (4.1) / (4.2), (4.4)
      Evaluation using (4.6)
      Update pBest
      Update leader
    end for
  end for
  $i$++
end while
```

rately by an individual swarm of particles $\boldsymbol{s}=[s_1, \ldots, s_{N_\mathrm{s}}]$, where δ is the number of dimensions for each swarm. The position of a particle i of swarm s is given by $\boldsymbol{p}_{s,i} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_\delta]$. The separation of the dimensions mitigates a drawback of the standard PSO algorithm: Since the standard PSO considers the full-dimensional vector in the update function, it allows that some dimensions move further away from the solution as long as the overall fitness value is improved. On the contrary, cooperative PSO is evaluating subsets of the D-dimensional vector. The probability that single components are deteriorated in favor of other dimensions is thus reduced.

If only one swarm is available, $N_\mathrm{s} = 1$, CPSO is equivalent to PSO since all dimensions are optimized by one swarm. In case of $N_\mathrm{s} > 1$, the evaluation of the optimization metric is no longer directly possible since a particle represents only a subset of dimensions of the optimization problem. Consequently, a context-vector $\boldsymbol{\phi}_{s,i}$ is necessary. In order to construct a D-dimensional vector, the $D-\delta$ missing dimensions are replaced by the global best positions of the remaining swarms: $\boldsymbol{\phi}_{s,i} = [\boldsymbol{p}_1^{\mathrm{GB}} \ldots \boldsymbol{p}_{s,i} \ldots \boldsymbol{p}_{D/\delta}^{\mathrm{GB}}]$. The optimization function (4.5) is changed accordingly:

$$p^{\mathrm{OPT}} = \arg \min_{\boldsymbol{\phi}_{s,i}} f\left(\boldsymbol{\phi}_{s,i}\right). \tag{4.6}$$

As said before, due to the change of the optimization function and the introduction of a context-vector, a central problem of the standard PSO is addressed, which allows the performance of one or more dimensions to be sacrificed as long as the overall fitness value is improved. An update of the global best position of a subswarm has to improve the fitness value with the given context vector. Hence, an improvement of the fitness value only affects the assigned dimensions of a subswarm. However, the separation of a D-dimensional problem into δ subdimensions has to be considered individually for each optimization function. In case subdimensions are correlated with each other, a separation might introduce so-called pseudo-minima, which again results in a premature convergence of the algorithm.

4.1.3 Bare Bones PSO

Despite the simplicity of the concept of PSO, it remains difficult to exactly describe the statistical behavior of PSO as well as the influence of each of the components. In order to simplify the algorithm and to facilitate a deeper understanding of the functional interaction between parameters, the so-called bare bones PSO (BBPSO) has been proposed by Kennedy in [Ken03]. The conducted analysis proposes to drop the velocity term and rely solely on the personal and global best position for a position update. Gaussian distributed random variables are generated, which provide the particle's updated position according to

$$\boldsymbol{p}_i' = \mathcal{N}\left(\frac{\boldsymbol{p}_i^{\mathrm{IB}} + \boldsymbol{p}^{\mathrm{GB}}}{2}, |\boldsymbol{p}_i^{\mathrm{IB}} - \boldsymbol{p}^{\mathrm{GB}}|\right). \tag{4.7}$$

As can be seen, the random variables are centered around a mean value $(\boldsymbol{p}_i^{\mathrm{IB}} + \boldsymbol{p}^{\mathrm{GB}})/2$ with a standard deviation of $|\boldsymbol{p}_i^{\mathrm{IB}} - \boldsymbol{p}^{\mathrm{GB}}|$. The standard deviation can be interpreted as a step size of a particle and is largest when the distance between the particle's personal best position and the global best position is large as well. A later study in [vdBE06] could prove that PSO will eventually converge to a weighted combination of the personal and global best, respectively:

$$\boldsymbol{p}_i' = \frac{c_1 \cdot \boldsymbol{p}_i^{\mathrm{IB}} + c_2 \cdot \boldsymbol{p}^{\mathrm{GB}}}{c_1 + c_2}. \tag{4.8}$$

In the concept of bare bones PSO, the weighting factors are set to $c_1 = c_2 = 1$. Similarly to PSO, bare bones PSO suffers from a premature convergence when handling multi-modal optimization problems [YH13]. Since bare bones PSO only replaces the conventional position update, it is easily integrated within the structural concept of cooperative PSO. A hybrid version which combines CPSO and BBPSO, termed CBBPSO, is proposed and evaluated in this thesis.

4.1.4 Multi-Objective PSO

As mentioned previously, the PSO algorithm including its variants is designed to solve single-objective optimization problems, i.e. an optimization metric is minimized with one specific solution $\boldsymbol{p}^{\mathrm{GB}}$. In the context of channel estimation, this scenario reflects a flat-fading channel. However, in case of a time-varying frequency-selective channel, there is not a single global solution minimizing the optimization function. More specifically, there is not only one optimization function but rather one specific function for each time and frequency index. Since OFDM is considered within this thesis, a straightforward solution is to apply a conventional PSO for each OFDM subcarrier, which is justified due to the absence of ISI, such that a direct separation of each channel coefficient is enabled. Obviously, this solution is inefficient, since a conventional PSO swarm needs a sufficient number of particles. Accordingly, this approach leads to a very high computational complexity, which scales linearly with the number of channel coefficients that need to be estimated.

A more promising solution is given by the so-called multi-objective PSO (MOPSO) proposed in [CL02, CPL04]. In a multi-objective optimization problem, $\mathit{\Lambda}$ objectives need to be minimized simultaneously:

$$p^{\mathrm{OPT}}[\lambda] = \arg \min_{\xi[\lambda]} f_\lambda \left(\xi[\lambda]\right), \quad 1 \leq \lambda \leq \mathit{\Lambda}. \tag{4.9}$$

In case of pilot-based channel estimation $\mathit{\Lambda}$ corresponds to the number of pilots. As can be seen from (4.9), $\mathit{\Lambda}$ fitness functions need to be optimized, whereas each function yields a unique solution, which minimizes the optimization metric for one specific OFDM subcarrier and/or OFDM symbol. As a matter of fact, the solution $\xi[\lambda]$ for the λth objective is not necessarily optimal for the ($\lambda\pm1$)th objective, due to e.g. fading in time and/or frequency. This means that one objective cannot be optimized without sacrificing the performance of at least one other objective. The MOPSO algorithm optimizes all objectives in parallel, where the so-called *nondominated* solutions are stored in an external archive. A particle ξ is said to dominate another particle ξ', denoted as $\xi \prec \xi'$, if and only if

$$\forall \lambda \in \{1, \ldots, \Lambda\} : f_\lambda(\xi[\lambda]) \leq f_\lambda(\xi'[\lambda]), \tag{4.10a}$$

$$\exists \lambda \in \{1, \ldots, \Lambda\} : f_\lambda(\xi[\lambda]) < f_\lambda(\xi'[\lambda]). \tag{4.10b}$$

Accordingly, a particle is stored within the archive if it provides a better solution in at least one objective without sacrificing the remaining objectives. In each iteration, a new solution is compared to the solutions in the archive. If a solution in the archive is dominated it will be replaced by the new solution. This set of solutions is termed Pareto set $F^\star$ and contains all particles ξ that are not dominated by another particle ξ':

$$F^\star \doteq \left\{ \xi \in \mathbb{R}^D | \nexists \xi' \in \mathbb{R}^D : \xi' \succ \xi \right\}. \tag{4.11}$$

One issue with the maintenance of an archive is that its size can increase to infinity, since solutions along the Pareto front can be infinitesimal close to each other. An intuitive solution to limit the size of an archive is given by the introduction of the so-called ϵ-dominance [LTDZ02]. As a result, particles are only added to the archive if they dominate another particle by a factor of ϵ. Optimally, particles are equi-distantly spaced with a factor of ϵ along the Pareto front. The maintenance of the archive as well as the Pareto front is explained in more detail in Example 4.1. Furthermore, in order to ensure that the particles within a swarm are not optimizing only one objective, diversity of the Pareto optimal solutions within the archive has to be maintained [RSC05, RSC06]. The selection of a leader is therefore of great importance. In contrast to single-objective PSO, this cannot be simply the local and/or global best position. One solution is to choose all entries from the archive as potential leaders, however, this increases the probability of early convergence. The amount of alternative solutions for the selection of a leader is again overwhelming. Popular methods are crowding [DPAM02] and the so-called sigma method [MT03]. In the former approach, the search space is divided into equally sized regions, whereas the particles within the least crowded region are favorable compared to other regions. For the latter, a so-called sigma value is calculated, which depends on how well the particle performs throughout all dimensions. A good overview of state-of-the-art methods for multi-objective PSO variants is given in [RSC06].

> **Example 4.1** This example illustrates the principle of Pareto optimal solutions on the basis of a two-dimensional optimization function. It is assumed that a minimum for the functions $f_1(\cdot)$ and $f_2(\cdot)$ has to be found. However, the two functions are contradictory, that is a value which minimizes $f_1(\cdot)$ does not minimize $f_2(\cdot)$. In fact, the optimum value minimizing $f_1(\cdot)$ maximizes $f_2(\cdot)$ and vice versa, as can be seen in Figure 4.2. Solutions along the Pareto front offer the optimum solution given a predefined trade-off.

A major drawback of the majority of multi-objective evolutionary algorithms is that they do not scale well with the number of objectives [ITN08, SLC11]. That means that the performance of the optimization algorithms degrades significantly when the number of objectives exceed a value of 6 to 7 objectives [PY07]. For channel estimation, this limit is easily surpassed in case the number of objectives correspond to the number of pilots.

Algorithm 3 Multi-objective PSO algorithm.

```
Initialize swarm
Locate leader in an external archive
i=1
while $i < i_{\max}$ or convergence do
    for each particle do
        Select leader from archive
        Update position using (4.1) / (4.2) , (4.4)
        Mutation
        Evaluation using (4.9)
        Update pBest
    end for
    Update leaders in the external archive
    i++
end while
```

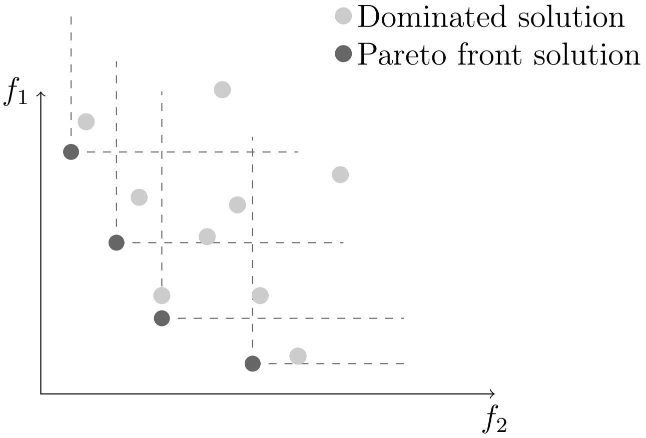


Figure 4.2: Set of Pareto optimum solutions.

In order to overcome this drawback, it is proposed in this thesis, that every entry inside the archive corresponds to one specific channel coefficient and inherently one objective. Hence, the number of particles within a MOPSO swarm and the number of particles stored within the archive are equal to the number of objectives. Partitioning the swarm to the number of objectives is related to the concept of cooperative PSO, where the swarm is partitioned into different dimensions. Unlike dimensions however, objectives—i.e. coefficients in time and/or frequency—are correlated, thus, a particle which serves as a leader for one objective is, at the same time, a good choice for other objectives as well. The problem of premature convergence, known for CPSO as well as conventional multi-objective EAs, is mitigated with this approach. As an additional advantage, this approach simplifies the selection of a leader significantly, since a particle only has to select the corresponding entry within the archive. Hence, one particle always has exactly one leader. The maintenance of the archive is described in more detail in Example 4.2.

After the selection of its leader, a particle has to update its position. The update function remains unchanged compared to PSO/CPSO. An additional mutation operator is recommended, because the MOPSO algorithm occasionally converges prematurely [CPL04]. Hereby, a randomly chosen dimension d_{rand} of a particle is reinitialized within newly determined boundaries which depend on a mutation rate m_r. The mutation range is determined as follows [CL02]:

$$\boldsymbol{S}_{\mathrm{mut}}[d_{\mathrm{rand}}] = (\boldsymbol{S}_{\mathrm{max}}[d_{\mathrm{rand}}] - \boldsymbol{S}_{\mathrm{min}}[d_{\mathrm{rand}}]) \cdot (1 - i/i_{\mathrm{max}})^{5/m_r}, \tag{4.12}$$

where m_r is commonly chosen to be 0.5. The dimensions of the search space with the mutation operator are resized according to

$$\boldsymbol{S}_{\mathrm{max}}[d_{\mathrm{rand}}] = \boldsymbol{p}_i[d_{\mathrm{rand}}] + \boldsymbol{S}_{\mathrm{mut}}[d_{\mathrm{rand}}] \tag{4.13}$$

$$\boldsymbol{S}_{\mathrm{min}}[d_{\mathrm{rand}}] = \boldsymbol{p}_i[d_{\mathrm{rand}}] - \boldsymbol{S}_{\mathrm{mut}}[d_{\mathrm{rand}}]. \tag{4.14}$$

The chosen dimension of the particle is randomly reset within these boundaries. The effect of the mutation range is strongest in the beginning of iterations and its influence decreases exponentially over the course of iterations. The principle of MOPSO is summarized by Algorithm 3. The maintenance of the archive and the additional mutation operator contribute to an increased complexity of the algorithm compared to the conventional PSO. Nevertheless, it is shown in Section 4.2.2 that MOPSO converges fast to a reasonable MSE. It is thus ideally suited to provide initial channel state information. In the following section, the performance of the PSO algorithm is evaluated with a flat-fading channel and a large number of antennas, related to the scenario of large-scale antenna implementations (cf. Section 2.3.3). Furthermore, MOPSO is tested in a time-varying frequency-selective channel. Four different scenarios of the WINNER channel model are considered, which provide different levels of diversity in frequency and/or time.

> **Example 4.2** This example illustrates the evolution of the archive which contains the non-dominated solutions. For this example, it is assumed that five objectives need to be optimized. Accordingly, the archive as well as the MOPSO swarm is composed of five particles. After each iteration, the particles are evaluated and stored within the archive if they provide a non-dominated solution for an objective. This procedure is illustrated in Table 4.1. During the initialization phase ($i = 0$), the archive is created but contains no solutions yet. It can be seen that after the first iteration ($i = 1$), the particle $\xi[0]$ dominates the remaining four particles. As a result, it will be used as a leader for all objectives. In following iterations, the particles of the swarm will be influenced by their corresponding entries inside the archive. During the exploration of the search space, particles will approach their optimum solution from different directions. It is therefore likely, that particles contribute solutions to different objectives apart from their assigned one. When the maximum number of iterations is reached ($i = i_{\mathrm{max}}$), a particle should provide the non-dominated solution of its objective, since this inherently means that the particle is converged to the optimum position. While this approach enables the convergence with many objectives, it is a strict requirement that solutions are sufficiently correlated. However, this assumption is typically fulfilled, when MOPSO is applied to channel estimation in time and/or frequency domain. See also Figure 4.6, which additionally illustrates the convergence of particles towards their Pareto optimum solution.

Table 4.1: Evolution of the archive as a function of iterations.

	$\xi[0]$	$\xi[1]$	$\xi[1]$		$\xi[0]$
	$\xi[0]$	$\xi[1]$	$\xi[1]$		$\xi[1]$
	$\xi[0]$	$\xi[0]$	$\xi[0]$	$\dots$	$\xi[2]$
	$\xi[0]$	$\xi[0]$	$\xi[0]$		$\xi[3]$
	$\xi[0]$	$\xi[0]$	$\xi[2]$		$\xi[4]$
$i = 0$	$i = 1$	$i = 2$	$i = 3$	$\dots$	$i = i_{\max}$

4.2 Performance Evaluation

The performance of PSO for MIMO channel estimation is evaluated in this section. Naturally for PSO and its single-objective variants, a quasi time-invariant channel is assumed in Section 4.2.1. Hereby, all particles aim to find one global objective within the multi-dimensional search space. The scenarios evaluated are related to modern large-scale antenna implementations. The effectiveness of the proposed solutions is illustrated in terms of MSE results.

Additionally, the multi-objective PSO is applied for two-dimensional channel estimation with a time-varying frequency-selective channel model in Section 4.2.2. Four different WINNER C2 channel models are used. A further performance improvement and the application to MIMO systems is achieved by implementing linear prediction into the fitness function.

4.2.1 Flat-Fading Channel

The MIMO system under consideration consists of N_{T} transmit and N_{R} receive antennas. The received signal vector at time index k, $\boldsymbol{y}[k] \in \mathbb{C}^{N_{\mathrm{R}} \times 1}$, is modeled as

$$\boldsymbol{y}[k] = \boldsymbol{H}\boldsymbol{x}[k] + \boldsymbol{n}[k], \tag{4.15}$$

where $\boldsymbol{x}[k] \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ is the transmitted signal vector at time index k. Without loss of generality, only the time domain is considered in this section. The application to the frequency domain is straightforward. The generalization to a fading channel in time and frequency is presented in the subsequent section. The entries of the channel matrix $\boldsymbol{H} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ are assumed to be independent and identically distributed (i.i.d.) according to $\mathcal{CN}(0, 1)$. Furthermore, $\boldsymbol{n}[k]$ denotes the noise vector at the receiver whose entries are i.i.d. modeled as $\mathcal{CN}(0, \sigma_n^2)$.

Pilots are transmitted to support pilot-aided channel estimation (PACE). Stacked in a matrix, the transmit vector $\boldsymbol{x}[k]$ can be written as $\boldsymbol{X} \in \mathbb{C}^{N_{\mathrm{T}} \times K}$. A minimum of $K = N_{\mathrm{T}}$ pilots are transmitted to ensure a full rank. The pilot matrix consists of orthogonal sequences subject to $\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}} = \mu \boldsymbol{I}_{N_{\mathrm{T}}}$,where μ is related to the signal power assigned to the pilot matrix [HH03].

In the following, it is assumed that the transmit vector $\boldsymbol{x}[k]$ of length $K = N_{\mathrm{T}}$ consists of pilots only. In case of a quasi-invariant (block-fading) channel, the maximum-likelihood metric (fitness function) for PSO can be written as follows:

$$f(\boldsymbol{p}_i) = \sum_{k=1}^{K=N_\mathrm{T}} \left\| \boldsymbol{y}[k] - \boldsymbol{P}_i \boldsymbol{x}[k] \right\|^2 . \tag{4.16}$$

The position of the ith particle $\boldsymbol{p}_i$ is used as a potential solution for the ML metric. For a consistent notation in line with (4.15), the previously used vector notation of the position of the particle is changed here to a matrix notation with $\boldsymbol{P}_i \in \mathbb{C}^{N_\mathrm{R} \times N_\mathrm{T}}$. Thus, a position of a particle represents a hypothesis of the channel matrix $\tilde{\boldsymbol{H}}$. It is of importance to note that each dimension of a particle is real-valued. Since a particle needs to estimate $N_\mathrm{R} \times N_\mathrm{T}$ complex-valued channel coefficients, the dimensions of the real-valued search space results in $D = 2 \cdot N_\mathrm{T} \cdot N_\mathrm{R}$.

The ML metric for CPSO is very similar to the PSO metric. As previously mentioned, a context vector is required for CPSO, since a single swarm does not longer estimate all dimensions of the channel matrix. The distributed dimensions are individually optimized by cooperative swarms. During iterations, the preliminary results are collected from each swarm as explained in Section 4.1.1. Instead of using a context vector $\boldsymbol{\phi}_{s,i}$, a context matrix $\boldsymbol{\Phi}_{s,i}$ is used for a consistent notation:

$$f(\boldsymbol{\Phi}_{s,i}) = \sum_{k=1}^{K=N_\mathrm{T}} \left\| \boldsymbol{y}[k] - \boldsymbol{\Phi}_{s,i} \boldsymbol{x}[k] \right\|^2 . \tag{4.17}$$

In case of MIMO channel estimation, $N_\mathrm{R} \cdot N_\mathrm{T}$ channel coefficients are estimated assuming a flat-fading time-invariant channel. As mentioned before, the performance of PSO deteriorates with increasing dimensions, which is determined by the number of transmit and receive antennas. One of the currently active research areas are large-scale antenna implementations with hundreds of antennas. Typical scenarios are referred to as *large-MIMO* [MZCR09] or *massive MIMO* [Mar10] depending on a symmetric or asymmetric antenna setup (cf. Section 2.3.3). In these scenarios, a large amount of channel coefficients needs to be estimated. The complexity of conventional channel estimation algorithms increases with the order of $\mathcal{O}\left(N_\mathrm{T}^3\right)$ in case of MMSE channel estimation. The application of PSO/CPSO to large-scale MIMO channel estimation is promising due to the omitted matrix inversion, which is typically applied for channel estimation (cf. Section 3.2) and thus, the inherent simplicity of calculating (4.16) and (4.17).

Figure 4.3 illustrates the difference between PSO and CPSO for channel estimation of a 2x2 MIMO system. PSO optimizes all channel coefficients with one swarm. CPSO is able to separate the $D = (2 \cdot N_\mathrm{R} \cdot N_\mathrm{T})$-dimensional problem into subsets and optimizes each subset with an individual swarm. In this example two swarms are shown, however, the number of possible subswarms is in the range of $N_\mathrm{s} = [1, D]$. In the case of $N_\mathrm{s} = D$ subswarms, a single swarm would optimize either the real or the imaginary part of one channel coefficient, whereas for $N_\mathrm{s} = 1$ CPSO is equivalent to PSO. While the number of subswarms N_s is directly related to the number of dimensions, there is no such relation for the number of particles. A minimum number of particles is needed for each subswarm in order to allow convergence. Additionally, the performance of both, PSO and CPSO, cannot be improved by increasing the number of particles once a threshold is reached.

The number of particles remain dependent on the optimization problem. A good trade-off between complexity per iteration and performance for PSO is to set the number of particles to $N_\mathrm{p} = 30$ [CD01]. For the cooperative approaches, a suitable number of particles per subswarm needs to be determined. The convergence speed for different numbers of particles for PSO/BBPSO as well as particles per swarm and numbers of subswarms for CPSO/CBBPSO is evaluated for an 8×1 MISO system in the following. As a matter of fact, the chosen antenna setup affects the complexity and the performance of the PSO algorithm. Hereby, three cases can be identified: (1) SIMO, (2) MISO, and (3) MIMO. The first case is considered to be trivial, since effectively N_R parallel SISO channels have to be estimated. The dimensionality of a particle is not necessarily affected, because PSO swarms could optimize the receive antennas in parallel, and more important, independent of each other. Such an independent optimization is not possible for the MISO case. Accordingly, the dimensionality of a particle is directly related to the number of transmit antennas. The MIMO setup is again a parallel optimization of N_R MISO setups. Hence, without loss of generality, the number of receive antennas is set to $N_\mathrm{R} = 1$. The important parameter for the complexity as well as performance is given by the dimensionality of the optimization problem, which, in this case, is given by the number of transmit antennas. The global best fitness value p^OPT is shown in Figure 4.4 as a function of iterations for different variants of particles and/or swarms at an SNR of 20 dB and $N_\mathrm{T} = 8$ transmit antennas. The overall number of particles is kept constant for PSO and CPSO to allow a fair comparison. Interestingly, the cooperative approaches perform better independent of the particle/swarm constellations. By comparing Figure 4.4b and Figure 4.4c, it can be seen that CPSO and CBBPSO converge faster with more subswarms. A further increase of the number of particles does not yield a significant gain w.r.t. the convergence speed (cf. Figure 4.4d). Concluding from these results, the number of subswarms is set to $N_\mathrm{s} = D/2 = N_\mathrm{T}$ with $N'_\mathrm{p} = 5$ particles per subswarm for CPSO and CBBPSO. Although these parameters are not optimized for one specific optimization problem, they show a good performance throughout all conducted simulations. Nevertheless, as stated by the no free lunch theorem, all parameters may be optimized for one specific problem while possibly deteriorating the performance for other problems.

With the chosen settings, MSE results as a function of SNR for the different PSO variants are given in Figure 4.5. Additionally, as a performance reference, the MSE

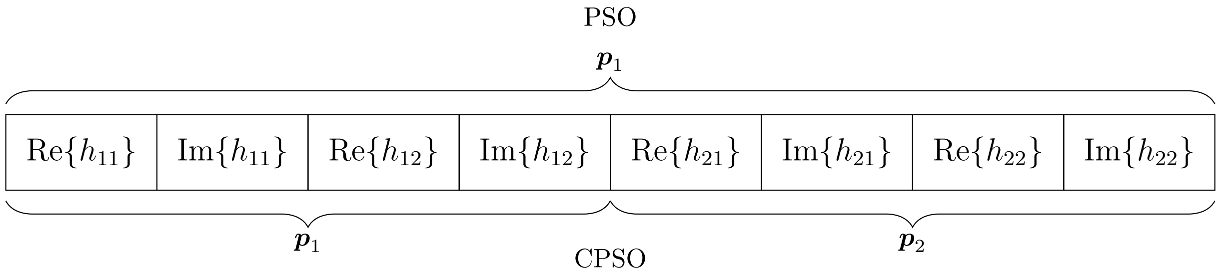
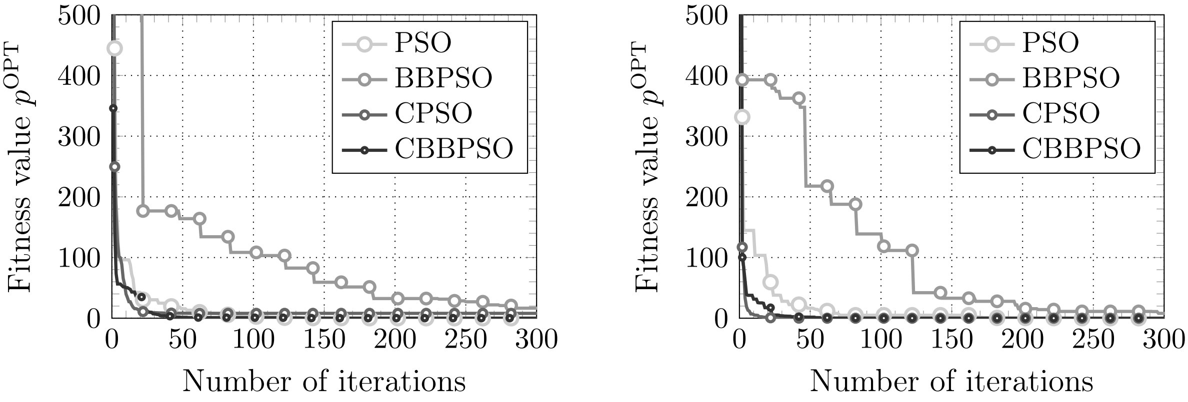
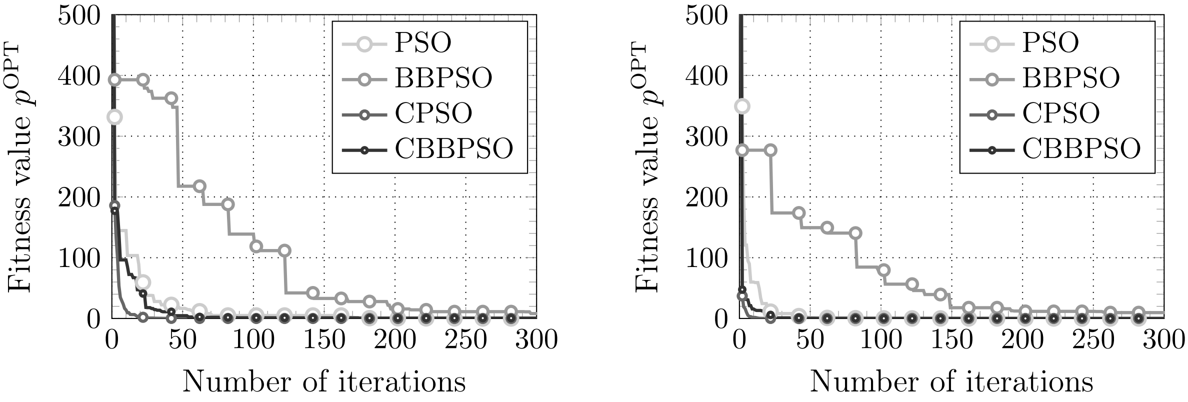


Figure 4.3: Possible separation of an 8-dimensional problem into a set of lower-dimensional problems by the CPSO compared to PSO.



(a) $N'_{\mathrm{p}} = 2$, $N_{\mathrm{s}} = 8$ and $N_{\mathrm{p}} = 16$

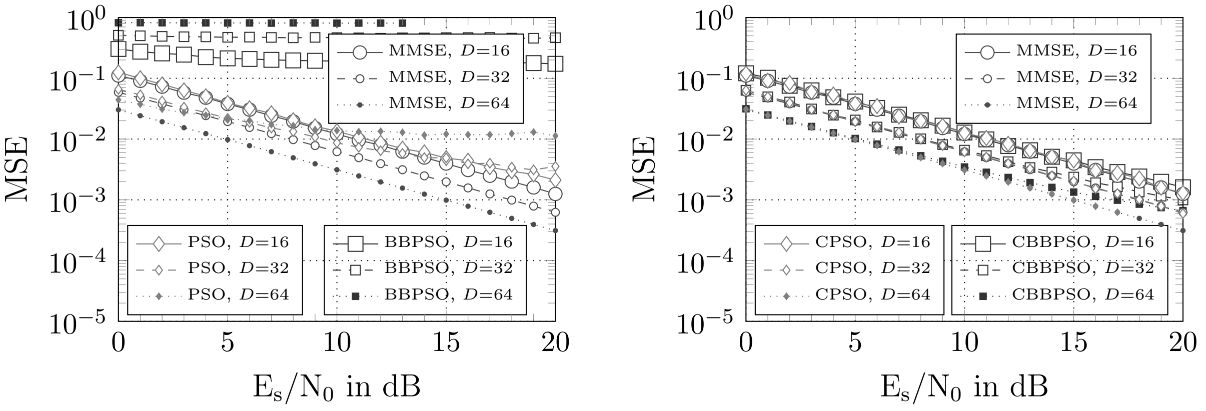
(b) $N'_{\mathrm{p}} = 5$, $N_{\mathrm{s}} = 8$ and $N_{\mathrm{p}} = 40$



(c) $N'_{\mathrm{p}} = 10$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{p}} = 40$

(d) $N'_{\mathrm{p}} = 10$, $N_{\mathrm{s}} = 8$ and $N_{\mathrm{p}} = 80$

Figure 4.4: Global best fitness value as a function of the number of iterations and the number of particles N'_{p} and swarms N_{s} at an SNR of 20 dB with $N_{\mathrm{T}} = 8$.



(a) Single-swarm PSO: PSO, BBPSO

(b) Multi-swarm PSO: CPSO, CBBPSO

Figure 4.5: MSE of PSO variants as a function of dimensions.

curves obtained by means of MMSE channel estimation are included. The single-swarm variants—PSO and BBPSO—are shown in Figure 4.5a. With increasing dimensions, PSO diverges from the MMSE performance and reaches an error floor of $3{\cdot}10^{-3}$ and $1{\cdot}10^{-2}$ for $D = 32$ and $D = 64$, respectively. A catastrophic performance is obtained by the bare bones PSO, which is unable to converge to the global optimum. This poor performance is caused by a limited maximum number of iterations of $i_{\max} = 2000$. As can be seen from previous results, shown in Figure 4.4, BBPSO is the slowest variant w.r.t. convergence and requires a large amount of iterations. This observation is in line with the MSE results. The cooperative approach to PSO and BBPSO yields significant gains for all dimensions. CPSO and CBBPSO converge to the global optimum and yield the same MSE performance as MMSE channel estimation. Still, the convergence speed is much slower for CBBPSO compared to CPSO. Although CBBPSO performs well for MIMO channel estimation, its increased complexity due to iterations renders it unsuitable for large-scale antenna implementations. While the assumption of a quasi-invariant channel is common for a massive MIMO system [RPL$^+$13, HtBD13], typically the channel is varying in time and/or frequency. In this case, PSO is rather inefficient. Hence, multi-objective PSO is evaluated in the following.

4.2.2 Time-Varying Frequency-Selective Channel

The major difference of MOPSO compared to PSO is the selection of a leader within the swarm. As mentioned before, an archive needs to be maintained, which contains the nondominated solutions used as leaders for the swarm. In this thesis, every entry within the archive corresponds to a channel coefficient of one OFDM subcarrier and/or OFDM symbol. Additionally, one particle is assigned to one entry of the archive. Since the channel coefficients are highly correlated in frequency and/or time, the topology of the swarm is adaptive and depends on the fading of the specific domain. In case of fast-fading, only a limited number of coefficients are sufficiently correlated, such that a solution for one channel coefficient can be used for its direct neighbors, which corresponds to the lbest topology. For slow-fading channels on the other hand, the topology is more related to the gbest structure (cf. Figure 4.1).

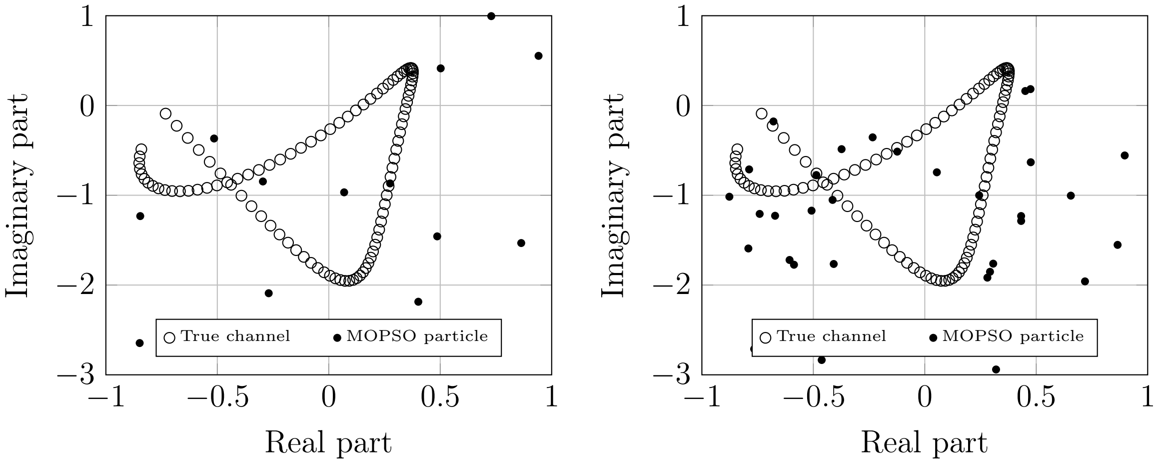
An extension of the single-objective fitness function of PSO (4.16), to a time-varying frequency-selective channel and a SISO setup is given by

$$f_\lambda\left(\xi[\lambda]\right) = \sum_{\lambda=1}^{\Lambda} \left\| y[\lambda] - \xi[\lambda]x[\lambda] \right\|^2 . \tag{4.18}$$

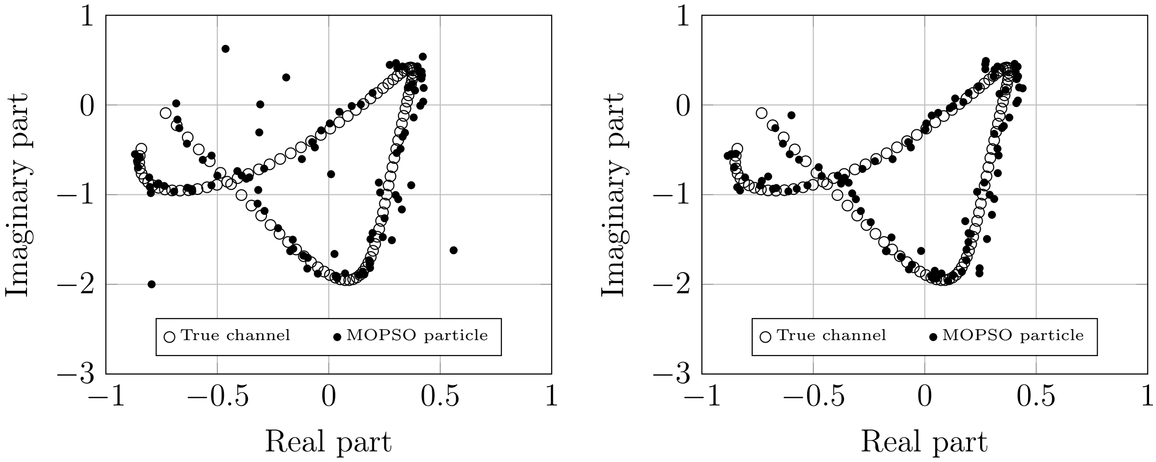
The nondominated solutions of the fitness function (4.18) are stored in the external archive. In case the MOPSO algorithm converges, the Pareto set $F^\star$ contains Λ possible solutions. The number of dimensions for one particle is set to $D = 2$, which means that a particle estimates one complex-valued channel coefficient. Obviously, the concept of cooperative PSO can be applied as well.

The movement or "flight" of particles during MOPSO iterations is shown in Figure 4.6. Herein, only a small section of the overall search space is highlighted. The true coefficients are depicted with white-filled markers. A total of 100 pilots are transmitted and correspondingly, 100 coefficients need to be estimated. Since the channel is varying with time and frequency, the markers change their positions, each indicating a different time and frequency position. Particles of the MOPSO swarm are shown in solid markers. Only a small subset of particles are initialized in the vicinity of the true coefficients, since their positions are initially unknown (cf. Figure 4.6a). Already with 10 iterations (cf. Figure 4.6b), particles are approaching the true coefficients from all directions, providing new entries within the archive and accordingly, attracting more particles of the swarm. As can be seen in Figure 4.6c, the majority of particles are close to the true channel coefficients. However, an offset of a few particles does not deteriorate the overall MSE performance. Effectively, such particles simply do not contribute to the Pareto set. Instead, other particles might have crossed their optimum positions. These positions are then used for the corresponding entries.

This behavior is further exemplified in Figure 4.7. On the right-hand side, the real term of the archive entries is plotted, together with the real channel coefficients as a function of OFDM subcarriers. The corresponding MSE is plotted on the right hand side. Only five individual particles contribute to the solutions in the Pareto set, as can be seen from the five distinct sections, shown in Figure 4.7a. Correspondingly, the MSE is around $1 \cdot 10^{-1}$ for the majority of OFDM subcarriers. The apparent poor decision of the archive entry for the OFDM subcarriers 75 to 81 is caused by the property of PSO to sacrifice one dimension in favor for an overall improvement. In this case, the imaginary part provides a better solution than the real dimension, which leads to an overall improved fitness value. With advancing iterations, the Pareto set gains in diversity, which in turn improves the MSE performance. After five iterations, the maximum MSE is at around $3 \cdot 10^{-2}$. Overall, the archive consists of 34 unique entries out of 100. With further iterations, more nondominated solutions will be added to the archive and the MSE performance will approach an average MSE of $1 \cdot 10^{-3}$. The fast convergence to a reasonable MSE is an ideal prerequisite for PACE. Especially for the initialization, a certain threshold needs to be met in order to allow convergence for a subsequent receiver (cf. Chapter 3.3.1). The MSE performance for the four different WINNER channel models is shown in Figure 4.8. The pilot sequence is distributed in time and frequency with 10 symbols for each domain. For these simulation results, a single-antenna system is employed. The reason for this selection will be explained in the following. The four selected channel scenarios feature different fading characteristics in frequency and are used up to their defined maximum velocities. The MSE performance of MOPSO is hereby independent of the fading in time and/or frequency and achieves the same performance as a least-squares channel estimator. This is not surprising, since the least-squares method is equal to the maximum likelihood criterion on the condition that the observation error is Gaussian distributed [Kay09]. More important is the fact, that within the variety of the WINNER channel models, the correlation among channel coefficients is sufficiently large such that particles assigned for one objective can be used as potential solutions for neighboring objectives. However, this result is not completely satisfying because the complexity of MOPSO is significantly larger than that of LS channel estimation and, furthermore, the fitness function (4.18) is not determined for a MIMO channel since only one observation is used for N_{T} transmit symbols. Hence, further improvement in terms of



(a) Iterations = 1 (b) Iterations = 10

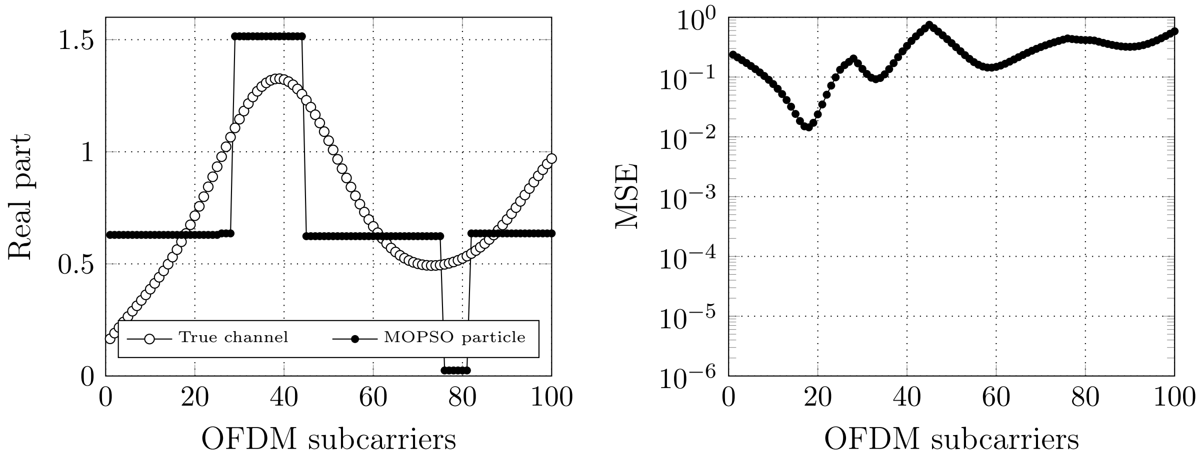


(c) Iterations = 50 (d) Iterations = 100

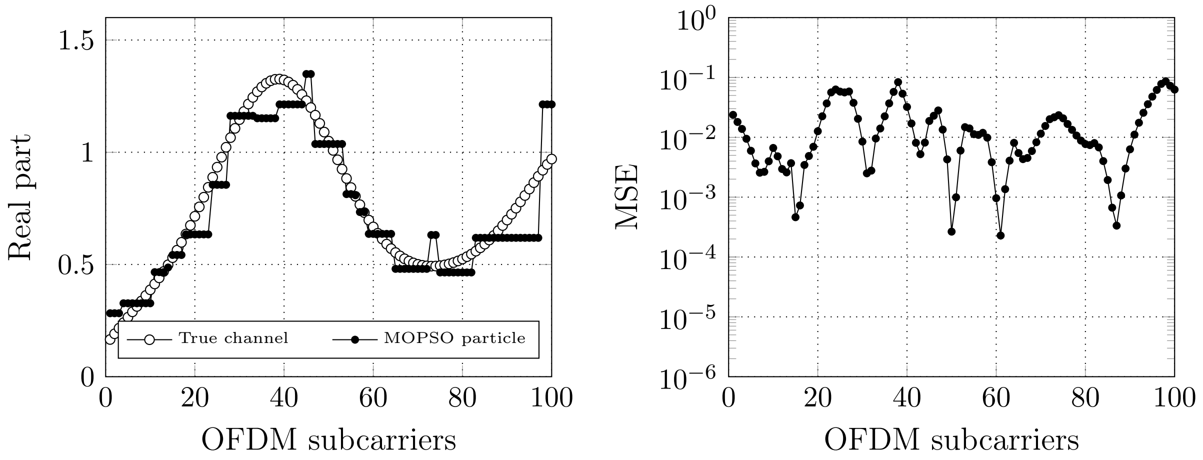
Figure 4.6: Movement of particles within the complex search space at different numbers of iteration at an SNR of 30 dB.

performance and the extension to MIMO is desirable. Typically for channel estimation, the correlation between channel coefficients is exploited to improve the channel estimation accuracy by means of interpolation/filtering. A corresponding post-processing of the channel estimates is proposed in [AK04] and could be combined with MOPSO. However, the complexity of this approach is significantly increased. Within the context of this thesis, linear prediction is combined with the fitness function to improve the performance as well as to facilitate MIMO channel estimation.

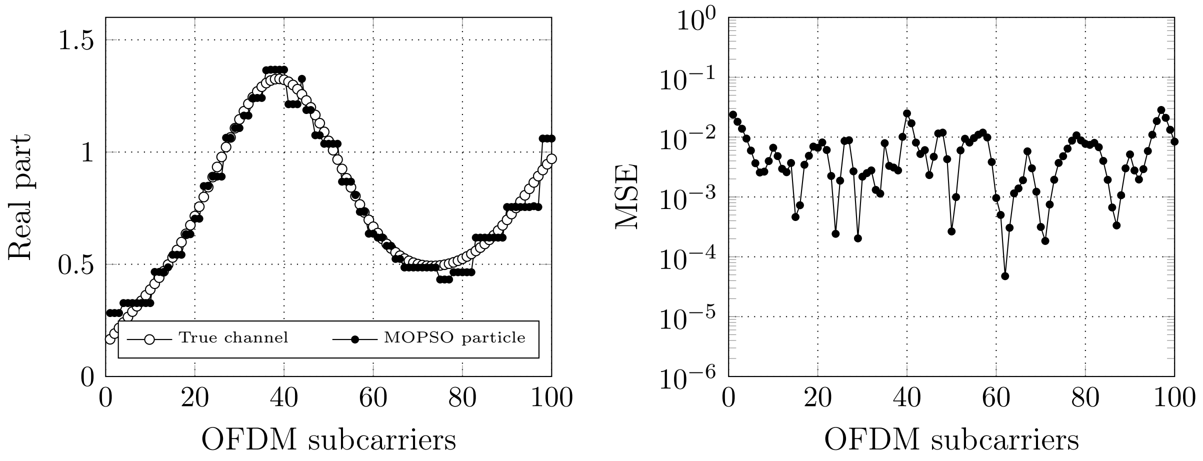
Commonly, linear prediction is used to estimate a channel coefficient given a history of previous channel coefficients [HL99]. It is important to note that linear prediction is not used to estimate a neighboring channel coefficient, but instead—as shown in Figure 4.9 for the time domain—used to map the current particle $\xi[\lambda]$ to neighboring objectives, i.e. OFDM symbols and OFDM subcarriers, respectively. Thus, the neighboring samples



(a) Iterations = 1 (b) Iterations = 1

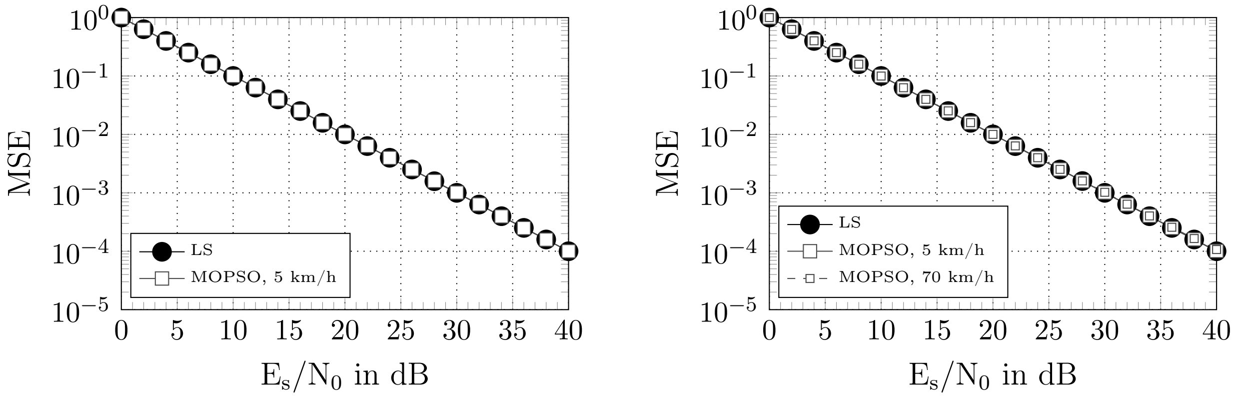


(c) Iterations = 3 (d) Iterations = 3



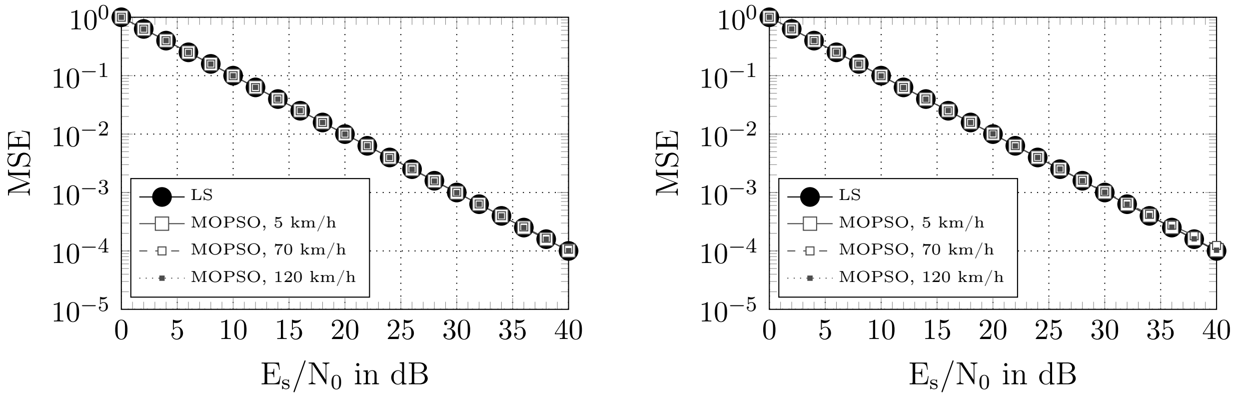
(e) Iterations = 5 (f) Iterations = 5

Figure 4.7: Movement of particles within the complex search space at different numbers of iteration at an SNR of 30 dB.



(a) WINNER A1 NLOS

(b) WINNER B1 NLOS



(c) WINNER C1 NLOS

(d) WINNER C2 NLOS

Figure 4.8: MSE performance of LS and MOPSO channel estimation for different WINNER channel scenarios with varying velocities.

(e.g. $y[\lambda'], y[\lambda''], \ldots$) can be used to either determine (4.18), and/or to further refine the estimate of $\xi[\lambda]$, since additional observations are used to estimate a single channel coefficient. The prediction depth is inherently limited to one, since only the current coefficient is mapped to its adjacent neighbors. The prediction coefficient p_1 for a prediction depth

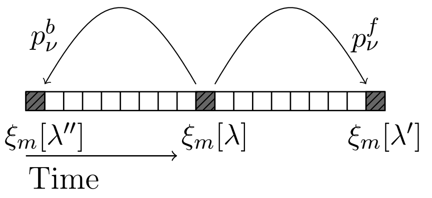


Figure 4.9: *Mapping of one coefficient to neighboring OFDM symbols by using linear prediction.*

of one, is obtained by solving the Wiener-Hopf equation:

$$r_0 \cdot p_1 = r_1. \tag{4.19}$$

The coefficient p_1 may be used for a forward and/or backward prediction, with $p_1^{\mathrm{f}} = p_1^{\mathrm{b}*}$. Assuming a Jakes power spectral density with maximum Doppler frequency $f_{\mathrm{D,max}}$, the correlation coefficient r_1 is determined by [HL99]: $r_0 = 1 + 2\sigma_n^2$ and the time correlation function given in (2.33):

$$\theta_{HH}\left(\Delta t\right) = J_0\left(2\pi f_{\mathrm{D,max}} \Delta t\right).$$

Hereby, r_1 corresponds to $r_1 = \theta_{HH}\left(D_{\mathrm{t}} T_{\mathrm{s}}\right)$. The proposed scheme can be directly implemented in the frequency domain as well, thus further extending the number of equations for one channel coefficient. Depending on the amount of a priori information of the power delay profile, different frequency correlation functions can be derived (cf. Section 2.1.2). Three different assumptions are evaluated in the following: (1) the PDP is assumed to be uniformly distributed between $[0, \tau_{\max}]$, (2) the PDP is exponentially decreasing between $[0, \tau_{\max}]$, and (3) exact knowledge of the PDP is given. The corresponding three frequency correlation functions are as follows:

$$\theta_{HH}^{\mathrm{uni}}\left(\Delta f\right) = \operatorname{sinc}\left(\tau_{\max} \Delta f\right) \cdot \exp\left(-j\pi\tau_{\max}\Delta f\right), \tag{4.20}$$

$$\theta_{HH}^{\mathrm{exp}}\left(\Delta f\right) = \frac{1}{1 + j2\pi\tau_{\mathrm{rms}} F}, \tag{4.21}$$

$$\theta_{HH}^{\mathrm{exact}}\left(\Delta f\right) = \sum_{c=1}^{M_{\mathrm{c}}} P_c \cdot \exp\left(-j2\pi\Delta f \tau_c\right). \tag{4.22}$$

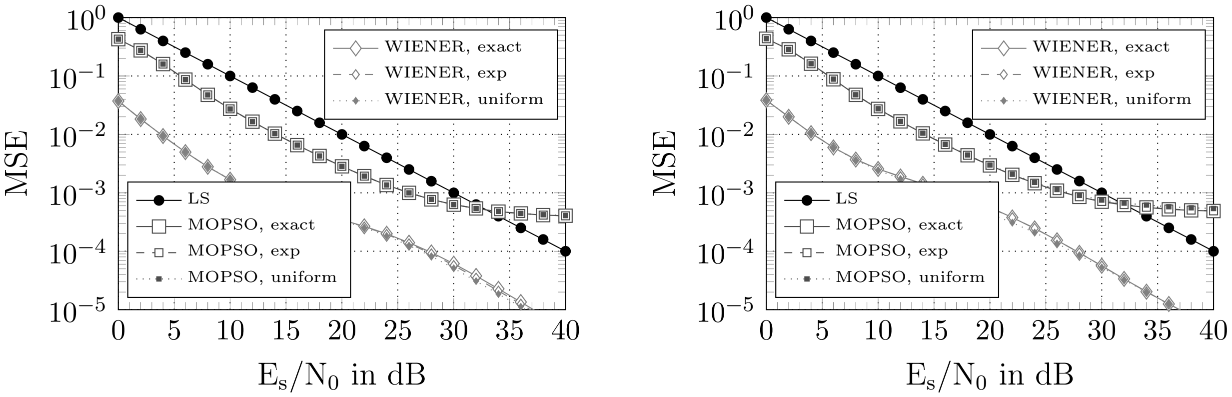
The correlation coefficient r_1 for the frequency domain is thus given by $r_1 = \theta_{HH}\left(D_{\mathrm{f}} F\right)$. The proposed fitness function incorporating the principles of linear prediction is now given by:

$$\begin{aligned} f_\lambda'(\boldsymbol{\xi}[\lambda]) = \Bigg| & y[\lambda] - \sum_{m=1}^{N_{\mathrm{T}}} \xi_m[\lambda] \tilde{x}_m[\lambda] + \\ & y[\lambda'] - \sum_{m=1}^{N_{\mathrm{T}}} p_1^{\mathrm{f}} \xi_m[\lambda] \tilde{x}_m[\lambda'] + y[\lambda''] - \sum_{m=1}^{N_{\mathrm{T}}} p_1^{\mathrm{b}} \xi_m[\lambda] \tilde{x}_m[\lambda''] \Bigg|^2, \end{aligned} \tag{4.23}$$

with p_1^{f} and p_1^{b} being the forward and backward prediction coefficients of either time or frequency domain. Furthermore, the particle $\boldsymbol{\xi}[\lambda]$ contains the N_{T} candidate solutions $\boldsymbol{\xi}[\lambda] = [\xi_1[\lambda], \ldots, \xi_m[\lambda], \ldots, \xi_{N_{\mathrm{T}}}[\lambda]]$ of one objective λ. The MSE performance of MOPSO with linear prediction in two domains is shown in Figure 4.10 for the four WINNER scenarios and a velocity of 5 km/h. A total of 10 internal iterations for the MOPSO algorithm is used. For comparison, symbol-wise LS channel estimation and a 2D Wiener filter are included. The three different frequency correlation functions are used for linear prediction as well as for the Wiener filter. The zeroth order Bessel function is always used for the time domain. Utilizing more observations within the fitness function of MOPSO with linear prediction improves the performance and is robust to approximation errors. In this case, the pilots are directly adjacent and the difference between the three autocorrelation functions is minor, as has been already mentioned in Section 2.1.2. Obviously, MOPSO and the Wiener filter are robust w.r.t the choice of frequency correlation function. This behavior is explained by the short sequence length in both time and frequency domain. More specifically, the different approximations for a correlation function are nearly equal if only a few OFDM subcarriers and/or OFDM symbols are considered. However, with increasing distance the functions very soon drift apart. For a more detailed illustration, the different approximations are plotted in Figure 5.26 for the considered channel scenarios.

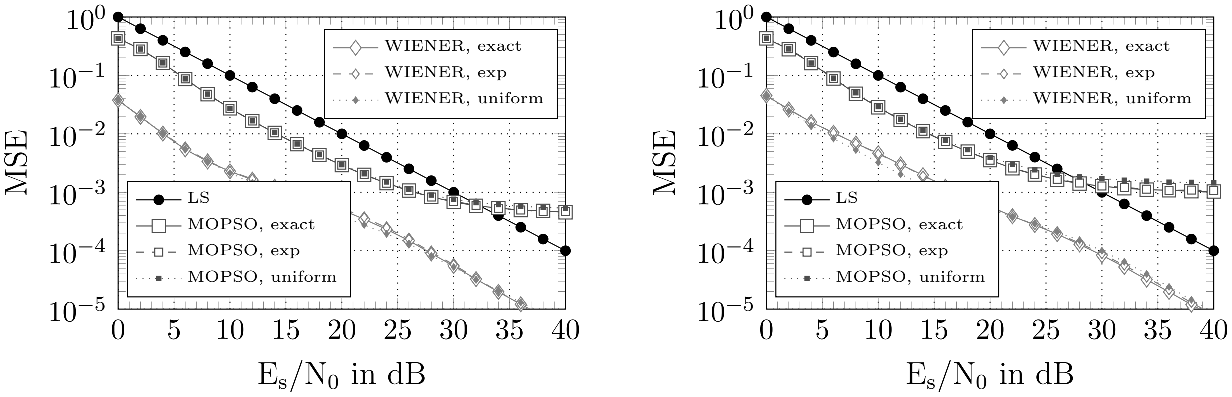
With increasing diversity in the frequency domain, the MSE of the MOPSO results in an error floor at around $1 \cdot 10^{-3}$ for all channel models. An equivalent behavior is observed with an increased diversity in time domain. For 70 km/h, shown in Figure 4.11, an error floor is obtained for all three scenarios. The A1 NLOS channel is excluded for this velocity as 70 km/h exceeds the maximum defined velocity. The error floor increases for a velocity of 120 km/h as can be seen in Figure 4.12. Again, the channel models A1 and B1 are excluded due to the exceeded maximum velocity. Apparently, the error is caused by ambiguities introduced by the linear prediction. With increasing SNR, the particles are not longer able to converge closer to the global optimum. This behavior is partly explained by the limited number of iterations. With more iterations, the MOPSO algorithm achieves a better MSE performance at higher SNR values. However, with increasing diversity in time and/or frequency domain, the precision is again limited by the linear prediction, which introduced ambiguities and thus prevents the MOPSO algorithm to converge closer to the global optimum. The results indicate that the performance and/or performance improvement w.r.t. LS channel estimation is best for the A1 NLOS channel and/or slow velocities and gradually decreasing for the remaining scenarios. However, in the SNR range of interest, between 0 dB and 25 dB, the MSE performance of MOPSO is always better compared to LS channel estimation.

Besides improving the performance, linear prediction enables MIMO channel estimation with MOPSO. The MSE performance of MOPSO with different number of transmit antennas is shown in Figure 4.13. The transmit sequence consists of $L = 10$ OFDM subcarriers and $K = 10$ OFDM symbols. Exact knowledge of the power delay profile is assumed for the calculation of the frequency correlation function. Furthermore, linear prediction is applied in time and frequency domain. However, the mapping of a particle to its neighbors is done for each domain separately. This means, observations used within the fitness function are adjacent in either time *or* frequency and not in time *and* frequency. Hence, a maximum of four additional observations can be used. This number reduces to two at the edges of the sequence. A maximum of five antennas is theoretically supported, however, due to the above mentioned edge effects, only four antennas can be estimated, as can be seen from the numerical results. Already with four transmit antennas, the amount of observations is limited and the resulting error floor can not be prevented. By a straightforward combination of prediction coefficients in time and frequency, the remaining four diagonal neighbors can be used. This would increase the number of supported transmit antennas to theoretically nine transmit antennas. However, even eight transmit antennas are rarely used in current wireless standards. And although defined in LTE-A, the in-



(a) WINNER A1 NLOS

(b) WINNER B1 NLOS



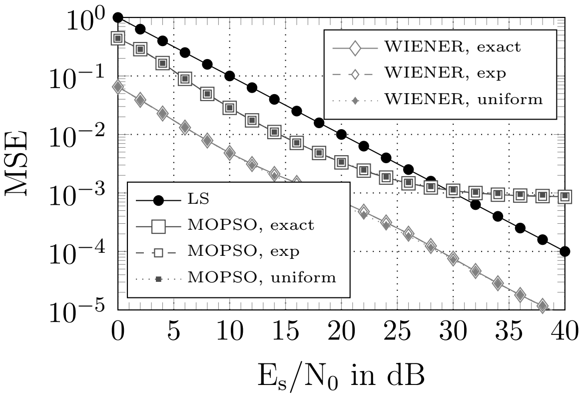
(c) WINNER C1 NLOS

(d) WINNER C2 NLOS

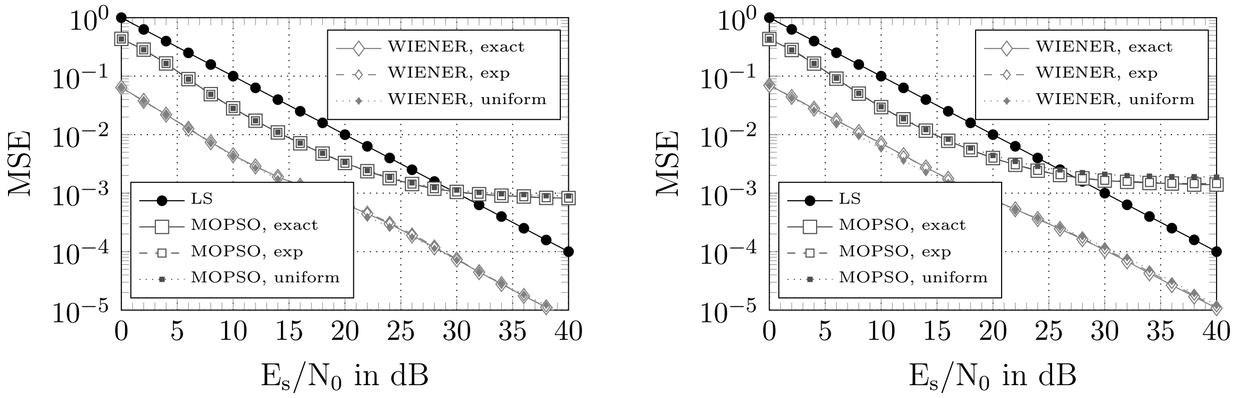
Figure 4.10: MSE for different WINNER channel scenarios for a velocity of 5 km/h.

tended use is for codebook-based beamforming, whereby the number of parallel transmit streams is still limited to four.

An additional advantage of MOPSO is given for multi-cell environments. Here, orthogonal sequences are required to distinguish between users of one cell. Depending on the pilot re-use factor, already directly adjacent cells may use the same set of orthogonal sequences. Typically, co-channel interference (CCI) strongly deteriorates the MSE performance and consequently BER performance as well. With MOPSO and linear prediction, the impact on the achievable performance is not pronounced for a single interferer within the SNR range of interest. The previous system setup with two transmit antennas is reused in the following. Instead of an orthogonal pilot sequence, a random sequence is generated and superimposed. Hereby, the power of the desired signal and the interference is equal, which represents the worst case scenario. As can be seen in Figure 4.14, the performance of MOPSO with a random pilot sequence results in an error floor, with varying height depending on the channel scenario. Up to an SNR of 15 dB the performance loss is negligible. Since the error floor remains below an MSE of $1{\cdot}10^{-2}$ for all scenarios, the influence on the BER performance is limited when MOPSO is used for the initialization [KHAT11b]. In case identical pilot sequences are superimposed, MOPSO is not able



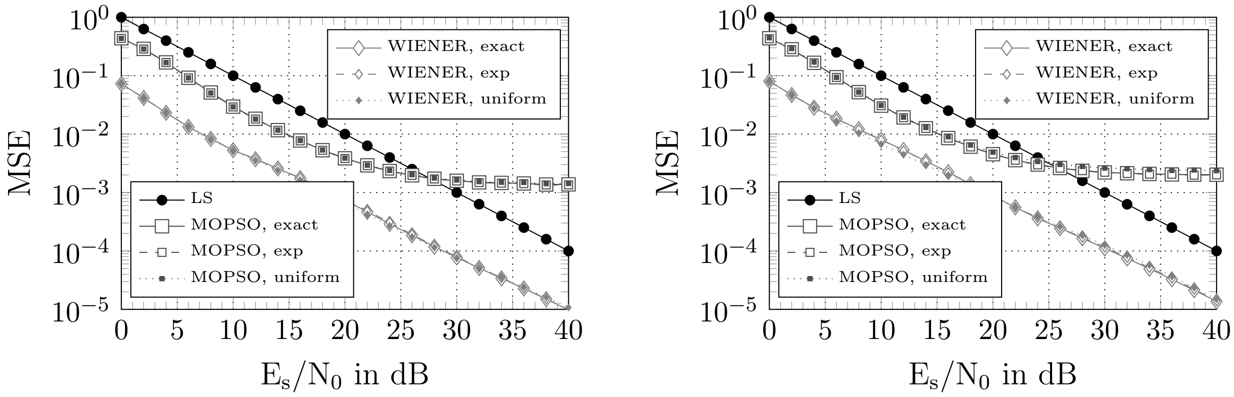
(a) WINNER B1 NLOS



(b) WINNER C1 NLOS

(c) WINNER C2 NLOS

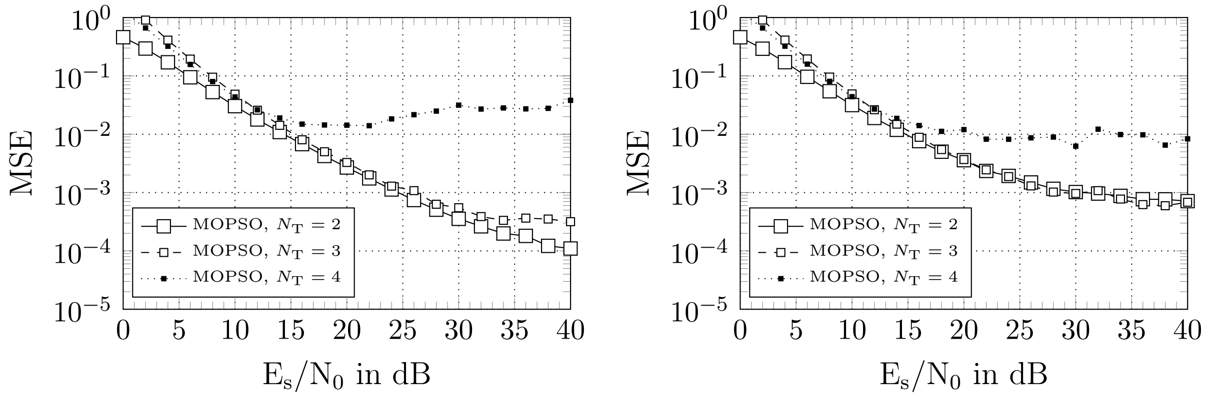
Figure 4.11: MSE for different WINNER channel scenarios for a velocity of 70 km/h.



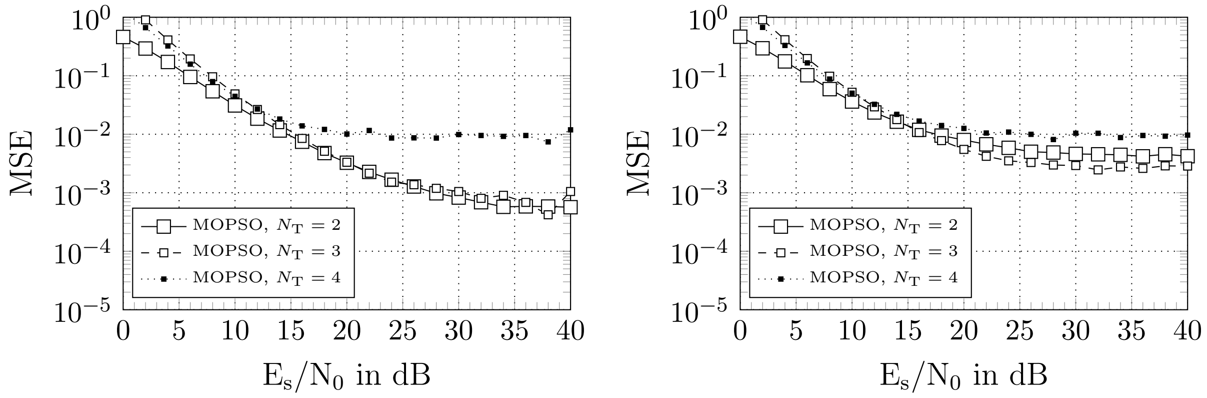
(a) WINNER C1 NLOS

(b) WINNER C2 NLOS

Figure 4.12: MSE for different WINNER channel scenarios for a velocity of 120 km/h.



(a) WINNER A1 NLOS (b) WINNER B1 NLOS



(c) WINNER C1 NLOS (d) WINNER C2 NLOS

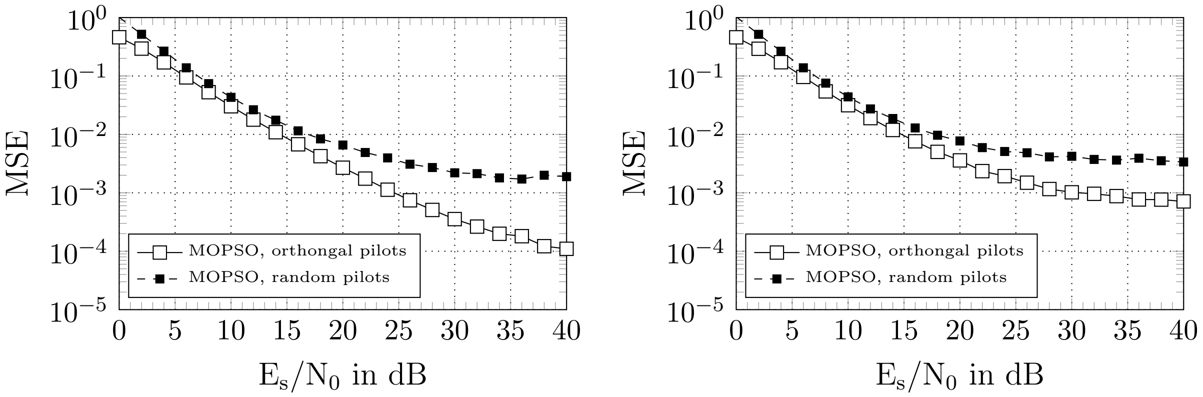
Figure 4.13: MSE for different WINNER channel scenarios for a velocity of 5 km/h as a function of the number of transmit antennas.

to converge, which is a rather unlikely case. As a result, the effects of CCI on the channel estimation performance can be mitigated with MOPSO.

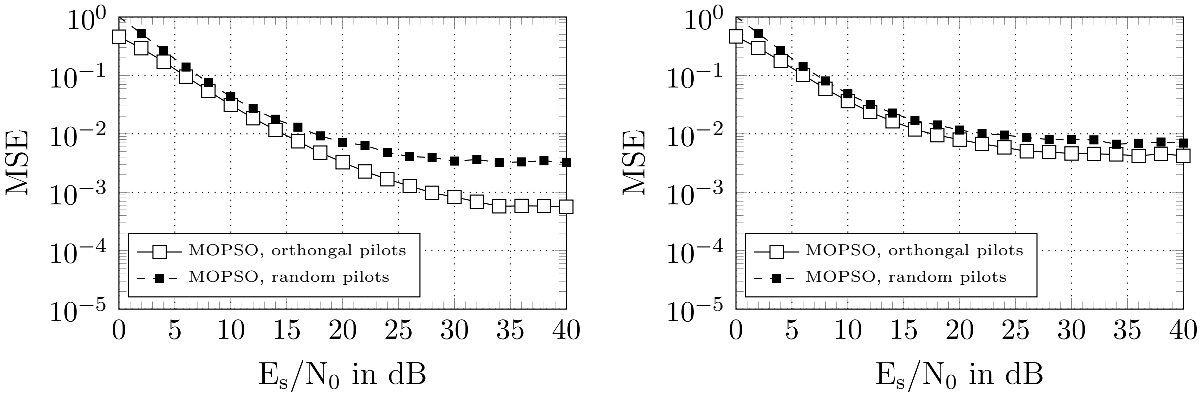
4.3 Complexity Analysis

One of the main advantages of PSO lies in its simple implementation. Since it does not require any gradient information of the optimization problem, the algorithm is versatile and can be applied to a variety of optimization problems. However, a disadvantage of PSO is that—being a meta-heuristic—a convergence to the global optimum is not guaranteed. Moreover, if the algorithm is able to converge to the global optimum, the required number of iterations cannot be predicted. In this chapter, the complexity of PSO/CPSO and MOPSO per iteration is analyzed. By using a generalized extreme value distribution, a maximum number of iterations can be determined semi-analytically, thus enabling a complete complexity assessment of PSO.

In general, the complexity of PSO/CPSO is determined by the number of particles,



(a) WINNER A1 NLOS (b) WINNER B1 NLOS



(c) WINNER C1 NLOS (d) WINNER C2 NLOS

Figure 4.14: MSE for different WINNER channel scenarios with a velocity of 5 km/h, two transmit antennas as a function of random pilot sequence.

subswarms, dimensions, and the required number of iterations for convergence. The number of particles and subswarms is a design parameter of the algorithm and is commonly chosen to achieve a good performance in terms of MSE for channel estimation. The number of dimensions is a fixed parameter depending on the optimization problem (e.g. number of transmit and receive antennas and/or channel memory length). In each iteration all particles N_p' of all subswarms N_s have to evaluate their current position and compare their current fitness value with their personal best as well as the global best, which results in a complexity of order

$$\mathrm{CPSO}^{(\mathrm{it})} = \mathcal{O}\left(N_\mathrm{p}' \cdot N_\mathrm{s} \cdot D\right) \tag{4.24}$$

per iteration. The overall number of particles influences the number of iterations needed to converge. In case of using only one particle the required number of iterations until convergence is maximized and computational complexity per iteration is minimized, while, on the other hand, using an infinite number of particles minimizes the number of iterations and maximizes the computational complexity per iteration. With an infinite number of particles, PSO is equivalent to exhaustive search. Hence, a trade-off between the overall size of PSO/CPSO and the number of iterations has to be found. Furthermore, the required minimum number of iterations is depending on the optimization metric as well. In general, the more complex (higher dimensional) the optimization problem is, the more iterations are needed and vice versa. Based on the previous simulation results, the parameters which determine the complexity of PSO/CPSO can be reformulated more precisely, i.e. the number of particles is set to $N_\mathrm{p}' = 5$ independent of the dimensions, the dimensionality of the optimization problem corresponds to $D = 2N_\mathrm{T}N_\mathrm{R}$, and the number of subswarms was chosen to be $N_\mathrm{s} = N_\mathrm{T}$. Hence, the resulting complexity is as follows:

$$\mathcal{O}\left(N_\mathrm{p}' \cdot N_\mathrm{s} \cdot D\right) = \mathcal{O}\left(5 \cdot N_\mathrm{T} \cdot 2N_\mathrm{T}N_\mathrm{R}\right) = \mathcal{O}\left(N_\mathrm{T}^2 N_\mathrm{R}\right). \tag{4.25}$$

Hereby, the number of particles per swarm N_p' is neglected since it has no influence on the asymptotic complexity. For comparison, the complexity of the MMSE channel estimation is dominated by the matrix inversion, which has a complexity of order

$$\mathrm{MMSE} = \mathcal{O}(N_\mathrm{T}^3). \tag{4.26}$$

Concluding from the comparison of (4.25) and (4.26), it is obvious that PSO/CPSO offers a complexity advantage for large MISO systems, which is the case for the downlink in a massive MIMO system given no CSIT [LTEM13]. Nevertheless, the overall complexity of CPSO depends on the complexity per iteration and the number of iterations:

$$\mathrm{CPSO}^{(\mathrm{total})} = \mathcal{O}\left(i_\mathrm{max}\right) \cdot \mathrm{CPSO}^{(\mathrm{it})}. \tag{4.27}$$

Similarly, the complexity can be determined for MOPSO. Hereby, the complexity of the algorithm is closely related to that of PSO, with the exception of the archive. For an update of the external archive, each particle has to compare its current position with the entries of the archive, which are equivalent to the number of objectives $\mathit{\Lambda}$. Thus, the complexity for the update process of the archive for each iteration results in $\mathcal{O}(N_\mathrm{p}\mathit{\Lambda})$. Accordingly, the complexity of MOPSO is increasing with the order of

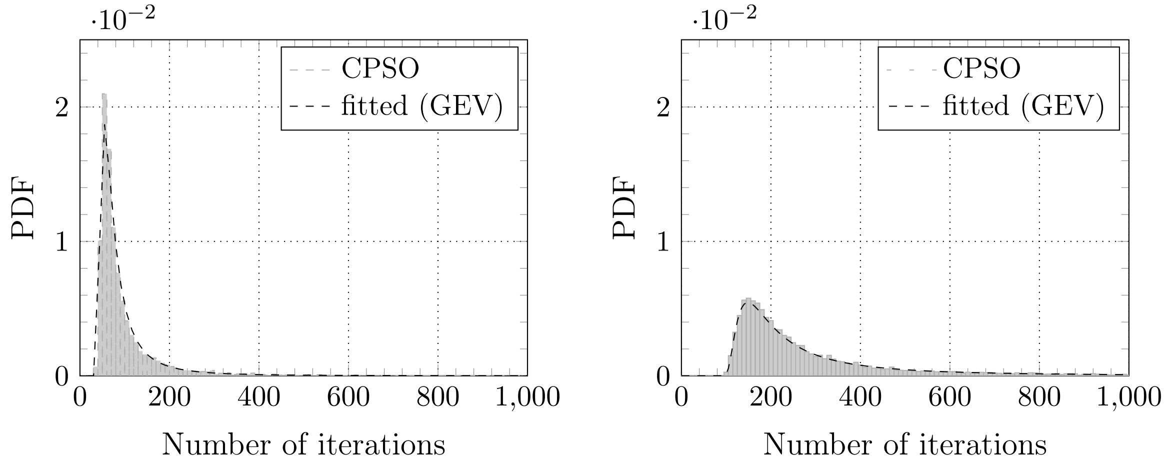
$$\mathrm{MOPSO}^{(\mathrm{total})} = \mathcal{O}\left(N_\mathrm{p}^2 \mathit{\Lambda} \cdot i_\mathrm{max}\right). \tag{4.28}$$

Due to the increased complexity of the archive maintenance, it was concluded in Section 4.2.2 that the virtue of MOPSO is its fast convergence to a reasonable MSE. As a result, the maximum number of iterations is typically set to a small number and hence, can be neglected for MOPSO.

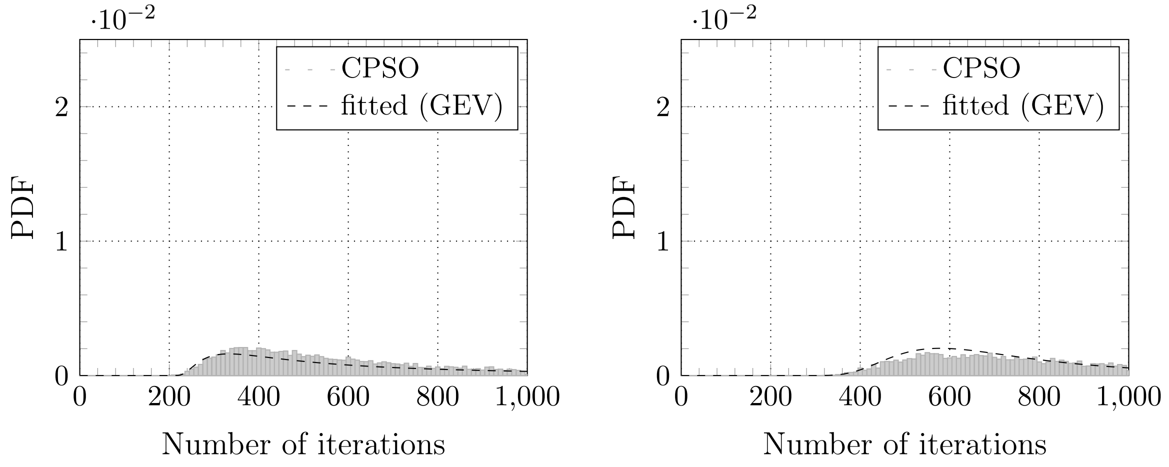
Nevertheless, iterations are an important parameter for PSO/CPSO. A strategy often used to determine the maximum number of iterations i_max is to find the minimum value of iterations at which the optimum MSE performance is reached. This approach requires extensive simulations over a variety of parameters in order to determine the optimum trade-off between complexity and iterations.

In the following, a general criterion to determine the maximum number of iterations based on the probability distribution function of the iterations required by PSO/CPSO for convergence is presented. The advantage of this strategy is that only a fraction of parameters need to be simulated while missing parameters can be reconstructed by means of an interpolation. PSO/CPSO is said to reach convergence if the fitness value p^{OPT} of (4.5)/ (4.6) is below a certain threshold t_{h} for γ iterations. In this case the threshold is set to $t_{\mathrm{h}} = 10^{-6}$ with $\gamma = 10$.

Monte Carlo simulations with a fixed parameter set for CPSO and varying number of transmit antennas are conducted, i.e. the number of subswarms is set to $N_{\mathrm{s}} = 8$ and the number of particles per swarm is $N_{\mathrm{p}}' = 5$. The iteration at which the stopping criterion is fulfilled is recorded. A histogram of the iterations fulfilling the stopping criterion for different dimensions is shown in Fig. 4.15. Each histogram is approximated by a generalized extreme value (GEV) distribution. The characteristic shape of the function is in the steep slope once a certain value is exceeded and a slow decline after the maximum is reached. In general, extreme value theory models events which occur with a very small probability. The corresponding distributions are often the limiting distributions for maximums or minimums of a sample of i.i.d. random variables [NA08]. They have been successfully



(a) Dimensions=8 (b) Dimensions=16



(c) Dimensions=32 (d) Dimensions=48

Figure 4.15: Histogram of the minimum number of iterations required to converge in dependence of the dimensionality.

applied to predict stock market crashes [GK06] and natural hazards [GYH+11]. Within the context of this thesis, the GEV distribution is used to predict the maximum number of iterations required to fulfill the chosen stopping criterion. As can be seen from the curves in Figure 4.15, the GEV distribution fits well for a wide range of dimensions. Given that the number of dimensions is equal to the number of swarms ($N_{\mathrm{T}} = D = 8$), a peak at the 63th iteration is very pronounced. Hence, CPSO is most likely to converge at this iteration. With increasing dimensions, the peak is shifted to more iterations as well as lowered in probability. For this setting, a likely convergence can still be predicted, but the loss of performance is larger when the maximum number of iterations is set closer to the maximum of the distribution. This effect is exacerbated with increasing dimensions. Accordingly, the observations gained from Figure 4.15a are confirmed, i.e. setting the number of subswarms in close relation to the number of dimensions yields a good trade-off between performance and complexity.

The probability density function (pdf) of the generalized extreme value distribution is described by (4.29). The distribution is characterized by three parameters, namely the shape parameter k, the scale parameter σ, and the location parameter μ:

$$p(k,\sigma,\mu) = \left(\frac{1}{\sigma}\right) \exp\left(-\left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{k}}\right)\left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-1-\frac{1}{k}}. \tag{4.29}$$

Given the pdf for a certain parameter set, the maximum number of iteration i_{max} can be defined to cover a certain percentage of the pdf. The amount to which the pdf is covered defines the trade-off between performance and complexity. Setting the maximum number of iterations too low reduces the complexity of the algorithm but also implies a performance loss due to a premature stop of the algorithm. Vice versa, setting the maximum number of iterations too large is increasing complexity without a gain in performance. In case of $D = 8$ (cf. Figure 4.15a), the location parameter results in $\mu = 63$, which resembles the most likely iteration at which the algorithm converges. In order to cover at least 90% of the required iterations the maximum number of iterations should be set to $i_{\mathrm{max}}^{90} \geq 180$. The parameters k, μ, and σ required for the GEV distribution to resemble the dimensions shown in Figure 4.15a are tabulated in Table 4.2.

The aforementioned trade-off between the number of particles/subswarms and the number of iterations is evaluated in the following. The maximum number of iterations required to cover 90% of the pdf is defined as i_{max}^{90} and is included in Table 4.2 as well. As indicated by the previous results, the distribution and thus the number of iterations

	k	σ	μ	i_{max}^{90}
$D = 8$	0.585151	22.7014	63.2505	180
$D = 16$	0.799056	87.1973	189.144	770
$D = 32$	1.15118	375.018	524.045	4650
$D = 48$	1.3971	837.99	918.205	>10000

Table 4.2: Parameters for the GEV distribution as a function of the dimensionality D with a swarm size of $N_{\mathrm{s}} = 8$ and $N'_{\mathrm{p}} = 5$.

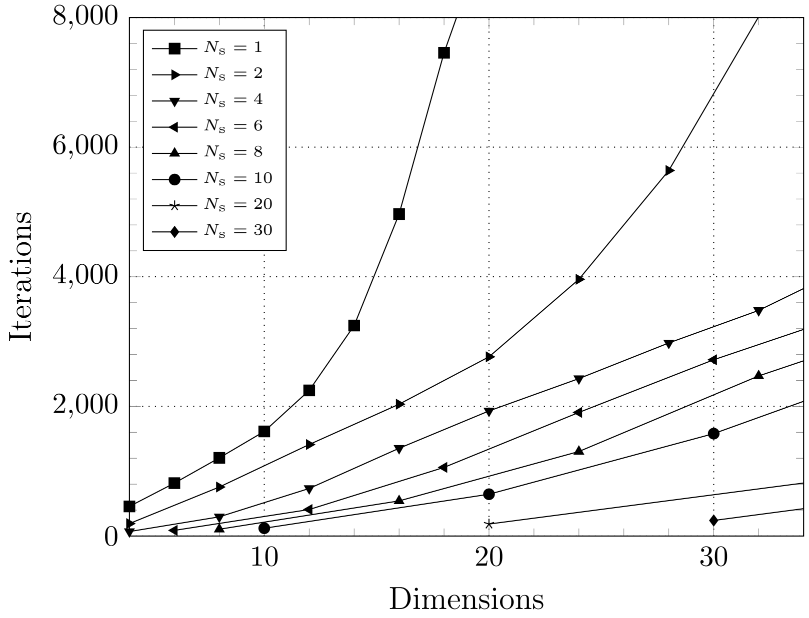


Figure 4.16: *Required number of iterations of different swarm sizes as a function of the number of dimensions at an SNR of 10 dB with* $N_\mathrm{s} = 8$ *and* $N'_\mathrm{p} = 5$.

required by PSO/CPSO until convergence depends on the number of dimensions of the optimization problem and the allocated number of swarms and inherently particles. In Figure 4.16, the required number of iterations depending on the dimensions of the optimization problem is given for different swarm sizes. With a constant swarm size the iterations are increasing quadratically with the dimensions. On the contrary, with increasing swarm sizes, the required iterations are nearly constant with increasing dimensions, as can be seen from the similar starting points of the curves. The required number of iterations for PSO ($N_\mathrm{s} = 1$) to converge, exceeds 8000 at 20 dimensions. Since the three parameters of the GEV distribution are correlated over the number of particles and subswarms, not all swarm sizes need to be simulated but can be calculated by means of interpolation. The optimum trade-off between swarm size and iterations can thus be determined with a minimum amount of simulations.

The derived convergence criterion of PSO/CPSO is designed to achieve the MMSE performance. In order to set the different numbers into relation, the maximum number of iterations is determined based on a fixed complexity. More specifically, the number of iterations should be sufficiently low such that the complexity is similar to that of MMSE channel estimation. Therefore, the value obtained by the Landau notation, given by (4.24) for CPSO and by (4.26) for MMSE are given in Table 4.3. These numbers do not represent an accurate information about the complexity but are rather shown to illustrate the importance of the maximum number of iterations on the overall complexity. The maximum number of iterations for which the total complexity of CPSO is similar to MMSE is shown as well. A maximum number of 4 and 16 iterations, respectively, is allowed for MISO system with $N_\mathrm{T} = 4$ and $N_\mathrm{T} = 16$ transmit antennas. Clearly this number is insufficient to achieve a reasonable MSE performance with CPSO. With increasing

	N_T=4, (D=8)	N_T=16, (D=32)	N_T=100, (D=200)	N_T=150, (D=300)
MMSE	64	4096	$1 \cdot 10^6$	$3.375 \cdot 10^6$
CPSO$^{(\mathrm{it})}$	16	256	1000	22500
$i_{\max}$	4	16	100	150

Table 4.3: Exemplary values obtained with the Landau notation for MMSE and PSO as a function of transmit antennas.

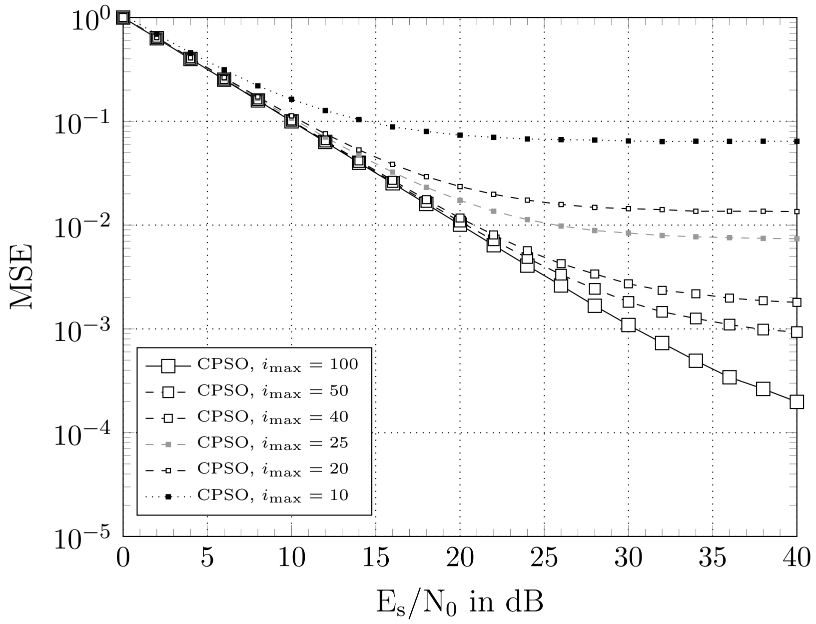


Figure 4.17: MSE of CPSO with varying $i_{\max}$ in a massive MIMO system with $N_T = 100, N_R = 1$ antennas.

number of transmit antennas, the maximum number of allowed iterations increases as well. Obviously, optimum performance cannot be achieved as well with 100 and/or 150 iterations. However, few iterations may suffice to provide accurate initial channel state information. Exemplary, the MSE performance of a massive MIMO system with $N_T = 100$ transmit antennas and $N_R = 1$ receive antenna is shown in Figure 4.17. Applying the semi-analytical analysis described above, the following parameters are obtained for the GEV distribution: $k = 0.0730982, \sigma = 154.294, \mu = 369.066$. In order to cover 90% of the distribution, the maximum number of iterations should be set to $i_{\max} = 747$. With this setting, the complexity of CPSO would exceed the complexity of MMSE significantly and is thus, not sustainable. In order to have a similar complexity, a maximum number of $i_{\max} = 100$ should be used, as shown in Table 4.3. Optimum performance cannot be achieved with $i_{\max} = 100$ as the curve slowly deviates into an error floor. For initialization, an MSE threshold of $1 \cdot 10^{-2}$ is usually sufficient, hence, within 25 iterations this target is achieved. Equivalent convergence behavior of CPSO is observed for even larger MIMO implementations. A fair comparison of the complexity is not straightforward, since not only the number of complex multiplications have to be considered. However, the simulation results indicate that both complexity and performance of PSO/CPSO depend on the number of iterations. Only for large-scale MIMO implementations a sufficient number of iterations is supported in order to achieve a reasonable performance. The fast converging nature of PSO is especially advantageous when used for initialization.

An additional advantage of PSO is that it is ideally suited for parallel computing. With the advent of so-called *many integrated core* architectures [Int13] and/or graphical processing units (GPUs) [Nvi13] with hundreds or even thousands of integrated cores, the need for algorithms that scale well with the number of available processing cores is evident. Hereby, each particle of a swarm could run in parallel, as proposed and evaluated in [MMS07, MDC11] for the use with GPUs. Performance improvements w.r.t. processing times are significant and scale well with the number of cores. Furthermore, PSO is robust to faulty nodes, which occur more often the more nodes are used.

On the contrary, applying parallel processing for the matrix inversion required by MMSE is not trivial and the gain due to parallelization depends on the method used to invert the matrix [Pea67].

4.4 Chapter Summary

In this chapter, particle swarm optimization is evaluated for the use of MIMO channel estimation. Novel approaches for MIMO channel estimation are proposed, such as CPSO for a flat-fading channel and MOPSO for a time-varying frequency-selective channel. It has been shown that the cooperative approaches are able to approach the performance of an MMSE estimator. Thus, for a potential implementation, the complexity and inherently the required number of iterations are of utmost importance. A thorough complexity analysis of PSO/CPSO revealed that it is particularly useful for asymmetric MIMO constellations, such as massive MIMO. Furthermore, the amount of iterations needed to converge to the optimum solution can be immense. A change of the underlying hardware to so-called many integrated core architectures may justify PSO/CPSO. However, a more useful application is the provision of initial channel state information. Here, PSO/CPSO has several advantages. First, only few iterations are required to converge to a reasonable MSE, even in very large dimensions. Second, nearly no a priori information are required, e.g. noise variance or the distribution of the power delay profile and/or Doppler spread. And third, no special pilot design is required in combination with MOPSO and linear prediction. This is especially useful in multi-cell scenarios with a small pilot re-use factor. Suitable pilots can be selected by means of linear prediction which improves channel estimation accuracy.

5

Graph-based Soft Iterative Receiver

DURING the past decades, digital communication systems evolved dramatically by incorporating MIMO, OFDM, as well as various smart antennas technologies such as beamforming. Among others, these technologies have been identified as key components to reach the challenging goals of a high spectral efficiency. With the advent of iterative processing based on the concept of Turbo codes, proposed by Berrou et al. [BGT93], the Shannon capacity for Gaussian channels is closely approached. Advanced iterative receivers adapated the concept of turbo processing to facilitate turbo equalization. Equally impressive gains formerly observed for Turbo codes are achieved by means of iterative equalization and decoding. Thereby, graphical models provide a way of an unified receiver design, eventually leading to low-complexity implementations, which offer close-to-optimum performance.

The development of the proposed graph-based receiver concept as well as the derivation of the messages required for soft channel estimation and soft data detection are presented in Section 5.1. A message exchange between coefficient nodes based on so-called transfer nodes is established. The exchange of messages is hereby based on a Gaussian random walk model, which offers a very good performance at lowest complexity. The transfer nodes can be used to facilitate a message exchange in arbitrary dimensions. The resulting receiver concept is dubbed *multi-dimensional graph-based soft iterative receiver*, abbreviated by MD-GSIR.

Due to cycles within the graph structure, a special scheduling needs to be designed, which is discussed in Section 5.2. It is shown that the achievable performance depends on the chosen schedule. However, the loss due to channel estimation remains large under certain channel conditions. A major reason for this poor performance can be traced back to the exchanged messages within the factor graph. Given certain channel conditions, these messages can become correlated, which violates the prerequisite of the sum-product algorithm that only extrinsic information is exchanged. By integrating the concept of correlated combining, which has been discussed in Section 3.4, the performance of the MD-GSIR is again substantially improved. Furthermore, the multi-objective PSO algorithm presented in Section 4.1.4 is used to provide improved a priori information of the channel state information. A remarkable performance w.r.t. BER is achieved without sacrificing the low complexity of the overall receiver structure. A detailed discussion on the impact of a priori information, i.e. its quality and quantity, is given in Section 5.4. It is shown, that the MD-GSIR does not require a particular and/or dense pilot grid as well as accurate a priori information of the channel statistics. An extensive performance evaluation for a wide range of modulation formats and code rates is presented in Section 5.5. Independent of the channel scenario, modulation format, and code rate, the MD-GSIR yields a good performance compared to an iterative state-of-the-art receiver. Section 5.6 highlights the versatility of the proposed receiver concept by integrating codebook-based beamforming to the message exchange of the MD-GSIR. It is shown that with conventional precoding strategies, substantial gains can be achieved without any changes to the graph structure and/or message generation. Moreover, additional improvements can be achieved by providing a priori information of the applied beam-weights. The conclusions of this chapter are drawn in Section 5.7

5.1 Receiver Structure and Associated Factor Graph

Factor graphs are powerful graphical tools that have their origins in coding theory [Loe04] but have been successfully applied to a large variety of problems in digital communications [KFL01, LDH$^+$07]. Generally, a factor graph is a bipartite graph, i.e. the nodes of a graph are partitioned into two disjoint sets $\mathcal{U}$ and $\mathcal{V}$ such that every edge connects a node $u \in \mathcal{U}$ with a node of $v \in \mathcal{V}$. A direct connection between nodes of the same set is forbidden. Moreover, a factor graph expresses the structure of the underlying factorization [KFL01]. The notation common for factor graphs is introduced in the following. It is shown that once the function of interest is factorized, the resulting factor graph structure is directly given. Consequently, the conditional pdf $p(\boldsymbol{x}|\boldsymbol{y})$, required for data detection, is factorized and the underlying factor graph structure enabling low complexity joint channel estimation and data detection is derived subsequently.

5.1.1 Preliminary Remarks

This section gives a brief overview of factor graphs and the sum-product algorithm. On the basis of a simple example, the general principle of both is illustrated. A global function $f\left(x_1, x_2, x_3, x_4\right)$ is assumed, which can be factorized according to

$$f\left(x_1, x_2, x_3, x_4\right) = f_A(x_1) f_B(x_1, x_2) f_C(x_1, x_3, x_4). \tag{5.1}$$

The factorization is represented by the graphical model shown in Figure 5.1. The two disjoint sets $\mathcal{U}$ and $\mathcal{V}$ are visualized by circles for the variable nodes and rectangles for the function nodes. Messages are exchanged within the factor graph between variable nodes

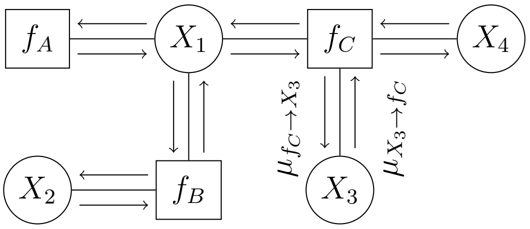


Figure 5.1: Graphical model of the factorization given by (5.1).

and function nodes on the basis of the sum-product algorithm. A message sent from a variable node x_1 to a function node f_C is given by

$$\mu_{X_1 \to f_C}(x_1) = \mu_{f_A \to X_1}(x_1) \cdot \mu_{f_B \to X_1}(x_1), \tag{5.2}$$

whereas a message from the function node f_C to the variable node X_1 is written as

$$\mu_{f_C \to X_1}(x_1) = \sum_{x_3, x_4} f_C(x_1, x_3, x_4) \cdot \mu_{X_3 \to f_C}(x_3) \cdot \mu_{X_4 \to f_C}(x_4). \tag{5.3}$$

The sum-product algorithm obtained its name due to the fact that the various messages are first multiplied and afterwords summed up, as can be seen in (5.3). Moreover, it can be seen from (5.2) and (5.3), that only extrinsic information is used for the message generation. That means that the message which is sent to a variable or function node does not contain any information from the node itself. In graphs without cycles, this rule is fulfilled and the computed marginals are exact. In the above example, the marginal of x_3 is calculated as follows

$$\begin{aligned} g_{X_3}(x_3) &= \mu_{f_C \to X_3}(x_3) \cdot \mu_{X_3 \to f_C}(x_3) \\ &= \sum_{x_1, x_2, x_4} f(x_1, x_2, x_3, x_4). \end{aligned} \tag{5.4}$$

The notation of the sum in (5.4) is replaced by the following short-notation

$$\sum_{x_1, x_2, x_4} f(x_1, x_2, x_3, x_4) = \sum_{\sim x_3} f(x_1, x_2, x_3, x_4). \tag{5.5}$$

Instead of denoting the variables which are summed up, the variable which is excluded is indicated.

5.1.2 Receiver Structure

Following the principles of the simple example given in the previous section, the complementary factor graph structure suitable for joint channel estimation and data detection is to be developed within the following sections. The MIMO system given by (3.1) is reused here

$$\boldsymbol{y}[l,k] = \boldsymbol{H}[l,k]\boldsymbol{x}[l,k] + \boldsymbol{n}[l,k],$$

whereas, according to (3.21), the received signal of the nth receive antenna is given by

$$\begin{aligned} y_n[l,k] &= h_{n,m}[l,k]x_m[l,k] + \sum_{\substack{i=1\\ i\neq m}}^{N_\mathrm{T}} h_{n,i}[l,k]x_i[l,k] + n_n[l,k] \\ &\approx h_{n,m}[l,k]x_m[l,k] + \zeta_{n,m}[l,k]. \end{aligned}$$

Based on the previous brief introduction, it is obvious that the probability density function required for data detection needs to be factorized. It is furthermore easy to agree that if the pdf is not factorized, the resulting factor graph consists of only a few nodes and vice versa the more the pdf is factorized the more nodes are introduced to the graph structure. The complexity of the message generation behaves inversely to the graph structure, which means, message generation is simple when done for a single random variable (many graph nodes) and difficult for a complete vector of random variables (few graph nodes). As said before, one of the objectives for this chapter is to develop a low-complexity graph-based solution. Therefore, the factorization of the pdf and the resulting factor graph structure is derived in the following.

The optimal MAP detector w.r.t. symbol-error probability as defined in (3.8) is revisited:

$$\hat{\boldsymbol{x}}^{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}\in\mathcal{X}} p\left(\boldsymbol{x}|\boldsymbol{y}\right),$$

where $\boldsymbol{x}$ corresponds to the transmitted burst of length $N_\mathrm{T}LK$ and $\boldsymbol{y}$ refers to the received burst with length $N_\mathrm{R}LK$. The trivial factor graph would consists of two nodes, $\boldsymbol{x}$ and $\boldsymbol{y}$, and the resulting complexity of the message generation is increasing exponentially, equivalently to the optimum MAP detector. If the channel is assumed to be unknown, the conditional pdf of (3.3) is changed according to

$$p\left(\boldsymbol{y}|\boldsymbol{x}\right) = \int p\left(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{H}\right)p\left(\boldsymbol{H}\right)d\boldsymbol{H}. \tag{5.6}$$

Assuming no or weak spatial correlation between the N_T transmit and N_R receive antennas, respectively, (5.6) can be simplified to

$$p\left(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{H}\right) \approx \prod_{n=1}^{N_\mathrm{R}}\prod_{m=1}^{N_\mathrm{T}} p\left(\boldsymbol{y}_n|\boldsymbol{x}_m,\boldsymbol{h}_{n,m}\right). \tag{5.7}$$

One of the virtues of OFDM is that the received data symbols are uncorrelated in time and frequency w.r.t. the transmitted data symbols given a sufficiently long interleaver as well as cyclic prefix. Accordingly, the transmitted burst can be written as

$$p\left(\boldsymbol{y}_n|\boldsymbol{x}_m,\boldsymbol{h}_{n,m}\right) = \prod_{l=1}^{L}\prod_{k=1}^{K} p\left(y_n[l,k]|x_m[l,k],h_{n,m}[l,k]\right). \tag{5.8}$$

The marginals of the pdf are significantly easier to calculate, i.e. the complexity is increasing linearly w.r.t. the number of OFDM symbols and/or OFDM subcarriers, as well

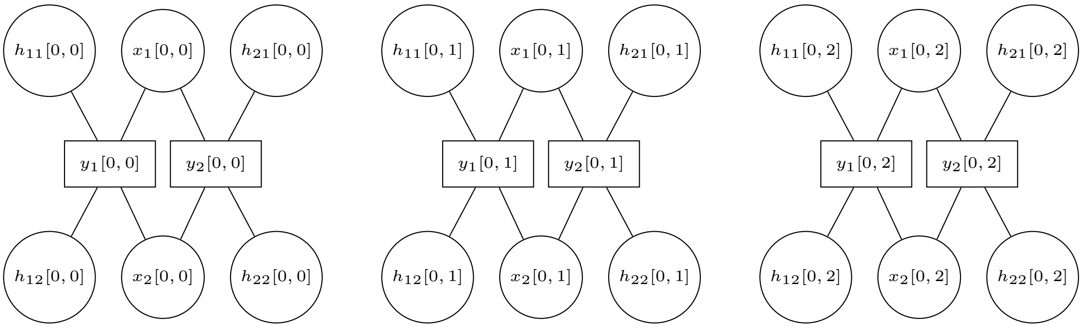
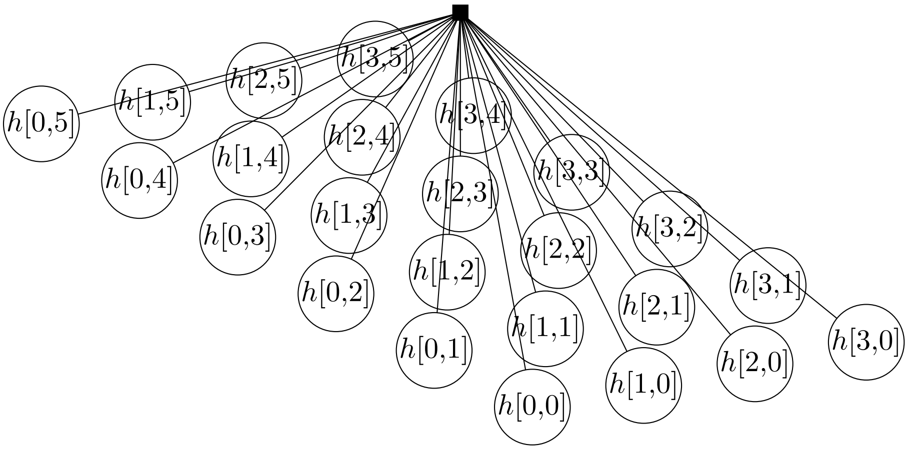
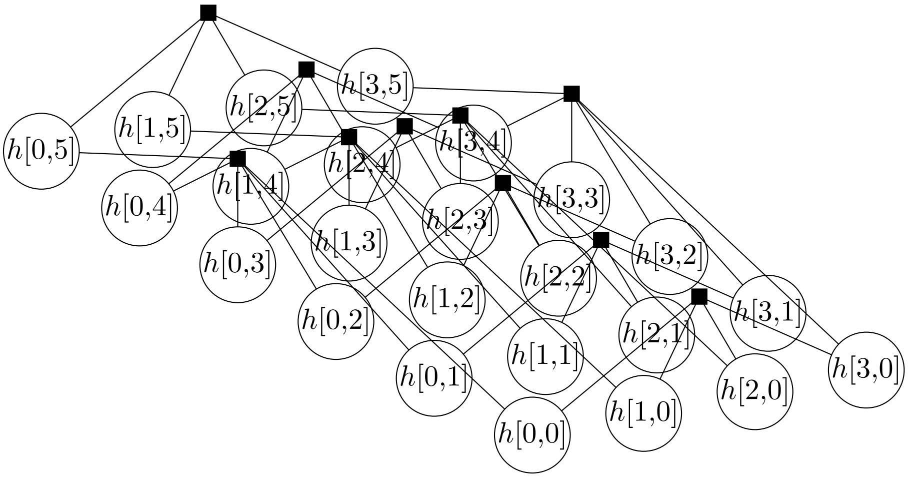


Figure 5.2: Associated factor graph for symbol-wise MIMO detection.

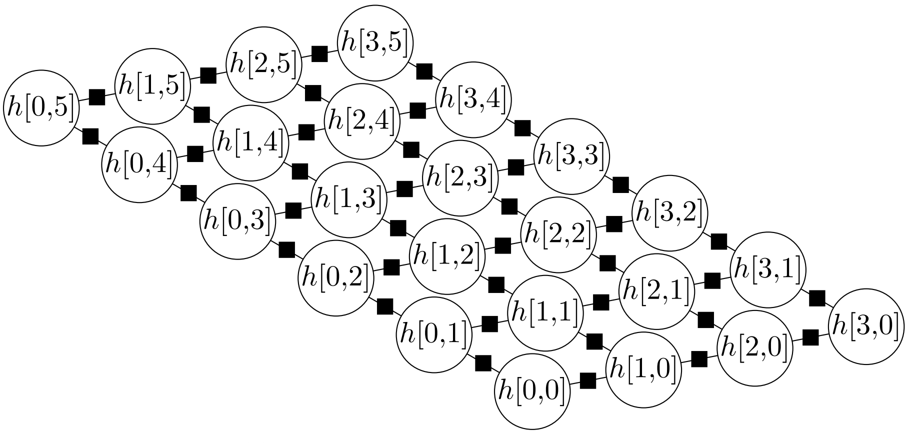
as the number of transmit and receive antennas, opposed to the exponential complexity of the MAP detector. The resulting factor graph structure for two transmit and two receive antennas with BPSK modulation is depicted in Figure 5.2. Up to this point, it is unclear how the message exchange between the single factor graph components is established. Given perfect channel state information, each symbol can be calculated individually. The corresponding soft information is passed to a channel code and fed back by means of turbo processing. However, in case of unknown channel state information, several components can not be calculated at all, since pilot symbols are only transmitted on certain subcarriers and thus, required information is missing. Fortunately, as mentioned in Section 2.2, channel coefficients are highly correlated in time and/or frequency. Under circumstances, even spatial correlation may be exploited in order to establish a message exchange. Hence, a suitable transfer node based on the channel correlation is to be derived. Three variants of such a transfer node are shown in Figure 5.3. Only a single coefficient node at different time and frequency indices is depicted. Without loss of generality, the indices of the transmit and receive antennas are omitted. A corresponding transfer node is indicated by the small black rectangle. The actual function within each transfer node varies, depending on the type of implementation and will be discussed subsequently. The three variants comprise (1) a multi-dimensional transfer node, which connects all channel coefficients in time and frequency, (2) multiple one-dimensional transfer nodes connect coefficient nodes of either time or frequency, and (3) so-called single-hop transfer nodes connect only directly adjacent channel coefficient in either time or frequency. The complexity of the message generation depends again on how many coefficients nodes are connected to a transfer node. Hereby, the multi-dimensional transfer node exhibits the highest complexity, a medium complexity is attained by the multiple one-dimensional transfer node and the single-hop transfer nodes have the lowest computational complexity in order to establish a message exchange. The advantage of the single-hop transfer node is that coefficients adjacent in arbitrary dimensions can be easily connected. The single-hop transfer nodes are chosen as the preferred implementation due to their favourable complexity and ease of adaptation to arbitrary dimensions. In the following they will be referred to as transfer nodes. Coefficient nodes connected in three dimensions via transfer nodes are shown in Figure 5.4. As can be seen, transfer nodes are represented by a triangle Δ. The index refers to the domain in which messages are exchanged, i.e.: Δ_t facilitates the message exchange in the time domain, Δ_f in the frequency domain, and Δ_s in the spatial



(a) Multi-dimensional transfer node



(b) Multiple one-dimensional transfer nodes



(c) Multiple single-hop transfer nodes

Figure 5.3: Variants of transfer nodes enabling the message exchange between factor graph components.

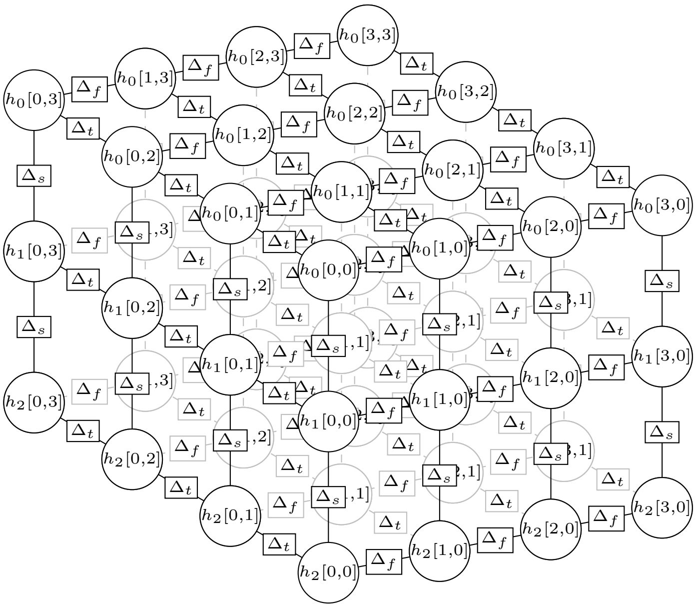


Figure 5.4: Three-dimensional factor graph with transfer nodes connecting coefficient nodes in time, frequency, and space.

domain. Following this concept, an arbitrary amount of dimensions can be added (like polarization, Tx antennas, Rx antennas, ...). The actual transfer function is explained in more detail in the subsequent section. The complete factor graph with transfer nodes and QPSK modulation with two transmit and two receive antennas is given in Figure 5.5. Additionally, messages exchanged between function nodes and variable nodes are included next to the corresponding edges. The nodes representing random variables comprise the unknown data bits (bit nodes), data symbols (symbol nodes), and channel coefficients (coefficient nodes). The received samples (observation nodes) connect symbol nodes and coefficient nodes. Mapping nodes establish the message exchange between bit nodes and symbol nodes according to the modulation format. The mapping nodes are depicted as a black dot within a rectangle.

An MD-GSIR iteration is outlined below, with each step being explained in detail in the corresponding sections.

1. At bit nodes, bit probabilities $P(c_i = \pm 1)$ are calculated and sent to mapping nodes, represented by the message $\mu_{C\to f_C}(c_i)$.

2. At the mapping nodes, corresponding symbol probabilities are calculated according to

$$P(x_m) = \prod_{i=1}^{N_\mathrm{b}} \mu_{C\to f_C}(c_i). \tag{5.9}$$

Subsequently, the symbol probabilities are sent to observation nodes with the message $\mu_{X\to f_Y}(x_m)$.

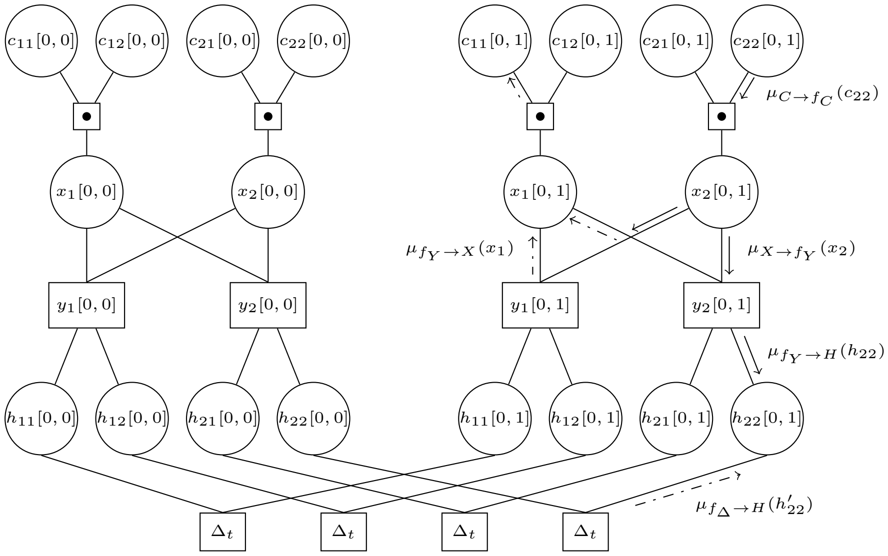


Figure 5.5: Factor graph structure of a 2×2 MIMO-OFDM system and QPSK modulation. Without loss of generality, the transfer nodes connect adjacent coefficient nodes in the time domain.

3. The symbol probabilities represented by the message $\mu_{X\to f_Y}(x_m)$ are used at the observation nodes in combination with the received samples, to calculate the corresponding channel coefficient estimates. The generation of the message $\mu_{f_Y\to H}(h_{n,m})$ is detailed in Section 5.1.3.

4. In order to facilitate message exchange throughout the entire frame, channel coefficients of neighboring OFDM subcarriers, OFDM symbols, transmit and receive antennas are connected via transfer nodes. A message sent from a coefficient node to a transfer node is denoted by $\mu_{H\to f_\Delta}(h_{n,m})$, whereas the message from a transfer node to coefficient node is given by $\mu_{f_\Delta\to H}(h'_{n,m})$. The message exchange within transfer nodes is investigated in Section 5.1.4.

5. By combining the messages of neighboring coefficient nodes, refined coefficient estimates are obtained and sent back to observation nodes, where in turn refined symbol probabilities are generated (cf. Section 5.1.6). The combination of messages is examined in detail in Section 5.3 for correlated as well as uncorrelated random variables.

6. Finally, the mapping nodes calculate LLR values for the bit nodes, which are passed to the channel decoder for further processing.

5.1.3 Soft Channel Estimation

The task of soft channel estimation—located within the observation nodes—is to compute the conditional pdf $p(y_n|h_{n,m})$. During initialization, only the information of pilot symbols is utilized, while additional information of data symbols is exploited in subsequent iterations. Without loss of generality, the OFDM symbol and subcarrier index k and l are omitted in the following to improve readability.

As described previously, to generate a message for a coefficient node $h_{n,m}$ all connected edges have to be considered, except the edge which connects the coefficient node $h_{n,m}$ itself. The corresponding message from an observation node to a coefficient node is given by

$$\mu_{f_Y \to H}(h_{n,m}) = \sum_{x \in \mathcal{X}} \int_{\sim h_{n,m}} \prod_{i=1}^{N_{\mathrm{T}}} p(y_n | h_{n,i}, x_i) \mu_{X \to f_Y}(x_i) \prod_{i=1, i \neq m}^{N_{\mathrm{T}}} \mu_{H \to f_Y}(h_{n,i}) d\boldsymbol{h}_{n,i}. \tag{5.10}$$

As a matter of fact, the message $\mu_{f_Y \to H}(h_{n,m})$ can be well represented by a Gaussian mixture model. However, the exchange of mixed Gaussian distributions on the basis of the sum-product algorithm would quickly be computationally intractable [BKM$^+$12]. Thus, to reduce the computational complexity, (5.10) is approximated by a Gaussian distribution: $p(y_n | h_{n,m}) \propto p(h_{n,m})$. The effective noise term $\zeta_{n,m}$ of (3.21) is approximated by a Gaussian variable, which is characterized by $\zeta_{n,m} \sim \mathcal{CN}(\mu_{\zeta_{n,m}}, \sigma^2_{\zeta_{n,m}})$. Note here, that this is the equivalent Gaussian approximation as done for data detection in Section 3.1.2. The message $\mu_{f_Y \to H}(h_{n,m})$ is calculated as follows:

$$\begin{aligned} \mu_{f_Y \to H}(h_{n,m}) &= \sum_{x_m \in \mathcal{S}} \mu_{X \to f_Y}(x_m) \cdot p\left(y_h | h_{n,m}, x_m\right) \\ &= \sum_{x_m \in \mathcal{S}} \mu_{X \to f_Y}(x_m) \cdot \frac{1}{\pi \sigma^2_{\zeta_{n,m}}} \cdot \exp\left(-\frac{|h_{n,m} - (y_n - \mu_{\zeta_{n,m}})/x_m|^2}{\sigma^2_{\zeta_{n,m}} / |x_m|^2} \right), \end{aligned} \tag{5.11}$$

where as before $\mathcal{S}$ refers to the symbol alphabet of one transmit antenna. If transmitted symbols are reliably detected during iterations, a symbol x_i is detected with a probability $P(x_i) \gg P(x_j), i \neq j$. Hence, if one symbol hypotheses dominates the multivariate Gaussian distribution, the pdf can be well approximated by a Gaussian distribution. The mean and variance of $\mu_{f_Y \to H}(h_{n,m})$ are thus given by [WHS12]:

$$\mu_{h,n,m} = \alpha \left(y_n - \mu_{\zeta_{n,m}} \right) \sum_{x_m \in \mathcal{S}} \frac{\mu_{X \to f_Y}(x_m)}{x_m |x_m|^2}, \tag{5.12}$$

$$\sigma^2_{h,n,m} = \alpha \left(\sigma^2_{\zeta_{n,m}} + |y_n - \mu_{\zeta_{n,m}}|^2 \right) \sum_{x_m \in \mathcal{S}} \frac{\mu_{X \to f_Y}(x_m)}{|x_m|^4} - |\mu_{h,n,m}|^2, \tag{5.13}$$

where α is a normalization factor defined as

$$\alpha = \frac{1}{\sum\limits_{x_m \in \mathcal{S}} \mu_{X \to f_Y}(x_m) / |x_m|^2}. \tag{5.14}$$

Correspondingly, the message which is sent from an observation node to a coefficient is represented as a Gaussian distribution according to

$$\mu_{f_Y \to H}(h_{n,m}) \sim \mathcal{CN}\left(\mu_{h,n,m}, \sigma^2_{h,n,m}\right), \tag{5.15}$$

whereas the mean value $\mu_{h,n,m}$ refers to the hard estimate of the channel coefficient and the variance $\sigma^2_{h,n,m}$ is interpreted as reliability information. The approximation of the mixed Gaussian random variable by a single Gaussian pdf reduces the complexity of the message passing algorithm significantly.

5.1.4 Transfer Nodes

In the absence of pilot information on all OFDM subcarriers it has been concluded earlier that a message exchange between the single components of the factor graph is necessary to provide the information throughout the entire burst. The proposed transfer nodes are the key element of the factor graph and facilitate a low complexity message exchange. Furthermore, the channel estimation accuracy may be improved by utilizing the reliability information of neighboring channel coefficients. Specifically, a transfer node describes the deviation between channel coefficients $h_{n,m}[l,k]$ and $h_{n',m'}[l+l',k+k']$, which are denoted by h and h' to simplify the notation.

Among others, two popular approaches exist in order to model the transfer function, namely a first-order autoregressive model AR(1) and/or a Gaussian random walk. The AR(1) model is the intuitive solution since it is a stationary model, opposed to the random walk which is non-stationary. Additionally, good performance has been reported in [ZGH09] for the AR(1) model. However, contrary to intuition, the random walk model is favored here due to its slightly lower complexity and better performance when used for the transfer of messages in time and frequency. A brief performance comparison of the AR(1) and the random walk model with a single antenna system in the time domain can be found in [WHS12].

In general, random walks resemble a motion by successively adding random variables. They have been successfully applied in many fields, including economics, psychology, computer science, and biology. In classical estimation theory, random walks resemble a non-stationary process, however in factor graphs taking the message combining step into account, this is not the case. In order to explain the aspect of the non-stationarity in more detail, the mathematical description of the AR(1) model and the Gaussian random walk are briefly discussed. The stationary AR(1) model is given by

$$X_{t+1} = c \cdot X_t + Z_t, \quad |c| < 1 \tag{5.16}$$

and the non-stationary Gaussian random walk is represented as

$$X_{t+1} = c \cdot X_t + Z_t, \quad c = 1. \tag{5.17}$$

For both models, X_t is assumed to be independent of Z_t and $Z_t \sim$ i.i.d $\mathcal{N}(0, \sigma_Z^2)$. Since the scaling factor c is equal to one for the Gaussian random walk, the influence of the variable X_t is not limited. Hence, as $t \to \infty$ the variance of X_{t+1} results in $\mathrm{var}(X_{t+1}) = t\sigma_X^2 \to \infty$. Combined with factor graphs, the behavior of the Gaussian random walk is changed fundamentally, since the variance of a message defines its influence during the message combining process. More specifically, at each node of a factor graph, messages are combined before being passed to the next node. Hereby, messages with a large variance are considered to be unreliable and thus contribute less to the combined message. Inherently, the influence of a message is limited. Exemplary, the Gaussian random walk is implemented in the transfer nodes shown in the simplified factor graph in Figure 5.6. The message exchange starts at X_0 and traverses the factor graph to the right side. At each node, two variables are combined. The variance of the combined variable is given in Figure 5.7 for a fixed variance $\sigma_Z = 0.1$ and varying variances of X_t as a function of

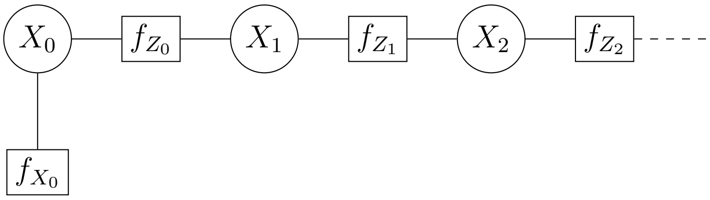


Figure 5.6: Simplified message exchange and combining within a factor graph.

combining steps. As can be seen, the variance of the combined message is asymptotic stationary. Moreover, the variance will converge to

$$\sigma_X^2 = \frac{\sqrt{4\sigma_Z^2\sigma_X^2 + (\sigma_Z^2)^2} - \sigma_Z^2}{2} \tag{5.18}$$

after a few combinations. The amount of combinations depends on the initial variance of X_t and the variance of Z_t. It can be concluded from these results, that the advantage of the Gaussian random walk compared to the AR(1) model within factor graphs is twofold, providing improved performance at lower complexity. Both models can be treated as stationary given a sufficient amount of combining steps.

The transfer of a variable via a transfer node can be written as

$$\Delta_{n',m'}[l',k'] \doteq h - \omega h', \quad |\omega| = 1. \tag{5.19}$$

The tuning factor $\omega \in \mathbb{C}$ depends on the correlation properties between adjacent coefficient nodes. For a symmetrically distributed spectrum, the correlation function is real valued such that $\omega = 1$, whereas the tuning factor $\omega = \exp(j\varphi)$ is complex valued for non-symmetric distributions. Moreover, a transfer node is approximated by a zero-mean

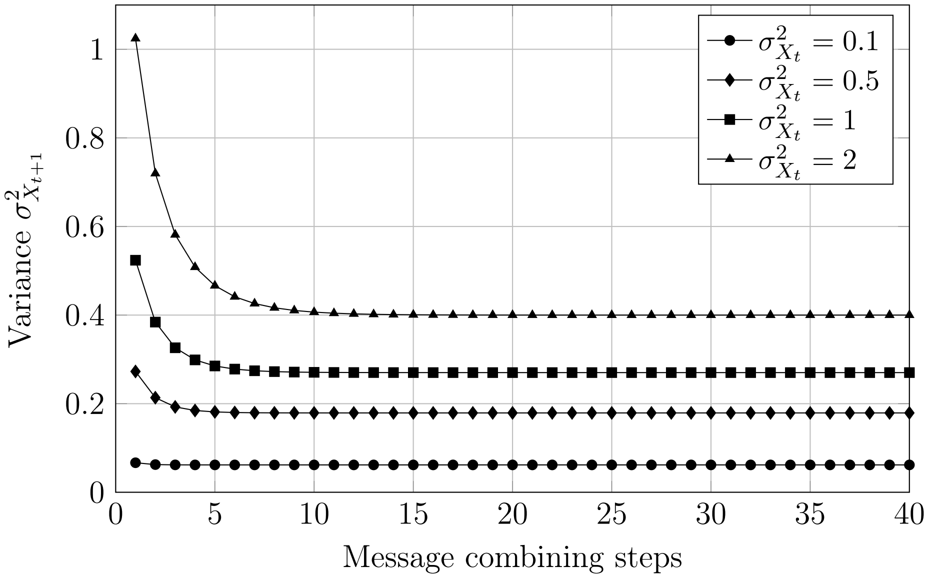


Figure 5.7: Variance of the outgoing message after combining process.

Gaussian pdf:

$$\Delta_{n',m'}[l',k'] \sim \mathcal{N}\left(0, \sigma^2_{\Delta,n',m'}[l',k']\right). \tag{5.20}$$

According to (5.19), information between adjacent channel coefficients is exchanged as follows:

$$\mu_{h'} = \omega \cdot \mu_h, \tag{5.21}$$

$$\sigma^2_{h'} = \sigma^2_h + \sigma^2_{\Delta,n',m'}[l',k']. \tag{5.22}$$

A message exchanged within a transfer node scales its mean $\mu_{h'}$ by a factor of ω, whereas its variance $\sigma^2_{h'}$ is increased by the variance of the domain-specific transfer node. Since the variance of a channel coefficient is interpreted as reliability information (cf. Section 5.1.3), the transfer function reduces the reliability of the message with each node; inherently decreasing the influence of this message on the overall message generation.

The variance of a transfer node is calculated as follows

$$\begin{aligned}\sigma^2_{\Delta,n',m'}[l',k'] &= \mathrm{E}\left\{|h-\omega h'|^2\right\}\\ &= \underbrace{\mathrm{E}\left\{|h|^2\right\}}_{\doteq 1} + \underbrace{\mathrm{E}\left\{|h'|^2\right\}}_{\doteq 1} - \mathrm{E}\left\{\omega h h'^*\right\} - \mathrm{E}\left\{\omega h^* h'\right\}\\ &= 2\left(1 - \mathrm{Re}\left[\omega\, \mathrm{E}\left\{h^* h'\right\}\right]\right),\end{aligned} \tag{5.23}$$

where $\mathrm{E}\{h^* h'\}$ refers to the multi-dimensional auto-correlation function $\theta_{HH}(\Delta x, \Delta f, \Delta t)$ between two channel coefficients h and h'.

In order to prevent short cycles and the exchange of intrinsic information, a transfer node only connects two neighboring channel coefficients of one domain, i.e. $|l'| + |k'| + |n'| + |m'| = 1$. The resulting grid is shown in two and three dimensions in Figure 5.3 and Figure 5.4, respectively. Short cycles refer to the situation in which a node receives a message which passes only a few nodes until it reaches its origin again. Accordingly, the arriving message will still contain a large fraction of the message that was sent since only few combinations have been performed. Nevertheless, an intelligent scheduling can mitigate the effect of short cycles, which is discussed in detail in Section 5.2. Additionally, since transfer nodes connect channel coefficients which are neighbors in the same domain, the required variance $\sigma^2_{\Delta,n',m'}[l',k']$ can be calculated independently for each domain; either space, frequency or time.

Spatial Domain

The variance for each domain depends on the distribution specific dispersion parameters. Revisiting (2.35), it can be seen that the the spatial domain depends on multiple parameters: Namely, the spacing between transmit and receive antennas d_{Tx} and d_{Rx}, the wavelength λ, and the distribution of the azimuth angle of departure φ as well as azimuth angle of arrival ϕ.

In order to derive the tuning factor ω_s of the spatial domain the transmitter side is chosen without loss of generality:

$$\sigma^2_{\Delta,s} = 2\left(1 - \underbrace{\mathrm{Re}\left[\mathbf{E}\left\{\omega_s \exp\left(j2\pi\frac{d_{\mathrm{Tx}}}{\lambda}\sin(\varphi)\right)\right\}\right]}_{\mathbf{C}}\right). \tag{5.24}$$

For the distributions of the azimuth angles and angular spreads discussed in Section 2.1.2, the autocorrelation functions are real-valued and thus, the tuning factor of the spatial domain ω_s results in a value of one. For the more complicated distributions used within the WINNER model, a uniform distribution of the angular spread ϑ within the interval $[\varphi - \vartheta/2, \varphi + \vartheta/2]$ per cluster is assumed. With the spatial autocorrelation function given in [SW94], $\mathbf{C}$ results in

$$\begin{aligned}\mathbf{C} = \mathrm{Re}\Bigg[\omega_s \cdot \Bigg(&\underbrace{J_0(z) + 2\cdot\sum_{m=1}^{\infty} J_{2m}(z)\cos(2m\varphi)\frac{\sin(2m\vartheta)}{2m\vartheta}}_{\mathbf{A}} + \\ &\underbrace{j\,2\cdot\sum_{m=0}^{\infty} J_{2m+1}(z)\sin((2m+1)\varphi)\frac{\sin((2m+1)\vartheta)}{(2m+1)\vartheta}}_{\mathbf{B}}\Bigg)\Bigg]\end{aligned} \tag{5.25}$$

with $z = 2\pi d_{\mathrm{Tx}}/\lambda$. To determine the spatial tuning factor $\omega_s = \exp(jx) = \cos(x) + j\sin(x)$, $\mathbf{C}$ can be rewritten as

$$\mathbf{C} = \mathrm{Re}\left[\omega_s\,(A + j\cdot B)\right] = \cos(x)\cdot A - \sin(x)\cdot B. \tag{5.26}$$

To find an extrema of (5.26), $(d/dx(\mathbf{C}) = 0)$ is derived:

$$\begin{aligned}&-\sin(x)\cdot A - \cos(x)\cdot B = 0\\ &\Rightarrow \tan(x) = -\frac{B}{A}\\ &\Rightarrow x = \tan^{-1}\left(-\frac{B}{A}\right).\end{aligned} \tag{5.27}$$

Which results in a tuning factor

$$\omega_s = \exp\left(j\tan^{-1}\left(-\frac{B}{A}\right)\right). \tag{5.28}$$

Revisiting (5.24) and inserting (5.28), yields

$$\sigma^2_{\Delta,s} = 2\left(1 - \cos(\tan^{-1}(-B/A))\cdot A - \sin(\tan^{-1}(-B/A))\cdot B\right). \tag{5.29}$$

Since

$$\cos(\tan^{-1}(-B/A)) = \frac{1}{\sqrt{1+(B/A)^2}}, \tag{5.30}$$

$$\sin(\tan^{-1}(-B/A)) = \frac{-B/A}{\sqrt{1+(B/A)^2}}, \tag{5.31}$$

(5.29) can be rewritten as

$$\sigma^2_{\Delta,\mathrm{s}} = 2\left(1 - \frac{|A|\sqrt{A^2+B^2}}{A}\right). \tag{5.32}$$

Moreover, for small angular spreads ($\vartheta \leq \pi/6$), (5.24) can be approximated by [Aue09]:

$$\sigma^2_{\Delta,\mathrm{s}} = 2\left(1 - \cos\left(z \cdot \sin(\varphi)\right) \cdot \mathrm{sinc}\left(\frac{d_{\mathrm{Tx}}}{\lambda}\vartheta\cos(\varphi)\right)\right). \tag{5.33}$$

Apart from the uniform distribution of the angular spread, typical distributions are the von Mises distribution or the Laplacian distribution [Fle00].

Frequency Domain

Detailed information about the correlation function is important to accurately predict the variance of the transfer nodes. Given the tabulated parameters of the WINNER channel models (cf. Appendix B), the frequency correlation function can be calculated exactly by means of a discrete Fourier transformation of the PDP:

$$\theta_{HH}\left(F_{\mathrm{s}}\right) = \sum_{c=1}^{M_{\mathrm{c}}} P_c \cdot \exp\left(-j2\pi F_{\mathrm{s}}\tau_c\right). \tag{5.34}$$

Inserting (5.34) into (5.23) yields

$$\sigma^2_{\Delta,\mathrm{f}} = 2\left(1 - \omega_f \cdot \left(\sum_{c=1}^{M_{\mathrm{c}}} P_c \cdot \exp\left(-j2\pi F_{\mathrm{s}}\tau_c\right)\right)\right), \tag{5.35}$$

whereas the tuning factor ω_f is obtained by an inverse DFT of the PDP. The knowledge of the exact PDP is not practical but rather serves as a benchmark.

In case exact knowledge of the PDP is not given, a suitable approximation has to be used. A commonly adopted distribution of the power delay profile is described by an exponentially decaying function given by (2.28):

$$\Theta_{SS}\left(\tau\right) = \begin{cases} \frac{1}{\tau_{\mathrm{rms}}}\exp\left(-\tau/\tau_{\mathrm{rms}}\right) & \text{for } 0 \leq \tau \leq \tau_{\max} \\ 0 & \text{else.} \end{cases}$$

With the corresponding autocorrelation function (2.29), the variance of the frequency domain transfer node can be written as

$$\sigma^2_{\Delta,\mathrm{f}} = 2\left(1 - \mathrm{Re}\left\{\frac{\omega_f}{1 + j2\pi\tau_{\mathrm{rms}}F_{\mathrm{s}}}\right\}\right). \tag{5.36}$$

By setting $\omega_f = 1/(1 - j2\pi\tau_{\mathrm{rms}}F_{\mathrm{s}})$, (5.36) is minimized and results in

$$\sigma^2_{\Delta,\mathrm{f}} = 2\left(1 - \frac{1}{1 + 4\pi^2\tau^2_{\mathrm{rms}}F^2_{\mathrm{s}}}\right). \tag{5.37}$$

In case almost no information about the PDP is available, a common approximation is to assume that propagation delays are uniformly distributed within the interval $[0, \tau_{\max}]$. Then, the variance of a transfer node in the frequency domain between adjacent subcarriers amounts to

$$\sigma^2_{\Delta,\mathrm{f}} = 2\left(1 - \mathrm{sinc}\left(\tau_{\max}F_{\mathrm{s}}\right)\mathrm{Re}\left\{\omega_f \exp(j2\pi\frac{\tau_{\max}}{2}F_{\mathrm{s}})\right\}\right). \tag{5.38}$$

The variance (5.38) is minimized when the real part is maximized. By setting the value of the tuning factor ω_f as

$$\omega_f = \exp(-j2\pi\frac{\tau_{\max}}{2}F_{\mathrm{s}}) = \exp(-j\pi\tau_{\max}F_{\mathrm{s}}), \tag{5.39}$$

the imaginary part in (5.38) diminishes. By substituting (5.39) into (5.38), the minimum variance for the frequency domain transfer node results in

$$\sigma^2_{\Delta,\mathrm{f}} = 2\left(1 - \mathrm{sinc}(\tau_{\max}F_{\mathrm{s}})\right). \tag{5.40}$$

Note that setting the tuning factor to $\omega_f = \exp(-j\pi\tau_{\max}F_{\mathrm{s}})$ is equivalent to shifting the power delay profile by $-\tau_{\max}/2$ in the delay domain [Hoe91, HKR97b]. Hence, the autocorrelation function with tuning factor is equal to (2.25).

Time Domain

Analogous to the spatial and frequency domain, channel coefficients of adjacent OFDM symbols are connected with a transfer node, modeling the variations of a time-varying channel. The variance of adjacent channel coefficients is determined by calculating (5.23) between two adjacent OFDM symbols.

Assuming a uniform Doppler power spectral density defined in the range $\pm f_{\mathrm{D,max}}$, the variance of $\sigma^2_{\Delta,\mathrm{t}}$ yields

$$\sigma^2_{\Delta,\mathrm{t}} = 2\left(1 - \mathrm{sinc}(2f_{\mathrm{D,max}}T_{\mathrm{s}})\right). \tag{5.41}$$

Due to the symmetric distribution of the Doppler frequencies, the time domain correlation function is real valued and thus, the tuning factor is set to $\omega_t = 1$.

A further commonly used distribution of the Doppler frequencies is given by the Jakes power spectral density, for which the variance of $\sigma^2_{\Delta,\mathrm{t}}$ is given by

$$\sigma^2_{\Delta,\mathrm{t}} = 2\left(1 - J_0(2\pi f_{\mathrm{D,max}}T_{\mathrm{s}})\right). \tag{5.42}$$

(a) Messages sent to a coefficient node (b) Message sent from a coefficient node

Figure 5.8: Message exchange at a coefficient node.

5.1.5 Information Exchange at Coefficient Nodes

In general, for the generation of a message at a node of a factor graph all connected edges need to be considered, except the edge for which the message is generated. For the MD-GSIR, a coefficient node is connected to two transfer nodes in each domain (space, frequency, and time) and one observation node, as illustrated in Figure 5.8. Thus, a coefficient node has to generate messages sent to transfer nodes

$$\mu_{H \to f_\Delta}(h_{n,m}^i) = \mu_{f_Y \to H}(h_{n,m}) \cdot \prod_{\sim h_{n,m}} \mu_{f_\Delta \to H}(h_{n',m'}) \tag{5.43}$$

and a message sent to an observation node

$$\mu_{H \to f_Y}(h_{n,m}^i) = \prod \mu_{f_\Delta \to H}(h_{n',m'}). \tag{5.44}$$

It is supposed that a channel coefficient receives the messages $\mu_{f_\Delta \to H}(h_j) \sim \mathcal{CN}(\mu_j, \sigma_j^2)$, $j \in \{1, \ldots, N\}$ from N connected edges. Given uncorrelated messages, the product of multiple Gaussian random variables results in a Gaussian random variable with mean and variance given by (3.51):

$$\hat{\mu} = \frac{\sum_{i=1}^{N} \frac{\hat{\mu}_i}{\hat{\sigma}_i^2}}{\sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2}}, \qquad \hat{\sigma}^2 = \frac{1}{\sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2}}.$$

Due to cycles within the factor graph structure as well as the probabilistic model, correlation may arise during the message exchange. Correspondingly, the combined mean and variance value of multiple correlated Gaussian random variables is given by (3.52) and (3.55), respectively:

$$\hat{\mu} = \sum_{i=1}^{N_{\mathrm{RM}}} \alpha_i \hat{\mu}_i. \qquad \hat{\sigma}^2 = \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\alpha}.$$

The combining of correlated random variables is described in detail in Section 3.4, while the application to the MD-GSIR is investigated in Section 5.3.

5.1.6 Soft Data Detection

Using the updated messages received from connected coefficient nodes $\mu_{H\to f_Y}(h^i_{n,m})$, an observation node y_n calculates

$$\mu_{f_Y\to X}(x_m) = p(y_n|x_m). \tag{5.45}$$

A selection of detection algorithms which are able to efficiently calculate the conditional probability $p(y_n|x_m)$ has been presented in Section 3.1. Exemplary, for the case of Gaussian detection, the message from an observation node to a symbol node is given by

$$\begin{aligned} p(y_n|x_m) &= \int p(y_n|h^i_{n,m}, x_m)p(h^i_{n,m})dh^i_{n,m} \\ &= \frac{1}{\pi(\sigma^2_{i,n,m}|x_m|^2 + \sigma^2_{\zeta_{n,m}})} \cdot \exp\left(-\frac{|y_n - (\mu_{i,n,m}x_m + \mu_{\zeta_{n,m}})|^2}{\sigma^2_{i,n,m}|x_m|^2 + \sigma^2_{\zeta_{n,m}}}\right). \end{aligned} \tag{5.46}$$

It is important to emphasize that besides the variance of the effective noise also the variance of the channel coefficient is considered in the denominator. Hence, an unreliable channel estimate reduces the log-likelihood ratio for the corresponding data symbol. Moreover, it can be seen from (5.46) that the conditional pdf is calculated for one specific observation node. It is a special property of the proposed factor graph structure that at each observation node only partial information of the channel matrix is available. By carefully examining Figure 5.2, it can be seen that only N_T channel coefficients out of $N_\mathrm{T} \cdot N_\mathrm{R}$ are connected to each observation node. Since information is only allowed to be passed along associated edges, the information of unconnected nodes is not available. Accordingly, detection methods which require the complete channel matrix—such as methods based on QR decomposition—can not be implemented within the factor graph without a change in the underlying structure. On the one hand, the chosen factorization allows a low complexity message generation, but on the other hand, the number of possible detection algorithms is limited, as mentioned already in Section 3.1.1. This limitation, however, is not severe in the presence of the Gaussian tree search algorithm, which offers a versatile trade-off between computational complexity and performance.

For the calculation of the extrinsic LLR values, N_R independent observations are provided at each bit node. The calculation of the log likelihood ratios is as follows:

$$\mathrm{L_P}(c_n|y_n) = \sum_{n=1}^{N_\mathrm{R}} \mu_{X_m\to f_C}(c_n), \tag{5.47}$$

whereas the single message is calculated again by many alternatives, as presented in Section 3.1. The Gaussian detector in combination with the max-log approximation given in (3.22) yields

$$\mathrm{L_P}(c_n|y_n) \approx \max_{x_m\in\mathcal{S}^i_{+1}} \left\{\Lambda^{\mathrm{GA}}(x_m)\right\} - \max_{x_m\in\mathcal{S}^i_{-1}} \left\{\Lambda^{\mathrm{GA}}(x_m)\right\}.$$

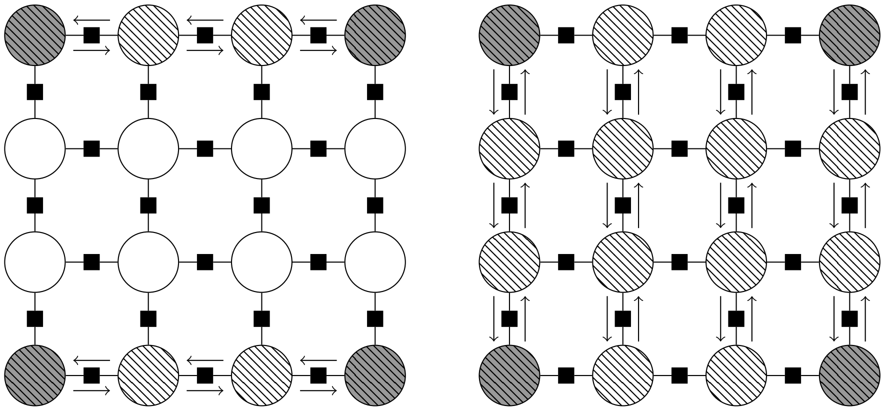
Obviously, the optimum APP detector or the Gaussian tree search detector can also be applied in a straightforward manner.

5.2 Cycles & Scheduling

The structure of the underlying factor graph is a trade-off between complexity and performance. The complexity of the message generation can only be reduced up to a certain extent before the achievable performance is impaired. In favor of complexity, cycles have been allowed in the structure of the factor graph, despite the fact that short cycles may deteriorate the performance of the sum-product algorithm. Short cycles occur when a message which leaves a node, "travels" only a few nodes until it is send back to its origin. Strategies have been devised that merge several nodes in order to prevent short cycles in the message exchange but they inherently involve higher computational complexity w.r.t. the message generation and exchange. Nevertheless, the insight that graphs with cycles can achieve high performances came as a revolution [FM97] and created the research area of loopy belief propagation [YFW00, YHB04, IFIW05]. It can be concluded that loopy belief propagation in combination with the proposed graph structure requires a suitable scheduling in order to achieve a good performance.

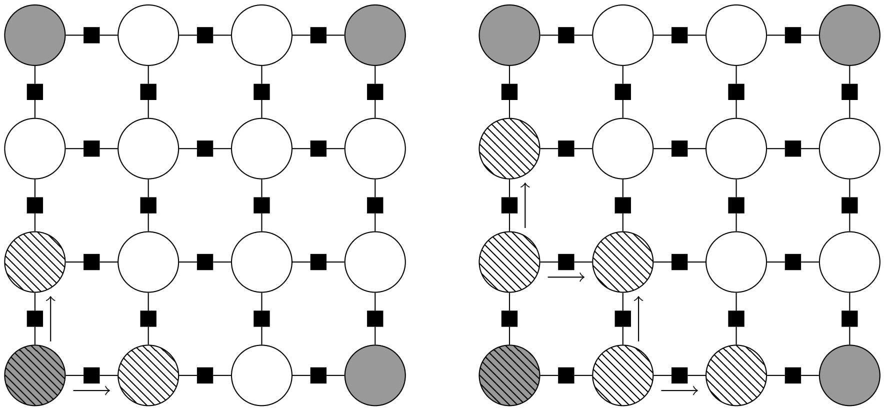
For a factor graph with cycles, message scheduling should ensure that the message which is received by a node, contains as few information of the node itself. Cycles exist at two locations in the proposed factor graph structure. The first cycle is between symbol nodes and observation nodes, as shown in Figure 5.2. The information of the first transmit antenna can be passed along the edges such that it is send back to its origin in only four steps. However, this cycle can be completely avoided by means of scheduling, which has been explained in detail in Section 5.1.2. The second cycle is located in the multi-dimensional grid created by the coefficient and transfer nodes. Here, cycles can not be completely avoided. Nevertheless, a suitable schedule can extend the length of a cycle such that the impairing effects are mitigated.

A possible schedule is to exchange messages of different domains successively, e.g., messages are exchanged in the time domain first, afterwards messages are exchanged in the frequency domain, or vice versa. The underlying principle is dubbed *two-way schedule* in [KF98]. A message is sent from a coefficient node to a transfer node and from there to the next coefficient node in the same domain; thus producing a burst that traverses in the direction of the desired domain. If the last coefficient node of the selected domain is reached, messages are propagated in the reversed direction. The two-way schedule is finished when the coefficient node that started the message exchange receives a message. The schedule hereby ensures that incoming messages are the combination of the maximum number of outgoing messages in one domain. The schedule is illustrated in Figure 5.9a and Figure 5.9b, respectively. A node that received and send a message is marked with a hatching, whereas a gray node refers to a pilot position. In this example, the horizontal domain is chosen to exchange its information first. In a second step, the vertical domain follows. Following the principle of the sum-product algorithm each node receives a message that does not contain information from itself, hence, only extrinsic messages are exchanged. An alternative scheduling is given by the *flooding schedule* [KF98], which distributes the messages simultaneously in all domains. The first and second step of the flooding schedule is shown in Figure 5.9c and Figure 5.9d. It can be seen that the information of the bottom left coefficient node is distributed to its neighbors adjacent in two domains. Already in the second step, the same information is combined at the



(a) Exchange of the first domain with the two-way schedule

(b) Exchange of the second domain with the two-way schedule



(c) First message exchange step with the flooding schedule

(d) Second message exchange step with the two-way schedule

Figure 5.9: Simplified factor graph to illustrate the message exchange of the two-way and flooding schedule.

neighboring upper right node. Obviously, messages arriving at a node contain intrinsic information, which leads to a degradation w.r.t. performance. Therefore, the two-way schedule is applied for the message exchange schedule throughout all simulations.

By adapting the two-way schedule, the domain, which starts to exchange its information first, has to be chosen as well. The aim of the schedule should be to combine the messages in an order such that the resulting message will have the lowest possible variance. Interestingly, the variance of the combined message does not only depend on the fading of the corresponding domains but also on the position of a node, which combines the arriving messages. In order to explain this effect in more detail it is assumed that the two-way schedule finished its first iteration and all coefficient nodes are provided with information. The corresponding situation is depicted in Figure 5.9b. Subsequently, the nodes combine all messages received at connected edges and send the resulting message to an observation node, which is not shown in the simplified graph structure. The variance of messages of one domain after combining is plotted in Figure 5.10 as a function of the

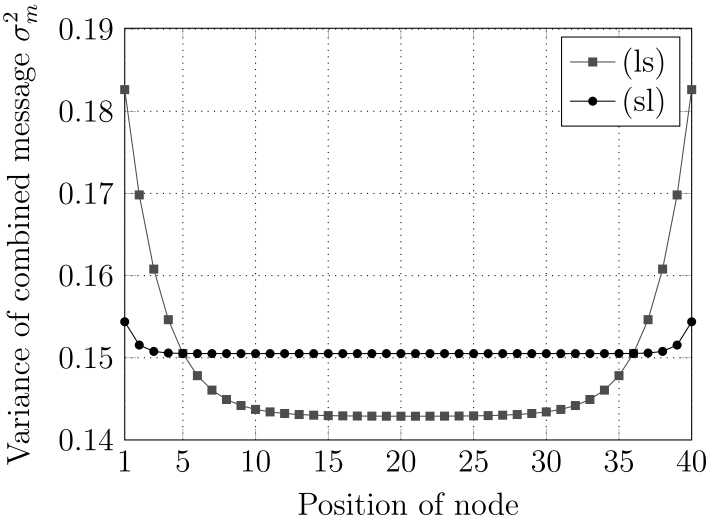
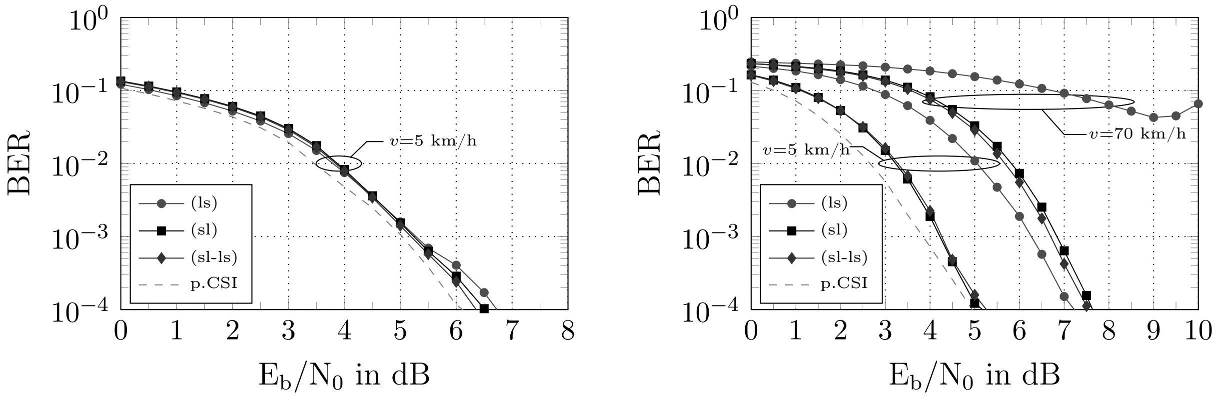


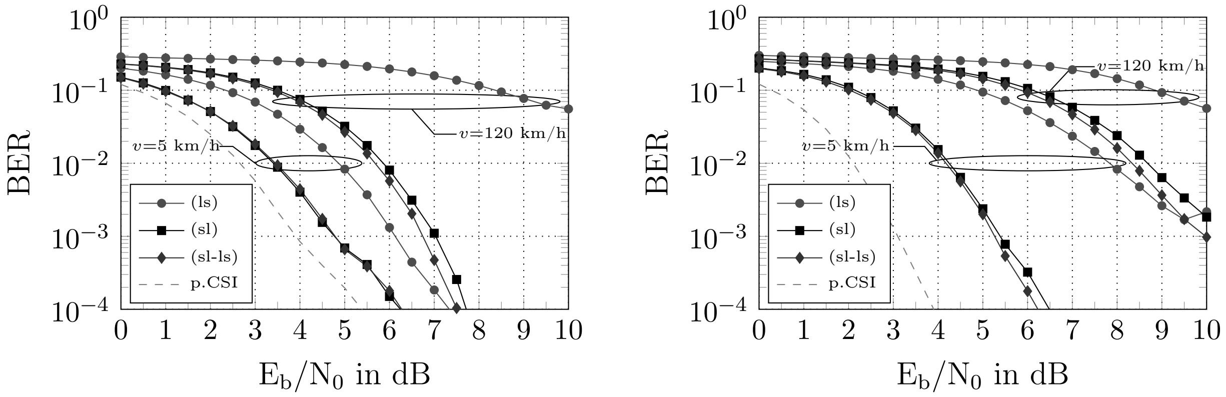
Figure 5.10: Effect of scheduling on the resulting variance of combined messages.

position within the burst. A total of 40 × 40 nodes have been simulated to exchange and combine their information. Without loss of generality, the horizontal domain is set to a larger variance compared to the vertical domain. The results are given for two variants of the two-way schedule: (1) the domain with the smaller variance is started first (sl) and (2) the domain with the larger variance is started first (ls). It can be seen that the overall smallest variance is obtained with the latter schedule, which exchanges the messages with the larger variance first. However, the resulting variances at the edges are much higher. Overall, the bowl shape character is strongly pronounced with the (ls) schedule. Although the minimum variance of the (sl) schedule is larger than the minimum of the (ls) schedule, the resulting variance is rather constant. A property which is also beneficial for the assumption of the asymptotic stationarity of the transfer nodes (cf. Section 5.1.4). BER results as a function of scheduling are given in Figure 5.11 for the four considered WINNER channel models with a low velocity of $v = 5$ km/h and a high velocity which depends on the maximum defined velocity of the corresponding channel model. In addition to the above mentioned two scheduling variants, a third variant is included, which changes the schedule after initialization, i.e. the (sl) schedule is applied during initialization and the (ls) schedule is used during subsequent iterations, correspondingly the schedule is abbreviated with (sl-ls). The rationale behind the third schedule is that a quasi-constant variance is obtained for all nodes after initialization. In the subsequent iterations, the variance of data symbols vary significantly due to fading. A constant variance is unlikely, independent of the underlying schedule. Hence, the aim of the schedule is to obtain the minimum variance instead of a constant variance. In order to determine a suitable schedule, the exchange of messages in the spatial domain is not considered yet. The reasons for this are two-fold: First, the number of OFDM symbols and/or OFDM subcarriers significantly exceeds the number of antennas. However, a large number of variables is needed to reach the asymptotic stationarity of the proposed message exchange with the transfer nodes. Second, the variance in the spatial domain is typically much higher, therefore messages are treated to be unreliable and, under circumstances, are not considered in the message combining at all. The aspect of 3D-channel estimation is elaborated in the following section in more detail.

Exemplary for the WINNER C1 NLOS channel, the schedule is explained with numerical values. Based on (5.35), the variance of transfer nodes in the frequency domain is given by $\sigma^2_{\Delta,\mathrm{f}} \approx 1.064 \cdot 10^{-04}$, whereas the variance of transfer nodes in the time domain for a velocity of $v = 5$ km/h and $v = 120$ km/h and a carrier frequency of $f_\mathrm{c} = 4$ GHz results in $\sigma^2_{\Delta,\mathrm{t}} \approx 3.454 \cdot 10^{-05}$ and $\sigma^2_{\Delta,\mathrm{t}} \approx 1.984 \cdot 10^{-02}$, respectively. Thus, with a low velocity, the transfer nodes in the frequency domain have a higher variance compared to the transfer nodes in time domain. This situation is reversed for a velocity of $v = 120$ km/h. As can be seen from the results shown in Figure 5.11, the third schedule achieves the best performance independent of the channel scenario. The performance of the three schedules is almost identical for the WINNER A1 NLOS channel and close to the performance with perfect channel state information (p.CSI). This is not surprising, since the A1 NLOS channel is not very frequency-selective. In combination with a slow fading in the time domain, the loss due to channel estimation is expected to be small. For the remaining



(a) WINNER A1 NLOS (b) WINNER B1 NLOS

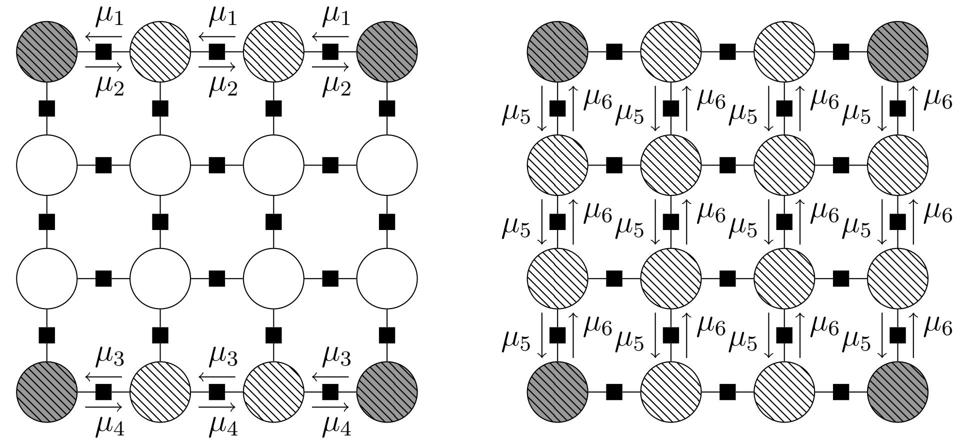


(c) WINNER C1 NLOS (d) WINNER C2 NLOS

Figure 5.11: BER performance as a function of message scheduling with a velocity of $v = 5$ km/h, $v = 120$ km/h, and WINNER C1 NLOS channel model.

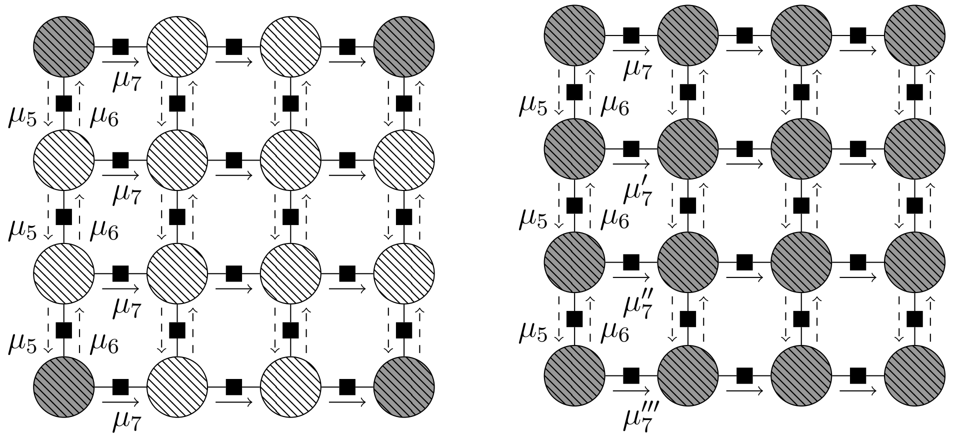
three scenarios, however, the performance of the (ls) schedule is catastrophic. The (sl-ls) schedule is slightly better than the (sl) schedule, whereas a larger gain is achieved with increasing diversity in time and/or frequency domain. Furthermore, the loss due to channel estimation varies for each channel model. Hereby, the C2 NLOS model has the largest loss with 2.5 dB and 7 dB for a velocity of $v = 5$ km/h and $v = 120$ km/h at an BER of 10^{-3}, respectively.

The comparably large loss of 7 dB for the C2 NLOS channel due to channel estimation is partly reasoned in the poor initialization, which is based on a symbol-wise LS approach. By means of an improved initialization, the BER performance improves as well. The aspect of initialization is investigated in Section 5.4. However, initialization does not explain the performance loss completely. Instead, due to the apparent cycles in the factor graph structure, intrinsic messages are exchanged under certain conditions which contribute significantly to the performance loss. Two different scenarios can be identified. In order to illustrate the difference between the two scenarios, the message exchange is again examined in Figure 5.12 . Hereby, the mean value of the exchanged messages is plotted along the corresponding edges. Furthermore, a gray colored circle denotes a coefficient node which contributes to the message exchange. Whether or not a node contributes to the message generation depends on the variance of the message. Pilot symbols and reliably detected data symbols will have a small variance and contribute strongly, whereas unreliably detected symbols are effectively ignored. A quasi-constant channel in time and frequency domain is assumed to clearly illustrate the intrinsic message exchange. During initialization, only pilot symbols provide a priori information and hence, only their information is distributed within the factor graph. In the first step, i.e. the exchange of messages in the horizontal domain, four messages are exchanged with their corresponding mean values represented by μ_i, with $i \in \{1, 2, 3, 4\}$ (cf. Figure 5.12a). In the second step, two messages are combined at each node. In the top row, μ_1 and μ_2 are combined to μ_5, whereas in the bottom row μ_3 and μ_4 generate μ_6 (cf. Figure 5.12b). The information of pilot symbols is now distributed and send to observation nodes where soft data detection is carried out. Subsequently, the channel decoder feeds back extrinsic a priori information, which are used to refine the channel coefficient estimates. Two different scenarios can be identified: (1) data detection is successful and provides reliable a priori information or (2) only unreliable information is available at symbol nodes. In the second case, the information obtained by an unreliable data estimate does not contribute to the overall message exchange since its variance is significantly larger than the messages of previous iterations, which was based on pilot information. This scenario is depicted in Figure 5.12c. Here, coefficient nodes at data positions do not contribute to the message exchange. The previous messages μ_5 and μ_6, depicted in dashed lines, are used to generate the updated message μ_7. It can be seen that the same message w.r.t. the mean value (μ_7) is send from the coefficient nodes at data positions for the horizontal message exchange. Already at the second column of coefficient nodes, the outgoing message μ_7 is combined again with the previous messages μ_5 and μ_6. Since μ_7 already contains these information, no new information is added. However, the message update rule is unaware of this situation and reduces the variance value of the updated messages. Although no new information is added to the message exchange, the variance value is reduced and thus



(a) Initialization - first step

(b) Initialization - second step



(c) First iteration - unreliable data estimates

(d) First iteration - reliable data estimates

Figure 5.12: Illustration of the message schedule in order to highlight weak and strong cycles.

the reliability of the messages is overly optimistic. In subsequent iterations, the messages of data symbols need to have an even lower variance value in order to be considered.

In case of reliable data estimates, each node contributes to the message exchange and sends individual messages (cf. Figure 5.12d). The effect of short cycles is thus mitigated and the exchanged messages are not longer overly confident. This observation is in line with the previous results. In case of slow fading, e.g. with the A1 NLOS channel, data symbols can be reliably detected and the effects of the cycles is not pronounced. With the strong fading of the C2 NLOS channel, the loss due to channel estimation is inherently larger and thus, data symbols are not as reliably detected as with the A1 NLOS channel. The situation is worsened with increasing velocity. Accordingly, the impact of the cycles is most severe for the C2 NLOS channel.

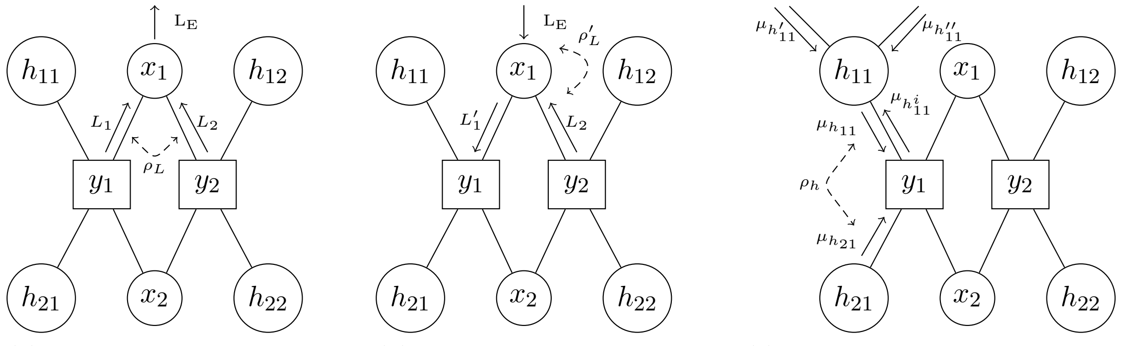
In order to improve the performance in the presence of cycles, the message combining is adapted in the following section to take the correlation between messages into account.

5.3 Correlated Combining

Message combining is an essential component within every factor graph. The message combining as well as the message exchange based on the sum-product algorithm requires extrinsic information as a prerequisite to reach its optimum performance. Correlation between random measures has a significant impact on the combining process as described in the previous section. The proposed factor graph connects adjacent coefficient nodes with transfer nodes to establish a message exchange. The resulting multi-dimensional grid is characterized by cycles. In order to provide the best estimate after combining, correlation has to be considered. While correlation between different random measures is obvious in some cases, it might be inadvertently introduced in other scenarios. It is explained in the following why correlation arises during the message exchange and how the combining of correlated random measures, which is described in Section 3.4, can be integrated within the proposed factor graph structure.

The previously described methods, how to combine multiple correlated observations (cf. Section 3.4.1) and how to combine correlated random variables (cf. Section 3.4.2), are applied at two different nodes in the factor graph. First, random measures in form of L-values of multiple receive antennas as well as a priori and a posteriori L-values have to be combined in the symbol nodes x_m, as illustrated in Figure 5.13a and Figure 5.13b. And second, messages arriving at a coefficient node have to be combined, which is shown in Figure 5.13c. In order to improve the readability of Figure 5.13 only the mean value μ_{h_i} of a message is given along the corresponding edges.

Correlation between L-values of different receive antennas, shown in Figure 5.13a, typically arises for correlated receive antennas. Correlation between a priori and a posteriori information, depicted in Figure 5.13b, may emerge from correlated data symbols due to short sequence lengths and/or cycles within the factor graph. The most prominent examples are graph-based LDPC decoding and a graph-based detector, which considers intersymbol interference. The corresponding correlation coefficients are given by ρ_L, and ρ'_L for messages at observation nodes and ρ_h for message at coefficient nodes, respectively.



(a) Correlation between receive antennas

(b) Correlation between a priori and a posteriori L-values

(c) Correlation between messages of coefficient nodes

Figure 5.13: Exchange of correlated messages in the factor graph.

The combination of messages arriving at a coefficient node is shown in Figure 5.13c. As mentioned previously, a message $\mu_{f_Y \to H}(h_{n,m})$ is represented by a Gaussian distribution and thereby characterized by $\mathcal{CN}(\mu_i, \sigma_i^2)$. Based on the sum-product algorithm, extrinsic information is exchanged between nodes. In a graph with cycles, however, a message does contain a fraction of the nodes' information. It is assumed in the following that the mean values $\mu_{h_{11}}$ and $\mu_{h_{21}}$ depicted in Figure 5.13c, are correlated, e.g. due to the probabilistic model and/or a loopy graph structure. Correspondingly, the mean values $\mu_{h_{11}}$ and $\mu_{h^i_{11}}$ will also be correlated, which has to be considered when $\mu_{f_Y \to H}(h'_{11})$ is combined with messages sent from adjacent transfer nodes. In general, it is difficult to determine the exact correlation between $\mu_{h_{11}}$ and $\mu_{h^i_{11}}$. However, it is closely related to the correlation between $\mu_{h_{11}}$ and $\mu_{h_{21}}$. For the channel scenarios under consideration, a medium to strong correlation between the transmit antennas exist (cf. Section 2.2). It has been previously explained that correlation between messages arises if the message exchange is dominated by the messages obtained by pilot symbols. For iterative MIMO detection, the worst case scenario is given when the transmit symbols of all antennas cannot be reliably detected. If the transmit symbols of one or more antenna can be well detected, their information helps improve the estimates of the remaining transmit symbols since the interference can be subtracted more accurately. Since the messages of pilot nodes are not updated during iterations, only data symbols provide diversity to the message exchange. Without their messages, the exchanged messages remain constant. Therefore, if the message exchange of all transmit antennas is dominated by pilot information and the transmit antennas are sufficiently correlated, the correlation between $\mu_{h_{11}}$ and $\mu_{h'_{11}}$ can be well approximated by using the mean values $\mu_{h_{11}}$ and $\mu_{h_{21}}$ of e.g. one OFDM symbol with L OFDM subcarriers. Obviously, the accuracy of the approximation depends on the length of L. Similarly, the correlation coefficient between L-values can be determined. The calculation of the correlation coefficient is exemplary illustrated in the following.

Two correlation coefficients between $\mu_{h_{11}}$ and $\mu_{h_{21}}$ can be obtained by averaging over time and frequency domain:

$$\varphi_k[k] \doteq \frac{\left| \mathrm{E}\left\{ \left(h_{11}[:,k] - \mu_{h_{11}[:,k]}\right) \cdot \left(h_{21}[:,k] - \mu_{h_{21}[:,k]}\right) \right\} \right|}{\sigma_{h_{11}[:,k]} \cdot \sigma_{h_{21}[:,k]}}, \tag{5.48}$$

$$\varphi_l[l] \doteq \frac{\left| \mathrm{E}\left\{ \left(h_{11}[l,:] - \mu_{h_{11}[l,:]}\right) \cdot \left(h_{21}[l,:] - \mu_{h_{21}[l,:]}\right) \right\} \right|}{\sigma_{h_{11}[l,:]} \cdot \sigma_{h_{21}[l,:]}}, \tag{5.49}$$

where φ_k contains K correlation coefficients and φ_l contains L coefficients. The expectation of each correlation coefficient is generated as follows:

$$\phi_k \doteq \mathrm{E}\left\{\varphi_k\right\} \qquad \phi_l \doteq \mathrm{E}\left\{\varphi_l\right\}. \tag{5.50}$$

Afterwords, the two vectors are combined and the mean is calculated:

$$\mathbf{\Phi} = \begin{bmatrix} (\phi_k[1] + \phi_l[1])/2 & (\phi_k[1] + \phi_l[2])/2 & \cdots & (\phi_k[1] + \phi_l[L])/2 \\ (\phi_k[2] + \phi_l[1])/2 & (\phi_k[2] + \phi_l[2])/2 & \cdots & (\phi_k[2] + \phi_l[L])/2 \\ \vdots & & \ddots & \vdots \\ (\phi_k[K] + \phi_l[1])/2 & (\phi_k[K] + \phi_l[2])/2 & \cdots & (\phi_k[K] + \phi_l[L])/2 \end{bmatrix}. \tag{5.51}$$

Figure 5.14 shows the absolute value of the average correlation between the two messages after initialization. Based on the previous conclusions, a high correlation is expected for scenarios where data detection is not able to provide reliable estimates within the first few iterations. In order to see the impact of correlated combining the WINNER C2 NLOS channel model is applied with a velocity of $v = 60$ km/h at an SNR of 6 dB. The LTE pilot grid, shown in Figure 5.24, is used with $L = 144$ OFDM subcarriers and $K = 14$ OFDM symbols.

The shape of the correlation exhibits an apparent symmetry with parallel waves along the OFDM subcarriers. A minimum can be identified in the middle of the burst with 6 peaks to both sides of it. The peaks are roughly aligned with the applied pilot grid. These observations are explained by the fact that in the absence of reliable data estimates the information of pilot symbols contribute significantly to the overall message exchange and thus, to the correlation of messages. In Figure 5.15 and Figure 5.16, the average correlation after 10 iterations is shown for conventional combining according to (3.51) and correlated combining given by (3.54), respectively. As can be seen, the correlation is successfully reduced, i.e. the maximum value of the peaks is at the previous minimum and the ripples between the peaks are reduced in height as well. However, the characteristic shape of the correlation remains. As explained previously, the correlation depends on several parameters, such as the strength of fading in time, frequency, and/or space, pilot grid, modulation order, etc.. The intention of the previous figures is to visualize the correlation between messages, to show that correlation is not negligible, but can be taken care of by an appropriate message combining.

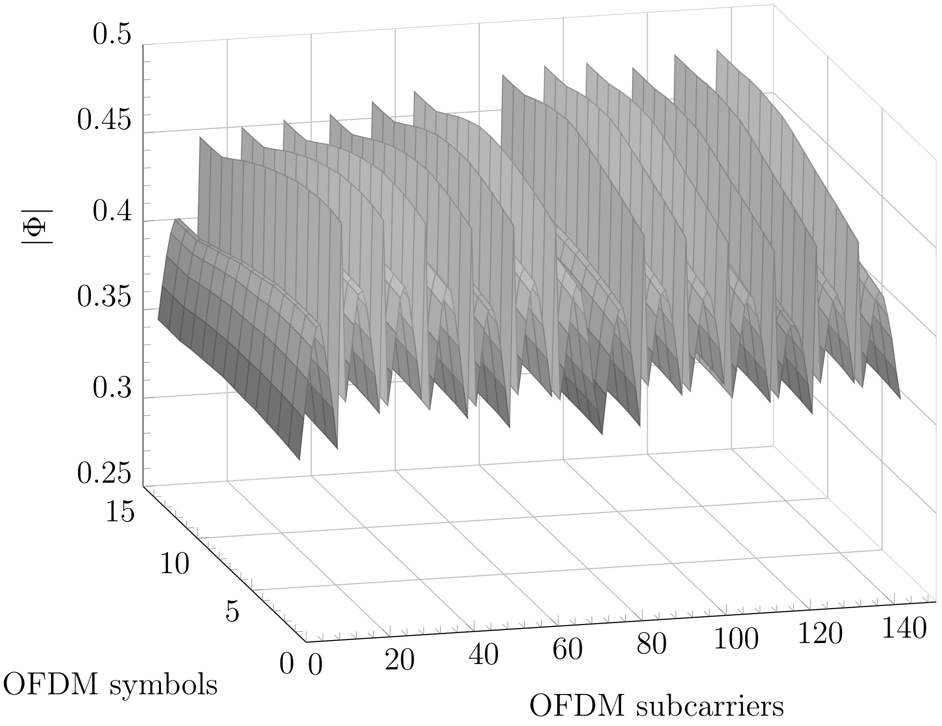


Figure 5.14: Averaged correlation between messages of coefficients after initialization.

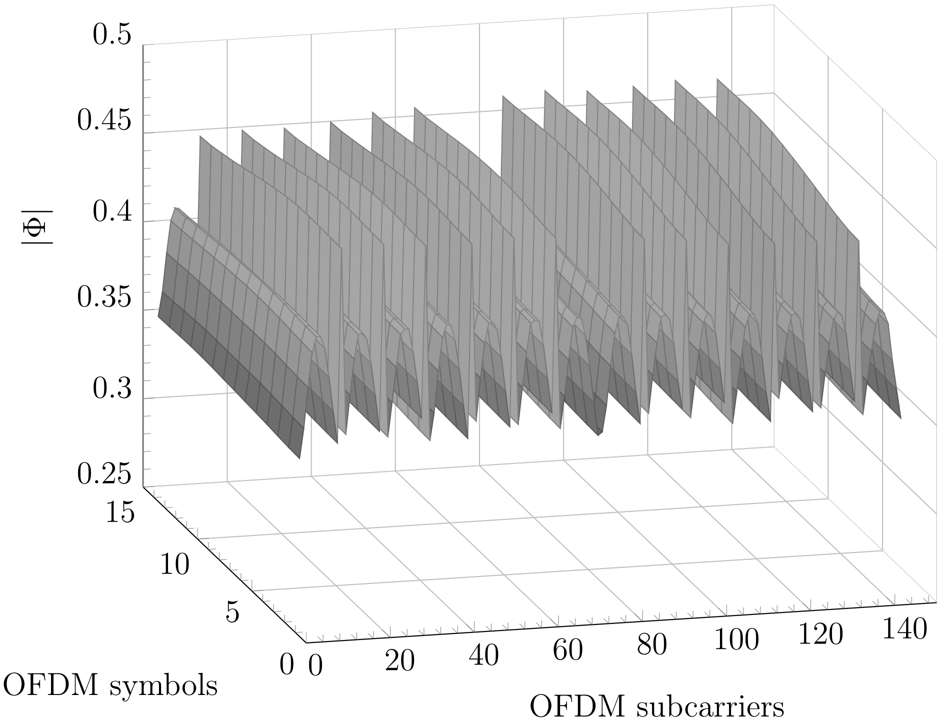


Figure 5.15: Averaged correlation between messages of coefficients after 10 iterations with conventional combining.

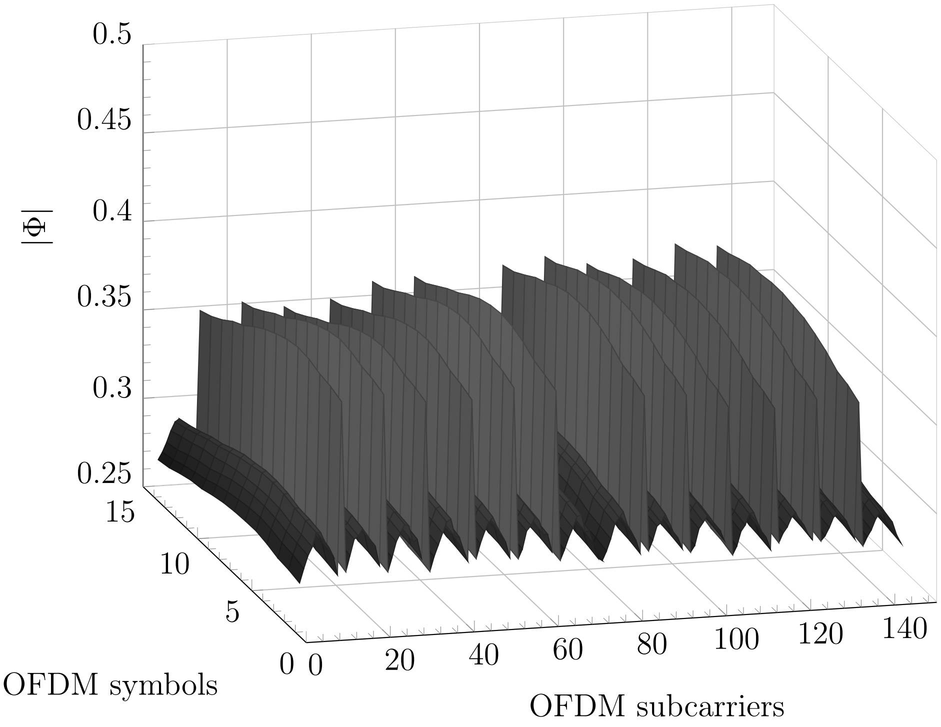
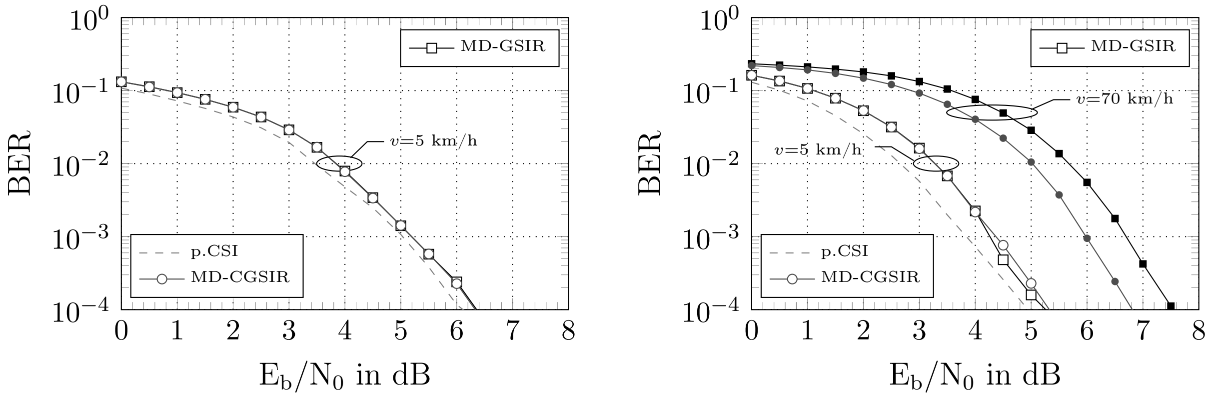
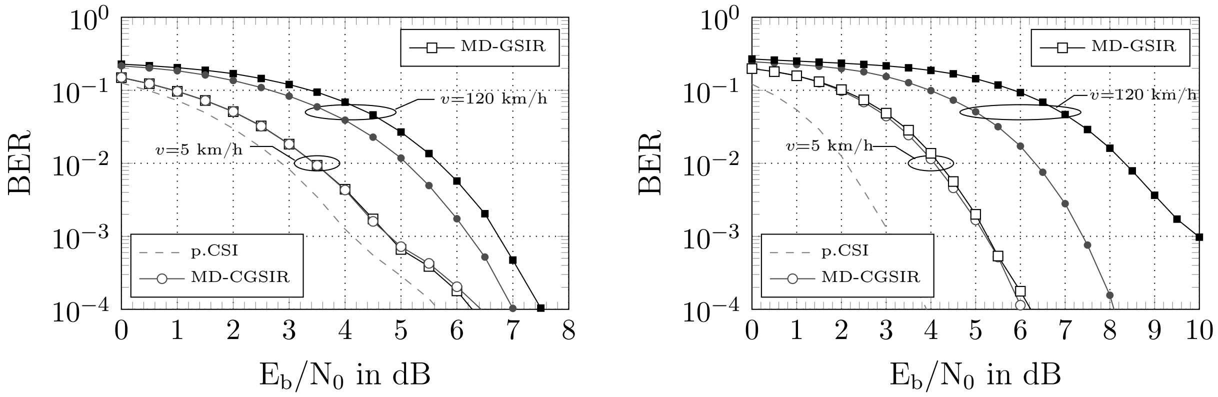


Figure 5.16: Averaged correlation between messages of coefficients after 10 iterations with correlated combining.



(a) WINNER A1 NLOS

(b) WINNER B1 NLOS



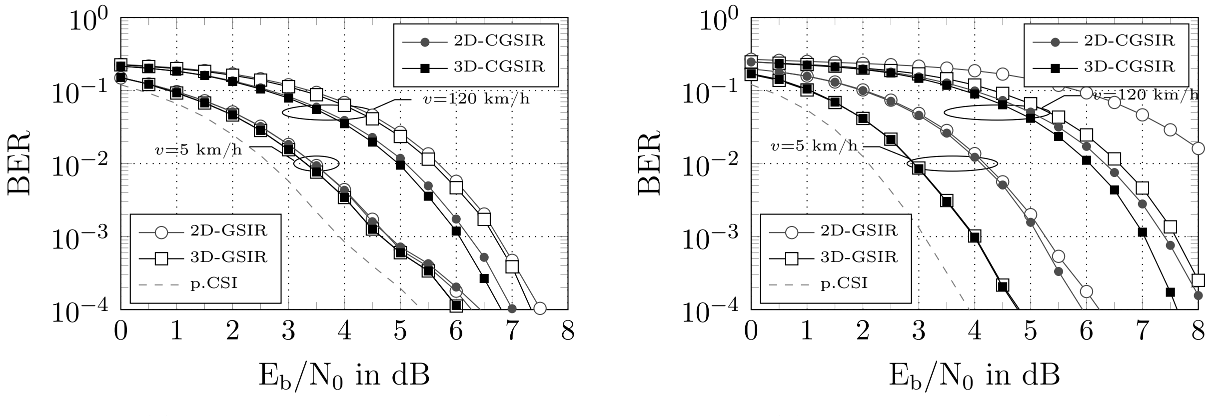
(c) WINNER C1 NLOS

(d) WINNER C2 NLOS

Figure 5.17: BER performance gain due to correlated combining for different channel models as well as varying velocities.

In order to highlight the achievable performance gain due to correlated combining, Figure 5.17 provides BER results of the MD-GSIR with correlated combining (MD-CGSIR) and the four considered WINNER channel models with varying velocities. A total of $L = 300$ OFDM subcarriers and $K = 14$ OFDM symbols are used. Pilot symbols are multiplexed in time and frequency according to the LTE pilot grid. It is expected that the gain due to correlated combining is most pronounced for channels with a strong fading in the frequency domain and a high velocity. Motivated by the results of the previous section, the (sl-ls) schedule is applied throughout all remaining simulations. For comparison, the MD-GSIR with conventional combining and with perfect channel state information (p.CSI) are included as well. As expected, the gain due to the correlated combining is highest for a channel with a strong fading in time and/or frequency domain. Almost no difference between MD-GSIR and MD-CGSIR is observed for a velocity of $v = 5$ km/h and all four considered WINNER channel models. A gain of about 0.5 dB is achieved for the WINNER B1 NLOS and the WINNER C1 NLOS at a velocity of $v = 70$ km/h and $v = 120$ km/h, respectively. The difference in velocity has no apparent influence on the gain of the MD-CGSIR. Moreover, the fading in frequency domain is similar which explains the almost equal gain. The strong influence of the frequency domain is reasoned in the burst structure, where the amount of OFDM subcarriers is significantly larger compared to the number of OFDM symbols. With a different burst structure a different behavior is expected. The loss due to channel estimation is also similar for the B1 and C1 channel model with about 1.6 dB. The MD-CGSIR yields a gain of 2.5 dB compared to the conventional MD-GSIR for the C2 NLOS channel and a velocity of $v = 120$ km/h. Unfortunately, the loss due to channel estimation is still the largest compared to the other scenarios, with a gap of about 2 and 4 dB to the perfect CSI curve.

As mentioned in the previous section, the spatial domain has been excluded from the message exchange because of the limited number of transmit antennas as well as the relatively lower correlation between them. Therefore, it does not surprise that 3D channel estimation does not improve the performance in general. In fact, a gain is only observed for the C1 and C2 NLOS channel, as can be seen in Figure 5.18. Due to the higher angular spread of the A1 and B1 NLOS channel, no improvement is observed. Therefore, the corresponding results are omitted. At low velocities, no gain is achieved for the C1 NLOS channel by using 3D channel estimation. Furthermore, only a small gain of about 0.15 dB is obtained for a velocity of $v = 120$ km/h with the MD-GSIR with both conventional as well as correlated combining. A surprising performance improvement is observed for the C2 NLOS channel. At low velocities, the 3D-GSIR is able to outperform 2D-GSIR as well as the 2D-CGSIR and achieves a similar performance as the 3D-CGSIR. Such a strong performance improvement is hardly explainable by the relatively weakly correlated spatial dimension. More insight is given by the variance of the messages and correlated combining. It has been identified as a problem that when the message exchange mainly relies on pilot information; the resulting variance of the distributed messages reduces to a very low value. Accordingly, messages are treated overly confident. Updated messages from data symbols are mostly ignored due to their larger variance. Exactly this situation is mitigated by means of 3D channel estimation. It is assumed that three messages have to be combined, one for each domain. Typically, the variances of the time and frequency



(a) WINNER C1 NLOS

(b) WINNER C2 NLOS

Figure 5.18: BER results with the 2D-GSIR and the 3D-GSIR with and without correlated combining.

domain are significantly smaller compared to the spatial domain. The combination of the three variance values yields a variance which is smaller than the combined value of the time and frequency domain. Hence, messages obtained by data symbols, which combine with the messages sent via spatial transfer nodes will have a lower overall variance. Accordingly, these messages are more likely to be considered in the following message combining process and the impairing effect of correlated messages is mitigated. Consequently, the 3D-CGSIR is not able to provide a substantial gain. With increasing velocity, correlated combining yields a gain of 0.5 dB and about 0.4 dB compared to the 3D-GSIR and the 2D-CGSIR, respectively. Still, the gain from the 3D-GSIR to the 2D-GSIR is about 1.5 dB.

5.4 Convergence, Initialization, and Robustness

So far, the performance of the MD-GSIR has been significantly improved by applying a suitable scheduling and by taking the correlation of messages during the combining step into account. For low velocities, the performance is close to the optimum with the exception of the C2 NLOS channel (cf. Figure 5.17), however, for high velocities the loss due to channel estimation is non-negligible independent of the channel scenario.

By providing improved a priori information, the MD-GSIR is able to improve its BER and MSE performance as well. For initialization, two options exists, either the channel state information for the complete burst is provided by means of interpolation or the a priori information only at pilot positions is improved. Improved a priori information hereby refers to the MSE performance compared to a symbol-wise LS channel estimation which has been used in the previous simulations. In order to maintain a low complexity solution, only the channel state information at pilot positions is refined. In addition to the symbol-wise LS initialization, two alternative methods are applied in the following, namely the MOPSO method and a Wiener filter and, presented in Section 4.1.4 and Section 6.1, respectively. The number of internal MOPSO iterations is restricted to 10 to reduce the complexity for the MOPSO approach. Motivated by the results of Section 4.2.2, the chosen maximum number of iterations yields a sufficiently accurate result in combination with the MD-GSIR. The common approach of LS and MOPSO-based initialization is to provide a priori information only at pilot positions, i.e. they are not used for interpolation. The Wiener filter on the other hand, is able to interpolate between two pilot positions. Since the pilot symbols of two transmit antennas are multiplexed in time and frequency, a muted pilot position is between two actual pilot positions of a transmit antenna (cf. Figure 5.24a). With the LS and MOPSO these positions remain silent since no a priori information is available. On the contrary, the Wiener filter is able to provide a priori information also on these muted positions by means of interpolation.

To provide a better insight on the effect of an improved initialization, EXIT charts are introduced in the following section. The obtained results are supported by additional Monte Carlo simulations, which provide BER results. Moreover, the pilot grid which is used for LTE and LTE-A is rather dense. A favorable property of a semi-blind iterative receiver is that it supports sparse pilot grids. It is analyzed in Chapter 6 to which extent pilot symbols can be separated such that a received sequence can still be reconstructed without errors. Here, the pilot density is reduced under the premise that the trade-off between performance and bandwidth efficiency is balanced. The achievable performance as a function of pilot grid and initialization is investigated in Section 5.4.2

5.4.1 EXIT Chart Analysis

The analysis of an iterative semi-blind receiver is typically very time-consuming due to the vast amount of possible parameter settings. Here, extrinsic information transfer (EXIT) charts are particularly useful, since a channel code can be evaluated separately from the iterative MIMO receiver. By monitoring the exchange of extrinsic information between the channel code and the MIMO receiver, the convergence behavior as a function of channel code and code rate can be predicted. EXIT charts have been initially developed for the convergence analysis of parallel concatenated channel codes in [tB01]. They also have been successfully applied for the convergence analysis of iterative receivers in [tBSS00, SSAR07, HRRE09]. Hereby, extrinsic information is exchanged between the outer channel code and the inner demapper. The demapper uses the received vector $\boldsymbol{y}$ and the a priori information $I_{\mathrm{A,DEM}}$ to estimate the channel and to generate the extrinsic information $I_{\mathrm{E,DEM}}$ of the transmitted vector $\boldsymbol{x}$. The extrinsic information is interleaved and becomes the a priori information of the outer channel code $I_{\mathrm{A,DEC}}$. Subsequently, extrinsic information of the outer coded bits $I_{\mathrm{E,DEC}}$ are calculated by the channel decoder. After interleaving, the extrinsic information is fed back to the MIMO demapper to serve as refined a priori information. The exchange of a priori and extrinsic information is shown in Figure 5.19. In order to simplify the computational complexity and to facilitate a tractable model, two main assumptions are done for EXIT charts. The first is the assumption of a sufficiently long sequence length or interleaver size and the second is that the a priori LLR values $\mathrm{L_A}$ can be modeled as i.i.d. Gaussian random variables [tB00a, Hoe13], i.e.:

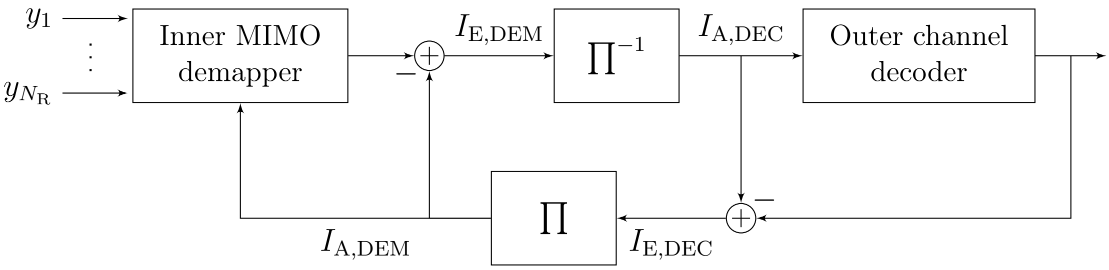


Figure 5.19: Exchange of a priori and extrinsic information between outer code and inner MIMO demapper.

$$\mathrm{L_A} = \mu_{\mathrm{L_A}} \cdot u + n_{\mathrm{A}}, \tag{5.52}$$

whereas σ_A^2 refers to the variance of the noise process n_{A} and $u \in \{+1, -1\}$ corresponds to the uncoded info bits. Since the log-likelihood ratio $\mathrm{L_A}$ is assumed to be based on a Gaussian distribution, the mean value must fulfill [Hoe13]

$$\mu_{\mathrm{L_A}} = \frac{\sigma_A^2}{2}. \tag{5.53}$$

Thus, the conditional pdf is given by

$$p\left(\mathrm{L}_\mathrm{A}|U=u\right) = \frac{1}{\sqrt{2\pi\sigma_A^2}} \cdot \exp\left(-\frac{\left(\mathrm{L}_\mathrm{A} - \frac{\sigma_A^2}{2}\cdot u\right)^2}{2\sigma_A^2}\right). \tag{5.54}$$

With the conditional pdf $p(\mathrm{L}_\mathrm{A}|U=u)$ given by (5.54) and equally probable information bits, the a priori mutual information for a discrete memoryless channel is as follows [Hoe13]:

$$I_\mathrm{A} = \frac{1}{2} \cdot \sum_{u=\{+1,-1\}} \int_{-\infty}^{\infty} p\left(\mathrm{L}_\mathrm{A}|U=u\right) \log_2 \frac{2p(\mathrm{L}_\mathrm{A}|U=u)}{p(\mathrm{L}_\mathrm{A}|U=+1)+p(\mathrm{L}_\mathrm{A}|U=-1)} d\mathrm{L}_\mathrm{A}. \tag{5.55}$$

Inserting (5.54) into (5.55) yields

$$\begin{aligned} I_\mathrm{A} &= 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{(\mathrm{L}_\mathrm{A} - (\sigma_A^2/2))^2}{2\sigma_A^2}\right) \log_2\left(1+\exp(-\mathrm{L}_\mathrm{A})\right) d\mathrm{L}_\mathrm{A} \\ &\doteq I_\mathrm{A}(\sigma_A). \end{aligned} \tag{5.56}$$

As can be seen from (5.56), the mutual information I_A is a monotonically increasing function, which depends only on the standard deviation σ_A. In order to simplify the notation, a function $J(\sigma) \doteq I_\mathrm{A}(\sigma_A = \sigma)$ is introduced [tBKA04]. Since $J(\sigma)$ is monotonically increasing, it can be inverted

$$\sigma_A = J^{-1}\left(I_\mathrm{A}\right). \tag{5.57}$$

Unfortunately, the function $J(\sigma)$ cannot be computed in closed form but can be well approximated in sections by a polynomial and an exponential fit, as described in [tBKA04]:

$$J(\sigma) \approx \begin{cases} -0.0421061\sigma^3 + 0.209252\sigma^2 - 0.00640081\sigma & \text{for } 0 \leq \sigma \leq 1.6363 \\ 1 - e^{0.00181491\sigma^3 - 0.142675\sigma^2 - 0.0822054\sigma + 0.0549608} & \text{for } 1.6363 < \sigma < 10 \\ 1 & \text{for } \sigma \geq 10 \end{cases}$$

$$J^{-1}(I_\mathrm{A}) \approx \begin{cases} 1.09542 I_\mathrm{A}^2 + 0.214217 I_\mathrm{A} + 2.33727\sqrt{IA} & \text{for } 0 \leq I_\mathrm{A} \leq 0.3646 \\ -0.706692\ln(0.386013(1-I_\mathrm{A})) + 1.75017 I_\mathrm{A} & \text{for } 0.3646 < I_\mathrm{A} < 1. \end{cases}$$

In order to generate a transfer characteristic of the outer channel code and/or inner MIMO demapper, the standard deviation σ_A is computed according to (5.57) for chosen values of $I_\mathrm{A} \in [0,1]$. For each value of σ_A, an a priori information is generated, which is used at the channel decoder and/or MIMO demapper. Subsequently, extrinsic LLR values L_E are generated, which are used to calculate the extrinsic mutual information

$$I_\mathrm{E} = \frac{1}{2} \cdot \sum_{x=\{+1,-1\}} \int_{-\infty}^{\infty} p\left(\mathrm{L}_\mathrm{E}|x\right) \log_2 \frac{2p(\mathrm{L}_\mathrm{E}|x)}{p(\mathrm{L}_\mathrm{E}|+1)+p(\mathrm{L}_\mathrm{E}|-1)} d\mathrm{L}_\mathrm{E}. \tag{5.58}$$

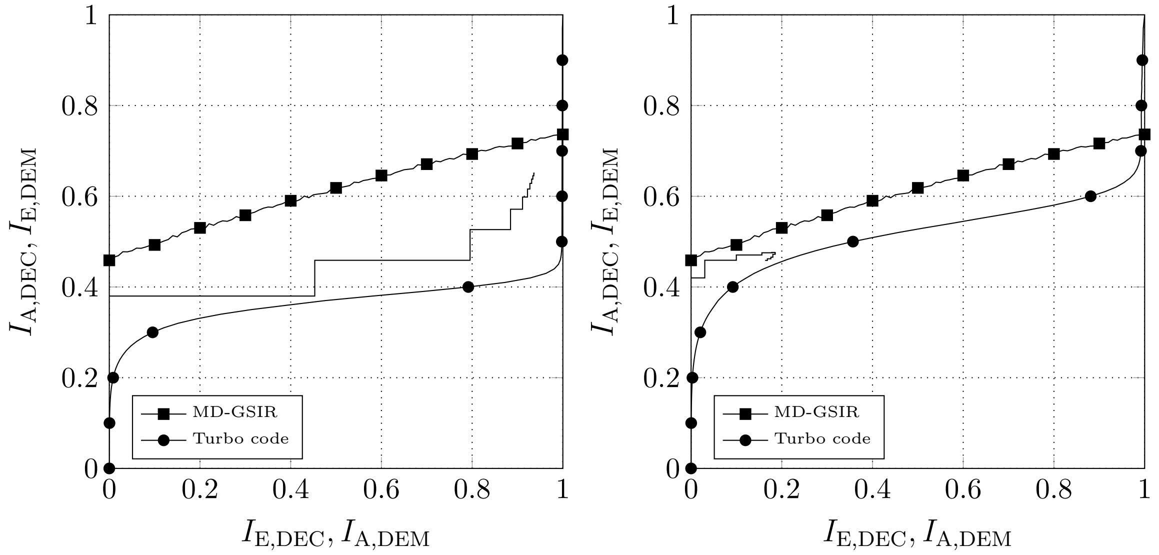
Typically, a histogram is used to obtain the pdf of $p(\mathrm{L_E}|x)$. Hereby, Monte Carlo simulations are conducted for each value σ_A for both inner and outer decoder [tB01]. A sufficiently long sequence length is required to obtain a good resolution of the pdf. This method has the advantage that it is generally applicable and does not assume any prerequisites on the distribution of $\mathrm{L_E}$ or the use of certain soft-input soft-output decoders [Hoe13]. Alternatively, several variants exist that compute (5.58) computationally more efficient [LHG04, Hag04].

Ideally, the convergence behavior of a MIMO demapper can be predicted for arbitrary channel codes and code rates, given its transfer characteristic. Not only the required number of iterations can be obtained but also the bit error probability for a given set of a priori and extrinsic mutual information [Hoe13]:

$$P_b = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{(J^{-1}(I_{\mathrm{E}}))^2 + (J^{-1}(I_{\mathrm{A}}))^2 + 8RE_b/N_0}{8}}\right). \tag{5.59}$$

The convergence of a coded MIMO demapper to a low error probability is enabled when the transfer characteristics of the demapper and the channel code do not intersect. The trajectory describes the behavior during iterations and thus the exchange of mutual information between demapper and decoder. Once the trajectory is stuck, further iterations cannot improve the performance. Moreover, large steps refer to a significantly improved output and indicate the number of iterations which have the most impact w.r.t. BER performance.

Exemplary for the WINNER C2 NLOS channel, a velocity of $v = 120$ km/h, and an SNR of 4 dB, the transfer characteristic of the LS-initialized MD-GSIR is depicted and compared to the transfer characteristic of a rate 1/3 and rate 1/2 turbo code in Figure 5.20. The trajectory is obtained by means of Monte Carlo simulations and gives insight about the accuracy of the EXIT chart simulations. An apparent difference to typical EXIT chart curves is the relatively wide tunnel, i.e. the spacing between the transfer characteristic of the demapper and the decoder. Typically it is sufficient that the two curves do not intersect in order to allow convergence. Here, the tunnel has to be wide open such that the receiver is able to converge to a low bit error probability. Moreover, a second difference is the inaccuracy of the trajectory, which should actually touch the transfer characteristics of the demapper. Here, they deviate up to a value of 0.2. Moreover, the transfer characteristics for the rate 1/2 turbo code indicate that a convergence is possible. However, the trajectory get stuck at an early point. The reasons for these discrepancies are two-fold. One reason is given by the limited sequence length and inaccuracies introduced by channel estimation errors. The second reason is more fundamental and is reasoned in the probabilistic behavior of the factor graph. The aim of EXIT charts is to analyze the behavior during iterations. Actually, no iterations are performed during EXIT chart simulations. Instead, the behavior is predicted based on the accuracy of a priori information. However, the performance as well as the behavior of the MD-GSIR changes significantly with iterations, .e.g. the previously described effect of correlated random measures only arises during iterations. Additionally, the performance of the message schedules change after the initial iteration as indicated by the results of the (sl) and (sl-ls) results given in Section 5.2. Accordingly, the results of the EXIT charts can



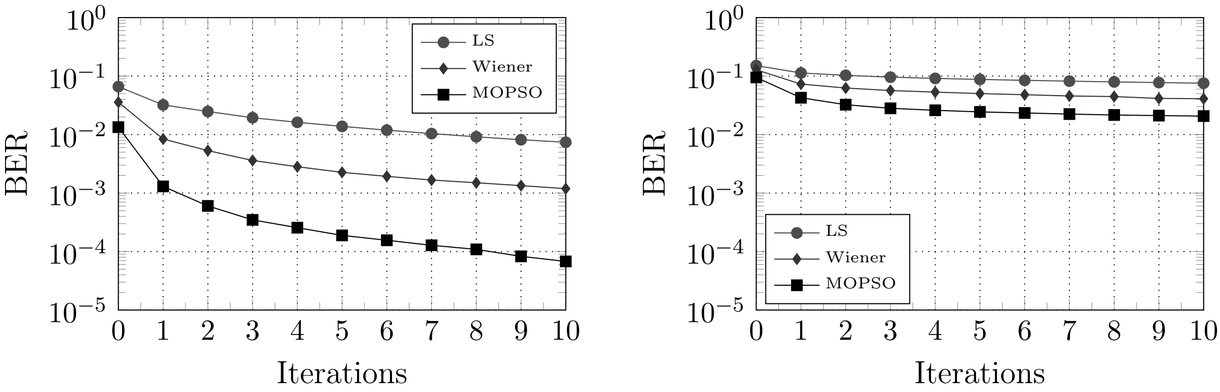
(a) Turbo code rate 1/3 (b) Turbo code rate 1/2

Figure 5.20: Trajectory of MD-GSIR with the WINNER C2 NLOS channel at an SNR of 4 dB.

be interpreted as that of an optimum MD-GSIR where the effects of cycles are mitigated. Up to now, no analysis tool exists, which can accurately predict the performance of an iterative semi-blind receiver. Keeping in mind the potential gains (cf. Chapter 6), accurate analysis tools are extremely helpful and an interesting future research topic.

Nevertheless, insights on the convergence behavior of the MD-GSIR can still be obtained. The results for the rate 1/3 turbo code indicate that within four to five iterations the achievable performance is reached. Further iterations do not improve the performance significantly. By carefully examining the results in Figure 5.21, the difference between LS and Wiener/MOPSO-based initialization becomes apparent. The advantages of an improved initialization are two-fold. Besides a direct improvement of the BER results after initialization, the gain after the first iteration is also larger. The MOPSO surprisingly performs better than the Wiener-initialized MD-GSIR. This observation is examined in more detail in the following.

All simulation results are conducted with five global iterations, which comprise one iteration for the MD-GSIR and one iteration for the turbo code. It can be seen that the trajectory of the MD-GSIR obtained by means of Monte Carlo simulations does give accurate results. The final point of the trajectory with code rate of $R = 1/2$, shown in Figure 5.20, is at ($I_{\mathrm{A,DEM}} = 0.1632, I_{\mathrm{E,DEM}} = 0.4586$). The calculated bit error probability based on (5.59) yields $P_b = 0.1027$, which is close the BER of 0.0888, measured with the Monte Carlo simulation. The final point of ($I_{\mathrm{A,DEM}} = 0.9376, I_{\mathrm{E,DEM}} = 0.6559$) for the rate $R = 1/3$ results in a bit error probability of $P_b = 5.03 \cdot 10^{-3}$, whereas the Monte Carlo simulations calculate a BER of $2.13 \cdot 10^{-2}$. The deviation between the two results is reasoned in (5.59) itself, which is only accurate for moderate bit error probabilities close to the waterfall region [Hoe13]. The EXIT chart results for the four considered



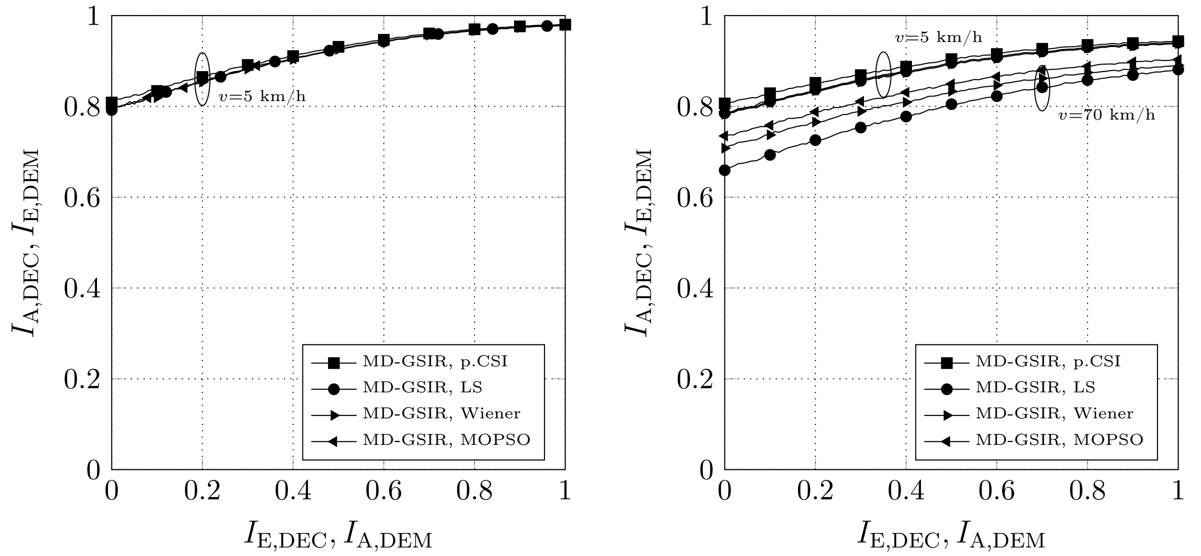
(a) Turbo code rate R=1/3 (b) Turbo code rate R=1/2

Figure 5.21: BER performance as a function of the number of iterations with the WINNER C2 NLOS channel at an SNR of 4 dB.

WINNER channel models and the MD-GSIR with varying initialization methods, namely LS, Wiener, and MOPSO, is given in Figure 5.22. Additionally, the transfer characteristics of the APP detector with perfect channel state information is given for comparison as well. In order to see a clear difference between the curves, the SNR is set to 6 dB. As before, a low velocity of $v = 5$ km/h and the respective maximum defined velocity is shown for each channel scenario.

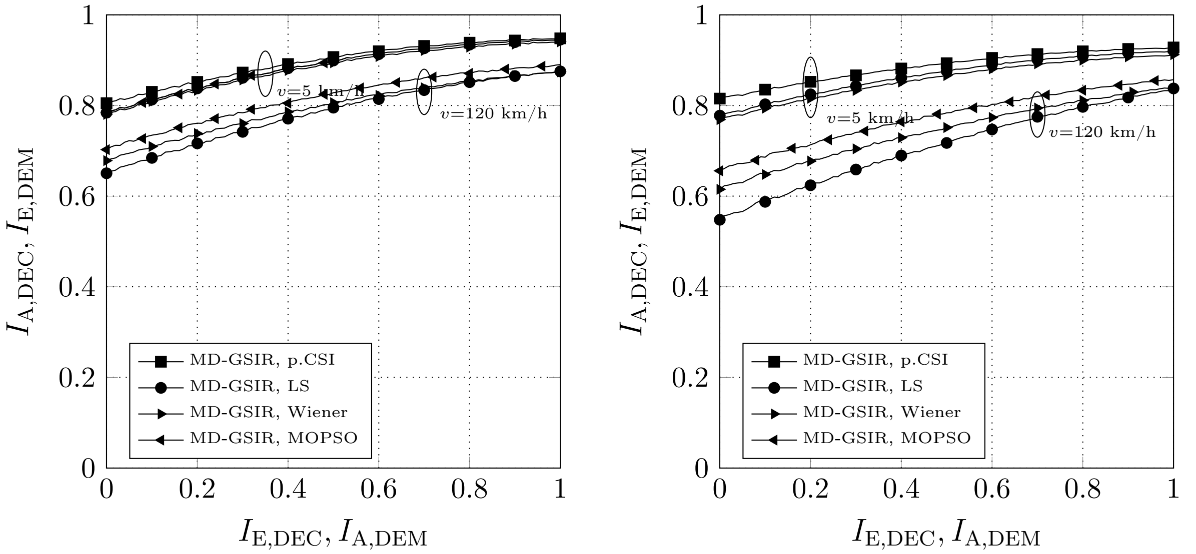
As expected, there is only a minor difference between all methods for the WINNER A1 NLOS channel. The curves start at relatively high position and end close to the (1,1) point, which indicates a bit error ratio of close to zero. With increasing diversity in time and/or frequency, the curves become separated, whereas the MOPSO-initialized and the Wiener-initialized MD-GSIR provide the results closest to the optimum p.CSI curve. The difference between MOPSO, Wiener and LS is most pronounced for the high velocities. Interestingly, all transfer characteristics do not have a constant slope over the a priori information $I_{\mathrm{A,DEM}}$ and become nearly flat for larger a priori values. On the one hand that means that improved a priori information can not improve the output of the demapper, on the other, the optimum result is achieved in less iterations. By comparing the figures, it can be seen that the gain due to an improved initialization is highest for an environment which is subject to rich scattering. At high velocities and/or a strong fading in the frequency domain, data symbols are not predicted accurately within the first few iterations. In these situations, pilot information dominate the message exchange. Improved a priori information directly translates into a gain w.r.t. BER. Although the Wiener filter yields more accurate results in terms of MSE (cf. Section 4.2), the Wiener-initialized MD-GSIR does not achieve the highest mutual information. Apparently, the MSE is not a sufficient metric to adequately describe the quality of a priori information. The strong influence of pilot information has been identified as a major component of correlated messages in Section 5.3. Providing reliable information at even more positions than LS/MOPSO seems to have a negative effect on the overall achievable performance.

In order to verify the results of the EXIT chart simulations, corresponding bit error results are given in Figure 5.23. Based on the previous results, the MD-GSIR with correlated combining is used with different initialization methods. It can be seen that the MOPSO-initialized MD-GSIR achieves the best performance independent of velocity and/or channel scenario. This impressive performance is somehow surprising, when compared to the MSE results presented in Section 4.2.2. There, MOPSO did achieve an improvement w.r.t. to LS channel estimation, however, the Wiener filter still outperformed the MOPSO algorithm. Nevertheless, the mean squared error is obviously not a



(a) WINNER A1 NLOS

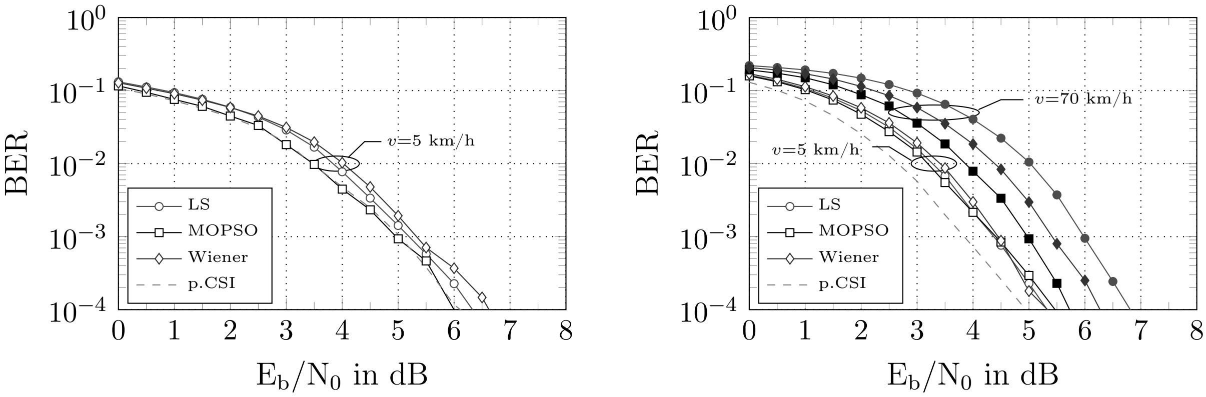
(b) WINNER B1 NLOS



(c) WINNER C1 NLOS

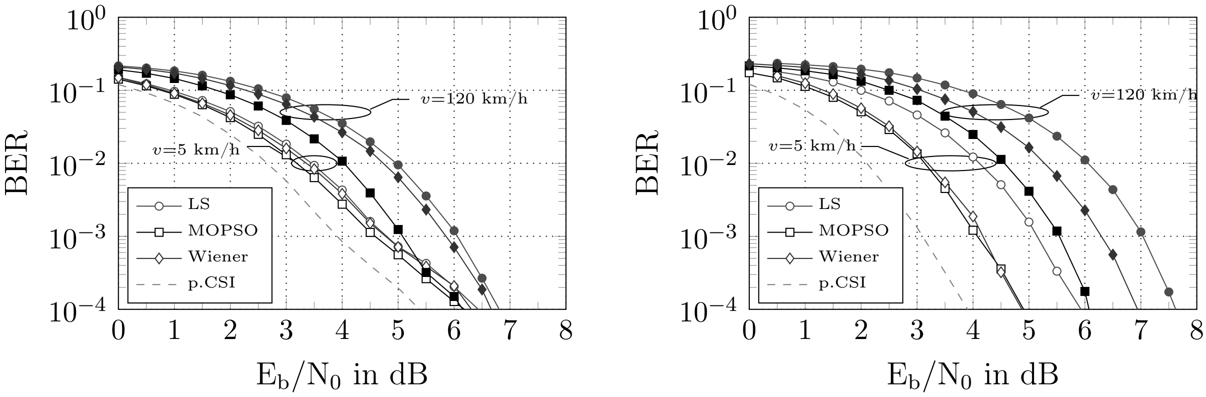
(d) WINNER C2 NLOS

Figure 5.22: EXIT charts of MD-GSIR with different initializations and varying channel models at an SNR of 6 dB.



(a) WINNER A1 NLOS

(b) WINNER B1 NLOS



(c) WINNER C1 NLOS

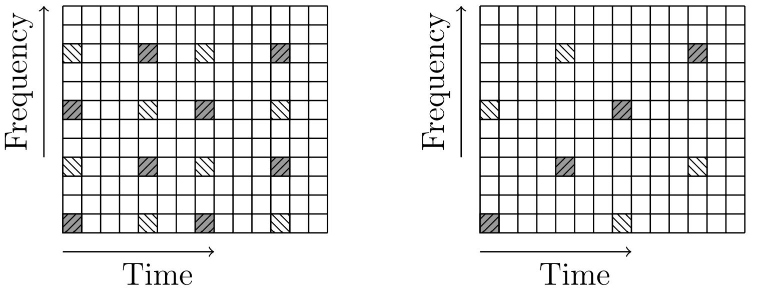
(d) WINNER C2 NLOS

Figure 5.23: BER performance of MD-GSIR with different initialization methods and WINNER channel models.

sufficient metric to characterize the quality of an initialization algorithm. Moreover, the statistical properties of the estimates are completely ignored. Since the Wiener filter uses the autocorrelation function in order to smooth and/or interpolate the hypotheses, the resulting estimates themselves follow a certain correlation. Concluding from the previous results, it can be said that a situation in which the information of pilot symbols dominate the message exchange should be avoided. This situation is more likely when very precise a priori information is provided and/or data symbols are not reliably detected. Obviously, it is not advisable to provide poor a priori information. Not the quality, i.e. the mean value, of the estimates is the origin of the poor performance but rather its strong influence, which is determined by its variance value. Hence, by artificially increasing the variance of the pilot information as a function of iteration, the overall performance is increased. Nevertheless, the performance of the MOPSO-initialization is not reached for the considered setups. For the A1 NLOS, the BER performance of the MOPSO-initialized MD-GSIR is almost identical to the performance of the APP with perfect CSI. The loss due to channel estimation with the B1 NLOS at a velocity of v = 70 km/h is reduced by 1 dB and only 0.7 dB away from perfect CSI. Similarly for the C1 NLOS channel, where the improvement by means of initialization as well as the distance to perfect CSI is about 0.7 dB. As before, the performance with the C2 NLOS is again significantly improved. For a velocity of $v = 5$ km/h the loss due to channel estimation is now 1 dB, whereas for a velocity of $v = 120$ km/h the loss is about 2.0 dB at a BER of 10^{-4}, which corresponds to an improvement of 1.5 dB compared to the LS-initialized MD-GSIR with correlated combining and a gain of more than 4 dB compared to the conventional MD-GSIR. In order to assess the performance in relation to a state-of-the receiver, the MOPSO-initialized MD-GSIR is compared to a Wiener-filter based channel estimation with iterative APP detection (Wiener+APP) in the following section. The performance for higher-order modulation is evaluated in Section 5.5.

5.4.2 Influence of Pilot Grid

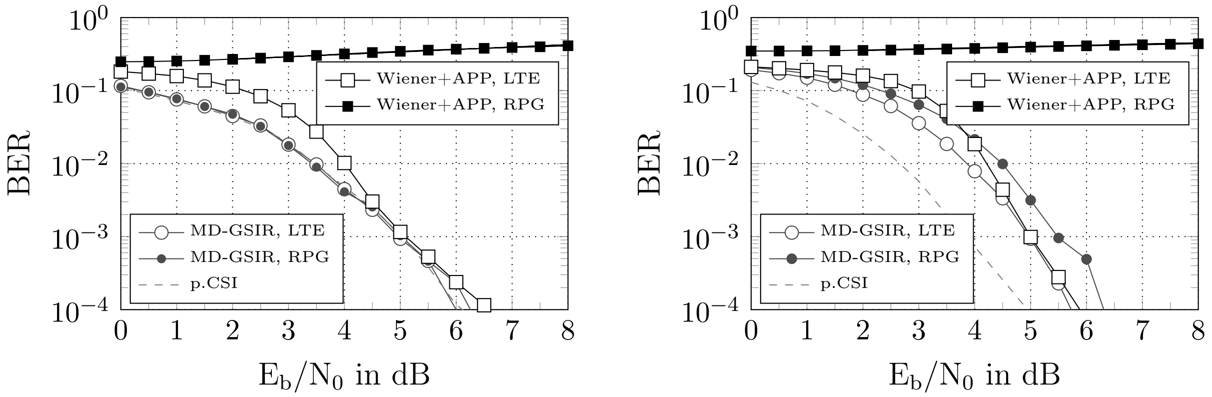
One of the main features of a semi-blind iterative receiver is that it achieves good performances with relatively few pilot symbols. The conventional pilot grid used in LTE and LTE-A for two transmit antennas is depicted in Figure 5.24a. Thereby, the pilot symbols of different transmit antennas are indicated by a different hatching as well as color. The pilot grid is designed to provide a good performance in high-mobility scenarios, hence, the pilot symbols are relatively close to each other. For comparison, the pilot grid shown in Figure 5.24b, uses only 50% of the pilot symbols. Additionally, the pilot symbols of one transmit antenna are distributed diagonally in time and frequency in order to track the variations in the frequency domain. A drawback of this pilot grid is that interpolation in the time domain is only possible with more than 14 OFDM symbols, since otherwise only one pilot is available in the time domain. It is therefore expected that the performance of the Wiener-filter is deteriorated.



(a) LTE pilot grid (b) Reduced pilot grid (RPG)

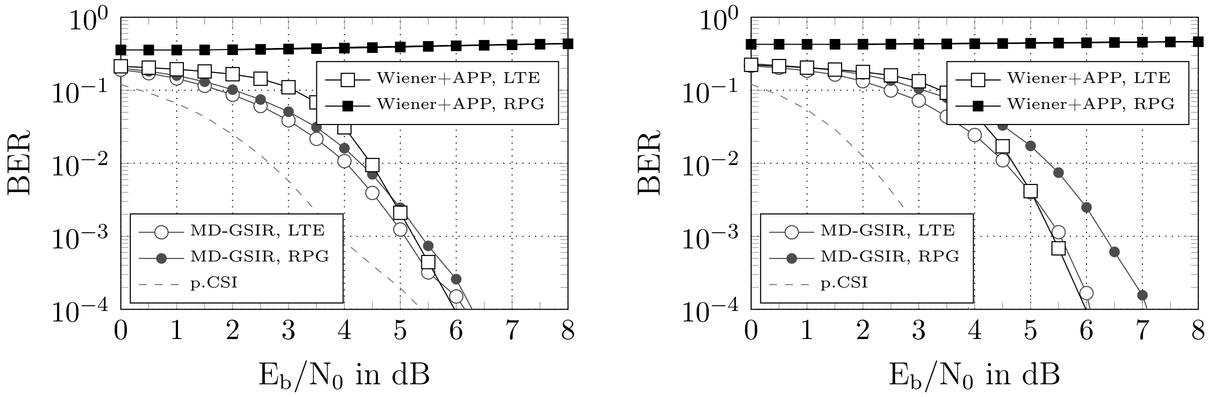
Figure 5.24: Conventional LTE-based pilot grid and proposed reduced size pilot grid (RPG).

The BER performance of the MOPSO-initialized MD-GSIR with correlated combining is compared to that of a Wiener-filter based channel estimation with iterative APP detection. An overview of the receiver structure is given in Figure 6.1. Additionally, the Wiener filter is introduced in Chapter 6.1. The above presented two pilot grids are evaluated. Only the maximum defined velocity is used for each specific channel model in order to improve readability. Most notably is that the performance of the MOPSO-initialized



(a) WINNER A1 NLOS

(b) WINNER B1 NLOS



(c) WINNER C1 NLOS

(d) WINNER C2 NLOS

Figure 5.25: BER performance of MOPSO-initialized MD-CGSIR for different WINNER channel models and two different pilot grids.

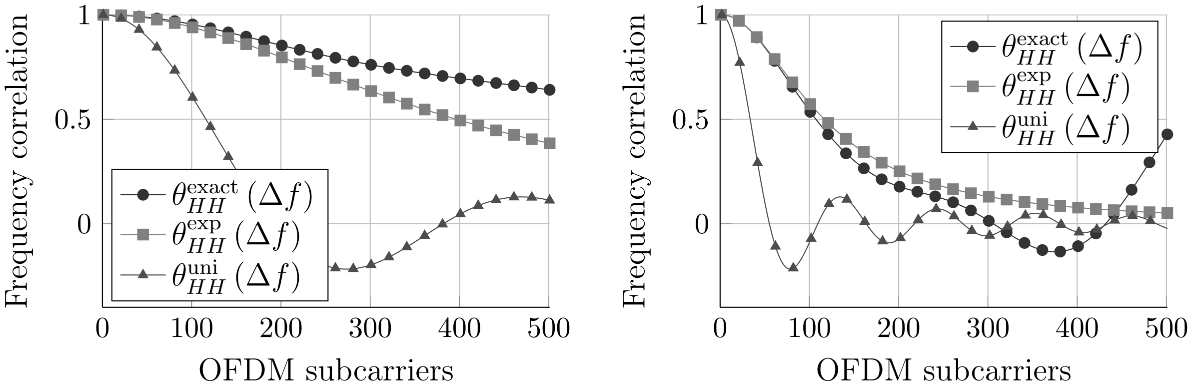
MD-GSIR is better or similar compared to the Wiener+APP receiver for all considered channel scenarios and the LTE pilot grid as can be seen in Figure 5.25. It is of importance to highlight the difference between the two receivers, which lies in the channel estimation only. Thereby, the Wiener filter has a relatively high complexity compared to the symbol-wise channel estimation within the MD-GSIR as well as the MOPSO-based initialization, which uses only 10 internal iterations. The performance of the Wiener+APP receiver with the reduced pilot grid (RPG) is catastrophic for all channel scenarios. The reason for this poor performance is mainly in the limited amount of OFDM symbols, which provides only one pilot in the time domain. The performance improves significantly with more OFDM symbols. However, the burst structure is chosen according to common LTE settings, which is $K = 14$ OFDM symbols. The MD-GSIR experiences no performance degradation for the A1 NLOS channel, a minor degradation of 0.5 dB and 0.2 dB for the B1 and C1 NLOS channel. A loss of about 1 dB is observed for the C2 NLOS channel. Again, these losses are for the maximum defined velocity and will reduce with lower velocities. Although the BER results for the majority of channel scenarios is deteriorated, a gain w.r.t. spectral efficiency is attained since the pilot overhead is reduced by a factor of two. The impact of the RPG on the achievable spectral efficiencies is investigated in Section 5.5.

5.4.3 Influence of A Priori Information

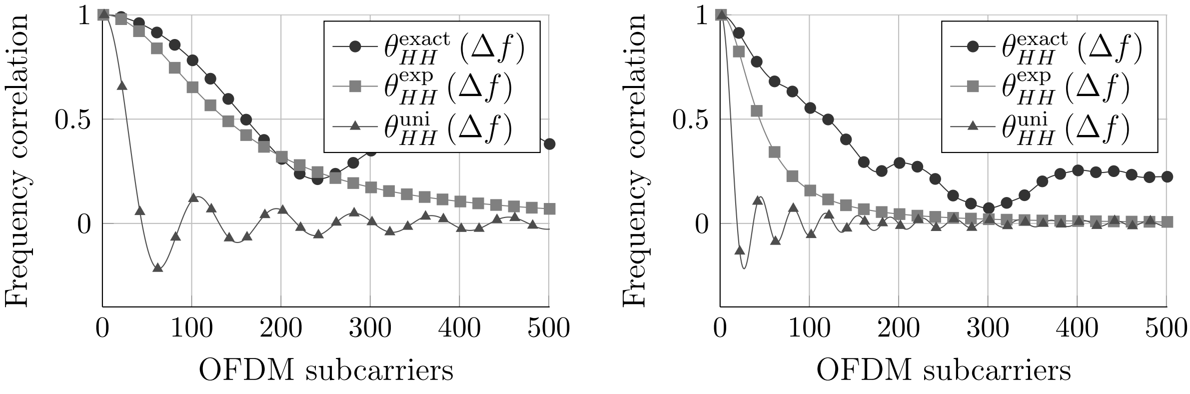
Pilot-based channel estimation by means of filtering/interpolation typically requires the knowledge of second-order statistics of the wireless channel, such as the r.m.s. delay spread, maximum delay spread, and/or Doppler frequency. Given detailed information, even the power delay profile and/or the Doppler power density spectrum can be used. The assumption of knowing the maximum values of the delay spread and Doppler frequency is realistic, since the design of the underlying system needs such maximum values itself. In the worst case scenario, these values may be used to determine the correlation functions, i.e. a uniform distribution is assumed within the given range. However, knoweldge of the exact PDP or Doppler power density spectrum is not always available, since they change dynamically and are different for each channel scenario. Unfortunately, the optimum performance of filter-based channel estimation algorithms is obtained with the exact knowledge of the channel statistics. For the assessment of a receiver under realistic conditions it is of importance to evaluate the robustness w.r.t. a priori information. Similarly to Section 4.2.2, three different assumptions are evaluated in the following: (1) the PDP is assumed to be uniformly distributed between $[0, \tau_{\max}]$, (2) the PDP is exponentially decreasing between $[0, \tau_{\max}]$, and (3) exact knowledge of the PDP is given. The corresponding correlation functions given by (4.20), (4.21), and (4.22) are recalled here:

$$\begin{aligned}
\theta_{HH}^{\text{uni}}(\Delta f) &= \operatorname{sinc}(\tau_{\max}\Delta f) \cdot \exp(-j\pi\tau_{\max}\Delta f), \\
\theta_{HH}^{\exp}(\Delta f) &= \frac{1}{1 + j2\pi\tau_{\text{rms}}F}, \\
\theta_{HH}^{\text{exact}}(\Delta f) &= \sum_{c=1}^{M_{\text{c}}} P_c \cdot \exp(-j2\pi\Delta f\tau_c).
\end{aligned}$$

The difference between the correlation functions in the time domain is more subtle. The two common correlation functions are the zeroth order Bessel function of first kind and the sinc-function. Both functions have a very similar shape for low Doppler frequencies. A comparison with the two functions is therefore omitted. The three frequency correlation functions are depicted for the four considered channel models in Figure 5.26. The frequency spacing Δf is set to multiples of the OFDM subcarrier spacing of $F_{\text{s}} = 15$ kHz. In consistency with the previous results, it can be seen that the A1 NLOS is highly correlated for more than 100 OFDM subcarriers. A high correlation is hereby defined to be equal or larger than 0.8. Hence, channel estimation is simplified and the BER performance approaches the perfect channel state information curve. The correlation function $\theta_{HH}^{\text{uni}}(\Delta f)$ decreases fast with increasing frequency spacing. The shape of the sinc function is clearly visible especially for the B1, C1, and C2 NLOS channel. Moreover, the shape of the functions $\theta_{HH}^{\exp}(\Delta f)$ and $\theta_{HH}^{\text{exact}}(\Delta f)$ are closely related up to a frequency spacing of about 200 OFDM subcarriers. Only for the C2 NLOS channel, the two functions deviate earlier.



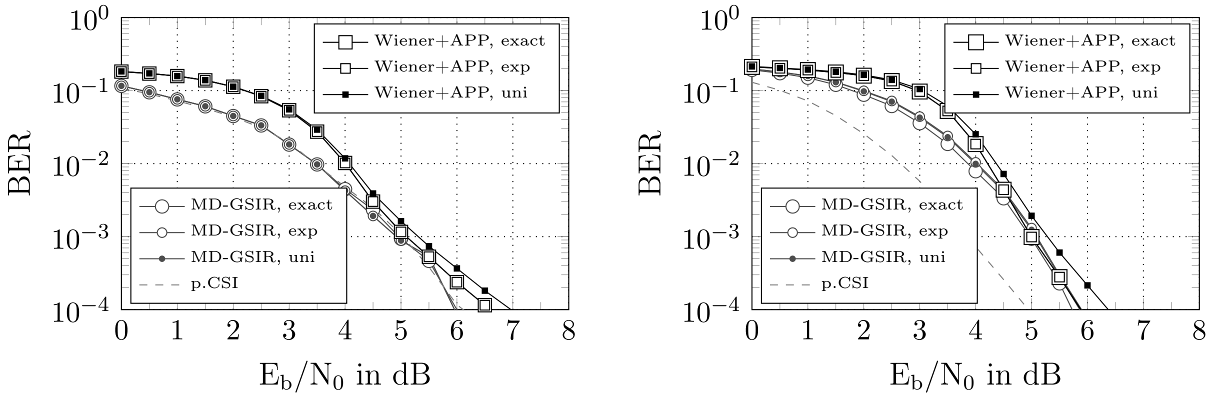
(a) WINNER A1 NLOS (b) WINNER B1 NLOS



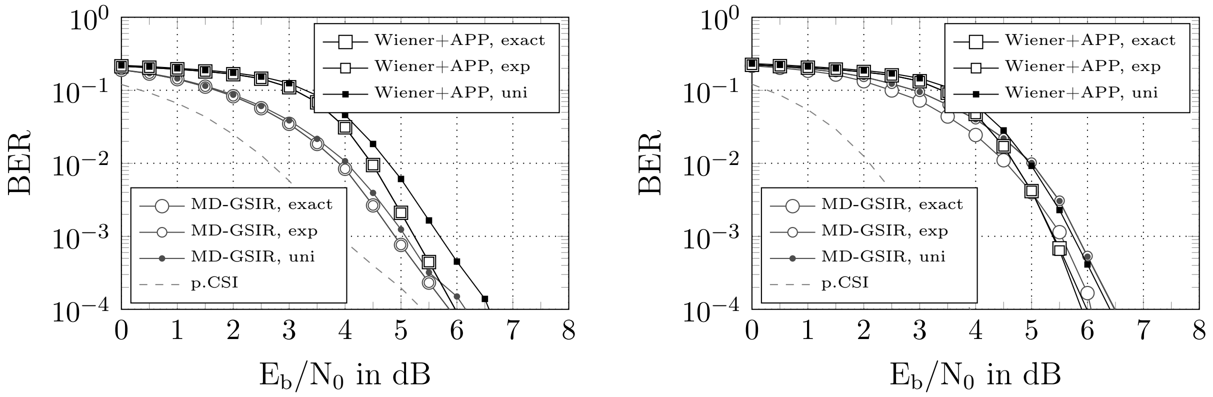
(c) WINNER C1 NLOS (d) WINNER C2 NLOS

Figure 5.26: Correlation functions as a function of a priori information of the channel statistics.

Additionally, the frequency correlation has an impact on the channel estimation performance. Based on simulations with the MD-GSIR, it can be shown that the minimum correlation between adjacent OFDM subcarriers needs to be at least 0.8 in order to yield a gain due to a message exchange. A property which has been reported independently in [OA07]. This means, if two adjacent subcarriers are correlated by a degree of 0.8 or less, the resulting variance of a message send via a transfer node is rather large. As a result, the message does not contribute to the combined message and is effectively ignored. Moreover, the assumption that the difference between two adjacent channel coefficients is zero-mean is less likely to be true. The required minimum correlation of 0.8 is fulfilled for a frequency range of 254 OFDM subcarriers for the A1 NLOS, 58 OFDM subcarriers for the B1 NLOS, 96 OFDM subcarriers for the C1, and 22 OFDM subcarriers for the C2 NLOS channel. In other words, a message of a pilot is spread for about 22 OFDM subcarriers for the C2 NLOS channel, until it has no influence on the remaining message combining process. In general, it is beneficial to consider as many pilot symbols as possible, hence, an inferior channel estimation accuracy is expected for the MD-GSIR in combination with the C2 NLOS channel. This observation is supported by the previous



(a) WINNER A1 NLOS (b) WINNER B1 NLOS



(c) WINNER C1 NLOS (d) WINNER C2 NLOS

Figure 5.27: BER performance of the MD-GSIR and Wiener+APP receiver as a function of a priori information.

simulation results where an increased loss due to channel estimation could be observed for the C2 channel compared to the remaining three channel scenarios. The BER performance of the MD-GSIR and the Wiener+APP receiver for the three different frequency correlation functions are given in Figure 5.27. The performance of the MD-GSIR for the A1, B1, and C1 NLOS channel is not affected by the choice of correlation function. A loss of 0.5 dB is observed for the C2 NLOS channel when the PDP is assumed to follow an uniform and/or exponentially decaying distribution. The Wiener+APP on the other hand experiences a performance loss for all channel scenarios with the uniform distribution of the PDP. Thus, concluding from these results, it can be said that the MD-GSIR is robust w.r.t. the correlation function. A similar or better performance is achieved compared to the Wiener+APP receiver independent of the channel scenario. The improvements introduced in the previous sections enable the MD-GSIR to achieve a high performance on a par with that of an iterative state-of-the-art receiver. Although the complexity of channel estimation is linear w.r.t. sequence length and the number of transmit antennas, the complexity of data detection is still large since an APP detector has been used. Motivated by the promising results presented in Section 3.1.3, the Gaussian tree search detection is integrated within the MD-GSIR. Their performance is evaluated for a wide range of modulation formats and code rates. The gain of the reduced pilot grid w.r.t. to spectral efficiency is also investigated.

5.5 Performance Evaluation of MD-GSIR

In most modern wireless systems, adaptive modulation and channel coding (AMC) is applied to achieve a high spectral efficiency. Hence, suitable receivers need to support a large variety of modulation and coding schemes (MCS) while maintaining a low complexity. The performance of the proposed multi-dimensional graph-based soft iterative receiver is evaluated in the following for a wide range of modulation formats and code rates. The previously presented improvements, such as an appropriate scheduling, correlated combining, and MOPSO initialization are applied for all simulation results obtained with the MD-GSIR. In order to further reduce the computational complexity of the MD-GSIR, the Gaussian tree search detection, presented in Section 3.1.3, is implemented within the framework of the MD-GSIR. For comparison, an iterative state-of-the-art receiver is included as well, namely a 2 × 1-D Wiener filter used for channel estimation and an APP MIMO detector. The same system settings, such as interleaver length, number of iterations, and assumptions about a priori information are made for the Wiener+APP receiver. One exception is that the Wiener filter is only executed once in the beginning of an iteration due to complexity reasons. The required complexity, in terms of memory consumption and runtime, would exceed the available resources by far when simulating a system similar to LTE. The important parameters for the link level simulations are listed in Table 5.1. The smallest addressable unit within an LTE system is dubbed resource element (RE) and contains a data and/or pilot symbol. The smallest assignable unit to a user is termed resource block (RB) and comprises $K_{\text{RB}} = 7$ OFDM symbols and $L_{\text{RB}} = 12$ OFDM subcarriers. Moreover, two RBs adjacent in time domain form a subframe with a duration of 1 ms. The overall bandwidth efficiency of the considered LTE system is impaired by multiple factors: (1) adjacent power leakage ratio (ACLR), (2) cyclic prefix, and (3) pilot symbols. The occupied bandwidth in the frequency domain has to be constrained such that the power leakage to adjacent channels is below a certain threshold.

Table 5.1: Parameters of the considered MIMO-OFDM system.

Parameter	Setting
Center frequency	$F = 4$ GHz
Channel bandwidth	$B_{\text{ch}} = 5$ MHz
FFT size	$\tilde{L} = 512$
Payload OFDM subcarriers	$L = 300$
OFDM subcarrier spacing	$F_{\text{s}} = 15$ kHz
Useful OFDM symbol duration	$T_{\text{s}} = 66.7\ \mu s$
Resource block duration	$T_{\text{b}} = 0.5$ ms

About 10% of the OFDM subcarriers at the band edges are used as guard carriers:

$$\eta_{\mathrm{ACLR}} = \frac{L \cdot F_{\mathrm{s}}}{B_{\mathrm{ch}}} = 0.9. \tag{5.60}$$

A cyclic prefix of length $T_{\mathrm{CP}} \approx 5.2\mu s$ for the first symbols and of length $T_{\mathrm{CP}} \approx 4.7\mu s$ for the remaining six symbols of one RB is employed to avoid intersymbol interference, which results in about 7% overhead due to the CP:

$$\eta_{\mathrm{CP}} = \frac{K \cdot T_{\mathrm{s}}}{T_{\mathrm{tot}}} = 0.9333. \tag{5.61}$$

In order to facilitate coherent detection, pilot symbols are inserted periodically in the data stream. The pilot density for $N_{\mathrm{T}} = 2$ is given by

$$\eta_{\mathrm{RS}} = 1 - \frac{8}{K_{\mathrm{RB}} \cdot L_{\mathrm{RB}}} = 0.9. \tag{5.62}$$

Hence, the overall bandwidth efficiency given the impairments described above results in

$$\eta_{\mathrm{BW}} = \eta_{\mathrm{ACLR}} \cdot \eta_{\mathrm{CP}} \cdot \eta_{\mathrm{RS}} \approx 0.756. \tag{5.63}$$

The spectral efficiency is calculated by

$$\eta = \eta_{\mathrm{BW}} \cdot N_{\mathrm{T}} \cdot R \cdot N_{\mathrm{b}} \cdot P_{\mathrm{bl}}, \tag{5.64}$$

where the probability that a codeword is transmitted successfully is defined as

$$P_{\mathrm{bl}} = 1 - \ \mathrm{BLER}. \tag{5.65}$$

Hereby, BLER refers to the block error rate. A list of the considered MCSs is given in Table 5.2, which is closely related to the applied MCSs of LTE and LTE-A. Besides the selection of code rates and modulation format is the maximal achievable spectral efficiency given for an optimal system without signaling overhead (opt.) and for a system with an overhead of η_{BW}, which is equal to the overhead of an LTE system. Ideally, a wireless system switches adaptively between the modulation and coding schemes depending on the instantaneous channel condition, i.e. a high modulation and code rate are chosen for a channel with a good reception in order to increase the throughput and vice versa, a low modulation and code rate is chosen to improve the block error rate. For this purpose, different quality measures are obtained for each resource element. A popular quality measure is given by the signal-to-interference-plus-noise ratio (SINR) [BAS$^+$05]. The individual SINR measurements are compressed to an effective SINR by means of different compression functions. With a suitable mapping function the effective SINR is mapped to a block error rate for an AWGN channel. Now, an MCS is selected according to a pre-defined threshold. Unfortunately, to our best knowledge, all publications which focus on adaptive modulation and coding consider linear receivers for which the SINR per resource element can be computed in closed-form. The proposed graph-based receiver resembles a non-linear receiver, thus, a straight-forward computation of the SINR after detection and estimation is not possible. The probabilistic approach makes it difficult to

Table 5.2: List of applied modulation and coding schemes.

MCS	Modulation	Code rate R	Max spectral efficiency opt.	Max spectral efficiency LTE
1	QPSK	1/3	1.32	1.0
2	QPSK	1/2	2	1.52
3	QPSK	2/3	2.64	2.0
4	QPSK	3/4	3	2.27
5	16-QAM	1/2	4	3.02
6	16-QAM	2/3	5.28	3.99
7	16-QAM	3/4	6	4.54
8	64-QAM	2/3	7.92	5.99
9	64-QAM	3/4	9	6.80

predict how the received signals of different receive antennas are combined. Up to now, the calculation of the SINR in combination with a nonlinear receiver remains an open research topic. The BLER results for the A1, B1, C1, and C2 NLOS channel are given in Figures 5.28 to 5.31 as a function of E_s/N_0. In order to improve the readability, the results are focused to the range of interest between a BLER of $1 \cdot 10^{-2}$ and $1 \cdot 10^{-3}$. The results of the four receiver variants for one modulation and coding scheme are encircled and denoted with the corresponding MCS index.

The general conclusion from these figures is that the proposed graph-based framework supports all evaluated modulation and coding schemes with all considered channel models.

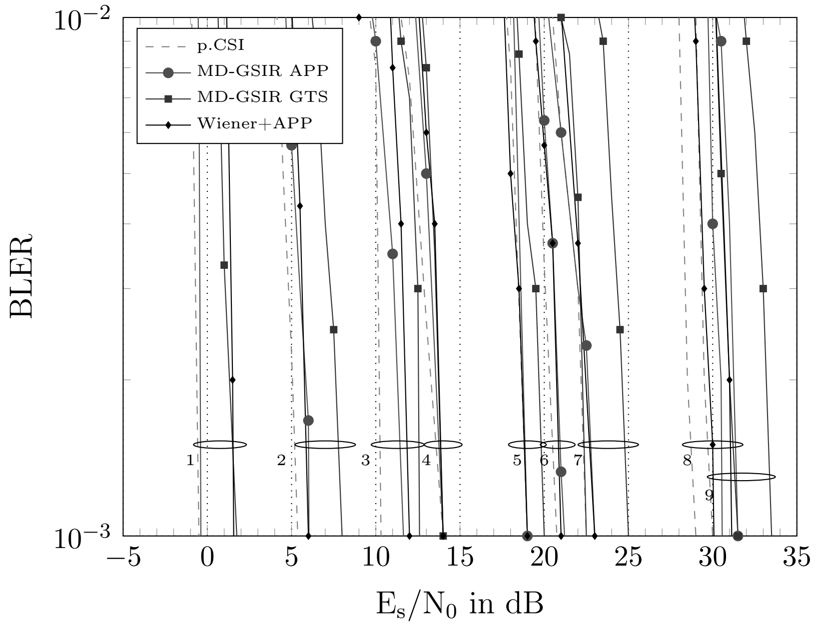


Figure 5.28: BLER results with the MD-GSIR with APP and GTS detection for all considered MCSs and WINNER A1 NLOS channel.

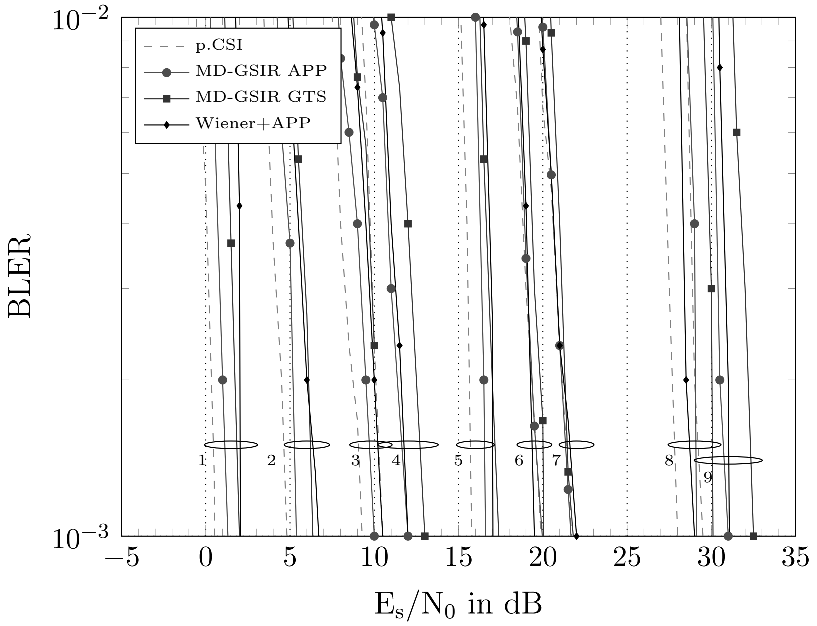


Figure 5.29: BLER results with the MD-GSIR with APP and GTS detection for all considered MCSs and WINNER B1 NLOS channel.

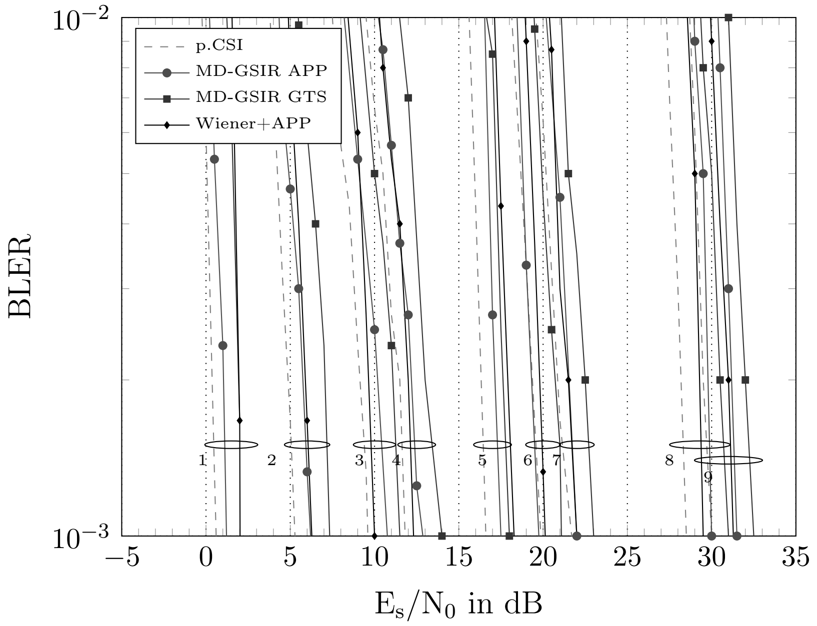


Figure 5.30: BLER results with the MD-GSIR with APP and GTS detection for all considered MCSs and WINNER C1 NLOS channel.

Naturally, the performance of the MD-GSIR is best with the optimum APP detection and competes well with the Wiener+APP receiver. The important fact here is that no sophisticated filtering algorithm is applied within the MD-GSIR. Instead, the proposed low-complexity transfer nodes facilitate the message exchange and yield accurate channel

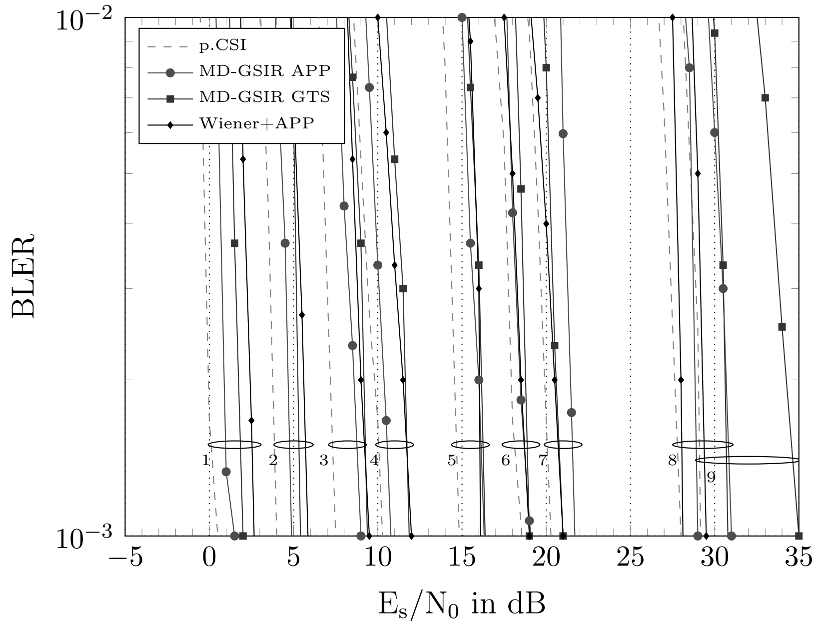


Figure 5.31: BLER results with the MD-GSIR with APP and GTS detection for all considered MCSs and WINNER C2 NLOS channel.

estimates. The Gaussian approximation which is applied for channel estimation reduces the complexity significantly, which is increasing linearly with the number of transmit and receive antennas as well as number of resource elements. The performance of the MD-GSIR with GTS detection is slightly worse for the majority of MCSs. The number of significant leaves M_{sl}, which determine the complexity and performance of the GTS, is chosen as presented in Table 3.1. That is, M_{sl} is set to $2, 8$, and 30 for QPSK, 16-QAM, and 64-QAM, respectively. A complexity reduction of up to 50% is achieved. An exception is the result obtained with 64-QAM, code rate 3/4 (MCS 9), and the C2 NLOS channel with the MD-GSIR GTS. A convergence with the chosen value of $M_{\mathrm{sl}} = 30$ was not possible and had to be increased significantly, i.e. a value of $M_{\mathrm{sl}} = 80$ is used. The importance of an initialization also for data detection becomes evident, which allows a reduction of the parameter M_{sl} without sacrificing the performance. The performance of the MD-GSIR with GTS detection w.r.t. the spectral efficiency is compared with the Wiener+APP receiver for the C1 NLOS channel in Figure 5.32. For comparison, two versions of an APP detector with perfect CSI are included. The first takes the bandwidth efficiency into account, while the second assumes an ideal system without any losses due to the signaling overhead. Additionally, the curves of the MD-GSIR with the reduced pilot pilot grid, shown in Figure 5.24b, is included to illustrate the achievable gain w.r.t. spectral efficiency by reducing the pilot overhead. Thereby, reducing the pilot overhead by 50% increases the bandwidth efficiency from 75.6% to 80%. It can be seen from the results that especially at low SNR, the iterative channel estimation turns out to be beneficial since the MD-GSIR with GTS detection outperforms the Wiener+APP receiver. With increasing SNR, however, the Wiener+APP receiver is able to achieve the same performance for the MCSs 3 and 4 and yields a better performance for the remaining

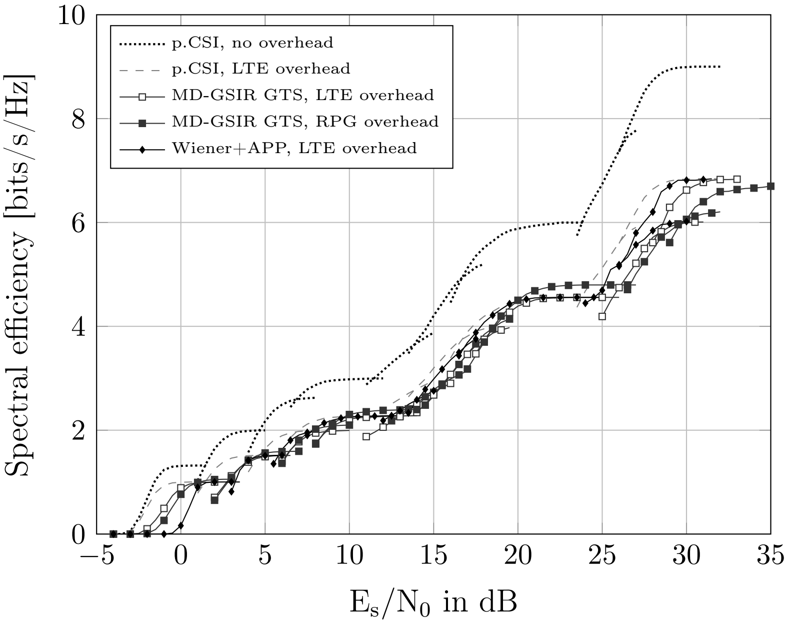


Figure 5.32: Spectral efficiency of the MD-GSIR with GTS detection for the WINNER C1 NLOS channel.

MCSs. The MD-GSIR with the reduced pilot grid achieves a higher spectral efficiency for the modulation and coding schemes 2 to 7. Especially with 16-QAM, the gain due to the reduced pilot overhead is largest and even surpasses the spectral efficiency of the APP detector with perfect channel state information. However, it can be seen that for low SNR and higher-order modulation, such as 64-QAM, the reduced pilot grid does not yield the required channel estimation accuracy. Nevertheless, as part of an extended modulation and coding scheme, which additionally takes the pilot grid into account, can increase the overall efficiency of the receiver.

In general, the performance loss due to the signaling overhead is significant and amounts to roughly 5 dB for 64-QAM. Obviously, this loss cannot be mitigated by just reducing the pilot overhead. Hence, it is proposed in [TM00, CYY09] to spare the cyclic prefix, which improves the bandwidth efficiency at the cost of increased intersymbol interference. An adaptation to the proposed graph-based receiver might further improve the efficiency.

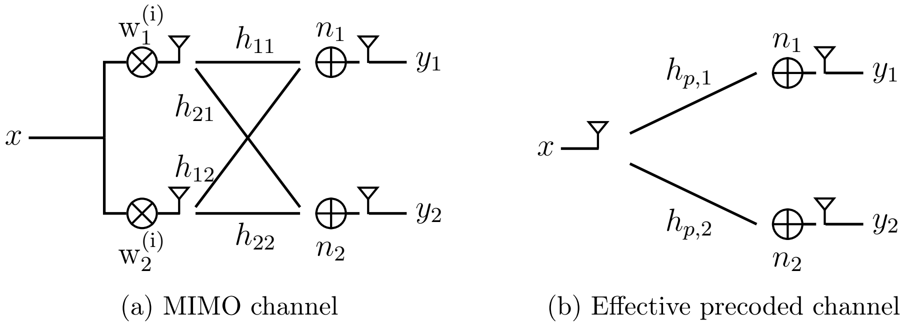
5.6 Codebook-based Beamforming

It has been shown in the previous sections that the proposed multi-dimensional graph-based receiver is able to achieve the challenging goals of high spectral efficiency as well as reliability for a variety of channel scenarios. Accurate channel state information at the receiver side are necessary to provide a low bit error ratio in combination with higher-order modulation and high code rates. Further improvements in terms of spectral efficiency and bit error performance can be achieved with channel state information at the transmitter side (CSIT). In combination with OFDM, each subcarrier can be precoded individually and thereby approach the theoretical capacity bound [SW09]. Precoded pilot symbols are inserted in the data stream in order to facilitate coherent detection of the precoded data symbols. Although beamforming introduces several advantages for data detection, channel estimation does not benefit to the same extent. More specifically, the proposed transfer nodes (cf. Section 5.1.4) model the difference between adjacent channel coefficients in order to establish a message exchange. They inherently assume that the channel is continuous in time and/or frequency, which is a common assumption since the channel is typically highly correlated in both domains. When beamforming is applied, the channel can be represented by a weighted superposition of the beamforming weights and the physical MIMO channel as exemplary depicted in Figure 5.33. The corresponding precoded system is given by

$$\begin{aligned} \boldsymbol{y}[l,k] &= \boldsymbol{H}[l,k]\mathbf{w}^{(i)}[l,k]x[l,k] + \boldsymbol{n}[l,k] && (5.66)\\ &= \mathbf{h}_{\mathrm{p}}[l,k]x[l,k] + \boldsymbol{n}[l,k]. && (5.67) \end{aligned}$$

With beamforming weights, $\mathbf{w}^{(i)}$, changing from subcarrier to subcarrier, the continuous channel response is not longer maintained. As a result, filter-based channel estimation algorithms degrade w.r.t. their achievable performance. The negative impact of beamforming on channel estimation has been reported in e.g. [SF08, SW09]. The common approach to mitigate the detrimental effects of a non-continuous channel is to choose the beamformer such that the effective channel maintains its “smoothness”. Obviously, codebook-based approaches can hardly be implemented following this idea since only a limited number of beamforming weights is available. The inherent quantization errors will destroy the smoothness of a channel. By increasing the codebook size and, thus, reducing the quantization error, the advantage of a reduced feedback overhead is lost. Accordingly, discontinuities of the effective beam-weighted channel are inevitable. Therefore, the transfer nodes presented in Section 5.1.4 are revised to take codebook-based beamforming into account. The definition of a transfer node as given by (5.19) is revisited:

$$\Delta_{n',m'}[l',k'] \doteq h - \omega h'.$$



(a) MIMO channel (b) Effective precoded channel

Figure 5.33: Physical MIMO channel and effective beam-weighted channel.

Considering the system described by (5.67), the transfer node for a precoded channel is given by

$$\Delta_{p,n',m'}[l',k'] \doteq h_p - \omega_p h'_p, \tag{5.68}$$

where the antenna index of a precoded channel h_p has been omitted to improve readability. The main difference between the precoded and the original transfer nodes is that the precoded channel can have a non-zero mean depending on the transmit beamforming scheme [PLL10]. The tuning factor ω_p, however, can be used to shift this mean value to zero and thus, fulfill the zero-mean approximation again. A calculation of the mean value is given in [PLL10]. As will be shown later, the resulting variance values will effectively prevent a message exchange and hence, the mean value of a message has no influence at all in the message combining process. Without loss of generality, the calculation of the variance for the precoded transfer nodes based on a zero-mean approximation is discussed in the following. Equivalently to (5.23), the variance of a precoded transfer node is given by

$$\begin{aligned}\sigma^2_{\Delta,p,n',m'}[l',k'] &= \mathrm{E}\left\{|h_p - \omega h'_p|^2\right\} \\ &= \mathrm{E}\left\{|h_p|^2\right\} + \mathrm{E}\left\{|h'_p|^2\right\} - \mathrm{E}\left\{|\omega h_p h'^*_p|^2\right\} - \mathrm{E}\left\{|\omega h^*_p h'_p|^2\right\}. \end{aligned} \tag{5.69}$$

Depending on the chosen codebook and beam-weights, (5.69) can be further simplified. Two cases can be identified: (1) the beam-weights of two adjacent OFDM subcarriers and/or symbols are equal and (2) they are chosen differently. A major difference to the conventional transfer nodes is that an individual variance value can be assigned to a transfer node for each OFDM subcarrier, OFDM symbol, and transmit antenna. Previously, the variance value was only determined by the dispersion parameters of the corresponding domains and thus did not change over frequency, time, and space. With beamforming the variance value assigned to a transfer node depends on the chosen beam-weight as well. By assuming that adjacent beam-weights are equal, a beam-weighted channel coefficient h_p is given by

$$\begin{aligned} h_p &= \mathrm{w}_1^{(i)} \cdot h_1 + \mathrm{w}_2^{(i)} \cdot h_2, \\ h'_p &= \mathrm{w}_1^{(i)} \cdot h'_1 + \mathrm{w}_2^{(i)} \cdot h'_2. \end{aligned}$$

With these assumptions, (5.69) results in

$$\begin{aligned}\sigma^2_{\Delta,p,n',m'}[l',k'] = {} & \mathrm{E}\left\{|\mathrm{w}_1^{(i)} h_1 + \mathrm{w}_2^{(i)} h_2|^2\right\} + \mathrm{E}\left\{|\mathrm{w}_1^{(i)} h'_1 + \mathrm{w}_2^{(i)} h'_2|^2\right\} \\ & - \mathrm{E}\left\{|\omega_p(\mathrm{w}_1^{(i)} h_1 + \mathrm{w}_2^{(i)} h_2)(\mathrm{w}_1^{(i)} h'_1 + \mathrm{w}_2^{(i)} h'_2)^*|^2\right\} \\ & - \mathrm{E}\left\{|\omega_p(\mathrm{w}_1^{(i)} h_1 + \mathrm{w}_2^{(i)} h_2)^*(\mathrm{w}_1^{(i)} h'_1 + \mathrm{w}_2^{(i)} h'_2)|^2\right\}. \end{aligned} \tag{5.70}$$

A property of the DFT is that the sum of the squared absolute value of the beamforming weights is equal to one, i.e. $|\mathrm{w}_1^{(i)}|^2 + |\mathrm{w}_2^{(i)}|^2 = 1$. The Grassmannian codebook fulfills this property only approximately, depending on the selected beam-weights. Nevertheless, for DFT and Grassmannian codebooks the calculation of the variance is further simplified to

$$\sigma^2_{\Delta,p,n',m'}[l',k'] = 2\left(1 - \mathrm{Re}\left[\mathrm{E}\left\{h^* h'\right\}\right]\right), \tag{5.71}$$

which is identical to the variance calculation of the original transfer nodes. This is especially beneficial for wireless systems such as LTE and/or LTE-A. There, the same beam-weight is applied in the frequency domain [Cox12]. As a consequence, the MD-GSIR does not require any adaptation to a codebook-based beamforming scheme. Certainly, the beam-weights can be adapted over time. However, since codebook based precoding requires feedback, the channel is typically assumed to be slowly-varying. It is therefore likely that the beam-weights change only on rare occasions in the time domain.

In case the beamforming weights differ for two adjacent OFDM subcarriers and/or OFDM symbols, the calculation of the variance is as follows:

$$\sigma^2_{\Delta,p,n',m'}[l',k'] = 2\left(1 - \left(\mathrm{Re}\left[\mathrm{w}_1^{(i)}(\mathrm{w}_1^{(j)})^* \,\mathrm{E}\left\{h_1^* h_1'\right\}\right] + \mathrm{Re}\left[\mathrm{w}_2^{(i)}(\mathrm{w}_2^{(j)})^* \,\mathrm{E}\left\{h_2^* h_2'\right\}\right]\right)\right). \tag{5.72}$$

It is easy to see that (5.72) reverts to (5.71) if $\mathrm{w}_1^{(i)}{=}\mathrm{w}_1^{(j)}$ and $\mathrm{w}_2^{(i)}{=}\mathrm{w}_2^{(j)}$ under the assumption that the two channels, h_1 and h_2, follow the same statistics. The resulting variance matrices in the frequency domain are given exemplary for the WINNER C2 NLOS channel and the DFT codebook as well as for the Grassmannian codebook:

$$\sigma^2_{\Delta,\mathrm{DFT}} = \begin{bmatrix} 0.0011 & 2.0000 & 1.0006 & 1.0006 \\ 2.0000 & 0.0011 & 1.0006 & 1.0006 \\ 1.0006 & 2.0000 & 0.0011 & 1.0006 \\ 1.0006 & 1.0006 & 2.0000 & 0.0011 \end{bmatrix}, \tag{5.73}$$

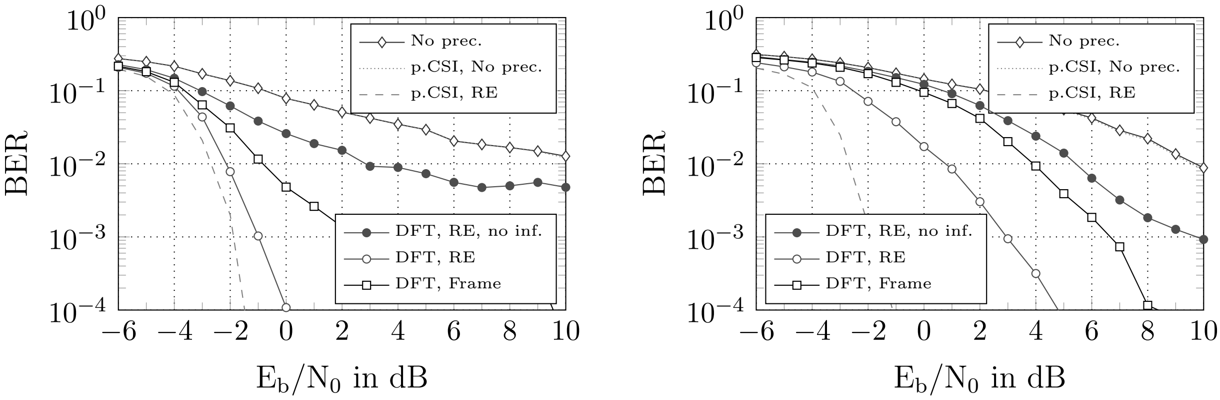
$$\sigma^2_{\Delta,\mathrm{GRASS}} = \begin{bmatrix} 0.0011 & 2.0000 & 2.0000 & 2.0000 \\ 2.0000 & 0.0011 & 2.0000 & 2.0000 \\ 2.0000 & 2.0000 & 0.0011 & 2.0000 \\ 2.0000 & 2.0000 & 2.0000 & 0.0011 \end{bmatrix}, \tag{5.74}$$

whereas the index of the columns denotes one beam-weight and the index of the rows refer to the beam-weight chosen for the adjacent OFDM subcarrier or OFDM symbol. Hence, for the entries along the main diagonal, the same beam-weight is chosen for two adjacent OFDM subcarriers. The resulting variance depends only on the physical channel statistics. A change in the beam-weight from e.g. the first weight $\mathrm{w}^{(1)}$ to the second weight $\mathrm{w}^{(2)}$ corresponds to the entry (1,2) within the matrix. As can be seen, the variance of a transfer node increases significantly with a change in the beam-weights. The variance of a message represents the reliability of the mean value, whereas a large value refers to an unreliable estimate and vice versa. Consequently, when two or more messages are combined, the message with the largest variance contributes the least to the resulting message. In case a message is sent to a coefficient node, which has been precoded with a different weight than its predecessor, the message is effectively ignored. Accordingly, the message exchange is interrupted. By comparing the two matrices, a difference between the DFT and the Grassmannian codebook becomes obvious. While a change in the beam-weights always leads to a variance of value 2 for the Grassmannian codebook, this is not the case with the DFT codebook. This effect is reasoned in the DFT beam-weights, shown in Table 2.2, which do not change for the first link h_{11} and/or h_{21}. Only the second channel link, h_{12} and/or h_{22}, experience a change in their phase in steps of 90°. Hereby a change of

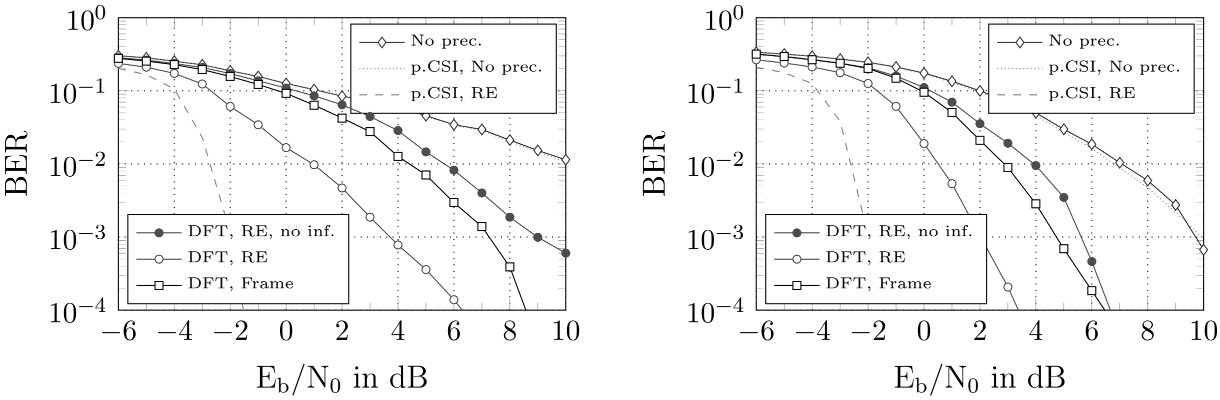
180° leads to a variance of 2 as well. It has to be mentioned that changing beamforming weights with each OFDM symbol has an additional impact on the channel estimation performance with the MD-GSIR. Since the message exchange is effectively interrupted between OFDM symbols with different weights it may happen that an OFDM symbol without pilot information is excluded from the overall message exchange. That is when the beam weights change twice between pilot symbols in the time domain. However, since this effect can only occur in fast-varying channels—where codebook-based beamforming is not applied—and rectangular pilot grids, it only has a minor influence. In order to circumvent this problem for slowly time-varying channels, the pilot grid can be adapted with a shift of pilot symbols in the time domain, which leads to the so-called diamond grid, displayed in Figure 5.24b. Alternatively, the update of beam-weights can be done per LTE frame. For time-varying channels often open-loop techniques, e.g. the Alamouti technique [Ala98], are preferred [Cox12].

A drawback of the proposed method is that knowledge of the applied beamforming weights is mandatory. Actually, only an indicator that the beam-weight has changed is required. However, given the actual weights, joint common and dedicated channel estimation could be performed as well. Unfortunately, it is solely the decision of the transmitter which exact precoding scheme it applies. Hence, knowledge of the applied beam-weight is typically not available. Nevertheless, it is shown in [PLL10] that the statistics of the beam-weighted channel can be estimated using the statistics of the common pilot channel. Accordingly, a change in the beamforming weights should be predictable. Furthermore, joint common and dedicated channel estimation without the knowledge of the applied beam-weights is presented in [MS03]. The actual estimation and/or prediction of beam-weights is beyond the scope of this thesis and remains an interesting topic for future research. For the remaining simulation results, the selection as well as the update interval of the beam-weights is optimal, i.e. the common channel is assumed to be perfectly known and feedback is send to the transmitter without any delays. All channels are assumed to be constant in time.

The BER results evaluating the gain due to codebook-based beamforming are presented in the following. Hereby, beamforming is applied in two different kinds. First, every resource element, i.e. OFDM subcarrier and/or OFDM symbol, is precoded with an individual weight. The corresponding results are denoted with “RE”. If the MD-GSIR is unaware of the applied weights, the curves are labelled with a “no inf.”. The second precoding option applies the same beamweight per LTE frame, that is all OFDM subcarriers and OFDM symbols. Knowledge of the applied weight is hereby not important since the channel maintains its continuity and the variance calculation of the transfer nodes remains unchanged. Additionally, an unprecoded MD-GSIR and an APP detector with perfect channel state information is included as well. The BER results with a DFT codebook are given in Figure 5.34. As expected, beamforming yields significant gains in all channel scenarios. Precoding per RE performs best, whereas the MD-GSIR requires the knowledge of the applied weights in order to have a good performance. Without knowledge of the applied weights, it is advisable to apply precoding on a frame-basis. The loss due to channel estimation is close to zero without precoding and up to several dB depending on the channel scenario. The loss is lowest for the A1 channel, which is



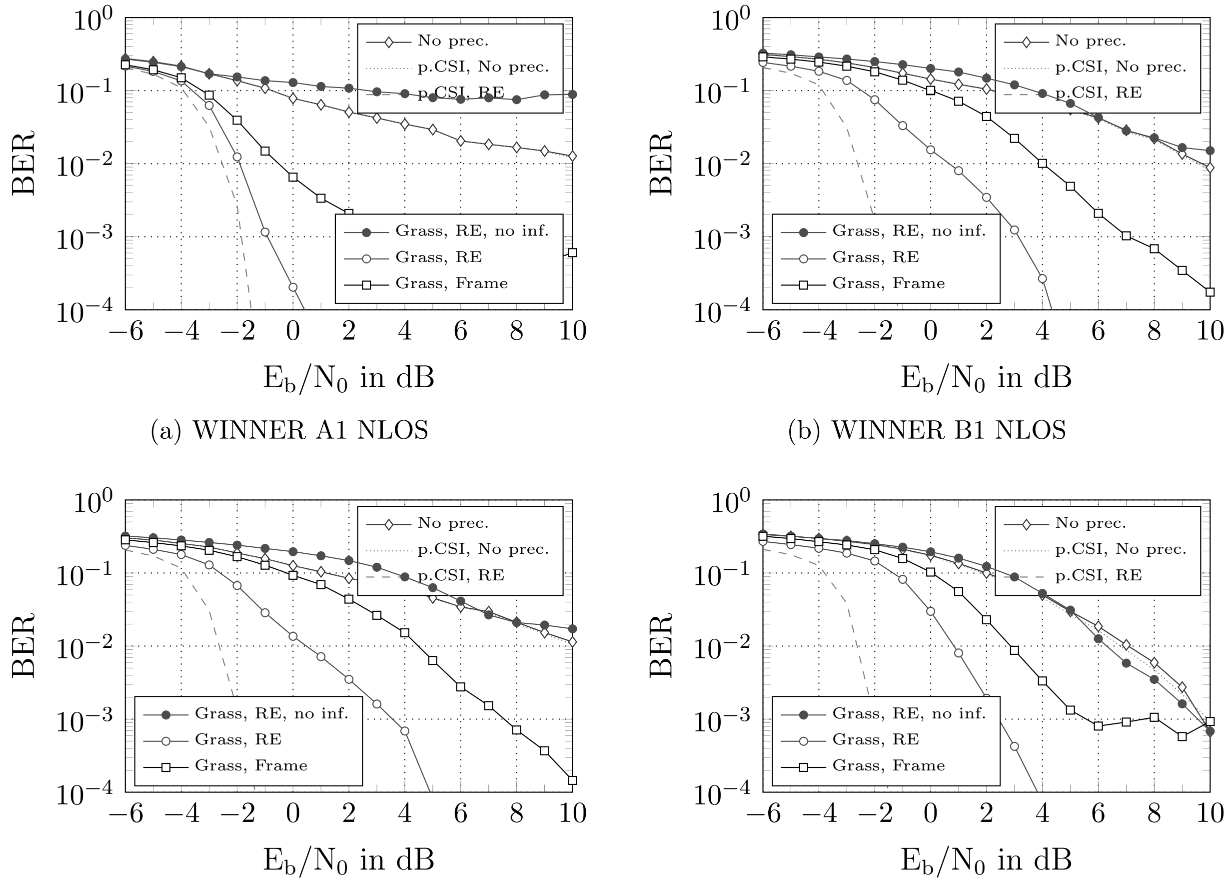
(a) WINNER A1 NLOS (b) WINNER B1 NLOS



(c) WINNER C1 NLOS (d) WINNER C2 NLOS

Figure 5.34: BER performance of MD-GSIR with DFT codebook-based beamforming.

anticipated since the diversity in the frequency domain is also lowest. A beam-weight is applied for a wide range of OFDM subcarriers. With increasing diversity, beam-weights change more often and the message exchange within the MD-GSIR is also increasingly limited. Nevertheless, a gain between 2 dB and 2.5 dB is observed for RE-based precoding compared to frame-based precoding. Similar results are observed in combination with a Grassmannian codebook, which can be seen in Figure 5.35. Given the a priori information of the applied beam-weights, the MD-GSIR is able to yield substantial BER improvements compared to an unprecoded system. With the exception of the A1 NLOS channel, the distance to a RE-precoded system with perfect channel state information is about 6 dB. A gain of about 4 dB is achieved compared to frame-based precoding. Overall, the performance of the frame-based precoding with the Grassmannian codebook is worse compared to the DFT codebook. The reason for the poor performance with the Grassmannian codebook lies in a miscalculation of the variance in the transfer nodes, which is caused by the assumption that the overall power of one codebook entry equals to one. In order to improve the performance, a more accurate variance calculation is required which needs again the knowledge of the applied beam-weight. The simple adap-



(a) WINNER A1 NLOS

(b) WINNER B1 NLOS

(c) WINNER C1 NLOS

(d) WINNER C2 NLOS

Figure 5.35: BER performance of MD-GSIR with Grassmannian codebook-based beamforming.

tation of the transfer nodes to codebook-based beamforming highlights the versatility of the proposed graph-based receiver concept.

5.7 Chapter Summary

A multi-dimensional graph-based receiver concept has been presented in this chapter. The design of the underlying factor graph structure follows the major premise of low complexity. By factorizing the conditional probability density function to its elementary components, i.e. single random variables, the complexity of the message generation is significantly reduced. However, the factorization of the pdf is based on approximations, which lead to a graph structure that contains cycles and thus, is suboptimal. In order to mitigate the deteriorating effects of the suboptimal structure, a special scheduling of the messages is designed. Although significant performance differences are observed for different schedules, the loss due to channel estimation remains rather large given certain channel conditions. An analysis of the exchanged messages revealed that the messages are correlated, which is contradictory to the major prerequisite of the sum-product algorithms, that is the exchange of extrinsic information. By taking the correlation in the message combining into account, the performance is again significantly improved. An additional gain is achieved by implementing the multi-objective PSO, presented in the previous chapter, to provide improved a priori information. With the Gaussian tree search detector, the complexity of the MD-GSIR can be further reduced. The BER and BLER performances are compared to a state-of-the-art iterative receiver based on a Wiener filter and an APP MIMO detector. It is shown that a similar performance is achieved given a dense pilot grid. However, the proposed concept is more robust, which means the MD-GSIR reaches a similar performance given less a priori information. Moreover, the required amount of pilot symbols is also reduced, which increases the spectral efficiency. The versatility of the graph-based receiver concept is illustrated by incorporating codebook-based beamforming. The factor graph structure as well as the message exchange based on transfer nodes does not need to be changed.

The MD-GSIR yields a remarkable performance on par with a state-of-the-art receiver, at a lower complexity. Due to its versatility it can be easily adopted to a various conditions, such as beamforming and/or irregular pilot grids.

6

Coded Sampling Bound

WIRELESS COMMUNICATION systems have evolved dramatically in recent years. In particular, the invention of bit-interleaved coded modulation (BICM) attracted the focus of research. It has been recognized that BICM with iterative processing (BICM-ID) [LR97, LR98] outperforms its non-iterative counterpart. Furthermore, it has been identified that all parts of the underlying systems—e.g. channel code, interleaver and mapping—need to be optimized to achieve near-capacity performance. However, the majority of "capacity-oriented" research assumes perfect channel knowledge. The performance of a coherent receiver significantly relies on the accuracy of channel state information, particularly if the channel is time-varying. The most common method to provide the receiver with channel state information (CSI) is to embed pilot symbols within the transmitted data stream [Cav91]. In order to reconstruct the CSI at the positions of the unknown data symbols, interpolation and filtering are often adopted [Hoe91, HKR97b, Aue03b]. If channel estimation is decoupled from channel decoding, the maximum pilot spacing in time domain is determine by the Nyquist-Shannon sampling theorem according to $D_{\mathrm{t,max}} = 1/(2f_{\mathrm{D,max}}T_{\mathrm{s}})$, where $f_{\mathrm{D,max}}T_{\mathrm{s}}$ is the normalized Doppler frequency. Although PACE is able to obtain accurate CSI, it inherently reduces the efficiency of a system, due to the large pilot overhead. It is therefore desirable to reduce the pilot overhead as much as possible without affecting the achievable performance.

Iterative joint channel estimation and detection aims to mitigate the trade-off between overhead and performance, and provides accurate CSI with a minimum of pilots. Interestingly, optimum pilot grids, sequences, and required pilot overhead are mainly studied for non-iterative pilot-based channel estimation [Li02, CL05, AC09a, AC09b]. Mainly, this is because the BER performance of an iterative coded receiver with channel estimation impairments is difficult to predict [MS02].

In several contributions it has been observed that the pilot spacing for iterative receivers is not strictly bounded by the Nyquist-Shannon sampling theorem [tBSS00, SJS03, KSHA10, XM11]. Although the sampling theorem is initially not fulfilled, the channel response can be reconstructed sufficiently well so that the convergence during following iterations is possible. In [SJS03, KSHA10, XM11], this behavior is observed for one particular channel code and is often attributed to special properties of the proposed receiver. Moreover, ten Brink observed in [tBSS00], that the sampling theorem is not strictly limiting the initial pilot spacings in combination with iterative channel estimation and data detection. It is concluded therein that by exploiting the redundancy of the code, pilots can be placed arbitrarily. So far, to our best knowledge no publication analyzed the impact of the channel code and/or the code rate on the maximum spacing of pilot symbols in combination with iterative receivers. Two strategies to design a pilot grid are often applied: 1. Under guidance of the Nyquist-Shannon sampling theorem, a pilot spacing is chosen which guarantees that the worst case scenario is sufficiently sampled. This typically results in a large pilot overhead which deteriorates the spectral efficiency. 2. The maximum pilot spacing is designed for a specific receiver setup such that a certain BER threshold is surpassed. The drawback of this method is that the effect of the pilot spacing can not be observed directly, as channel estimation errors do not directly translate into bit errors. It is therefore difficult to judge whether a certain BER is caused by noise or by an insufficient pilot spacing.

Furthermore, it is shown that the maximum pilot spacing can be significantly extended compared to the Nyquist-Shannon sampling theorem when iterative joint channel estimation and data detection is applied. In coded transmission systems employing pilot symbols two forms of redundancy exist: The redundancy due to channel coding and the redundancy due to pilot symbols. By combining both types, a so-called *coded sampling bound* is derived in this chapter. This bound provides a limit on the pilot density given iterative semi-blind channel estimation (SBCE) based on linear FIR filters. However, the coded sampling bound is a semi-analytical bound, rather than an analytical bound because receiver sub-optimalities such as a finite sequence length, finite number of iterations, sub-optimum receiver structure, and others, can not be expressed in a closed form solution so far. Accordingly, the coded sampling bound depends on the type of receiver, the signal-to-noise ratio (SNR) as well as the channel code and the code rate. It is shown that with a suitable channel code and appropriate code rate the maximum spacing of pilot symbols can be set arbitrarily large. In combination with a posteriori probability (APP) detection and quasi-noise-free transmission, the obtained results can be interpreted as a lower bound. Moreover, the results indicate that for the majority of wireless systems, neither the channel codes nor the pilot density are specifically suited for iterative semi-blind channel estimation and thus are not able to fully exploit the achievable gains with respect to the spectral efficiency.

The aim of this chapter is to answer the following two questions: (1) How much training is required for an iterative semi-blind receiver? And (2), is the maximum pilot spacing upper limited? Towards these goals, in Section 6.1, the fundamentals of a linear semi-blind channel estimation are reviewed and a linear FIR filter which takes reliability information into account is developed. In order to be able to quantify the maximum pilot spacing in terms of MSE, Section 6.2 introduces a parametrized version of the general MSE. On the basis of the parametrized MSE, upper bounds for the spacing for a non-iterative as well as iterative receiver are given in Section 6.3 and Section 6.4, respectively. Additionally, an EXIT chart analysis of the iterative receiver provides insights on the convergence behavior as a function of the pilot density. Implications for a practical receiver design are outlined in Section 6.4.3 Finally, the conclusions of this chapter are drawn in Section 6.5.

6.1 Fundamentals of Iterative Channel Estimation

In order to ease the derivation of the coded sampling, the focus will be on the time domain throughout this chapter. An extension to other domains (e.g. the frequency domain) and/or multi-dimensional estimators is straightforward. Considering a frequency-flat time-varying channel, the channel model can be written in complex baseband notation as

$$y[k] = x[k]h[k] + n[k],$$

where $y[k]$ is the kth observation, $x[k]$ a data or pilot symbol, $h[k]$ the time-varying channel coefficient (referred to as CSI), and $n[k]$ a sample of a white Gaussian noise process. The time-variation of the channel is modeled by the so-called Jakes Doppler power spectrum, which is non-zero within $[-f_{\mathrm{D,max}}, f_{\mathrm{D,max}}]$. Moreover, the distribution is known to the receiver. The corresponding model in vector-matrix form is given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{n}. \tag{6.1}$$

The diagonal matrix $\boldsymbol{X}$, which refers to the transmitted sequence, is of size $(N_{\mathrm{D}}+N_{\mathrm{P}}) \times (N_{\mathrm{D}}+N_{\mathrm{P}})$, where N_{D} refers to the number of data symbols and N_{P} to the number of pilot symbols per data block. Throughout this paper, It is assumed that the pilot sequence is equi-distantly multiplexed into the data stream with a pilot spacing D_{t}. Although other pilot grids are possible, it has been shown that the equi-distant spacing minimizes the mean squared error (MSE) [NC98] of the channel estimates as well as maximizes the capacity [ATV02]. The channel vector $\boldsymbol{h}$ is assumed to be wide-sense stationary complex Gaussian with zero mean.

After initial calculation of the channel response at pilot positions $\mathcal{P}$, for example, by means of least-squares channel estimation, a Wiener interpolation filter is commonly applied to obtain an estimate of $h[k]$ for the unknown data symbols [HKR97b, Aue04], i.e.:

$$\hat{h}[k] = \sum_{k' \in \mathcal{P}} w[k;k'] \cdot \left([\boldsymbol{X}]_{k',k'}\right)^{-1} \cdot y[k'], \tag{6.2}$$

where $[\cdot]_{k',k'}$ refers to the element of the k'th row and column. The coefficients $\mathbf{w}_k = [w[k;0], w[k;1], \ldots, w[k;N_{\mathrm{P}}]]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{P}} \times 1}$ of the Wiener filter are given by

$$\begin{aligned} \mathbf{w}_k &= \arg\min_{\mathbf{w}'_k} \mathrm{E}\left\{|h[k] - \hat{h}[k]|^2\right\} \\ &= \arg\min_{\mathbf{w}'_k} \mathrm{E}\left\{|h[k] - \mathbf{w}'^{\mathrm{T}}_k \boldsymbol{X}_{\mathrm{P}}^{-1} \boldsymbol{y}_{\mathrm{P}}|^2\right\}. \end{aligned} \tag{6.3}$$

Hereby, $\boldsymbol{X}_\mathrm{P}$ corresponds to a diagonal matrix which contains only the N_P pilot symbols. The well known Wiener-Hopf equations can be used to calculate the filter coefficients in (6.3) as

$$\mathbf{w}_k^\mathrm{T} = \boldsymbol{\theta}_{\boldsymbol{hy}}^\mathrm{T}[k]\boldsymbol{\theta}_{\boldsymbol{yy}}^{-1}, \tag{6.4}$$

where

$$\boldsymbol{\theta}_{h\boldsymbol{y}}[k] = \mathrm{E}\left\{\boldsymbol{y}_\mathrm{P} h^*[k]\right\}, \qquad \boldsymbol{\theta}_{\boldsymbol{yy}} = \mathrm{E}\left\{\boldsymbol{y}_\mathrm{P}\boldsymbol{y}_\mathrm{P}^H\right\}, \tag{6.5}$$

correspond to the cross-correlation vector between the received sequence $\boldsymbol{y}$ and the desired response $h[k]$, and the auto-correlation matrix of the received pilots, respectively. Optimally, filter coefficients are calculated by (6.3) during runtime. However, in order to reduce the computational complexity, coefficients can be pre-calculated and stored in tables [HKR97b].

Only pilot symbols are used to determine the filter coefficients based on (6.3), which is the case during the initialization of an iterative receiver. As mentioned previously, after initialization, reliably detected symbols may be used as pseudo-pilots to further refine the channel estimation accuracy. The approach in this chapter is similar to [MS05], where soft bits are used to represent the reliability of a soft decision. Hereby, a soft bit is motivated by the following definitions

$$P_{c_k|\mathbf{y}}(c_k = +1|\mathbf{y}) = \frac{\exp(\mathrm{L_A}(c_k))}{1+\exp(\mathrm{L_A}(c_k))}, \tag{6.6}$$

$$P_{c_k|\mathbf{y}}(c_k = -1|\mathbf{y}) = \frac{\exp(-\mathrm{L_A}(c_k))}{1+\exp(-\mathrm{L_A}(c_k))}. \tag{6.7}$$

Based on these equations, a soft-bit is given by [HOP96]

$$\begin{aligned}\lambda_k^{(i)} &= (+1)\cdot\frac{\exp(\mathrm{L}_\mathrm{A}^{(i)}(c_k))}{1+\exp(\mathrm{L}_\mathrm{A}^{(i)}(c_k))} + (-1)\cdot\frac{\exp(-\mathrm{L}_\mathrm{A}^{(i)}(c_k))}{1+\exp(-\mathrm{L}_\mathrm{A}^{(i)}(c_k))} \\ &= \frac{\exp(\mathrm{L}_\mathrm{A}^{(i)}(c_k))-1}{\exp(\mathrm{L}_\mathrm{A}^{(i)}(c_k))+1} = \tanh\left(\frac{\mathrm{L}_\mathrm{A}^{(i)}(c_k)}{2}\right),\end{aligned} \tag{6.8}$$

where i and k denote the ith iteration and the kth time index, respectively. In case of BPSK modulation—which is assumed in the following to simplify the notation—the soft decision of the kth information bit (c_k) is equivalent to the soft decision of the kth transmitted symbol ($x[k]$). A diagonal reliability matrix can be constructed as follows:

$$\left[\boldsymbol{P}^{(i)}\right]_{n,n} = \begin{cases} 1, & n\in\mathcal{P} \\ |\lambda_n^{(i)}|, & n\notin\mathcal{P} \\ 0, & \text{else.}\end{cases} \tag{6.9}$$

Naturally, the reliability of pilot symbols is equal to one, whereas the reliability of detected data symbols depend on their L-values. As data symbols are detected more and more reliably with the number of iterations, the reliability matrix $\boldsymbol{P}^{(i)}$ has to be recalculated for each iteration. Accordingly, the Wiener-Hopf equations have to be modified taking reliability information into account. Revisiting (6.3), the diagonal matrix $\boldsymbol{X}$ is weighted with the reliability matrix $\boldsymbol{P}^{(i)}$, which accounts for the fact that reliably detected data symbols are used as pseudo pilot symbols:

$$\mathbf{w}_k = \arg\min_{\mathbf{w}'_k} \mathrm{E}\left\{|h[k] - \mathbf{w}'^{\mathrm{T}}_k \boldsymbol{P}^{(i)}\boldsymbol{y}|^2\right\}. \tag{6.10}$$

Although $\boldsymbol{P}^{(i)}(\boldsymbol{X})^{-1}$ is not strictly an identity matrix, it is treated as such, since the entries which are unknown (i.e. data symbols with missing a priori information ($\mathrm{L_A} = 0$)) correspond to zero values in the coefficient vector $\mathbf{w}_k$. Accordingly, the filter ignores the missing/unkown symbols [MS05]. Without a priori information, the reliability matrix has only non-zero entries on pilot positions, which results in the pilot-based channel estimator described above. In case data symbols can be reconstructed perfectly, the reliability matrix has ones along its main diagonal and, thus, $N_\mathrm{P} + N_\mathrm{D}$ filter coefficients are used to estimate the channel coefficient $h[k]$. The corresponding Wiener filter acts as a smoothing filter. The cross-correlation vector as well as the auto-correlation matrix used within the Wiener-Hopf equations, given in (6.3), are changed to take reliability information into account:

$$\boldsymbol{\theta}^{(i)}_{h\boldsymbol{y}}[k] = \mathrm{E}\left\{\boldsymbol{P}^{(i)}\boldsymbol{y}h^*[k]\right\}, \tag{6.11a}$$

$$\boldsymbol{\theta}^{(i)}_{\boldsymbol{yy}} = \mathrm{E}\left\{\boldsymbol{P}^{(i)}\boldsymbol{y}(\boldsymbol{P}^{(i)}\boldsymbol{y})^{\mathrm{H}}\right\}. \tag{6.11b}$$

Following the approach in [MS05], a reliability function

$$f^{(i)}(n) = \left[\boldsymbol{P}^{(i)}\right]_{n,n} \tag{6.12}$$

is introduced. Accordingly, the cross-correlation vector in (6.11a) can be rewritten as

$$\left[\boldsymbol{\theta}^{(i)}_{h\boldsymbol{y}}[k]\right]_n = f^{(i)}(n)\,\theta_{HH}(n-k), \tag{6.13}$$

where $\theta_{HH}(n-k)$ is the domain-specific auto-correlation function as defined in (2.15) for the time domain. Furthermore, the auto-correlation matrix in (6.11b) is given by

$$\left[\boldsymbol{\theta}^{(i)}_{\boldsymbol{yy}}\right]_{n,m} = f^{(i)}(n)\cdot f^{(i)}(m)\cdot\left\{\frac{1}{\xi}\delta(n-m) + \theta_{HH}(n-m)\right\}. \tag{6.14}$$

Hereby, ξ refers to the SNR at pilot and data positions: $\xi = E_\mathrm{p}/N_0 = S_\mathrm{p}E_\mathrm{d}/N_0$, where E_p and E_d refer to the energy per pilot and data symbol, respectively. Throughout this thesis, it is assumed that pilots have the same average power as data symbol, i.e. the pilot boost is set to $S_p = 1.0$.

In Figure 6.1, the structure of an iterative semi-blind receiver is shown. The received sequence in combination with the a priori information of pilot symbols is used to calculate initial channel estimates $\hat{\boldsymbol{h}}$. Subsequently, a soft-input soft-output (SiSo) detector uses the initial channel estimates together with the variance of the channel estimates, represented as [MS05]

$$\left[\boldsymbol{\sigma}_h^2\right]_k = \theta_{HH}\left(0\right) - \boldsymbol{\theta}_{\boldsymbol{hy}}^{\mathrm{T}}[k]\mathbf{w}_k, \tag{6.15}$$

to calculate the a posteriori LLR values $\mathrm{L}_{\mathrm{L},1}$. After channel decoding, the information is interleaved and fed back to both, Wiener filter and MIMO detector. Refined channel estimates and additional a priori information, $\mathrm{L}_{\mathrm{A},1}$, for the data detection yield improved data estimates. Thus, iterative joint channel estimation and data detection results in an overall improved performance. By using the reliably detected data symbols as pseudo pilots, the initial constraints of pilots given by the sampling theorem can be relaxed, which reduces the overhead introduced by pilots, as will be shown in Section 6.4.

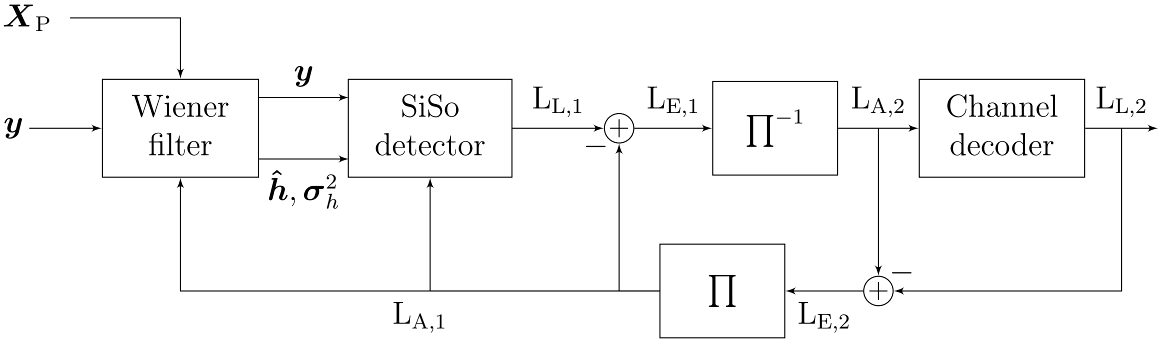


Figure 6.1: Iterative receiver structure with APP detector and Wiener filtering.

6.2 Decomposition of the MSE

For the assessment of the channel estimation impairment caused by an insufficient pilot density, the mean squared error of the overall estimation error $\varepsilon[k] = h[k] - \hat{h}[k]$ is used. While the MSE is typically applied to express the general performance of a channel estimation algorithm, it is also possible to separate it into a noise term and an interpolation term, which enables the identification of distortions caused by an inappropriate pilot density.

In general, the MSE of a linear estimator is given by [HKR97b]

$$\begin{aligned}\sigma_\varepsilon^2[k] &= \mathrm{E}\left\{|\varepsilon[k]|^2\right\} = \mathrm{E}\left\{|h[k] - \hat{h}[k]|^2\right\} \\ &= \mathrm{E}\left\{|h[k]|^2\right\} - 2\,\mathrm{Re}\left\{\mathbf{w}_k^{\mathrm{H}}\boldsymbol{\theta}_{\boldsymbol{hy}}[k]\right\} + \mathbf{w}_k^{\mathrm{H}}\boldsymbol{\theta}_{\boldsymbol{yy}}\mathbf{w}_k. \end{aligned} \tag{6.16}$$

The MSE in (6.16) depends on the time index k, but is averaged in the following in order to simplify the model, i.e.: $\sigma_\varepsilon^2[k] \to \sigma_\varepsilon^2$. Furthermore, the channel estimates of a linear receiver can be decomposed into a signal part and a noise part, denoted by $\hat{h}[k] = \mathbf{w}_k^{\mathrm{H}}\boldsymbol{h} + \mathbf{w}_k^{\mathrm{H}}\boldsymbol{n}$. Under the assumption that the noise and the channel response are uncorrelated, the MSE is separated into a noise and an interpolation error as follows [CA07]:

$$\begin{aligned}\sigma_\varepsilon^2 &= \mathrm{E}\left\{|h[k] - \mathbf{w}_k^{\mathrm{H}}\boldsymbol{h} - \mathbf{w}_k^{\mathrm{H}}\boldsymbol{n}|^2\right\}\\ &= \underbrace{\mathrm{E}\left\{|h[k] - \mathbf{w}_k^{\mathrm{H}}\boldsymbol{h}|^2\right\}}_{\sigma_i^2} + \underbrace{\mathrm{E}\left\{|\mathbf{w}_k^{\mathrm{H}}\boldsymbol{n}|^2\right\}}_{\sigma_n^2}.\end{aligned} \tag{6.17}$$

The interpolation term is determined by

$$\sigma_i^2 = \mathrm{E}\left\{|h[k] - \mathbf{w}_k^{\mathrm{H}}\boldsymbol{h}|^2\right\} \tag{6.18}$$

$$= \mathrm{E}\left\{|h[k]|^2\right\} - 2\,\mathrm{Re}\left\{\mathbf{w}_k^{\mathrm{H}}\boldsymbol{\theta}_{\boldsymbol{hh}}\right\} + \mathbf{w}_k^{\mathrm{H}}\boldsymbol{\theta}_{\boldsymbol{hh}}\mathbf{w}_k \tag{6.19}$$

and the noise part is given by

$$\sigma_n^2 = \mathrm{E}\left\{|\mathbf{w}_k^{\mathrm{H}}\boldsymbol{n}|^2\right\} = \frac{\mathbf{w}_k^{\mathrm{H}}\mathbf{w}_k}{\xi}. \tag{6.20}$$

6.3 Maximum Pilot Spacing for Noniterative PACE

The decomposition of the MSE into a noise term and an interpolation term is used in the following to determine a maximum spacing of pilot symbols for which channel estimation yields the best results given noisy observations. As said before (cf. Section 3.3.1), the Nyquist-Shannon theorem determines the maximum spacing of pilot symbols assuming an infinite sequence length. However, practical simulations can hardly reach this limit, neither is it simple to quantify. That is, a sequence generated by means of a Monte Carlo simulation will always be of finite length and due to numerical issues (e.g. rounding errors) an MSE larger than zero is typically attained. A semi-analytical bound is determined by the above mentioned MSE separation. More specifically, when the interpolation error is larger than the noise error, the coded sampling bound (CSB) is said to be violated and, hence, the pilot spacing is too large:

$$D_{\mathrm{s,CSB}} \doteq \max\left\{D_{\mathrm{s}} : \sigma_i^2 < \sigma_n^2\right\}. \tag{6.21}$$

In order to relate the chosen pilot spacing to the maximum pilot spacing given by the Nyquist-Shannon sampling theorem (NSB), a sampling ratio is defined as

$$\kappa = \frac{D_{\mathrm{t,max}}}{D_{\mathrm{t}}}. \tag{6.22}$$

When the channel response is sampled according to the Nyquist-Shannon theorem, then $\kappa = 1$. It is larger than one for oversampling and smaller than one for undersampling. For an iterative receiver which exploits the redundancy of pilot symbols as well as channel coding, typically $D_{\mathrm{t,CSB}} \geq D_{\mathrm{t,max}}$.

Unfortunately, the finite sequence length has a significant influence on the achievable performance, as can be seen in Figure 6.2. Hereby, the normalized maximum Doppler frequency is set to $f_{\mathrm{D,max}} = 0.02$, which results in a maximum pilot spacing of $D_{\mathrm{t,max}} = 25$. With the chosen pilot spacing $D_{\mathrm{t}} = 23$, the sampling ratio results in $\kappa \approx 1.09$. A minimum

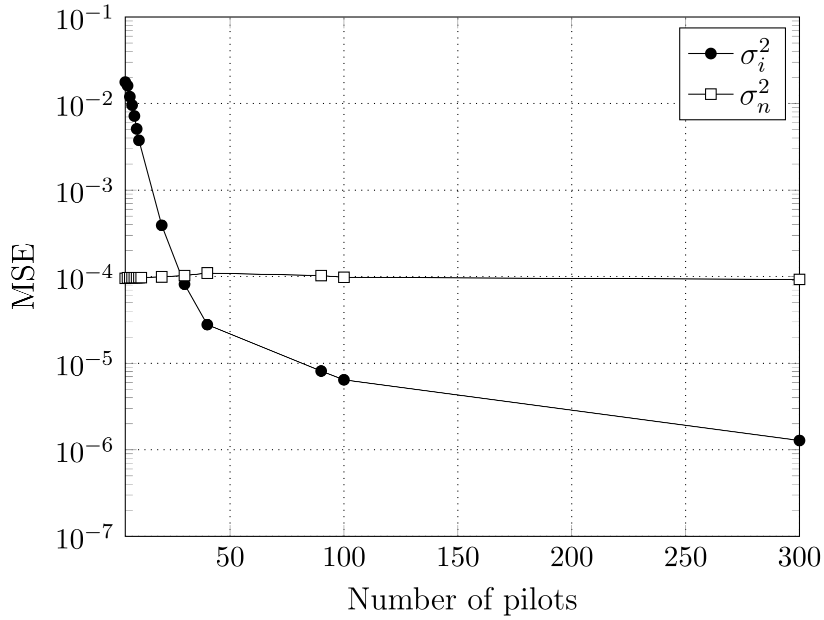
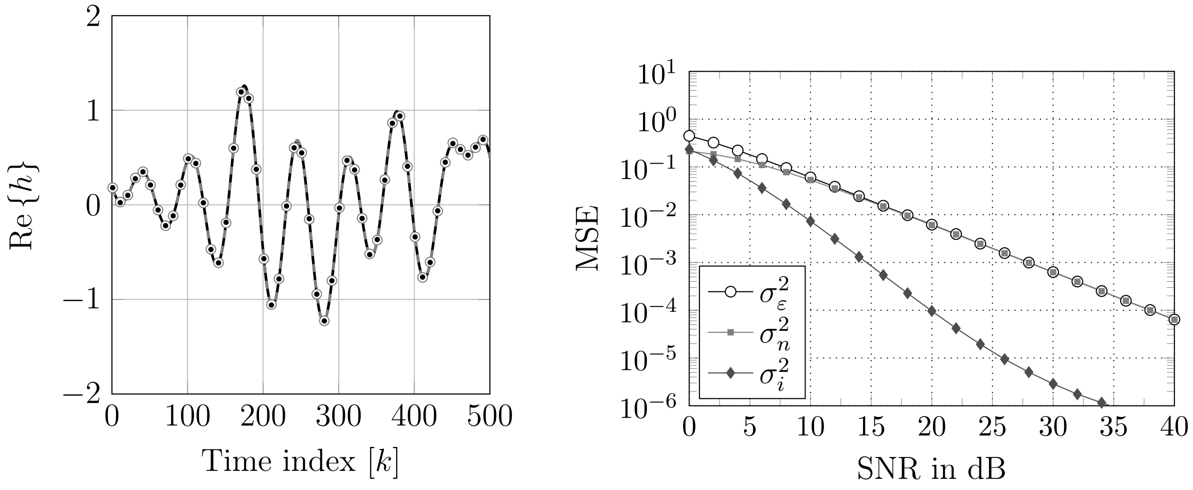


Figure 6.2: MSE performance as a function of the number of pilots at an SNR of 40 dB and a sampling ratio of $\kappa \approx 1.09$.

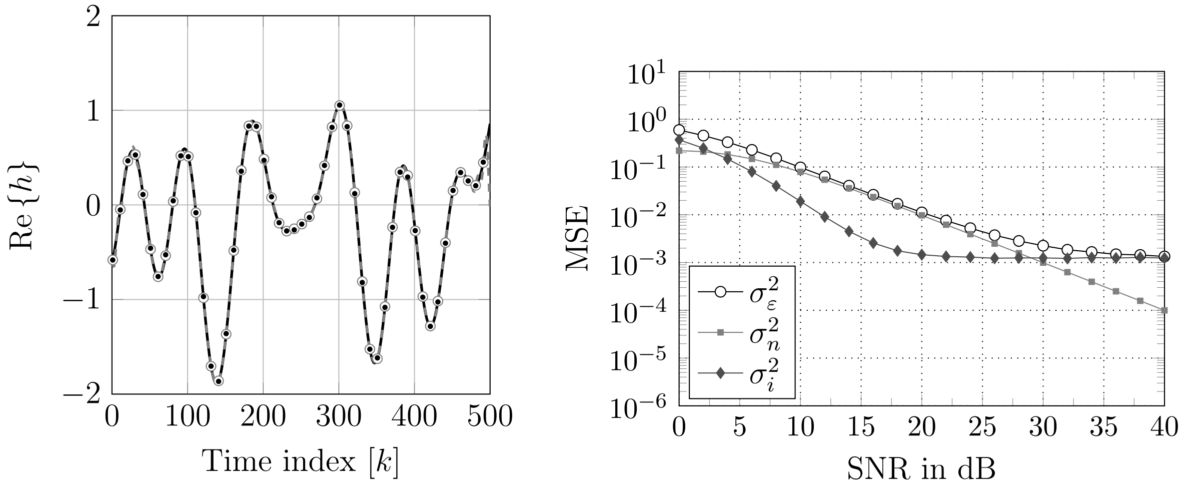
of approximately 30 pilot symbols is required to achieve an interpolation error lower than the noise error. With a short sequence length, edge effects are more pronounced. Hereby, edge effects refer to the fact that at the beginning and the end of a sequence the quality of the estimates degrade. For a joint receiver, edge effects of channel estimation also affect the achievable BER since every channel coefficient is used to detect the data symbols and can not be ignored for the MSE calculation.

In Figure 6.3, the filtered channel response is shown as a function of pilot spacing D_t. The normalized maximum Doppler frequency is set to $f_{\mathrm{D,max}}T_\mathrm{s} = 0.1$ and the number of pilot symbols is set to $N_\mathrm{P} = 100$. According to the Nyquist-Shannon sampling theorem, the maximum allowed spacing is $D_{\mathrm{t,max}} = 1/(2f_{\mathrm{D,max}}T_\mathrm{s}) = 5$. On the left side of Figure 6.3, the real part of the true coefficients is shown in black filled circles whereas the estimated coefficient is depicted in dashed lines with white circles. Every tenth symbol is accentuated with a marker. As can be seen, from Figure 6.3a, the channel response is reconstructed very well. The corresponding MSE, given in Figure 6.3b, shows that the interpolation error is below the noise error for an SNR above 2 dB. With increasing spacing of pilots, the interpolation error increases as well and slight imperfections of the reconstructed channel response can be seen at the edges in Figure 6.3c.

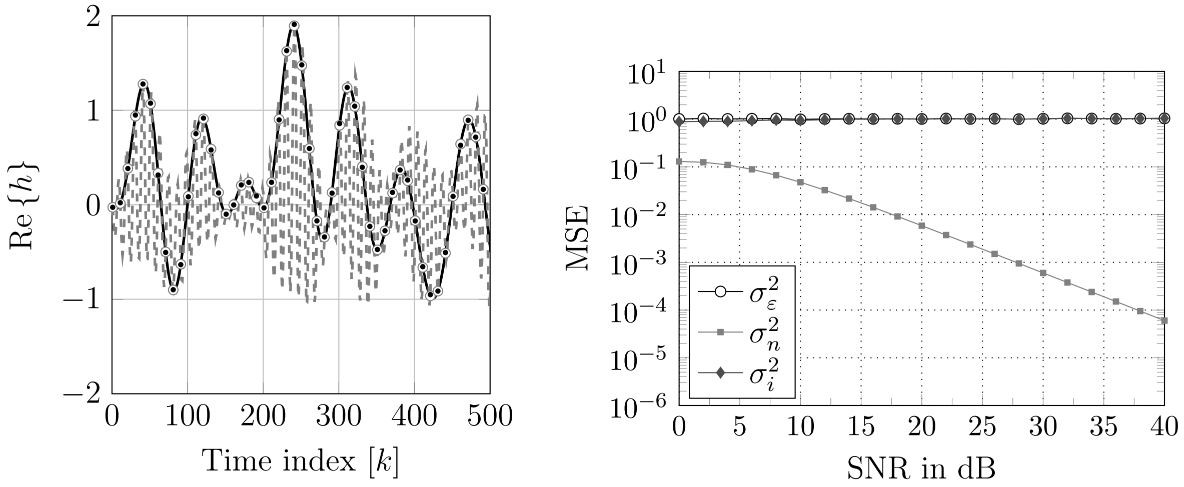
Although the Nyquist-Shannon sampling theorem is just fulfilled, the CSB is violated, since $\sigma_i^2 > \sigma_n^2$ at SNR < 2.5 dB and SNR > 28 dB, as can be seen in Figure 6.3d. This is partly due to the finite sequence length. With a longer sequence the error floor can be reduced as interpolation errors occur mainly at the edges. However, it cannot be completely removed due to the restrictions of a simulation and hence, the interpolation error will always be larger than the noise error for a sufficiently high SNR. In the remainder of this chapter the SNR is set to 40 dB. Once the sampling theorem is significantly violated,



(a) $D_\mathrm{t} = 3$, SNR=40 dB (b) $D_\mathrm{t} = 3$



(c) $D_\mathrm{t} = 5$, SNR=40 dB (d) $D_\mathrm{t} = 5$



(e) $D_\mathrm{t} = 10$, SNR=40 dB (f) $D_\mathrm{t} = 10$

Figure 6.3: Filtering of the channel impulse response as a function of training spacing with $f_{\mathrm{D,max}}T_\mathrm{s} = 0.1$ and $D_{\mathrm{t,max}} = 5$.

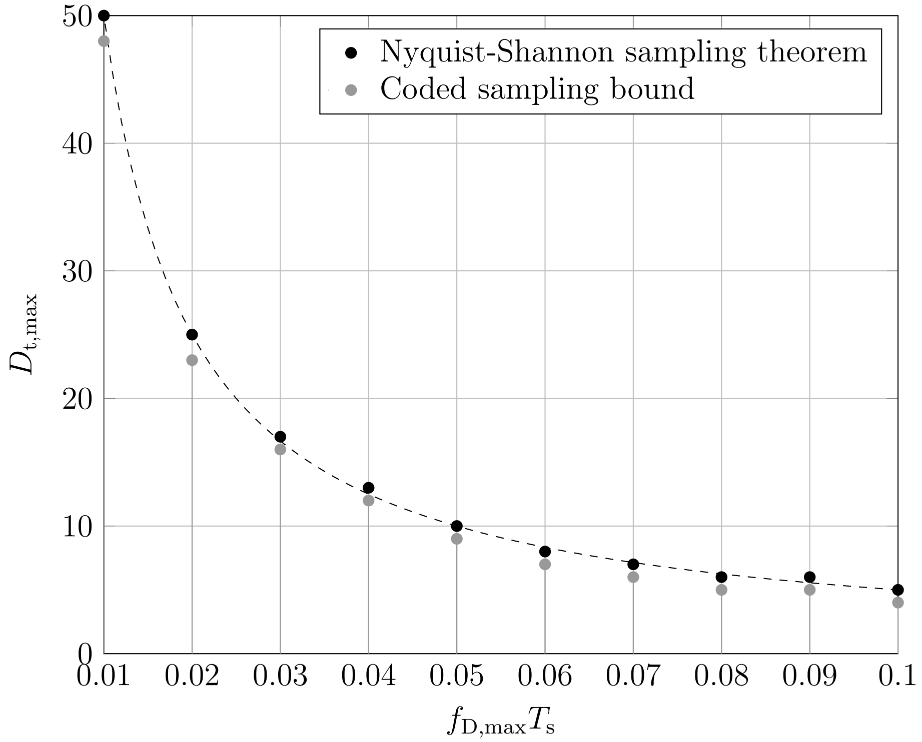


Figure 6.4: Maximum pilot spacing determined by the CSB and the Nyquist-Shannon sampling theorem for PACE.

the channel response can not be reconstructed by means of PACE, as can be seen from the oscillating behavior of the estimated response shown in Figure 6.3e. The MSE is dominated by the interpolation error for the complete SNR range (cf. Figure 6.3f).

The maximum pilot spacing determined according to the Nyquist-Shannon bound is compared with the coded sampling bound in Figure 6.4. As can be seen, the CSB closely resembles the strict bound given by the sampling theorem and deviates by roughly one symbol. The separation of interpolation error and noise error is thus a useful tool for the analysis of a lower bound for iterative semi-blind channel estimation.

6.4 Maximum Pilot Spacing for Iterative SBCE

Motivated by the results of the previous section, the coded sampling bound for iterative channel estimation is determined by means of an SNR analysis in Section 6.4.1. Additionally, the obtained results are evaluated with an EXIT chart analysis in Section 6.4.2. Based on the obtained results, implications for an iterative receiver design are drawn in Section 6.4.3.

6.4.1 MSE Analysis

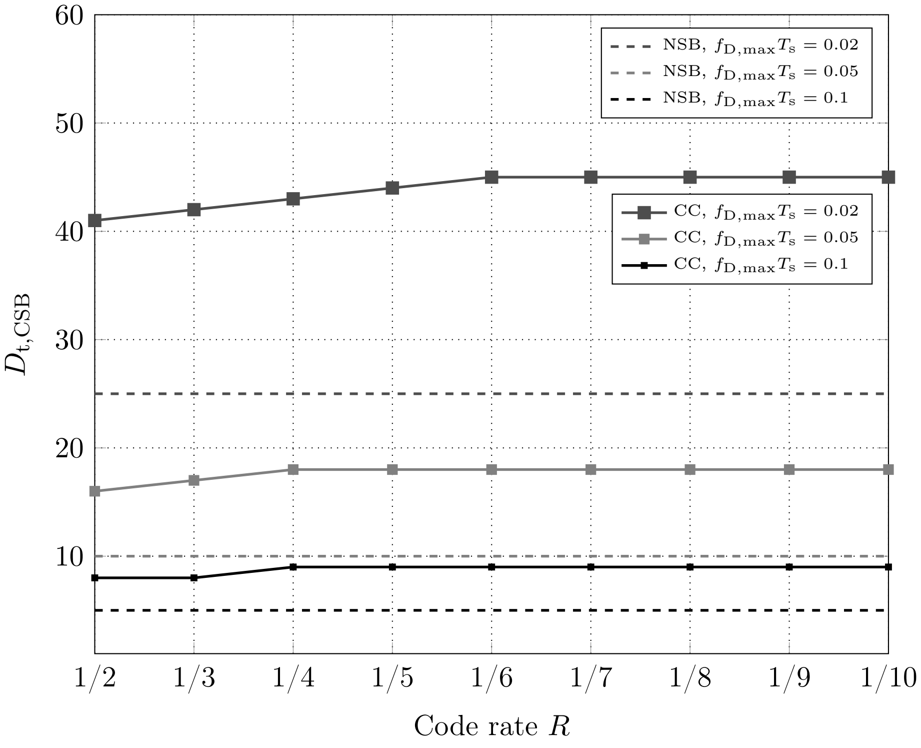
For an iterative receiver, the definition for the CSB given in (6.21) has to be extended to take possible phase inversions (cycle slips) into account. Given a sufficiently large pilot spacing, the probability of a phase inversion increases rapidly, which means that the available pilot symbols can not ensure that the correct phase is assigned to the detected data symbols and/or the estimated channel coefficients. In the event that the complete sequence is inverted, the obtained filter coefficients are equal to the non-inverted case and thus, the MSE analysis fails to detect the insufficient pilot spacing. Nevertheless, phase inversion can easily be tracked by incorporating an additional constraint to the CSB, which is that the sum of the interpolation and noise error has to be equal to the overall estimation error. Hence, the CSB is also violated if

$$\sigma_n^2 + \sigma_i^2 \not\approx \sigma_\varepsilon^2. \tag{6.23}$$

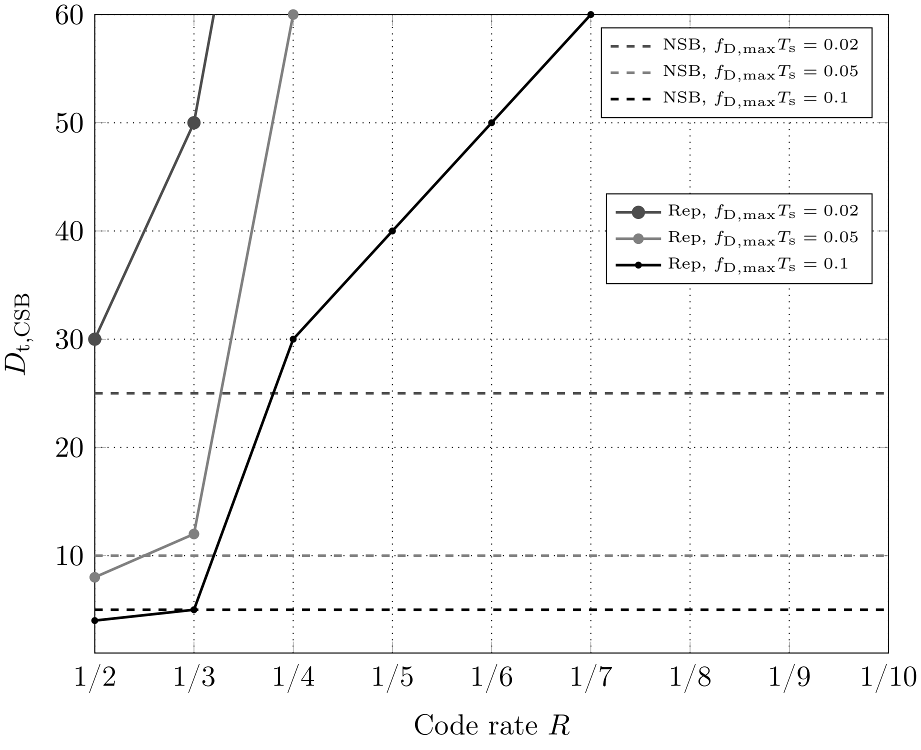
Due to rounding errors, the sum of the two terms may not be exactly equal to the MSE. However, in case of a phase inversion the two terms differ significantly and can thus easily be identified. Unless a channel code with spatial structure (such as an asymmetric code [WLH08]) is applied, a phase inverted sequence results in outage.

The maximum pilot spacing determined by (6.21) and (6.23) by means of Monte-Carlo simulations is shown in Figure 6.5 as a function of the code rate. As before, a sequence length was chosen to accommodate at least $N_\mathrm{p} = 100$ pilot symbols. Simulations were conducted at an SNR of 40 dB and repeated 10000 times for each pilot spacing. The bounds given by the Nyquist-Shannon sampling theorem (NS-ST) are depicted with dashed lines for the normalized maximum Doppler frequency. Examples are plotted for three values, i.e. $f_\mathrm{D,max}T_\mathrm{s} = [0.1, 0.05, 0.02]$ with a resulting maximum spacing of $D_\mathrm{t,max} = [5, 10, 25]$. Two channel codes are considered, namely a convolutional code (CC) and a repetition code (Rep). The generator polynomials used within the simulations are given in Table 6.1. While the first provides redundancy by means of parity information, the latter is using repetitions. The two codes differ significantly in the maximum supported pilot spacings. As can be seen in Figure 6.5a, the maximum spacing in combination with a convolutional code increases linearly with decreasing code rate until a threshold is reached and an upper bound—the coded sampling bound— is reached. The results obtained for a turbo are similar to the results of the convolutional code but are omitted to improve readability.

The maximum pilot spacing of a system with a convolutional code approaches the CSB faster as a function of the code rate given fast fading. The repetition code, on the other hand, requires a sufficiently low code rate to support a pilot spacing close the sampling theorem. At a code rate of $R = 1/4$, the pilot spacing $D_\mathrm{t,CSB}$ suddenly increases and continues to extend linearly with decreasing code rate. The differences of the two codes are not surprising. The optimum result in terms of MSE is attained with a perfectly reconstructed data sequence, since every symbol (pilot symbols and data) of the transmit sequence can be used to estimate the channel. With a strong channel code, e.g. a convolutional code, data symbols are more reliably detected compared to a repetition code with the same code rate. Apparently, at a code rate around $R = 1/2$, the convolutional code supports higher pilot spacings than the repetition code. Once



(a) Convolutional code (CC)



(b) Repetition code (Rep)

Figure 6.5: Coded sampling bound for iterative channel estimation for a convolutional code (CC) and a repetition code (Rep) for different Doppler frequencies as a function of code rate.

data symbols are reliably detected and can be used as pseudo-pilot symbols, the channel impulse response is in fact oversampled. It is of importance to remember that the Nyquist-Shannon sampling theorem is actually not violated in general, but rather are the initial constraints for the pilot symbols relaxed to the extent that convergence of the iterative receiver is enabled. Furthermore, a repetition code with code rate of $R = 1/2$ represents a special case: Extrinsic information is only exchanged for the two code bits. Hence, in case of hard detection, errors can not be corrected. Although soft values are able to identify and correct detection errors, fair and unbiased log likelihood ratios are necessary. In the presence of channel estimation errors, this prerequisite is not necessarily fulfilled and hence, a sufficiently close pilot spacing has to be provided. The reasons why the convolutional code saturates at a certain level while the repetition code allows increasing pilot spacings is examined in the following section in more detail.

Two main conclusions can be drawn from these results: First, an iterative semi-blind receiver is not strictly bound to the pilot spacing given by the Nyquist-Shannon sampling theorem. In fact, the initial pilot spacing can be extended by nearly a factor of two for the channel codes under consideration. And second, given a suitable channel code and code rate, an arbitrary pilot spacing is supported. By comparing the linear parts obtained from the two channel codes at low code rates, a linear regression is performed. The resulting curves are shown in Figure 6.6. Obviously, they have a common starting point. It has to be mentioned that a code rate of $R = 1$ does not refer to the uncoded case, since there the pilot spacing is strictly limited by the sampling theorem. While the function is constant over the code rate for convolutional codes, it increases linearly for repetition codes. Given a sufficiently low code rate, an arbitrary pilot spacing is supported with repetition codes, as mentioned previously. The simulation results are slightly below these curves, which can be explained by the finite sequence length and is similarly observed for the non-iterative case. However, with an increasing number of pilot symbols the CSB is asymptotically approached. Revisiting the Nyquist-Shannon sampling theorem given by (3.35c)

$$D_{\mathrm{t,max}} < \frac{1}{2 f_{\mathrm{D,max}} T_{\mathrm{s}}},$$

Table 6.1: Tabulated generator polynomials taken from [Bos98].

R	$O_G^{(1)}$	$O_G^{(2)}$	$O_G^{(3)}$	$O_G^{(4)}$	$O_G^{(5)}$	$O_G^{(6)}$	$O_G^{(7)}$	$O_G^{(8)}$	$O_G^{(9)}$	$O_G^{(10)}$
1/2	5	7								
1/3	25	33	37							
1/4	25	27	33	37						
1/5	25	27	33	35	37					
1/6	25	27	33	35	35	37				
1/7	25	27	27	33	35	35	37			
1/8	25	25	27	33	33	35	37	37		
1/9	25	25	27	33	33	35	35	37	37	
1/10	25	25	25	33	33	33	35	37	37	37

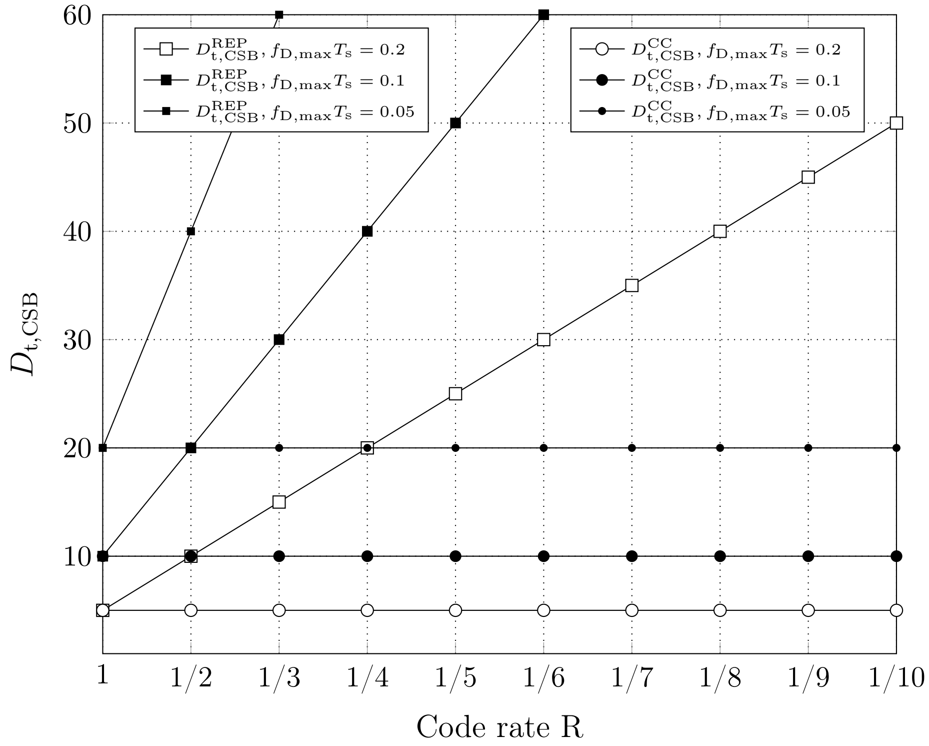


Figure 6.6: CSB obtained by the semi-analytical MSE analysis.

a similar functional relation can be inferred for the CSB from the obtained simulation results. Depending on the channel code, two functions are obtained: First, for convolutional codes

$$D_{\mathrm{t,CSB}}^{\mathrm{CC}} < \frac{1}{f_{\mathrm{D,max}}T_{\mathrm{s}}} = 2 \cdot D_{\mathrm{t,max}} \tag{6.24}$$

and second, for repetition codes

$$D_{\mathrm{t,CSB}}^{\mathrm{REP}} < \frac{\mathrm{SF}}{f_{\mathrm{D,max}}T_{\mathrm{s}}} = 2 \cdot \mathrm{SF} \cdot D_{\mathrm{t,max}}, \tag{6.25}$$

where SF refers to the spreading factor $\mathrm{SF}{=}\mathrm{R}^{-1}$ of a repetition code. The obtained equations are specific for the chosen receiver and system setup, but can be interpreted as lower bounds. Exemplary, the results for a convolutional code as well as for a repetition code with various spreading factors are given in Figure 6.7.

To this extent, the CSB is evaluated for a theoretical system, which ignores noise as well as the demands of spectrally efficient systems. Although a large pilot spacing increases the spectral efficiency, a low code rate is required to support the chosen spacing which deteriorates the spectral efficiency again. It is therefore of interest to provide an additional insight of the supported pilot spacings also for high code rates. Typically, the convergence behavior of coded systems is conveniently evaluated by means of an EXIT chart analysis. Although the effects of channel estimation are typically ignored, the incorporation of an estimated channel is straightforward.

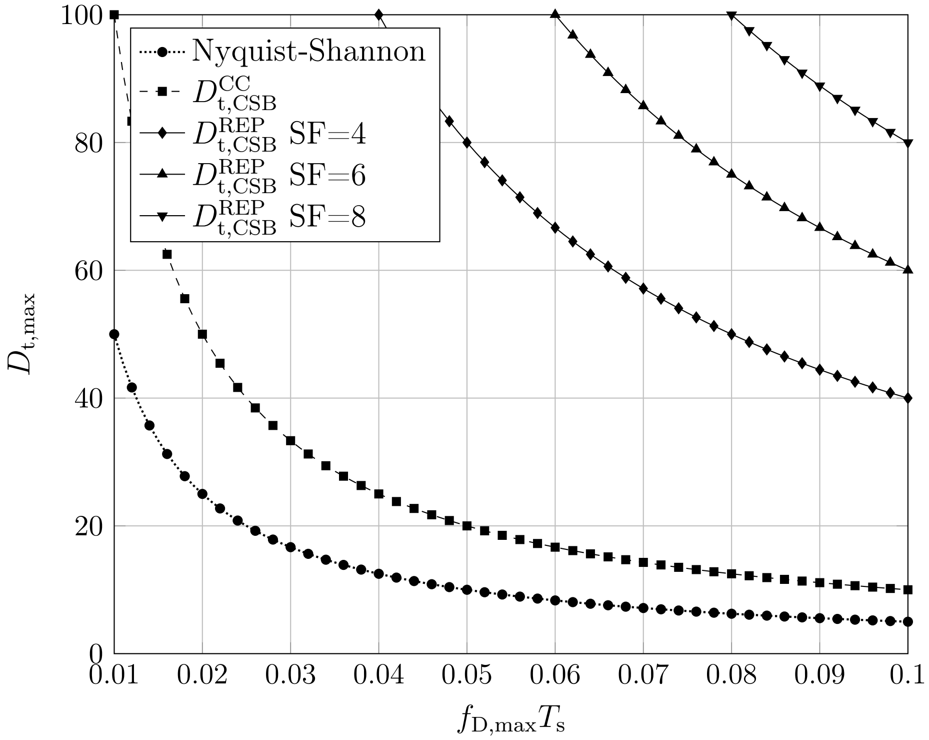


Figure 6.7: Maximum pilot spacing determined by the CSB for a convolutional code and repetition code with varying spreading factors for SBCE as a function of maximum normalized Doppler frequency.

6.4.2 EXIT Chart Analysis

Predicting the performance of an iterative receiver in the presence of channel estimation errors is, in general, challenging. Often, tools like EXIT charts [tB00b, tB01], density evolution [RSU01], and variance transfer charts [AGR98] are used. However, their accuracy varies and rather give an impression of the expected performance. Furthermore, the focus of these tools is to assess the performance of data detection instead of channel estimation, which is of interest in this chapter. Additionally, bit error rates do not necessarily reflect the channel estimation performance. Especially for lower-order modulation, such as BPSK, a rough estimation of the channel is sufficient for which error-free data detection is enabled. Nevertheless, EXIT charts are widely and successfully applied for the analysis of iterative semi-blind receivers, e.g. in [tBSS00, SJS03] as well as in Section 5.4.1. Accordingly, the applicability of EXIT charts to determine the maximum spacing of pilots is examined in the following.

Typically, the transfer characteristic of a soft-input soft-output detector is increasing with improved a priori information. An unexpected result in combination with channel estimation is, that the curve may decrease with poor a priori information at first and increase later with sufficiently good a priori information. As can be seen in Figure 6.8, for $I_{\mathrm{A}} = 0$ the extrinsic mutual information I_{E} is larger than the remaining values of I_{A}. This means that having no a priori information is better than having poor informa-

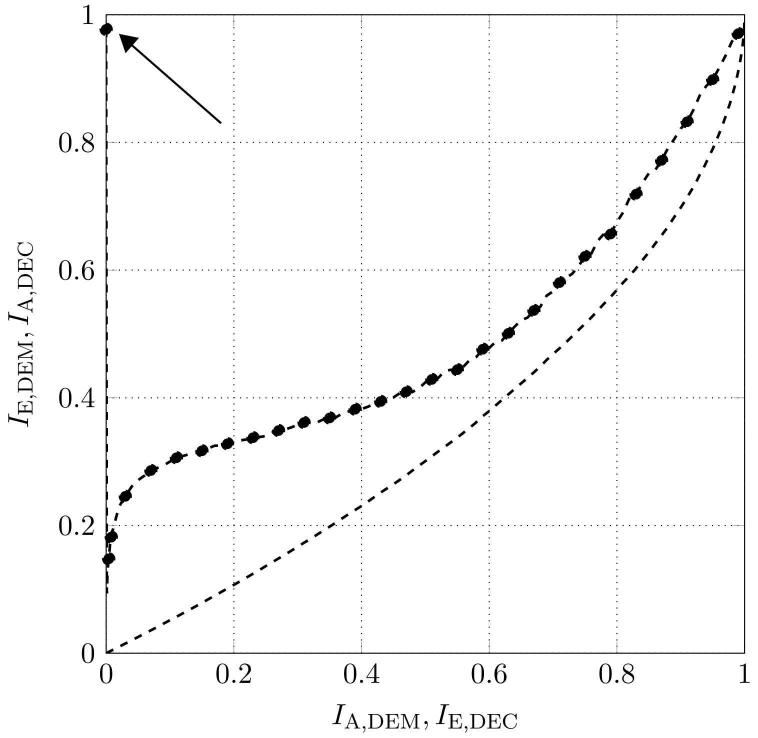
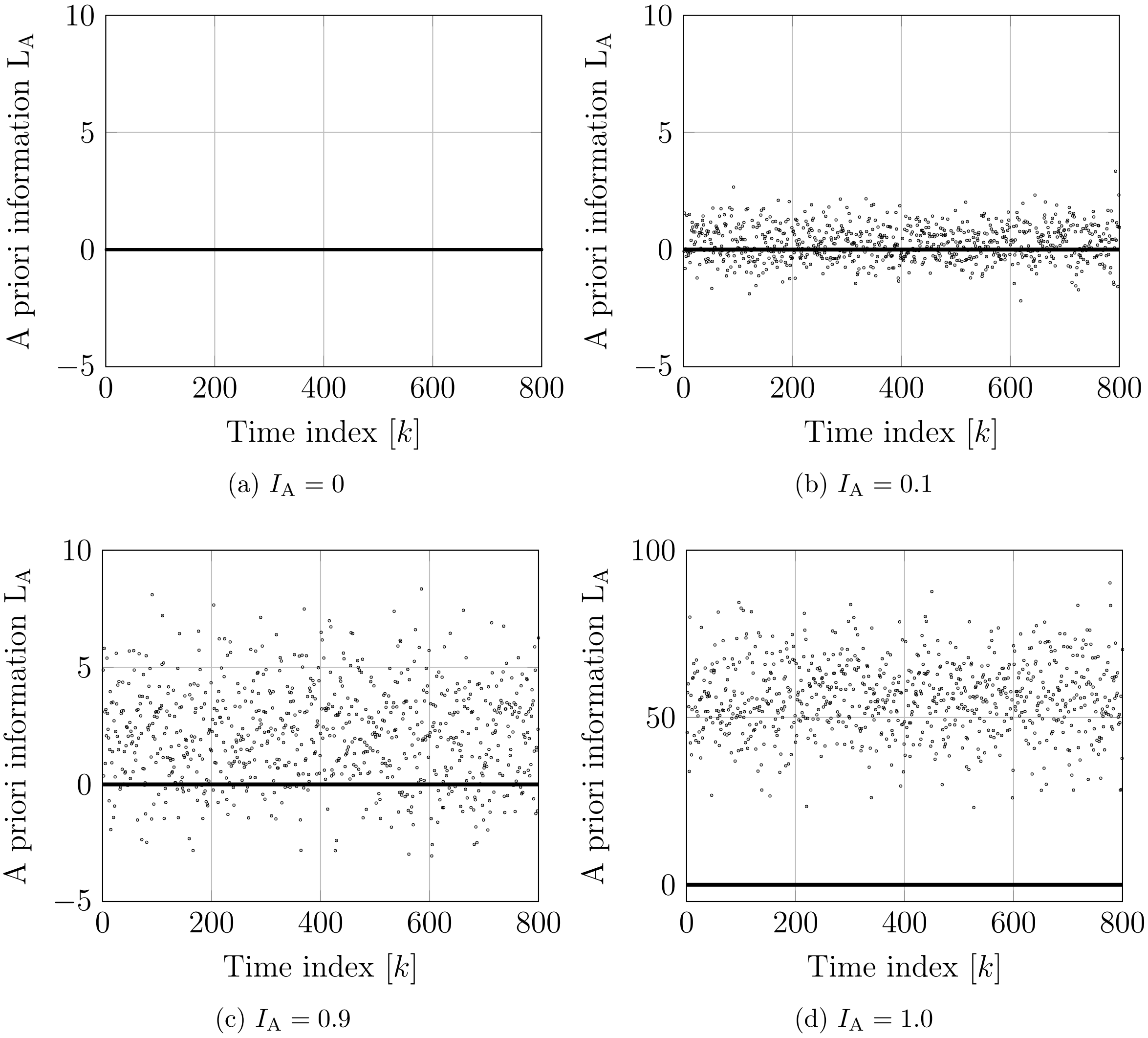


Figure 6.8: EXIT chart simulation of the iterative receiver as a function of a priori mutual information for a sampling ratio of $\kappa = 1$ and an SNR of 40 dB. Additionally, the transfer characteristic of a rate-1/3 repetition code is included.

tion [SSAR07]. In order to explain this effect in more detail, the L-values $\mathrm{L_A}$ generated for a vector of ones are depicted in Figure 6.9. Hence, a positive L-value represents a correct estimate while a negative L-value corresponds to a wrong decision. The reliability of a decision is given by the magnitude. A higher magnitude should reflect a more reliable estimate and vice versa. The idea of soft values is that wrong decisions can be made as long as they are identified to be unreliable, i.e. the magnitude is small. As can be seen, this is not the case, as also L-values with similar magnitude exist on both sides. Obviously, having no a priori information results in a priori L-values which are all zero. Channel estimation can only rely on pilot symbols, however, all information is correct and can be trusted. With an a priori information of $I_{\mathrm{A}} = 0.1$, shown in Figure 6.9b, a large amount of wrong a priori information is introduced. Unfortunately, the magnitude of negative L-values is nearly as large as for positive L-values. Accordingly, the estimated sequence will be heavily distorted. With increasing a priori information, the amount of wrong a priori information is reduced, as can be seen in Figure 6.9c. Only a few negative L-values exist and their magnitude is much smaller compared to that of the positive L-values. Having perfect a priori information, $I_{\mathrm{A}} = 1$, the complete transmitted sequence provides reliable information, cf. Figure 6.9d. During iterations between the detector and the decoder, the resulting trajectory moves between the inner and outer curve, i.e. the transfer characteristic of the detector and the decoder, respectively. If the two transfer characteristics do not intersect, the receiver is able to converge. The area of the resulting



(a) $I_A = 0$

(b) $I_A = 0.1$

(c) $I_A = 0.9$

(d) $I_A = 1.0$

Figure 6.9: A priori L-values as a function of a priori mutual information I_A.

convergence tunnel hereby refers to the capacity loss of the receiver [AKt04]. Hence, a common design goal is to reduce the area as much as possible while the remaining gap has to be sufficiently large. In general, non-iterative receivers exhibit a flat characteristic since they do not consider any a priori information. On the other hand, the transfer characteristic of an iterative receiver has a slope, as can be seen in Figure 6.8. Moreover, the initial starting point without a priori information ($I_A = 0$) will be lowered as a function of the pilot spacing. In Figure 6.10, the initial value of the receivers transfer characteristic is plotted as a function of the sampling ratio. Hereby, the region of interest is between a sampling ratios of $\kappa = 0.5$ and $\kappa = 0.6$. As can be seen, the output extrinsic information I_E is close to 1 for a sampling ratio larger than 0.6. Although the sampling theorem is disobeyed for the pilot spacings, the received sequence can be reconstructed nearly perfect. This is not surprising, since channel estimation errors are not necessarily reflected in BER performance. Moreover, BPSK is rather robust against channel estimation errors.

Obviously, these curves are shifted to higher sampling ratios for higher order modulation formats. By further reducing the sampling ratio, it can be seen, that once a certain threshold is surpassed, the output characteristic drops exponentially with decreasing sampling ratio. Below a sampling ratio of $\kappa = 0.6$, the transfer characteristic of a channel code has a significant influence on the convergence behavior of the overall receiver. Exemplary the transfer characteristic for repetition and convolutional codes are given in Figure 6.11 for selected code rates. By comparing the initial values of the receiver with that of the channel codes, the advantage of the repetition code is clearly visible. Already at a low values of $I_{\mathrm{A,DEC}}$, the output of the decoder is improved ($I_{\mathrm{E,DEC}} > I_{\mathrm{A,DEC}}$). Hence, for a wide range of code rates the convergence tunnel is open. However, the resulting area and thus the capacity loss is large. The convolutional code, on the other hand, requires a sufficiently large a priori information in order to improve its output. Therefore, only low code rates are able to open the convergence tunnel. As before, the resulting loss of capacity is large.

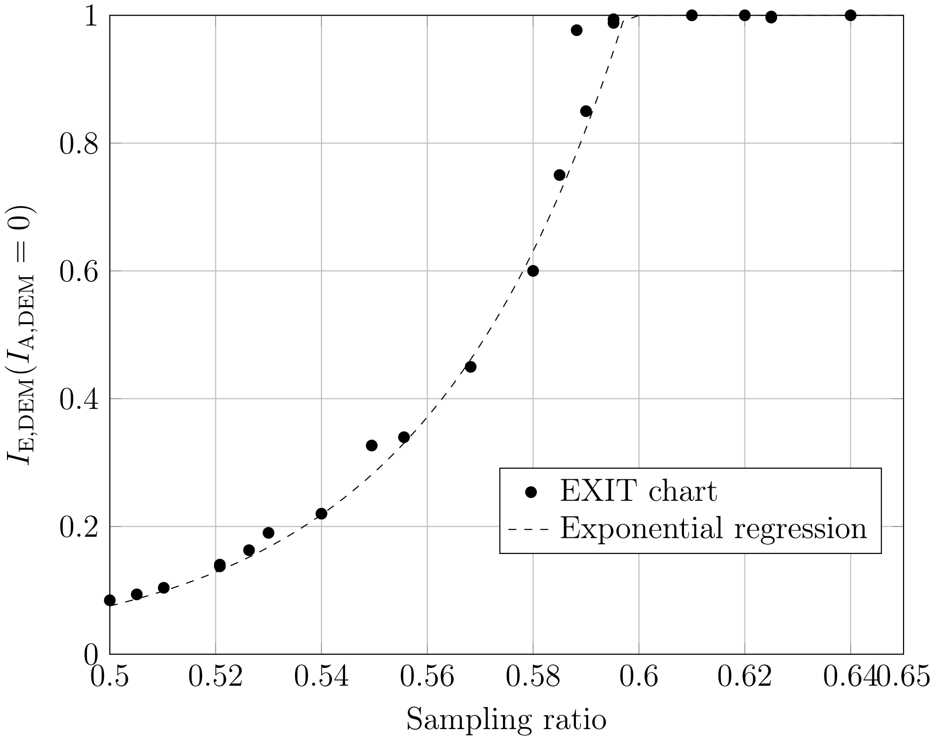
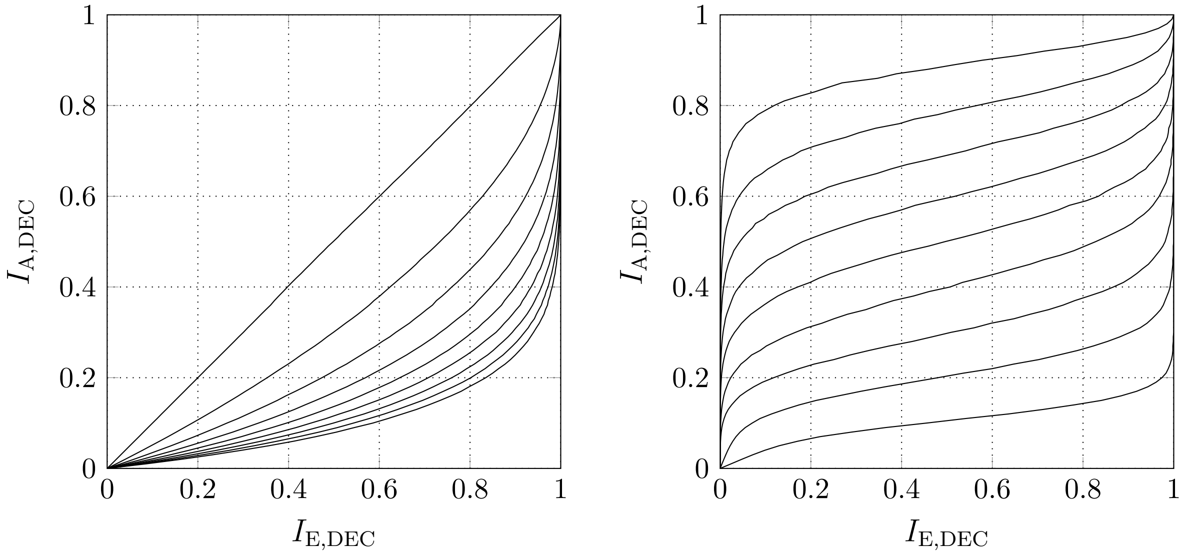


Figure 6.10: Extrinsic mutual information output with no a priori information available as a function of sampling ratio.

The dominance of convolutional, turbo, and LDPC-codes, which all have a rather low slope, indicate that the majority of wireless systems is designed for non-iterative receivers. In order to exploit the possible gains of a joint channel estimation and data detection as well as the reduction of pilot overhead, carefully designed channel codes are needed which closely resemble the transfer characteristic of an iterative receiver. Examples for a corresponding channel code design are for example given in [TH04] where irregular convolutional codes are serially concatenated to match the transfer characteristic of the



(a) Repetition code with code rates ranging from 0.1 to 0.5 in steps of 0.1

(b) Convolutional code with code rates ranging from 0.1 to 0.9 in steps of 0.1

Figure 6.11: Transfer characteristics of a repetition and a convolutional code.

receiver.

6.4.3 Implications for Receiver Design

The implications of the coded sampling bound on receiver design are two-fold and depends whether an adapted channel code (matched to the transfer characteristic of the receiver) is used or existing codes, like convolutional, Turbo and/or LDPC codes. The former requires new analysis tools which take the impact of imperfect channel estimation into account and provide accurate insight of the behavior of a semi-blind receiver. The latter aims to change the paradigms of pilot grid design, which is mostly done for non-iterative receivers.

Analysis tools, such as EXIT charts, density evolution, etc., are currently not able to accurately model the effects of channel estimation errors. Although multi-dimensional EXIT charts exist which separate the a priori information for data detection and channel estimation [HRRE09], they are not able to model the a priori information of estimated channel coefficients directly. As a consequence, the predicted convergence behavior is not precise. This effect is exacerbated when channel estimation performance is very poor, e.g. for very large pilot spacings. Thus, an optimized code design for semi-blind iterative receivers can yield substantial gains in terms of improved bandwidth efficiencies without performance impairments, as indicated by the results with repetition codes.

For the design of a pilot grid, certain design parameters are chosen under which a receiver should be able to recover the transmitted sequence. LTE as well as LTE-A are designed for high mobility scenarios, i.e. a maximum velocity of 500 km/h is supported at a carrier frequency of $f_c = 1.8$ GHz [STB11]. Hence, a maximum pilot spacing of

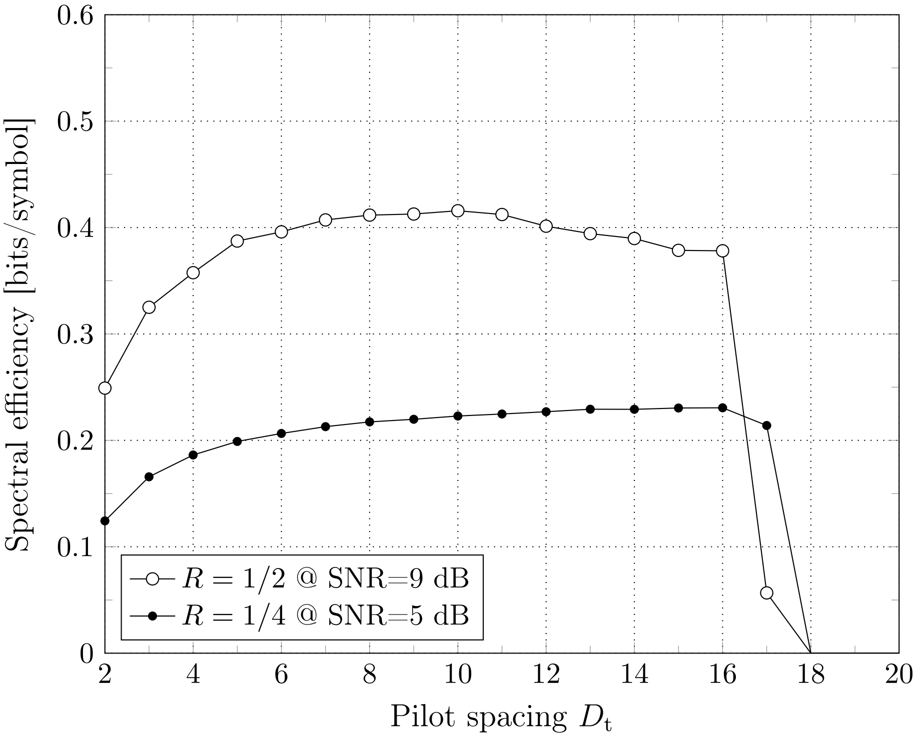


Figure 6.12: Spectral efficiency as a function of pilot spacing D_t for different convolutional code rates.

$D_\mathrm{t,max} = 1/(2f_\mathrm{D,max}T_\mathrm{s}) = 1/0.119 = 8.4$ should be obeyed. Comparing the maximum pilot spacing with the pilot grid of LTE-A (shown in Figure 5.24a), it can be seen that the maximum spacing of pilots is set to $D_\mathrm{t} = 7$, which results in a sampling ratio of $\kappa = 1.2$. Given an iterative semi-blind receiver, the pilot spacing can be reduced by 50% to a sampling ratio of $\kappa = 0.6$ with $D_\mathrm{t} = 14$ while still ensuring convergence. Keeping in mind that 500 km/h is defined as a worst-case condition and that the fastest train in the world currently reaches a maximum velocity of 420 km/h, a reduction of pilot overhead would have a negligible performance impact for current users. New trains which exceed 500 km/h are expected in the year 2027 [The13]. Hence, without additional changes and only minor performance impairments, the efficiency is improved.

The following example illustrates the practical application of the coded sampling bound to an iterative receiver operating at the lower SNR regime. For this example, a fixed sequence length of $K_\mathrm{s} = 320$ symbols and 5 iterations for the iterative receiver are chosen. Hereby, one iteration is applied for data detection and channel estimation as well as channel decoding. A rate-1/2 as well as a rate-1/4 convolutional code is applied, whereas the normalized maximum Doppler frequency is assumed to be $f_\mathrm{D,max}T_\mathrm{s} = 0.05$. Hence, the maximum pilot spacing according to the Nyquist-Shannon sampling theorem is $D_\mathrm{t,max} = 10$. The results for the repetition code are omitted here since it fails to converge for the chosen parameters. More specifically, a higher SNR is required in comparison to a convolutional code. The evaluation of the coded sampling bound for the two setups yield $D_\mathrm{t,CSB} = 14$ for the rate-1/2 code at an SNR of $\mathrm{E_d/N_0} = 9$ dB and $D_\mathrm{t,CSB} = 12$

for the rate-1/4 code rate at an SNR of $\mathrm{E_d/N_0} = 5$ dB. Without loss of generality, the corresponding SNR operating points are chosen to be within the waterfall region, i.e. the ideal case of a perfectly reconstructed data sequence is not reached. In order to illustrate the link between pilot overhead and bit error ratio, the spectral efficiency is shown in Figure 6.12. Hereby, the spectral efficiency given by (5.64) is revisited:

$$\eta = R \cdot \eta_{\mathrm{RS}} \cdot (1 - \mathrm{BLER}),$$

where η_{RS} refers to the pilot overhead defined as $\eta_{\mathrm{RS}} = N_{\mathrm{P}}/K_{\mathrm{s}}$ and BLER denotes the block error rate. As can be seen from the results, once a certain threshold in terms of MSE is surpassed, the transmitted sequence can not be reconstructed and the spectral efficiency quickly deteriorates. Interestingly, the pilot spacing beyond which the spectral efficiency drops to zero does not coincide with the coded sampling bound. This is reasoned in the choice of the modulation format, which is insensitive to estimation errors. For higher-order modulation schemes, the pilot spacing as determined by the coded sampling bound and the pilot spacing for which the spectral efficiency strongly deteriorates will converge to the same point.

Moreover, it can be observed that the spectral efficiency of the rate-1/2 encoded system slowly deteriorates when the pilot spacing approaches $D_{\mathrm{t,CSB}}$. This effect can be observed as well for the Nyquist-Shannon sampling theorem, as both bounds state the maximum pilot spacing opposed to the optimum amount of pilot symbols. Accordingly, oversampling is recommended for both non-iterative as well as iterative receivers.

6.5 Chapter summary

Among the first components of a wireless receiver, channel codes have been implemented in an iterative fashion in order to yield substantial performance gains. Similarly, it is well known that iterative receivers promise significant gains in terms of BER and MSE performance at often reduced computational complexity. Moreover, by changing the modulation format, to e.g. superposition modulation [Hoe13], a capacity gain compared to conventional modulation formats such as QAM is yield, but it requires an iterative demodulator to reach that gain. Apparently, the research community identified several components which can improve the performance and/or efficiency of a system by means of iterative processing. Yet, this insight had no influence on the design of pilot grids. So far, no publication exists which investigates the possible reduction of pilot overhead in combination with an iterative semi-blind receiver.

In this chapter, a coded sampling bound is formulated, which states the maximum spacing of pilots up to which an iterative coded receiver is able to reconstruct the estimated sequence. By separating the MSE of a linear Wiener filter, into a noise and an interpolation part, the maximum pilot spacing is identified. With a suitable channel code and code rate, an arbitrary pilot spacing is supported. An additional EXIT chart analysis revealed that the pilot spacing can be extended by a factor of 1.6 compared to the Nyquist-Shannon sampling theorem, independent of the channel code and code rate.

7

Summary and Conclusions

FOR THE DESIGN of a wireless system, various, partly contradictory requirements have to be balanced. A high spectral efficiency of the system is desired while simultaneously the computational complexity of the receiver as well as the transmit power should be rather low. Given the current growth of mobile data traffic, the spectral efficiency of current systems has to be increased by a factor of 1000. Several key technologies, such as MIMO, OFDM, smart antennas, etc., have been identified to meet these challenging goals. A detailed overview of these technologies is given in Chapter 2.

Concurrently, advances in signal processing revealed several approaches to actually approach the promised gains of MIMO-OFDM systems. However, optimum detection is considered to be computationally infeasible and thus, sub-optimum solutions are needed. Iterative joint data detection and channel estimation is a viable solution to balance the trade-off between performance and complexity. The design, evaluation, and extension of a multi-dimensional graph-based soft iterative receiver is the core part of this thesis.

Contributions of the Thesis

The literature on MIMO-OFDM receivers is extensive and a large variety of detection and estimation algorithms exists. Since optimum detection with an APP detector is computationally complex, suboptimal MIMO detection approaches are needed. One class of suboptimal detectors is particularly popular, i.e. tree-search algorithms based on the QR decomposition such as the sphere detector and/or the QRD-M detector. However, due to the chosen factor graph structure, QR decomposition cannot be directly implemented within the MD-GSIR. Furthermore, it is typically not flexible enough to support arbitrary MIMO configurations, such as massive MIMO. More specifically, QR decomposition requires the number of transmit antennas to be lower or equal to the number of receive antennas: $N_\mathrm{T} \leq N_\mathrm{R}$. Two alternative MIMO detectors suitable for iterative processing are presented in Chapter 3: The Gaussian detector and the Gaussian tree search detector. The former approximates the multi-antenna interference by a Gaussian random variable while the latter combines the Gaussian approximation with tree-based detection. The complexity as well as the performance of the Gaussian tree search depends on the number of significant leaves M_sl, that is the amount of leaves which are kept alive during the traversal of the tree. By evaluating M_sl hypotheses in parallel, the poor performance of the Gaussian detector for higher-order modulation is mitigated. Thereby, the overall number of evaluations is significantly smaller compared to the APP detector. It is shown that GTS detection is able to approach close-to-optimum performance at lower complexity. The gain w.r.t. to complexity improvement is largest for a large number of transmit antennas and higher-order modulation.

In the second part of Chapter 3, pilot-based channel estimation is evaluated in terms of estimation algorithms and pilot allocation schemes. The Nyquist-Shannon sampling theorem is introduced, which states the maximum separation of pilots for which an error-free reconstruction of the channel impulse response is still possible. A trade-off between estimation accuracy and pilot overhead has to be found.

A key component of iterative processing and factor graphs in general, is the exchange and the combining of random measures, i.e. for turbo decoding the exchange and combination of L-values, while for joint estimation and detection also probability density function are typically exchanged and combined. The common assumption is that the individual random measures are uncorrelated, which is achieved by a sufficiently long interleaver length. But correlation cannot be avoided in all cases and hence, correlated combining explains how to combine correlated observations as well as correlated variables. More accurate results are achieved by taking the correlation into account. As a result, a receiver achieves a higher estimation and/or detection accuracy and improves its overall performance.

In Chapter 4 the applicability of particle swarm optimization for MIMO channel estimation w.r.t. the achievable performance, convergence speed, and complexity is evaluated in detail. The research community of particle swarm optimization is extremely active and new variants of the algorithm appear continuously. Most of these advances are specific for an optimization problem and need to be adapted individually. Nevertheless, several contributions are reported, which yield a performance gain w.r.t. convergence speed and/or optimization precision for a wide range of optimization problems. Due to this vast amount of PSO-related literature, the majority of publications in the area of wireless communications ignore these improvements. Therefore, the general algorithm is reviewed and suitable parameter sets for MIMO channel estimation are presented. Furthermore, the application of PSO to MIMO channel estimation with a flat fading channel is studied. Cooperative versions of PSO are proposed and compared to non-cooperative approaches. Hereby, the cooperative approaches separate the high-dimensional optimization problem into multiple sets of lower dimension. It is shown that due to cooperation the convergence speed of the algorithm is significantly improved and outperforms non-cooperative approaches. Conventional PSO, however, is limited to single-objective optimization problems. For the generalization to a time-varying frequency-selective channel, a multi-objective PSO is introduced. Accordingly, every OFDM subcarrier and OFDM symbol represents an objective. With the proposed leader selection and archive maintenance, the convergence is enabled even for so-called many-objective problems. The MOPSO shows an improved performance compared to LS channel estimation and a lower complexity compared to a Wiener-filter.

The simplicity with which problems can be implemented with PSO is one of its main advantages. Moreover, the algorithm does not require the knowledge of a particular a priori information, nor a special initialization. However, a simple adaptation of the algorithm should not be mistaken with a low computational complexity. Instead, the overall complexity depends significantly on the maximum number of iterations. The restriction to optimization methods as well as the selection of parameters, which does not need to be fine-tuned to optimum performance for each individual optimization problem, enables the assessment of the overall complexity of PSO. By a thorough analysis of PSO and MOPSO, it is shown that the properties and the advantages of the algorithms are best exploited when used to provide initial channel state information. Furthermore, massive MIMO system are particularly interesting due to their asymmetric channel matrix and the resulting complexity advantage of PSO compared to conventional MMSE channel estimation.

The multi-dimensional graph-based soft iterative receiver is derived in Chapter 5. The factorization of the conditional probability density function $p\left(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{H}\right)$ is explained and the underlying graphical model is developed. The major premise of the factor graph design is to achieve a very low computational complexity, which affects the message generation as well as the message exchange. By applying the Gaussian approximation for the multi-antenna interference for channel estimation, a complexity which depends linearly on the number of antennas is attained. It is shown that the detection algorithms in Chapter 3.1 can be easily integrated within the graph-based framework. In order to facilitate a low complexity message exchange so-called transfer nodes based on a Gaussian random walk model are developed. Although the Gaussian random walk is non-stationary, it is shown that in combination with message combining, the overall process can be seen as asymptotic stationary and is therefore well suited to model the fading characteristics of a mobile radio channel. The additional advantage of transfer nodes is that an arbitrary amount of dimensions can be connected with only a linear increase of the complexity. Due to the presence of cycles, the scheduling of the message exchange is important since it affects the variance of the distributed messages. The impact on the achievable performance as a function of schedule is illustrated.

However, the achievable performance of the MD-GSIR is suboptimal under certain conditions. An analysis of the exchanged messages revealed that correlation arises due to the probabilistic message exchange and the loopy graph structure. Correlated combining, which has been derived in Section 3.4 is implemented and yields significant performance improvements. Further improvements are achieved by means of a more sophisticated initialization. A Wiener-filter as well as the MOPSO algorithm of Chapter 4 are evaluated, whereas the MOPSO-initialized MD-GSIR achieves the best performance in combination with QPSK modulation and a 2×2 MIMO system. The performance is similar compared to state-of-the-art receiver given by Wiener-filter based channel estimation and iterative

APP detection (Wiener+APP), whereas the complexity is significantly lower.

Furthermore, the robustness of the receiver in terms of a priori information is evaluated. For the majority of channel scenarios, almost no performance loss is observed when detailed information of the fading statistics are missing. A 0.5 dB loss is reported for the C2 NLOS channel model. Additionally, it is illustrated that the pilot overhead can be reduced by a factor of two, which results in a performance loss between 0 and 1 dB for the MD-GSIR depending on the channel scenario. Subsequently, the achievable performance of the MD-GSIR for various modulation formats and code rates is studied and compared to the Wiener+APP receiver. The main conclusion from these results is that the channel estimation established by transfer node competes well with the Wiener filter independent of the chosen modulation and coding scheme as well as channel scenario. The complexity of the MD-GSIR can further be reduced by implementing the Gaussian tree search detection, which has been derived in Chapter 3.1.3. The versatility of the transfer nodes is highlighted by the adaptation to a codebook-based beamforming scheme. Without any changes to the message exchange, subcarrier-wise precoding can be performed, which previously resembled a major problem for conventional filter-based channel estimation algorithms.

Finally, the maximum pilot spacing for iterative joint channel estimation and data detection is studied in Chapter 6. It has been observed in multiple publications that the initial pilot spacing is not longer restricted to the Nyquist-Shannon sampling theorem when iterative semi-blind channel estimation is performed. In this thesis, the relation to the channel code as well as code rate is established. By separating the MSE into a noise and an interpolation part, a so-called coded sampling bound is defined. The coded sampling bound is verified for the non-iterative channel estimation by a comparison to the known bound given by the Nyquist-Shannon sampling theorem. It is shown that the maximum spacing in the non-iterative case can be extended by a factor of two, independent of the channel code and code rate. Furthermore, given a repetition code, the maximum pilot spacing additionally depends on the chosen spreading factor and thus, can be further extended.

Directions for future work

In this thesis, a multi-dimensional graph-based iterative receiver has been developed. For the design and the assessment of the proposed MD-GSIR, the range of available parameters has to be restricted. Analysis tools, such as EXIT charts, have shown to be very effective in the analysis of iterative channel codes as well as iterative equalizers. In the presence of channel estimation errors, however, their accuracy diminishes and an exact prediction of the convergence behavior is not longer possible. With precise analysis tools for iterative joint channel estimation and data detection algorithms, the complexity as well as efficiency of a receiver can be improved since a receiver can be adjusted to a specific channel code and the required pilot overhead can be optimized.

Codebook-based beamforming is one of the key techniques to achieve high spectral efficiencies. It has been shown that precoding on a subcarrier basis is beneficial in terms of BER. However, the precoded channel looses its “smoothness”, i.e. with changing beamweights the channel impulse response changes as well. Hence, channel estimation algorithms based on interpolation/filtering typically enforce a precoding scheme for which a complete block of subcarriers is precoded with the same weight. As a proof of concept, it is illustrated that the MD-GSIR supports subcarrier-wise precoding when the index of the selected weights is known to the receiver. For a reasonable implementation, the selected beam-weights need to be estimated on the basis of the common channel. Furthermore, by extending the factor graph, joint common and dedicated channel estimation can be facilitated.

Due to the vast amount of variables within a wireless system, certain restrictions have been assumed. For example, the system is limited to a single-user system. However, in a multi-user multi-cell system, the receiver has to deal with intercell-interference. One approach is to assume the interference as an additional noise term. More promising is to model the interference within the factor graph structure. However, this may lead to short cycles. Hence, either message exchange and combining need to be adapted and/or the graph structure itself.

Correlated combining is a promising approach to acknowledge a suboptimal graph structure without affecting the complexity. However, knowledge of the correlation between messages is crucial. In case of the MD-GSIR, it has been shown that correlation between messages can be approximated. Other graph-based algorithms, such as e.g. LDPC decoding, experience the same impairments of the performance due to correlated messages when for example a short sequence length is considered. There, the calculation of the correlation matrix on the basis of the generator polynomial would be desirable.

A

Notation

Functions and Operators

$\mathrm{x}(\cdot)$	Function with continuous argument $\cdot$
$\mathrm{x}[\cdot]$	Function with discrete argument $\cdot$
$(\cdot)^{\mathrm{T}}$	Transpose of vector/matrix variable
$(\cdot)^{-1}$	Inverse of vector/matrix variable
$(\cdot)^{\mathrm{H}}$	Hermitian transpose of vector/matrix variable
$\arg\max\,(\cdot)$	Argument of maximum
$\arg\min\,(\cdot)$	Argument of minimum
$\cos(\cdot)$	Cosine function
$\mathrm{E}\left\{\cdot\right\}$	Expectation of a random variable
$\exp(\cdot)$	Exponential function
$\mathcal{F}\left(\cdot\right)$	Fourier transform
$\ln(\cdot)$	Natural logarithm (to the base e)
$\log(\cdot)$	Logarithm to the base 10
$\max(\cdot)$	Maximum
$\min(\cdot)$	Minimum
$\mathrm{Re}\{\cdot\}$	Real term of a complex variable
$\sin(\cdot)$	Sine function
$\mathrm{sinc}(\cdot)$	Cardinal sine function

List of Variables

x	Scalar variable

$\boldsymbol{x}$	Vector variable
$\boldsymbol{X}$	Matrix variable
$\mathbb{X}$	Set
x	Constant

Variables used for System Modeling

c	Speed of light
$(\Delta x)_c$	Coherence distance
$(\Delta t)_c$	Coherence time
$(\Delta f)_c$	Coherence bandwidth
D_t	Spacing of training symbols in time domain
D_f	Spacing of training symbols in frequency domain
D_s	Spacing of training symbols in spatial domain
F_s	OFDM subcarrier spacing
$h_{n,m}$	Channel weighting function
$\boldsymbol{H}$	Channel matrix
k	Discrete time index
K	Number of OFDM symbols
K_RB	OFDM symbols within one LTE resource block
l	Discrete frequency index
L	Number of OFDM subcarriers
L_CP	Length of cyclic prefix
L_RB	OFDM subcarriers within one LTE resource block
N_D	Number of data symbols
N_P	Number of training symbols
N_R	Number of receive antennas
N_T	Number of transmit antennas
σ_n^2	Variance of noise
T_s	OFDM symbol duration
v	Velocity of mobile station
$\boldsymbol{n}$	White Gaussian noise vector
$\mathbf{w}^{(i)}$	Beamforming vector
$\boldsymbol{x}$	Transmit vector
ξ	Signal-to-noise ratio
$\boldsymbol{y}$	Receive vector

Variables used for Channel Modeling

c	Cluster index
δf	Discrete spacing of two frequencies
Δt	Continuous spacing in time
Δf	Continuous spacing in frequency
Δx	Continuous spacing in direction

d_{Tx}	Spacing between transmit antenna elements
d_{Rx}	Spacing between receive antenna elements
$f_{\mathrm{D,c,r}}$	Doppler frequency of cluster c and ray r
M_{c}	Number of clusters
M_{r}	Number of rays
ϕ_c	Mean angle of arrival of cluster c
φ_c	Mean angle of departure of cluster c
$\Phi_{c,r}$	Random phase
Φ_{AoA}	Composite angular spread of arriving rays
Φ_{AoD}	Composite angular spread of departing rays
Ψ_{AoA}	Mean composite angle of arriving rays
Ψ_{AoD}	Mean composite angle of departing rays
r	Ray index
τ_c	Propagation delay of cluster c

Variables used for Detection and Estimation Algorithms

a_i	Weighting factor for the combining of correlated L-values
α_i	Weighting factor for the combining of correlated pdfs
$\hat{\boldsymbol{h}}_{\mathrm{LS}}$	LS estimate of h
$\hat{\boldsymbol{h}}_{\mathrm{MMSE}}$	MMSE estimate of h
$\boldsymbol{i}_{N_{\mathrm{RM}}}$	Unit vector of length N_{RM}
$\boldsymbol{\mathcal{L}}$	Constellation set used for GTS detection
κ	Sampling ratio compared to the Shannon-Nyquist sampling theorem
$\mathrm{L_A}$	A priori log-likelihood ratio
$\mathrm{L_E}$	Extrinsic log-likelihood ratio
$\mathrm{L_P}$	A posteriori log-likelihood ratio
Λ^{APP}	APP metric
Λ^{GA}	Metric of Gaussian detector
Λ^{GTS}	Metric of Gaussian tree search
$\Lambda^{\mathrm{GTS}}_{\mathrm{par}}$	Partial path metric of Gaussian tree search
Λ^{QR}	QR-based metric
M_{sl}	Number of significant leaves
M_{vn}	Number of visited nodes
μ_x	Mean value of x
$\mu_{\zeta_{n,m}}$	Mean value of interference for the nth receive antenna and mth transmit antenna
N_{b}	Number of bits per symbol
N_{m}	Modulation order
N_{RM}	Number of random variables
R_{c}	Ratio of metric calculations compared to an APP detector
$\boldsymbol{\mathcal{S}}$	Full constellation set
$\boldsymbol{\Sigma}$	Covariance matrix of random variables
σ^2_{ε}	Overall MSE of time index k

σ_i^2	MSE caused by interpolation error
σ_n^2	MSE caused by noise
σ_x^2	Variance value of x
$\sigma_{\zeta_{n,m}}^2$	Variance value of interference for the nth receive antenna and mth transmit antenna
$\mathbf{w}_k$	Filter coefficient for time index k
$\tilde{\boldsymbol{x}}$	Hypothesis of x
$\hat{\boldsymbol{x}}^{\mathrm{ML}}$	ML estimate of x
$\hat{\boldsymbol{x}}^{\mathrm{MAP}}$	MAP estimate of x
$\mathcal{X}$	Constellation set used for APP detection
$\zeta_{n,m}$	Effective Gaussian interference of the nth receive and mth transmit antenna

Variables used for Particle Swarm Optimization

c_1	Acceleration coefficient towards personal best
c_2	Acceleration coefficient towards global best
$\boldsymbol{\varepsilon}$	Random numbers in the range [0,1]
i_{max}	Maximum number of iterations
N_{p}	Number of particles within one swarm
N_{p}'	Number of particles within one subswarm
N_{s}	Number of subswarms
$\boldsymbol{\Omega}$	Inertia weight
$\boldsymbol{p}_i$	Current position of a particle
$\boldsymbol{p}_i^{\mathrm{IB}}$	Personal best position of a particle
$\boldsymbol{p}^{\mathrm{GB}}$	Global best position of a swarm
p^{OPT}	Optimal fitness value
$\boldsymbol{p}_{s,i}$	Current position of particle i of swarm s
$\boldsymbol{p}_1^{\mathrm{GB}}$	Partial global best of the first swarm
$\boldsymbol{S}_{\mathrm{min}}$	Minimum border of the search space
$\boldsymbol{S}_{\mathrm{max}}$	Maximum border of the search space
$\boldsymbol{S}_{\mathrm{mut}}$	Mutation range
t_{h}	Threshold limit to stop the iterative process of PSO
$\boldsymbol{v}_i$	Velocity vector of particle i
$\boldsymbol{V}_{\mathrm{min}}$	Minimum velocity
$\boldsymbol{V}_{\mathrm{max}}$	Maximum velocity

Variables used for Graph-based Soft Iterative Receiver

η	Spectral efficiency
η_{ACLR}	Bandwidth efficiency w.r.t. adjacent channel leakage ratio
η_{BW}	Overall bandwidth efficiency
η_{CP}	Bandwidth efficiency w.r.t. cyclic prefix
η_{RS}	Bandwidth efficiency w.r.t. pilot overhead

I_{A}	A priori mutual information
$I_{\mathrm{A,DEC}}$	A priori information used by a channel decoder
$I_{\mathrm{A,DEM}}$	A priori information used by a MIMO detector
I_{E}	Extrinsic mutual information
$I_{\mathrm{E,DEC}}$	Extrinsic information provided by a channel decoder
$I_{\mathrm{E,DEM}}$	Extrinsic information provided by a MIMO detector
$\mu_{C\rightarrow f_C}(c_i)$	Message sent from a bit node to mapping node
$\mu_{H\rightarrow f_\Delta}(h_{n,m})$	Message sent from a coefficient node to a transfer node
$\mu_{f_\Delta\rightarrow H}(h_{n,m})$	Message sent from a transfer node to a coefficient node
$\mu_{f_Y\rightarrow H}(h_{n,m})$	Message sent from an observation node to a coefficient node
$\mu_{X\rightarrow f_Y}(x_m)$	Message sent from a symbol node to an observation node
$\mu_{h,n,m}$	Mean value of the message $\mu_{f_Y\rightarrow H}(h_{n,m})$
$\sigma^2_{\Delta,n',m'}[l',k']$	Variance of a domain-specific transfer node
$\sigma^2_{h,n,m}$	Variance of the message $\mu_{f_Y\rightarrow H}(h_{n,m})$
P_{bl}	Probability that a codeword is transmitted successfully

Abbreviations

1G	First generation mobile system
2G	Second generation mobile system
3G	Third generation mobile system
4G	Fourth generation mobile system
ACLR	Adjacent channel leakage ratio
AMC	Adaptive modulation and coding
AoA	Angle of arrival
AoD	Angle of departure
APP	A posteriori probability
AWGN	Additive white Gaussian noise
BBPSO	Bare bones PSO
BER	Bit error rate
BICM	Bit-interleaved coded modulation
BLER	Block error rate
BLUE	Best linear unbiased estimator
BPSK	Binary phase shift keying
CBBPSO	Cooperative bare bone PSO
CCI	Co-channel interference
CDF	Cumulative distribution function
CDL	Clustered delay line
CGSIR	GSIR with correlated combining
CPSO	Cooperative PSO
CSB	Coded sampling bound
CSI	Channel state information
CSIT	Transmitter channel state information
CP	Cyclic prefix

CQI	Channel quality information
DFT	Discrete Fourier transform
EA	Evolutionary algorithm
EM	Expectation maximization
ESE	Elementary signel estimator
EXIT	Extrinsic information transfer
FFT	Fourier transform
GA	Gaussian approximation
GAO	Genetic algorithms
GEV	Generalized extreme value
GPU	Graphical processing unit
GSIR	Graph-based soft iterative receiver
GTS	Gaussian tree search
IDFT	Inverse discrete Fourier transform
IDMA	Interleave-division multiple access
IFFT	Inverse fast Fourier transform
ICI	Intercarrier interference
ISI	Intersymbol interference
LLR	Log likelihood ratio
LS	Least square
LTE	Long-term evolution
LTE-A	Long-term evolution-advanced
LTV	Linear time-variant
MAC	Medium access control
MAP	Maximum a posteriori
MCS	Modulation and coding scheme
MIMO	Multiple-input multiple-output
MISO	Multiple-input single-output
ML	Maximum likelihood
MSE	Mean squared error
MMSE	Minimum mean squared error
MOPSO	Multi-objective particle swarm optimization
NLOS	Non line-of-sight
OFDM	Orthogonal frequency division multiplexing
PACE	Pilot aided channel estimation
PAPR	Peak to average power ratio
PDF	Probability density function
PDP	Power delay profile
PSO	Particle swarm optimization
PHY	Physical layer
P/S	Parallel-to-serial conversion
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase shift keying
QoS	Quality of service

RE	Resource element
RB	Resource block
RPG	Reduced pilot grid
SA	Simulated annealing
SAGE	Space alternating generalized expectation maximization
SBCE	Semi-blind channel estimation
SDMA	Space division multiple access
SiSo	Soft-input soft-output
SISO	Single-input single-output
SINR	Signal-to-interference plus noise ratio
SNR	Signal-to-noise ratio
SM	Superposition modulation
S/P	Serial-to-parallel conversion
WINNER	Wireless world initiative new radio
WiMAX	Worldwide interoperability for microwave access

B

Parameters of WINNER Channel Models

The tabulated parameters of each scenario comprise the discrete power delay profile as well as the azimuth angle of departure and arrival, respectively.

Ray number r	Offset angles γ_r
1, 2	±0.0447
3, 4	±0.1413
5, 6	±0.2492
7, 8	±0.3715
9, 10	±0.5129
11, 12	±0.6797
13, 14	±0.8844
15, 16	±1.1481
17, 18	±1.5195
19, 20	±2.1551

Table B.1: Ray offset angles within a cluster

WINNER A1 NLOS

Short description of the WINNER A1 NLOS scenario:

> "Base stations (Access Points) are assumed to be in corridor, thus LOS case is corridor-to-corridor and NLOS case is corridor-to-room. In the NLOS case

Table B.2: Indoor office / residential model, WINNER A1 NLOS

cluster c #	1	2	3	4	5	6	7	8	9	10
delay τ_c [ns]	0	5	5	5	15	15	15	20	25	40
AoD φ_c [°]	45	77	43	72	54	-65	-60	85	0	-104
AoA ϕ_c [°]	41	-70	39	66	-49	59	-55	-78	0	95
power P_c [dB]	-15.2	-19.7	-15.1	-18.8	-16.3	-17.7	-17.1	-21.2	-13.0	-14.6
cluster c #	11	12	13	14	15	16				
delay τ_c [ns]	80	85	110	115	150	175				
AoD φ_c [°]	95	-104	-105	103	-135	-122				
AoA ϕ_c [°]	86	95	-96	-94	123	-111				
power [dB]	-23.0	-25.1	-25.4	-24.8	-33.4	-29.6				

the basic path-loss is calculated into the rooms adjacent to the corridor where the AP is situated. For rooms farther away from the corridor wall-losses must be applied for the walls parallel to the corridors. E.g. for the UE at the bottom wall of the lay-out in the Figure 2-1 there are three walls to be taken into account. Finally, we have to model the Floor Loss (FL) for propagation from floor to floor. It is assumed that all the floors are identical. The Floor Loss is constant for the same distance between floors, but increases with the floor separation and has to be added to the path-loss calculated for the same floor." [IST07, p. 16]

WINNER B1 NLOS

Table B.3: Typical urban micro-cell model, WINNER B1 NLOS

cluster c #	1	2	3	4	5	6	7	8	9	10
delay τ_c [ns]	0	95	105	115	230	240	245	285	390	430
AoD φ_c [°]	8	0	-24	-24	-24	29	29	30	-37	41
AoA ϕ_c [°]	-20	0	57	-55	57	67	-68	70	-86	-95
power P_c [dB]	-14.0	-13.0	-13.9	-21.1	-21.6	-24.7	-25.0	-25.9	-32.6	-36.9
cluster c#	11	12	13	14	15	16				
delay τ_c [ns]	460	505	515	595	600	615				
AoD φ_c [°]	-39	-42	-40	47	47	46				
AoA ϕ_c [°]	-92	-99	94	111	110	-107				
power P_c [dB]	-35.1	-38.6	-36.4	-45.2	-44.7	-42.9				

Short description of the WINNER B1 NLOS scenario:

"In urban micro-cell scenarios the height of both the antenna at the BS and at the MS is assumed to be well below the tops of surrounding buildings. Both antennas are assumed to be outdoors in an area where streets are laid out in a Manhattan-like grid. The streets in the coverage area are classified as 'the main street', where there is the LOS from all locations to the BS, with the possible exception in cases where the LOS is temporarily blocked by traffic (e.g. trucks and busses) on the street. Streets that intersect the main street are referred to as perpendicular streets, and those that run parallel to it are referred to as parallel streets. This scenario is defined for both the LOS and the NLOS cases. Cell shapes are defined by the surrounding buildings, and energy reaches NLOS streets as a result of the propagation around corners, through buildings, and between them." [IST07, p. 17]

WINNER C1 NLOS

Table B.4: Suburban macro-cell model, WINNER C1 NLOS

cluster c #	1	2	3	4	5	6	7	8	9	10	11	12	13	14
delay τ_c [ns]	5	25	35	35	50	65	65	75	145	160	195	200	205	770
AoD φ_c [°]	0	13	-15	-8	12	-17	12	-8	-10	-13	12	8	14	22
AoA ϕ_c [°]	0	-71	-84	46	-66	-97	-66	-46	-56	73	70	-46	-80	123
power P_c [dB]	-13.0	-20.8	-23.5	-16.2	-16.1	-27.0	-19.4	-16.1	-17.6	-21.0	-20.2	-16.1	-22.5	-35.4

Short description of the WINNER C1 NLOS scenario:

"In suburban macro-cells base stations are located well above the rooftops to allow wide area coverage, and mobile stations are outdoors at street level. Buildings are typically low residential detached houses with one or two floors, or blocks of flats with a few floors. Occasional open areas such as parks or playgrounds between the houses make the environment rather open. Streets do not form urban-like regular strict grid structure. Vegetation is modest." [IST07, p. 19]

WINNER C2 NLOS

Short description of the WINNER C2 NLOS scenario:

"In typical urban macro-cell mobile station is located outdoors at street level and fixed base station clearly above surrounding building heights. As for propagation conditions, non- or obstructed line-of-sight is a common case, since street level is often reached by a single diffraction over the rooftop. The

Table B.5: Urban Macro Channel model, WINNER C2 NLOS

cluster c #	1	2	3	4	5	6	7	8	9	10
delay τ_c [ns]	0	60	75	150	150	190	225	335	370	430
AoD φ_c [°]	11	-8	-6	0	6	8	-12	-9	-12	-12
AoA ϕ_c [°]	61	44	-34	0	33	-44	-67	52	-67	-67
power P_c [dB]	-6.4	-3.4	-2.0	-5.2	-1.9	-3.4	-5.6	-4.6	-7.8	-7.8
cluster c #	11	12	13	14	15	16	17	18	19	20
delay τ_c [ns]	510	685	725	735	800	960	1020	1100	1210	1845
AoD φ_c [°]	13	15	-12	-1	-14	19	-16	15	18	17
AoA ϕ_c [°]	-73	-83	-70	87	80	109	91	-82	99	98
power P_c [dB]	-9.3	-12.0	-8.5	-13.2	-11.2	-20.8	-14.5	-11.7	-17.2	-16.7

building blocks can form either a regular Manhattan type of grid, or have more irregular locations. Typical building heights in urban environments are over four floors. Buildings height and density in typical urban macro-cell are mostly homogenous." [IST07, p. 19]

C

Codebook-Based Beamforming

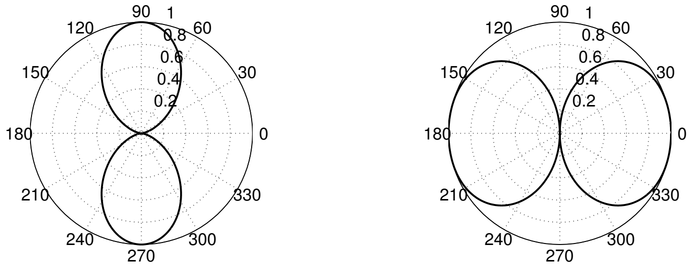
In order to facilitate beamforming at least two antennas ($N_{\mathrm{T}} \geq 2$) are required. They are typically positioned according a certain geometry, i.e. in a linear or in a circular array, etc.. Of importance is the distance $\delta = \lambda/2$, which is often given in multiples of the wavelength. The wavefronts emanating from each antenna superimpose constructively or destructively depending on their individual phase. By a suitable selection of a weighting factor, the phase and/or amplitude of the antennas can be controlled in order to design the resulting interference pattern. Hence, a directed transmission of the signal is possible. The characteristic pattern of such a *smart antenna* is described by the array factor

$$AF^{(i)} = \frac{1}{|\mathbf{w}|_{\max}} \sum_{m=0}^{N_{\mathrm{T}}-1} w_m^{(i)} \exp\left(-j\mathbf{k}^{\mathrm{T}}\mathbf{d}_m\right) \tag{C.1}$$

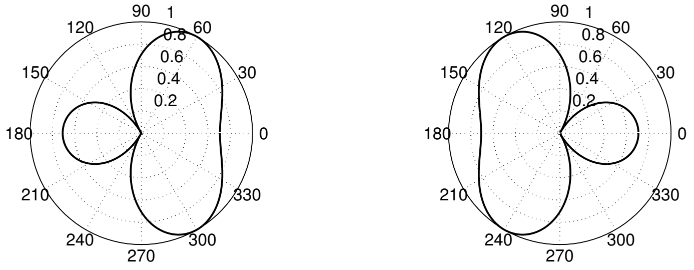
whereas $w_m \in \mathbb{C}$ refers to the scalar weighting factor of the mth antenna element, $m \in \{0, 1, \ldots, N_{\mathrm{T}} - 1\}$. The so called wave vector is given by

$$\mathbf{k} \doteq 1\pi/\lambda \cdot \left[\cos(\phi)\cos(\varphi), \sin(\phi)\cos(\varphi), \sin(\varphi)\right]^{\mathrm{T}}. \tag{C.2}$$

The advantages of beamforming are an improved signal-to-interference-plus-noise ratio, better cell coverage, and an increase of channel capacity [Hoe13]. The major drawback of beamforming is that it requires channel state information at the transmitter side. Under circumstances, the uplink and downlink of a radio channel are reciprocal, e.g. given time-division duplex. Hence, the channel can be directly estimated at the base-station. However, with frequency-division duplex, reciprocity of the uplink and downlink channel is not fulfilled. As a consequence, the channel quality information need to be send to the transmitter. In order to reduce the feedback overhead, codebooks, known to the transmitter and the receiver, are used. The entry of a codebook which maximizes for example



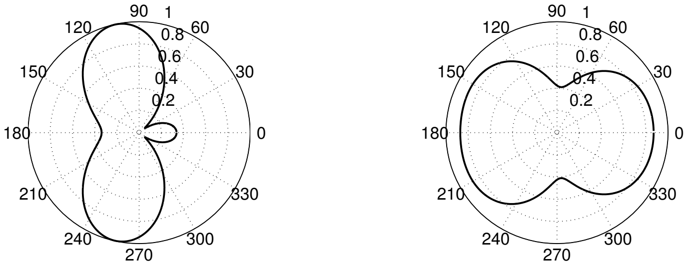
(a) $|AF|$ for $w^{(0)}$ (b) $|AF|$ for $w^{(1)}$



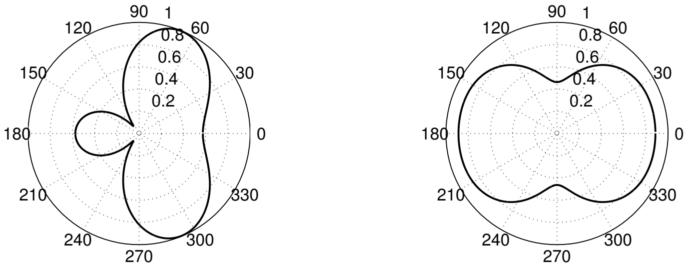
(c) $|AF|$ for $w^{(2)}$ (d) $|AF|$ for $w^{(3)}$

Figure C.1: Array factor of DFT-Codebook weights.

the capacity is chosen. Only the corresponding index is sent to the transmitter instead of the complex beam-weight. Popular codebooks are the DFT codebook and the Grassmannian codebook. The beam-weights of the two codebooks are given in Table 2.2 and Table 2.3, respectively. The array factors for the DFT codebook and the Grassmannian codebook are shown in Figure C.1 and Figure C.2. The most notable difference is that the array factors of the Grassmannian codebook do not reach the maximum antenna gain of the normalized value of one. Opposed to the array factors of the DFT codebook, where two desired directions for each array factor have the maximum antenna gain of one. The achievable performance of codebook-based beamforming with the proposed MD-GSIR is evaluated in Section 5.6.



(a) $|AF|$ for $w^{(0)}$ (b) $|AF|$ for $w^{(1)}$



(c) $|AF|$ for $w^{(2)}$ (d) $|AF|$ for $w^{(3)}$

Figure C.2: Array factor of Grassmannian-Codebook weights.

D

Own Publications Related to the Thesis

Journal Papers and Letters

- C. Knievel, and P.A. Hoeher, “Coded Sampling Bound - How much training is needed for iterative semi-blind channel estimation?,” accepted for publication in *IEEE Transactions on Communications*, 2014.

- C. Knievel, P.A. Hoeher, and G. Auer, “On the Combining of Correlated Random Measures with Application to Graph-Based Receivers,” *IEEE Communications Letters*, vol. 16, no. 12, pp. 1996-1999, Dec. 2012.

- C. Knievel, P.A. Hoeher, A. Tyrrell, and G. Auer, “Multi-Dimensional Graph-Based Soft Iterative Receiver for MIMO-OFDM,” *IEEE Transactions on Communications*, vol. 60, no. 6, pp.1599-1906, June 2012.

- C. Knievel and P.A. Hoeher, “On Particle Swarm Optimization for MIMO Channel Estimation,” *Journal of Electrical and Computer Engineering*, vol. 2012, Article ID 614384, 10 pages, 2012.

Conference Papers

- C. Knievel, D. Hao, P.A. Hoeher, P. Weitkemper, and H. Taoka, “Evaluation and Extension of a Multi-Dimensional Graph-Based Receiver Concept for MIMO-OFDM,” in *Proc. IEEE International Conference on Communications (ICC)*, Budapest, Hungary, June 2013.

- C. Knievel, P.A. Hoeher, A. Tyrrell, and G. Auer, “Improving Multi-Dimensional Graph-Based Soft Channel Estimation,” in *Proc. IEEE Vehicular Technology Conference*, Yokohama, Japan, May 2012.
- C. Knievel, M. Noemm, and P.A. Hoeher, “Low-Complexity Receiver for Large-MIMO Space-Time Coded Systems,” in *Proc. IEEE Vehicular Technology Conference*, San Francisco, USA, Sep. 2011.
- C. Knievel, P.A. Hoher, A. Tyrrell, and G. Auer, “Particle Swarm Enhanced Graph-Based Channel Estimation for MIMO-OFDM,” in *Proc. IEEE Vehicular Technology Conference*, Budapest, Hungary, May 2011.
- C. Knievel, Z. Shi, P.A. Hoeher, and G. Auer, “2D Graph-Based Soft Channel Estimation for MIMO-OFDM,” in *Proc. IEEE International Conference on Communications*, Capetown, South Africa, May 2010.
- K. Schmeink, R. Block, C. Knievel, and P.A. Hoeher, “Joint Channel and Parameter Estimation for Joint Communication and Navigation using Particle Swarm Optimization,” in *Proc. Workshop on Positioning, Navigation and Communication (WPNC)*, Dresden, Germany, March 2012.

Patents

- G. Auer, P.A. Hoeher, C. Knievel, Z. Shi, and T. Wo, “Apparatus and method for estimating a channel coefficient of a data subchannel of a radio channel”, European Patent 2 555 479, Feb. 06, 2013.
- G. Auer, A. Tyrrell, and P.A. Hoeher, and C. Knievel, “A radio channel estimation using reliability information”, European Patent 2 293 503, Mar. 09, 2011.

Bibliography

[3GP08] 3GPP, "Physical layer - general description," 3rd Generation Partnership Project (3GPP), TS, Jun. 2008.

[AB07] G. Auer and J. Bonnet, "Threshold controlled iterative channel estimation for coded OFDM," in *Proc. IEEE Vehicular Technology Conf. (VTC Spring)*, Dublin, Ireland, Apr. 2007, pp. 1737–1741.

[AC09a] G. Auer and I. Cosovic, "Pilot design of MIMO-OFDM with beamforming," in *Proc. IEEE International Conference on Communications (ICC)*, Dresden, Germany, Jun. 2009.

[AC09b] ——, "Pilot design for multi-user MIMO," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 3621–3624.

[ACH05] M. Alias, S. Chen, and L. Hanzo, "Multiple-antenna-aided OFDM employing genetic-algorithm-assisted minimum bit error rate multiuser detection," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 5, pp. 1713–1721, Sep. 2005.

[AEVZ02] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.

[AGR98] P. D. Alexander, A. J. Grant, and M. C. Reed, "Performance analysis of an iterative decoder for code-division multiple access," *European Transactions on Telecommunications (ETT)*, vol. 9, no. 5, pp. 419–426, Sep./Oct. 1998.

[AK04] G. Auer and E. Karipidis, "Pilot aided channel estimation for OFDM: a separated approach for smoothing and interpolation," in *Proc. IEEE International Conference on Communications (ICC)*, Seoul, South Korea, May 2004.

[AKt04] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[Ala98] S. M. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[ATV02] S. Adireddy, L. Tong, and S. Vishwanath, “Optimal placement of training for frequency-selective block-fading channels,” *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2338–2353, Aug. 2002.

[Aue03a] G. Auer, “Channel estimation for OFDM systems with multiple transmit antennas by filtering in time and frequency,” in *Proc. IEEE Vehicular Technology Conf. (VTC Fall)*, Orlando, Florida, USA, Oct. 2003, pp. 1204–1208.

[Aue03b] ——, “Channel estimation in two dimensions for OFDM systems with multiple transmit antennas,” in *Proc. IEEE Global Communications Conference (GLOBECOM)*, San Francisco, USA, Dec. 2003, pp. 322–326.

[Aue04] ——, “Analysis of pilot-symbol aided channel estimation for OFDM systems with multiple transmit antennas,” in *Proc. IEEE International Conference on Communications (ICC)*, Paris, Jun. 2004, pp. 3221–3225.

[Aue09] ——, “3D pilot aided channel estimation,” in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Budapest, Hungary, Apr. 2009.

[Aue12] ——, “3D MIMO-OFDM channel estimation,” *IEEE Transactions on Communications*, vol. 60, no. 4, pp. 972–985, Apr. 2012.

[BAS+05] K. Brueninghaus, D. Astely, T. Saelzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, “Link performance models for system level simulations of broadband radio access systems,” in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Berlin, Germany, Sep. 2005.

[BCC+07] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications.* Cambride University Press, 2007.

[BCJR74] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[BE07] D. Bratton and R. C. Eberhart, “Defining a standard for particle swarm optimisation,” in *Proc. Swarm Intelligence Symposium (SIS 2007)*, Apr. 2007, pp. 120–127.

[Bel63] P. Bello, “Characterization of randomly time-variant linear channels,” *IEEE Transactions on Communications*, vol. 11, no. 4, pp. 360–393, Dec. 1963.

[Bel64] ——, "Time-frequency duality," *IEEE Transactions on Information Theory*, vol. 10, no. 1, pp. 18–33, Jan. 1964.

[BGT93] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting: Coding and decoding: Turbo-codes," in *Proc. IEEE International Conference on Communications (ICC)*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[BKM$^+$12] M.-A. Badiu, G. E. Kirkelund, C. Manchon, E. Riegler, and B. H. Fleury, "Message-passing algorithms for channel estimation and decoding using approximate inferencce," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Boston, USA, Jul. 2012.

[Bos98] M. Bossert, *Kanalcodierung*, 2nd ed. Germany: Vieweg and Teubner, 1998.

[BPG$^+$09] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference coordination and cancellation for 4G networks," *IEEE Communications Magazine*, vol. 74, no. 4, pp. 74–81, Apr. 2009.

[BR03] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," *ACM Computing Surveys*, vol. 35, no. 3, pp. 268–308, Sep. 2003.

[BTA$^+$07] H. Bodur, C. A. Tunc, D. Aktas, V. B. Erturk, and A. Altintas, "Particle swarm optimization for SAGE maximization step in channel parameter estimation," in *2nd European Conference on Antennas and Propagation (EuCAP)*, Ankara, Turkey, Nov. 2007.

[CA07] I. Cosovic and G. Auer, "Capacity of MIMO-OFDM with pilot aided channel estimation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, Article ID 32460, 12 pages, 2007.

[CA08] V. Chandrasekhar and J. G. Andrews, "Femtocell networks: A survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, Sep. 2008.

[Cav91] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[CD01] A. Carlisle and G. Dozier, "An off-the-shelf PSO," in *Proc. of the Particle Swarm Optimization Workshop*, Apr. 2001, pp. 1–6.

[CH03] C. Cozzo and B. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Transactions on Communications*, vol. 51, no. 8, pp. 1266–1270, Aug. 2003.

[CHH$^+$10] C.-H. Cheng, H.-C. Hsu, Y.-F. Huang, J.-H. Wen, and L.-C. Hsu, "Performance of an adaptive PSO parallel interference canceller for CDMA communication systems," in *Proc. 5th Annual ICST Wireless Internet Conference (WICON)*, Singapore, Mar. 2010.

[CL02] C. A. C. Coello and M. S. Lechuga, "MOPSO: A proposal for multiple objective particle swarm optimization," in *Proc. of the 2002 Congress of Evolutionary Computation part of the 2002 IEEE World Congress on Computational Intelligence.* IEEE Press, May 2002, pp. 1051–1056.

[CL05] J.-W. Choi and Y.-H. Lee, "Optimum pilot pattern for channel estimation in OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2083–2088, Sep. 2005.

[CL07] ——, "Complexity-reduced channel estimation in spatially correlated MIMO-OFDM systems," *IEICE Transactions on Communications*, vol. E90-B, no. 9, pp. 2609–2612, Sep. 2007.

[Cox12] C. Cox, *An Introduction to LTE.* John Wiley & Sons Inc., 2012.

[CPL04] C. A. C. Coello, G. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimisation," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 256–279, June 2004.

[CS00] J. Chuang and N. Sollenberger, "Beyond 3G: Wideband wireless data access based on OFDM and dynamic packet assignment," *IEEE Communications Magazine*, vol. 38, no. 7, pp. 78–87, Jul. 2000.

[CYY09] Z. Chen, C. Yongyu, and D. Yang, "Low-complexity turbo equalization for MIMO-OFDM system without cyclic prefix," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Tokyo, Japan, Sep. 2009.

[Dah08] W. Dahmen, *Numerik für Ingenieure und Naturwissenschaftler*, 2nd ed. Berlin Heidelberg: Springer, 2008.

[DJB$^+$95] C. Douillard, M. Jézéquel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo-equalization," *European Trans. on Telecommun. (ETT)*, vol. 6, no. 5, pp. 507–511, Sep./Oct. 1995.

[DLL08] W. Dong, J. Li, and Z. Lu, "Joint frequency offset and channel estimation for MIMO systems based on particle swarm optimization," in *Proc. IEEE Vehicular Technology Conf. (VTC Spring)*, Singapore, May 2008.

[DM84] D. Dudgeon and R. Mersereau, *Multidimensional Digital Signal Processing.* Prentice Hall, Inc., 1984.

[D'O08] L. D'Orazio, "Study and development of novel techniques for PHY-layer optimization of smart terminals in the context of next-generation mobile communications," Ph.D. dissertation, University of Trento, Italy, Nov. 2008.

[DPAM02] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[Eri13] Ericsson, "Ericsson mobility report," Sep. 2013. [Online]. Available: http://www.ericsson.com/ericsson-mobility-report

[FG98] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[Fle00] B. H. Fleury, "First and second-order characterization of direction dispersion and space selectivity in the radio channel," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2027–2044, Sep. 2000.

[FM97] B. J. Frey and D. J. MacKay, "A revolution: Belief propagation in graphs with cycles," in *Proc. Neural Information Processing Systems*, Denver, USA, Dec. 1997.

[FSMH05] J. C. Fricke, M. Sandell, J. Mietzner, and P. A. Hoeher, "Impact of the Gaussian approximation on the performance of the probabilistic data association MIMO decoder," *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 5, pp. 796–800, Oct. 2005.

[Ges12] D. Gesbert, "Cooperation over interference-limited networks: Distributed approaches," in *Proc. IEEE Workshop on Advanced Information Processing For Wireless Communication Systems*, Copenhagen, Denmark, May 2012.

[GJJV03] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[GK06] M. Gilli and E. Kellezi, "An application of extreme value theory for measuring financial risk," *Computational Economics*, vol. 27, no. 2-3, pp. 207–228, May 2006.

[GLHL07] Y. Gao, Z. Li, X. Hu, and H. Liu, "A multi-population particle swarm optimizer and its application to blind multichannel estimation," in *Proc. Third Int. Conf. on Natural Computation (ICNC)*, Haikou, China, Aug. 2007.

[Gol89] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, MA: Addison-Wesley, 1989.

[Gol05] A. Goldsmith, *Wireless Communications.* Cambridge University Press, 2005.

[GYH+11] M. Ghil, P. Yiou, S. Hallegatte, B. Malamud, B. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov, O. Mestre, C. Nicolis, H. Rust, P. Shebalin, M. Vrac, A. Witt, and I. Zaliapin, "Exteme events: dynamics, statistics and prediction," *Nonlinear Processes in Geophysics*, vol. 18, pp. 295–350, May 2011.

[Hag04] J. Hagenauer, "The EXIT chart - introduction to extrinsic information transfer," in *Proc. 12th European Signal Processing Conference*, Vienna, Austria, Sep. 2004.

[HAW11] L. Hanzo, Y. Akhtman, and L. Wang, *MIMO-OFDM for LTE, Wi-Fi and WiMAX*. John Wiley & Sons Inc., 2011.

[HEHA11] L. Hanzo, M. El-Hajjar, and O. Alamri, "Near-capacity wireless transceivers and cooperative communications in the MIMO era: Evolution of standards, waveform design, and future perspectives," *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1343–1385, Aug. 2011.

[Hen09] T. Hendtlass, "Particle swarm optimization and high dimensional problem spaces," in *Proc. IEEE Congress on Evolutionary Computation (CEC)*, Trondheim, Norway, May 2009.

[HH03] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[HH12] D. Hao and P. Hoeher, "A low-complexity tree search detection algorithm for superposition modulation," in *Proc. 7th International Symposium on Turbo Codes & Iterative Information Processing*, Gothenburg, Sweden, Aug. 2012.

[HHI$^+$12] L. Hanzo, H. Haas, S. Imre, D. O´ Brien, M. Rupp, and L. Gyongyosi, "Wireless myths, realities, and futures: From 3G/4G to optical and quantum wireless," *Proceedings of the IEEE*, vol. 100, pp. 1853–1888, May 2012.

[HKR97a] P. Hoeher, S. Kaiser, and P. Robertson, "Pilot-symbol-aided channel estimation in time and frequency," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Phoenix, Arizona, USA, Nov. 1997, pp. 90–96.

[HKR97b] ——, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Munich, Germany, Apr. 1997, pp. 1845–1848.

[HL99] P. Hoeher and J. Lodge, "Turbo DPSK: Iterative differential PSK demodulation and channel decoding," *IEEE Transactions on Communications*, vol. 47, pp. 837–843, June 1999.

[HM04] T. Huang and S. A. Mohan, "Significance of neighborhood topologies for the reconstruction of microwave images using particle swarm optimization," in *Proc. 2004 Asia-Pacific Microwave Conference*, Dec. 2004, pp. 237–240.

[Hoe91] P. Hoeher, "TCM on frequency-selective land-mobile fading channels," in *Proc. 5th Tirrenia Int. Workshop on Digital Commun.*, E. Biglieri and M. Luise, Eds., Coded Modulation and Bandwidth-Efficient Transmission. Amsterdam: Elsevier Science Publishers, Sep. 1991, pp. 317–328.

[Hoe13] P. A. Hoeher, *Grundlagen der digitalen Informationsübertragung*, 2nd ed. Springer Vieweg, 2013.

[HOP96] J. Hagenauer, E. Offer, and L. Papke, “Iterative decoding of binary block and convolutional codes,” *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

[HRRE09] P. Hammarberg, F. Rusek, P. S. Rossi, and O. Edfors, “EXIT chart evaluation of a receiver structure for multi-user multi-antenna OFDM systems,” in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Honolulu, Hawaii, Nov. 2009.

[Ht03] B. M. Hochwald and S. ten Brink, “Achieving near-capacity on a multiple-antenna channel,” *IEEE Transactions on Communications*, vol. 51, no. 2, pp. 389–399, Mar. 2003.

[HtBD13] J. Hoydis, S. ten Brink, and M. Debbah, “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[Hub96] J. Huber, “Mehrfachzugriffsverfahren für Mobilkommunikationssysteme,” in *Ferien-Akademie.* Erlangen-Nürnberg: University Erlangen-Nürnberg, Sep. 1996.

[IEE06] IEEE, “Air interface for fixed and mobile broadband wireless access systems amendment 2: Physical and medium access control layers for combined fixed and mobile operation in licensed bands,” IEEE Standard 802.16e-2005, Tech. Rep., Feb. 2006.

[IEE09] ——, “IEEE standard for information technology – local and metropolitan are networks – specific requirements – part 11: Wireless lan medium access control (MAC) and physical layer (PHY) specifications amendment 5: Enhancements for higher throughput,” IEEE Standard 802.11n-2009, Tech. Rep., Oct. 2009.

[IFIW05] A. T. Ihler, J. W. Fisher III, and A. S. Willsky, “Loopy belief propagation: Convergence and effects of message errors,” *Journal of Machine Learning Research*, vol. 6, pp. 905–936, May 2005.

[IH09] T. Inoue and R. Heath, “Kerdock codes for limited feedback precoded MIMO systems,” *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3711–3716, Sep. 2009.

[Int13] Intel, “Intel many integrated core architecture,” Jul. 2013. [Online]. Available: http://www.intel.com/content/www/us/en/architecture-and-technology/many-integrated-core/intel-many-integrated-core-architecture.html

[IST07] IST4-027756 WINNER II, *D1.1.2 WINNER II Channel Models*, Sep. 2007.

[ITN08] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization: A short review," in *IEEE Congress on Evolutionary Computation*, Hong Kong, China, Jun. 2008.

[Jak75] W. C. Jakes, *Microwave Mobile Communications.* John Wiley & Sons Inc., February 1975.

[JH07] M. Jiang and L. Hanzo, "Multiuser MIMO-OFDM for next-generation wireless systems," *Proceedings of the IEEE*, vol. 95, no. 7, pp. 1430–1469, Jul. 2007.

[JO05] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.

[Kay09] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, 17th ed. Upper Saddle River, New Jersey, USA: Prentice Hall, Inc., 2009.

[KB06] M. Khalighi and J. J. Boutros, "Semi-blind channel estimation using the EM algorithm in iterative MIMO APP detectors," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3165–3173, Nov. 2006.

[KE95] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. on Neural Networks*, 1995, pp. 1942–1948.

[KE01] J. Kennedy and R. C. Eberhart, *Swarm Intelligence*, D. E. M. Penrose, Ed. Morgan Kaufmann, 2001.

[Ken03] J. Kennedy, "Bare bones particle swarms," in *Proc. of the 2003 IEEE Swarm Intelligence Symposium*, Indianapolis, USA, Apr. 2003, pp. 80 –87.

[KF98] F. R. Kschischang and B. J. Frey, "Iterative decoding of compound codes by probability propagation in graphical models," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 219–230, Feb. 1998.

[KFL01] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[KH12] C. Knievel and P. A. Hoeher, "On particle swarm optimization for MIMO channel estimation," *Journal of Electrical and Computer Engineering*, vol. 2012, Article ID 614384, pp. 1–10, 2012.

[KHA12] C. Knievel, P. A. Hoeher, and G. Auer, "On the combining of correlated random measures with application to graph-based receivers," *IEEE Communications Letters*, vol. 16, no. 12, pp. 1996–1999, Dec. 2012.

[KHAT11a] C. Knievel, P. A. Hoeher, G. Auer, and A. Tyrrell, "Particle swarm enhanced graph-based channel estimation for MIMO-OFDM," in *Proc. IEEE Vehicular Technology Conf. (VTC Spring)*, Budapest, Hungary, May 2011.

[KHAT11b] ——, “Particle swarm enhanced graph-based channel estimation for MIMO-OFDM,” in *Proc. IEEE Vehicular Technology Conf. (VTC Spring)*, Budapest, Hungary, May 2011.

[KHH+13] C. Knievel, D. Hao, P. A. Hoeher, P. Weitkemper, and H. Taoka, “Evaluation and extension of a multi-dimensional graph-based receiver concept for MIMO-OFDM,” in *Proc. IEEE International Conference on Communications (ICC)*, Budapest, Hungary, Jun. 2013.

[KHTA12a] C. Knievel, P. Hoeher, A. Tyrrell, and G. Auer, “Improving multi-dimensional graph-based soft channel estimation,” in *Proc. IEEE Vehicular Technology Conference (VTC-Spring)*, Yokohama, Japan, May 2012.

[KHTA12b] ——, “Multi-dimensional graph-based iterative receiver for MIMO-OFDM,” *IEEE Transactions on Communications*, vol. 60, no. 6, pp. 1599 – 1906, Jun. 2012.

[KNH11] C. Knievel, M. Noemm, and P. A. Hoeher, “Low-complexity receiver for Large-MIMO space-time coded systems,” in *Proc. IEEE Vehicular Technology Conf. (VTC Fall)*, San Francisco, USA, Sep. 2011.

[KSHA10] C. Knievel, Z. Shi, P. A. Hoeher, and G. Auer, “2D graph-based soft channel estimation for MIMO-OFDM,” in *Proc. IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010.

[KST04] R. Koetter, A. Singer, and M. Tuechler, “Turbo equalization,” *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 67–80, Jan. 2004.

[KYIG05] K. J. Kim, J. Yue, R. A. Iltis, and J. D. Gibson, “A QRD-M/Kalman filter-based detection and channel estimation algorithm for MIMO-OFDM systems,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 710–721, Mar. 2005.

[Lar09] E. G. Larsson, “MIMO detection methods: How they work,” *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 91–95, May 2009.

[LDH+07] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, “The factor graph approach to model-based signal processing,” *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007.

[LGC88] L. Lyons, D. Gibaut, and P. Clifford, “How to combine correlated estimates of a single physical quantity,” *Nuclear Instruments and Methods in Physics Research*, vol. A270, pp. 110–117, 1988.

[LHG04] I. Land, P. A. Hoeher, and S. Gligorevic, “Computation of symbol-wise mutual information in transmission systems with LogAPP decoders and application to EXIT charts,” in *Proc. of the International ITG Conference on Source and Channel Coding*, Erlangen, Germany, 2004.

[LHL+08] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, “An overview of limited feedback in wireless communication systems,” *IEEE Communications Surveys & Tutorials*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[LHS03] D. J. Love, R. W. Heath, and T. Strohmer, “Grassmannian beamforming for multiple-input multiple-output wireless systems,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[Li00] Y. G. Li, “Optimum training sequences for OFDM systems with multiple transmit antennas,” in *Proc. IEEE Global Communications Conference (GLOBECOM)*, San Francisco, USA, Nov. 2000.

[Li02] ——, “Simplified channel estimation for OFDM systems with multiple transmit antennas,” *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 67–75, Jan. 2002.

[Loe04] H.-A. Loeliger, “An introduction to factor graphs,” *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 28–41, Jan. 2004.

[LR97] X. Li and J. Ritcey, “Bit-interleaved coded modulation with iterative decoding,” *IEEE Communications Letters*, vol. 1, no. 6, pp. 169–171, Nov. 1997.

[LR98] ——, “Bit-interleaved coded modulation with iterative decoding using soft feedback,” *Electronics Letters*, vol. 34, no. 10, pp. 942–943, May 1998.

[LS08] N. Liangfang and D. Sidan, “Evolutionary particle swarm algorithm based on higher-order cumulant fitting for blind channel identification,” in *Proc. on Wireless Communications, Networking and Mobile Computing (WiCOM)*, Dalian, China, Oct. 2008.

[LTDZ02] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, “Combining convergence and diversity in evolutionary multiobjective optimization,” *Journal of Evolutionary Computation*, vol. 10, no. 3, pp. 263–282, Fall 2002.

[LTEM13] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *arXiv:1304.6690*, pp. 1–19, May 2013.

[LWL01] B. Lu, X. Wang, and Y. Li, “Iterative receivers for space-time block coded OFDM systems in dispersive fading channels,” in *Proc. IEEE Global Communications Conference (GLOBECOM)*, San Antonio, USA, Nov. 2001, pp. 514–518.

[Mar10] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[Mat05] G. Matz, “Statistical characterization of non-WSSUS mobile radio channels,” *e & i Elektrotechnik und Informationstechnik*, vol. 122, no. 3, pp. 80–84, 2005.

[May79] P. S. Maybeck, *Stochastic models, estimation, and control.* Academic Press, Inc., 1979, vol. 1, ch. 1, pp. 1–16.

[MDC11] L. Mussi, F. Daolib, and S. Cagnoni, "Evaluation of parallel particle swarm optimization algorithms within the CUDA architecture," *Information Sciences*, vol. 181, no. 20, pp. 4642–4657, Oct. 2011.

[MHC$^+$05] M. Myllyla, J.-H. Hintikka, J. Cavallaro, M. Juntti, M. Limingoja, and A. Byman, "Complexity analysis of MMSE detector architectures for MIMO-OFDM systems," in *Asilomar Conference on Signals, Systems and Computers (ACSSC)*, Pacific Grove, USA, Oct. 2005, pp. 75–81.

[MJ05] H. Miao and M. J. Juntti, "Space-time channel estimation and performance analysis for wireless MIMO-OFDM systems with spatial correlation," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 6, pp. 2003–2016, Nov. 2005.

[MKN03] R. Mendes, J. Kennedy, and J. Neves, "Watch thy neighbor or how the swarm can learn from its environment," in *Proc. IEEE Swarm Intelligence Symposium*, Apr. 2003, pp. 88–94.

[MMS07] A. W. McNaab, C. K. Monson, and K. D. Seppi, "Parallel PSO using mapreduce," in *Proc. of the Congress on Evolutionary Computation*, Brisbane, Australia, Jun. 2007.

[Moo94] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Transactions on Communications*, vol. 42, no. 10, pp. 2908–2914, Oct. 1994.

[MS02] B. Mielczarek and A. Svensson, "Timing error recovery in turbo-coded systems on AWGN channels," *IEEE Transactions on Communications*, vol. 50, no. 10, pp. 1584 – 1592, Oct. 2002.

[MS03] G. Montalbano and D. Slock, "Joint common-dedicated pilots based estimation of time-varying channels for W-CDMA receivers," in *Proc. IEEE Vehicular Technology Conf. (VTC Fall)*, Orlando, USA, Oct. 2003.

[MS05] B. Mielczarek and A. Svensson, "Modeling fading channel-estimation errors in pilot-symbol-assisted systems, with application to turbo codes," *IEEE Transactions on Communications*, vol. 53, no. 11, pp. 1822–1832, Nov. 2005.

[MSL$^+$09] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications - a comprehensive literature survey," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 2, pp. 87–105, 2nd quarter 2009.

[MT03] S. Mostaghim and J. Teich, "The role of epsilon-dominance in multi-objective particle swarm optimization methods," in *IEEE Congress on Evolutionary Computation (CEC)*, Canberra, Australia, Dec. 2003.

[MZCR09] S. K. Mohammed, A. Zaki, A. Chockalingam, and B. S. Raja, “High-rate space-time coded Large-MIMO systems: Low-complexity detection and channel estimation,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 6, pp. 958–974, Dec. 2009.

[NA08] C. Neves and I. F. Alves, “Testing extreme value conditions—an overview and recent approaches,” *RevStat - Statistical Journal*, vol. 6, pp. 83–100, 2008.

[NC98] R. Negi and J. Cioffi, “Pilot tone selection for channel estimation in a mobile OFDM system,” *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 1122–1128, 1998.

[NMH09] C. Novak, G. Matz, and F. Hlawatsch, “Factor graph based design of an OFDM-IDMA receiver performing joint data detection, channel estimation, and channel length selection,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2561–2564.

[NSRL05] H. Niu, M. Shen, J. Ritcey, and H. Liu, “A factor graph approach to iterative channel estimation and LDPC decoding over fading channels,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1345–1350, Jul. 2005.

[Nvi13] Nvidia, “Tesla c2050/c2070 GPU computing processor,” Jul. 2013. [Online]. Available: http://www.nvidia.com/object/tesla-supercomputing-solutions.html

[OA07] M. K. Ozdemir and H. Arslan, “Channel estimation for wireless OFDM systems,” *IEEE Communications Surveys & Tutorials*, vol. 9, no. 2, pp. 18–48, 2nd quarter 2007.

[PCYH09] H. Palally, S. Chen, W. Yao, and L. Hanzo, “Particle swarm optimisation aided semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems,” in *Proc. of the 15th IEEE Workshop on Statistical Signal Processing (SSP ’09)*, Cardiff, United Kingdom, Sep. 2009, pp. 309–312.

[Pea67] M. C. Pease, “Matrix inversion using parallel processing,” *Journal of the Association for Computing Machinery*, vol. 14, no. 4, pp. 757–764, Oct. 1967.

[PLL03] L. Ping, L. Liu, and W. K. Leung, “A simple approach to near-optimal multiuser detection: Interleave-division multiple access,” in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, New Orleans, Louisana, USA, Mar. 2003, pp. 391–396.

[PLL10] H.-J. Park, K.-W. Lee, and Y.-H. Lee, “Channel estimation using dedicated pilot for transmit beamforming in OFDM systems,” in *Proc. IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010.

[PMF00] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 2, pp. 437–447, Mar. 2000.

[Pro00] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.

[PRV96] L. Papke, P. Robertson, and E. Villebrun, "Improved decoding with the SOVA in a parallel concatenated (turbo-code) scheme," in *Proc. IEEE International Conference on Communications (ICC)*, Dallas, TX, USA, Jun. 1996.

[PY07] K. Praditwong and X. Yao, "How well do multi-objective evolutionary algorithms scale to large problems," in *IEEE Congress Evolutionary Computation (CEC)*, Singapore, Singapore, Sep. 2007.

[QJJ08] W. Qiang, Z. Jiashu, and Y. Jing, "Identification of nonlinear communication channel using an novel particle swarm optimization technique," in *Proc. Int. Conf. on Computer Science and Software Engineering*, 2008, pp. 1162–1165.

[RHV97] P. Robertson, P. Hoeher, and E. Villebrun, "Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding," *European Trans. on Telecommun. (ETT)*, vol. 8, no. 2, pp. 119–125, Mar./Apr. 1997.

[RK99] P. Robertson and S. Kaiser, "The effects of doppler spreads in OFDM(A) mobile radio systems," in *Proc. IEEE Vehicular Technology Conf. (VTC Fall)*, Amsterdam, Netherlands, Sep. 1999.

[RPL$^+$13] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[RSC05] M. Reyes-Sierra and C. A. C. Coello, "Improving PSO-based multi-objective optimization using crowding, mutation, and epsilon-dominance," in *Third International Conference on Evolutionary Multi-Criterion Optimization*, Guanajuato, Mexico, Mar. 2005, pp. 505–519.

[RSC06] ——, "Multi-objective particle swarm optimizers: A survey of the state-of-the-art," *International Journal of Computational Intelligence Research*, vol. 2, no. 3, pp. 287–308, 2006.

[RSU01] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity check codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[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 *7th Workshop on Positioning Navigation and Communication (WPNC)*, Dresden, Germany, Mar. 2010.

[SBM+04] G. L. Stueber, J. R. Barry, S. W. McLaughlin, Y. G. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proceedings of the IEEE*, vol. 2, pp. 271–294, 2004.

[SE98] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *Proc. IEEE World Congress on Computational Intelligence*, Anchorage, USA, May 1998.

[SF08] C. Shen and M. P. Fitz, "MIMO-OFDM beamforming for improved channel estimation," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 6, pp. 948–959, Aug. 2008.

[SG08] H. Saligheh and S. Gazor, "The impact of non-iostropic scattering and directional antennas on MIMO multicarrier mobile communication channels," *IEEE Transactions on Communications*, vol. 56, no. 4, pp. 642–652, Apr. 2008.

[Sha48] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Technical J.*, vol. 27, pp. 379–423 and 623–656, Jul./Oct. 1948.

[SJS03] F. Sanzi, S. Jelting, and J. Speidel, "A comparative study of iterative channel estimators for mobile OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 849–859, Sep. 2003.

[Skl97] B. Sklar, "Rayleigh fading channels in mobile digital communication systems part I: Characterization," *IEEE Communications Magazine*, vol. 35, pp. 90–100, Jul. 1997.

[SLC11] O. Schutze, A. Lara, and C. A. C. Coello, "On the influence of the number of objectives on the hardness of a multiobjective optimization problem," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 4, pp. 444–455, Aug. 2011.

[SSAR07] D. P. Shepherd, Z. Shi, M. Anderson, and M. C. Reed, "EXIT chart analysis of an iterative receiver with channel estimation," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Washington, DC, USA, Nov. 2007.

[SSC+07] K. Soo, Y. Siu, W. Chan, L. Yang, and R. Chen, "Particle-swarm-optimization-based multiuser detector for CDMA communications," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 3006–3013, Sep. 2007.

[SSO⁺07] M. Sternad, T. Svensson, T. Ottosson, A. Ahlen, A. Svensson, and A. Brunstrom, “Towards systems beyond 3G based on adaptive OFDMA transmission,” *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2432–2455, Dec. 2007.

[STB09] S. Sesia, I. Toufik, and M. Baker, *LTE The UMTS Long Term Evolution.* John Wiley & Sons Inc., 2009.

[STB11] ——, *LTE - The UMTS Long Term Evolution*, 2nd ed. John Wiley & Sons Inc., 2011.

[SW94] J. Salz and J. Winters, “Effect of fading correlation on adaptive arrays in digital mobile radio,” *IEEE Transactions on Vehicular Technology*, vol. 43, no. 4, pp. 1049–1057, Nov. 1994.

[SW09] R. Schober and R. Wichmann, “MIMO-OFDM channel estimation with eigenbeamforming and user-specific reference signals,” in *Proc. IEEE Vehicular Technology Conf. (VTC Spring)*, Barcelona, Spain, Apr. 2009.

[SZF02] M. Stege, P. Zillmann, and G. Fettweis, “MIMO channel estimation with dimension reduction,” in *Proc. of the 5th Int. Symposium on Wireless Personal Multimedia Communications*, Honolulu, Hawaii, Oct. 2002, pp. 417–421.

[tB00a] S. ten Brink, “Designing iterative decoding schemes with the extrinsic information transfer chart,” *AEÜ International Journal of Electronics and Communications*, vol. 54, no. 6, pp. 389–398, Nov. 2000.

[tB00b] ——, “Iterative decoding trajectories of parallel concatenated codes,” in *3rd IEEE ITG Conf. Source and Channel Coding*, Munich, Germany, Jan. 2000, pp. 75–80.

[tB01] ——, “Convergence behavior of iteratively decoded parallel concatenated codes,” *IEEE Transactions on Communications*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[tBKA04] S. ten Brink, G. Kramer, and A. Ashikhmin, “Design of low-density parity-check codes for modulation and detection,” *IEEE Transactions on Communications*, vol. 52, no. 4, pp. 670–678, Apr. 2004.

[tBSS00] S. ten Brink, F. Sanzi, and J. Speidel, “Two-dimensional iterative APP channel estimation and decoding for OFDM systems,” in *Proc. IEEE Global Communications Conference (GLOBECOM)*, San Francicso, USA, Dec. 2000.

[Tel99] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Trans. on Telecommun. (ETT)*, vol. 10, no. 6, pp. 1–28, Nov./Dec. 1999.

[TH04] M. Tuechler and J. Hagenauer, “Design of serially concatenated systems depending on the block length,” *IEEE Transactions on Communications*, vol. 52, no. 3, pp. 209–218, Feb. 2004.

[The13] The Independent, “Maglev bullet train,” Website, 2013, available online at http://www.independent.co.uk/life-style/gadgets-and-tech/news/japanese-311mph-maglev-bullet-train-tested-successfully-8645369.html; visited on 12. July 2013.

[TJK07] S. Tao, X. Jiadong, and Z. Kai, “Blind MIMO identification using particle swarm algorithm,” in *Proc. Int. Conf. on Wireless Communications, Networking and Mobile Computing (WiCom)*, Shanghai, China, Sep. 2007.

[TM00] M. Toeltsch and A. Molisch, “Efficient OFDM transmission without cyclic prefix over frequency-selective channels,” in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, London, England, Sep. 2000.

[VB99] E. Viterbo and J. Boutros, “A universal lattice code decoder for fading channels,” *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1639–1642, Jul. 1999.

[vdBE04] F. van den Bergh and A. Engelbrecht, “A cooperative approach to particle swarm optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 225–239, Jun. 2004.

[vdBE06] ——, “A study of particle swarm optimization particle trajectories,” *Information Sciences*, vol. 176, no. 8, pp. 937–971, Apr. 2006.

[VMCR08] K. V. Vardhan, S. K. Mohammed, A. Chockalingam, and B. S. Rajan, “A low-complexity detector for large MIMO systems and multicarrier CDMA systems,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 3, pp. 473–485, Apr. 2008.

[VW01] M. C. Valenti and B. D. Woerner, “Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels,” *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1697–1705, Sep. 2001.

[WHS12] T. Wo, P. Hoeher, and Z. Shi, “Graph-based soft channel estimation for fast fading channels,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 12, pp. 4243–4251, Dec. 2012.

[WLH08] T. Wo, C. Liu, and P. A. Hoeher, “Graph-Based Soft Channel and Data Estimation for MIMO Systems with Asymmetric LDPC Codes,” in *Proc. IEEE International Conference on Communications (ICC)*, Beijing, China, 2008, pp. 620–624.

[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.

[WS01] A. P. Worthen and W. E. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 843–849, Feb. 2001.

[XG03] Y. Xie and C. N. Georghiades, "Two EM-type channel estimation algorithms for OFDM with transmitter diversity," *IEEE Transactions on Communications*, vol. 51, no. 1, pp. 106–115, Jan. 2003.

[XM11] X. Xu and R. Mathar, "Low complexity joint channel estimation and decoding for LDPC coded MIMO-OFDM systems," in *Proc. IEEE Vehicular Technology Conf. (VTC Spring)*, Budapest, Hungary, May 2011.

[YFW00] J. Yedidia, W. Freeman, and Y. Weiss, "Generalized belief propagation," *Advances in Neural Information Processing Systems*, vol. 13, pp. 689–695, Dec. 2000.

[YH13] J. Yao and D. Han, "Improved barebones particle swarm optimization with neighborhood search and its application on ship design," *Mathematical Problems in Engineering*, vol. 2013, Article ID 175848, pp. 1–12, 2013.

[YHB04] M. R. Yazdani, S. Hemati, and A. H. Banihashemi, "Improving belief propagation on graphs with cycles," *IEEE Communications Letters*, vol. 8, no. 1, pp. 57–59, Jan. 2204.

[YJ09] J. Ylioinas and M. Juntti, "Iterative joint detection, decoding, and channel estimation in turbo-coded MIMO-OFDM," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1784–1796, May 2009.

[ZGH09] Y. Zhu, D. Guo, and M. L. Honig, "A message-passing approach for joint channel estimation, interference mitigation and decoding," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 6008–6018, Dec. 2009.

[ZL07] K. Zielinksi and R. Laur, "Stopping criteria for a constrained single-objective particle swarm optimization algorithm," *Informatica*, vol. 31, pp. 51–59, 2007.

[ZT03] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[ZZLC09] Z.-H. Zhan, J. Zhang, Y. Li, and H. S.-H. Chung, "Adaptive particle swarm optimization," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 39, no. 6, pp. 227–234, Dec. 2009.