Design and Analysis of Low–Complexity Noncoherent Detection Schemes

Volker Pauli

May 7, 2007

Abstract

Signal detection without the need for channel state information at the receiver, so–called *noncoherent detection*, constitutes an interesting alternative to the widely–used concatenated scheme of channel estimation and subsequent detection, so–called *coherent detection*, in adverse fading channel environments. However, existing approaches to noncoherent detection are either too complex or fail to achieve satisfactory power efficiencies under general fading conditions.

This thesis deals with the design and analysis of power–efficient, yet low–complexity noncoherent detection schemes for point–to–point *multiple–input multiple–output (MIMO)* communication systems. The starting point of this work is *multiple–symbol differential detection (MSDD)*, which simultaneously processes blocks of $N > 2$ received samples to estimate the transmitted data. While MSDD is known to be capable of achieving power efficiencies close to that of coherent detection with perfect channel state information if N is large, it quickly becomes computationally intractable, as the candidate–signal space is $(N-1)$–dimensional, i.e. the number of possible transmit sequences grows exponentially in N.

The application of tree–search algorithms, that have attracted considerable attention in the recent communications literature, to overcome the complexity limitations of MSDD is investigated. Furthermore, a nested MSDD structure consisting of an outer and a number of inner tree–search decoders is developed, which renders MSDD feasible over wide ranges of system parameters.

A second approach to low–complexity MSDD based on methods from combinatorial geometry is proposed for the interesting special cases of *differential phase–shift keying (DPSK)*. This approach is particularly appealing due to the fact, that its complexity is practically constant, whereas tree–search based methods may have very high instantaneous complexities.

Inspired by *decision–feedback differential detection (DFDD)* and the observation that decisions in the different positions of the MSDD observation window are not equally reliable, a new noncoherent detection scheme, referred to as *decision–feedback MSDD (DF–MSDD)* is devised. DF–MSDD achieves power efficiencies comparable to those of MSDD while the dimension of the candidate–signal space and thereby the decoder complexity is reduced significantly. Here, the tree–search methods developed for conventional MSDD are still applicable, such that a computationally highly efficient noncoherent detector results.

Following the development of the detection methods based on a generic MIMO channel model, their application to transmission over time–selective and frequency–nonselective MIMO channels and to transmission using *orthogonal frequency–division multiplexing (OFDM)* over time– and frequency–selective channels is studied. While well–known *differential space–time modulation (DSTM)* is applied for transmission over frequency–nonselective MIMO channels, a new signal allocation scheme for OFDM–based transmission over frequency–selective channels is devised, which makes use of both spatial and / or spectral (multipath) diversity and is particularly apt for power–efficient, low–delay MSDD. For this transmission scenario the application of a two–dimensional observation window to MSDD to exploit channel correlations in both time and frequency direction is investigated.

These practical aspects of this work are complemented by analytical studies regarding the achievable power efficiency and computational complexity of the different detection schemes. These investigations provide interesting insights into the connections between the performances of the different detection schemes and their dependence on system and channel parameters and into the behavior of the decoder complexity as a function of the system and channel parameters, respectively. In consequence, they provide valuable guidelines for quick decoder design and make system simulations largely expendable.

In summary, this work shows how power efficiencies very close to that of idealized coherent detection assuming perfect channel state information can be achieved by means of noncoherent detection with moderate computational complexity, even in adverse fading channel scenarios.

Contents

Notation

For the sake of clarity, we apply the following rules on notation throughout this thesis:

- Lower case symbols in boldface denote vectors, e.g. $\boldsymbol{x}$. Their elements are addressed using subscripts, e.g. x_n denotes the nth entry of the vector $\boldsymbol{x}$. Conversely, an N–dimensional vector can be defined as $\boldsymbol{x} = [x_n]_{n=1,\dots,N}$.

- Upper case symbols in boldface denote matrices, e.g. $\boldsymbol{X}$. Their elements are addressed using subscripts, e.g. $x_{n,m}$ denotes the element in the nth row and mth column of $\boldsymbol{X}$. For simplicity of notation, we at times also denote the element in the nth row and mth column of a matrix $\boldsymbol{X}$ as $[\boldsymbol{X}]_{n,m}$. The nth row and the mth column of $\boldsymbol{X}$ are denoted as $[\boldsymbol{X}]_{n,:}$ and $[\boldsymbol{X}]_{:,m}$, respectively. Conversely, an $(N \times M)$–dimensional matrix can be defined as $\boldsymbol{X} = [x_{n,m}]_{\substack{n=1,\dots,N \\ m=1,\dots,M}}$.

- Upper case symbols in calligraphic typeface / boldface, denote sets of scalars / matrices, respectively, e.g. $\mathcal{X}$ / $\boldsymbol{\mathcal{X}}$.

- To clearly distinguish between continuous– and discrete–time signals, we deploy round and square brackets, respectively, to illustrate their dependence of the respective time variable, e.g. $s(t)$, $t \in \mathbb{R}$ and $s[k]$, $k \in \mathbb{Z}$.

Further, in order to keep the notation as simple as possible, we do not distinguish between random variables and particular realizations thereof. We also adopt the simplified notation that the argument of a probability density function (PDF) corresponds to the random variable. We emphasize that this simplified notation implies that the conditioned PDFs $p(\boldsymbol{x} \,|\, \boldsymbol{y})$ and $p(\boldsymbol{x} \,|\, \boldsymbol{z})$ of $\boldsymbol{x}$ conditioned on $\boldsymbol{y}$ and on $\boldsymbol{z}$, respectively, denote different functions, if $\boldsymbol{y}$ and $\boldsymbol{z}$ represent different random variables.

A list of important symbols and frequently used mnemonics can be found in Appendix A.

Chapter 1

Introduction and Outline

In many practical and especially in wireless communication systems, transmission channels typically are time variant due to oscillator instabilities, phase noise and motion of transmitter, receiver and / or scatterers. In digital communications, there are essentially two approaches to signal detection:

Coherent Detection: Here, the instantaneous channel state is estimated explicitly, usually based on the transmission of pilot symbols (cf. e.g. [MB84, ML89, Cav91, MD97]), and signal detection is subsequently performed using the resulting channel estimate.

Noncoherent Detection: Here, the explicit estimation of the instantaneous channel state is avoided. Instead channel estimation is either performed implicitly in signal detection (cf. e.g. [WFM89, DS90, DSS90, LP91, LP92, HF92, DS94]), or even avoided entirely (cf. e.g. [BN62, Hug00a, HM00]).

While the coherent approach to signal detection based on the (ad hoc) separation of the detection problem into explicit channel estimation and signal detection is most commonly deployed in digital transmission systems, the noncoherent approach appears to be more natural, since the receiver is usually primarily interested in the transmitted information, but not in information about the current state of the channel. Furthermore, noncoherent detection schemes are more robust in rapidly varying transmission scenarios than their coherent counterparts, which rely on the accuracy of the externally obtained channel estimates. They are therefore particularly apt for (i) discontinuous transmission, where coherent transmission would require a relatively large portion of pilot symbols for accurate channel estimation, (ii) systems with low–cost high–frequency components, where e.g. strong fluctuations of phase and frequency of local oscillators may occur, and (iii) systems with time–variant interference, e.g. if transmission takes place in unlicensed frequency bands such as the "industrial, scientific and medical" (ISM) bands, where channel estimation would have to be repeated frequently.

The reason for the popularity of coherent schemes lies mainly in the fact that existing noncoherent detection schemes are characterized either by poor power efficiency or by high computational complexity for high power efficiency. In order to achieve good power efficiency by means of noncoherent detection simultaneous processing of blocks of N received samples based on knowledge of the *statistical* properties, such as power spectral density, power–delay profile and noise variance, of the channel are required. So, given a transmit–symbol alphabet of size L such block–based noncoherent detection involves finding the best out of L^N possible transmit sequences, whereas coherent detection with external channel estimation can be performed on a symbol–by–symbol basis. Such block–based noncoherent detectors are commonly referred to as *multiple–symbol (differential) detectors (MS(D)D)* and the block length N is called the *observation window length*, cf. e.g. [WFM89, DS90, DSS90, LP91, LP92, HF92, DS94, VCBT97].

Ever since the introduction of MSDD the search for noncoherent detectors that achieve a comparable power efficiency at manageable computational complexity has been a subject of continued research in the communications community. While Mackenthun proposed an MSDD algorithm whose complexity is of the order $N\log(N)$ for the time–invariant channel in [Mac94], other authors developed algorithms based on restricted tree search, cf. e.g. [AS89, LW90], or reduced–state Viterbi decoding, cf. e.g. [MF90, LM90, YP95, VT95a, Rap96a, Ada96, CR99]. In particular the complete reduction of the states in the trellis has lead to popular decision–feedback differential detection (DFDD), cf. e.g. [LP88, Edb92, AS93, AS95, SGH99, SL02]. However, these schemes leave room for improvement regarding the performance–complexity tradeoff and the introduction of decision–feedback strategies impedes the application of such detectors in coded transmission systems.

In this thesis, we deal with the design and analysis of noncoherent detection schemes, whose power efficiencies are close to that of idealized coherent detection with perfect instantaneous channel state information (CSI), while their complexities are very well comparable to those of existing less power–efficient schemes such as DFDD. In particular, we consider highly time– and / or frequency–selective multiple–input multiple–output (MIMO) channels, where coherent detection becomes unattractive, as pilot symbols, which are required for frequent estimation of a large number of channel coefficients, consume a substantial portion of the ever more precious commodities signaling bandwidth and transmit power, cf. e.g. [Mar99].

Chapter 2 introduces a generic MIMO system model that shall serve as basis for our considerations on computationally and power–efficient noncoherent detectors in Chapter 3. In particular, we review a number of important unitary–matrix signal constellations, that were originally developed for differential space–time modulation (DSTM) and are designed to exploit the diversity provided by the MIMO channel, while facilitating noncoherent detection. Following the introduction of DSTM we review a number of standard coherent and noncoherent detection strategies considered in this work. Based on the identification of a relationship

between MSDD and linear minimum–mean squared error (MMSE) interpolation novel variants of MSDD are introduced that will be shown to achieve excellent complexity–performance tradeoffs.

Chapter 3 is dedicated to the development of computationally highly efficient algorithms to (approximately) solve the MSDD problem. Here, we consider two different approaches: The first one makes use of tree–search algorithms, that were developed in the context of sequential decoding of convolutional codes and in lattice theory. Based on a revision and classification of the various tree–search methods, two promising algorithms are selected for application to MSDD with arbitrary DSTM constellations. Further significant complexity savings are obtained through optimization of the receiver structures for individual DSTM constellations. The second approach is based on methods from combinatorial geometry and is applicable to single–transmit antenna schemes employing differential phase–shift keying (DPSK) or simple repetition transmit diversity schemes in conjunction with possibly multiple receive antennas.

In Chapter 4, we present a general time–variant MIMO channel model for a system with N_{T} transmit and N_{R} receive antennas. This model serves as basis for our considerations on detection for transmission over frequency–nonselective and frequency–selective MIMO channels in Chapters 5 and 6, respectively.[1]

In Chapter 5 we relate the generic system model of Chapter 2 to a DSTM–based system for transmission over a frequency–nonselective channel, which is derived from the general channel model of Chapter 4. Here, we furthermore present the results of our in–depth analytical investigations regarding both the computational complexity and the achievable performance of the novel noncoherent detection schemes of Chapter 3 when applied to DSTM–based transmission over frequency–nonselective channels.

In Chapter 6 transmission over frequency–selective MIMO channels is considered, and we investigate the usefulness of our methods in a system using orthogonal frequency division multiplexing (OFDM). We also develop a novel signal–allocation scheme, that allows us to benefit from both spatial and spectral diversity. For power–efficient noncoherent detection we propose to use MSDD with a two–dimensional observation window, such that correlation in both time and frequency can be exploited in the detection process.

Chapter 7 summarizes the contributions of this thesis and provides a brief outlook to possible directions for further related research.

Parts of the material presented in this work have been published in [PL05, PL07b, PL06, PLH07b, PLSF06, PSL06, PL07a, PLH07a, PLSF07].

Finally, we would like to mention that in this work we concentrate on transmission without forward error correction (FEC). However, extension to FEC coded transmission can be ac-

[1]Note that all results presented in this work trivially include popular DPSK with or without receive diversity as special case with $N_{\mathrm{T}} = 1$. This case is therefore not treated separately in this work.

complished by transforming the proposed detectors into soft–output, or soft–input soft–output modules for iterative decoding, as shown in e.g. [PLS06] for the special case of DPSK transmission.

Chapter 2

Differentially Encoded Space–Time Transmission and Detection

In this section, we introduce a generic discrete–time multiple–input multiple–output (MIMO) system model that will serve as basis for our considerations on low–complexity noncoherent detection strategies in Chapter 3. This model is based on unitary–matrix signal constellations, that were originally developed for differential space–time modulation (DSTM) and will be reviewed in Section 2.1, together with differential encoding as discussed in Section 2.2. The MIMO channel model will be introduced in Section 2.3. In Section 2.4, we will describe the detection schemes considered in this work and present their respective decision rules.

At this, it is our full intention to remain fairly abstract, e.g. not relating the inputs and outputs of the MIMO channel to transmit and receive antennas, respectively, in order to simultaneously cover the common aspects of the different transmission scenarios considered in Chapters 5 and 6. While the MIMO channel model is introduced in a rather ad–hoc way in this chapter and contains some assumptions made by the detection schemes, it will be related to a physically motivated channel model in Chapter 4. The rather abstract variables introduced in this chapter and Chapter 3 will be filled with life in Chapters 5 and 6, where we will consider single–carrier transmission over frequency non–selective channels using DSTM and transmission over frequency selective channels based on orthogonal frequency division multiplexing (OFDM), respectively.

All of our considerations are set in the equivalent complex baseband (ECB) domain [Tre71], i.e. all quantities involved in the channel model are in general represented by complex–valued variables. In particular, we employ an energy–invariant baseband transformation, cf. e.g. [Tre71].

2.1 Signal Constellations

Differential space–time modulation (DSTM) using signal constellations

$$\boldsymbol{\mathcal{V}} \stackrel{\triangle}{=} \left\{ \boldsymbol{V}^{(l)} \mid l \in \{1, \ldots, L\} \right\} \tag{2.1}$$

($\stackrel{\triangle}{=}$: definition) consisting of L ($N_{\mathrm{S}} \times N_{\mathrm{S}}$)–dimensional *unitary* matrices $\boldsymbol{V}^{(l)}$, $1 \leq l \leq L$, has been introduced in [TJC99, Hug00a, HS00, HH02].[1] Ever since, the design and analysis of good constellations has been a field of very active research and various authors have presented results on numerous signal constellations in e.g. [Hug99a, Hug99b, MHH00, ARU00, HM00, HMR$^+$00, CMH00, Hug00b, LX00, HHSS00, SHHS01, ARU01, LX01, HM01, Sho01b, Sho01a, LLLL01a, LLLL01b, TC01, GS02, DADSC02, LX02, MHH02, Hug03, SY03, HSL05a, WWM05]. These can be coarsely classified into two classes: (i) so–called group codes, where the set $\boldsymbol{\mathcal{V}}$ forms a group with respect to matrix multiplication and (ii) so–called non–group codes, where this is not the case, i.e. where the product of two elements $\boldsymbol{V}^{(l_1)}$ and $\boldsymbol{V}^{(l_2)}$ of $\boldsymbol{\mathcal{V}}$ does not necessarily result in an element from $\boldsymbol{\mathcal{V}}$. Given this abundance of signal constellations, we chose to restrict our attention to the four important representatives, reviewed briefly in the following.

2.1.1 Constellations from Group Codes

2.1.1.1 Cyclic (Diagonal) Codes

Cyclic DSTM codes (also often referred to as *diagonal* DSTM codes) were originally proposed independently in [Hug00a] and [HS00]. Here, $\boldsymbol{\mathcal{V}}$ is defined as

$$\boldsymbol{\mathcal{V}} \stackrel{\triangle}{=} \left\{ \boldsymbol{V}^{(l)} = \mathrm{diag}\left\{ \mathrm{e}^{\mathrm{j}\frac{2\pi}{L}c_1}, \ldots, \mathrm{e}^{\mathrm{j}\frac{2\pi}{L}c_{N_{\mathrm{S}}}} \right\}^l \,\middle|\, l \in \{1, \ldots, L\} \right\} \tag{2.2}$$

($\mathrm{diag}\{x_1, \ldots, x_M\}$: $(M \times M)$–dimensional diagonal matrix with the x_i, $1 \leq i \leq M$, on its main diagonal, e^x, $\exp(x)$: exponential function, $\mathrm{j} \stackrel{\triangle}{=} \sqrt{-1}$: imaginary unit). The integer coefficients $c_i \in \{1, \ldots, L-1\}$, $1 \leq i \leq N_{\mathrm{S}}$, are optimized with respect to asymptotic[2] power efficiency of conventional differential detection (CDD, cf. Section 2.4.1) under the assumption that the channel does not change significantly over time. Optimized coefficients for various values of N_{S} can be found in e.g. [HS00, Table I]. An example of a cyclic DSTM constellation with $N_{\mathrm{S}} = 2$ and $L = 16$ designed in [HS00] is given by the set

$$\boldsymbol{\mathcal{V}} = \left\{ \boldsymbol{V}^{(l)} = \begin{bmatrix} \mathrm{e}^{\mathrm{j}\frac{\pi}{8}} & 0 \\ 0 & \mathrm{e}^{\mathrm{j}\frac{3\pi}{8}} \end{bmatrix}^l \,\middle|\, l \in \{1, \ldots, 16\} \right\} \tag{2.3}$$

[1] Many authors also refer to DSTM based on unitary–matrix signal constellations as unitary DSTM (UDSTM).

[2] "Asymptotic" in the sense of high signal–to–noise ratio (SNR).

of unitary matrices.

Note that differential phase–shift keying (DPSK, cf. e.g. [Pro00]) can be viewed as special case of cyclic DSTM with $N_{\mathrm{S}} = 1$ and $c_1 = 1$. Therefore, all results presented in this work are in an obvious way also applicable to DPSK.

2.1.1.2 Dicyclic Codes

For N_{S} even, dicyclic DSTM codes (also referred to as *generalized quaternion* DSTM codes) were originally proposed in [Hug03]. Here, an $(L/2)$–ary variable l and a binary variable m select an unitary matrix $\boldsymbol{V}^{(l+(m-1)L/2)}$ from the set $\boldsymbol{\mathcal{V}}$ of cardinality $|\boldsymbol{\mathcal{V}}| = L$ defined as

$$\boldsymbol{\mathcal{V}} \triangleq \left\{ \boldsymbol{V}^{(l+(m-1)\frac{L}{2})} = \mathrm{diag}\left\{ \mathrm{e}^{\mathrm{j}\frac{4\pi}{L}c_1}, \ldots, \mathrm{e}^{\mathrm{j}\frac{4\pi}{L}c_{N_{\mathrm{S}}}} \right\}^l \cdot \begin{bmatrix} \boldsymbol{0}_{N_{\mathrm{S}}/2,N_{\mathrm{S}}/2} & -\boldsymbol{I}_{N_{\mathrm{S}}/2} \\ \boldsymbol{I}_{N_{\mathrm{S}}/2} & \boldsymbol{0}_{N_{\mathrm{S}}/2,N_{\mathrm{S}}/2} \end{bmatrix}^m \middle| \begin{array}{l} l \in \{1, \ldots, L/2\} \\ m \in \{0,1\} \end{array} \right\}, \tag{2.4}$$

where $c_{N_{\mathrm{S}}/2+l} = -c_l$, $1 \leq l \leq N_{\mathrm{S}}/2$, and $\boldsymbol{X}^0 = \boldsymbol{I}_N$ for any $(N \times N)$–dimensional matrix $\boldsymbol{X}$ ($\boldsymbol{I}_N$: $(N \times N)$–dimensional identity matrix, $\boldsymbol{0}_{M,N}$: $(M \times N)$–dimensional all–zeros matrix). Integer coefficients $c_l \in \{1, \ldots, L/2-1\}$ optimized with respect to asymptotic power efficiency under CDD in very–slow–fading high–SNR regimes are listed in [Hug03, Table III]. An example of a dicyclic DSTM constellation with $N_{\mathrm{S}} = 4$ and $L = 16$ designed in [Hug03] is given by the set

$$\boldsymbol{\mathcal{V}} = \left\{ \boldsymbol{V}^{(l+8(m-1))} = \begin{bmatrix} \mathrm{e}^{\mathrm{j}\frac{\pi}{8}} & 0 & 0 & 0 \\ 0 & \mathrm{e}^{\mathrm{j}\frac{3\pi}{8}} & 0 & 0 \\ 0 & 0 & \mathrm{e}^{-\mathrm{j}\frac{\pi}{8}} & 0 \\ 0 & 0 & 0 & \mathrm{e}^{-\mathrm{j}\frac{3\pi}{8}} \end{bmatrix}^l \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}^m \middle| \begin{array}{l} l \in \{1, \ldots, 8\} \\ m \in \{0,1\} \end{array} \right\} \tag{2.5}$$

of unitary matrices.

Note that for L being a power of two, every full–rank group DSTM code, i.e. every unitary DSTM code that achieves full transmit diversity, is equivalent to either a cyclic or a dicyclic DSTM code, cf. [Hug03, Theorem 2].

2.1.2 Constellations from Non–Group Codes

2.1.2.1 Orthogonal Codes

A DSTM scheme that is frequently considered in the literature is that of orthogonal DSTM codes originally proposed in [TJC99]. While it is straightforward to extend our subsequently presented results regarding orthogonal codes to $N_{\mathrm{S}} > 2$, we will restrict our attention to the interesting

special case of $N_\mathrm{S} = 2$, where the orthogonal codes of [TJC99] coincide with Alamouti's code [Ala98]. Here, the set $\boldsymbol{\mathcal{V}}$ is defined as

$$\boldsymbol{\mathcal{V}} \triangleq \left\{ \boldsymbol{V}^{(l)} = \frac{1}{\sqrt{2}} \begin{bmatrix} a & -b^* \\ b & a^* \end{bmatrix} \middle| a, b \in \left\{ 1, \mathrm{e}^{\mathrm{j}\frac{2\pi}{\sqrt{L}}}, \ldots, \mathrm{e}^{\mathrm{j}\frac{2\pi(\sqrt{L}-1)}{\sqrt{L}}} \right\} \right\}. \tag{2.6}$$

($\cdot^*$: (elementwise) complex conjugate)

2.1.2.2 Cayley Codes

As second important representative of non–group codes we consider Cayley codes [HH02, WWM05], which are particularly apt for high–rate data transmission. Here, the transmit data is split into Q parallel streams and mapped to Q real–valued coefficients α_q, $1 \leq q \leq Q$, from a $\sqrt[Q]{L}$–ary set $\mathcal{A}_\mathrm{cay}$. Together with Q predefined Hermitian symmetric ($N_\mathrm{S} \times N_\mathrm{S}$)–dimensional "basis matrices" $\boldsymbol{A}_q$ they define L Hermitian symmetric matrices

$$\boldsymbol{A}_\mathrm{cay}^{(l)} \triangleq \sum_{q=1}^{Q} \alpha_q \boldsymbol{A}_q. \tag{2.7}$$

Applying the Cayley transform to the skew–Hermitian symmetric matrices $\mathrm{j}\boldsymbol{A}_\mathrm{cay}^{(l)}$ leads to the L–ary unitary DSTM constellation

$$\boldsymbol{\mathcal{V}} \triangleq \left\{ \boldsymbol{V}^{(l)} = \left(\boldsymbol{I}_{N_\mathrm{S}} + \mathrm{j}\boldsymbol{A}_\mathrm{cay}^{(l)} \right)^{-1} \left(\boldsymbol{I}_{N_\mathrm{S}} - \mathrm{j}\boldsymbol{A}_\mathrm{cay}^{(l)} \right) \middle| l \in \{1, \ldots, L\} \right\}. \tag{2.8}$$

An example of a 4096–ary constellation with $N_\mathrm{S} = 2$ for which we will show numerical results in later chapters is specified by the set of coefficients (cf. [HH02])

$$\mathcal{A}_\mathrm{cay} = \{\pm 5.0273, \pm 1.4966, \pm 0.6682, \pm 0.1989\} \tag{2.9}$$

and basis matrices

$$\boldsymbol{A}_1 = \begin{bmatrix} 0.1785 & 0.0510 + \mathrm{j}0.1340 \\ 0.0510 - \mathrm{j}0.1340 & 0.0321 \end{bmatrix} \quad \boldsymbol{A}_2 = \begin{bmatrix} -0.1902 & 0.1230 + \mathrm{j}0.0495 \\ 0.1230 - \mathrm{j}0.0495 & -0.0512 \end{bmatrix}$$
$$\boldsymbol{A}_3 = \begin{bmatrix} -0.2350 & 0.0515 - \mathrm{j}0.0139 \\ 0.0515 + \mathrm{j}0.0139 & 0.1142 \end{bmatrix} \quad \boldsymbol{A}_4 = \begin{bmatrix} 0.0208 & 0.1143 - \mathrm{j}0.1532 \\ 0.1143 + \mathrm{j}0.1532 & 0.0220 \end{bmatrix}. \tag{2.10}$$

2.2 Differential Encoding

As we consider detection without instantaneous channel state information (CSI), i.e. without knowledge of the current state of the channel at the receiver, it is not possible to successfully

transmit information by mapping blocks of $\log_2(L)$ bits to symbols $\boldsymbol{V}[k] \in \boldsymbol{\mathcal{V}}$ and transmit them directly. As a remedy, a differential encoding technique for unitary–matrix signal constellations (cf. Section 2.1) was proposed in [HS00, Hug00a]. It can be viewed as extension of classical DPSK to unitary matrices. Here, blocks of $\log_2(L)$ bits are mapped to matrix *data* symbols $\boldsymbol{V}[k] \in \boldsymbol{\mathcal{V}}$, which are differentially encoded into matrix *transmit* symbols $\boldsymbol{S}[k]$ via

$$\boldsymbol{S}[k] = \boldsymbol{V}[k]\boldsymbol{S}[k-1], \quad \boldsymbol{S}[0] = \boldsymbol{I}_{N_\mathrm{S}}. \tag{2.11}$$

Consequently, the information to be transmitted is contained in the "differences" between successive transmit symbols $\boldsymbol{S}[k]$ and can —provided that the channels over which $\boldsymbol{S}[k]$ and $\boldsymbol{S}[k-1]$ are transmitted are not statistically independent— be recovered without the need for CSI at the receiver.

While the meaning of k may vary depending on the application derived from this generic system model, we will in the following refer to it as "time index" of matrix symbols.

2.3 Generic Channel Model

The detection algorithms presented in Section 3 are based on the following generic "temporally" correlated MIMO Rayleigh–fading channel model. The received signal $\boldsymbol{R}[k]$ corresponding to the transmission of an $(N_\mathrm{S} \times N_\mathrm{S})$–dimensional transmit symbol $\boldsymbol{S}[k]$ is organized in an $(N_\mathrm{S} \times N_\mathrm{R})$–dimensional matrix

$$\boldsymbol{R}[k] = \boldsymbol{S}[k]\boldsymbol{G}[k] + \boldsymbol{N}[k]. \tag{2.12}$$

Here, the elements $g_{i,j}[k]$ in the ith row and jth column of $(N_\mathrm{S} \times N_\mathrm{R})$–dimensional $\boldsymbol{G}[k]$ representing the MIMO channel are modeled as independent identically distributed (iid) Rayleigh fading processes, i.e. as iid $\mathcal{N}_\mathrm{c}(0,1)$ random variables ($\mathcal{N}_\mathrm{c}(m,\sigma^2)$: circularly symmetric complex Gaussian distribution with mean m and variance σ^2). Their temporal correlation is described by

$$\mathcal{E}\big\{g_{i_1,j_1}[k+\kappa]g^*_{i_2,j_2}[k]\big\} = \begin{cases} \psi_{gg}[\kappa] & i_1 = i_2, j_1 = j_2 \\ 0 & \text{otherwise} \end{cases}, \tag{2.13}$$

($\mathcal{E}_x\{f(x)\}$: expectation of $f(x)$ with respect to random variable x, where the subscript $_x$ is omitted whenever possible). It models the temporally correlated non–amplifying channel, which is assumed to be constant during the transmission of $\boldsymbol{S}[k]$. A remark on denomination: Interpreting k as "time index" we speak of (i) "rapid", (ii) "slow" or (iii) "static" fading or time–variance of the fading process, if $\boldsymbol{G}[k]$ statistically changes (i) significantly, (ii) hardly or (iii) not at all from matrix symbol to matrix symbol.

$(N_\mathrm{S} \times N_\mathrm{R})$–dimensional $\boldsymbol{N}[k]$ contains iid temporally uncorrelated $\mathcal{N}_\mathrm{c}(0, \sigma_n^2)$ random variables $n_{i,j}[k]$, i.e.

$$\mathcal{E}\{n_{i_1,j_1}[k+\kappa] n^*_{i_2,j_2}[k]\} = \begin{cases} \sigma_n^2 & i_1 = i_2, j_1 = j_2, \kappa = 0 \\ 0 & \text{otherwise} \end{cases} . \tag{2.14}$$

It models the additive white Gaussian noise (AWGN) perturbing the reception of $\boldsymbol{S}[k]$.

Note that the organization of the received signal corresponding to the transmission of $\boldsymbol{S}[k]$ in an $(N_\mathrm{S} \times N_\mathrm{R})$–dimensional matrix rather than in e.g. a vector of length $N_\mathrm{S} N_\mathrm{R}$ is in principle arbitrary. It follows from considerations on simplicity of notation and from intuitive arguments regarding the interesting special case considered in Chapter 5, where $(N_\mathrm{S} \times N_\mathrm{R})$–dimensional $\boldsymbol{G}[k]$ represents a time–variant spatially uncorrelated MIMO channel in a transmission system employing $N_\mathrm{T} = N_\mathrm{S}$ transmit and N_R receive antennas.

2.4 Differential Detection

At the receiver various known detection schemes for noncoherent communication are considered. For completeness, we briefly review them in the following.

2.4.1 Conventional Differential Detection (CDD)

Conventional differential detection (CDD) is most often considered in the abovementioned publications on DSTM constellations. It is based on the assumption that temporal variations between two successive channel realizations are negligible, i.e. that $\boldsymbol{G}[k] = \boldsymbol{G}[k-1]$. On the basis of this assumption and the Rayleigh–fading channel model of Section 2.3 the maximum–likelihood (ML) decision rule for CDD can be deduced from the conditional probability density function (PDF) (cf. e.g. [Mil74])

$$\begin{aligned} &p(\boldsymbol{R}[k-1], \boldsymbol{R}[k] \,|\, \boldsymbol{S}[k-1], \boldsymbol{S}[k]) = \\ &\frac{1}{((\pi N_\mathrm{R})^2 \sigma_n^2 (2+\sigma_n^2))^{N_\mathrm{S} N_\mathrm{R}}} \cdot \exp\left(-\frac{1}{N_\mathrm{R}} \mathrm{tr}\left\{ \left[\boldsymbol{R}^\mathsf{H}[k-1], \boldsymbol{R}^\mathsf{H}[k]\right] \begin{bmatrix} \boldsymbol{S}[k-1] & \boldsymbol{0}_{N_\mathrm{S},N_\mathrm{S}} \\ \boldsymbol{0}_{N_\mathrm{S},N_\mathrm{S}} & \boldsymbol{S}[k] \end{bmatrix} \right.\right. \\ &\left.\left. \left(\begin{bmatrix} 1+\sigma_n^2 & 1 \\ 1 & 1+\sigma_n^2 \end{bmatrix} \otimes \boldsymbol{I}_{N_\mathrm{S}} \right)^{-1} \begin{bmatrix} \boldsymbol{S}^\mathsf{H}[k-1] & \boldsymbol{0}_{N_\mathrm{S},N_\mathrm{S}} \\ \boldsymbol{0}_{N_\mathrm{S},N_\mathrm{S}} & \boldsymbol{S}^\mathsf{H}[k] \end{bmatrix} \begin{bmatrix} \boldsymbol{R}[k-1] \\ \boldsymbol{R}[k] \end{bmatrix} \right\} \right) \end{aligned} \tag{2.15}$$

($\cdot^\mathsf{H}$: Hermitian transposition, $\otimes$: Kronecker product (cf. e.g. [HJ91]), and $\mathrm{tr}\{\cdot\}$: trace). Inserting (2.11) into (2.15) and recalling that $\boldsymbol{S}[k-1]$ is unitary one can see that $p(\boldsymbol{R}[k-1], \boldsymbol{R}[k] \,|$

$\boldsymbol{S}[k-1], \boldsymbol{S}[k])$ depends only on $\boldsymbol{V}[k]$ and we can write

$$p(\boldsymbol{R}[k-1], \boldsymbol{R}[k] \,|\, \boldsymbol{S}[k-1], \boldsymbol{S}[k]) = p(\boldsymbol{R}[k-1], \boldsymbol{R}[k] \,|\, \boldsymbol{V}[k]) = \tag{2.16}$$
$$\frac{1}{((\pi N_{\mathrm{R}})^2 \sigma_n^2 (2+\sigma_n^2))^{N_{\mathrm{S}} N_{\mathrm{R}}}} \cdot \exp\left(-\frac{1}{N_{\mathrm{R}}} \mathrm{tr}\left\{ \left[\boldsymbol{R}^{\mathsf{H}}[k-1], \boldsymbol{R}^{\mathsf{H}}[k]\right] \begin{bmatrix} \boldsymbol{I}_{N_{\mathrm{S}}} & \boldsymbol{0}_{N_{\mathrm{S}},N_{\mathrm{S}}} \\ \boldsymbol{0}_{N_{\mathrm{S}},N_{\mathrm{S}}} & \boldsymbol{V}[k] \end{bmatrix} \right.\right.$$
$$\left.\left. \left(\begin{bmatrix} 1+\sigma_n^2 & 1 \\ 1 & 1+\sigma_n^2 \end{bmatrix}^{-1} \otimes \boldsymbol{I}_{N_{\mathrm{S}}} \right) \begin{bmatrix} \boldsymbol{I}_{N_{\mathrm{S}}} & \boldsymbol{0}_{N_{\mathrm{S}},N_{\mathrm{S}}} \\ \boldsymbol{0}_{N_{\mathrm{S}},N_{\mathrm{S}}} & \boldsymbol{V}^{\mathsf{H}}[k] \end{bmatrix} \begin{bmatrix} \boldsymbol{R}[k-1] \\ \boldsymbol{R}[k] \end{bmatrix} \right\} \right)$$

and averaging with respect to $\boldsymbol{S}[k-1]$ is not required. Using the identity $(\boldsymbol{1}_{N,N} + x\boldsymbol{I}_N)^{-1} = \frac{1}{x}(\boldsymbol{I}_N - \frac{1}{N+x}\boldsymbol{1}_{N,N})$ ($\boldsymbol{1}_{N,N}$: $(N \times N)$–dimensional all–ones matrix), taking into account that $\exp(\cdot)$ is a monotonous function and neglecting all terms that do not depend on $\boldsymbol{V}[k]$ one obtains the ML decision rule for estimation of the data symbol $\boldsymbol{V}[k]$ as (cf. e.g. [Hug00a])

$$\hat{\boldsymbol{V}}[k] = \underset{\tilde{\boldsymbol{V}}[k] \in \boldsymbol{\mathcal{V}}}{\mathrm{argmax}} \Big\{ \mathrm{Re}\left\{ \mathrm{tr}\Big\{ \tilde{\boldsymbol{V}}[k] \boldsymbol{R}[k-1] \boldsymbol{R}^{\mathsf{H}}[k] \Big\} \right\} \Big\} \tag{2.17}$$

($\mathrm{argmax}_{x \in \mathcal{X}}\{f(x)\}$: returns that element of a set $\mathcal{X}$ that maximizes the function $f(x)$, $\mathrm{Re}\,\{\cdot\}$, $\mathrm{Im}\,\{\cdot\}$: real, imaginary part).

While CDD is capable of achieving reasonable performance when $\boldsymbol{G}[k] = \boldsymbol{G}[k-1]$ (cf. e.g. [HS00]), growing temporal variations of the channel lead to poorer power efficiency of CDD reflecting in an increasingly high error floor at high SNR, cf. [PS03, DB06]. To overcome these limitations more sophisticated detection schemes such as multiple–symbol differential detection (MSDD, cf. Section 2.4.2) or low–complexity derivates thereof such as decision–feedback differential detection (DFDD, cf. Section 2.4.3) or the algorithms described in Chapter 3 are required.

2.4.2 Multiple–Symbol Differential Detection (MSDD)

MSDD processes blocks of N successively received symbols $\boldsymbol{R}[k-\kappa]$, $0 \leq \kappa \leq N-1$, to detect the corresponding N transmit symbols $\boldsymbol{S}[k-\kappa]$, $0 \leq \kappa \leq N-1$, or equivalently by reversal of (2.11) $N-1$ differentially encoded data symbols $\boldsymbol{V}[k-\kappa]$, $0 \leq \kappa \leq N-2$, cf. e.g. [WFM89, DS90, DSS90, LP91, LP92, HF92, DS94, KL94]. At this, N is referred to as *observation window length*. Let us collect the data, transmit and received symbols and the corresponding channel matrices $\boldsymbol{G}[k]$ and noise matrices $\boldsymbol{N}[k]$ in respective block matrices $\bar{\boldsymbol{V}}$, $\bar{\boldsymbol{S}}$, $\bar{\boldsymbol{G}}$, $\bar{\boldsymbol{N}}$ and $\bar{\boldsymbol{R}}$, and define the $((N-1)N_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional matrix

$$\bar{\boldsymbol{V}}[k] \;\triangleq\; \left[\boldsymbol{V}^{\mathsf{H}}[k-N+2], \ldots, \boldsymbol{V}^{\mathsf{H}}[k]\right]^{\mathsf{H}} \tag{2.18}$$

the $(NN_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional matrices

$$\bar{\boldsymbol{S}}[k] \;\triangleq\; \left[\boldsymbol{S}^{\mathsf{H}}[k-N+1], \ldots, \boldsymbol{S}^{\mathsf{H}}[k]\right]^{\mathsf{H}} \tag{2.19}$$

$(NN_{\mathrm{S}} \times N_{\mathrm{R}})$–dimensional

$$\bar{\boldsymbol{G}}[k] \triangleq \left[\boldsymbol{G}^{\mathsf{H}}[k-N+1],\ldots,\boldsymbol{G}^{\mathsf{H}}[k]\right]^{\mathsf{H}} \tag{2.20}$$

$$\bar{\boldsymbol{N}}[k] \triangleq \left[\boldsymbol{N}^{\mathsf{H}}[k-N+1],\ldots,\boldsymbol{N}^{\mathsf{H}}[k]\right]^{\mathsf{H}} \tag{2.21}$$

$$\bar{\boldsymbol{R}}[k] \triangleq \left[\boldsymbol{R}^{\mathsf{H}}[k-N+1],\ldots,\boldsymbol{R}^{\mathsf{H}}[k]\right]^{\mathsf{H}} \tag{2.22}$$

and the $(NN_{\mathrm{S}} \times NN_{\mathrm{S}})$–dimensional unitary block–diagonal matrix

$$\bar{\boldsymbol{S}}_{\mathrm{D}}[k] \triangleq \mathrm{diag}\{\bar{\boldsymbol{S}}[k]\} = \begin{bmatrix} \boldsymbol{S}[k-N+1] & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{S}[k] \end{bmatrix}. \tag{2.23}$$

With this notation, the channel can be described by

$$\bar{\boldsymbol{R}}[k] = \bar{\boldsymbol{S}}_{\mathrm{D}}[k]\bar{\boldsymbol{G}}[k] + \bar{\boldsymbol{N}}[k]. \tag{2.24}$$

For the sake of readability, we will in the following drop the reference $[k]$ to time wherever possible.

Based on the above Rayleigh–fading channel model the corresponding conditional PDF $p\big(\bar{\boldsymbol{R}} \,\big|\, \bar{\boldsymbol{S}}\big)$ is given by (cf. e.g. [Mil74])

$$p\big(\bar{\boldsymbol{R}} \,\big|\, \bar{\boldsymbol{S}}\big) = \frac{1}{\det\{\pi \boldsymbol{\Psi}_{\bar{\boldsymbol{R}}\bar{\boldsymbol{R}}|\bar{\boldsymbol{S}}}\}^{N_{\mathrm{R}}}} \cdot \exp\Big(-\mathrm{tr}\Big\{\bar{\boldsymbol{R}}^{\mathsf{H}} \boldsymbol{\Psi}^{-1}_{\bar{\boldsymbol{R}}\bar{\boldsymbol{R}}|\bar{\boldsymbol{S}}} \bar{\boldsymbol{R}}\Big\}\Big) \tag{2.25}$$

($\det\{\cdot\}$: determinant). For the autocorrelation matrix $\boldsymbol{\Psi}_{\bar{\boldsymbol{R}}\bar{\boldsymbol{R}}|\bar{\boldsymbol{S}}}$ of $\bar{\boldsymbol{R}}$ conditioned on $\bar{\boldsymbol{S}}$ we can exploit the independence of zero–mean $\bar{\boldsymbol{G}}$ and $\bar{\boldsymbol{N}}$ and write

$$\boldsymbol{\Psi}_{\bar{\boldsymbol{R}}\bar{\boldsymbol{R}}|\bar{\boldsymbol{S}}} \triangleq \mathcal{E}\big\{\bar{\boldsymbol{R}}\bar{\boldsymbol{R}}^{\mathsf{H}} \,\big|\, \bar{\boldsymbol{S}}\big\} \tag{2.26}$$

$$= \mathcal{E}\Big\{\big(\bar{\boldsymbol{S}}_{\mathrm{D}}\bar{\boldsymbol{G}} + \bar{\boldsymbol{N}}\big) \cdot \Big(\bar{\boldsymbol{G}}^{\mathsf{H}}\bar{\boldsymbol{S}}_{\mathrm{D}}^{\mathsf{H}} + \bar{\boldsymbol{N}}^{\mathsf{H}}\Big)\Big\} \tag{2.27}$$

$$= \bar{\boldsymbol{S}}_{\mathrm{D}}\mathcal{E}\big\{\bar{\boldsymbol{G}}\bar{\boldsymbol{G}}^{\mathsf{H}}\big\}\,\bar{\boldsymbol{S}}_{\mathrm{D}}^{\mathsf{H}} + \sigma_n^2 \boldsymbol{I}_{NN_{\mathrm{S}}} N_{\mathrm{R}}. \tag{2.28}$$

Taking further the correlation structure (2.13) of the fading channel into account we obtain for the autocorrelation matrix of $\bar{\boldsymbol{G}}$

$$\boldsymbol{\Psi}_{\bar{\boldsymbol{G}}\bar{\boldsymbol{G}}} \triangleq \mathcal{E}\Big\{\bar{\boldsymbol{G}}\bar{\boldsymbol{G}}^{\mathsf{H}}\Big\} \tag{2.29}$$

$$= \mathrm{toeplitz}\{\psi_{gg}[0],\ldots,\psi_{gg}[N-1]\} \otimes \boldsymbol{I}_{N_{\mathrm{S}}N_{\mathrm{R}}}, \tag{2.30}$$

$$\triangleq \boldsymbol{\Psi}_{gg} \otimes \boldsymbol{I}_{N_{\mathrm{S}}N_{\mathrm{R}}}, \tag{2.31}$$

($\mathrm{toeplitz}\{x_1,\ldots,x_N\}$: $(N\times N)$–dimensional Hermitian symmetric Toeplitz matrix [Gra71] with $[x_1,\ldots,x_N]^{\mathsf{T}}$ as its first column, $\cdot^{\mathsf{T}}$: transposition) with $\psi_{gg}[\kappa]$ as defined in (2.13). Finally,

plugging (2.31) into (2.28) and using the identity $\det\{\boldsymbol{A}\boldsymbol{B}\} = \det\{\boldsymbol{A}\}\det\{\boldsymbol{B}\}$ for arbitrary square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ we find that

$$\begin{aligned}\det\{\boldsymbol{\Psi}_{\bar{\boldsymbol{R}}\bar{\boldsymbol{R}}|\bar{\boldsymbol{S}}}\} &= \det\{\bar{\boldsymbol{S}}_{\mathrm{D}}\left(\left(\boldsymbol{\Psi}_{gg}+\sigma_n^2\boldsymbol{I}_N\right)\otimes\boldsymbol{I}_{N_{\mathrm{S}}}\right)\bar{\boldsymbol{S}}_{\mathrm{D}}^{\mathsf{H}}\} &(2.32)\\ &= \det\{\bar{\boldsymbol{S}}_{\mathrm{D}}\}\cdot\det\{\left(\boldsymbol{\Psi}_{gg}+\sigma_n^2\boldsymbol{I}_N\right)\otimes\boldsymbol{I}_{N_{\mathrm{S}}}\}\cdot\det\{\bar{\boldsymbol{S}}_{\mathrm{D}}^{\mathsf{H}}\} &(2.33)\\ &= \det\{\boldsymbol{\Psi}_{gg}+\sigma_n^2\boldsymbol{I}_N\}^{N_{\mathrm{S}}} &(2.34)\end{aligned}$$

and with (2.25)

$$p\left(\bar{\boldsymbol{R}}\,\middle|\,\bar{\boldsymbol{S}}\right)=\frac{\exp\left(-\frac{1}{N_{\mathrm{R}}}\mathrm{tr}\left\{\bar{\boldsymbol{R}}^{\mathsf{H}}\bar{\boldsymbol{S}}_{\mathrm{D}}\left(\left(\boldsymbol{\Psi}_{gg}+\sigma_n^2\boldsymbol{I}_N\right)^{-1}\otimes\boldsymbol{I}_{N_{\mathrm{S}}}\right)\bar{\boldsymbol{S}}_{\mathrm{D}}^{\mathsf{H}}\bar{\boldsymbol{R}}\right\}\right)}{\left(\left(\pi N_{\mathrm{R}}\right)^N\det\{\boldsymbol{\Psi}_{gg}+\sigma_n^2\boldsymbol{I}_N\}\right)^{N_{\mathrm{R}}N_{\mathrm{S}}}}. \tag{2.35}$$

Neglecting terms that are irrelevant for the maximization, realizing that $\exp(\cdot)$ is a monotonous function, and recalling the relation between $\bar{\boldsymbol{S}}$ and $\bar{\boldsymbol{V}}$ due to the differential encoding [cf. (2.11)] the ML–MSDD decision rule with respect to $\bar{\boldsymbol{V}}$ can be written as

$$\hat{\bar{\boldsymbol{V}}}=\underset{\tilde{\bar{\boldsymbol{V}}}\in\boldsymbol{\mathcal{V}}^{N-1}}{\mathrm{argmin}}\left\{\mathrm{tr}\left\{\bar{\boldsymbol{R}}^{\mathsf{H}}\tilde{\bar{\boldsymbol{S}}}_{\mathrm{D}}\left(\boldsymbol{M}\otimes\boldsymbol{I}_{N_{\mathrm{S}}}\right)\tilde{\bar{\boldsymbol{S}}}_{\mathrm{D}}^{\mathsf{H}}\bar{\boldsymbol{R}}\right\}\right\}, \tag{2.36}$$

with the $(N\times N)$–dimensional matrix

$$\boldsymbol{M}\triangleq\left(\boldsymbol{\Psi}_{gg}+\sigma_n^2\boldsymbol{I}_N\right)^{-1} \tag{2.37}$$

and the $(NN_{\mathrm{S}}\times NN_{\mathrm{S}})$–dimensional unitary block–diagonal matrix

$$\tilde{\bar{\boldsymbol{S}}}_{\mathrm{D}}\triangleq\mathrm{diag}\left\{\left[\prod_{\kappa=0}^{N-2}\tilde{\boldsymbol{V}}[k-\kappa]\right]^{\mathsf{H}},\left[\prod_{\kappa=0}^{N-3}\tilde{\boldsymbol{V}}[k-\kappa]\right]^{\mathsf{H}},\ \ldots\ ,\left[\prod_{\kappa=0}^{0}\tilde{\boldsymbol{V}}[k-\kappa]\right]^{\mathsf{H}},\boldsymbol{I}_{N_{\mathrm{S}}}\right\}, \tag{2.38}$$

($\prod_{i=l}^{u}\boldsymbol{X}[i]\triangleq\boldsymbol{X}[l]\boldsymbol{X}[l+1]\cdots\boldsymbol{X}[u]$, $\mathrm{argmin}_{x\in\mathcal{X}}\{f(x)\}$: returns that element of a set $\mathcal{X}$ that minimizes the function $f(x)$). Note that the last $(N_{\mathrm{S}}\times N_{\mathrm{S}})$–dimensional block–diagonal element of $\tilde{\bar{\boldsymbol{S}}}$ can be fixed as $\boldsymbol{I}_{N_{\mathrm{S}}}$ without loss of optimality. The reason for this lies in the fact that the noncoherent detector can only determine the transmit signal up to a common unitary transformation of all transmit symbols inside the MSDD observation window. This is however not a problem due to the differential encoding (2.11), i.e. right–multiplication of $\tilde{\bar{\boldsymbol{S}}}$ with an arbitrary unitary $\tilde{\boldsymbol{S}}[k]$ would not alter the decoding result (cf. also the step leading from (2.15) to (2.16)). Note also that throughout this work we use accents to distinguish between (i) the true data and transmit sequences $\bar{\boldsymbol{V}}$ and $\bar{\boldsymbol{S}}$, (ii) candidate sequences $\tilde{\bar{\boldsymbol{V}}}$ and $\tilde{\bar{\boldsymbol{S}}}$, and (iii) decoder output sequences $\hat{\bar{\boldsymbol{V}}}$ and $\hat{\bar{\boldsymbol{S}}}$.

As MSDD returns $N-1$ decisions on data symbols $\boldsymbol{V}[k-\kappa]$, $0\leq\kappa\leq N-2$, successive matrices $\bar{\boldsymbol{R}}[k]$ must overlap by one symbol $\boldsymbol{R}[k]$, i.e. the so–called MSDD observation window of length N must only slide forward by $N-1$ symbols at a time, i.e. $k:=k+N-1$ ($x:=y$: assignment of a value y to a variable x). For illustration see Fig. 2.1.

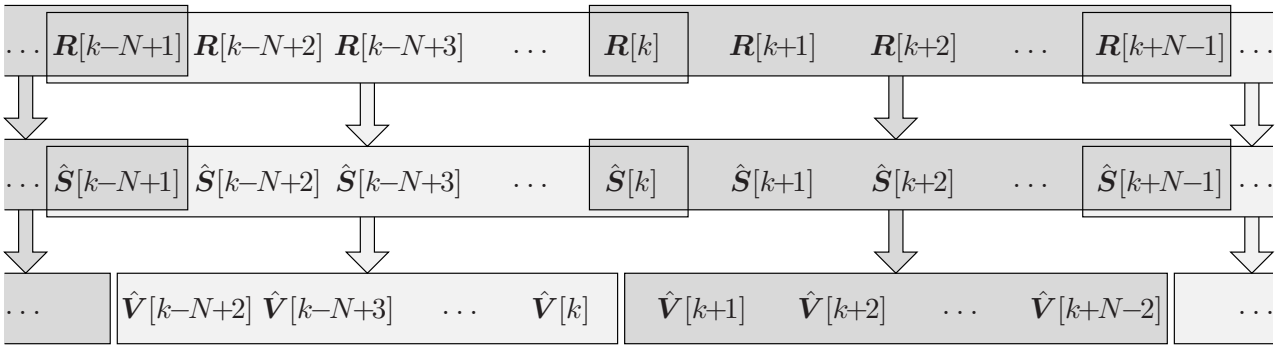

Figure 2.1: Illustration of multiple–symbol differential detection (MSDD).

2.4.2.1 Relation between ML MSDD and Linear MMSE Interpolation

Let us briefly review a result from [PK88]. Given a vector $\boldsymbol{x}$ of N samples x_n of a random process with autocorrelation matrix $\boldsymbol{\Psi}_{\boldsymbol{xx}} \stackrel{\triangle}{=} \mathcal{E}\{\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}\}$ the interpolation error filter $\boldsymbol{b}_n$ that leads to the minimum–mean squared error (MMSE) variance $\mathcal{E}\{|\breve{x}_n|^2\}$ with

$$\breve{x}_n \stackrel{\triangle}{=} x_n - \hat{x}_n = \boldsymbol{b}_n^{\mathsf{H}}\boldsymbol{x} \tag{2.39}$$

in the interpolation of an $\hat{x}_n$ from the remaining $N-1$ samples x_ν, $1 \le \nu \le N$, $\nu \neq n$, is given by

$$\boldsymbol{b}_n = \frac{1}{\left[\boldsymbol{\Psi}_{\boldsymbol{xx}}^{-1}\right]_{n,n}} \left[\boldsymbol{\Psi}_{\boldsymbol{xx}}^{-1}\right]_{n,:}. \tag{2.40}$$

Thus, the interpolation error vector $\breve{\boldsymbol{x}}$ if all N samples x_n are interpolated from the respective remaining $N-1$ samples x_ν, $1 \le \nu \le N$, $\nu \neq n$, can be written as

$$\begin{aligned} \breve{\boldsymbol{x}} &\stackrel{\triangle}{=} \boldsymbol{x} - \hat{\boldsymbol{x}} && (2.41)\\ &= \operatorname{diag}\left\{ \left[\boldsymbol{\Psi}_{\boldsymbol{xx}}^{-1}\right]_{1,1}, \ldots, \left[\boldsymbol{\Psi}_{\boldsymbol{xx}}^{-1}\right]_{N,N} \right\}^{-1} \boldsymbol{\Psi}_{\boldsymbol{xx}}^{-1}\boldsymbol{x}. && (2.42) \end{aligned}$$

When considering

$$\mathcal{E}\{\breve{\boldsymbol{x}}\breve{\boldsymbol{x}}^{\mathsf{H}}\} = \operatorname{diag}\left\{ \left[\boldsymbol{\Psi}_{\boldsymbol{xx}}^{-1}\right]_{1,1}, \ldots, \left[\boldsymbol{\Psi}_{\boldsymbol{xx}}^{-1}\right]_{N,N} \right\}^{-2} \boldsymbol{\Psi}_{\boldsymbol{xx}}^{-1} \tag{2.43}$$

we see that (i) the entries $\breve{x}_n$ of the interpolation error $\breve{\boldsymbol{x}}$ are correlated and (ii) the interpolation error variances $\sigma_{\mathrm{i},n}^2$ in the different positions n are equal to the respective diagonal elements of $\boldsymbol{\Psi}_{\boldsymbol{xx}}^{-1}$, i.e.

$$\sigma_{\mathrm{i},n}^2 = \frac{1}{\left[\boldsymbol{\Psi}_{\boldsymbol{xx}}^{-1}\right]_{n,n}}. \tag{2.44}$$

In view of the desired relation between linear MMSE interpolation and MSDD it is further interesting to consider the correlation between the sample vector $\boldsymbol{x}$ and the interpolation error

vector $\breve{\boldsymbol{x}}$. Using (2.42), (2.44) and the identity $\boldsymbol{a}^{\mathsf{H}}\boldsymbol{b} = \mathrm{tr}\{\boldsymbol{b}\boldsymbol{a}^{\mathsf{H}}\}$ for arbitrary vectors $\boldsymbol{a}, \boldsymbol{b}$ of equal length we can write

$$\begin{aligned}
\mathcal{E}\{\boldsymbol{x}^{\mathsf{H}}\breve{\boldsymbol{x}}\} &= \mathcal{E}\{\boldsymbol{x}^{\mathsf{H}}\mathrm{diag}\{\sigma_{\mathrm{i},1}^2, \ldots, \sigma_{\mathrm{i},N}^2\}\, \boldsymbol{\Psi}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\boldsymbol{x}\} && (2.45)\\
&= \mathcal{E}\{\mathrm{tr}\{\mathrm{diag}\{\sigma_{\mathrm{i},1}^2, \ldots, \sigma_{\mathrm{i},N}^2\}\, \boldsymbol{\Psi}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}\}\} && (2.46)\\
&= \mathrm{tr}\{\mathrm{diag}\{\sigma_{\mathrm{i},1}^2, \ldots, \sigma_{\mathrm{i},N}^2\}\} && (2.47)\\
&= \sum_{n=1}^{N} \sigma_{\mathrm{i},n}^2. && (2.48)
\end{aligned}$$

In MSDD on the other hand, the candidate $\tilde{\tilde{\boldsymbol{S}}}$, which minimizes [cf. (2.36)]

$$\mathrm{tr}\Big\{\bar{\boldsymbol{R}}^{\mathsf{H}}\boldsymbol{\Psi}_{\bar{\boldsymbol{R}}\bar{\boldsymbol{R}}|\tilde{\tilde{\boldsymbol{S}}}}^{-1}\bar{\boldsymbol{R}}\Big\} \tag{2.49}$$

where [cf. (2.28) with (2.31) and (2.37)]

$$\boldsymbol{\Psi}_{\bar{\boldsymbol{R}}\bar{\boldsymbol{R}}|\tilde{\tilde{\boldsymbol{S}}}} = \tilde{\tilde{\boldsymbol{S}}}_{\mathrm{D}}\,(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{S}}})\,\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{D}}^{\mathsf{H}}, \tag{2.50}$$

is chosen as decoder output $\hat{\tilde{\tilde{\boldsymbol{S}}}}$. With (2.42) we can now see that $\boldsymbol{\Psi}_{\bar{\boldsymbol{R}}\bar{\boldsymbol{R}}|\tilde{\tilde{\boldsymbol{S}}}}^{-1}\bar{\boldsymbol{R}}$ in (2.49) can be interpreted as the $(NN_{\mathrm{S}} \times N_{\mathrm{R}})$–dimensional matrix of interpolation errors under the hypothesis that $\tilde{\tilde{\boldsymbol{S}}}$ was transmitted, its entries being normalized by the respective interpolation error variances.[3] In other words, ML MSDD corresponds to choosing the hypothesis $\tilde{\tilde{\boldsymbol{S}}}$ such that the instantaneous correlation between the matrix $\bar{\boldsymbol{R}}$ of received samples and the matrix $\boldsymbol{\Psi}_{\bar{\boldsymbol{R}}\bar{\boldsymbol{R}}|\tilde{\tilde{\boldsymbol{S}}}}^{-1}\bar{\boldsymbol{R}}$ of hypothetical linear MMSE interpolation errors is minimized.

Finally, we see by plugging (2.50) and (2.24) into (2.49) that for $\tilde{\tilde{\boldsymbol{V}}} = \bar{\boldsymbol{V}}$ the ML–MSDD metric becomes

$$\mathrm{tr}\Big\{\big(\bar{\boldsymbol{G}} + \bar{\boldsymbol{N}}\big)^{\mathsf{H}}\,(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{S}}})\,\big(\bar{\boldsymbol{G}} + \bar{\boldsymbol{N}}\big)\Big\}, \tag{2.51}$$

where $\boldsymbol{M}$ [cf. (2.37)] denotes the inverse of the autocorrelation matrix that is common to the $N_{\mathrm{S}}N_{\mathrm{R}}$ iid fading–plus–noise processes $g_{i,j}[k] + n_{i,j}[k]$, $1 \le i \le N_{\mathrm{S}}$, $1 \le j \le N_{\mathrm{R}}$. From (2.44) it then follows that the inverse main diagonal elements $m_{n,n}^{-1}$, $1 \le n \le N$, of $\boldsymbol{M}$ are the corresponding interpolation–error variances for the fading–plus–noise process.

Thus, we can summarize: MSDD can be interpreted as a concurrent interpolation of the fading–plus–noise process for each of the N samples from the remaining $N-1$ samples in the observation window. At this, symbols $\hat{\boldsymbol{V}}[k-N+1+n]$, $1 \le n \le N-1$, are chosen such that the sum of the squared interpolation errors in the N positions for each of the $N_{\mathrm{S}}N_{\mathrm{R}}$ individual subchannels is minimized. The inverse main diagonal elements $m_{n,n}^{-1}$, $1 \le n \le N$, of the $(N \times N)$–dimensional matrix $\boldsymbol{M}$ [cf. (2.37)] are the interpolation–error variances for

[3] We note that if $\tilde{\tilde{\boldsymbol{S}}} \neq \bar{\boldsymbol{S}}$, this interpolation is mismatched.

interpolating $g_{i,j}[k-N+n]+n_{i,j}[k-N+n]$ from $g_{i,j}[k-N+\nu]+n_{i,j}[k-N+\nu]$, $1 \leq \nu \leq N$, $\nu \neq n$.

This connection between MSDD and linear interpolation is a new result of this work and will be quite useful in the following. On the one hand, it allows us to exploit well–known results from interpolation theory for the performance analysis of MSDD in later sections. On the other hand, it motivates the introduction of a variant of MSDD we refer to as subset MSDD (S–MSDD), cf. Section 2.4.2.2.

2.4.2.2 Subset MSDD (S–MSDD)

Based on the abovementioned relationship between MSDD and linear MMSE interpolation it is intuitive that symbol decisions on the $N-1$ data symbols in $\bar{\boldsymbol{V}}$ are not equally reliable. Especially in relatively fast fading environments it can be expected that symbols located in the center of the observation window can be detected more reliably than those at the edges.

This observation strongly suggests a variant of MSDD, which we refer to as subset MSDD (S–MSDD). Like regular MSDD it processes blocks $\bar{\boldsymbol{R}}[k]$ of N matrix symbols to find estimates $\hat{\bar{\boldsymbol{S}}}[k]$ for corresponding blocks $\bar{\boldsymbol{S}}[k]$ of N transmit symbols. Contrary to regular MSDD, however, it only returns estimates

$$\hat{\boldsymbol{V}}[k-n], \quad \lceil (N-N')/2 \rceil - 1 \leq n \leq N-2-\lfloor (N-N')/2 \rfloor, \tag{2.52}$$

of $N' \leq N-1$ data symbols located in the center of the observation window, i.e. it discards $(N-N'-1)/2$ decisions at each end of the observation window. Accordingly, the observation window must slide forward in steps of N', i.e. $k := k+N'$, and the decoding complexity compared to regular MSDD is increased by a factor of $(N-1)/N'$.

Note that S–MSDD can be viewed as generalization of regular MSDD as the latter is included as special case with $N' = N-1$.

2.4.2.3 Discussion

Note that the performance of MSDD improves if N is increased and approaches that of coherent detection with perfect CSI as the channel memory is taken into account more and more completely. Unfortunately, its complexity increases exponentially in N as the number of relevant candidate sequences is L^{N-1}. Therefore, for arbitrary fading scenarios only relatively small values of N were feasible. Only for certain fading scenarios more efficient implementations have been developed, such as the algorithm of [Mac94] (cf. also [Swe01]), whose complexity is of the order $N\log(N)$, for DPSK and a time–invariant fading channel.

Other authors have considered noncoherent sequence detection based on sparse tree search, cf. e.g. [AS89, LW90] or by trellis search, cf. e.g. [MF90, LM90, MMB94, YP95, VT95a, VT95b,

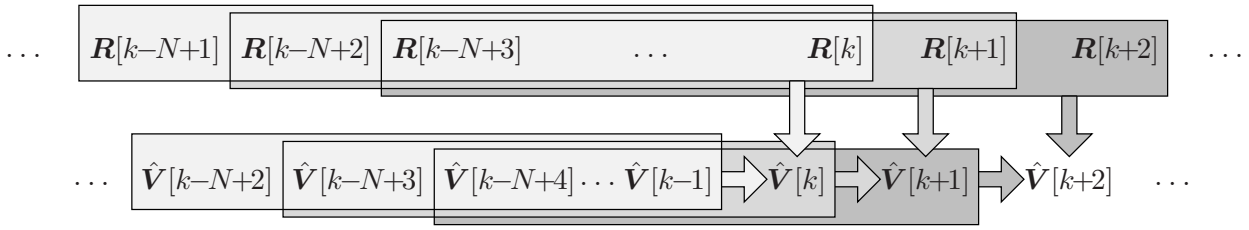

Figure 2.2: Illustration of decision–feedback differential detection (DFDD).

Rap96a, Ada96, CR99]. If the number of states in the trellis, in which detection is performed, is reduced to its minimum L, decision–feedback differential detection (DFDD) results, cf. [LP88, Edb92, AS93, AS95, SGH99, SL02].

2.4.3 Decision–Feedback Differential Detection (DFDD)

This detection scheme was developed for transmission over the single–input single–output (SISO) AWGN channel in [LP88, Edb92, AS93, AS95]. In [Sve94, Ada98, SGH99] and [SL02] it was then extended to the interesting scenarios of fading SISO and MIMO channels with DPSK and DSTM, respectively.

DFDD is derived from MSDD by feeding back $N-2$ previously decided symbols $\hat{\boldsymbol{V}}[k-\kappa]$, $1 \leq \kappa \leq N-2$, into the ML–MSDD metric in (2.36) and deciding only on $\hat{\boldsymbol{V}}[k]$. Neglecting all terms that do not depend on $\boldsymbol{V}[k]$, the DFDD decision rule reduces to [SL02]

$$\hat{\boldsymbol{V}}[k] = \underset{\tilde{\boldsymbol{V}}[k] \in \boldsymbol{\mathcal{V}}}{\operatorname{argmax}} \Big\{ \operatorname{Re} \Big\{ \operatorname{tr} \Big\{ \tilde{\boldsymbol{V}}[k] \boldsymbol{R}_{\mathrm{ref}}[k-1] \boldsymbol{R}^{\mathsf{H}}[k] \Big\} \Big\} \Big\}, \tag{2.53}$$

where

$$\boldsymbol{R}_{\mathrm{ref}}[k-1] \triangleq \sum_{\kappa=1}^{N-1} p_{\mathrm{F},\kappa}^{(N-1)} \prod_{i=1}^{\kappa-1} \hat{\boldsymbol{V}}[k-i] \boldsymbol{R}[k-\kappa]. \tag{2.54}$$

The linear forward MMSE prediction filter coefficients $\boldsymbol{p}_{\mathrm{F}}^{(N-1)} \triangleq \big[p_{\mathrm{F},1}^{(N-1)}, \ldots, p_{\mathrm{F},N-1}^{(N-1)}\big]^{\mathsf{T}}$ can be obtained (i) through solution of the corresponding $(N-1)$–dimensional Yule–Walker equation [Hay96]

$$\begin{bmatrix} \psi_{gg}[0]+\sigma_n^2 & \psi_{gg}[-1] & \ldots & \psi_{gg}[2-N] \\ \psi_{gg}[1] & \psi_{gg}[0]+\sigma_n^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \psi_{gg}[-1] \\ \psi_{gg}[N-2] & \ldots & \psi_{gg}[1] & \psi_{gg}[0]+\sigma_n^2 \end{bmatrix} \Big(\boldsymbol{p}_{\mathrm{F}}^{(N-1)}\Big)^* = \begin{bmatrix} \psi_{gg}[1] \\ \psi_{gg}[2] \\ \vdots \\ \psi_{gg}[N-1] \end{bmatrix}, \tag{2.55}$$

with $\psi_{gg}[\kappa]$ as defined in (2.13), (ii) equivalently as the sub–diagonal elements of the first column of $(N \times N)$–dimensional $\boldsymbol{M}$ [cf. (2.37)], or (iii) adaptively e.g. via the recursive least–squares

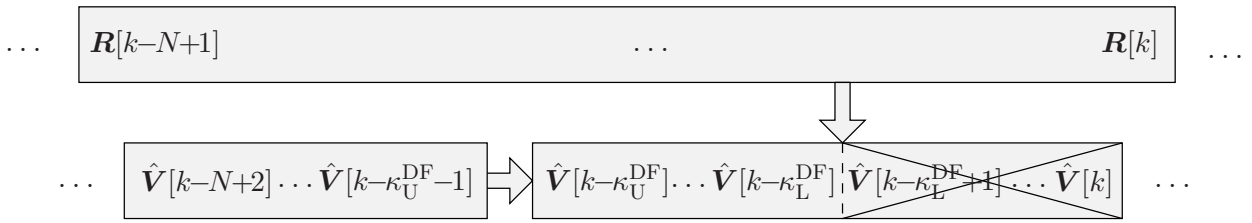

Figure 2.3: Illustration of decision–feedback multiple–symbol differential detection (DF–MSDD). Observation window slides forward in steps of $\kappa_{\mathrm{U}}^{\mathrm{DF}} - \kappa_{\mathrm{L}}^{\mathrm{DF}} + 1$, i.e. $k := k + \kappa_{\mathrm{U}}^{\mathrm{DF}} - \kappa_{\mathrm{L}}^{\mathrm{DF}} + 1$.

(RLS) algorithm [YL95, SG00, SGH01]. Since DFDD returns only a single estimate $\hat{\boldsymbol{V}}[k]$ per decoder run, the observation window of length N comprising the N received matrices $\boldsymbol{R}[k-\kappa]$, $0 \leq \kappa \leq N-1$, must slide forward by only one matrix–symbol at a time, i.e. $k := k+1$. For illustration see Fig. 2.2

A comparison of (2.15), (2.17), (2.36) and (2.53) confirms the intuitively expected result that CDD, MSDD with $N = 2$ and DFDD with $N = 2$ are equivalent.

2.4.4 Decision–Feedback Multiple–Symbol Differential Detection (DF–MSDD)

Decision–feedback differential detection (DFDD, cf. Section 2.4.3) allows for significant performance gains compared to conventional differential detection (CDD, cf. Section 2.4.1), because it can —due to its relation with MSDD— take information about the statistical properties of the fading channel into account. However, given the relationship between MSDD and linear MMSE interpolation (cf. Section 2.4.2.1) one can expect that in most cases the decisions at the very edges of the MSDD observation window are the ones that are least reliable. DFDD returns a decision on the last symbol of the observation window, thereby leaving room for improvement.

Therefore, it appears reasonable to combine ideas of DFDD and subset MSDD (S–MSDD, cf. Section 2.4.2.2) by feeding back some previously decided symbols into the MSDD metric, and returning decisions only on symbols that do not lie at the very edges of the observation window.

More specifically, we propose the following noncoherent detection scheme: As in (S–)MSDD and DFDD, we deploy an observation window extending over N received symbols summarized in $\bar{\boldsymbol{R}}[k] = \left[\boldsymbol{R}^{\mathsf{H}}[k-N+1], \ldots, \boldsymbol{R}^{\mathsf{H}}[k]\right]^{\mathsf{H}}$. Instead of optimizing the ML–MSDD metric, i.e. the argument of the argmin function in (2.36), over all $N-1$ corresponding data symbols $\boldsymbol{V}[k-\kappa]$, $0 \leq \kappa \leq N-2$, we feed back $N - \kappa_{\mathrm{U}}^{\mathrm{DF}} - 2$ previous decisions $\hat{\boldsymbol{V}}[k-\kappa]$, $\kappa_{\mathrm{U}}^{\mathrm{DF}} + 1 \leq \kappa \leq N-2$, and optimize the ML–MSDD metric only over the remaining $\kappa_{\mathrm{U}}^{\mathrm{DF}} + 1$ symbols $\boldsymbol{V}[k-\kappa]$, $0 \leq \kappa \leq \kappa_{\mathrm{U}}^{\mathrm{DF}}$. In order to exclude the often unreliable symbols at the end of the observation window, the

decoder does not return decisions on all $\kappa_{\mathrm{U}}^{\mathrm{DF}}+1$ symbols $\boldsymbol{V}[k-\kappa]$, $0 \leq \kappa \leq \kappa_{\mathrm{U}}^{\mathrm{DF}}$, but only on $\kappa_{\mathrm{U}}^{\mathrm{DF}}-\kappa_{\mathrm{L}}^{\mathrm{DF}}+1$ symbols $\boldsymbol{V}[k-\kappa]$, $\kappa_{\mathrm{L}}^{\mathrm{DF}} \leq \kappa \leq \kappa_{\mathrm{U}}^{\mathrm{DF}}$, and discards the remaining $\kappa_{\mathrm{L}}^{\mathrm{DF}}$ decisions at the end of the observation window. In consequence, the observation window must slide forward in steps of $\kappa_{\mathrm{U}}^{\mathrm{DF}}-\kappa_{\mathrm{L}}^{\mathrm{DF}}+1$ symbols at a time, i.e. $k := k+\kappa_{\mathrm{U}}^{\mathrm{DF}}-\kappa_{\mathrm{L}}^{\mathrm{DF}}+1$. For illustration see Fig. 2.3.

Clearly, it can be expected that the power efficiency of this decoder is superior compare to that of DFDD. At the same time, computational complexity is —through expedient choice of the parameters $\kappa_{\mathrm{U}}^{\mathrm{DF}}$ and $\kappa_{\mathrm{L}}^{\mathrm{DF}}$— significantly reduced compared to (S–)MSDD as the dimension of the search space is $\kappa_{\mathrm{U}}^{\mathrm{DF}}+1$ instead of $N-1$. In fact, we will see that with $\kappa_{\mathrm{U}}^{\mathrm{DF}}=\kappa_{\mathrm{L}}^{\mathrm{DF}}=1$, i.e. a decision is returned only on the second to last symbol $\boldsymbol{V}[k-1]$ in the current observation window, leads to a power efficiency very close to that of S–MSDD, i.e. at times even better than that of regular MSDD. At the same time, the dimension of the search space is reduced from $N-1$ to 2, i.e. there are L^2 instead of L^{N-1} relevant candidates to be examined.

2.4.5 (Differentially) Coherent Detection

As ultimate benchmark decoders[4] for the above noncoherent detection schemes, we consider idealized symbol–by–symbol coherent detection *with perfect channel state information (CSI)* at the receiver and with and without differential encoding at the transmitter.

Since, according to the above channel model, the entries of $\boldsymbol{N}[k]$ are iid $\mathcal{N}_{\mathrm{c}}(0,\sigma_n^2)$ distributed random variables, the coherent symbol–by–symbol ML decision rule is obtained from the conditional PDF

$$p(\boldsymbol{R}[k] \,|\, \boldsymbol{S}[k]) = \frac{1}{(\pi\sigma_n^2)^{N_{\mathrm{R}}N_{\mathrm{S}}}} \cdot \exp\left(-\frac{||\boldsymbol{R}[k]-\boldsymbol{S}[k]\boldsymbol{G}[k]||^2}{\sigma_n^2}\right), \tag{2.56}$$

($||\cdot||$: Frobenius norm).

2.4.5.1 Coherent Detection

In a communication system, that does *not* deploy differential encoding at the transmitter, i.e. the data to be transmitted is mapped directly to DSTM matrix symbols $\boldsymbol{S}[k]=\boldsymbol{V}[k] \in \boldsymbol{\mathcal{V}}$, coherent ML symbol–by–symbol detection is performed via

$$\hat{\boldsymbol{V}}[k] = \hat{\boldsymbol{S}}[k] = \underset{\tilde{\boldsymbol{S}}[k]\in\boldsymbol{\mathcal{V}}}{\operatorname{argmin}} \left\{ \left|\left| \boldsymbol{R}[k]-\tilde{\boldsymbol{S}}[k]\boldsymbol{G}[k] \right|\right|^2 \right\}. \tag{2.57}$$

[4]We note that good noncoherent receivers may approach the power efficiency of the corresponding coherent detectors, however they can not outperform the latter [SHL94].

2.4.5.2 Differentially Coherent Detection

Differentially encoded transmission with DSTM group–codes in conjunction with differentially coherent detection may also be an interesting alternative, because it resolves L–ary "phase ambiguities", i.e. rotations of $\boldsymbol{G}[k]$ by any one of the L unitary DSTM symbols. Here, the detector must maintain an estimate $\hat{\boldsymbol{S}}[k-1]$ of $\boldsymbol{S}[k-1]$ to perform detection via

$$\hat{\boldsymbol{V}}[k] = \underset{\tilde{\boldsymbol{V}}[k]\in\boldsymbol{\mathcal{V}}}{\operatorname{argmin}} \left\{ \left|\left| \boldsymbol{R}[k] - \tilde{\boldsymbol{V}}[k]\hat{\boldsymbol{S}}[k-1]\boldsymbol{G}[k] \right|\right|^2 \right\}. \tag{2.58}$$

Given the above idealized settings this entails an increase in the symbol–error rate (SER) due to error propagation by a factor of approximately two compared to the non–differential coherent detector for group–code DSTM constellations, whereas for non–group DSTM constellations such a symbol–by–symbol detector incurs severe error propagation.

Chapter 3

Low–Complexity Multiple–Symbol Differential Detection

In the previous chapter, we derived the decision rule for ML–MSDD ([cf. (2.36)] and also introduced DF–MSDD (cf. Section 2.4.4). Because the evaluation of the ML–MSDD decision rule involves a search in an $(N-1)$–dimensional space of L–ary variables, the brute–force approach of finding the ML–MSDD solution $\hat{\tilde{\boldsymbol{V}}}$ by computing the ML–MSDD metric for all L^{N-1} candidates $\tilde{\tilde{\boldsymbol{V}}} \in \boldsymbol{\mathcal{V}}^{N-1}$ quickly becomes intractable as N grows. Similarly, the complexity of the brute–force approach to DF–MSDD is of the order $L^{\kappa_{\mathrm{U}}^{\mathrm{DF}}+1}$. While DFDD (cf. Section 2.4.3) achieves significantly better performance than CDD (cf. Section 2.4.1) at a comparable complexity it still leaves ample room for improvement especially in fast fading scenarios. In this chapter, we will therefore deal with the development of algorithms that (approximately) solve ML MSDD at a computational receiver complexity that is comparable to those of CDD and DFDD and significantly reduced compared to the abovementioned brute–force approach.

To this end, we consider two different approaches: The first one is based on a representation of the MSDD problem in a tree and uses methods from tree–search decoding which encompasses methods from (i) sequential decoding developed for the decoding of convolutional codes with high memory, and (ii) algorithms developed for closest–point / shortest–vector search in lattice theory, which include the frequently considered class of socalled sphere decoders (SpD). It is presented in Section 3.1 along with more detailed background information on the origins and methods of tree–search decoding.

The second approach makes use of methods from combinatorial geometry. Here, the $(N-1)$–dimensional MSDD problem is cast into an N_λ–dimensional space, where N_λ denotes the rank of the fading correlation matrix $\boldsymbol{\Psi}_{gg}$ [cf. (2.31)]. Using a number of $(N_\lambda-1)$–dimensional hyperplanes this space is partitioned into disjoint cells, each of which corresponds to one candidate $\bar{\boldsymbol{V}}$. While the average complexity of this approach is in most scenarios higher than that of the tree–search based approach, it has the appealing property that its computational com-

plexity is (i) practically independent of the SNR, (ii) practically independent of the particular channel state and (iii) polynomial in N if the rank N_λ of $\mathbf{\Psi}_{gg}$ is fixed. This approach will be presented in Section 3.2.

The system model of Chapter 2 shall serve as basis for our considerations throughout this chapter.

3.1 MSDD Based on Tree–Search Decoding

Tree–search methods have attracted considerable attention in the recent communications literature, as they have been found well apt for solving multi–dimensional optimization problems. In this section, we shall investigate the application of tree–search methods to achieve power–efficient noncoherent detection based on MSDD with low receiver complexity. To this end, we begin by reviewing some preliminaries and important algorithms of tree–search decoding in Section 3.1.1, before we develop representations of the MSDD metric amenable to tree–search decoding in Section 3.1.2. In Section 3.1.3 we will then present highly efficient implementations of MSDD based on tree–search decoding and further optimize them regarding individual DSTM constellations in Section 3.1.4.

3.1.1 Preliminaries from Tree–Search Decoding

3.1.1.1 Trees

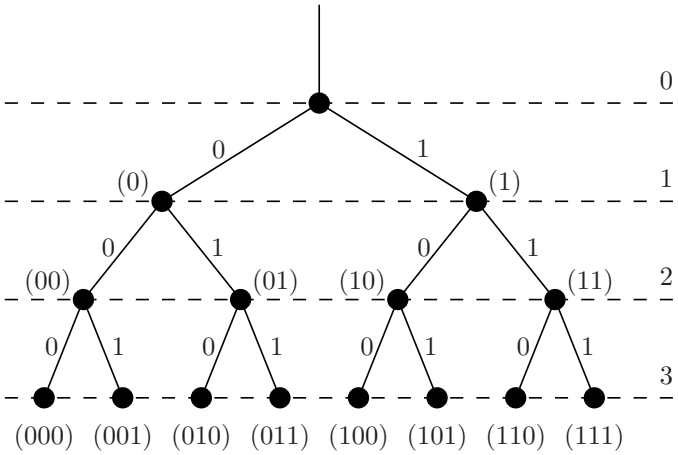

Figure 3.1: Binary tree of maximal depth three. The branch and node / path labels are indicated next to the branches and in brackets near the nodes, respectively. The depth of a node is indicated on the right.

A *tree* consists of a single *root(–node)*, *branches* and *nodes* at different *depths* of the tree.

For illustration see Fig. 3.1. The root lies at depth zero of the tree. From every node at depth i, $i = 0, 1, 2, \ldots$, a number of branches emanate, each ending at a node at depth $i + 1$. While in many applications the same number of branches emanate from each node, this is not a prerequisite for the tree–search algorithms discussed in the following to be applicable. Nodes at the end of the tree, i.e. nodes from which no branches emanate to nodes at greater depths, are referred to as *leaves* and all nodes between the root and the leaves as *intermediate nodes*. Again, while in the applications considered in this work all leaves lie at the same depth of the tree, this is by no means necessary for the tree–search algorithms to be applicable. A *path* is a sequence of branches that connect a node of the tree to its root. The number of branches that make up a path to a node, i.e. its depth, is also referred to as *length* of this path. A path is referred to as *child* of another path, if it results from the latter *parent* by extending it by one branch. We refer to two paths as *sisters* if they have the same length and differ only in the last branch. The same relations apply to nodes. Each branch shall be associated with a real–valued *branch metric* and each path / node with a *additive path metric* given as the sum of the corresponding branch metrics that form this path.[1] In addition the different branches emanating from the same node shall be labeled with different symbols, the meaning of which depends on the application and is irrelevant at this point. Hence, each path / node of the tree is uniquely identified by the sequence of symbols corresponding to the branches making up this path / leading to this node.

3.1.1.2 Origins of Tree–Search Decoding

Many decoding and other optimization problems from various fields of scientific research can be stated in tree structures as described above such that the optimization is transformed into a search for the path from the root to a leaf of the tree that has optimal (minimal or maximal) path metric. Two areas where major contributions have been made to the evolution of tree–search decoding are *lattice theory* and *sequential decoding*.

3.1.1.2.1 Lattice Theory Research in this area evolved from number theory and was initially concerned with convex quadratic optimization, cf. e.g. [Her50, Vor09, Min11]. For a fundamental treatment of lattice theory and related topics cf. e.g. [vzGG99, Jou93, CS99] or the more recent semi–tutorial paper [AEVZ02].

In lattice theory, a lattice is defined via

$$\mathcal{L} \triangleq \{\boldsymbol{G}_{\mathcal{L}}\boldsymbol{u} \mid \boldsymbol{u} \in \mathbb{Z}^n\}\,, \tag{3.1}$$

[1] Note that these "metrics" may not be metrics in the strongest mathematical sense of the word. However, we still use this expression in place of the more appropriate expression *cost function* to conform with the relevant literature, e.g. [AM84, JZ99].

where the n linearly independent columns of the $(m \times n)$–dimensional *generator matrix* $\boldsymbol{G}_{\mathcal{L}} \in \mathbb{R}^{m \times n}$, $m \geq n$, constitute a basis of the lattice $\mathcal{L}$ ($\mathbb{R}$: set of real numbers, $\mathbb{Z}$: set of integer numbers).

A frequently considered problem in this context is the so–called *closest lattice point* problem for the special case of $m = n$, where given some vector $\boldsymbol{x} \in \mathbb{R}^n$ one is interested in the vector $\hat{\boldsymbol{u}} \in \mathbb{Z}^n$ such that of all $\boldsymbol{c} \in \mathcal{L}$, $\hat{\boldsymbol{c}} = \boldsymbol{G}_{\mathcal{L}}\hat{\boldsymbol{u}}$ lies closest to $\boldsymbol{x}$ with respect to Euclidean distance, i.e.

$$\hat{\boldsymbol{u}} = \underset{\boldsymbol{u} \in \mathbb{Z}^n}{\operatorname{argmin}} \left\{ ||\boldsymbol{G}_{\mathcal{L}}\boldsymbol{u} - \boldsymbol{x}||^2 \right\}. \tag{3.2}$$

In the special case of $\boldsymbol{x} = \boldsymbol{0}_{n,1}$ and $\boldsymbol{u} \in \mathbb{Z}^n \backslash \{\boldsymbol{0}_{n,1}\}$ one obtains the closely related *shortest–vector* problem. Using e.g. a QR–decomposition [GvL96] of $\boldsymbol{G}_{\mathcal{L}}$ the optimization problem (3.2) can be brought into a form amenable to tree–search decoding. Algorithms such as those of [Die75, FP85, Bab86, SE94, VB99, AEVZ02] that find an (approximate) solution to the shortest–vector / closest–point problem therefore fall into the broader class of tree–search algorithms.

3.1.1.2.2 Sequential Decoding Independently of the above lattice theory a class of algorithms unifyingly referred to as sequential decoding algorithms has been developed for solving tree–search problems in digital communications, cf. e.g. [AM84, Bau92] for an overview. Original work in this field [Woz57, Fan63, Zig66, Jel69] was done for the decoding of convolutional codes (cf. e.g. [JZ99]). Later these methods where enhanced and employed in various types of source coding algorithms, cf. e.g. [Gal74, JBM75, MBA81]. Regarding the application to decoding of convolutional codes sequential decoding was later displaced by the Viterbi algorithm [Vit67] and received only minor attention until recently when the algorithms from lattice theory where introduced into various fields of modern communications and researchers noticed the close relation between the two classes of tree–search algorithms, cf. e.g. [MGDC06] for an overview.

3.1.1.2.3 Classification of Tree–Search Algorithms Tree–search algorithms are also often referred to as *branch–and–bound* algorithms, cf. e.g. [LW66]. The basic principle followed by all of these algorithms is to consider paths to intermediate nodes of the tree and decide based on a comparison of the corresponding path metric to some *threshold* ρ whether or not this path is likely to be part of the optimal path and should therefore be extended or discarded. Accordingly, the decoder either extends (*branch*) this path repeating the comparison for its children (*bound*) or does not extend this path thereby pruning the entire subtree emanating from the corresponding node from the decoding tree.

Whether or not the decoder can be guaranteed to find the optimal path or maybe only a close–to–optimal path depends on the definition of the metric and whether the decoder uses

other criteria for discarding paths. To assure that a tree–search decoder finds the optimal path through the tree, it must not use any discard criteria besides the comparison of path metric and threshold and the branch metrics must be strictly non–negative / non–positive if minimization / maximization of the path metric is desired, since in this case a shorter path whose metric is already larger / smaller than that of a full–length path can never —upon extension to full length— lead to a smaller / larger path metric. Accordingly, we will in the following distinguish between *optimal* and *suboptimal* tree–search decoders, where a tree–search algorithm is referred to as "optimal", if —given an appropriate metric definition— it finds the optimal leaf at all times.

The individual algorithms then differ essentially in their strategies regarding candidate enumeration and threshold adaptation. In [AM84] Anderson et al. classified the various sequential decoding algorithms with respect to two criteria: (i) *breadth first search* versus *depth first search* versus *metric first search* and (ii) *sorting* versus *non–sorting*. A similar classification with respect to the first criterion was also presented in [MGDC06] based on a generic tree–search algorithm.

The first criterion refers to the basic strategy a decoder uses in searching the tree, i.e. the order in which it examines candidates and the use of discard criteria.

Breadth First Search (BFS): Algorithms that fall into this category use *fixed discard criteria*.

An example for a breadth–first algorithm is the Fincke–Pohst sphere decoder (FP–SpD) [FP85] which examines all paths up to the point where the metric of all of their children lie above / below the fixed threshold ρ if minimization / maximization of metric is desired. Further examples are the M– and the T–algorithm of [AH77] and [MA91], respectively, and Wozencraft's sequential decoder [Woz57].

Algorithms of this type typically have a relatively high average complexity as they examine a relatively large number of (intermediate) nodes. On the upside this characteristic makes them well apt for applications where soft output is required, cf. e.g. [Kuh06].

Depth First Search (DFS): The prominent characteristic of DFS–type algorithms is that they are designed to reach a leaf of the tree as quickly as possible, even though this leaf may not be the decoder output of an optimal tree–search algorithm. At this many of these algorithms *tighten / relax the threshold adaptively* in the process of decoding.

Babai's nearest plane algorithm [Bab86] for lattice decoding, which is equivalent to decision–feedback equalization (DFE), cf. e.g. [Pro00], is an example of a suboptimal DFS algorithm. It moves straight from the root to a leaf of the tree by always extending the single path under consideration to its best child without ever checking any other branches. An example of an

optimal DFS algorithms is the sphere decoder (SpD) of Agrell et al. [AEVZ02]. This algorithm builds upon DFE in that it uses the DFE output as first preliminary result subsequently searching the tree for leaves that have a better metric. At this, the threshold for discarding further candidates is tightened whenever a better leaf has been found. For a detailed description of this and the closely related single–stack algorithm [Gal74] see Section 3.1.1.3.3.

DFS–type algorithms usually investigate fewer nodes than comparable BFS algorithms. In consequence, they are apt for hard output rather than soft output applications. Due to their strategy of determining preliminary decoding results and subsequently checking for better candidates these algorithms are particularly useful if the maximal decoder complexity is limited.

Metric First Search (MFS): Algorithms that use the MFS strategy typically do not use a threshold at all, but maintain a list of candidate paths to intermediate nodes and from this list always extend the path that presently has the best metric.[2] As soon as a path of full length is found to be the best of the current list, the search is terminated.

Examples of algorithms that follow this strategy are the stack algorithm [Zig66] and variants thereof such as the bucket algorithm [Jel69] and the algorithm of [Vin84]. All of these algorithms are discussed in detail in Section 3.1.1.3.1.

Among all optimal tree search algorithms, optimal MFS algorithms have the lowest complexity in terms of considered nodes, since paths are only extended up to the point where their metric is worse than that of the decoder output.

The main drawback of these algorithms is that theoretically an indefinite amount of memory must be provided to keep track of the abovementioned list of candidates.

Sorting: Apart from classifying tree–search algorithms according to breadth / depth / metric first they can also be categorized as sorting and non–sorting procedures.

Tree–search algorithms that fall into this category consider a number of candidate paths simultaneously, and sort them according to their metric. Based on this sorting they decide which paths are to be extended, stored for later extension or terminated.

Examples for sorting algorithms are the stack algorithm and its variants, where sorting is used to identify the path to be extended in the next decoding step while the others are stored for possible later extension, or the M– / T–algorithms, where a fixed / variable number of best paths from a list of candidate paths is extended, respectively, while the remaining are terminated.

Non–Sorting: Algorithms that do not compare paths to each other are referred to as non–sorting algorithms. Algorithms of this type only consider one path at a time and search

[2] Accordingly, this strategy is also referred to as *best first search*, cf. e.g. [MGDC06].

	Breadth first search	Depths first search	Metric first search
sorting	M–alg. [AH77] T–alg. [MA91]		Stack alg. [Zig66] Bucket alg. [Jel69] Merge Alg. [AM84]
non–sorting	Fincke–Pohst SpD [FP85]	——Fano Alg. [Fan63]—— Single–stack alg. [Gal74] A–SpD [AEVZ02] Nearest plane alg. [Bab86]	

Table 3.1: Classification of various important tree–search algorithms.

the tree by extending and moving back along this path in response to the value of its current path metric relative to the threshold. They are therefore much more efficient with respect to required memory than their sorting counterparts.

Examples for algorithms of this type are the single–stack algorithm, the sphere decoder and Fano's algorithm, all of which are discussed below.

Summary: The classification of various important tree–search algorithms according to BFS versus DFS versus MFS and sorting versus non–sorting is summarized in Table 3.1. The popular sequential decoding algorithm due to Fano [Fan63] discussed in detail below falls into the category of non–sorting algorithms, but can not be classified as purely BFS, MFS or DFS.

3.1.1.3 Selected Tree–Search Algorithms

From the plethora of tree–search algorithms that have been devised by various researchers over the past five decades we will in the following briefly describe a small number of algorithms that turned out to be most useful for the applications considered in this work. While these algorithms are applicable regardless of whether the path metric is to be maximized or minimized, the following descriptions assume —in anticipation of the applications considered in this work— *minimization* of the path metric.

3.1.1.3.1 (Modified) Stack Algorithm The stack algorithm is among the simplest yet very efficient tree–search algorithms. It was originally proposed by Zigangirov in [Zig66] and later by Jelinek in [Jel69] and is therefore also frequently referred to as Zigangirov–Jelinek algorithm.

This algorithm, whose flowchart is presented in Fig. 3.2, maintains a list of candidate paths from the root to intermediate nodes of the tree sorted in order of increasing path metric. At the

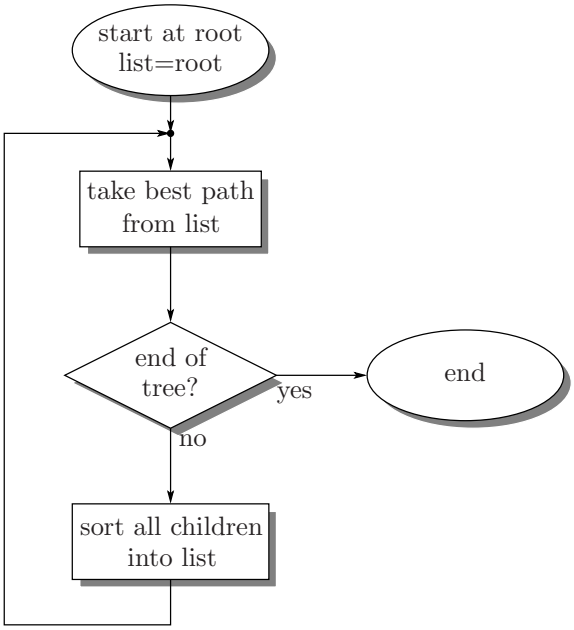

Figure 3.2: Flowchart of the stack decoder.

beginning of the search process the list is initialized with the root of the tree whose metric is without loss of generality chosen as zero. The algorithm then repeatedly extracts the currently best path from the top of the sorted list and sorts *all children* into the list. This process is terminated when the path at the top of the list has full length, i.e. when a leaf of the tree has been found, whose metric is smaller than those of all other candidates in the current list, and this candidate is returned as decoding result.

From this, it should be clear that the stack algorithm is optimal in the sense of Section 3.1.1.2.3, i.e. provided that the branch metrics are strictly non–negative this algorithm find the leaf that has them smallest metric of all leaves.

Despite its low complexity in terms of average number of examined branches, this simple algorithm has the drawback that the list of paths it maintains may become very long especially when the number of branches emanating from a node and / or the depth of the tree are large. Apart from the fact that a large amount of storage has to be provided in order to not impair the performance of this decoder this also poses a problem when it comes to sorting extensions into the list. As remedy for the latter problem Jelinek proposed the *bucket algorithm* in [Jel69]. This variant of the stack algorithm splits the support for the path metric into a number of disjoint intervals (buckets) and merely sorts candidates into these fixed buckets.[3] Instead of extracting the best path of a long list, this algorithm then takes one of the paths in the first (in order of

[3]It should be clear, that through expedient choice of the boundaries between the buckets the problem of finding the appropriate bucket for a candidate can be implemented such that the complexity of this sorting operation is independent of the number of buckets.

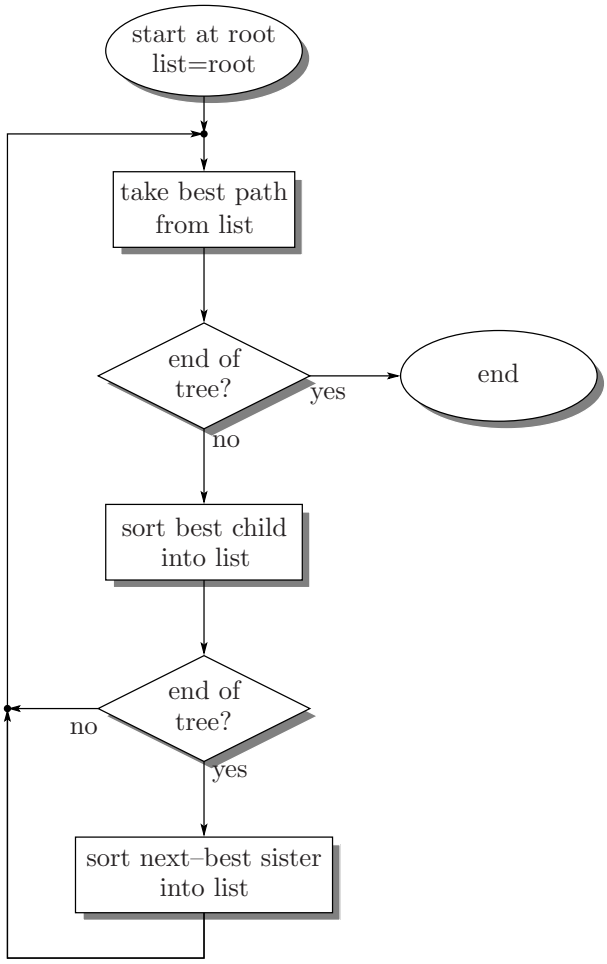

Figure 3.3: Flowchart of the modified stack decoder.

increasing metric) non–empty bucket and sorts its extensions into the buckets. If candidates are not sorted within these buckets the performance is somewhat degraded and complexity is increased slightly compared to the regular stack algorithm. When sorting is applied within the individual buckets, the bucket algorithm can be viewed as an implementation of the stack algorithm that uses a procedure similar to "quicksort" [PTV02] to sort the candidates into the stack.

In order to reduce the number of candidates examined by the stack or the bucket algorithm, i.e. to reduces the length of the list of candidates maintained by these decoders and thereby lower the complexity of sorting operations, Vinck [Vin84] proposed another clever variant of the stack algorithm. It differs from the regular stack algorithm only in the following way: Instead of replacing a path from the list with all of its children, this algorithm only sorts the path's *best child* and —unless all sisters have been examined previously— the *next–best sister* of this path into the list. For its flowchart see Fig. 3.3. Note that this enumeration strategy does not lead to a degradation of decoder performance compared to the regular stack algorithm and is

particularly useful for trees where a large number of branches emanate from each node. In fact, this algorithm examines the least number of nodes required to achieve optimal performance at all times.

3.1.1.3.2 Fano Algorithm The Fano algorithm was proposed together with the so–called Fano metric for sequential decoding of convolutional codes in [Fan63]. Instead of maintaining a list of candidate paths this algorithm considers only a single path at a time and searches the tree by extending and backtracking along this single path. It is therefore advantageous compared to the stack algorithm or its variants in cases where decoder memory is limited.

The flowchart of this algorithm is presented in Fig. 3.4. Here d_c, d_b and d_f denote the path metric of the current path, its parent ("look back") and one of its children ("look forward"), respectively. The essential idea behind this algorithm is that contrary to the stack algorithm, which always extends the most promising path from a given list of paths, this algorithm pursues a path from the root towards a leaf of the tree as long as it appears "promising" thereby *tightening a threshold* ρ such that $\rho := \lceil d_\mathrm{c}/\Delta_\mathrm{F} \rceil \Delta_\mathrm{F}$ ($\lceil x \rceil$: ceiling function) with a *stepsize* $\Delta_\mathrm{F} > 0$ and otherwise backtracks testing alternative paths.[4] At this the question whether a path is promising or not is decided adaptively using this variable threshold ρ, i.e. a path is extended (*move forward*) as long as the path metric d_f of its child currently under consideration (*look forward*) does not exceed the current threshold ρ. If at some point the decoder decides due to $d_\mathrm{f} > \rho$ that a child of this path might not be that promising after all, it backtracks along this path thereby looking for sister paths that might be more auspicious. This backtracking is performed in a way that the decoder first *looks back* to see whether the path metric d_b of the parent lies below the current threshold. If this is true, it will *move back*, i.e. the parent of the current path will become the new current path, and —if there are any left— will *look forward to the next–best child* again pursuing an alternative path as long as $d_\mathrm{f} \leq \rho$.[5] Thus, the decoder may come to the point where it can move neither forward nor backward, in which case it has to *relax the threshold*, i.e. $\rho := \rho + \Delta_\mathrm{F}$, and return to *looking forward to the best child*. If this happens, it means that the decoder has found that there are no paths in the tree that appear to be more promising than the path where it started backtracking for the last time and therefore moves forward again along this path to see whether it can be extended further given the relaxed threshold. At this point it should be clear that the above *tightening of the threshold* is to be performed only if a particular path is *visited* for the *first* time. This is the case when $d_\mathrm{b} > \rho - \Delta_\mathrm{F}$.[6] When the decoder reaches a leaf (*end of the tree*) the search is terminated and the current path is returned as decoder output.

Apparently, compared to the stack algorithm the savings in memory requirements are traded

[4]Note that in the step *move forward* $d_\mathrm{c} := d_\mathrm{f}$, as the child of the current path becomes the new current path.

[5]To keep the decoder from moving back beyond the root of the tree, $d_\mathrm{b} = \infty$ is assumed if the decoder looks

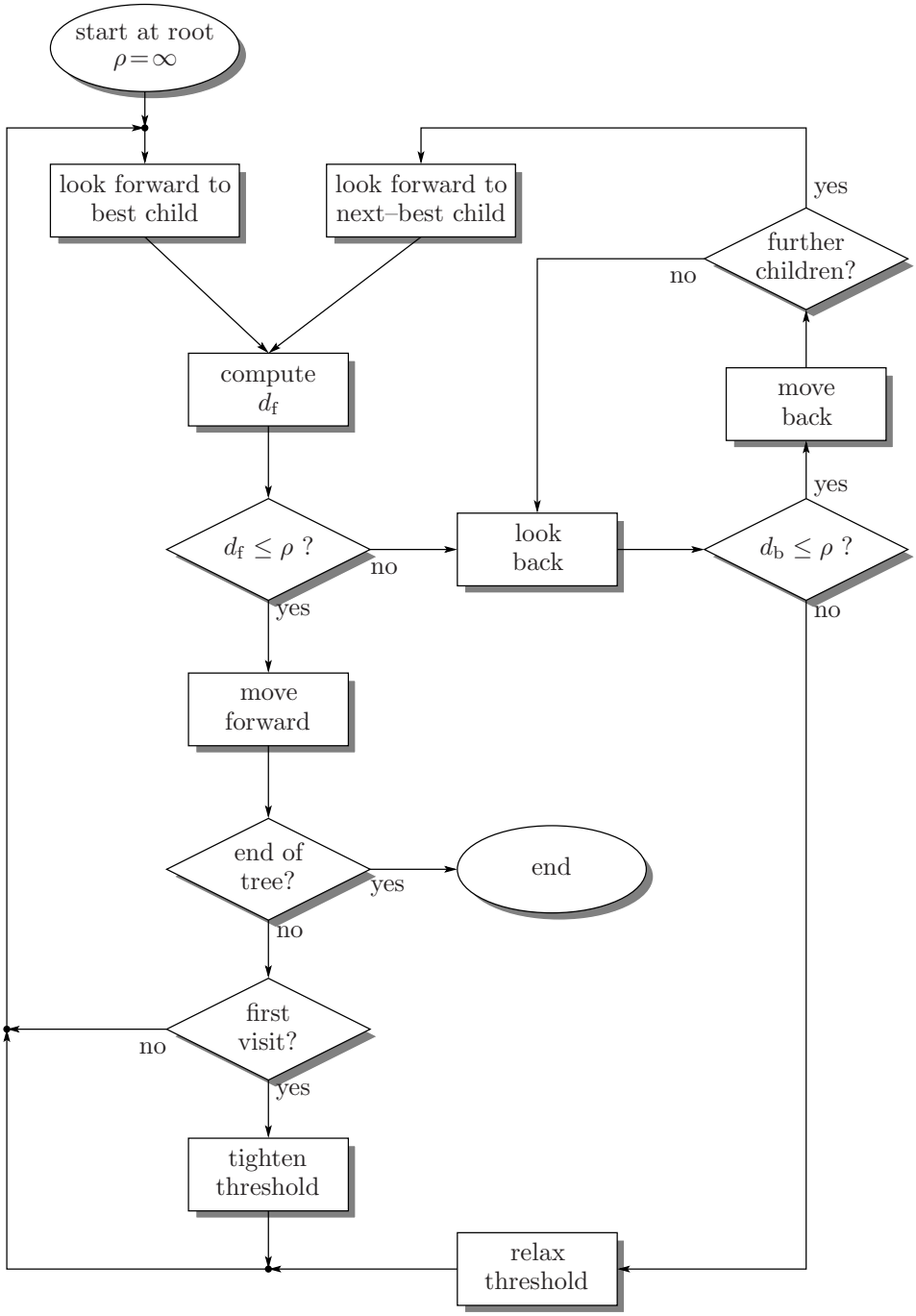

Figure 3.4: Flowchart of the Fano algorithm [JZ99].

for an increased computational complexity as some paths may be considered repeatedly. The latter effect is particularly pronounced if the path metric of the correct path has a positive drift as the decoder is likely to backtrack to the root of the tree after every move forward even if it proceeds along the correct path.

Due to the quantization of the threshold in multiples of $\Delta_F > 0$, there is a non–zero probability that the Fano decoder leads to a suboptimal decoder output even when the branch metrics are strictly non–negative. In [Gei73] a modified Fano algorithm was proposed, in which the quantization of the threshold was eliminated, such that the decoder is optimal. However, the difference in performance is usually negligible and the latter variant is more complex and therefore not considered in the following.

3.1.1.3.3 Agrell Sphere Decoder The last algorithm we want to consider here is a rather straightforward generalization of the algorithm presented by Agrell et al. in [AEVZ02] for finding the closest point in an infinite lattice to tree search with arbitrary, i.e. (possibly varying) finite or infinite, numbers of branches emanating from every node. It essentially combines the work of Fincke and Pohst [Poh81, FP85], Schnorr and Euchner [SE94] and Babai [Bab86]. Since the geometric interpretation of this algorithm is that it restrains the search to hyperspheres of possibly decreasing radii we chose to subsequently refer to this algorithm as *sphere decoder* (SpD), a term frequently used for this type of algorithm in the recent literature, cf. e.g. [VB99, AEVZ02, BGBF03, DEC03, CT04a, GN04, JO05b, LXW+05, HV05, MGDC06, SVH06]. For distinction from the Fincke–Pohst SpD (FP–SpD, cf. Section 3.1.1.2.3) we add the prefix *Agrell*, i.e. subsequently refer to this algorithm as *Agrell sphere decoder* (*A–SpD*). The flowchart of this tree–search algorithm is presented in Fig. 3.5. In order to highlight the algorithmic similarities and differences when compared to the Fano algorithm we chose a graphical representation based on the same functional blocks and variables as in the flowchart of the Fano algorithm, cf. Fig. 3.4.

The sphere decoder is similar to the Fano algorithm in various ways:

- It does not maintain a list of candidate paths but only considers a single candidate path, which it extends and along which it backtracks, plus a possibly previously found preliminary result.
- A candidate path is extended (*move forward*) as long as the path metric d_f of its child under consideration does not exceed the threshold, i.e. as long as $d_f \leq \rho$.

back from the root.

[6] In the description of the Fano algorithm in [JZ99] an additional constraint corresponding to $d_c \leq \rho - \Delta_F$ was introduced. While it does not provide any information regarding the first–visit question, it serves the purpose of avoiding the computation of $\rho := \lceil d_c/\Delta_F \rceil \Delta_F$ that is superfluous if $\rho - \Delta < d_c \leq \rho$ as the threshold ρ is already tight in this case.

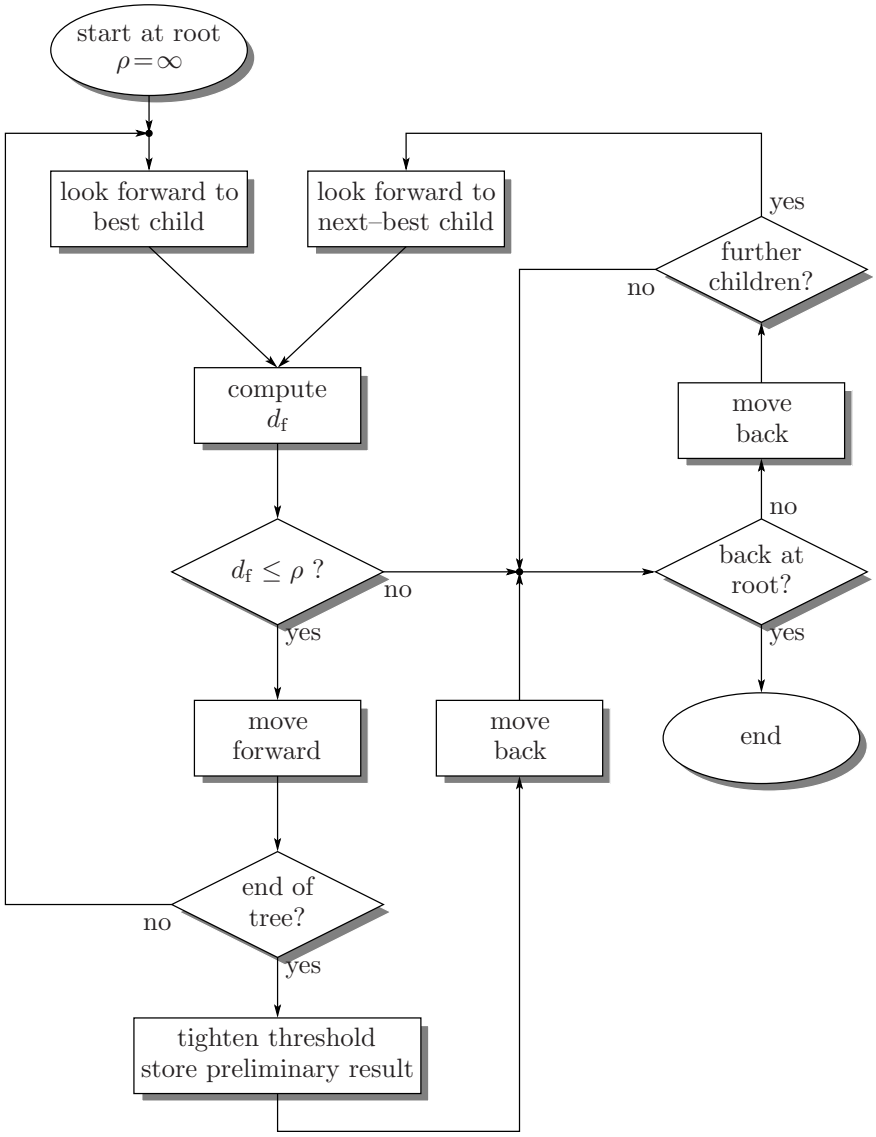

Figure 3.5: Flowchart of the Agrell sphere decoder (A–SpD).

- When looking forward from a particular node, the algorithm always tests the children in order of increasing branch metric, i.e. the best child is tested before the second best and so on. In the lattice / sphere decoding literature this strategy is commonly referred to as Schnorr–Euchner enumeration as it was introduced in this field of research by Schnorr and Euchner in [SE94]. Note however, that this enumeration strategy has been introduced into the general field of tree–search decoding much earlier in e.g. Fano's algorithm [Fan63] and Gallager's single–stack algorithm [Gal74].

However, the A–SpD differs from the Fano algorithm in the following major aspects:

- The threshold (in this context frequently referred to as *sphere radius*) ρ is tightened via $\rho := d_\mathrm{c}$ only when the decoder has found a (new) path of full length with $d_\mathrm{c} \leq \rho$.

- Rather than terminating the search at this point, the current path is stored as preliminary result and the search continues by backtracking along the current path and pursuing alternative paths until the metrics d_{f} of their best children exceed the current threshold ρ. The search is terminated when the decoder has returned to the root and either (i) $d_{\mathrm{f}} > \rho$, or (ii) it has examined all branches emanating from the root.

- Since the threshold ρ is never increased during the search process the decoder never examines branches more than once.

- As there is no quantization of the threshold ρ this algorithm is an optimal tree–search algorithm.

The probably most prominent advantage of this algorithm compared to the above algorithms from sequential decoding lies in the fact that it very quickly determines a preliminary decoding result by moving directly, i.e. without any backtracking, from the root to a leaf of the tree always choosing the best child. This candidate is in the lattice / sphere decoding literature usually referred to as "Babai nearest plane point" as this part of the sphere decoder coincides with an algorithm proposed by Babai in [Bab86]. In the context of communications this greedy decoding process would be referred to as a sort of decision–feedback equalization (DFE), cf. e.g. [Pro00]. After having determined the DFE solution the A–SpD continues to look for better candidates until it either terminates as described above or is terminated externally e.g. when a prescribed maximal decoding complexity is exceeded. Consequently, the decoder will contrary to the above algorithms always produce an output when the maximal complexity is limited.[7]

For further illustration, Fig. 3.6 shows a random example of a search–tree generated by the A–SpD in a quaternary tree of depth five. The branch metrics are assumed to be strictly non–negative, i.e. an optimal tree–search decoder finds the leaf with minimal path metric. Branches examined by the decoder are marked as solid lines and path metrics are written near the corresponding nodes, while the numbers next to the branches indicate the order in which branches have been examined. The full–length path with the smallest path metric is highlighted using the bold line. The dotted lines represent unexamined sisters of examined paths. One can observe that —even though the first path of full length found by the A–SpD is not the optimal path— the algorithm in this example quickly terminates after examining 19 branches having found the optimal path in step 15. For comparison the search–tree consisting only of branches that must be examined to find the optimal leaf and to be sure of it is depicted in Fig. 3.7. This search–tree would be generated e.g. by Vinck's modified stack algorithm (cf. Section 3.1.1.3.1).

[7]Note that the minimal complexity in terms of the number of examined nodes of the Vinck–stack and the Fano algorithm is equal to the complexity of the A–SpD determining the DFE solution. Limiting the complexity to smaller values than this is meaningless.

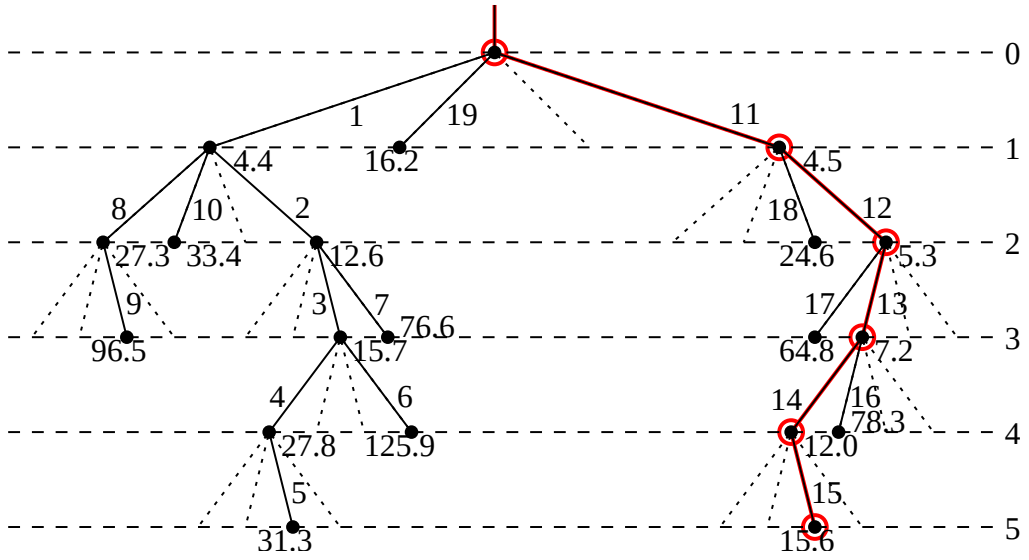

Figure 3.6: Example of a search–tree generated by the A–SpD for the example of a quaternary tree of depth five.

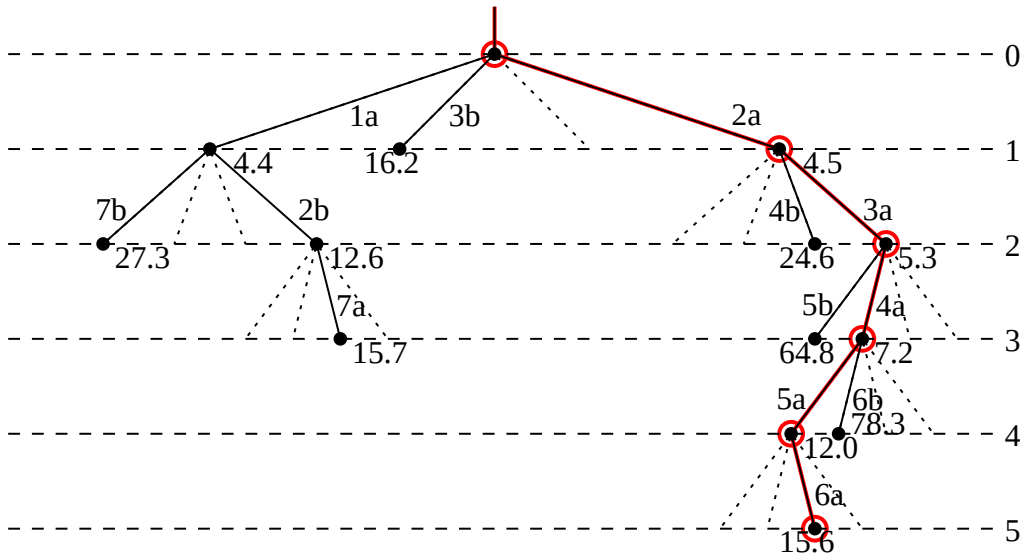

Figure 3.7: Minimal search–tree of an optimal tree–search algorithm corresponding to the example of Fig. 3.6.

Labels xa and xb mark branches generated by Vinck's algorithm as next–best sister and best child of the path at the top of the stack in iteration $x \in \{1, \ldots, 7\}$ of the tree–search process, respectively. It can be seen that the complexity of the A–SpD (19 examined branches) is quite close to that of the most efficient optimal algorithm (13 branches) and significantly lower than the number of leaves of the tree (1024 leaves).

Modification: Finite Initial Threshold: Due to the fact that the A–SpD heads straight for the DFE solution in the first steps of the decoding process a finite threshold ρ is not required at initialization. However, if the DFE path deviates early in the decision process, i.e. close to the root, from the correct path, the first finite value of ρ will usually be relatively large and thus the decoder will only converge slowly to the true solution. It may therefore be desirable to "guide" the tree–search towards the correct path by using a finite initial radius $\rho := \rho_{\text{init}} < \infty$ chosen as small as possible, but such that the true solution has a path metric $d_{\text{c}} \leq \rho$ with high probability and restart the sphere decoder with an increased threshold ρ in case no leaf was found with $d_{\text{c}} \leq \rho$, cf. e.g. [DEC03]. The appropriate choice of the finite initial threshold ρ_{init} when applying this algorithm to MSDD will be discussed in detail in Section 5.5.1.2.1.

Remark: The A–SpD as depicted in Fig. 3.5 is also quite closely related to Gallager's single–stack algorithm [Gal74], whose flowchart is depicted in Fig. 3.8. Recall that the A–SpD repeatedly updates its threshold ρ and continues its search until all but one candidate path of full length have been eliminated from the decoding tree. The single–stack algorithm on the other hand uses a fixed threshold, and searches the tree until it has either found one path of full length whose metric lies below the threshold ρ (*end (success)*) or has pursued all paths in the tree up to the point where $d_{\text{f}} > \rho$ without finding a path of full length, whose metric lies below the threshold (*end (failure)*). Clearly, this algorithm is suboptimal and is discussed here only due to its close relation to the A–SpD.[8]

3.1.1.3.4 Summary Let us briefly summarize the advantages and disadvantages of the algorithms considered above.

Performance: The stack algorithm, its variant due to Vinck and the A–SpD as discussed in Sections 3.1.1.3.1 and 3.1.1.3.3, respectively, are optimal in the sense that they find the leaf with minimal metric, provided that the branch metrics are strictly non–negative and that neither storage nor decoding time are limited. Due to the non–zero stepsize Δ_{F} the Fano algorithm operates at a loss in performance, which however is usually negligible.

Complexity: Among the above optimal tree–search algorithms, i.e. stack, Fano and A–

[8]When proposing this algorithm in [Gal74] Gallager was concerned with tree–search–based source coding and content with finding a representation of the source data such that the distortion after decoding did not exceed a desired maximal distortion.

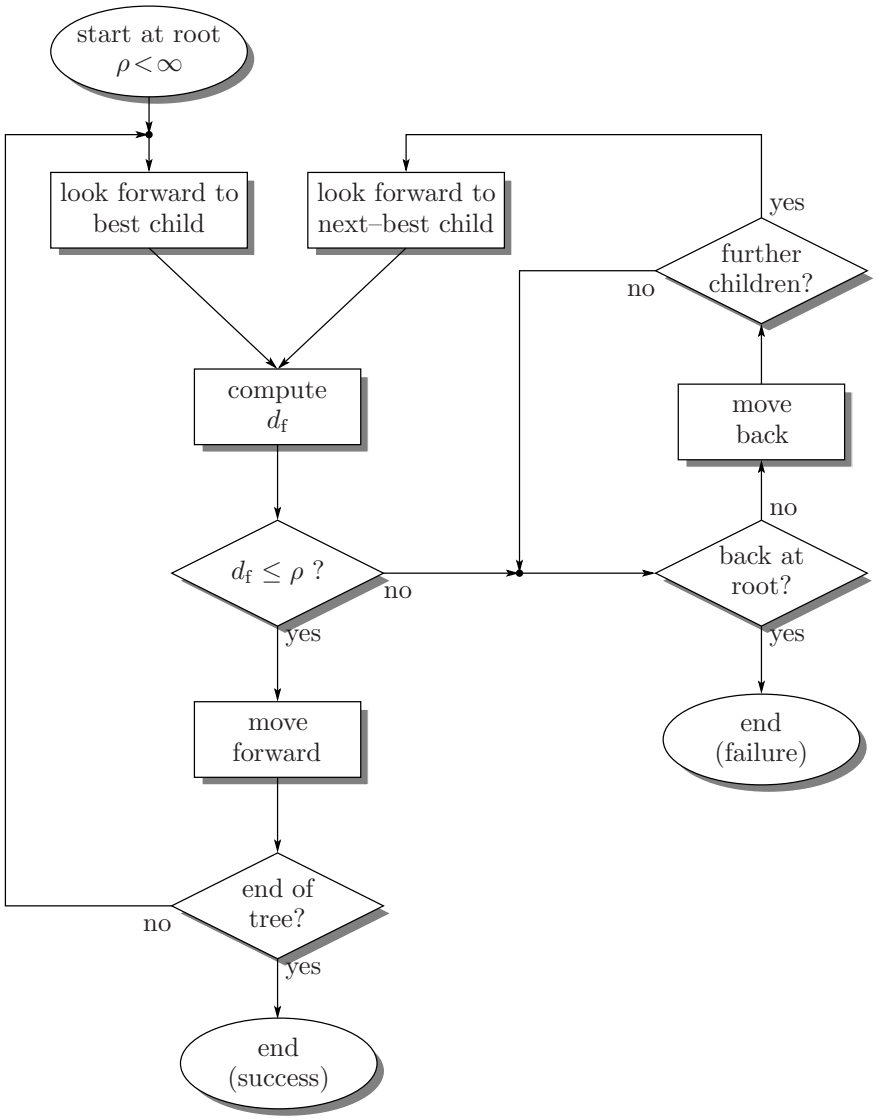

Figure 3.8: Flowchart of Gallager's single–stack algorithm.

SpD algorithm, the stack algorithm has the lowest complexity in terms of average number of examined branches, because it only extends paths whose metric lies below that of the final decoder output, cf. also [XWZW04]. However, when storage space is limited the Fano and the A–SpD algorithms are interesting alternatives especially for trees with a large number of branches emanating from individual nodes. When the maximal decoder complexity is limited the A–SpD is clearly preferable due to his strategy of successive refinement of preliminary decoding results. Compared to the Fano algorithm the A–SpD may suffer from a too large initial threshold, whereas it is advantageous in that it does not consider candidates repeatedly. The earlier problem can be ameliorated by initializing the A–SpD with a finite threshold ρ_{init} that is increased later if no path of full length, whose metric lies below the initial threshold, is found.

3.1.2 Metric Calculation for Tree–Search MSDD

Having reviewed some fundamentals of tree–search decoding we will in this section derive a representation of the MSDD decision rule (cf. Section 2.4.2) that is amenable to the application of tree–search algorithms. More specifically, following some elementary definitions we will in Section 3.1.2.2 present a metric structure, that allows for efficient *ML* MSDD based on tree–search decoding. There, we also establish a relationship between the ML–MSDD metric and linear MMSE prediction. The insights that can be derived from this connection lead to a modification of the ML metric that is presented in Section 3.1.2.3 and leads to further often significant savings in computational complexity at the expense of a slight loss in power efficiency.

3.1.2.1 Notation

Recall, that in Section 2.4.2 we collected the data, transmit and receive symbols, channel coefficients and noise samples involved in the decision process for an MSDD block in block–matrices $\bar{\boldsymbol{V}}$, $\bar{\boldsymbol{S}}$, $\bar{\boldsymbol{R}}$, $\bar{\boldsymbol{G}}$ and $\bar{\boldsymbol{N}}$ of dimensions $((N-1)N_{\mathrm{S}} \times N_{\mathrm{S}})$, $(NN_{\mathrm{S}} \times N_{\mathrm{S}})$, $(NN_{\mathrm{S}} \times N_{\mathrm{R}})$, $(NN_{\mathrm{S}} \times N_{\mathrm{R}})$ and $(NN_{\mathrm{S}} \times N_{\mathrm{R}})$, respectively, cf. (2.18)–(2.22). For the sake of readability, in the following we use subscripts to address $(N_{\mathrm{S}} \times N_{\mathrm{S}})$–, $(N_{\mathrm{S}} \times N_{\mathrm{R}})$–dimensional sub–matrices $\boldsymbol{V}_i$, $\boldsymbol{S}_i$ and $\boldsymbol{R}_i$, $\boldsymbol{G}_i$, $\boldsymbol{N}_i$ of the above block–matrices $\bar{\boldsymbol{V}}$, $\bar{\boldsymbol{S}}$ and $\bar{\boldsymbol{R}}$, $\bar{\boldsymbol{G}}$, $\bar{\boldsymbol{N}}$, respectively. This way, we have e.g. $\bar{\boldsymbol{S}} = \left[\boldsymbol{S}_1^{\mathsf{H}}, \ldots, \boldsymbol{S}_N^{\mathsf{H}}\right]^{\mathsf{H}}$, i.e. $\boldsymbol{S}_n \triangleq \boldsymbol{S}[k-N+n]$, $1 \leq n \leq N$, but $\bar{\boldsymbol{V}} = \left[\boldsymbol{V}_1^{\mathsf{H}}, \ldots, \boldsymbol{V}_{N-1}^{\mathsf{H}}\right]^{\mathsf{H}}$ with $\boldsymbol{V}_n \triangleq \boldsymbol{V}[k-N+1+n]$, $1 \leq n \leq N-1$, and

$$\boldsymbol{S}_{n+1} = \boldsymbol{V}_n \boldsymbol{S}_n. \tag{3.3}$$

Furthermore, we introduce the notation of

$$\tilde{n} \triangleq N - n + 1, \tag{3.4}$$

$((\tilde{n}-1)N_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional

$$\bar{\boldsymbol{V}}_n \triangleq \left[\boldsymbol{V}_n^{\mathsf{H}}, \ldots, \boldsymbol{V}_{N-1}^{\mathsf{H}}\right]^{\mathsf{H}}, \tag{3.5}$$

and $(\tilde{n}N_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional

$$\bar{\boldsymbol{S}}_n \triangleq \left[\boldsymbol{S}_n^{\mathsf{H}}, \ldots, \boldsymbol{S}_N^{\mathsf{H}}\right]^{\mathsf{H}}. \tag{3.6}$$

In generalization of (2.38) we further define the $(\tilde{n}N_{\mathrm{S}} \times \tilde{n}N_{\mathrm{S}})$–dimensional block–diagonal matrix

$$\tilde{\bar{\boldsymbol{S}}}_{\mathrm{D},n} \triangleq \mathrm{diag}\left\{\tilde{\bar{\boldsymbol{S}}}_n\right\} = \begin{bmatrix} \tilde{\boldsymbol{S}}_n & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \tilde{\boldsymbol{S}}_N \end{bmatrix}, \tag{3.7}$$

with $(\tilde{n}N_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional

$$\tilde{\bar{\boldsymbol{S}}}_n \triangleq \left[\tilde{\boldsymbol{S}}_n^{\mathsf{H}}, \ldots, \tilde{\boldsymbol{S}}_N^{\mathsf{H}}\right]^{\mathsf{H}} \tag{3.8}$$

$$\triangleq \left[\left(\prod_{i=n}^{N-1}\tilde{\boldsymbol{V}}_i^{\mathsf{H}}\right)^{\mathsf{H}}, \left(\prod_{i=n+1}^{N-1}\tilde{\boldsymbol{V}}_i^{\mathsf{H}}\right)^{\mathsf{H}}, \ldots, \left(\prod_{i=N-1}^{N-1}\tilde{\boldsymbol{V}}_i^{\mathsf{H}}\right)^{\mathsf{H}}, \boldsymbol{I}_{N_{\mathrm{S}}}\right]^{\mathsf{H}} \tag{3.9}$$

$$= \left[\prod_{i=1}^{N-n}\tilde{\boldsymbol{V}}_{N-i}, \prod_{i=1}^{N-n-1}\tilde{\boldsymbol{V}}_{N-i}, \ldots, \prod_{i=1}^{2}\tilde{\boldsymbol{V}}_{N-i}, \tilde{\boldsymbol{V}}_{N-1}, \boldsymbol{I}_{N_{\mathrm{S}}}\right]^{\mathsf{H}}. \tag{3.10}$$

These include $(NN_{\mathrm{S}} \times NN_{\mathrm{S}})$–dimensional $\tilde{\bar{\boldsymbol{S}}}_{\mathrm{D}} = \tilde{\bar{\boldsymbol{S}}}_{\mathrm{D},1}$ and $(NN_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional $\tilde{\bar{\boldsymbol{S}}} = \tilde{\bar{\boldsymbol{S}}}_1$ as special cases, respectively.

3.1.2.2 ML–MSDD Metric

In Section 2.4.2 we showed that in ML MSDD the expression [cf. (2.36)]

$$d_1\left(\tilde{\bar{\boldsymbol{S}}}\right) \triangleq \mathrm{tr}\left\{\bar{\boldsymbol{R}}^{\mathsf{H}}\tilde{\bar{\boldsymbol{S}}}_{\mathrm{D}}\left(\boldsymbol{M}\otimes\boldsymbol{I}_{N_{\mathrm{S}}}\right)\tilde{\bar{\boldsymbol{S}}}_{\mathrm{D}}^{\mathsf{H}}\bar{\boldsymbol{R}}\right\} \tag{3.11}$$

is to be minimized over all L^{N-1} $\tilde{\bar{\boldsymbol{S}}}$ corresponding to different data sequences $\tilde{\bar{\boldsymbol{V}}} \in \boldsymbol{\mathcal{V}}^{N-1}$. In the sequel, we will refer to $d_1\left(\tilde{\bar{\boldsymbol{S}}}\right)$ as ML–MSDD metric (of $\tilde{\bar{\boldsymbol{S}}}$).

In order to make ML–MSDD amenable to tree–search decoding, we apply the Cholesky decomposition (cf. e.g. [GvL96]) to the matrix $\boldsymbol{M}$ [cf. (2.37)]

$$\boldsymbol{M} = \boldsymbol{C}^{\mathsf{H}}\boldsymbol{C}, \tag{3.12}$$

which yields an $(N \times N)$–dimensional upper–right triangular matrix $\boldsymbol{C}$. Using $(\boldsymbol{XY})\otimes\boldsymbol{Z} = (\boldsymbol{X}\otimes\boldsymbol{Z})\cdot(\boldsymbol{Y}\otimes\boldsymbol{Z})$ for arbitrary matrices $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$ of appropriate dimensions and $\mathrm{tr}\{\boldsymbol{X}^{\mathsf{H}}\boldsymbol{X}\} = ||\boldsymbol{X}||^2$ for any matrix $\boldsymbol{X}$ we obtain

$$d_1\left(\tilde{\bar{\boldsymbol{S}}}\right) = \left|\left|\left(\boldsymbol{C}\otimes\boldsymbol{I}_{N_{\mathrm{S}}}\right)\tilde{\bar{\boldsymbol{S}}}_{\mathrm{D}}^{\mathsf{H}}\bar{\boldsymbol{R}}\right|\right|^2 \tag{3.13}$$

$$= \sum_{n=1}^{N}\left|\left|\sum_{j=n}^{N} c_{n,j}\tilde{\boldsymbol{S}}_j^{\mathsf{H}}\boldsymbol{R}_j\right|\right|^2 \tag{3.14}$$

$$= \sum_{n=1}^{N}\left|\left|\breve{\boldsymbol{R}}_{n,n}^{\mathsf{H}}\tilde{\boldsymbol{S}}_n + \boldsymbol{X}_n\right|\right|^2 \tag{3.15}$$

$$\triangleq \sum_{n=1}^{N}\delta_n\left(\tilde{\bar{\boldsymbol{S}}}_n\right) \tag{3.16}$$

where

$$\breve{\boldsymbol{R}}_{n,j} \triangleq c_{n,j}\boldsymbol{R}_j, \quad \text{and} \tag{3.17}$$

$$\boldsymbol{X}_n \triangleq \sum_{j=n+1}^{N}\breve{\boldsymbol{R}}_{n,j}^{\mathsf{H}}\tilde{\boldsymbol{S}}_j, \tag{3.18}$$

for $1 \leq n \leq N,\ n \leq j \leq N$.

It can be observed that the ML–MSDD metric in (3.16) is a *sum of N strictly non–negative scalar terms*

$$\delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) \stackrel{\triangle}{=} \Big|\Big|\breve{\boldsymbol{R}}_{n,n}^{\mathsf{H}}\tilde{\boldsymbol{S}}_n + \boldsymbol{X}_n\Big|\Big|^2, \quad 1 \leq n \leq N, \tag{3.19}$$

which depend only on $\tilde{n} = N - n + 1$ symbols $\tilde{\boldsymbol{S}}_j,\ n \leq j \leq N$, [cf. (3.18)]. Thus, ML MSDD can be solved by means of tree–search decoding in a tree of maximal depth $(N-1)$. At this, $\tilde{\boldsymbol{S}}_n$ (or equivalently $\tilde{\boldsymbol{V}}_n$) are the labels of branches leading to nodes at depth $(N-n)$, $1 \leq n \leq N-1$, of the tree, and a partial candidate sequence $\tilde{\tilde{\boldsymbol{S}}}_n$ as defined in (3.8) (or equivalently $\tilde{\tilde{\boldsymbol{V}}}_n \stackrel{\triangle}{=} \big[\tilde{\boldsymbol{V}}_n^{\mathsf{H}}, \ldots, \tilde{\boldsymbol{V}}_{N-1}^{\mathsf{H}}\big]^{\mathsf{H}}$) represents a path to a node at depth $(N-n)$ of the tree. It is important to mark this reversed assignment between the subscript n of $\tilde{\tilde{\boldsymbol{S}}}_n$ and the depth $(N-n)$ of the tree, i.e. the root of the tree corresponds to $\tilde{\boldsymbol{S}}_N = \boldsymbol{I}_{N_\mathrm{S}}$, whereas nodes corresponding to candidates $\tilde{\tilde{\boldsymbol{S}}}_n$ lie at depth $(N-n)$ of the tree. In particular, leaves of the tree represent "full–length" candidates $\tilde{\tilde{\boldsymbol{S}}}$. This is due to the *upper* triangular structure of the matrix $\boldsymbol{C}$ [cf. (3.12)]. In this context $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$, $1 \leq n \leq N-1$, in (3.19) represents the metric of a branch labeled by a symbol $\tilde{\boldsymbol{S}}_n$ emanating from a node corresponding to $\tilde{\tilde{\boldsymbol{S}}}_{n+1}$. The path metric corresponding to a path $\tilde{\tilde{\boldsymbol{S}}}_n$ is then given by

$$\begin{aligned} d_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) &\stackrel{\triangle}{=} \sum_{i=n}^{N} \Big|\Big|\breve{\boldsymbol{R}}_{i,i}^{\mathsf{H}}\tilde{\boldsymbol{S}}_i + \boldsymbol{X}_i\Big|\Big|^2 && (3.20)\\ &= d_{n+1}\Big(\tilde{\tilde{\boldsymbol{S}}}_{n+1}\Big) + \delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big), \quad N-1 \geq n \geq 1, && (3.21) \end{aligned}$$

where $d_N\big(\tilde{\tilde{\boldsymbol{S}}}_N\big) = \delta_N\big(\tilde{\tilde{\boldsymbol{S}}}_N\big) = \big|\big|\breve{\boldsymbol{R}}_{N,N}\big|\big|^2$.[9] In these settings the leaf with the minimal metric $d_1\big(\tilde{\tilde{\boldsymbol{S}}}\big)$ corresponds to the ML–MSDD solution $\hat{\tilde{\boldsymbol{S}}}$ or equivalently —through reversal of (2.11)— $\hat{\boldsymbol{V}}$. The application of tree–search algorithms to MSDD will be discussed in more detail in Section 3.1.3.

Finally, note that since the branch metrics $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ are strictly non–negative tree–search algorithms that are optimal in the sense of Section 3.1.1.2.3 will achieve ML–MSDD performance.

Relation between ML MSDD and Linear MMSE Prediction: It is also worth pointing out that ML MSDD is related to linear backward MMSE prediction in that

$$c_{n,j} = \frac{p_{\mathrm{B},j-n}^{(N-n)}}{\sigma_{\mathrm{p}}^{(N-n)}}, \tag{3.22}$$

[9] Note that one could just as well choose $d_N\big(\tilde{\tilde{\boldsymbol{S}}}_N\big) = 0$, without any impact on the complexity or performance of the tree–search decoders. $d_N\big(\tilde{\tilde{\boldsymbol{S}}}_N\big) = \big|\big|\breve{\boldsymbol{R}}_{N,N}\big|\big|^2$ was chosen to facilitate the analysis of performance and complexity of decoders in later sections as $d_1\big(\tilde{\tilde{\boldsymbol{S}}}\big) = \mathrm{tr}\Big\{\bar{\boldsymbol{R}}^{\mathsf{H}}\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{D}}\left(\boldsymbol{M}\otimes\boldsymbol{I}_{N_\mathrm{S}}\right)\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{D}}^{\mathsf{H}}\bar{\boldsymbol{R}}\Big\}$ in this case.

where $p_{\mathrm{B},j}^{(n)}$, $1 \leq j \leq n$, and $\sigma_{\mathrm{p}}^{(n)}$ denote the jth coefficient of the nth–order linear backward MMSE predictor for the discrete time fading–plus–noise random process $\boldsymbol{G}[k] + \boldsymbol{N}[k]$ and the corresponding standard deviation of the prediction error, respectively, and $p_{\mathrm{B},0}^{(n)} = -1, \forall n$, cf. e.g. [VT95a]. I.e. the non–zero entries $c_{n,j}$, $n \leq j \leq N$ in the nth row of $\boldsymbol{C}$ are the coefficients of the $(N-n)$th order prediction error filter normalized by the corresponding standard deviation of the prediction error. This means that —provided that all $\tilde{\boldsymbol{S}}_j$, $n+1 \leq j \leq N$, are correct up to a common rotation with an arbitrary unitary matrix— the quantity $\boldsymbol{X}_n \sigma_{\mathrm{p}}^{(N-n)}$ [cf. (3.18)] represents the $(N-n)$th–order MMSE prediction of $\boldsymbol{G}_n^{\mathsf{H}} + \boldsymbol{N}_n^{\mathsf{H}}$ and $\breve{\boldsymbol{R}}_{n,n}^{\mathsf{H}} \tilde{\boldsymbol{S}}_n + \boldsymbol{X}_n$ can be viewed as the corresponding prediction error. Since $\breve{\boldsymbol{R}}_{n,n}^{\mathsf{H}} \tilde{\boldsymbol{S}}_n = \big(\boldsymbol{G}_n^{\mathsf{H}} \boldsymbol{S}_n^{\mathsf{H}} + \boldsymbol{N}_n^{\mathsf{H}}\big) \tilde{\boldsymbol{S}}_n$ and the distribution of the AWGN matrix $\boldsymbol{N}_n$ is invariant to unitary transformations, choosing $\tilde{\boldsymbol{S}}_n$ such that $\delta_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$ in (3.19) is minimized, corresponds to $(N-n)$th–order linear backward MMSE prediction.

From this two conclusions can be derived that will prove useful later on. First, consider two different values of N, namely N_1 and $N_2 > N_1$. Then it is clear from the above discussion that —under the assumption of identical statistical properties of the channel ($\psi_{gg}[\kappa]$ and σ_n^2)— the $(N_1 \times N_1)$–dimensional matrix $\boldsymbol{C}_1$ obtained from the $(N_1 \times N_1)$–dimensional inverse fading–plus–noise correlation matrix $\boldsymbol{M}_1$ via Cholesky factorization (3.12) is a lower–right corner submatrix of $(N_2 \times N_2)$–dimensional $\boldsymbol{C}_2$ obtained from the respective $(N_2 \times N_2)$–dimensional matrix $\boldsymbol{M}_2$, even though $\boldsymbol{M}_1$ is not a submatrix of $\boldsymbol{M}_2$. This means that results on full–length candidate sequences $\tilde{\bar{\boldsymbol{S}}}$ and the statistical properties of $d_1\big(\tilde{\bar{\boldsymbol{S}}}\big)$ that can be derived based on the representation $d_1\big(\tilde{\bar{\boldsymbol{S}}}\big) = \mathrm{tr}\big\{\bar{\boldsymbol{R}}^{\mathsf{H}} \tilde{\bar{\boldsymbol{S}}}_{\mathrm{D}} \left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{S}}}\right) \tilde{\bar{\boldsymbol{S}}}_{\mathrm{D}}^{\mathsf{H}} \bar{\boldsymbol{R}}\big\}$ apply equally to partial candidates $\tilde{\bar{\boldsymbol{S}}}_n$, $2 \leq n \leq N-1$, and their metrics $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$.

Second, it follows from the normalization of the prediction coefficients $p_{\mathrm{B},j}^{(n)}$ with the corresponding prediction–error standard deviation $\sigma_{\mathrm{p}}^{(n)}$ [cf. (3.22)] that the elements of $\breve{\boldsymbol{R}}_{n,n}^{\mathsf{H}} \tilde{\boldsymbol{S}}_n + \boldsymbol{X}_n$ [cf. (3.19)] along the *correct path*, i.e. with $\tilde{\bar{\boldsymbol{S}}} = \bar{\boldsymbol{S}}$, are iid $\mathcal{N}_{\mathrm{c}}(0,1)$ distributed random variables. Hence, the $\delta_n\big(\bar{\boldsymbol{S}}_n\big)$ are $\chi^2(N_{\mathrm{S}} N_{\mathrm{R}}, 2N_{\mathrm{S}} N_{\mathrm{R}})$ distributed ($\chi^2(\sigma^2, K)$: central χ^2 distribution with variance σ^2 and K degrees of freedom), such that the expected branch metrics along the *correct path* are given by

$$\mathcal{E}\big\{\delta_n\big(\bar{\boldsymbol{S}}_n\big)\big\} = N_{\mathrm{S}} N_{\mathrm{R}}. \tag{3.23}$$

It is important to note that this holds *regardless of n and N, the SNR or the temporal correlation $\psi_{gg}[\kappa]$ of the fading process*. Accordingly, the expected path metric along the correct path grows linearly with $\tilde{n}$ [cf. (3.4)], i.e.

$$\mathcal{E}\big\{d_n\big(\bar{\boldsymbol{S}}_n\big)\big\} = \tilde{n} N_{\mathrm{S}} N_{\mathrm{R}}. \tag{3.24}$$

This has detrimental effects on the complexity of tree–search decoding, as it leads to an increasing probability that other branches besides those that form the correct path are investigated towards lower levels of the tree. Especially, towards lower SNR, where the normalized predictor

coefficients $p_{\mathrm{B},j}^{(n)}/\sigma_{\mathrm{P}}^{(n)}$, $n+1 \leq j \leq N$, become very small, $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$ becomes less dependent on tentative decisions $\tilde{\boldsymbol{V}}_j$, $n+1 \leq j \leq N-1$ represented by $\tilde{\bar{\boldsymbol{S}}}_{n+1}$. Consequently, towards lower SNR the probability $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \leq d_1\big(\bar{\boldsymbol{S}}\big)\big)$ is increasingly close to one for all $\tilde{\bar{\boldsymbol{S}}}_n$ and $2 \leq n \leq N-1$, which means that "optimal" tree–search decoders, that must (at least) examine all paths with $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \leq d_1\big(\bar{\boldsymbol{S}}\big)$, become computationally very inefficient. It is therefore advisable to use a modified metric in tree–search decoding that takes the length of a path into account.

3.1.2.3 Fano–Type Metric

The problem with tree–search decoding based on the ML decision rule is that the probability of a path is a monotonically decreasing function of its length, cf. Section 3.1.2.2. Thus, tree–search decoders are likely to investigate shorter paths even though they have —considering their shortness— already a rather poor metric and are therefore unlikely to be part of the ML path. In order to solve this problem, i.e. to have a metric that allows for fair comparisons between paths of different lengths, Fano proposed a metric for tree–search decoding of convolutional codes in [Fan63], the well–known "Fano metric". While his motivation appeared to be rather heuristic, it was later justified by Massey [Mas72] by means of a probabilistic derivation to be the appropriate means for comparing paths of different lengths.

The essential idea is to consider the a–posteriori probability $\Pr\big(\tilde{\bar{\boldsymbol{S}}}_n \mid \bar{\boldsymbol{R}}_n\big)$, even though a–priori information, that is commonly included in the a–posteriori probability, is not taken into account. In particular, we consider $-\log\big(\Pr\big(\tilde{\bar{\boldsymbol{S}}}_n \mid \bar{\boldsymbol{R}}_n\big)\big)$, and write using Bayes' rule (cf. e.g. [CT91])

$$-\log\Big(\Pr\Big(\tilde{\bar{\boldsymbol{S}}}_n \,\Big|\, \bar{\boldsymbol{R}}_n\Big)\Big) = -\log\left(\frac{p\Big(\bar{\boldsymbol{R}}_n \,\Big|\, \tilde{\bar{\boldsymbol{S}}}_n\Big)\Pr\Big(\tilde{\bar{\boldsymbol{S}}}_n\Big)}{\sum_{\forall \breve{\bar{\boldsymbol{S}}}_n} p\Big(\bar{\boldsymbol{R}}_n \,\Big|\, \breve{\bar{\boldsymbol{S}}}_n\Big)\Pr\Big(\breve{\bar{\boldsymbol{S}}}_n\Big)}\right), \quad 1 \leq n \leq N-1, \tag{3.25}$$

where $\breve{\bar{\boldsymbol{S}}}_n$ is defined as $\tilde{\bar{\boldsymbol{S}}}_n$ [cf. (3.8)] and introduced only to avoid ambiguities in the formulas. Assuming that all possible $\breve{\bar{\boldsymbol{S}}}_n$ are a priori equally probable, i.e. $\Pr\big(\breve{\bar{\boldsymbol{S}}}_n\big) = L^{-N+n}$, $\forall \breve{\bar{\boldsymbol{S}}}_n$ and $1 \leq n \leq N-1$, we obtain

$$= -\log\Big(p\Big(\bar{\boldsymbol{R}}_n \,\Big|\, \tilde{\bar{\boldsymbol{S}}}_n\Big)\Big) + \log\left(\sum_{\forall \breve{\bar{\boldsymbol{S}}}_n} p\Big(\bar{\boldsymbol{R}}_n \,\Big|\, \breve{\bar{\boldsymbol{S}}}_n\Big)\right). \tag{3.26}$$

Note that minimizing (3.26) for $n = 1$ over all $\tilde{\bar{\boldsymbol{S}}}$ leads to the ML–MSDD solution. While the second term in (3.26) does not influence the ML decision it plays an important role in comparing paths of different lengths in tree–search decoding (see below).

Still, using the expression in (3.26) as metric in a tree–search decoder is obviously impracticable, as the evaluation of the second term of (3.26) involves the computation of L^{N-n} probabilities $\Pr\big(\bar{\boldsymbol{R}}_n \mid \breve{\tilde{\boldsymbol{S}}}_n\big)$, i.e. the complexity of evaluating (3.26) for a single $\tilde{\tilde{\boldsymbol{S}}}_n$ is of the same order as a search over all L^{N-n} different $\tilde{\tilde{\boldsymbol{S}}}_n$. However, it is reasonable to assume that the actually transmitted signal $\bar{\boldsymbol{S}}_n$ has a relatively high probability $p\big(\bar{\boldsymbol{R}}_n \mid \bar{\boldsymbol{S}}_n\big)$ such that

$$-\log\Big(p\Big(\bar{\boldsymbol{R}}_n \,\Big|\, \tilde{\tilde{\boldsymbol{S}}}_n\Big)\Big) + \log\big(p\big(\bar{\boldsymbol{R}}_n \,\big|\, \bar{\boldsymbol{S}}_n\big)\big) \tag{3.27}$$

appears to be a good approximation of (3.26). Still, this is not helpful as $\bar{\boldsymbol{S}}_n$ is unknown. Therefore, we approximate (3.27) by taking the expectation of the second term of (3.27) with respect to channel states and noise, i.e. we use

$$-\log\Big(p\Big(\bar{\boldsymbol{R}}_n \,\Big|\, \tilde{\tilde{\boldsymbol{S}}}_n\Big)\Big) + \mathcal{E}\big\{\log\big(p\big(\bar{\boldsymbol{R}}_n \,\big|\, \bar{\boldsymbol{S}}_n\big)\big)\big\} \tag{3.28}$$

as an approximation of (3.26). Using the results of Sections 2.4.2 and 3.1.2.2 [cf. (3.24)] we obtain the *Fano–type* path metric

$$d_{\mathrm{F},n}\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) \triangleq \left[\sum_{i=n}^{N}\Big\|\breve{\boldsymbol{R}}_{i,i}^{\mathsf{H}}\tilde{\boldsymbol{S}}_i + \boldsymbol{X}_i\Big\|^2\right] - \tilde{n}N_{\mathrm{S}}N_{\mathrm{R}} \tag{3.29}$$

$$= d_{\mathrm{F},n+1}\Big(\tilde{\tilde{\boldsymbol{S}}}_{n+1}\Big) + \delta_{\mathrm{F},n}\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) \tag{3.30}$$

with $\breve{\boldsymbol{R}}_{n,n}$ and $\boldsymbol{X}_n$ as defined in (3.17) and (3.18), respectively. Based on (3.23), which tells us that all branches along the correct path on average contribute equally to the path metric of the correct path, the Fano–type branch metric reads

$$\delta_{\mathrm{F},n}\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) \triangleq \Big\|\breve{\boldsymbol{R}}_{n,n}^{\mathsf{H}}\tilde{\boldsymbol{S}}_n + \boldsymbol{X}_n\Big\|^2 - N_{\mathrm{S}}N_{\mathrm{R}}. \tag{3.31}$$

Due to the additive nature of this path metric, we can still directly employ the tree–search algorithms of Section 3.1.1 with this Fano–type metric.

Note that we obtain the intuitively quite reasonable result that the Fano–type metric of a candidate $\tilde{\tilde{\boldsymbol{S}}}_n$ is equal to the difference between the regular ML–MSDD metric and the expected ML–MSDD metric of the correct sequence $\bar{\boldsymbol{S}}_n$, i.e.

$$d_{\mathrm{F},n}\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) = d_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) - \mathcal{E}\big\{d_n\big(\bar{\boldsymbol{S}}_n\big)\big\} \tag{3.32}$$

$$= d_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) - \tilde{n}N_{\mathrm{S}}N_{\mathrm{R}} \tag{3.33}$$

and

$$\delta_{\mathrm{F},n}\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) = \delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) - \mathcal{E}\big\{\delta_n\big(\bar{\boldsymbol{S}}_n\big)\big\} \tag{3.34}$$

$$= \delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) - N_{\mathrm{S}}N_{\mathrm{R}}, \tag{3.35}$$

i.e. the path metric is *biased* such that the expected metric of the correct path is always zero,

$$\mathcal{E}\{d_{\mathrm{F},n}(\bar{\boldsymbol{S}}_n)\} = 0 \quad \text{and} \tag{3.36}$$

$$\mathcal{E}\{\delta_{\mathrm{F},n}(\bar{\boldsymbol{S}}_n)\} = 0 \tag{3.37}$$

for $1 \leq n \leq N-1$, while it tends to grow beyond zero for incorrect paths $\tilde{\bar{\boldsymbol{S}}}_n \neq \bar{\boldsymbol{S}}_n$ after diverging from the correct path. This way, a path in the decoding tree can be identified as "not promising" and possibly terminated as soon as the metric $d_{\mathrm{F},n}(\tilde{\bar{\boldsymbol{S}}}_n)$ of the current candidate $\tilde{\bar{\boldsymbol{S}}}_n$ exceeds the threshold ρ, as it can be *expected* due to (3.37), that the path metric of a path $\tilde{\bar{\boldsymbol{S}}}$ generated upon extension of $\tilde{\bar{\boldsymbol{S}}}_n$ will at best —meaning if $\tilde{\bar{\boldsymbol{S}}}_n = \bar{\boldsymbol{S}}_n$— have a path metric that is equal to that of $\tilde{\bar{\boldsymbol{S}}}_n$. The other way around, a tree–search decoder based on this metric can be interpreted to not simply compare the current ML–MSDD metric $d_n(\tilde{\bar{\boldsymbol{S}}}_n)$ to a threshold, but to compare the current ML–MSDD metric $d_n(\tilde{\bar{\boldsymbol{S}}}_n)$ augmented by the expected metric $(n-1)N_{\mathrm{S}}N_{\mathrm{R}}$ of the remainder of this path —if this were the correct path— to the threshold. This strategy is very beneficial with respect to decoder complexity in the low–SNR scenario discussed at the end of Section 3.1.2.2.

On the downside, it is possible that the path that might eventually have turned out to be optimal is terminated prematurely thereby leading to a performance that is suboptimal in the ML–sense, even if a tree–search decoder that is optimal in the sense of Section 3.1.1.2.3 is employed.

While the above branch metric bias $b = N_{\mathrm{S}}N_{\mathrm{R}}$ was motivated based on probabilistic arguments, it is possible to heuristically optimize the bias b, i.e. to use a branch metric $\delta_n(\tilde{\bar{\boldsymbol{S}}}_n) - b$ with an arbitrary (positive) b, cf. e.g. [PH05b, PH05a]. In general, a smaller bias leads to better performance but higher complexity while a larger bias decreases decoder complexity at the expense of degraded performance. The extreme cases are (i) $b = 0$, in which case ML MSDD is performed, and (ii) $b \to \infty$, where any optimal tree–search decoder degenerates to a DFE–type algorithm. In the following, we will however restrict our considerations to the above case of the probabilistically motivated branch metric bias $b = N_{\mathrm{S}}N_{\mathrm{R}}$.

It is worth mentioning that recently a comparable approach was considered for tree–search decoding in coherent detection of space–time codes in e.g. [GN04, MGDC06, ZF06, BZRF06].

3.1.3 Algorithms for Tree–Search MSDD

For efficient (fast) tree search for MSDD and novel DF–MSDD, we concentrate on the application of depth–first search (DFS) and metric–first search (MFS) algorithms, which achieve the best performance–complexity tradeoff (cf. also the comparative study [MGDC06] for coherent MIMO detection). In particular, we apply the Agrell Sphere Decoder (A–SpD) and the Fano algorithm of Sections 3.1.1.3.3 and 3.1.1.3.2, respectively.

Let us point out again that both algorithms can in principle be operated with either of the metrics considered in Section 3.1.2, even though the combination of Fano algorithm and ML metric (without bias term) is usually not very meaningful (see below).

3.1.3.1 Multiple–Symbol Differential Sphere Decoding (MSDSD)

We refer to the application of A–SpD to accomplish MSDD as multiple–symbol differential sphere decoding (MSDSD), cf. [LSPW04] for DPSK. MSDD based on the A–SpD and the Fano–type metric (cf. Section 3.1.2.3) is subsequently referred to as MSDSD–FM.

Let us for clarity very briefly summarize the application of A–SpD to MSDD, based on the notation of ML branch and path metrics $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ and $d_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$, cf. (3.19) and (3.20), respectively. For illustration see the flowchart of the A–SpD in Fig. 3.5. Starting at $n = N$, the SpD selects candidates $\tilde{\boldsymbol{V}}_{n-1}$ (or equivalently $\tilde{\boldsymbol{S}}_{n-1}$ according to (3.8)) based on tentative decisions $\tilde{\boldsymbol{V}}_j$, $n \leq j \leq N-1$ represented by $\tilde{\tilde{\boldsymbol{S}}}_n$, and continues to decrement n, i.e. $n := n-1$, as long as the metric $d_\mathrm{f} = d_{n-1}\big(\tilde{\tilde{\boldsymbol{S}}}_{n-1}\big)$ of the child $\tilde{\tilde{\boldsymbol{S}}}_{n-1}$ currently under consideration does not exceed a given threshold (radius) ρ, i.e.

$$d_\mathrm{f} = d_{n-1}\Big(\tilde{\tilde{\boldsymbol{S}}}_{n-1}\Big) \leq \rho. \tag{3.38}$$

If the decoder reaches the end of the tree, i.e. $n = 1$, $\tilde{\tilde{\boldsymbol{S}}}$ is stored as (preliminary) decoding result $\hat{\tilde{\boldsymbol{S}}}$, its metric is used to further reduce the size of the search space by updating $\rho := d_1\big(\tilde{\tilde{\boldsymbol{S}}}\big)$, and the decoder increments n by two, i.e. $n := n+2$, searching for alternative paths whose metric lies below the updated threshold ρ. If $d_{n-1}\big(\tilde{\tilde{\boldsymbol{S}}}_{n-1}\big)$ exceeds ρ for any value of n, n is incremented, i.e. $n := n+1$, and a new candidate for $\tilde{\boldsymbol{V}}_n$ is examined. If the decoder returns to $n = N$, it means that all paths $\tilde{\tilde{\boldsymbol{S}}}_n$ have been pursued up to the point, where $d_{n-1}\big(\tilde{\tilde{\boldsymbol{S}}}_{n-1}\big) \geq \rho$. Consequently, the search is terminated and the decoder returns $\hat{\tilde{\boldsymbol{S}}} = \tilde{\tilde{\boldsymbol{S}}}$ or through reversal of (2.11) $\hat{\tilde{\boldsymbol{V}}}$. For the ordering of candidates for any $\boldsymbol{V}_n$, $1 \leq n \leq N-1$, the Schnorr–Euchner (SE) enumeration strategy is employed, i.e. candidates are checked in order of increasing branch metric $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$, as this allows for an initialization with $\rho \to \infty$ and an (usually) fast convergence of the search process, cf. e.g. [LSPW04, AEVZ02, DEC03].

As pointed out in Section 3.1.1.3.3 it may still be useful to initialize the threshold ρ with a *finite* value ρ_init in order to "guide" the SpD towards the best path of full length and thereby accelerate the convergence of the tree search. Determination of an optimal initial threshold is not trivial. On the one hand, it should be chosen as small as possible in order to achieve optimal "guidance". On the other hand, if the initial threshold is chosen too small the decoder will have to be restarted with an increased initial threshold frequently, such that the overall complexity may be even higher than starting with $\rho \to \infty$ in the first place. Since for MSDSD the metric $d_1\big(\bar{\boldsymbol{S}}\big)$ of the correct path is $\chi^2(NN_\mathrm{S}N_\mathrm{R}, 2NN_\mathrm{S}N_\mathrm{R})$ distributed, i.e. has mean $NN_\mathrm{S}N_\mathrm{R}$

and variance $NN_{\mathrm{S}}N_{\mathrm{R}}$, regardless of the SNR or the temporal correlation, the initial threshold should certainly be chosen proportional to $NN_{\mathrm{S}}N_{\mathrm{R}}$, i.e. as

$$\begin{array}{ll} \text{MSDSD:} & \rho_{\mathrm{init}} = NN_{\mathrm{S}}N_{\mathrm{R}}c_\rho \\ \text{MSDSD–FM:} & \rho_{\mathrm{init}} = NN_{\mathrm{S}}N_{\mathrm{R}}(c_\rho - 1) \end{array} \qquad c_\rho > 0. \tag{3.39}$$

The result for MSDSD–FM follows directly from the above discussion based on the observation that $d_{\mathrm{F},1}\big(\bar{\boldsymbol{S}}\big) = d_1\big(\bar{\boldsymbol{S}}\big) - NN_{\mathrm{S}}N_{\mathrm{R}}$. In case no candidate is found below the finite initial threshold ρ_{init}, the search is repeatedly restarted with an increased initial threshold

$$\rho_{\mathrm{init}} := \rho_{\mathrm{init}} + NN_{\mathrm{S}}N_{\mathrm{R}} \tag{3.40}$$

until at least one candidate with $d_1\big(\tilde{\bar{\boldsymbol{S}}}\big) \leq \rho_{\mathrm{init}}$ is found.

For small values of N the optimal[10] choice of c_ρ can only be determined based on simulations. For large N on the other hand, it appears reasonable to choose $c_\rho = 1 + \varepsilon_\rho$, with an arbitrarily small positive constant ε_ρ, since by the strong law of large numbers (cf. e.g. [CT91]) the probability of $d_1\big(\bar{\boldsymbol{S}}\big)/N$ deviating by more than an arbitrarily small positive constant ε_ρ from its mean $N_{\mathrm{S}}N_{\mathrm{R}}$ decreases exponentially in N, such that we obtain in the limit of $N \to \infty$

$$\lim_{N\to\infty} \Pr\big(d_1\big(\bar{\boldsymbol{S}}\big) \leq NN_{\mathrm{S}}N_{\mathrm{R}}(1-\varepsilon_\rho) \,\big|\, \bar{\boldsymbol{S}}\big) = 0 \quad \text{and} \tag{3.41}$$

$$\lim_{N\to\infty} \Pr\big(d_1\big(\bar{\boldsymbol{S}}\big) \leq NN_{\mathrm{S}}N_{\mathrm{R}}(1+\varepsilon_\rho) \,\big|\, \bar{\boldsymbol{S}}\big) = 1. \tag{3.42}$$

For a detailed analysis of the dependence of $\Pr\big(d_1\big(\bar{\boldsymbol{S}}\big) \leq \rho \mid \bar{\boldsymbol{S}}\big)$ on the system parameters see Section 5.5.1.2.2.

3.1.3.2 Fano Multiple–Symbol Differential Detection (Fano–MSDD)

We refer to the application of the Fano algorithm to accomplish MSDD as Fano–MSDD.

Having described the Fano algorithm in detail in Section 3.1.1.3.2 and established the connection between tree–search decoding and MSDD using the example of A–SpD in MSDSD the application of the Fano algorithm to MSDD should be clear at this point. We therefore restrict ourselves to identifying the variables of the MSDD metric with those of the description of the general Fano algorithm in Section 3.1.1.3.2: If the current path of the decoder corresponds to $\tilde{\bar{\boldsymbol{S}}}_n$, we have

$$d_{\mathrm{b}} = d_{\mathrm{F},n+1}\Big(\tilde{\bar{\boldsymbol{S}}}_{n+1}\Big) \tag{3.43}$$

$$d_{\mathrm{c}} = d_{\mathrm{F},n}\Big(\tilde{\bar{\boldsymbol{S}}}_{n}\Big) \tag{3.44}$$

$$d_{\mathrm{f}} = d_{\mathrm{F},n-1}\Big(\tilde{\bar{\boldsymbol{S}}}_{n-1}\Big)\,. \tag{3.45}$$

[10] Optimal with respect to minimal decoder complexity.

For numerical results in later sections, we use a stepsize of $\Delta_{\mathrm{F}} = N_{\mathrm{T}}N_{\mathrm{R}}$. For this intermediate value there is neither a noticeable effect of an increased complexity nor a performance degradation that occur if Δ_{F} is chosen too small or too large, respectively.

The combination of Fano–MSDD with the ML–MSDD metric is not meaningful, as this results in a computationally quite inefficient decoder. The reason for this originates from the fact that $d_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ is a strictly non–decreasing function of the path length $N-n$ —even for $\tilde{\tilde{\boldsymbol{S}}}_n = \bar{\boldsymbol{S}}_n$. Due to this property of the metric the Fano algorithm would be likely to move back after every look–forward operation and examine all paths up to the point where their metrics exceed the current threshold.[11] Hence, we only consider Fano–MSDD based on the Fano–type metric.

3.1.3.3 Decision–Feedback Multiple–Symbol Differential Sphere Decoding (DF–MSDSD)

In this section, we apply tree–search algorithms to efficiently implement decision–feedback multiple–symbol differential detection (DF–MSDD, cf. Section 2.4.4). We refer to the application of A–SpD to accomplish DF–MSDD as decision–feedback multiple–symbol differential sphere decoding (DF–MSDSD).

Recall that DF–MSDD is derived from MSDD by feeding back $N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-2$ previously obtained decisions $\hat{\boldsymbol{V}}[k-\kappa]$, $\kappa_{\mathrm{U}}^{\mathrm{DF}} \leq \kappa \leq N-2$, into the MSDD metric, optimizing the metric only with respect to symbols $\hat{\boldsymbol{V}}[k-\kappa]$, $0 \leq \kappa \leq \kappa_{\mathrm{U}}^{\mathrm{DF}}$, and returning estimates $\hat{\boldsymbol{V}}[k-\kappa]$, $\kappa_{\mathrm{L}}^{\mathrm{DF}} \leq \kappa \leq \kappa_{\mathrm{U}}^{\mathrm{DF}}$, as decoder output whereas estimates $\hat{\boldsymbol{V}}[k-\kappa]$, $0 \leq \kappa \leq \kappa_{\mathrm{L}}^{\mathrm{DF}}-1$, are discarded. The observation window slides forward in steps of $\kappa_{\mathrm{U}}^{\mathrm{DF}}-\kappa_{\mathrm{L}}^{\mathrm{DF}}+1$, i.e. $k := k+\kappa_{\mathrm{U}}^{\mathrm{DF}}-\kappa_{\mathrm{L}}^{\mathrm{DF}}+1$.

A direct approach to DF–MSDSD would be to insert the feedback symbols in the ML–MSDD metric representation of (3.16), such that the tree–search is performed only over the $\kappa_{\mathrm{U}}^{\mathrm{DF}}+1$ lowest levels of the tree, whereas the remaining $N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-2$ levels are fixed. However, in terms of computational complexity this is not a very efficient approach, as the branch metrics $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$, $n \leq N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-2$, have to be recomputed for every investigated candidate $\big[\tilde{\boldsymbol{V}}[k-\kappa_{\mathrm{U}}^{\mathrm{DF}}],\ldots,\tilde{\boldsymbol{V}}[k]\big]$. It therefore appears reasonable to reverse the orientation of the decoding tree, such that the $N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-2$ feedback symbols fix a node at depth $N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-2$ of the tree and the decoder has to perform its tree search only over a small subtree of depth $\kappa_{\mathrm{U}}^{\mathrm{DF}}+1$ emanating from this fixed node.

To derive the appropriate metric representation, we recall the relationship between the elements of $\boldsymbol{C}$ [cf. (3.12)] and coefficients of linear backward MMSE predictors established in Section 3.1.2.2. There, we showed that $c_{n,j}$ is the $(j-n)$th coefficient of the $(N-n)$th–order

[11] Note that since $d_{\mathrm{b}} \leq d_{\mathrm{c}} \leq \rho$ always for the ML–MSDD metric, the decoder would always move back all the way to the root.

linear backward MMSE predictor normalized by the standard deviation of the corresponding prediction error [cf. (3.22)]. It is then easy to see that $c_{n,j}^*$ is the $(N+1-j)$th coefficient $p_{\mathrm{F},N+1-j}^{(N-n)}$ of the $(N-n)$th–order linear *forward* MMSE predictor normalized by the standard deviation $\sigma_{\mathrm{p}}^{(N-n)}$ of the corresponding prediction error (cf. also [Hay96]), i.e.

$$c_{n,j} = \left(\frac{p_{\mathrm{F},N+1-j}^{(N-n)}}{\sigma_{\mathrm{p}}^{(N-n)}} \right)^* . \tag{3.46}$$

This means that we can alternatively obtain the *regular* ML–MSDD estimate by solving [cf. (3.13) with (3.22) and (3.46)]

$$\hat{\tilde{\boldsymbol{V}}} = \underset{\tilde{\boldsymbol{V}} \in \boldsymbol{\mathcal{V}}^{N-1}}{\operatorname{argmin}} \left\{ \sum_{n=1}^{N} \left\| \sum_{j=1}^{n} \frac{p_{\mathrm{F},j}^{(n-1)}}{\sigma_{\mathrm{p}}^{(n-1)}} \tilde{\boldsymbol{S}}_{\mathrm{forw},j}^{\mathsf{H}} \boldsymbol{R}_j \right\|^2 \right\} \tag{3.47}$$

$$= \underset{\tilde{\boldsymbol{V}} \in \boldsymbol{\mathcal{V}}^{N-1}}{\operatorname{argmin}} \left\{ \sum_{n=1}^{N} \left\| \sum_{j=1}^{n} c_{N+1-n,N+1-j} \boldsymbol{R}_j^{\mathsf{H}} \tilde{\boldsymbol{S}}_{\mathrm{forw},j} \right\|^2 \right\} \tag{3.48}$$

with

$$\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{forw}} \triangleq \left[\tilde{\boldsymbol{S}}_{\mathrm{forw},1}^{\mathsf{H}}, \ldots, \tilde{\boldsymbol{S}}_{\mathrm{forw},N}^{\mathsf{H}} \right]^{\mathsf{H}} \tag{3.49}$$

and

$$\tilde{\boldsymbol{S}}_{\mathrm{forw},n} \triangleq \prod_{\kappa=N-n}^{N-2} \tilde{\boldsymbol{V}}[k-\kappa]. \tag{3.50}$$

In this context, *regular* ML MSDSD could be performed in an $(N-1)$–dimensional tree, where $\tilde{\boldsymbol{S}}_{\mathrm{forw},1} = \boldsymbol{I}_{N_{\mathrm{S}}}$ and the $(NN_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional $\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{forw}}$ correspond to the root and to the leaves of the tree, respectively.

In *DF*–MSDSD on the other hand, the feedback of $N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-2$ previous decisions $\hat{\boldsymbol{V}}[k-\kappa]$, $\kappa_{\mathrm{U}}^{\mathrm{DF}}+1 \leq \kappa \leq N-2$, corresponds to fixing a node at depth $N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-2$ by fixing $\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{forw},n} \triangleq \left[\tilde{\boldsymbol{S}}_{\mathrm{forw},1}^{\mathsf{H}}, \ldots, \tilde{\boldsymbol{S}}_{\mathrm{forw},n}^{\mathsf{H}} \right]^{\mathsf{H}}$ as

$$\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{forw},n} = \tilde{\tilde{\boldsymbol{S}}}_{\mathrm{DF},n} \triangleq \left[\tilde{\boldsymbol{S}}_{\mathrm{DF},1}^{\mathsf{H}}, \ldots, \tilde{\boldsymbol{S}}_{\mathrm{DF},n}^{\mathsf{H}} \right]^{\mathsf{H}} \tag{3.51}$$

with

$$\tilde{\boldsymbol{S}}_{\mathrm{DF},n} \triangleq \prod_{\kappa=N-n}^{\kappa_{\mathrm{U}}^{\mathrm{DF}}} \tilde{\boldsymbol{V}}[k-\kappa] \cdot \prod_{\kappa=\max(N-n,\, \kappa_{\mathrm{U}}^{\mathrm{DF}}+1)}^{N-2} \hat{\boldsymbol{V}}[k-\kappa]. \tag{3.52}$$

This node can be viewed as root of a $(\kappa_{\mathrm{U}}^{\mathrm{DF}}+1)$–dimensional tree, and the remaining $\kappa_{\mathrm{U}}^{\mathrm{DF}}+1$ symbols $\hat{\boldsymbol{V}}[k-\kappa]$, $0 \leq \kappa \leq \kappa_{\mathrm{U}}^{\mathrm{DF}}$, are found by means of tree–search decoding in this $(\kappa_{\mathrm{U}}^{\mathrm{DF}}+1)$–dimensional tree.

To provide more detail on the implementation, we plug (3.51) into the ML–MSDD decision rule (3.47) and obtain the DF–MSDD decision rule as

$$\Big[\hat{\boldsymbol{V}}[k-\kappa_{\mathrm{U}}^{\mathrm{DF}}],\ldots,\hat{\boldsymbol{V}}[k]\Big] = \underset{\substack{\tilde{\boldsymbol{V}}[k-\kappa]\in\boldsymbol{\mathcal{V}}\\ \forall\kappa\in\left\{0,\ldots,\kappa_{\mathrm{U}}^{\mathrm{DF}}\right\}}}{\operatorname{argmin}} \left\{d_{\mathrm{DF},N}\Big(\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{DF},N}\Big)\right\} \tag{3.53}$$

with path metrics

$$d_{\mathrm{DF},n}\Big(\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{DF},n}\Big) \triangleq d_{\mathrm{DF},n-1}\Big(\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{DF},n-1}\Big) + \delta_{\mathrm{DF},n}\Big(\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{DF},n}\Big), \quad N-\kappa_{\mathrm{U}}^{\mathrm{DF}} \le n \le N, \tag{3.54}$$

and branch metrics

$$\begin{aligned}
\delta_{\mathrm{DF},n}\Big(\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{DF},n}\Big) &\triangleq \left\|\sum_{j=1}^{n} c_{N+1-n,N+1-j}\boldsymbol{R}_j^{\mathsf{H}}\tilde{\boldsymbol{S}}_{\mathrm{DF},j}\right\|^2 && (3.55)\\
&= \left\|c_{N+1-n,N+1-n}\boldsymbol{R}_n^{\mathsf{H}}\tilde{\boldsymbol{S}}_{\mathrm{DF},n} + \sum_{j=1}^{n-1} c_{N+1-n,N+1-j}\boldsymbol{R}_j^{\mathsf{H}}\tilde{\boldsymbol{S}}_{\mathrm{DF},j}\right\|^2 && (3.56)\\
&= \left\|\breve{\boldsymbol{R}}_{\mathrm{DF},n,n}^{\mathsf{H}}\tilde{\boldsymbol{S}}_{\mathrm{DF},n} + \boldsymbol{X}_{\mathrm{DF},n}\right\|^2, \quad N-\kappa_{\mathrm{U}}^{\mathrm{DF}} \le n \le N, && (3.57)
\end{aligned}$$

where we defined

$$\begin{aligned}
\breve{\boldsymbol{R}}_{\mathrm{DF},n,j} &\triangleq c_{N+1-n,N+1-j}^{*}\boldsymbol{R}_j && (3.58)\\
\boldsymbol{X}_{\mathrm{DF},n} &\triangleq \sum_{j=1}^{n-1}\breve{\boldsymbol{R}}_{\mathrm{DF},n,j}^{\mathsf{H}}\tilde{\boldsymbol{S}}_{\mathrm{DF},j} && (3.59)\\
&= \sum_{j=N-\kappa_{\mathrm{U}}^{\mathrm{DF}}}^{n-1}\breve{\boldsymbol{R}}_{\mathrm{DF},n,j}^{\mathsf{H}}\tilde{\boldsymbol{S}}_{\mathrm{DF},j} + \boldsymbol{Y}_{\mathrm{DF},n} && (3.60)\\
\boldsymbol{Y}_{\mathrm{DF},n} &\triangleq \sum_{j=1}^{N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-1}\breve{\boldsymbol{R}}_{\mathrm{DF},n,j}^{\mathsf{H}}\tilde{\boldsymbol{S}}_{\mathrm{DF},j}. && (3.61)
\end{aligned}$$

At this point a few observations can be made: (i) This metric structure is of the same form as that of ML–MSDD [cf. (3.20)], where $\boldsymbol{X}_{\mathrm{DF},n}$ (like $\boldsymbol{X}_n$) reflects the impact of the path to a node on the metric of a branch emanating from this node. (ii) The metric $d_{\mathrm{DF},N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-1}\big(\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{DF},N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-1}\big)$ of the root of the $(\kappa_{\mathrm{U}}^{\mathrm{DF}}+1)$–dimensional tree need not be computed and can be fixed without loss of optimality as $d_{\mathrm{DF},N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-1}\big(\tilde{\tilde{\boldsymbol{S}}}_{\mathrm{DF},N-\kappa_{\mathrm{U}}^{\mathrm{DF}}-1}\big) \equiv 0$. (iii) The matrices $\boldsymbol{Y}_{\mathrm{DF},n}$, $N-\kappa_{\mathrm{U}}^{\mathrm{DF}} \le n \le N$, [cf. (3.61)] are independent of the particular candidate under consideration, which means that they have to be computed only once at the beginning of the decoding process.

Compared to MSDSD the *minimal* complexity of DF–MSDSD is slightly increased if $\kappa_{\mathrm{L}}^{\mathrm{DF}} > 0$, as the depth of the search tree is greater than the number of decisions returned per decoder use. This is however more than compensated for by the fact that the depth of the search tree

in DF–MSDSD can usually be chosen much smaller than $N-1$, which leads to tremendous complexity savings under adverse channel conditions.

It should be clear, that all methods that can be applied to solve the ML–MSDD problem more efficiently, such as finite initial search radius or Fano–type metric, can be applied equally well to DF–MSDD. In particular, following the above conventions on denomination, we subsequently refer to DF–MSDSD based on the Fano–type metric as DF–MSDSD–FM and to DF–MSDD based on the Fano algorithm (with Fano–type metric) as DF–Fano–MSDD.

3.1.4 Optimization for Various Signal Constellations

The application of tree–search algorithms to (DF–)MSDD as described in Section 3.1.3 helps to significantly reduce the complexity of (DF–)MSDD by efficiently solving the $(N-1)$– / $(\kappa_{\mathrm{U}}^{\mathrm{DF}}+1)$–dimensional optimization problem. Note that these MSDD algorithms are universally applicable to unitary DSTM constellations. However, the question of how the so–called Schnorr–Euchner (SE) candidate enumeration strategy, where branches emanating from a particular node are processed in order of increasing branch metrics, is to be implemented in an efficient way remains to be answered. In this section, we will address this issue first introducing a brute–force approach, that can be applied with arbitrary DSTM constellations, and subsequently presenting computationally more efficient algorithms tailored to the four signal constellations discussed in Section 2.1. At this, we will for clarity restrict our attention to MSDD, but note that all methods described in the following are directly applicable to DF–MSDD, as well. With the exception of orthogonal DSTM constellations, this leads to a nested structure of search algorithms, where the *outer* decoder is implemented using one of the algorithms introduced in Section 3.1.3, i.e. it solves the $(N-1)$–dimensional tree–search problem, whereas the *inner* decoder is responsible for SE candidate enumeration.

A simple approach that we subsequently refer to as *full–search* (FS) symbol enumeration is the following: Whenever *the best* child $\tilde{\tilde{\boldsymbol{S}}}_{n-1}$ of a node corresponding to $\tilde{\tilde{\boldsymbol{S}}}_{n}$ is required the decoder computes the branch metrics of *all* L children $\tilde{\tilde{\boldsymbol{S}}}_{n-1}$ of $\tilde{\tilde{\boldsymbol{S}}}_{n}$, stores them in a list sorted in order of non–decreasing branch metric and continues its tree search with the best branch from this list. If the decoder —at some later stage of the tree–search process— has to examine further branches emanating from the same parent node $\tilde{\tilde{\boldsymbol{S}}}_{n}$, it merely has to take them from the list generated earlier. Note, that in the case of MSDSD it is *not* possible that a particular node is visited *again* by the decoder after another node *at the same depth* has been examined since the first visit to that node. Consequently, the decoder only has to maintain (at most) $(N-2)$ lists of length L of branches $\tilde{\boldsymbol{S}}_{i}$, $2 \leq n \leq i \leq N-1$, forming the currently investigated path in the tree and their respective $(L-1)$ sisters, which constitutes an affordable memory requirement.[12]

[12]Note that for $n=1$, i.e. at the leaves of the tree, the decoder does not need to store the list, as further

In the case of Fano–MSDD it is possible that individual nodes are visited more than once. In order to avoid uncontrollable memory requirements by storing all lists corresponding to visited nodes, we stick to maintaining only the lists corresponding to nodes along the current path and recompute them if a node should be visited again. We identify tree–search decoders that use this full–search approach by adding a suffix "–FS" to their abbreviation, i.e. MSDSD–FS(–FM) and Fano–MSDD–FS. In this case, MSDD benefits from the efficiency of the outer tree–search algorithm, but is still rather inefficient when L is large and particularly in high SNR, because here it would in most cases be sufficient for the tree–search decoder to examine only two candidates $\tilde{\boldsymbol{S}}_n$, $2 \leq n \leq N-1$, and one candidate $\tilde{\boldsymbol{S}}_1$ to identify the ML–MSDD solution.

In [LSPW04] an efficient implementation of the SE enumeration strategy was proposed for DPSK, cf. also Section 3.1.4.1.1. There, it was possible to determine the appropriate order of candidates, i.e. in order of non–decreasing branch metric, for individual symbols directly by mere examination of the argument of a single complex number, i.e. *without explicit sorting* of candidates.

In order to see what can be done to reduce the complexity of symbol search in the general DSTM case, we take a closer look at the branch metric $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$. Using $||\boldsymbol{X}||^2 = \mathrm{tr}\big\{\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}}\big\}$ and $\mathrm{tr}\{\boldsymbol{A}\boldsymbol{B}\} = \mathrm{tr}\{\boldsymbol{B}\boldsymbol{A}\}$ we can write [cf. (3.19)]

$$\delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) = \Big|\Big|\breve{\boldsymbol{R}}_{n,n}^{\mathsf{H}}\tilde{\boldsymbol{S}}_n + \boldsymbol{X}_n\Big|\Big|^2 \tag{3.62}$$

$$= \Big|\Big|\breve{\boldsymbol{R}}_{n,n}\Big|\Big|^2 + ||\boldsymbol{X}_n||^2 + 2\mathrm{Re}\left\{\mathrm{tr}\Big\{\tilde{\boldsymbol{S}}_n\boldsymbol{X}_n^{\mathsf{H}}\breve{\boldsymbol{R}}_{n,n}^{\mathsf{H}}\Big\}\right\}. \tag{3.63}$$

Since $\big|\big|\breve{\boldsymbol{R}}_{n,n}\big|\big|^2$ and $\big|\big|\boldsymbol{X}_n\big|\big|^2$ do not depend on $\tilde{\boldsymbol{S}}_n$ we can see that the problem of finding the $\tilde{\boldsymbol{S}}_n$ that minimizes $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ is of the same form as CDD, cf. (2.17). This observation allows us to implement MSDD for DSTM using a nested structure of an outer and $N-1$ identical inner CDD–type decoders. The outer decoder, which solves (3.16), initializes an inner decoder at stage n, $1 \leq n \leq N-1$, with matrices $\breve{\boldsymbol{R}}_{n,n}$ and $\boldsymbol{X}_n$ to find a new candidate $\tilde{\boldsymbol{S}}_n$ that minimizes $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ given $\breve{\boldsymbol{R}}_{n,n}$ and $\boldsymbol{X}_n$. It further provides the inner decoder with (i) a list of candidates $\tilde{\boldsymbol{S}}_n$ that have been examined previously given the same parent path $\tilde{\tilde{\boldsymbol{S}}}_{n+1}$ and thus are to be excluded from the search, and (ii) a threshold $\rho - d_{n+1}\big(\tilde{\tilde{\boldsymbol{S}}}_{n+1}\big)$ to limit the symbol search to candidates with

$$\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big) \leq \rho - d_{n+1}\Big(\tilde{\tilde{\boldsymbol{S}}}_{n+1}\Big), \tag{3.64}$$

i.e. that satisfy $d_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big) \leq \rho$, cf. (3.38).

Due to (3.33) and (3.35) it is immediately clear that this approach is equally applicable to the Fano–type metric $\delta_{\mathrm{F},n}\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$, i.e. can be used as inner decoder in a nested structure with

branches besides the best one are of no interest here.

either MSDSD(–FM) or Fano–MSDD as outer decoder. In the following, we will —for the sake of notational simplicity— only use the notation of the regular ML metric, keeping in mind that $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ and $d_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ can be replaced with $\delta_{\mathrm{F},n}\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ and $d_{\mathrm{F},n}\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ throughout this section.

While the approach of applying appropriately optimized variants of CDD algorithms as inner decoders for symbol search is applicable to any DSTM constellation, we will in the following only derive inner decoders for the most common DSTM codes, which were introduced in Section 2.1.

3.1.4.1 Cyclic Codes

In Section 2.1.1.1 we pointed out that cyclic DSTM codes are a generalization of DPSK. Thus one might wonder whether there is a similarly simple rule for enumerating candidates $\tilde{\boldsymbol{S}}_n$ for DSTM as was proposed for DPSK in [LSPW04]. Unfortunately, however, this is in general, i.e. for arbitrary cyclic DSTM constellations, not the case. Only for the special case where all diagonal elements of the matrix symbols are equal is this feasible. This special case is therefore considered separately in Section 3.1.4.1.1 before two efficient candidate symbol enumeration strategies for arbitrary cyclic codes based on adaptations of two different CDD algorithms for cyclic codes are discussed in Sections 3.1.4.1.2 and 3.1.4.1.3.

3.1.4.1.1 Special Case $[c_1,\ldots,c_{N_\mathrm{S}}]=[1,\ldots,1]$ The subclass of cyclic DSTM constellations, where $[c_1,\ldots,c_{N_\mathrm{S}}]=[1,\ldots,1]$, i.e. where

$$\tilde{\boldsymbol{S}}_n = \mathrm{e}^{\mathrm{j}\frac{2\pi}{L}\tilde{l}_n}\boldsymbol{I}_{N_\mathrm{S}}, \tag{3.65}$$

is optimal for $N_\mathrm{S}=2$ and $L=4$ (cf. [Hug00a, HS00]). Substitution of (3.65) into (3.63) leads to

$$\begin{aligned}\delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) &= \Big|\Big|\breve{\boldsymbol{R}}_{n,n}\Big|\Big|^2 + ||\boldsymbol{X}_n||^2 + 2\mathrm{Re}\Big\{\mathrm{e}^{\mathrm{j}\frac{2\pi}{L}\tilde{l}_n}\mathrm{tr}\Big\{\breve{\boldsymbol{R}}_{n,n}^{\mathsf{H}}\boldsymbol{X}_n^{\mathsf{H}}\Big\}\Big\} &(3.66)\\ &= \gamma_n + |\alpha_n|\cos\left(\frac{2\pi}{L}\tilde{l}_n - \mathrm{angle}(\alpha_n)\right), &(3.67)\end{aligned}$$

($\mathrm{angle}(x)$: from $x=|x|\exp(\mathrm{j}\cdot\mathrm{angle}(x))$ for $x\in\mathbb{C}\backslash\{0\}$, $\mathbb{C}$: complex numbers) where

$$\begin{aligned}\alpha_n &\triangleq 2\mathrm{tr}\Big\{\breve{\boldsymbol{R}}_{n,n}\boldsymbol{X}_n\Big\} &(3.68)\\ \gamma_n &\triangleq \Big|\Big|\breve{\boldsymbol{R}}_{n,n}\Big|\Big|^2 + \Big|\Big|\boldsymbol{X}_n\Big|\Big|^2. &(3.69)\end{aligned}$$

We observe that the MSDD decoding problem with respect to unitary–matrix symbols is reduced to an $(N-1)$–dimensional one with respect to L–PSK symbols, which means that the SE strategy can be implemented as in the single–antenna case. I.e. given an $\breve{\boldsymbol{R}}_{n,n}$ and $\boldsymbol{X}_n$ the decoder computes the "unconstrained solution"

$$\begin{aligned}l_n^{\mathrm{u}} &= \frac{L}{2\pi}\left(\mathrm{angle}(\alpha_n)+\pi\right) &(3.70)\\ &= \frac{L}{2\pi}\mathrm{angle}(-\alpha_n) &(3.71)\end{aligned}$$

for $\tilde{l}_n$ such that $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ attains its minimum $\gamma_n - |\alpha_n|$ [cf. (3.67)]. It then examines candidates $\tilde{\boldsymbol{S}}_n = \exp\big(\mathrm{j}\frac{2\pi}{L}\tilde{l}_n\big)\boldsymbol{I}_{N_\mathrm{S}}$ by choosing $\tilde{l}_n$ as (cf. [LSPW04])

$$\tilde{l}_n = \lfloor l_n^{\mathrm{u}} \rceil,\ \lfloor l_n^{\mathrm{u}} \rceil + 1,\ \lfloor l_n^{\mathrm{u}} \rceil - 1,\ \lfloor l_n^{\mathrm{u}} \rceil + 2,\ \ldots \tag{3.72}$$

if $l_n^{\mathrm{u}} > \lfloor l_n^{\mathrm{u}} \rceil$ and

$$\tilde{l}_n = \lfloor l_n^{\mathrm{u}} \rceil,\ \lfloor l_n^{\mathrm{u}} \rceil - 1,\ \lfloor l_n^{\mathrm{u}} \rceil + 1,\ \lfloor l_n^{\mathrm{u}} \rceil - 2,\ \ldots \tag{3.73}$$

if $l_n^{\mathrm{u}} \leq \lfloor l_n^{\mathrm{u}} \rceil$ ($\lfloor x \rceil$: integer closest to $x \in \mathbb{R}$). This way candidates $\tilde{\boldsymbol{S}}_n$ are *implicitly* sorted in order of non–decreasing $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$, i.e. the decoder does not have to compute an often unnecessarily large number of branch metrics at the time of the first visit to the parent node and sort them to find the appropriate order, which is done in the FS approach.

3.1.4.1.2 Lattice–Decoder Symbol Search For arbitrary coefficients $[c_1, \ldots, c_{N_\mathrm{S}}]$, an equally simple sorting is not feasible. To see this and to derive efficient symbol enumeration strategy, we consider the branch metric $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ of a candidate $\tilde{\tilde{\boldsymbol{S}}}_n = \big[\tilde{\boldsymbol{S}}_n^{\mathsf{H}}, \tilde{\tilde{\boldsymbol{S}}}_{n+1}^{\mathsf{H}}\big]^{\mathsf{H}}$, where $\tilde{\boldsymbol{S}}_n$ equals the $\tilde{l}_n$th element of the group signal constellation $\boldsymbol{\mathcal{V}}$ [cf. (2.2)], i.e.

$$\tilde{\boldsymbol{S}}_n = \mathrm{diag}\Big\{\mathrm{e}^{\mathrm{j}\frac{2\pi}{L}c_1}, \ldots, \mathrm{e}^{\mathrm{j}\frac{2\pi}{L}c_{N_\mathrm{S}}}\Big\}^{\tilde{l}_n}. \tag{3.74}$$

Starting again from (3.63) we can write

$$\delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) = \Big|\Big|\breve{\boldsymbol{R}}_{n,n}\Big|\Big|^2 + ||\boldsymbol{X}_n||^2 + 2\sum_{i=1}^{N_\mathrm{S}}\sum_{j=1}^{N_\mathrm{R}} |\breve{r}_{n,n,i,j}| \cdot |x_{n,j,i}| \cos\left(\frac{2\pi}{L}c_i\tilde{l}_n - \mathrm{angle}(\breve{r}_{n,n,i,j}x_{n,j,i})\right) \tag{3.75}$$

with $\breve{r}_{n,n,i,j}$ and $x_{n,i,j}$ denoting the elements in the ith row and jth column of $\breve{\boldsymbol{R}}_{n,n}$ and $\boldsymbol{X}_n$, respectively. Based on the general identity

$$\sum_{i=1}^{N} |a_i| \cos(x + \mathrm{angle}(a_i)) = \left|\sum_{i=1}^{N} a_i\right| \cos\left(x + \mathrm{angle}\left(\sum_{i=1}^{N} a_i\right)\right), \quad x \in \mathbb{C},\ a_i \in \mathbb{C},\ 1 \leq i \leq N, \tag{3.76}$$

we can combine the cosines in (3.75) in groups of N_R, which leads to

$$\delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) = \Big|\Big|\breve{\boldsymbol{R}}_{n,n}\Big|\Big|^2 + ||\boldsymbol{X}_n||^2 + 2\sum_{i=1}^{N_\mathrm{S}} \left|\sum_{j=1}^{N_\mathrm{R}} \breve{r}_{n,n,i,j}x_{n,j,i}\right| \cos\left(\frac{2\pi}{L}c_i\tilde{l}_n - \mathrm{angle}\left(\sum_{j=1}^{N_\mathrm{R}} \breve{r}_{n,n,i,j}x_{n,j,i}\right)\right). \tag{3.77}$$

At this point, it is in the general case of distinct coefficients c_i , $1 \leq i \leq N_\mathrm{S}$, not possible to combine the remaining cosine terms into a single cosine based on (3.76). This would however

be a prerequisite for an implicit candidate sorting rule as discussed in Section 3.1.4.1.1 for the special case $c_1 = \ldots = c_{N_\mathrm{S}} = 1$.

In order to reduce the complexity of the symbol search in this general case we make use of the observation discussed at the beginning of Section 3.1.4, that the problem of finding the optimal $\tilde{\boldsymbol{S}}_n$ given a certain $\breve{\boldsymbol{R}}_{n,n}$ and $\boldsymbol{X}_n$ is of the same form as CDD. In this context, two approaches appear particularly well apt. One of them is based on the bound–intersect–detect (BID) algorithm (cf. [CT05a, CT06]) and discussed in Section 3.1.4.1.3. The other considered in this section is based on an adaptation of the lattice–based CDD algorithm proposed in [CSZ01] (cf. also [LMLK05]). We will refer to this approach as inner lattice decoding (LD) and indicate its deployment through a suffix "–LD", i.e. MSDSD–LD(–FM) and Fano–MSDD–LD.

In order to derive the appropriate representation we continue from (3.77) introducing for notational brevity

$$A_{n,i} \triangleq \sqrt{\left|\sum_{j=1}^{N_\mathrm{R}} \breve{r}_{n,n,i,j} x_{n,j,i}\right|}, \quad 1 \le i \le N_\mathrm{S} \tag{3.78}$$

$$\phi_{n,i} \triangleq \frac{L}{2\pi}\mathrm{angle}\left(\sum_{j=1}^{N_\mathrm{R}} \breve{r}_{n,n,i,j} x_{n,j,i}\right), \quad 1 \le i \le N_\mathrm{S} \tag{3.79}$$

and write

$$\delta_n\left(\tilde{\tilde{\boldsymbol{S}}}_n\right) = \left|\left|\breve{\boldsymbol{R}}_{n,n}\right|\right|^2 + ||\boldsymbol{X}_n||^2 + 2\sum_{i=1}^{N_\mathrm{S}} A_{n,i}^2 \cos\left(\frac{2\pi}{L}\left(c_i\tilde{l}_n - \phi_{n,i}\right)\right) \tag{3.80}$$

$$= \left|\left|\breve{\boldsymbol{R}}_{n,n}\right|\right|^2 + ||\boldsymbol{X}_n||^2 - 2\sum_{i=1}^{N_\mathrm{S}} A_{n,i}^2 \cos\left(\frac{2\pi}{L}\left(c_i\tilde{l}_n - \phi_{n,i} - \frac{L}{2}\right)\right) \tag{3.81}$$

$$= \left|\left|\breve{\boldsymbol{R}}_{n,n}\right|\right|^2 + ||\boldsymbol{X}_n||^2 - 2\sum_{i=1}^{N_\mathrm{S}} A_{n,i}^2 \cos\left(\frac{2\pi}{L}\mathrm{mod}^*\left(c_i\tilde{l}_n - \phi_{n,i} - \frac{L}{2}, L\right)\right). \tag{3.82}$$

where we used $\cos(x) = -\cos(x-\pi)$ in (3.81), and exploited the periodicity of the cosine introducing a modified modulo function $\mathrm{mod}^*(x,L) = \mathrm{mod}(x + L/2, L) - L/2$ that returns values in the interval $[-L/2, L/2)$ ($\mathrm{mod}(x,L) \triangleq x - \lfloor x/L \rfloor L$: regular modulo function) in (3.82). Since a candidate $\tilde{\boldsymbol{S}}_n$ is more likely to be part of the desired candidate $\hat{\tilde{\boldsymbol{S}}}$ if its branch metric $\delta_n\left(\tilde{\tilde{\boldsymbol{S}}}_n\right)$ is small, i.e. if the argument of the cosine in (3.82) is small, the deployment of the cosine approximation for small arguments $\cos(x) \approx 1 - x^2/2$ in (3.82) appears reasonable. With this, we obtain

$$\approx \left|\left|\breve{\boldsymbol{R}}_{n,n}\right|\right|^2 + ||\boldsymbol{X}_n||^2 - 2\sum_{i=1}^{N_\mathrm{S}} A_{n,i}^2 + \frac{4\pi^2}{L^2}\sum_{i=1}^{N_\mathrm{S}}\left[A_{n,i}\mathrm{mod}^*\left(c_i\tilde{l}_n - \phi_{n,i} - \frac{L}{2}, L\right)\right]^2. \tag{3.83}$$

As $y = \mathrm{mod}^*(x, L)$ corresponds to adding integer–multiples of L to x such that $|y|$ is minimized, the problem of finding the minimizer for $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ can be viewed as an N_S–dimensional lattice–decoding problem (cf. [CSZ01, LMLK05] for the application to CDD)

$$\hat{\boldsymbol{x}} = \underset{\tilde{\boldsymbol{x}} \in \boldsymbol{Z}^{N_\mathrm{S}}}{\mathrm{argmin}} \big\{ ||\boldsymbol{L}\tilde{\boldsymbol{x}} - \boldsymbol{t}||^2 \big\}, \tag{3.84}$$

with

$$\boldsymbol{L} \triangleq \begin{bmatrix} A_{n,1}c_1 & 0 & \dots & \dots & 0 \\ A_{n,2}c_2 & A_{n,2}L & \ddots & & \vdots \\ A_{n,3}c_3 & 0 & A_{n,3}L & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ A_{n,N_\mathrm{S}}c_{N_\mathrm{S}} & 0 & \dots & 0 & A_{n,N_\mathrm{S}}L \end{bmatrix} \tag{3.85}$$

$$\boldsymbol{t} \triangleq \begin{bmatrix} A_{n,1}(\phi_{n,1} + L/2) \\ A_{n,2}(\phi_{n,2} + L/2) \\ A_{n,3}(\phi_{n,3} + L/2) \\ \vdots \\ A_{n,N_\mathrm{S}}(\phi_{n,N_\mathrm{S}} + L/2) \end{bmatrix} \tag{3.86}$$

and $\tilde{\boldsymbol{S}}_n = \mathrm{diag}\big\{\exp\big(\mathrm{j}\frac{2\pi}{L}c_1\big), \dots, \exp\big(\mathrm{j}\frac{2\pi}{L}c_{N_\mathrm{S}}\big)\big\}^{\hat{x}_1}$.

Note that for candidates $\tilde{\tilde{\boldsymbol{S}}}_n$, for which the argument of the cosine in (3.82) is relatively large, the small–argument approximation of the cosine increases the branch metric $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ even further. This way branches in the outer decoder may be pruned from the tree, although they might actually have lead to the (ML) solution. Thus, the performance of a MSDD decoder using this approach to symbol search is degraded. However, we will see later that only very small performance degradations result.

Clearly, (3.84) can be solved by means of tree–search decoding. The slightly extended task compared to a regular tree–search problem is to find the optimal solution $\hat{\boldsymbol{x}}$ of (3.84) at first call of the inner decoder by the outer decoder, the second best value for $\boldsymbol{x}$ at a possible second call etc. Based on the inequality (3.64) that must be satisfied by a candidate $\tilde{\tilde{\boldsymbol{S}}}_n$ for the latter to be eligible for further consideration the search for (further) candidates $\boldsymbol{x}$ can be limited to those, for which

$$||\boldsymbol{L}\tilde{\boldsymbol{x}} - \boldsymbol{t}||^2 \leq r^{\mathrm{LD}}. \tag{3.87}$$

Plugging (3.83) into (3.64) one can see that r^{LD} is given by

$$r^{\mathrm{LD}} \triangleq \left(\rho - d_{n+1}\big(\tilde{\tilde{\boldsymbol{S}}}_{n+1}\big) - \Big|\Big|\breve{\boldsymbol{R}}_{n,n}\Big|\Big|^2 - ||\boldsymbol{X}_n||^2 + 2\sum_{i=1}^{N_\mathrm{S}} A_{n,i}^2 \right) \frac{L^2}{4\pi^2}. \tag{3.88}$$

If this inner tree–search decoder does not find a (further) candidate given this threshold r^{LD}, the outer tree–search decoder need not pursue the corresponding branch in its decoding tree.

In [CSZ01, LMLK05] the use of lattice reduction (cf. e.g. [LLL82, Coh93, AEVZ02, Win04] and references therein) was advocated to reduce the complexity of finding $\hat{\boldsymbol{x}}$ in (3.84). However, we found that in most scenarios this only increases decoder complexity since it destroys the unique sparse structure of $\boldsymbol{L}$. Note that $\boldsymbol{L}$ only has entries in the first column and on the main diagonal. Consequently, decisions on x_i, $2 \leq i \leq N_{\mathrm{S}}$, are mutually independent and (3.84) is not exactly an N_{S}–dimensional lattice decoding problem, but rather a degenerate version thereof. This reflects in the fact that given a candidate $\tilde{x}_1$ further candidates $\tilde{x}_i$, $2 \leq i \leq N_{\mathrm{S}}$, besides the best candidate

$$\hat{x}_i \stackrel{\triangle}{=} \lfloor (t_i - l_{i,1}\tilde{x}_1)/l_{i,i} \rceil, \quad 2 \leq i \leq N_{\mathrm{S}}, \tag{3.89}$$

where $l_{i,j}$ and t_i denote the elements of $\boldsymbol{L}$ and $\boldsymbol{t}$ in (3.85) and (3.86), respectively, need not be considered. We therefore consider two tree–search algorithms tailored to this particular decoding problem.

Sphere Decoding Algorithm: The decoder may use the SE strategy for finding candidates $\tilde{x}_1 \in \mathbb{Z}$, i.e. based on the unconstrained solution

$$x_1^{\mathrm{u}} = t_1/l_{1,1} \tag{3.90}$$

choose $\tilde{x}_1$ according to

$$\tilde{x}_1 = \lfloor x_1^{\mathrm{u}} \rceil, \lfloor x_1^{\mathrm{u}} \rceil + 1, \lfloor x_1^{\mathrm{u}} \rceil - 1, \lfloor x_1^{\mathrm{u}} \rceil + 2, \ldots \tag{3.91}$$

if $x_1^{\mathrm{u}} > \lfloor x_1^{\mathrm{u}} \rceil$ and

$$\tilde{x}_1 = \lfloor x_1^{\mathrm{u}} \rceil, \lfloor x_1^{\mathrm{u}} \rceil - 1, \lfloor x_1^{\mathrm{u}} \rceil + 1, \lfloor x_1^{\mathrm{u}} \rceil - 2, \ldots \tag{3.92}$$

if $x_1^{\mathrm{u}} \leq \lfloor x_1^{\mathrm{u}} \rceil$, and increase i, $2 \leq i \leq N_{\mathrm{S}}$, as long as

$$\mu_i \stackrel{\triangle}{=} |l_{1,1}\tilde{x}_1 - t_1|^2 + \sum_{j=2}^{i} |l_{j,1}\tilde{x}_1 + l_{j,j}\hat{x}_j - t_j|^2 < r^{\mathrm{LD}} \tag{3.93}$$

with r^{LD} and $\hat{x}_j$ according to (3.88) and (3.89), respectively. Contrary to the regular SpD as described in Section 3.1.1.3.3 this decoder jumps back to the root, i.e. examines the next–best admissible[13] candidate x_1, if $\mu_i > r^{\mathrm{LD}}$ for $2 \leq i \leq N_{\mathrm{S}}$. If $\mu_{N_{\mathrm{S}}} < r^{\mathrm{LD}}$ the current x_1 is stored as tentative decoding result and the decoder checks the next candidate for x_1 with updated

[13]Besides previously examined candidates, candidates x_1 that are congruent modulo L with the former are excluded from the search as well.

$r^{\mathrm{LD}} := \mu_{N_\mathrm{S}}$. The decoder terminates, when $|l_{1,1}\tilde{x}_1 - t_1|^2 > r^{\mathrm{LD}}$ for some candidate $\tilde{x}_1$. Thus the decoder effectively searches only one dimension of the alleged N_S–dimensional search space.

We also note that the *lower*–triangular structure of $\boldsymbol{L}$ is particularly advantageous for this application, since inadmissible candidates are excluded from the search directly at the root, which —for incorrect candidates— leads to fast growing metrics close to the root and thereby rapid termination of the search process.

Stack Decoding Algorithm: One drawback of the SpD algorithm is that the search restarts from $i = 1$ and $\tilde{x}_1 = \lfloor x_1^{\mathrm{u}} \rceil$ [cf.(3.90)] every time the inner decoder is called to provide the next–best candidate $\tilde{\boldsymbol{S}}_n$ given the same $\boldsymbol{X}_n$. Here, the stack algorithm appears to be an appealing alternative for the inner decoder, since it maintains a sorted list of examined partial paths and decoding can be continued if the inner decoder is called again.

In particular a degenerate variant of Vinck's algorithm (cf. Section 3.1.1.3.1) is well apt for the problem. This variant of the stack algorithm replaces the path $\boldsymbol{x}_i = [x_1, \ldots, x_i]$ currently at the top of the stack with its best extension $\boldsymbol{x}^{(i+1)}$ and —due to the sparse structure of the matrix $\boldsymbol{L}$— the next–best admissible candidate for x_i only when $i = 1$. The search can be terminated without loss of optimality, if the metric of the path at the top of the stack exceeds the threshold r^{LD} defined in (3.88). As for the regular stack algorithm, the first path of length N_S to reach the top of the stack is the (next–)best path corresponding to the (next–)best candidate for $\tilde{\boldsymbol{S}}_n$. It remains to store the stack and to continue with this stack when the next candidate given the same $\boldsymbol{X}_n$ is required.

In total, we therefore have to keep track of $N-1$ stacks, one for each tree depth in the outer decoder. While, theoretically, limitation of the stack size may lead to decoding errors, we found that stacks of maximal size $2L$ are sufficient to not impair decoder performance noticeably.

Remark: Numerical tests revealed that the use of a Fano–type metric in the inner decoder *in*– rather than *de*creases the overall decoder complexity. This observation can be explained by the fact that a Fano–type metric at times leads to suboptimal decoding results, which means that the outer decoder has an increased risk of being needlessly mislead by the inner decoder to incorrect paths.

3.1.4.1.3 Bound–Intersect–Detect Symbol Search As alternative to the above lattice–decoder based algorithms for candidate enumeration (inner decoder in the nested structure) we consider symbol search based on the bound–intersect–detect (BID) algorithm of Cui et al., cf. [CT05a, CT06].

The underlying idea is to form a set of promising candidates $\tilde{\boldsymbol{S}}_n = \mathrm{diag}\big\{\exp\big(\mathrm{j}\frac{2\pi}{L}c_1\big), \ldots, \exp\big(\mathrm{j}\frac{2\pi}{L}c_{N_\mathrm{S}}\big)\big\}^{\tilde{l}_n}$ by determining N_S pairs of upper and lower *bounds* on $\mathrm{mod}\big(c_i\tilde{l}_n, L\big)$, $1 \leq i \leq N_\mathrm{S}$, and *intersecting* the resulting N_S sets of candidates $\tilde{l}_n$ to obtain a single smaller set

of candidates that are returned to the outer tree–search decoder as potential candidates for further examination.

Starting from (3.80) we can write

$$\delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) = \sum_{i=1}^{N_\mathrm{S}} B_{n,i}^2 + 2A_{n,i}^2 \cos\left(\frac{2\pi}{L}\left(c_i\tilde{l}_n - \phi_{n,i}\right)\right), \tag{3.94}$$

for the branch metric of a candidate $\tilde{\boldsymbol{S}}_n = \mathrm{diag}\big\{\exp\big(\mathrm{j}\frac{2\pi}{L}c_1\big), \ldots, \exp\big(\mathrm{j}\frac{2\pi}{L}c_{N_\mathrm{S}}\big)\big\}^{\tilde{l}_n}$ with $A_{n,i}$ and $\phi_{n,i}$, as defined in (3.78) and (3.79), respectively, and

$$B_{n,i} \stackrel{\triangle}{=} \sqrt{\sum_{j=1}^{N_\mathrm{R}} \left|\breve{r}_{n,n,i,j}\right|^2 + \left|x_{n,j,i}\right|^2}, \quad 1 \le i \le N_\mathrm{S}. \tag{3.95}$$

Bound: In [CT05a] it is argued, that a necessary condition for

$$\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big) \le r^{\mathrm{BID}} \stackrel{\triangle}{=} \rho - d_{n+1}\Big(\tilde{\tilde{\boldsymbol{S}}}_{n+1}\Big) \tag{3.96}$$

to hold is that

$$B_{n,i}^2 + 2A_{n,i}^2 \cos\left(\frac{2\pi}{L}\left(c_i\tilde{l}_n - \phi_{n,i}\right)\right) \le r^{\mathrm{BID}} \tag{3.97}$$

must hold for all $i \in \{1, \ldots, N_\mathrm{S}\}$. With the definition of N_S pairs of lower and upper bounds

$$B_{\mathrm{L},n,i}^{\mathrm{BID}} \stackrel{\triangle}{=} \left\lceil \frac{L}{2\pi}\arccos\left(\frac{r^{\mathrm{BID}} - B_{n,i}^2}{2A_{n,i}^2}\right) + \phi_{n,i} \right\rceil \tag{3.98}$$

$$B_{\mathrm{U},n,i}^{\mathrm{BID}} \stackrel{\triangle}{=} \left\lfloor L - \frac{L}{2\pi}\arccos\left(\frac{r^{\mathrm{BID}} - B_{n,i}^2}{2A_{n,i}^2}\right) + \phi_{n,i} \right\rfloor, \tag{3.99}$$

$1 \le i \le N_\mathrm{S}$, and a mapping function

$$\mathrm{map}(i, \mathrm{mod}(c_i j, L)) = j, \quad 1 \le j \le L, \tag{3.100}$$

which can be implemented using a look–up table, one can see that a set $\mathcal{L}_{n,i}^{\mathrm{BID}}$ of all $\tilde{l}_n$ that satisfy (3.97) for a particular $i \in \{1, \ldots, N_\mathrm{S}\}$ is given by

$$\mathcal{L}_{n,i}^{\mathrm{BID}} \stackrel{\triangle}{=} \begin{cases} \{\} & \text{if } B_{n,i}^2 - 2A_{n,i}^2 > r^{\mathrm{BID}} \\ \{0, 1, \ldots, L-1\} & \text{if } B_{n,i}^2 + 2A_{n,i}^2 < r^{\mathrm{BID}} \\ \big\{\tilde{l}_n = \mathrm{map}(i, \mathrm{mod}(j, L)) \mid j \in \{B_{\mathrm{L},n,i}^{\mathrm{BID}}, \ldots, B_{\mathrm{U},n,i}^{\mathrm{BID}}\}\big\} & \text{otherwise} \end{cases}. \tag{3.101}$$

Intersect: In a second step the intersection is taken of all candidate sets $\mathcal{L}_{n,i}^{\mathrm{BID}}$, $1 \leq i \leq N_{\mathrm{S}}$, i.e.

$$\mathcal{L}_n^{\mathrm{BID}} = \bigcap_{i=1}^{N_{\mathrm{S}}} \mathcal{L}_{n,i}^{\mathrm{BID}}, \tag{3.102}$$

to determine a set $\mathcal{L}_n^{\mathrm{BID}}$, whose elements $\tilde{l}_n$ satisfy (3.97) for all $i \in \{1, \ldots, N_{\mathrm{S}}\}$. For all candidates in this set the metric $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ is computed and the candidates are sorted in order of increasing $\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$ to facilitate SE enumeration in the outer decoder.

Detect: This part of BID is not required here. Instead the inner BID decoder passes the set $\mathcal{L}_n^{\mathrm{BID}}$ along with the corresponding branch metrics to the outer decoder which stores the set and processes its elements in order of increasing branch metric.

Note that BID symbol search can —like the symbol search algorithms proposed in Section 3.1.4.1.2— be used in combination with any of the above algorithms as outer tree–search decoder operating either on the ML or the Fano–type metric. We indicate its deployment by means of a suffix "–BID", e.g. MSDSD–BID(–FM). Note also that, since (i) $\mathcal{L}_n^{\mathrm{BID}}$ contains *at least* all values of $\tilde{l}_n$ for which (3.96) is fulfilled, and (ii) there is no approximation involved as in the case of LD symbol search, the deployment of BID symbol search does not inhere a performance degradation compared to FS symbol search.

3.1.4.2 Dicyclic Codes

Due to the close relationship between dicyclic (cf. Section 2.1.1.2) and cyclic signal constellations, we can make use of the various methods described in Section 3.1.4.1 for cyclic codes. In addition, we can exploit that $c_{N_{\mathrm{S}}/2+i} = -c_i$, $1 \leq i \leq N_{\mathrm{S}}/2$, in a way that the symbol search can be performed using two separate decoders as for $(N_{\mathrm{S}}/2)$–dimensional cyclic signal constellations, cf. Section 3.1.4.1, one for each value of $m \in \{0, 1\}$ [cf. (2.4)]. These two decoders should run concurrently with the same threshold (cf. (3.88) and (3.96) for LD– and BID–based symbol search, respectively), as the metric of the decoder assuming the correct value of $m \in \{0, 1\}$ is likely to be much smaller than that of the other, which leads to a quicker termination of decoding.

3.1.4.3 Orthogonal Codes

Here, we consider orthogonal DSTM codes with $N_{\mathrm{S}} = 2$, i.e. Alamouti's space–time code, cf. Section 2.1.2.1.

Since —differently from DSTM based on group codes— transmit symbols are in general not chosen from a finite set, it is advisable to search directly for the data symbols $\boldsymbol{V}_n$. Thus taking

the Hermitian transpose of the argument of the Frobenius norm in (3.62), multiplying it from the left with unitary $\tilde{\boldsymbol{S}}_{n+1}$ and using the differential encoding rule (3.3), we can write

$$\delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) = \Big|\Big|\tilde{\boldsymbol{V}}_n\breve{\boldsymbol{R}}_{n,n} + \boldsymbol{Y}_n\Big|\Big|^2, \tag{3.103}$$

where

$$\boldsymbol{Y}_n \stackrel{\triangle}{=} \tilde{\boldsymbol{S}}_{n+1}\sum_{j=n+1}^{N}\tilde{\boldsymbol{S}}_j^{\mathsf{H}}\breve{\boldsymbol{R}}_{n,j}. \tag{3.104}$$

Using a number of fairly straightforward manipulations we eventually obtain

$$\begin{aligned}\delta_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) &= \Big|\Big|\breve{\boldsymbol{R}}_{n,n}\Big|\Big|^2 + ||\boldsymbol{Y}_n||^2 + \mathrm{Re}\left\{\tilde{a}_n\alpha_n^*\right\} + \mathrm{Re}\left\{\tilde{b}_n\beta_n^*\right\} &(3.105)\\ &= \Big|\Big|\breve{\boldsymbol{R}}_{n,n}\Big|\Big|^2 + ||\boldsymbol{Y}_n||^2 + |\alpha_n|\cos\left(\frac{2\pi}{L}\tilde{l}_{a,n} - \mathrm{angle}(\alpha_n)\right) + |\beta_n|\cos\left(\frac{2\pi}{L}\tilde{l}_{b,n} - \mathrm{angle}(\beta_n)\right), &(3.106)\end{aligned}$$

for $1 \le n \le N-1$, and $\delta_N\big(\tilde{\tilde{\boldsymbol{S}}}_N\big) = \big|\big|\breve{\boldsymbol{R}}_{N,N}\big|\big|^2$, where

$$\begin{aligned}\alpha_n &\stackrel{\triangle}{=} \sqrt{2}\sum_{j=1}^{N_{\mathrm{R}}}\breve{r}_{n,n,1,j}^* y_{n,1,j} + \breve{r}_{n,n,2,j}y_{n,2,j}^*, &(3.107)\\ \beta_n &\stackrel{\triangle}{=} \sqrt{2}\sum_{j=1}^{N_{\mathrm{R}}}\breve{r}_{n,n,1,j}^* y_{n,2,j} - \breve{r}_{n,n,2,j}y_{n,1,j}^*, &(3.108)\end{aligned}$$

$\breve{r}_{n,n,i,j}$ and $y_{n,i,j}$ denote the elements in the ith row and jth column of $\breve{\boldsymbol{R}}_{n,n}$ and $\boldsymbol{Y}_n$, respectively, and

$$\begin{aligned}\tilde{a}_n &= \mathrm{e}^{\mathrm{j}\frac{2\pi}{\sqrt{L}}\tilde{l}_{a,n}} &(3.109)\\ \tilde{b}_n &= \mathrm{e}^{\mathrm{j}\frac{2\pi}{\sqrt{L}}\tilde{l}_{b,n}} &(3.110)\end{aligned}$$

represent the two $\sqrt{L}$–PSK symbols that form $\tilde{\boldsymbol{V}}_n$, cf. (2.6). Apparently, the matrix–symbol search decomposes into two independent SE enumerations regarding $\sqrt{L}$–PSK symbols $\tilde{a}_n$ and $\tilde{b}_n$, and outer and inner decoder for MSDD can be integrated into a single decoder searching a $(2N-2)$–dimensional tree of $\sqrt{L}$–PSK symbols, where decision ons $\tilde{a}_n$ and $\tilde{b}_n$ may be made e.g. at depths $\nu = 2n-1$ and $\nu = 2n$, $1 \le n \le N-1$, respectively. Therefore, the tree–search algorithms as discussed in Section 3.1.1 can be applied directly.

Clearly, the SE candidate enumeration for $\tilde{a}_n$ and $\tilde{b}_n$ can be implemented in analogy to the single–antenna case, i.e. based on the unconstrained solution

$$l_{a,n}^{\mathrm{u}} = \frac{\sqrt{L}}{2\pi}\mathrm{angle}(-\alpha_n) \tag{3.111}$$

the order of candidates $\tilde{a}_n = \exp\left(\mathrm{j}\frac{2\pi}{\sqrt{L}}\tilde{l}_{a,n}\right)$ is determined by sequences

$$\tilde{l}_{a,n} = \lfloor l^{\mathrm{u}}_{a,n} \rceil, \lfloor l^{\mathrm{u}}_{a,n} \rceil + 1, \lfloor l^{\mathrm{u}}_{a,n} \rceil - 1, \lfloor l^{\mathrm{u}}_{a,n} \rceil + 2, \ldots \tag{3.112}$$

if $l^{\mathrm{u}}_{a,n} > \lfloor l^{\mathrm{u}}_{a,n} \rceil$ and

$$\tilde{l}_{a,n} = \lfloor l^{\mathrm{u}}_{a,n} \rceil, \lfloor l^{\mathrm{u}}_{a,n} \rceil - 1, \lfloor l^{\mathrm{u}}_{a,n} \rceil + 1, \lfloor l^{\mathrm{u}}_{a,n} \rceil - 2, \ldots \tag{3.113}$$

if $l^{\mathrm{u}}_{a,n} \leq \lfloor l^{\mathrm{u}}_{a,n} \rceil$, and equivalently for $\tilde{b}_n$ based on β_n.

3.1.4.4 Cayley Codes

In the case of Cayley codes, the inner decoder can either perform a full–scale (FS) or lattice–decoder (LD) based symbol search. The LD is based on an efficient CDD decoder for this constellation proposed in [HH02], which relies on a "linear approximation" (LA) and thus is suboptimal. The lattice decoder solves the Q–dimensional minimization problem [HH02, Eqs. (19), (21)] with respect to the real–valued coefficients $\boldsymbol{\alpha} \triangleq [\alpha_1, \ldots, \alpha_Q]^{\mathsf{T}}$ (cf. Section 2.1.2.2)

$$\hat{\boldsymbol{\alpha}} = \underset{\tilde{\boldsymbol{\alpha}} \in \mathcal{A}^Q_{\mathrm{cay}}}{\operatorname{argmin}} \left\{ \left|\left| \boldsymbol{B}_{\mathrm{cay}} \tilde{\boldsymbol{\alpha}} - \boldsymbol{t}_{\mathrm{cay}} \right|\right|^2 \right\}, \tag{3.114}$$

with a $(2N_{\mathrm{S}}N_{\mathrm{R}} \times Q)$–dimensional, non–sparse matrix $\boldsymbol{B}_{\mathrm{cay}}$ (see [HH02] for details).[14] Different LD algorithms consisting of a preprocessing and a tree–search stage could be applied, cf. e.g. [AEVZ02, MGDC06]. Since, we found that the computational burden of preprocessing dominates the overall complexity, we concentrate on minimal preprocessing performing the QR decomposition $\boldsymbol{B}_{\mathrm{cay}} = \boldsymbol{Q}_{\mathrm{cay}}\boldsymbol{R}_{\mathrm{cay}}$ with upper–right triangular $\boldsymbol{R}_{\mathrm{cay}}$. Based on this, it is convenient to apply the modification of Vinck's stack decoder as described in Section 3.1.4.1.2 to find candidates $\tilde{\boldsymbol{\alpha}} \in \mathcal{A}^Q_{\mathrm{cay}}$ in order of non–decreasing values of

$$\left|\left| \boldsymbol{R}_{\mathrm{cay}} \tilde{\boldsymbol{\alpha}} - \breve{\boldsymbol{t}}_{\mathrm{cay}} \right|\right|^2 \quad \text{with} \quad \breve{\boldsymbol{t}}_{\mathrm{cay}} \triangleq \left(\boldsymbol{Q}_{\mathrm{cay}}\right)^{\mathsf{T}} \boldsymbol{t}_{\mathrm{cay}}. \tag{3.115}$$

Finally, it is worth pointing out that it is helpful to let the inner decoder return the true value of $\delta_n\left(\tilde{\tilde{\boldsymbol{S}}}_n\right)$ to the outer decoder rather than the result obtained from the linear approximation. This both accelerates the convergence of the entire MSDD decoder and improves the error performance as it reduces the effect of the approximation involved in (3.114).[15]

[14] $\boldsymbol{B}_{\mathrm{cay}}$ and $\boldsymbol{t}_{\mathrm{cay}}$ are computed as described in Section II-D of [HH02] by replacing —what is there denoted as— $\boldsymbol{X}_\tau$ and $\boldsymbol{X}_{\tau-1}$ with $-\boldsymbol{Y}_n$ and $\breve{\boldsymbol{R}}_{n,n}$, cf. (3.104) and (3.17) respectively, when searching for candidates for $\boldsymbol{V}_n$.

[15] Modifying the inner sphere decoder such that it computes the true $\delta_n\left(\tilde{\tilde{\boldsymbol{S}}}_n\right)$ whenever a new full candidate vector $\boldsymbol{\alpha}$ is found and only updates its tentative best candidate when $\delta_n\left(\tilde{\tilde{\boldsymbol{S}}}_n\right)$ of the current candidate is smaller than that of the best so far, does not lead to further improvements in the performance of the overall decoder.

3.2 MSDD Based on Combinatorial Geometry

The tree–search based MSDD algorithms presented in Section 3.1 succeed nicely in drastically reducing the complexity required to perform MSDD over wide ranges of SNR and observation window lengths N. However, these algorithms also have a number of drawbacks. (i) Their *instantaneous* computational complexity is a random variable, which is considered disadvantageous for practical implementation. If the ML–MSDD solution is desired at all times their (ii) *average* computational complexity may become prohibitively large in low SNRs and (iii) *worst–case* complexity increases exponentially with N.

In this section, we therefore take a completely different approach to MSDD. We are inspired by the recent work by Manglani and Chaturvedi [MC06], who expressed the multiuser detection problem in code–division multiple access (CDMA) as the quadratic maximization problem

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\{0,1\}^n}{\operatorname{argmax}} \left\{ \boldsymbol{x}^{\mathsf{T}} \boldsymbol{Q} \boldsymbol{x} \right\}, \tag{3.116}$$

with an $(n \times n)$–dimensional positive semidefinite matrix $\boldsymbol{Q}$ of rank $d \leq n$, and applied methods from combinatorial geometry (CG) [Ede87] to solve it. In particular, Ferrez et al. [FFL05] proposed an efficient algorithm that solves this quadratic maximization problem in *polynomial* time $O(n^{d-1})$. In this section, we show that in the popular special case of binary and quaternary phase–shift keying (BPSK and QPSK) modulation, i.e. for

$$N_{\mathrm{S}} = 1, \tag{3.117}$$

also MSDD can be cast in the form of (3.116). The matrix $\boldsymbol{Q}$ is then related to the inverse $\boldsymbol{M}$ of the $(N \times N)$–dimensional fading–plus–noise autocorrelation matrix. By applying the algorithm in [FFL05] the MSDD problem is solved in practically constant —meaning independent both of the SNR and the instantaneous channel state— *polynomial* time with respect to N, provided that the rank of the fading autocorrelation matrix is fixed, i.e. independent of N. We refer to this MSDD making use of the algorithm from [FFL05] as CG–MSDD. What makes this approach to MSDD attractive, is the fact that the rank of the autocorrelation matrix $\boldsymbol{\Psi}_{gg}$ of the fading process, which is closely related to the above parameter d, is usually much smaller than N.

After a brief introduction into the fundamentals of convex quadratic maximization based on combinatorial geometry in Section 3.2.1, we apply these methods to MSDD in Section 3.2.2.

3.2.1 Preliminaries from Convex Quadratic Maximization

In [FFL05] the unconstrained quadratic maximization problem

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\{0,1\}^n}{\operatorname{argmax}} \left\{ \boldsymbol{x}^{\mathsf{T}} \boldsymbol{Q} \boldsymbol{x} \right\} \tag{3.118}$$

with a real–valued, symmetric, and positive semidefinite $(n \times n)$–dimensional matrix $\boldsymbol{Q}$ of fixed rank d was considered and shown to be solvable in time polynomial in n. The essential steps towards this result are as follows. Since the matrix $\boldsymbol{Q}$ is normal, we can factorize it —e.g. by means of an eigenvalue decomposition— as

$$\boldsymbol{Q} = \boldsymbol{T}^{\mathsf{T}}\boldsymbol{T} \tag{3.119}$$

with $(d \times n)$–dimensional $\boldsymbol{T}$. Furthermore, taking the convexity of the Frobenius norm into account [FFL05], (3.118) is equivalent to maximizing $||\boldsymbol{T}\boldsymbol{x}||$ over the *hypercube* $[0,\,1]^n$, i.e.

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in[0,1]^n}{\operatorname{argmax}}\left\{ ||\boldsymbol{T}\boldsymbol{x}||^2 \right\}. \tag{3.120}$$

The image

$$\boldsymbol{\mathcal{Z}}(\boldsymbol{T}) \triangleq \{\boldsymbol{z} = \boldsymbol{T}\boldsymbol{x} \mid \boldsymbol{x} \in [0,1]^n\} \subset \mathbb{R}^d \tag{3.121}$$

of the hypercube $[0,1]^n$ is a convex polytope also referred to as *zonotope*, cf. e.g. [FFL05]. In these terms the optimization (3.118) is equivalent to finding the element $\boldsymbol{z} = \boldsymbol{T}\boldsymbol{x}$ of $\boldsymbol{\mathcal{Z}}$ that lies farthest (in terms of Euclidean distance) from the origin. Due to the definition of $\boldsymbol{\mathcal{Z}}$ the maximum of $||\boldsymbol{T}\boldsymbol{x}||$ lies at some corner (or *extreme point*) of the zonotope $\boldsymbol{\mathcal{Z}}$, each of which corresponds one–to–one to some vector $\boldsymbol{x} \in \{0,1\}^n$. Consequently, the optimization problem (3.118) can be solved by enumerating the extreme points of the zonotope and evaluating the Frobenius norm for these points. Note however, that the number of extreme points of the zonotope is not equal to 2^n, but of the order $O(n^{d-1})$, as some of the linear combinations of the columns of $\boldsymbol{T}$ lead to points that lie inside the convex zonotope.

For illustration Fig. 3.9 shows the zonotope corresponding to a randomly chosen (3×4)–dimensional matrix

$$\boldsymbol{T} = \begin{bmatrix} 7 & 11 & -12 & 14 \\ -2 & 16 & 13 & -2 \\ 8 & -3 & 15 & 7 \end{bmatrix}. \tag{3.122}$$

"Visible" edges are drawn as solid lines, "hidden" edges as dashed lines. For clarity, we highlighted the four edges of the zonotope corresponding to the four columns

$$\boldsymbol{t}_i \triangleq \boldsymbol{T}_{:,i},\ 1 \leq i \leq n, \tag{3.123}$$

($\boldsymbol{X}_{:,i}$: ith column of a matrix $\boldsymbol{X}$) of $\boldsymbol{T}$ (*generators* of $\boldsymbol{\mathcal{Z}}$) and exemplarily marked the extreme point of $\boldsymbol{\mathcal{Z}}$ corresponding to $\boldsymbol{x} = [1,1,1,0]^{\mathsf{T}}$.

In combinatorial geometry many results are based on the dual representation in a *central arrangement*

$$\mathcal{A}(\boldsymbol{T}) \triangleq \{\boldsymbol{\mathcal{P}}_i \mid 1 \leq i \leq n\} \tag{3.124}$$

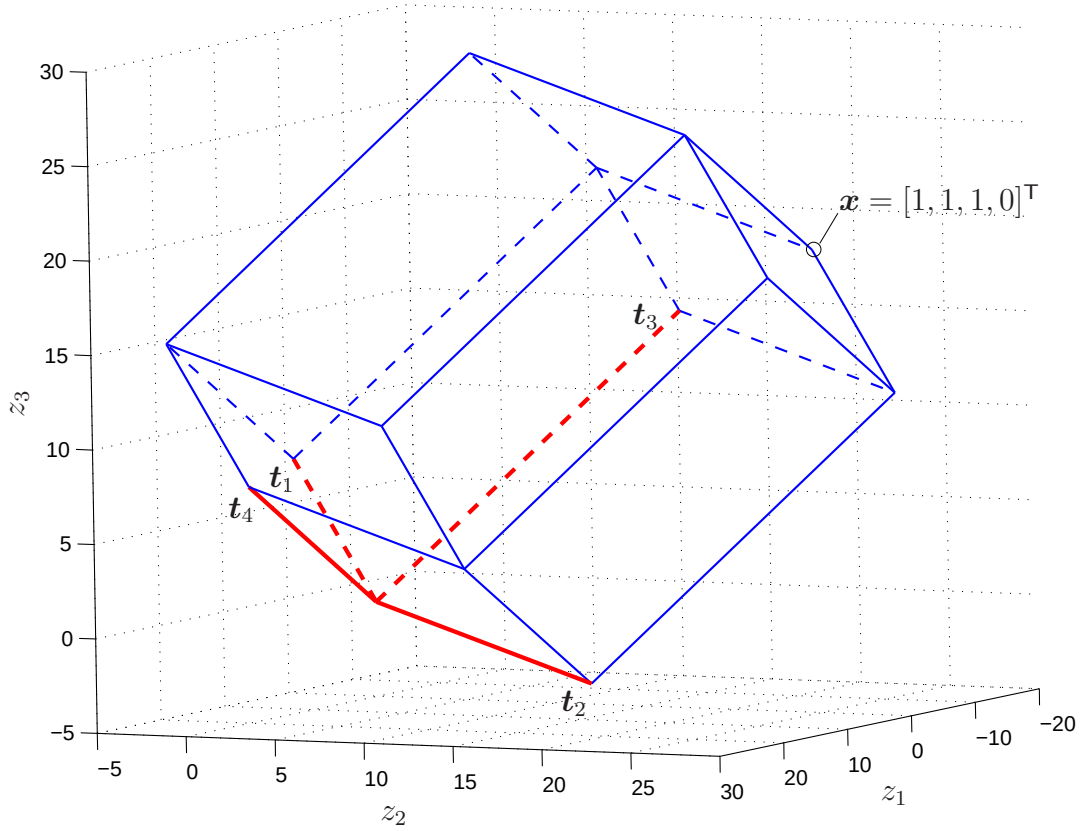

Figure 3.9: Zonotope $\boldsymbol{\mathcal{Z}}$ corresponding to (3×4)–dimensional $\boldsymbol{T}$.

of n hyperplanes $\boldsymbol{\mathcal{P}}_i$ through the origin of the $\mathbb{R}^d$ with $\boldsymbol{t}_i$ as normal vector, i.e.

$$\boldsymbol{\mathcal{P}}_i \triangleq \left\{ \boldsymbol{y} \in \mathbb{R}^d \mid \boldsymbol{y}^{\mathsf{T}} \boldsymbol{t}_i = 0 \right\}. \tag{3.125}$$

These hyperplanes partition the $\mathbb{R}^d$ into a number of convex *faces*

$$\boldsymbol{\mathcal{C}}_{\boldsymbol{\sigma}} \triangleq \left\{ \boldsymbol{y} \in \mathbb{R}^d \mid \operatorname{sgn}\left(\boldsymbol{y}^{\mathsf{T}} \boldsymbol{t}_i\right) = \sigma_i, \forall i \in \{1, \ldots, n\} \right\}, \tag{3.126}$$

($\operatorname{sgn}(x)$: sign of x) where $\boldsymbol{\sigma} \in \{-1, 0, 1\}^n$. Values of -1, 1 and 0 for σ_i indicate that the elements of a face $\boldsymbol{\mathcal{C}}_{\boldsymbol{\sigma}}$ lie on one side or the other of, or directly on the hyperplane $\boldsymbol{\mathcal{P}}_i$ corresponding to $\boldsymbol{t}_i$, respectively. This results in $(d+1)$ types of faces of different dimensions (point, lines, planes, etc.) depending on the number of indeces i, for which $\sigma_i = 0$. However, the d–dimensional faces, subsequently referred to as *cells*, are of particular interest. Here, $\boldsymbol{\sigma} \in \{-1, 1\}^n$ indicates, on which side of each hyperplane $\boldsymbol{\mathcal{P}}_i$, $1 \leq i \leq n$, the cell $\boldsymbol{\mathcal{C}}_{\boldsymbol{\sigma}}$ lies. For these cells a one–to–one correspondence to one extreme point of the zonotope can be proven, cf. [Zie94]. More specifically, the extreme point corresponding to a cell $\boldsymbol{\mathcal{C}}_{\boldsymbol{\sigma}}$ specified by the sign vector $\boldsymbol{\sigma}$ is given by the sum over all columns $\boldsymbol{t}_i$ of $\boldsymbol{T}$ corresponding to hyperplanes $\boldsymbol{\mathcal{P}}_i$, for which the sign σ_i is

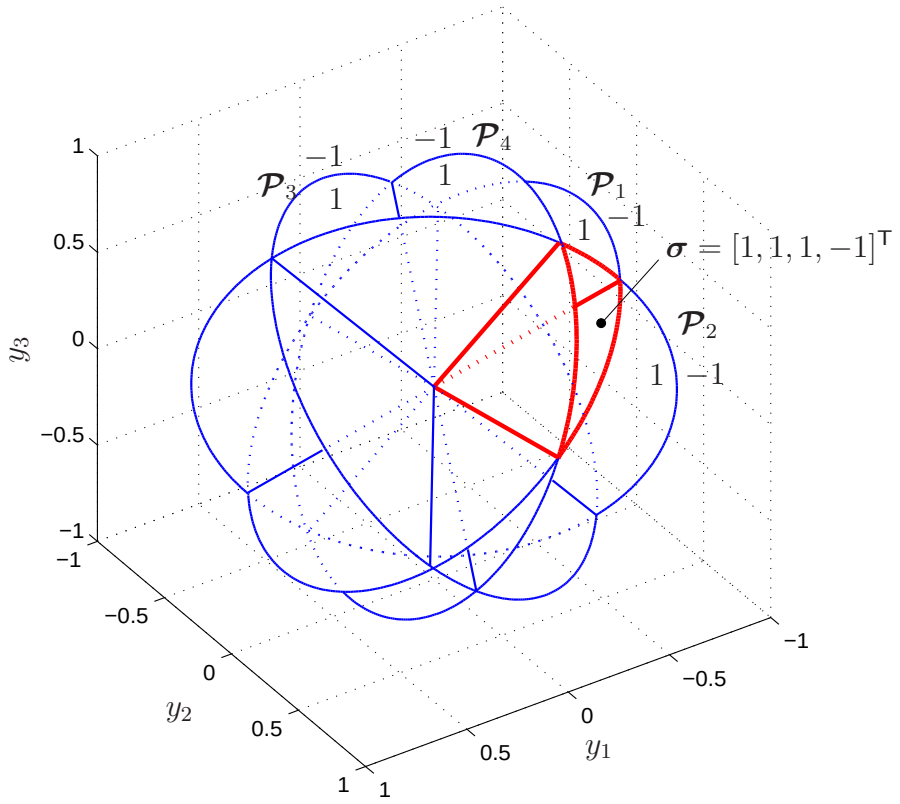

Figure 3.10: Central arrangement $\mathcal{A}(\boldsymbol{T})$ corresponding to (3×4)–dimensional $\boldsymbol{T}$ (shown as cut section with the unit–sphere).

positive. I.e. the corresponding extreme point is given by [Zie94]

$$\sum_{i=1}^{n} \boldsymbol{t}_i(\sigma_i + 1)/2. \tag{3.127}$$

Thus, by definition of the zonotope [cf. (3.121)] a one–to–one correspondence between sign vectors $\boldsymbol{\sigma}$ and binary vectors $\boldsymbol{x}$ exists via

$$x_i = (\sigma_i + 1)/2, \quad 1 \leq i \leq n, \tag{3.128}$$

and the hyperplanes $\boldsymbol{\mathcal{P}}_i$ can be viewed as decision boundaries for x_i, but in the $\mathbb{R}^d$ rather than in the $\mathbb{R}^n$.

For illustration see Fig. 3.10, which shows the central arrangement $\mathcal{A}(\boldsymbol{T})$ corresponding to the zonotope of Fig. 3.9. Visible contours are plotted as solid, whereas hidden ones are visualized as dotted lines. The n hyperplanes are depicted through the contour of their intersection with the unit–sphere and for each a pair of $-1, 1$ for σ_i is given to indicate the orientation of $\boldsymbol{t}_i$. We also highlighted the cell corresponding to the extreme point of the zonotope in Fig. 3.9, i.e. to $\boldsymbol{x} = [1, 1, 1, 0]^{\mathsf{T}}$.

Based on this duality, the problem of optimizing (3.118) can also be solved through enumeration of the cells of the central arrangement and evaluation of the Frobenius norm of the corresponding extreme points of the zonotope [cf. (3.127)].

In order for this approach to lead to the solution in polynomial time, essentially two things are required. (i) The number of cells must be polynomial in n, since the complexity of evaluating the Frobenius norm of an individual candidate is linear in n if d is fixed. (ii) The complexity of finding these cells must also be polynomial in n.

The first prerequisite has been known to be fulfilled for a long time. It has been shown in [Buc43, Zas75] that the number of cells in the central arrangement of n hyperplanes in the $\mathbb{R}^d$ or equivalently the number of extreme points of a zonotope generated by n column vectors in the $\mathbb{R}^d$ is upper bounded by

$$2\sum_{i=0}^{d-1}\binom{n-1}{i}, \tag{3.129}$$

and that this upper bound is achieved with equality if the columns $\boldsymbol{t}_i$ of $\boldsymbol{T}$ are in general position, meaning if any d–tuple of columns $\boldsymbol{t}_{i_j}$, $1 \leq j \leq d$, of $\boldsymbol{T}$ spans the $\mathbb{R}^d$. The probability of the latter event equals one as the algorithm operates on real numbers. Thus, the number of cells is given by (3.129), i.e. it is of order $O(n^{d-1})$.

Second, algorithms to enumerate these cells in polynomial, more specifically $O(n^{d-1})$ time can be found in e.g. [EOS86, Ede87, FFL05].[16] Having enumerated all cells, it remains to compute $||\boldsymbol{T}\boldsymbol{x}||$ for these $O(n^{d-1})$ candidates to find the best one. As the complexity of evaluating $||\boldsymbol{T}\boldsymbol{x}||$ is linear in n if d is fixed, the overall complexity is of order

$$O(n^d). \tag{3.130}$$

For more details on the complexity analysis of these algorithms when applied to noncoherent transmission over fading channels see Section 5.5.2.

Remark: For CG–MSDD for BPSK and QPSK, we shall apply the above result to the case that the support $\{0,1\}^n$ of $\boldsymbol{x}$ is replaced by $\{-1,+1\}^n$ in (3.118). It is important to note that one should not change variables to deal with this case, as such an operation might increase the rank d of the associated matrix $\boldsymbol{Q}$, i.e. the number of columns in $\boldsymbol{T}$ and thereby the complexity of the decoder. Fortunately, we can treat this case directly. In fact, the image $\boldsymbol{\mathcal{Z}}'$ of the linear map $\boldsymbol{T}\boldsymbol{x}$ of the centrally symmetric cube $\boldsymbol{x} \in \{-1,+1\}^n$ is simply a scaled and shifted version of the zonotope $\boldsymbol{\mathcal{Z}}$ in (3.121), more specifically

$$\boldsymbol{\mathcal{Z}}' = 2\boldsymbol{\mathcal{Z}} - \sum_{i=1}^{d}\boldsymbol{t}_i. \tag{3.131}$$

[16] Implementation of the cell enumeration algorithm by Fukuda et al. in C can be found under [FFL, FF].

Thus, the extreme points of $\boldsymbol{\mathcal{Z}}'$ can be computed in exactly the same manner, i.e. based on the same central arrangement $\mathcal{A}$, as those of $\boldsymbol{\mathcal{Z}}$.

3.2.2 Application to MSDD

In order to efficiently apply the above methods of combinatorial geometry, we proceed as follows. Let the $(N \times N)$–dimensional fading autocorrelation matrix $\boldsymbol{\Psi}_{gg}$ [cf. (2.31)] be real–valued and of rank

$$N_\lambda \stackrel{\triangle}{=} \operatorname{rank}\{\boldsymbol{\Psi}_{gg}\}\,. \tag{3.132}$$

Then we can factorize $\boldsymbol{\Psi}_{gg}$ using the eigenvalue decomposition as

$$\boldsymbol{\Psi}_{gg} = \boldsymbol{U}_{\boldsymbol{\Psi}_{gg}} \boldsymbol{\Lambda}_{\boldsymbol{\Psi}_{gg}} \boldsymbol{U}_{\boldsymbol{\Psi}_{gg}}^{\mathsf{T}} \tag{3.133}$$

($\boldsymbol{\Lambda}_{\boldsymbol{X}}$: $(\operatorname{rank}\{\boldsymbol{X}\} \times \operatorname{rank}\{\boldsymbol{X}\})$–dimensional matrix containing the non–zero eigenvalues $\lambda_{\boldsymbol{X},i}$, $1 \leq i \leq \operatorname{rank}\{\boldsymbol{X}\}$ of $\boldsymbol{X}$, sorted in order of decreasing magnitude on its main diagonal, $\boldsymbol{U}_{\boldsymbol{X}}$: $(N \times \operatorname{rank}\{\boldsymbol{X}\})$–dimensional unitary matrix whose columns are given by the eigenvectors of $(N \times N)$–dimensional $\boldsymbol{X}$). With this and with the help of Woodbury's identity for matrix inversion (cf. e.g. [HJ85]) the inverse of the fading–plus–noise autocorrelation matrix $\boldsymbol{M}$ [cf. (2.37)] can be written as

$$\boldsymbol{M} = \sigma_n^{-2} \left(\boldsymbol{I}_N - \boldsymbol{U}_{\boldsymbol{\Psi}_{gg}} \boldsymbol{A} \boldsymbol{U}_{\boldsymbol{\Psi}_{gg}}^{\mathsf{T}} \right), \tag{3.134}$$

where

$$\boldsymbol{A} \stackrel{\triangle}{=} \boldsymbol{\Lambda}_{\boldsymbol{\Psi}_{gg}} \left(\boldsymbol{\Lambda}_{\boldsymbol{\Psi}_{gg}} + \sigma_n^2 \boldsymbol{I}_{N_\lambda} \right)^{-1}. \tag{3.135}$$

Consequently, we can express the ML–MSDD decision rule as

$$\hat{\tilde{\boldsymbol{S}}} = \underset{\tilde{\boldsymbol{S}} \in \boldsymbol{\mathcal{V}}^N, \tilde{\boldsymbol{S}}_N = 1}{\operatorname{argmax}} \left\{ \tilde{\tilde{\boldsymbol{S}}}^{\mathsf{H}} \boldsymbol{Q}_{\mathrm{c}} \tilde{\tilde{\boldsymbol{S}}} \right\}, \text{[17]} \tag{3.136}$$

where

$$\boldsymbol{Q}_{\mathrm{c}} \stackrel{\triangle}{=} \bar{\boldsymbol{R}}_{\mathrm{D}} \boldsymbol{U}_{\boldsymbol{\Psi}_{gg}} \boldsymbol{A} \boldsymbol{U}_{\boldsymbol{\Psi}_{gg}}^{\mathsf{T}} \bar{\boldsymbol{R}}_{\mathrm{D}}^{\mathsf{H}} \tag{3.137}$$

is an $(N \times N)$–dimensional positive semidefinite matrix of rank N_λ,

$$\bar{\boldsymbol{R}}_{\mathrm{D}} \stackrel{\triangle}{=} \operatorname{diag}\{\bar{\boldsymbol{R}}\}\,, \tag{3.138}$$

and $\bar{\boldsymbol{R}}$ as N–dimensional vector of received samples [cf. (2.22)].

For the following derivation of CG–MSDD, we consider the two cases of BPSK ($L = 2$) and QPSK ($L = 4$) separately and subsequently address the issue of transmit and receive diversity.

[17] Note, that although many quantities have singular dimensions in the single–antenna case, e.g. $\bar{\boldsymbol{S}}$ is now a vector of length N and $\boldsymbol{S}_n$ is a scalar, we stick to the general denomination of the MIMO system, to avoid confusion by introducing new variable names.

3.2.2.1 Binary Phase–Shift Keying

In the case of BPSK the imaginary part of $\boldsymbol{Q}_{\mathrm{c}}$ is irrelevant. Therefore, defining

$$\boldsymbol{Q} \stackrel{\triangle}{=} \mathrm{Re}\left\{\boldsymbol{Q}_{\mathrm{c}}\right\}, \tag{3.139}$$

the decision rule (3.136) is equivalent to

$$\hat{\tilde{\boldsymbol{S}}} = \underset{\tilde{\boldsymbol{S}}\in\{-1,+1\}^N, \tilde{\boldsymbol{S}}_N=1}{\mathrm{argmax}} \left\{\tilde{\boldsymbol{S}}^{\mathsf{T}}\boldsymbol{Q}\tilde{\boldsymbol{S}}\right\}. \tag{3.140}$$

It is not difficult to show that $\boldsymbol{Q} = \boldsymbol{T}^{\mathsf{T}}\boldsymbol{T}$ with the $(2N_\lambda \times N)$–dimensional matrix

$$\boldsymbol{T} = \left(\boldsymbol{I}_2 \otimes \sqrt{\boldsymbol{A}}\boldsymbol{U}_{\boldsymbol{\Psi}_{gg}}^{\mathsf{T}}\right) \cdot \begin{bmatrix} \mathrm{Re}\left\{\bar{\boldsymbol{R}}_{\mathrm{D}}\right\} \\ \mathrm{Im}\left\{\bar{\boldsymbol{R}}_{\mathrm{D}}\right\} \end{bmatrix}, \tag{3.141}$$

and $\bar{\boldsymbol{R}}_{\mathrm{D}}$, $\boldsymbol{U}_{\boldsymbol{\Psi}_{gg}}$ and $\boldsymbol{A}$ as defined in (3.138), (3.134) and (3.135), respectively. Hence, the ML–MSDD problem can be represented using a central arrangement of N hyperplanes in the $\mathbb{R}^{2N_\lambda}$, and we can solve it applying the methods from combinatorial geometry described in Section 3.2.1. Note that the complexity of the eigenvalue decomposition of $\boldsymbol{\Psi}_{gg}$ is negligible as it needs to be performed only once provided that the channel statistics do not change. Thus decoder complexity is of order $O(N^{2N_\lambda})$ [cf. (3.130) with $n = N$ and $d = 2N_\lambda$].

Remark: It should be noted that if $2N_\lambda > N$ a decomposition $\boldsymbol{Q} = \boldsymbol{T}^{\mathsf{T}}\boldsymbol{T}$ with an $(N \times N)$–dimensional matrix $\boldsymbol{T}$ different from that in (3.141) can be found. In this case however, the computational complexity of CG–MSDD is of the same order as that of brute–force MSDD, and thus it is not of interest here.

3.2.2.2 Quaternary Phase–Shift Keying

For QPSK the transmit signal $\bar{\boldsymbol{S}}$ can be expressed as $\bar{\boldsymbol{S}} = \mathrm{e}^{\mathrm{j}\frac{\pi}{4}}(\boldsymbol{s}_{\mathrm{Re}} + \mathrm{j}\boldsymbol{s}_{\mathrm{Im}})/\sqrt{2}$ with $\boldsymbol{s}_{\mathrm{Re}}, \boldsymbol{s}_{\mathrm{Im}} \in \{-1,+1\}^N$, and the MSDD decision rule becomes

$$\hat{\tilde{\boldsymbol{S}}} = \underset{\substack{[\tilde{\boldsymbol{s}}_{\mathrm{Re}},\tilde{\boldsymbol{s}}_{\mathrm{Im}}]\in\{-1,+1\}^{2N} \\ \tilde{s}_{\mathrm{Re},N}=1, \tilde{s}_{\mathrm{Im},N}=0}}{\mathrm{argmax}} \left\{ \begin{bmatrix} \tilde{\boldsymbol{s}}_{\mathrm{Re}}^{\mathsf{T}} & \tilde{\boldsymbol{s}}_{\mathrm{Im}}^{\mathsf{T}} \end{bmatrix} \boldsymbol{Q} \begin{bmatrix} \tilde{\boldsymbol{s}}_{\mathrm{Re}} \\ \tilde{\boldsymbol{s}}_{\mathrm{Im}} \end{bmatrix} \right\}, \tag{3.142}$$

where

$$\boldsymbol{Q} \stackrel{\triangle}{=} \begin{bmatrix} \mathrm{Re}\left\{\boldsymbol{Q}_{\mathrm{c}}\right\} & -\mathrm{Im}\left\{\boldsymbol{Q}_{\mathrm{c}}\right\} \\ \mathrm{Im}\left\{\boldsymbol{Q}_{\mathrm{c}}\right\} & \mathrm{Re}\left\{\boldsymbol{Q}_{\mathrm{c}}\right\} \end{bmatrix} \tag{3.143}$$

and $\boldsymbol{Q}_{\mathrm{c}}$ according to (3.137). The factorization $\boldsymbol{Q} = \boldsymbol{T}^{\mathsf{T}}\boldsymbol{T}$ leads to the $(2N_\lambda \times 2N)$–dimensional matrix

$$\boldsymbol{T} = \left(\boldsymbol{I}_2 \otimes \sqrt{\boldsymbol{A}}\boldsymbol{U}_{\boldsymbol{\Psi}_{gg}}^{\mathsf{T}}\right) \cdot \begin{bmatrix} \mathrm{Re}\left\{\bar{\boldsymbol{R}}_{\mathrm{D}}\right\} & \mathrm{Im}\left\{\bar{\boldsymbol{R}}_{\mathrm{D}}\right\} \\ -\mathrm{Im}\left\{\bar{\boldsymbol{R}}_{\mathrm{D}}\right\} & \mathrm{Re}\left\{\bar{\boldsymbol{R}}_{\mathrm{D}}\right\} \end{bmatrix}, \tag{3.144}$$

with $\bar{\boldsymbol{R}}_{\mathrm{D}}$, $\boldsymbol{U}_{\boldsymbol{\Psi}_{gg}}$ and $\boldsymbol{A}$ as defined in (3.138), (3.134) and (3.135), respectively. Similar to the case of BPSK, this factorization is based on the eigenvalue decomposition of $\boldsymbol{\Psi}_{gg}$, which is performed only once assuming that the channel is stationary. Hence, (3.142) can be solved using the methods from combinatorial geometry based on a central arrangement of $2N$ hyperplanes in the $\mathbb{R}^{2N_\lambda}$ and the complexity of CG–MSDD is of the order $O((2N)^{2N_\lambda})$ [cf. (3.130) with $n = 2N$ and $d = 2N_\lambda$].

3.2.2.3 Transmit and Receive Diversity

CG–MSDD can also be applied to repetition transmit diversity in a system with N_{S} inputs to the channel, such that the same PSK symbol is fed into each of the N_{S} inputs exactly once while the remaining $N_{\mathrm{S}} - 1$ inputs are inactive (compare cyclic DSTM codes as reviewed in Section 2.1.1.1). In addition receive diversity with $N_{\mathrm{R}} > 1$ receive antennas. Based on the assumption of $N_{\mathrm{S}}N_{\mathrm{R}}$ iid subchannels between different input–output pairs it is fairly straightforward to extend the derivations above to the MIMO case. Let the $(N \times 1)$–dimensional $\boldsymbol{r}_{i,j} = \big[r_{i,j}[k-N+1], \ldots, r_{i,j}[k]\big]^{\mathsf{T}}$ with $r_{i,j}[k-\kappa]$ as the element in the ith row and jth column of $\boldsymbol{R}[k-\kappa]$ [cf. (2.12)] denote the signal received at channel output j upon transmission from channel input i, $\bar{\boldsymbol{R}}_{\mathrm{D},i,j} \triangleq \mathrm{diag}\{\boldsymbol{r}_{i,j}\}$ and $\tilde{\boldsymbol{R}}_{\mathrm{D}} \triangleq [\bar{\boldsymbol{R}}_{\mathrm{D},1,1}, \ldots, \bar{\boldsymbol{R}}_{\mathrm{D},N_{\mathrm{S}},1}, \bar{\boldsymbol{R}}_{\mathrm{D},1,2}, \ldots, \bar{\boldsymbol{R}}_{\mathrm{D},N_{\mathrm{S}},N_{\mathrm{R}}}]^{\mathsf{T}}$. Then $\boldsymbol{T}$ in $\boldsymbol{Q} = \boldsymbol{T}^{\mathsf{T}}\boldsymbol{T}$ can be expressed as $(2N_{\mathrm{S}}N_{\mathrm{R}}N_\lambda \times N)$–dimensional

$$\boldsymbol{T} = \left(\boldsymbol{I}_{2N_{\mathrm{S}}N_{\mathrm{R}}} \otimes \sqrt{\boldsymbol{A}}\boldsymbol{U}_{\boldsymbol{\Psi}_{gg}}^{\mathsf{T}}\right) \cdot \begin{bmatrix} \mathrm{Re}\{\tilde{\boldsymbol{R}}_{\mathrm{D}}\} \\ \mathrm{Im}\{\tilde{\boldsymbol{R}}_{\mathrm{D}}\} \end{bmatrix}, \tag{3.145}$$

for BPSK and $(2N_{\mathrm{S}}N_{\mathrm{R}}N_\lambda \times 2N)$–dimensional

$$\boldsymbol{T} = \left(\boldsymbol{I}_{2N_{\mathrm{S}}N_{\mathrm{R}}} \otimes \sqrt{\boldsymbol{A}}\boldsymbol{U}_{\boldsymbol{\Psi}_{gg}}^{\mathsf{T}}\right) \cdot \begin{bmatrix} \mathrm{Re}\{\tilde{\boldsymbol{R}}_{\mathrm{D}}\} & \mathrm{Im}\{\tilde{\boldsymbol{R}}_{\mathrm{D}}\} \\ -\mathrm{Im}\{\tilde{\boldsymbol{R}}_{\mathrm{D}}\} & \mathrm{Re}\{\tilde{\boldsymbol{R}}_{\mathrm{D}}\} \end{bmatrix}. \tag{3.146}$$

for QPSK, respectively, where $\boldsymbol{U}_{\boldsymbol{\Psi}_{gg}}$ and $\boldsymbol{A}$ are computed as before from the $(N \times N)$–dimensional autocorrelation matrix $\boldsymbol{\Psi}_{gg}$ of the channel coefficients corresponding to the received samples collected in $\boldsymbol{r}_{i,j}$ [cf. (3.133) and (3.135)]. One can see that the dimension of the parameter space in which the central arrangement is set and thereby the complexity exponent is increased by a factor of $N_{\mathrm{S}}N_{\mathrm{R}}$, which was to be expected as the number of degrees of freedom in the channel is increased by the factor $N_{\mathrm{S}}N_{\mathrm{R}}$. Thus, the decoder complexity is of the order $O((\log_2(L)N)^{2N_\lambda N_{\mathrm{S}}N_{\mathrm{R}}})$, $L \in \{2, 4\}$ for BPSK and QPSK, respectively.

3.2.2.4 Relation to MAPSqD from [MAKA07]

In [MAKA07, Section II] Motedayen–Aval et al. proposed an MSDD algorithm, referred to as maximum a–posteriori sequence detection (MAPSqD), for PSK modulation that relies on

the same principles as the algorithm presented above. As in our case the detection problem is moved from the N–dimensional signal space to an N_λ–dimensional parameter space, where N_λ is the rank of the channel autocorrelation matrix $\mathbf{\Psi}_{gg}$. Rather than exploiting the binary structure underlying BPSK and QPSK to define a central arrangement of hyperplanes based on its relation to a zonotope, Motedayen–Aval et al. define hyperplanes in the parameter space separating cells based on the metric difference between the respective candidate transmit sequences associated with the individual cells. Depending on the sign of the metric difference, a point is identified as lying on one side of a hyperplane or the other (cf. meaning of σ_i in Section 3.2.1), and detection is essentially performed based on combinatorial geometry in this N_λ–dimensional space. While MAPSqD is more general in that it is also applicable to PSK modulation with $L > 4$, CG–MSDD is not a special case of MAPSqD, because the definitions of the N_λ–dimensional parameter spaces are different. For example, CG–MSDD is advantageous over MAPSqD for QPSK since the resulting central arrangement has only $2N$ hyperplanes instead of $6N$ for MAPSqD, which corresponds to significant savings in computational complexity [cf. (3.130)].

Chapter 4

Multiple–Input Multiple–Output Channel Model

Recall, that in Section 2.3 we have introduced in a rather ad–hoc fashion a generic $(N_{\mathrm{S}} \times N_{\mathrm{R}})$–dimensional multiple–input multiple–output (MIMO) channel model for the derivation of the detection schemes in Section 2.4 and Chapter 3. In this chapter, we now present the general MIMO channel model for a system with N_{T} transmit and N_{R} receive antennas, that serves as basis for the discrete–time channel models considered in Chapters 5 and 6. The connection between the $(N_{\mathrm{S}} \times N_{\mathrm{R}})$–dimensional and the $(N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional channel models of Section 2.3 and this section, respectively, will be established separately for the two transmission scenarios considered in Chapters 5 and 6.

In general all of our considerations shall be set in the equivalent complex baseband (ECB) [Tre71], i.e. all quantities involved in the channel model are in general represented by complex–valued variables.

In the following, we consider the general case of transmission using $N_{\mathrm{T}} \geq 1$ transmit and $N_{\mathrm{R}} \geq 1$ receive antennas over time–variant spatially correlated MIMO channels perturbed by additive white Gaussian noise (AWGN). The classical single–input single–output (SISO) case, where both transmitter and receiver use only a single antenna is apparently included as special case for $N_{\mathrm{T}} = N_{\mathrm{R}} = 1$ and all results presented subsequently are therefore equally applicable here.

Fig. 4.1 shows the MIMO channel model under consideration. Information is mapped to transmit symbols $x_i[k]$, $1 \leq i \leq N_{\mathrm{T}}$, $k \in \mathbb{N}$, in a way that is to be specified later. After transmit pulse shaping with a $\sqrt{\text{Nyquist}}$ [Pro00] pulse $h_{\mathrm{T}}(t)$ the $x_i[k]$ are transmitted at a rate $1/T$, where T is referred to as modulation interval, in the form of the continuous–time signal

$$x_i(t) = \sum_{k'=0}^{\infty} x_i[k'] h_{\mathrm{T}}(t - k'T). \tag{4.1}$$

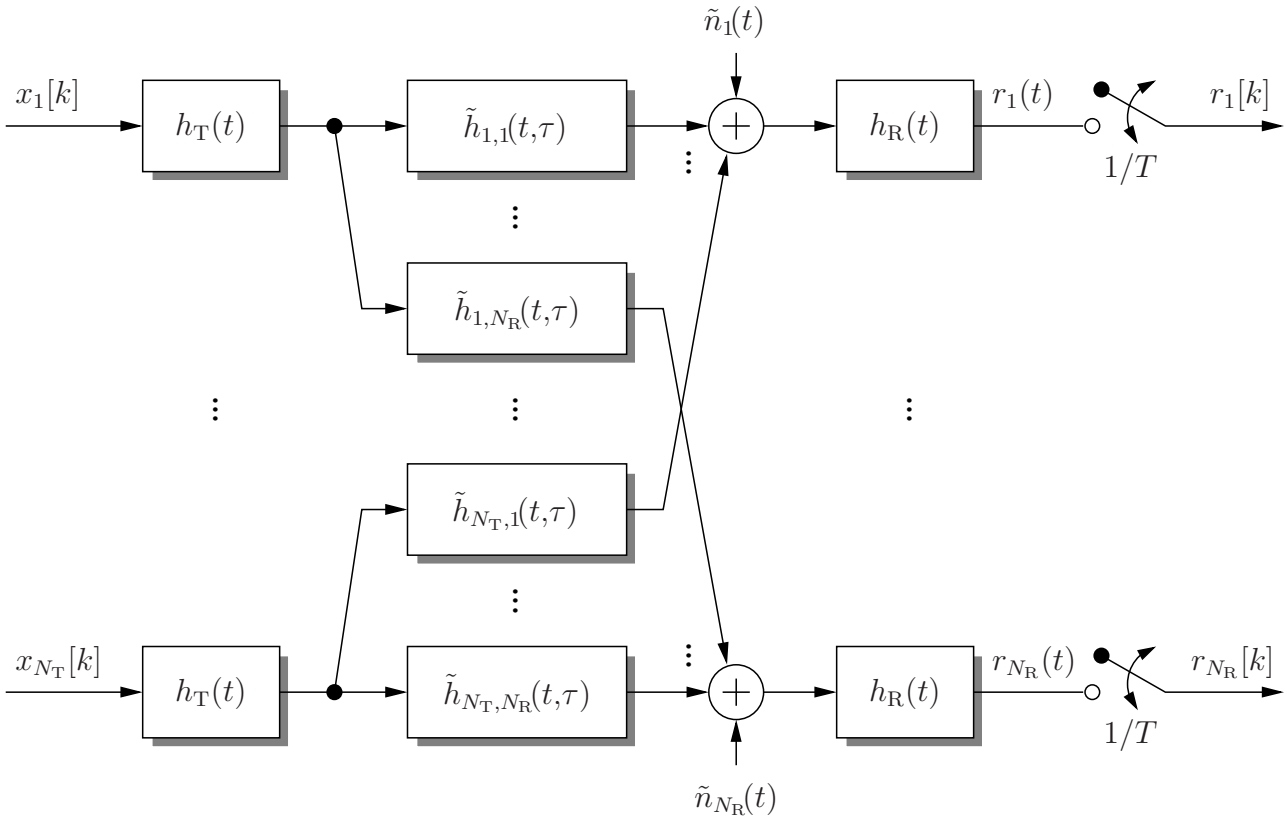

Figure 4.1: Multiple–input multiple–output channel model.

The received signal $r_j(t)$ at receive antenna j after matched filtering with

$$h_{\mathrm{R}} \triangleq c_{\mathrm{norm}} h_{\mathrm{T}}^*(-t), \tag{4.2}$$

reads

$$\begin{aligned} r_j(t) &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \sum_{i=1}^{N_{\mathrm{T}}} x_i(\eta-\tau) \tilde{h}_{i,j}(\eta,\tau)\,\mathrm{d}\tau + \tilde{n}_j(\eta) \right] h_{\mathrm{R}}(t-\eta)\,\mathrm{d}\eta \qquad (4.3) \\ &= \sum_{i=1}^{N_{\mathrm{T}}} \sum_{k'=0}^{\infty} x_i[k'] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{\mathrm{T}}(\eta-\tau-k'T) \tilde{h}_{i,j}(\eta,\tau) h_{\mathrm{R}}(t-\eta)\,\mathrm{d}\tau\,\mathrm{d}\eta \\ &\quad + \int_{-\infty}^{\infty} \tilde{n}_j(\eta) h_{\mathrm{R}}(t-\eta)\,\mathrm{d}\eta, \qquad (4.4) \end{aligned}$$

where $\tilde{h}_{i,j}(t,\tau)$ denotes the complex–valued time–variant input delay–spread function [Ste99, Hub97] of the channel between transmit antenna i and receive antenna j, $\tilde{n}_j(t)$ represents the circularly symmetric complex additive spatially and temporally white Gaussian noise with double–sided power spectral density (PSD) $\mathcal{N}_0$ effective at receive antenna j,[1] and c_{norm} is a

[1]Note that based on Grettenberg's theorem [Mil74] this corresponds to a zero–mean real–valued AWGN passband process with single–sided PSD $\mathcal{N}_0$.

normalization constant to be discussed below. For a list PSDs and corresponding temporal correlation functions of common fading models, see Appendix 4.A. We assume that $\tilde{h}_{i,j}(t,\tau)$ remains approximately constant for twice the period of time, where $h_{\mathrm{T}}(t)$ deviates significantly from zero. This assumption is justified if the normalized single–sided bandwidth B_hT of the fading process $\tilde{h}_{i,j}(t,\tau)$ satisfies $B_hT \leq 0.03 \ldots 0.05$, where the value of this upper bound depends on the roll–off factor of the transmit pulse, cf. [CH92]. With this, we can express the received signal as

$$r_j(t) = \sum_{i=1}^{N_{\mathrm{T}}} \sum_{k'=0}^{\infty} x_i[k'] \int_{-\infty}^{\infty} \tilde{h}_{i,j}(t,\tau) \int_{-\infty}^{\infty} h_{\mathrm{T}}(\eta - \tau - k'T) h_{\mathrm{R}}(t-\eta)\,\mathrm{d}\tau\,\mathrm{d}\eta + n_j(t), \tag{4.5}$$

where we introduced

$$n_j(t) \triangleq \tilde{n}_j(t) * h_{\mathrm{R}}(t) \tag{4.6}$$

($x(t) * y(t) \triangleq \int_{-\infty}^{\infty} x(\tau) y(t-\tau)\,\mathrm{d}\tau$: convolution).

In mobile communications a general frequency–selective channel is typically modeled using a sequence of L_h mutually independent time–variant random channel taps $h_{i,j}^{(l)}(t)$ at delays $\tau_{l,i,j}$, $1 \leq l \leq L_h$, $1 \leq i \leq N_{\mathrm{T}}$, $1 \leq j \leq N_{\mathrm{R}}$, [Ste87, Rap96b, Hub97]. For simplicity, we assume that all spatial subchannels have identical power delay profiles, i.e. the variances of $h_{i,j}^{(l)}(t)$ and delays $\tau_{l,i,j}$ are independent of i and j. Thus, we simply write τ_l for the delay of the lth channel tap, such that the channel is modeled by

$$\tilde{h}_{i,j}(t,\tau) = \sum_{l=1}^{L_h} h_{i,j}^{(l)}(t)\,\delta(\tau - \tau_l) \tag{4.7}$$

($\delta(x-y)$: Dirac–δ function). To describe the spatial correlation of the MIMO channel we use the frequently considered Kronecker correlation model (cf. e.g. [SFGK00, CRTP03]), i.e. we jointly model matrices of channel coefficients as

$$\boldsymbol{H}^{(l)}(t) \triangleq \left[h_{i,j}^{(l)}(t)\right]_{\substack{i=1,\ldots,N_{\mathrm{T}} \\ j=1,\ldots,N_{\mathrm{R}}}} = \sqrt{\boldsymbol{\Psi}^{\mathrm{Tx}}}\,\boldsymbol{W}^{(l)}(t)\sqrt{\boldsymbol{\Psi}^{\mathrm{Rx}}}^{\mathsf{T}}, \tag{4.8}$$

($\sqrt{\boldsymbol{X}}$: from $\boldsymbol{X} = \sqrt{\boldsymbol{X}}\sqrt{\boldsymbol{X}}^{\mathsf{H}}$)[2] with constant ($N_{\mathrm{T}} \times N_{\mathrm{T}}$)–dimensional and ($N_{\mathrm{R}} \times N_{\mathrm{R}}$)–dimensional matrices

$$\boldsymbol{\Psi}^{\mathrm{Tx}} \triangleq \left[\psi^{\mathrm{Tx}}[i,j]\right]_{\substack{i=1,\ldots,N_{\mathrm{T}} \\ j=1,\ldots,N_{\mathrm{T}}}} \tag{4.9}$$

$$\boldsymbol{\Psi}^{\mathrm{Rx}} \triangleq \left[\psi^{\mathrm{Rx}}[i,j]\right]_{\substack{i=1,\ldots,N_{\mathrm{R}} \\ j=1,\ldots,N_{\mathrm{R}}}} \tag{4.10}$$

[2]Note that this square–root decomposition is not uniquely defined. The Cholesky decomposition is one example of such a square–root matrix decomposition.

representing spatial correlation at transmitter and receiver side, respectively. In particular, $\psi^{\mathrm{Tx}}[i,j]$ represents the correlation between spatial subchannels from two different transmit antennas i and j to the same receive antenna. For $\psi^{\mathrm{Rx}}[i,j]$ a dual meaning exists. The L_h mutually independent $(N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional matrices $\boldsymbol{W}^{(l)}(t)$, $1 \leq l \leq L_h$, contain iid circularly symmetric complex Gaussian random processes with identical temporal correlation functions. Consequently, the L_h matrices $\boldsymbol{H}^{(l)}(t)$ contain $N_{\mathrm{T}}N_{\mathrm{R}}$ circularly symmetric complex Gaussian random processes and have the same spatial correlation structure. Based on the vector operator

$$\mathrm{vec}\left\{\boldsymbol{X}\right\} \triangleq \left[(\boldsymbol{X}_{:,1})^{\mathsf{H}}, (\boldsymbol{X}_{:,2})^{\mathsf{H}}, \ldots, (\boldsymbol{X}_{:,m})^{\mathsf{H}}\right]^{\mathsf{H}} \tag{4.11}$$

and the identity

$$\mathrm{vec}\left\{\boldsymbol{X}\boldsymbol{Y}\right\} = (\boldsymbol{I}_n \otimes \boldsymbol{X})\mathrm{vec}\left\{\boldsymbol{Y}\right\} \tag{4.12}$$

holding for arbitrary matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$, where m and n denote the numbers of columns of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, the spatial correlation can be described using the $(N_{\mathrm{T}}N_{\mathrm{R}} \times N_{\mathrm{T}}N_{\mathrm{R}})$–dimensional spatial correlation matrix

$$\boldsymbol{\Psi}_{\boldsymbol{HH}} \triangleq \frac{\mathcal{E}_t\left\{\mathrm{vec}\left\{\boldsymbol{H}^{(l)}(t)\right\}\mathrm{vec}\left\{\boldsymbol{H}^{(l)}(t)\right\}^{\mathsf{H}}\right\}}{\mathcal{E}_t\left\{\left|h_{i,j}^{(l)}(t)\right|^2\right\}} = \boldsymbol{\Psi}^{\mathrm{Rx}} \otimes \boldsymbol{\Psi}^{\mathrm{Tx}}, \tag{4.13}$$

with $\boldsymbol{\Psi}^{\mathrm{Rx}}$ and $\boldsymbol{\Psi}^{\mathrm{Tx}}$ as defined in (4.10) and (4.9), respectively. The normalized temporal correlation function that is common to all channel taps is denoted by

$$\psi^{\mathrm{t}}(\tau) \triangleq \frac{\mathcal{E}_t\left\{\left(h_{i,j}^{(l)}(t)\right)^* h_{i,j}^{(l)}(t+\tau)\right\}}{\mathcal{E}_t\left\{\left|h_{i,j}^{(l)}(t)\right|^2\right\}} = \frac{\mathcal{E}_t\left\{\left(w_{i,j}^{(l)}(t)\right)^* w_{i,j}^{(l)}(t+\tau)\right\}}{\mathcal{E}_t\left\{\left|w_{i,j}^{(l)}(t)\right|^2\right\}}. \tag{4.14}$$

Plugging (4.7) into (4.5) and with the definition of

$$h_{\mathrm{G}}(t) \triangleq h_{\mathrm{T}}(t) * h_{\mathrm{R}}(t), \tag{4.15}$$

which due to the fact that $h_{\mathrm{T}}(t)$ and $h_{\mathrm{R}}(t)$ are $\sqrt{\mathrm{Nyquist}}$ pulses is a Nyquist pulse, we obtain for the continuous–time received signal at antenna j

$$r_j(t) = \sum_{i=1}^{N_{\mathrm{T}}} \sum_{k'=0}^{\infty} x_i[k'] \sum_{l=1}^{L_h} h_{i,j}^{(l)}(t) h_{\mathrm{G}}(t - \tau_l - k'T) + n_j(t). \tag{4.16}$$

After sampling of $r_j(t)$ at time instants $t = kT + \tau_1$ for appropriate symbol synchronization the discrete–time received signal reads

$$\begin{aligned} r_j[k] &\triangleq r_j(kT + \tau_1) && (4.17)\\ &= \sum_{i=1}^{N_{\mathrm{T}}} \sum_{k'=0}^{\infty} x_i[k'] \sum_{l=1}^{L_h} h_{i,j}^{(l)}(kT + \tau_1) h_{\mathrm{G}}((k-k')T + \tau_1 - \tau_l) + n_j[k], && (4.18) \end{aligned}$$

where

$$n_j[k] \stackrel{\triangle}{=} n_j(kT + \tau_1). \tag{4.19}$$

Based on (i) the assumption that the continuous–time process $\tilde{n}_j(t)$ is circularly symmetric complex spatially and temporally white Gaussian distributed and (ii) the fact that the matched filter $h_{\mathrm{R}}(t)$ at the receiver is a $\sqrt{\text{Nyquist}}$ pulse, the noise process $n_j[k]$ is discrete–time circularly symmetric complex spatially and temporally white Gaussian distributed.

In accordance with the organization of the present work we distinguish between two channel models, namely frequency non–selective and frequency selective Rayleigh fading channels if $L_h = 1$ and $L_h > 1$, respectively.[3] These are discussed separately at the beginning of Chapters 5 and 6, respectively.

[3]Throughout this work, the attributes "frequency selective" and "dispersive" (channel), "frequency non–selective", "flat fading" and "non–dispersive" (channel), and "time varying", "time selective" and "fading" (channel) are used interchangeably.

4.A Fading Models

In this appendix, we list the power spectral densities (PSD) $\Psi_{\mathrm{c}}^{\mathrm{t}}(f)$ and temporal correlation functions $\psi^{\mathrm{t}}(t)$ of a number of fading models considered in this work. B_h denotes the fading bandwidth also referred to as Doppler spread.

Rectangular model:

$$\Psi_{\mathrm{c}}^{\mathrm{t}}(f) = \begin{cases} \dfrac{1}{2B_h}, & |f| \leq B_h \\ 0, & B_h < |f| \end{cases} \tag{4.20}$$

$$\psi^{\mathrm{t}}(t) = \frac{\sin(2\pi B_h t)}{2\pi B_h t} \tag{4.21}$$

Clarke's model (land–mobile model):

$$\Psi_{\mathrm{c}}^{\mathrm{t}}(f) = \begin{cases} \dfrac{1}{\pi\sqrt{B_h^2 - f^2}}, & |f| \leq B_h \\ 0, & B_h < |f| \end{cases} \tag{4.22}$$

$$\psi^{\mathrm{t}}(t) = \mathrm{J}_0(2\pi B_h t) \tag{4.23}$$

Gaussian PSD model:

$$\Psi_{\mathrm{c}}^{\mathrm{t}}(f) = \frac{1}{\sqrt{\pi}B_h}\exp\left(-\left(\frac{f}{B_h}\right)^2\right) \tag{4.24}$$

$$\psi^{\mathrm{t}}(t) = \exp\left(-(\pi B_h t)^2\right) \tag{4.25}$$

First–order Butterworth PSD model:

$$\Psi_{\mathrm{c}}^{\mathrm{t}}(f) = \frac{1}{\pi B_h \left(1 + \frac{f}{B_h}\right)^2} \tag{4.26}$$

$$\psi^{\mathrm{t}}(t) = \exp(-2\pi\,|B_h t|) \tag{4.27}$$

Second–order Butterworth PSD model:

$$\Psi_{\mathrm{c}}^{\mathrm{t}}(f) = \frac{2\sqrt{2}}{\pi B_h \left(1 + \left(2\frac{f}{B_h}\right)^4\right)} \tag{4.28}$$

$$\psi^{\mathrm{t}}(t) = \exp\left(-\frac{\pi\,|B_h t|}{\sqrt{2}}\right) \cdot \left(\cos\left(\frac{\pi B_h t}{\sqrt{2}}\right) + \sin\left(\frac{\pi\,|B_h t|}{\sqrt{2}}\right)\right) \tag{4.29}$$

Table 4.1: Power spectral densities (PSD) $\Psi_{\mathrm{c}}^{\mathrm{t}}(f)$ and temporal correlation functions $\psi^{\mathrm{t}}(t)$ for various fading models [SA05].

Chapter 5

Multiple–Symbol Differential Detection for Frequency–Nonselective Channels

In narrowband digital communications the transmission channel can often be modeled as frequency nonselective. For this scenario the unitary–matrix signal constellations reviewed in Section 2.1 have —in conjunction with differential encoding (cf. Section 2.2)— been developed as a means for power–efficient transmission without the need for accurate channel state information (CSI) at the receiver. As the unitary–matrix symbols extend over time and space this transmission scheme whose details are given below is commonly referred to as (unitary) differential space–time modulation (U)DSTM.

In this chapter, we apply the methods of Chapter 3 to the scenario of DSTM–based transmission over frequency–nonselective channels. In Sections 5.1 and 5.2 we present the system model, whose connection with the generic system model of Chapter 2 will be established in Section 5.3.

Section 5.4 provides an in–depth analysis of the performance of the various detection schemes considered in this work under spatially uncorrelated and correlated frequency–nonselective fading. To analyze the case of spatially correlated channels we resort to standard methods based on characteristic functions. For the spatially uncorrelated case we derive expressions for the error rates of the various detection schemes that facilitate a number of interesting and intuitive insights, such as interdependencies between the various detection schemes and dependence of their performance on system parameters such as SNR, fading bandwidth, numbers of antennas at both ends of the channel and observation window length.

Section 5.5 is concerned with a complexity analysis of the detection schemes based on tree–search decoding and combinatorial geometry considered in Sections 3.1 and 3.2, respectively. While we will prove that for standard fading models the complexity of both approaches to MSDD is in general exponential in the length N of the observation window, we will show that that tree–search methods are more suitable for fast fading scenarios, whereas the approach

based on combinatorial geometry is preferably used in slower fading scenarios.

5.1 Differential Space–Time Modulation (DSTM)

DSTM for noncoherent transmission over frequency–nonselective MIMO fading channels assumes

$$N_{\mathrm{T}} = N_{\mathrm{S}} \tag{5.1}$$

transmit antennas. Here, groups of $N_{\mathrm{T}}R$ bits are mapped to $(N_{\mathrm{T}} \times N_{\mathrm{T}})$–dimensional signal matrices $\boldsymbol{V}[k]$ taken from a set $\boldsymbol{\mathcal{V}}$ of

$$L = 2^{N_{\mathrm{T}}R} \tag{5.2}$$

unitary matrices, which are differentially encoded according to (2.11) to yield $(N_{\mathrm{T}} \times N_{\mathrm{T}})$–dimensional unitary transmit matrices $\boldsymbol{S}[k]$. These matrices are transmitted in a row–by–row fashion over the $(N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional MIMO channel in N_{T} modulation intervals. Consequently, R denotes the information transmission rate in bit per MIMO channel use, i.e. per modulation interval not per DSTM symbol.

5.2 Time–Selective Channel Model

In this section, we derive —building upon Chapter 4— the time–selective and frequency–nonselective discrete–time model of the communication system deploying N_{T} transmit and N_{R} receive antennas considered in this chapter.

The $(N_{\mathrm{T}} \times N_{\mathrm{T}})$–dimensional unitary transmit matrices $\boldsymbol{S}[k]$ are transmitted in a row–by–row fashion in N_{T} modulation intervals over the $(N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional MIMO channel (cf. Chapter 4) such that $s_{\nu,i}[k]$ is transmitted from antenna i at discrete time $(kN_{\mathrm{T}} + \nu)T$, i.e. (cf. Fig. 4.1)

$$x_i[kN_{\mathrm{T}} + \nu] = s_{\nu,i}[k]. \tag{5.3}$$

Since the channel is frequency–nonselective, we have from (4.7) that

$$\tilde{h}_{i,j}(t,\tau) = h_{i,j}^{(1)}(t)\delta(\tau - \tau_1)\,. \tag{5.4}$$

Plugging this into (4.17) and recalling that $h_{\mathrm{G}}(t)$ is a Nyquist pulse, the discrete–time intersymbol–interference–free received signal $r_j[kN_{\mathrm{T}} + \nu]$ at antenna j sampled at time $(kN_{\mathrm{T}} + \nu)T + \tau_1$ is given by

$$r_j[kN_{\mathrm{T}} + \nu] = \sum_{i=1}^{N_{\mathrm{T}}} s_{\nu,i}[k]h_{i,j}[kN_{\mathrm{T}} + \nu] + n_j[kN_{\mathrm{T}} + \nu], \tag{5.5}$$

where

$$h_{i,j}[k] \triangleq h_{i,j}^{(1)}(kT + \tau_1) h_{\mathrm{G}}(0) \tag{5.6}$$

and $n_j[k]$ as defined in (4.19).

In order to obtain dimensionless discrete–time signals, the normalization constant c_{norm} of the receiver's matched filter $h_{\mathrm{R}}(t)$ [cf. (4.2)] is chosen such that the variance of the noise–free component of the received signal $r_j[k]$ is one. Thus, the variance of the discrete–time AWGN $n_j[k]$ is given by $\sigma_n^2 \triangleq \mathcal{E}\{|n_j[k]|^2\} = \mathcal{N}_0/E_{\mathrm{s}} = \mathcal{N}_0/(E_{\mathrm{b}}R)$ (E_{s}: average received energy per MIMO–channel use, E_{b}: average received energy per transmitted bit, R: transmission rate in bit per MIMO–channel use), i.e. the signal–to–noise ratio (SNR) will subsequently be quantified by

$$\frac{E_{\mathrm{b}}}{\mathcal{N}_0} = \frac{E_{\mathrm{s}}}{\mathcal{N}_0 R} = \frac{1}{\sigma_n^2 R}. \tag{5.7}$$

Based on the spatial correlation model (4.8) the $(N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional MIMO–channel matrix $\boldsymbol{H}_{\mathrm{c}}[k]$ at discrete time kT can be modeled as

$$\boldsymbol{H}_{\mathrm{c}}[k] \triangleq [h_{i,j}[k]]_{\substack{i=1,\ldots,N_{\mathrm{T}} \\ \mathrm{j}=1,\ldots,N_{\mathrm{R}}}} = \sqrt{\boldsymbol{\Psi}^{\mathrm{Tx}}}\boldsymbol{W}[k]\sqrt{\boldsymbol{\Psi}^{\mathrm{Rx}}}^{\mathsf{T}}, \tag{5.8}$$

where constant $(N_{\mathrm{T}} \times N_{\mathrm{T}})$–dimensional $\boldsymbol{\Psi}^{\mathrm{Tx}}$ and $(N_{\mathrm{R}} \times N_{\mathrm{R}})$–dimensional $\boldsymbol{\Psi}^{\mathrm{Rx}}$ as defined in (4.9) and (4.10) model the correlation at transmitter and receiver side, respectively. Due to the normalization of the received signal, the $(N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional matrix $\boldsymbol{W}[k]$ contains iid $\mathcal{N}_{\mathrm{c}}(0,1)$ distributed random variables with identical temporal correlation functions [cf. (4.14)]

$$\psi^{\mathrm{t}}[\kappa] \triangleq \mathcal{E}\big\{h_{i,j}[k+\kappa]h_{i,j}^*[k]\big\} = \mathcal{E}\big\{w_{i,j}[k+\kappa]w_{i,j}^*[k]\big\} = \psi^{\mathrm{t}}(\kappa T). \tag{5.9}$$

It is shown in Appendix 5.A.1 that the autocorrelation function of $h_{i,j}[\kappa]$ is separable into a temporal and two spatial components as

$$\psi_{hh}[\kappa, i_1, i_2, j_1, j_2] \triangleq \mathcal{E}\big\{h_{i_1,j_1}[k+\kappa]h_{i_2,j_2}^*[k]\big\} = \psi^{\mathrm{t}}[\kappa]\psi^{\mathrm{Tx}}[i_1, i_2]\psi^{\mathrm{Rx}}[j_1, j_2], \tag{5.10}$$

where $\psi^{\mathrm{t}}[\kappa]$ as introduced in (5.9) specifies the temporal correlation of the $N_{\mathrm{T}}N_{\mathrm{R}}$ individual fading processes, and $\psi^{\mathrm{Tx}}[i,j]$ and $\psi^{\mathrm{Rx}}[i,j]$ are the elements in the ith row and jth column of $\boldsymbol{\Psi}^{\mathrm{Tx}}$ and $\boldsymbol{\Psi}^{\mathrm{Rx}}$ [cf. (4.9) and (4.10)] representing spatial correlation at transmitter and receiver side, respectively. Note, that we did not make any simplifying assumptions on the geometry of the transmit and receive antenna arrays. This leads to the general form of the autocorrelation function with pairs of indeces i_1, i_2 and j_1, j_2 to model the correlation between the spatial subchannel from transmit antenna i_1 to receive antenna j_1 and the spatial subchannel from transmit antenna i_2 to receive antenna j_2.

5.2.1 Continuous–Fading Channel Model

Recall that the MIMO fading channel changes "continuously" over time, i.e. the channel coefficients $h_{i,j}[k]$ change from modulation interval to modulation interval. In order to describe the transmission of one DSTM symbol $\boldsymbol{S}[k]$ we therefore proceed as follows. Let $\boldsymbol{R}[k]$, $\boldsymbol{H}_{\mathrm{c}}[kN_{\mathrm{T}}+\nu]$ and $\boldsymbol{N}[k]$ each contain $r_j[kN_{\mathrm{T}}+i]$, $h_{i,j}[kN_{\mathrm{T}}+\nu]$ and $n_j[kN_{\mathrm{T}}+i]$ in the ith row and jth column, $1 \leq i \leq N_{\mathrm{T}}$, $1 \leq j \leq N_{\mathrm{R}}$. Since the νth row of $\boldsymbol{S}[k]$ is transmitted over the $(N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional MIMO channel at discrete time $(kN_{\mathrm{T}}+\nu)T$, i.e. over $\boldsymbol{H}_{\mathrm{c}}[kN_{\mathrm{T}}+\nu]$, we obtain the $(N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional received matrix $\boldsymbol{R}[k]$ corresponding to the transmission of one space–time symbol $\boldsymbol{S}[k]$ as [cf. (5.5)]

$$\boldsymbol{R}[k] = \boldsymbol{S}_{\mathrm{c}}[k]\breve{\boldsymbol{H}}_{\mathrm{c}}[k] + \boldsymbol{N}[k] \,, \tag{5.11}$$

where we defined $(N_{\mathrm{T}}^2 \times N_{\mathrm{R}})$–dimensional

$$\breve{\boldsymbol{H}}_{\mathrm{c}}[k] \stackrel{\triangle}{=} \left[\boldsymbol{H}_{\mathrm{c}}^{\mathsf{H}}[kN_{\mathrm{T}}+1], \ldots, \boldsymbol{H}_{\mathrm{c}}^{\mathsf{H}}[(k+1)N_{\mathrm{T}}]\right]^{\mathsf{H}} \tag{5.12}$$

and $(N_{\mathrm{T}} \times N_{\mathrm{T}}^2)$–dimensional block–diagonal matrix

$$\boldsymbol{S}_{\mathrm{c}}[k] \stackrel{\triangle}{=} \begin{bmatrix} \boldsymbol{S}_{1,:}[k] & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{S}_{N_{\mathrm{T}},:}[k] \end{bmatrix} \tag{5.13}$$

($\boldsymbol{X}_{i,:}$: ith row of a matrix $\boldsymbol{X}$).

Furthermore, the $(NN_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional block matrix $\bar{\boldsymbol{R}}[k]$ of N consecutively received matrix symbols, that are processed simultaneously in MSDD and DFDD, is easily obtained as

$$\bar{\boldsymbol{R}}[k] \stackrel{\triangle}{=} \left[\boldsymbol{R}^{\mathsf{H}}[k-N+1], \ldots, \boldsymbol{R}^{\mathsf{H}}[k]\right]^{\mathsf{H}} \tag{5.14}$$

$$= \bar{\boldsymbol{S}}_{\mathrm{D,c}}[k]\bar{\boldsymbol{H}}_{\mathrm{c}}[k] + \bar{\boldsymbol{N}}[k], \tag{5.15}$$

with the $(NN_{\mathrm{T}} \times NN_{\mathrm{T}}^2)$–dimensional block–diagonal matrix of transmit signals

$$\bar{\boldsymbol{S}}_{\mathrm{D,c}}[k] \stackrel{\triangle}{=} \begin{bmatrix} \boldsymbol{S}_{\mathrm{c}}[k-N+1] & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{S}_{\mathrm{c}}[k] \end{bmatrix}, \tag{5.16}$$

the $(NN_{\mathrm{T}}^2 \times N_{\mathrm{R}})$–dimensional channel matrix

$$\bar{\boldsymbol{H}}_{\mathrm{c}}[k] \stackrel{\triangle}{=} \left[\breve{\boldsymbol{H}}_{\mathrm{c}}^{\mathsf{H}}[k-N+1], \ldots, \breve{\boldsymbol{H}}_{\mathrm{c}}^{\mathsf{H}}[k]\right]^{\mathsf{H}} \tag{5.17}$$

and the $(NN_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional AWGN matrix

$$\bar{\boldsymbol{N}}[k] \stackrel{\triangle}{=} \left[\boldsymbol{N}^{\mathsf{H}}[k-N+1], \ldots, \boldsymbol{N}^{\mathsf{H}}[k]\right]^{\mathsf{H}}. \tag{5.18}$$

In order to distinguish this fading model from other fading models considered below, we refer to this model as *continuous–fading channel* (CFC) model, an expression used in [HM00].

5.2.2 Quasi–Static–Fading Channel Model

While the continuous–fading channel model will be used for analytical purposes and performance evaluation in later sections, the derivation of low–complexity DSTM detectors requires the assumption of a *quasi–static fading channel* (QSFC), i.e. the channel is assumed to be constant during N_{T} consecutive modulation intervals (cf. e.g. [BS02, CT05b]).[1] Defining the $(N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional QSFC matrix

$$\boldsymbol{H}[k] \stackrel{\triangle}{=} [h_{i,j}[kN_{\mathrm{T}}]]_{\substack{i=1,\ldots,N_{\mathrm{T}} \\ j=1,\ldots,N_{\mathrm{R}}}}, \tag{5.19}$$

transmission over the QSFC is described by

$$\boldsymbol{R}[k] = \boldsymbol{S}[k]\boldsymbol{H}[k] + \boldsymbol{N}[k]. \tag{5.20}$$

With the $(NN_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional QSFC matrix

$$\bar{\boldsymbol{H}}[k] \stackrel{\triangle}{=} \left[\boldsymbol{H}^{\mathsf{H}}[k-N+1], \ldots, \boldsymbol{H}^{\mathsf{H}}[k]\right]^{\mathsf{H}} \tag{5.21}$$

$(NN_{\mathrm{T}} \times NN_{\mathrm{T}})$–dimensional block–diagonal matrix [cf. (2.23)]

$$\bar{\boldsymbol{S}}_{\mathrm{D}}[k] \stackrel{\triangle}{=} \begin{bmatrix} \boldsymbol{S}[k-N+1] & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{S}[k] \end{bmatrix}, \tag{5.22}$$

of N transmit symbols and $\bar{\boldsymbol{N}}[k]$ as defined in (5.18), the $(NN_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional block matrix $\bar{\boldsymbol{R}}[k] \stackrel{\triangle}{=} \left[\boldsymbol{R}^{\mathsf{H}}[k-N+1], \ldots, \boldsymbol{R}^{\mathsf{H}}[k]\right]^{\mathsf{H}}$ of N consecutively received matrix symbols as introduced in Section 2.4.2 is given by

$$\bar{\boldsymbol{R}}[k] = \bar{\boldsymbol{S}}_{\mathrm{D}}[k]\bar{\boldsymbol{H}}[k] + \bar{\boldsymbol{N}}[k]. \tag{5.23}$$

The assumption of the channel being approximately constant for N_{T} consecutive modulation intervals is justified for $B_h T N_{\mathrm{T}} \leq 0.01$ only. In faster fading scenarios the noncoherent detection schemes operating based on (5.23) suffer a performance loss due to a metric mismatch, cf. also Section 5.4.5. However, it is important to note that for cyclic and dicyclic DSTM, where every spatial subchannel is used exactly once per DSTM symbol, the CFC and QSFC models coincide and thus, the QSFC model accurately describes transmission over continuously fading channels when cyclic and dicyclic constellations are employed for DSTM. This reflects in the fact that all but N_{T} out of the N_{T}^2 rows of $\boldsymbol{S}_{\mathrm{c}}[k]$ [cf. (5.13)] are exclusively filled with zeros. If one

[1] The reason for this is that while the PDF $p\big(\bar{\boldsymbol{R}} \,\big|\, \bar{\boldsymbol{S}}\big)$ can be given in a form similar to (2.25) the ML–MSDD decision rule can not be brought into the form of (2.36) since a simplification as in (2.32)–(2.34) is not feasible for the CFC.

removes the corresponding rows from $(N_{\mathrm{T}}^2 \times N_{\mathrm{R}})$–dimensional $\boldsymbol{H}_{\mathrm{c}}[k]$ [cf. (5.12)] an $(N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional channel matrix $\boldsymbol{H}[k]$ results. In this case, the QSFC model is —again based on the arguments of [Cav92]— valid for $B_h T \leq 0.03 \ldots 0.05$ and using this model as basis for MSDD or DFDD does not incur a metric mismatch and thereby a performance degradation if $B_h T N_{\mathrm{T}} > 0.01$.

Note however, that different from cyclic DSTM, for dicyclic DSTM the time between two consecutive transmissions from the same antenna may vary depending on the parameter m [cf. (2.4)]. Consequently, optimal MSDD would have to adapt the prediction coefficients according to the parameters m associated with tentative decisions $\tilde{\boldsymbol{V}}_j$, $n+1 \leq j \leq N-1$, and the current candidate $\tilde{\boldsymbol{V}}_n$, which would complicate the detection process significantly. Alternatively, prediction coefficients can be based on an average time between consecutive transmissions from the same antenna, which is equivalent to using the QSFC model.

We would also like to point out that the QSFC model considered in (5.20) is different from the *block–fading channel* (BFC) model (cf. also [Hug00a, HS00, TJ00, HH02, HMR+00]) which stipulates that the channel remains constant during the transmission of N consecutive DSTM symbols and changes from block to block.

5.3 Application of Efficient MSDD Algorithms

Note that (5.23) is exactly of the form of (2.22). This means that we can directly apply the receiver algorithms of Chapter 3 based on the generic system model of Chapter 2. At this $\bar{\boldsymbol{H}}[k]$ [cf. (5.21)] now plays the role of $\bar{\boldsymbol{G}}[k]$ [cf. (2.20)]. With the help of Appendix 5.A.1, where an expression for $\mathcal{E}\{h^*_{i_1,j_1}[k] h_{i_2,j_2}[k+\kappa]\}$ is derived, it is fairly straightforward to see that the $(NN_{\mathrm{T}} \times NN_{\mathrm{T}})$–dimensional correlation matrix of $\bar{\boldsymbol{H}}[k]$ is given by

$$\boldsymbol{\Psi}_{\bar{\boldsymbol{H}}\bar{\boldsymbol{H}}} \stackrel{\triangle}{=} \mathcal{E}\Big\{\bar{\boldsymbol{H}}\bar{\boldsymbol{H}}^{\mathsf{H}}\Big\} = \boldsymbol{\Psi}^{\mathrm{t}} \otimes \boldsymbol{\Psi}^{\mathrm{Tx}} N_{\mathrm{R}}, \tag{5.24}$$

with $\boldsymbol{\Psi}^{\mathrm{Tx}}$ as defined in (4.9) and $(N \times N)$–dimensional $\boldsymbol{\Psi}^{\mathrm{t}}$ modeling temporal correlation.[2]

For computational efficiency of the receivers we only consider algorithms that operate based on the assumption of absence of spatial correlation. In this case $\boldsymbol{\Psi}_{\bar{\boldsymbol{H}}\bar{\boldsymbol{H}}}$ reduces to $\boldsymbol{\Psi}^{\mathrm{t}} \otimes \boldsymbol{I}_{N_{\mathrm{T}}} N_{\mathrm{R}}$ and the $(N \times N)$–dimensional matrix $\boldsymbol{M}$ that appears in the ML–MSDD metric [cf. (2.36)] can be identified as

$$\boldsymbol{M} = \left(\boldsymbol{\Psi}^{\mathrm{t}} + \sigma_n^2 \boldsymbol{I}_N\right)^{-1}, \tag{5.25}$$

thus allowing for application of the algorithms described in Chapter 3.

[2]Note that spatial correlation at the receiver side does not have an impact on this expression, because in $\bar{\boldsymbol{H}}\bar{\boldsymbol{H}}^{\mathsf{H}}$ only channel coefficients corresponding to the same receive antenna are multiplied with each other.

5.3.1 Spatial Correlation

Note that in the presence of spatial correlation ML MSDD using tree–search decoding is still feasible. However, the $c_{i,j}$ in (3.14) obtained from the Cholesky decomposition of $\boldsymbol{M}$ would have to be replaced with $(N_{\mathrm{T}} \times N_{\mathrm{T}})$–dimensional matrices $\boldsymbol{C}_{i,j}$ computed via the Cholesky decomposition $\left(\boldsymbol{\Psi}^{\mathrm{t}} \otimes \boldsymbol{\Psi}^{\mathrm{Tx}} + \sigma_n^2 \boldsymbol{I}_{NN_{\mathrm{T}}}\right)^{-1} = \tilde{\boldsymbol{C}}^{\mathsf{H}} \tilde{\boldsymbol{C}}$, where upper–right triangular $\tilde{\boldsymbol{C}} = [\boldsymbol{C}_{i,j}]_{\substack{i=1,\ldots,N \\ j=1,\ldots,N}}$, to describe correlation at the transmitter side. As matrices are in general not commutative under regular matrix multiplication, the introduction of $\breve{\boldsymbol{R}}_{i,j}$, $1 \leq i,j \leq N$, [cf. (3.17)] which needs to be computed only once at the beginning of the decoding process, would not be possible. In addition, the efficient symbol–search strategies as discussed in Section 3.1.4 would not be feasible, such that decoder complexity would be increased significantly. We shall therefore not pursue this further in this work, and only present some numerical results in Section 5.4 that illustrate the minor performance loss due to spatial correlation of the proposed detection algorithms compared to detectors that take spatial correlation into account.

5.3.2 Effective Fading Process

Note that the use of regular DSTM as introduced in Section 5.1 increases the normalized fading bandwidth relevant to the noncoherent detector by a factor of N_{T}, cf. [SL02, DB06]. This is due to the fact that transmissions of symbols $s_{i,j}[k]$ and $s_{i,j}[k+1]$ are separated by N_{T} modulation intervals. While this temporal "channel decorrelation" impedes power–efficient detection it can be avoided through application of *blockwise* DSTM (BDSTM) [DB06] based on cyclic DSTM. Here the elements $s_{i,i}[k]$ are blockwise interleaved prior to transmission, such that $s_{i,i}[k]$ and $s_{i,i}[k+1]$ are transmitted in successive modulation intervals. More details on BDSTM can be found in Appendix 5.C.

In order to avoid repetitive distinction between DSTM and BDSTM, we will subsequently consider an *effective* discrete–time fading process with temporal autocorrelation matrix $\boldsymbol{\Psi}^{\mathrm{t}}$

$$\boldsymbol{\Psi}^{\mathrm{t}} = \mathrm{toeplitz}\left\{\psi^{\mathrm{t}}[0], \psi^{\mathrm{t}}[x], \ldots, \psi^{\mathrm{t}}[x(N-1)]\right\}, \quad \begin{cases} x = N_{\mathrm{T}}, & \text{DSTM} \\ x = 1, & \text{BDSTM} \end{cases} \tag{5.26}$$

with $\psi^{\mathrm{t}}[\kappa]$ as introduced in (5.9). Accordingly, its PSD is given by

$$\Psi^{\mathrm{t}}(f) \triangleq \sum_{\kappa=-\infty}^{\infty} \psi^{\mathrm{t}}[\kappa x] \mathrm{e}^{-\mathrm{j}2\pi\kappa f T} = \frac{1}{T} \sum_{k=-\infty}^{\infty} \Psi_{\mathrm{c}}^{\mathrm{t}}(f/x - k/T), \quad \begin{cases} x = N_{\mathrm{T}}, & \text{DSTM} \\ x = 1, & \text{BDSTM} \end{cases}, \tag{5.27}$$

where $\Psi_{\mathrm{c}}^{\mathrm{t}}(f)$ denotes the PSD of the continuous–time fading process with bandwidth B_h underlying the discrete–time fading process $h_{i,j}[k]$ [cf. (5.6)], and an *effective* normalized fading bandwidth

$$B_{h,\mathrm{eff}} T \triangleq x B_h T, \quad \begin{cases} x = N_{\mathrm{T}}, & \text{DSTM} \\ x = 1, & \text{BDSTM} \end{cases} \tag{5.28}$$

can be introduced.

5.4 Performance Analysis

In this section, we provide tight symbol–error rate (SER) approximations for the different detection schemes, which are computed using truncated union bounds over the pairwise error probabilities (PEP) of the dominant error events. While a cast–iron proof is yet to be found it is generally accepted, cf. e.g. [HF92, LLK04, PL07b], that single–transmit–symbol errors, which have been shown in [DS90] to dominate the error rate of MSDD when transmitting over the AWGN channel, are also the most probable error events when transmitting over fading channels (cf. Section 5.4.4 for more details). Therefore, we will in the performance analysis of MSDD restrict our attention to candidates of the form

$$\hat{\tilde{\boldsymbol{S}}}\big(l,\hat{l}\big) = \big[\boldsymbol{S}_1^{\mathsf{H}},\ldots,\boldsymbol{S}_{n-1}^{\mathsf{H}},\boldsymbol{S}_n^{\mathsf{H}}\cdot\big(\boldsymbol{V}^{(l)}\big)^{\mathsf{H}}\cdot\boldsymbol{V}^{(\hat{l})},\boldsymbol{S}_{n+1}^{\mathsf{H}},\ldots,\boldsymbol{S}_N^{\mathsf{H}}\big]^{\mathsf{H}}, \tag{5.29}$$

with $\boldsymbol{V}^{(l)}$, $\boldsymbol{V}^{(\hat{l})} \in \boldsymbol{\mathcal{V}}$ and $\boldsymbol{V}^{(\hat{l})} \neq \boldsymbol{V}^{(l)}$.

In the following analysis, we distinguish three different cases. (i) In cases where the QSFC model is inaccurate, we compute the PEPs based on the CFC model, using standard methods for the analysis of quadratic forms. This is done in Section 5.4.1. (ii) In Section 5.4.2 we present a novel analysis for the QSFC model which represents an accurate image of reality for slow and moderately fast fading ($B_{h,\mathrm{eff}}T \leq 0.01$) when using non–(di–)cyclic DSTM codes and also for fast fading ($B_{h,\mathrm{eff}}T \leq 0.03N_{\mathrm{T}}\ldots0.05N_{\mathrm{T}}$) when (di–)cyclic DSTM codes are considered, cf. discussion of the validity of the QSFC model at the end of Section 5.2.2. This analysis provides us with many novel and interesting theoretical insights, e.g. regarding asymptotic performance for high SNR and / or large N and the comparison between the different detection schemes based on the definition of an effective SNR. (iii) The PEPs for suboptimal MSDD based on combinatorial geometry, if the autocorrelation matrix $\boldsymbol{\Psi}^{\mathrm{t}}$ of the fading process is approximated by a matrix of smaller rank to reduce decoder complexity, are computed in Section 5.4.3.

The SER approximations for the various detection schemes considered in this work will then be presented in Section 5.4.4, followed by numerical examples to illustrate our analytical findings in Section 5.4.5.

We note, that the PEPs for DF–MSDD and S–MSDD are the same as those for regular ML MSDD. Separate expressions for the approximation of their SERs will be given in Section 5.4.4.

5.4.1 Pairwise Error Probabilities Based on the CFC Model

To analyze the PEPs for the general case of a spatially correlated continuous–fading channel, we apply standard methods based on the characteristic function of Hermitian quadratic forms, cf. e.g. [Bar87, HF92, BCTV98, BV01, SFG02].

5.4.1.1 (DF–)MSDD

To compute the PEPs based on the CFC, it is convenient to first vectorize the channel model (5.15) as

$$\bar{\boldsymbol{r}} \triangleq \operatorname{vec}\left\{\bar{\boldsymbol{R}}\right\} \tag{5.30}$$

$$= \operatorname{vec}\left\{\bar{\boldsymbol{S}}_{\mathrm{D,c}}\bar{\boldsymbol{H}}_{\mathrm{c}} + \bar{\boldsymbol{N}}\right\}. \tag{5.31}$$

Using the identity (4.12) we obtain

$$= \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \bar{\boldsymbol{S}}_{\mathrm{D,c}}\right)\bar{\boldsymbol{h}}_{\mathrm{c}} + \bar{\boldsymbol{n}}, \tag{5.32}$$

where

$$\bar{\boldsymbol{h}}_{\mathrm{c}} \triangleq \operatorname{vec}\left\{\bar{\boldsymbol{H}}_{\mathrm{c}}\right\} \tag{5.33}$$

$$\bar{\boldsymbol{n}} \triangleq \operatorname{vec}\left\{\bar{\boldsymbol{N}}\right\}. \tag{5.34}$$

With $\bar{\boldsymbol{S}}_{\mathrm{D,c}}\bar{\boldsymbol{S}}_{\mathrm{D,c}}^{\mathsf{H}} = \boldsymbol{I}_{NN_{\mathrm{T}}}$ it is fairly straightforward to show that the conditioned correlation matrix $\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}|\bar{\boldsymbol{S}}}$ of the received vector $\bar{\boldsymbol{r}}$ is given by

$$\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}|\bar{\boldsymbol{S}}} \triangleq \mathcal{E}\left\{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}^{\mathsf{H}} \mid \bar{\boldsymbol{S}}\right\} \tag{5.35}$$

$$= \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \bar{\boldsymbol{S}}_{\mathrm{D,c}}\right)\boldsymbol{T}_{\mathrm{c}}\left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \bar{\boldsymbol{S}}_{\mathrm{D,c}}^{\mathsf{H}}\right) \tag{5.36}$$

with the $(NN_{\mathrm{T}}^2N_{\mathrm{R}} \times NN_{\mathrm{T}}^2N_{\mathrm{R}})$–dimensional channel–plus–noise autocorrelation matrix

$$\boldsymbol{T}_{\mathrm{c}} \triangleq \boldsymbol{\Psi}^{\mathrm{Rx}} \otimes \boldsymbol{\Psi}_{\mathrm{c}}^{\mathrm{t}} \otimes \boldsymbol{\Psi}^{\mathrm{Tx}} + \sigma_n^2\boldsymbol{I}_{NN_{\mathrm{T}}^2N_{\mathrm{R}}}, \tag{5.37}$$

with $\boldsymbol{\Psi}^{\mathrm{Tx}}$ and $\boldsymbol{\Psi}^{\mathrm{Rx}}$ [cf. (4.9) and (4.10)] describing spatial correlation at transmitter and receiver, respectively, and the $(NN_{\mathrm{T}} \times NN_{\mathrm{T}})$–dimensional temporal fading autocorrelation matrix

$$\boldsymbol{\Psi}_{\mathrm{c}}^{\mathrm{t}} \triangleq \operatorname{toeplitz}\left\{\psi^{\mathrm{t}}[0], \psi^{\mathrm{t}}[1], \ldots, \psi^{\mathrm{t}}[NN_{\mathrm{T}}-1]\right\}.^{3} \tag{5.38}$$

Similarly, it is easy to see that the metric of the ML estimate $\hat{\bar{\boldsymbol{S}}}$ can be written as [cf. (3.11)][4]

$$d_1\left(\hat{\bar{\boldsymbol{S}}}\right) = \bar{\boldsymbol{r}}^{\mathsf{H}}\left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \left(\hat{\bar{\boldsymbol{S}}}_{\mathrm{D}}\left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\hat{\bar{\boldsymbol{S}}}_{\mathrm{D}}^{\mathsf{H}}\right)\right)\bar{\boldsymbol{r}}. \tag{5.39}$$

Thus, if we define

$$\boldsymbol{F} \triangleq \hat{\bar{\boldsymbol{S}}}_{\mathrm{D}}\left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\hat{\bar{\boldsymbol{S}}}_{\mathrm{D}}^{\mathsf{H}} - \bar{\boldsymbol{S}}_{\mathrm{D}}\left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\bar{\boldsymbol{S}}_{\mathrm{D}}^{\mathsf{H}} \tag{5.40}$$

[3] In the case of BDSTM with blocklength L^{BDSTM} the temporal correlation is described by $\boldsymbol{\Psi}_{\mathrm{c}}^{\mathrm{t}} = \operatorname{toeplitz}\left\{\boldsymbol{\Psi}_{\mathrm{c},0}^{\mathrm{t}}, \boldsymbol{\Psi}_{\mathrm{c},1}^{\mathrm{t}}, \ldots, \boldsymbol{\Psi}_{\mathrm{c},N-1}^{\mathrm{t}}\right\}$ with $\boldsymbol{\Psi}_{\mathrm{c},i}^{\mathrm{t}} \triangleq \operatorname{toeplitz}\left\{\psi^{\mathrm{t}}[i], \psi^{\mathrm{t}}\left[i+L^{\mathrm{BDSTM}}\right], \ldots, \psi^{\mathrm{t}}\left[i+(N_{\mathrm{T}}-1)L^{\mathrm{BDSTM}}\right]\right\}$.

[4] Clearly, this holds for any $\tilde{\bar{\boldsymbol{S}}}$, but we write $\hat{\bar{\boldsymbol{S}}}$ here, because we are concerned with the probability that the decoder *decides* in favor of $\hat{\bar{\boldsymbol{S}}}$ rather than $\bar{\boldsymbol{S}}$.

we can write

$$\Delta \triangleq d_1\Big(\hat{\bar{\boldsymbol{S}}}\Big) - d_1\big(\bar{\boldsymbol{S}}\big) \tag{5.41}$$

$$= \bar{\boldsymbol{r}}^{\mathsf{H}}\left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{F}\right)\bar{\boldsymbol{r}} \tag{5.42}$$

for the difference between the metrics of the ML estimate $\hat{\bar{\boldsymbol{S}}}$ and the true signal $\bar{\boldsymbol{S}}$. Thus, we can see that the pairwise error probability $\mathrm{PEP}\big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\big)$, i.e. the probability of the decoder deciding in favor of $\hat{\bar{\boldsymbol{S}}} \neq \bar{\boldsymbol{S}}$ in a binary decision between $\bar{\boldsymbol{S}}$ and $\hat{\bar{\boldsymbol{S}}}$ disregarding all other candidates $\tilde{\bar{\boldsymbol{S}}} \notin \big\{\bar{\boldsymbol{S}}, \hat{\bar{\boldsymbol{S}}}\big\}$, can be computed via

$$\mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big) = \Pr(\Delta \leq 0)\,. \tag{5.43}$$

We observe that Δ is a Hermitian quadratic form in zero–mean complex Gaussian random variables. Hence, the characteristic function $\Phi_\Delta(v) \triangleq \mathcal{E}\big\{\mathrm{e}^{\mathrm{j}v\Delta}\big\}$ of Δ is given by [SBS66]

$$\Phi_\Delta(v) = \det\big\{\boldsymbol{I}_{NN_{\mathrm{T}}N_{\mathrm{R}}} - \mathrm{j}v\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}|\bar{\boldsymbol{S}}}(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{F})\big\}^{-1} \tag{5.44}$$

$$= \prod_{j=1}^{NN_{\mathrm{T}}N_{\mathrm{R}}} \Big(1 - \mathrm{j}v\lambda_{\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}|\bar{\boldsymbol{S}}}(\boldsymbol{I}_{N_{\mathrm{R}}}\otimes\boldsymbol{F}),j}\Big)^{-1}\,, \tag{5.45}$$

($\lambda_{\boldsymbol{X},i}$: ith eigenvalue of $\boldsymbol{X}$) with $\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}|\bar{\boldsymbol{S}}}$ and $\boldsymbol{F}$ as defined in (5.35) and (5.40), respectively. Based on (5.43) and (5.44) the PEP can then be computed via (cf. e.g. [Pro00])

$$\mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big) = -\frac{1}{2\pi\mathrm{j}}\int_{-\infty+\mathrm{j}c}^{\infty+\mathrm{j}c} \frac{\Phi_\Delta(v)}{v}\,\mathrm{d}v \tag{5.46}$$

$$= -\sum_{i=1}^{N_v} \operatorname*{Res}_{v=\mathrm{j}v_i^+} \left\{\frac{\Phi_\Delta(v)}{v}\right\} \tag{5.47}$$

$$= -\sum_{i=1}^{N_v} \operatorname*{Res}_{v=\mathrm{j}v_i^+} \left\{\frac{1}{v}\prod_{l=1}^{NN_{\mathrm{T}}N_{\mathrm{R}}} \frac{-\mathrm{j}v_l}{v-\mathrm{j}v_l}\right\}, \tag{5.48}$$

($\mathrm{Res}_{x=\hat{x}}\{f(x)\}$: residue of $f(x)$ corresponding to the pole at $x=\hat{x}$) with

$$v_l \triangleq -\frac{1}{\lambda_{\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}|\bar{\boldsymbol{S}}}\boldsymbol{F},l}}, \quad 1 \leq l \leq NN_{\mathrm{T}}N_{\mathrm{R}}, \tag{5.49}$$

and $\mathrm{j}v_i^+$, $1 \leq i \leq N_v \leq NN_{\mathrm{T}}N_{\mathrm{R}}$, denoting those N_v poles of $\Phi_\Delta(v)$ that lie in the upper half of the complex v–plane. The constant $c \in \mathbb{R}$ that must satisfy $0 \leq c \leq \min_{1\leq i\leq N_v}(\mathrm{Re}\big\{v_i^+\big\})$ is introduced to move the path of integration away from the pole at $v=0$. Consequently, the PEP can be evaluated —with arbitrarily high accuracy as $n \to \infty$— through numerical integration of (5.46) using a Gauss–Chebyshev quadrature rule [BCTV98, HF92]

$$\mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big) \approx \frac{1}{n}\sum_{k=1}^{n/2} \mathrm{Re}\left\{\Phi_\Delta(c\tau_k + \mathrm{j}c)\right\} - \tau_k \mathrm{Im}\left\{\Phi_\Delta(c\tau_k + \mathrm{j}c)\right\} \tag{5.50}$$

with

$$\tau_k \stackrel{\triangle}{=} \tan\left(\frac{(2k-1)\,\pi}{2n}\right). \tag{5.51}$$

Alternatively, the PEP can be computed based on (5.48) using the standard formula (cf. e.g. [BV01])

$$\operatorname*{Res}_{v=\mathrm{j}v_i^+}\left\{\frac{\Phi_\Delta(v)}{v}\right\} = \lim_{v\to\mathrm{j}v_i^+}\frac{1}{\Gamma(l_i)}\frac{\mathrm{d}^{l_i-1}}{\mathrm{d}v^{l_i-1}}\frac{\Phi_\Delta(v)(v-\mathrm{j}v_i^+)^{l_i}}{v} \tag{5.52}$$

($\Gamma(y) \stackrel{\triangle}{=} \int_0^\infty x^{y-1}\mathrm{e}^{-x}\,\mathrm{d}x$ for $y \in \mathbb{R}$ and $\Gamma(y) = (y-1)!$ for $y \in \mathbb{N}$: Gamma function) to compute the residue corresponding to a pole $\mathrm{j}v_i^+$ of multiplicity l_i. As this may be quite difficult to evaluate if a pole $\mathrm{j}v_i^+$ has multiplicity $l_i > 1$, Siwamogsatham et al. [SFG02] proposed a simple yet very accurate approximation technique, where different small constants are added to $\mathrm{j}v_i^+$ to disperse the multiple pole into a small cloud of distinct single poles and the individual residues are evaluated using (5.52) with $l_i = 1$.

5.4.1.2 CDD and DFDD

Due to the close relation between CDD, DFDD and MSDD, the expressions derived above for the PEPs of MSDD extend to CDD ($N = 2$) and DFDD as well.

5.4.1.3 (Differentially) Coherent Detection

The PEPs for coherent detection can be derived in complete analogy to the case of MSDD and a detailed derivation is therefore omitted. Based on a vectorization of the channel model as in (5.30) the difference between the metrics of the true transmit signal $\boldsymbol{S}[k]$ and an estimate $\hat{\boldsymbol{S}}[k]$ reads [cf. (2.56)]

$$\Delta_{\mathrm{coh}} = \bar{\boldsymbol{z}}^{\mathsf{H}}\left(\boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{F}'_{\mathrm{coh}}\right)\bar{\boldsymbol{z}} \tag{5.53}$$

with

$$\bar{\boldsymbol{z}} \stackrel{\triangle}{=} \begin{bmatrix} \mathrm{vec}\left\{\boldsymbol{H}[k]\right\} \\ \mathrm{vec}\left\{\boldsymbol{R}[k]\right\} \end{bmatrix} \tag{5.54}$$

$$\boldsymbol{F}_{\mathrm{coh}} \stackrel{\triangle}{=} \begin{bmatrix} \boldsymbol{0}_{N_\mathrm{T},N_\mathrm{T}} & \left(\boldsymbol{S}[k]-\hat{\boldsymbol{S}}[k]\right)^{\mathsf{H}} \\ \boldsymbol{S}[k]-\hat{\boldsymbol{S}}[k] & \boldsymbol{0}_{N_\mathrm{T},N_\mathrm{T}} \end{bmatrix}. \tag{5.55}$$

One can see that Δ_{coh} is again a Hermitian quadratic form in zero–mean complex Gaussian random variables $\bar{\boldsymbol{z}}$ whose autocorrelation matrix can be found as

$$\begin{aligned} \boldsymbol{\Psi}_{\bar{\boldsymbol{z}}\bar{\boldsymbol{z}}|\boldsymbol{S}} &\stackrel{\triangle}{=} \mathcal{E}\left\{\bar{\boldsymbol{z}}\bar{\boldsymbol{z}}^{\mathsf{H}} \mid \boldsymbol{S}[k]\right\} \\ &= \begin{bmatrix} \boldsymbol{\Psi}_{\boldsymbol{HH}} & \boldsymbol{\Psi}_{\boldsymbol{HH}}\left(\boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{S}^{\mathsf{H}}[k]\right) \\ \left(\boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{S}[k]\right)\boldsymbol{\Psi}_{\boldsymbol{HH}} & \left(\boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{S}[k]\right)\boldsymbol{\Psi}_{\boldsymbol{HH}}\left(\boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{S}^{\mathsf{H}}[k]\right) + \sigma_n^2\boldsymbol{I}_{N_\mathrm{T}N_\mathrm{R}} \end{bmatrix}, \end{aligned} \tag{5.56}$$

where [cf. (4.13)]

$$\boldsymbol{\Psi}_{\boldsymbol{HH}} = \boldsymbol{\Psi}^{\mathrm{Rx}} \otimes \boldsymbol{\Psi}^{\mathrm{Tx}}, \tag{5.57}$$

with $\boldsymbol{\Psi}^{\mathrm{Tx}}$ and $\boldsymbol{\Psi}^{\mathrm{Rx}}$ [cf. (4.9) and (4.10)] describing spatial correlation at transmitter and receiver, respectively. Therefore, the characteristic function of Δ_{coh} reads

$$\begin{aligned} \Phi_{\Delta_{\mathrm{coh}}}(v) &= \det\left\{ \boldsymbol{I}_{N_{\mathrm{T}}N_{\mathrm{R}}} - \mathrm{j}v\boldsymbol{\Psi}_{\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}|\boldsymbol{S}} \boldsymbol{F}_{\mathrm{coh}} \right\}^{-1} && (5.58) \\ &= \prod_{j=1}^{N_{\mathrm{T}}N_{\mathrm{R}}} \left(1 - \mathrm{j}v\lambda_{\boldsymbol{\Psi}_{\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}|\boldsymbol{S}}\boldsymbol{F}_{\mathrm{coh}},j} \right)^{-1}, && (5.59) \end{aligned}$$

and the PEPs can be computed using one of the methods described in Section 5.4.1.1.

Note that the individual PEPs of coherent and differentially coherent detection are the same. The difference in average SER is caused by error propagation in differentially coherent detection, which is accounted for in the SER approximation, cf. Section 5.4.4.

5.4.2 Pairwise Error Probabilities Based on the QSFC Model

The PEP analysis presented in Section 5.4.1 is valid for arbitrary MIMO fading channel scenarios, i.e. regardless of the fading bandwidth and in the presence of spatial correlation. However, it does not permit much theoretical and intuitive insight regarding issues such as diversity order, possible error floors at high SNR or comparisons between the different detection schemes.

In order to gain such insights, we present a novel approach to evaluate the PEPs based on the QSFC, which is a valid model for arbitrary DSTM with slow to moderately fast fading and with (di–)cyclic DSTM for fast fading as well (cf. discussion in Section 5.2.2). In particular, we consider the interesting special case of a MIMO channel without spatial correlation, i.e.

$$\boldsymbol{\Psi}^{\mathrm{Tx}} = \boldsymbol{I}_{N_{\mathrm{T}}} \quad \text{and} \quad \boldsymbol{\Psi}^{\mathrm{Rx}} = \boldsymbol{I}_{N_{\mathrm{R}}}. \tag{5.60}$$

As basis for our considerations serves a generalization of a classical result from [Pro00, Appendix B] for sums of Gaussian quadratic forms to the case, where the individual terms in the sums are multiplied by *non–identical* weights. This is presented in Appendix 5.B.

5.4.2.1 Exact PEP of Various Detection Algorithms

5.4.2.1.1 (DF–)MSDD As mentioned at the beginning of Section 5.4, we are only interested in the *dominant* error events, which correspond to single–transmit symbol errors, i.e. to candidates $\hat{\bar{\boldsymbol{S}}}\big(l, \hat{l}\big)$ of the form of (5.29). The difference Δ_n between the ML–MSDD metrics of such a candidate $\hat{\bar{\boldsymbol{S}}}\big(l, \hat{l}\big)$ with an error in position n and the true transmit signal $\bar{\boldsymbol{S}}$ can be

derived from (5.41) using a number of rather straightforward manipulations as

$$\Delta_n \stackrel{\triangle}{=} d_1\Big(\hat{\bar{\boldsymbol{S}}}\big(l,\hat{l}\big)\Big) - d_1\big(\bar{\boldsymbol{S}}\big) \tag{5.61}$$

$$= 2\sum_{i=1}^{N_{\mathrm{T}}} \mathrm{Re}\left\{\theta_i\big(l,\hat{l}\big)\sum_{j=1}^{N_{\mathrm{R}}}[\boldsymbol{F}_n]_{i,j}\sum_{\mu=1,\mu\neq n}^{N} m_{\mu,n}[\boldsymbol{E}_{n,\mu}]^*_{i,j}\right\}, \tag{5.62}$$

where $m_{i,j}$ are the elements of $\boldsymbol{M}$ [cf. (5.25)],

$$\boldsymbol{F}_n \stackrel{\triangle}{=} \big(\boldsymbol{Q}\big(l,\hat{l}\big)\big)^{\mathsf{H}}\boldsymbol{R}_n \tag{5.63}$$

$$\boldsymbol{E}_{n,\mu} \stackrel{\triangle}{=} \big(\boldsymbol{Q}\big(l,\hat{l}\big)\big)^{\mathsf{H}}\boldsymbol{S}_n\boldsymbol{S}_\mu^{\mathsf{H}}\boldsymbol{R}_\mu \tag{5.64}$$

and where $\theta_i\big(l,\hat{l}\big)$ and unitary $\boldsymbol{Q}\big(l,\hat{l}\big)$ are obtained from the eigenvalue decomposition

$$(\boldsymbol{V}^{(l)})^{\mathsf{H}}\boldsymbol{V}^{(\hat{l})} - \boldsymbol{I}_{N_{\mathrm{T}}} \stackrel{\triangle}{=} \boldsymbol{Q}\big(l,\hat{l}\big)\mathrm{diag}\Big\{\theta_1\big(l,\hat{l}\big),\ldots,\theta_{N_{\mathrm{T}}}\big(l,\hat{l}\big)\Big\}\big(\boldsymbol{Q}\big(l,\hat{l}\big)\big)^{\mathsf{H}}.^{5} \tag{5.65}$$

Then $\mathrm{PEP}\big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\big)$, which will be briefly denoted as PEP_n, is given by

$$\mathrm{PEP}_n = \Pr(\Delta_n \leq 0)\,. \tag{5.66}$$

Identifying

$$x_{i,j} \stackrel{\triangle}{=} \sum_{\mu=1,\mu\neq n}^{N} m_{n,\mu}[\boldsymbol{E}_{n,\mu}]^*_{i,j} \tag{5.67}$$

$$y_{i,j} \stackrel{\triangle}{=} [\boldsymbol{F}_n]^*_{i,j} \tag{5.68}$$

it can be shown (cf. Appendix 5.A.2) that $\{x_{i,j}, y_{i,j}\}$ are —regardless of $\boldsymbol{Q}\big(l,\hat{l}\big)$— iid pairs of correlated zero–mean complex Gaussian random variables with variances and covariance [cf. (5.325)]

$$\sigma_x^2 = m_{n,n}[m_{n,n}(1+\sigma_n^2) - 1] \tag{5.69}$$

$$\sigma_y^2 = 1 + \sigma_n^2 \tag{5.70}$$

$$\mu_{xy} = 1 - m_{n,n}(1+\sigma_n^2). \tag{5.71}$$

Further, let $N_\Theta \leq N_{\mathrm{T}}$ denote the number of different values $\theta_\iota\big(l,\hat{l}\big)$, $1 \leq \iota \leq N_{\mathrm{T}}$, and let $|\Theta_i|$ and l_i, $1 \leq i \leq N_\Theta$, be defined as these different values and as the numbers of eigenvalues $\theta_\iota\big(l,\hat{l}\big)$, $1 \leq \iota \leq N_{\mathrm{T}}$, that lead to the same value of $|\Theta_i|$, respectively. Then, we can see that (5.62) is of the form of the general quadratic form (5.324). Thus, using the results of Appendix 5.B the

[5] We note that the related literature usually considers the *singular value decomposition* of $(\boldsymbol{V}^{(l)})^{\mathsf{H}}\boldsymbol{V}^{(\hat{l})}$, cf. e.g. [SHHS01, HS00]. However, for our purpose the eigenvalue decomposition of $(\boldsymbol{V}^{(l)})^{\mathsf{H}}\boldsymbol{V}^{(\hat{l})} - \boldsymbol{I}_{N_{\mathrm{T}}}$ is more convenient and leads to a diagonal matrix $\big(\boldsymbol{Q}\big(l,\hat{l}\big)\big)^{\mathsf{H}}\big((\boldsymbol{V}^{(l)})^{\mathsf{H}}\boldsymbol{V}^{(\hat{l})} - \boldsymbol{I}_{N_{\mathrm{T}}}\big)\boldsymbol{Q}\big(l,\hat{l}\big)$ since $(\boldsymbol{V}^{(l)})^{\mathsf{H}}\boldsymbol{V}^{(\hat{l})}$ is unitary [HJ85].

exact pairwise error probability PEP_n can be computed in closed form and without eigenvalue decompositions of large matrices via [cf. (5.326)]

$$\Pr(\Delta<0) = \sum_{i=1}^{N_\Theta}(-1)^{\sum_{\substack{j=1\\ j\neq i}}^{N_\Theta} l_j} \sum_{k_{1,0}=0}^{l_i-1}\sum_{k_{1,1}=0}^{k_{1,0}} \cdots \sum_{k_{1,N_\Theta}=0}^{k_{1,N_\Theta-1}}\sum_{k_{2,1}=0}^{k_{2,0}} \cdots \sum_{k_{2,i-1}=0}^{k_{1,i-2}}\sum_{k_{2,i+1}=0}^{k_{1,i-1}} \cdots \sum_{k_{2,N_\Theta-1}=0}^{k_{1,N_\Theta-2}} \tag{5.72}$$
$$v_{2,i}^{k_{1,0}} \frac{\binom{l_i+k_{1,i-1}-k_{1,i}-1}{l_i-1} v_{1,i}^{l_i}}{(v_{2,i}+v_{1,i})^{l_i+k_{1,i-1}-k_{1,i}}} \prod_{\substack{j=1\\ j\neq i}}^{N_\Theta} \frac{\binom{l_j+k_{1,j-1}-k_{1,j}-1}{l_j-1}\binom{l_j+k_{2,j-1}-k_{2,j}-1}{l_j-1}(v_{1,j}v_{2,j})^{l_j}}{(v_{2,i}+v_{1,j})^{l_j+k_{1,j-1}-k_{1,j}}(v_{2,i}-v_{2,j})^{l_j+k_{2,j-1}-k_{2,j}}},$$

with

$$v_{1|2,i} = \frac{1}{2}\left(\sqrt{1+\frac{4}{|\Theta_i|^2\rho_n}} \mp 1\right) \tag{5.73}$$
$$\rho_n \triangleq \frac{|\mu_{xy}|^2}{\sigma_x^2\sigma_y^2-|\mu_{xy}|^2} = m_{n,n}(1+\sigma_n^2)-1, \tag{5.74}$$

where the "$-$" and "$+$" of "$\mp$" in (5.73) correspond to $v_{1,i}$ and $v_{2,i}$, respectively. Interestingly, (5.73)–(5.74) reveal that PEP_n depends on the statistical properties of the channel only through the diagonal element $m_{n,n}$ of the inverse autocorrelation matrix $\boldsymbol{M}$ of the fading–plus–noise process and the noise variance.[6]

Special Case DPSK: For DPSK (5.72) can be considerably simplified. Assuming the general DPSK case with N_R receive antennas, only one residue corresponding to a pole of order N_R needs to be computed, i.e. $N_\Theta = 1$ and $l_1 = N_\mathrm{R}$, and (5.72) reduces to

$$\mathrm{PEP}_n = \left[\frac{1}{2}\left(1-\frac{1}{\sqrt{1+\frac{4}{|v^{(\hat{l})}-v^{(l)}|^2\rho_n}}}\right)\right]^{N_\mathrm{R}} \sum_{k=0}^{N_\mathrm{R}-1}\binom{N_\mathrm{R}+k-1}{N_\mathrm{R}-1}\frac{1}{2^k}\left(1+\frac{1}{\sqrt{1+\frac{4}{|v^{(\hat{l})}-v^{(l)}|^2\rho_n}}}\right)^k, \tag{5.75}$$

where $v^{(l)}$ and $v^{(\hat{l})}$ represent the true and the estimated data PSK symbol, respectively. For $N_\mathrm{R}=1$ (5.75) is a closed–form solution for [HF92, Eq. (9)] for the dominant error event. For $N_\mathrm{R}\geq 1$ (5.75) is a generalization of the results in [SA98, SA01, KHR05] which are limited to static fading.

5.4.2.1.2 CDD and DFDD The analytical performance evaluation of DFDD is usually performed by considering "genie–aided" DFDD, i.e. under the assumption that previous decisions that are fed back into the metric (2.53) are always correct, and accounting for error

[6]Note, that this derivation also extends to arbitrary matrices $\boldsymbol{M} \neq \left(\boldsymbol{\Psi}^\mathrm{t}+\sigma_n^2\boldsymbol{I}_N\right)^{-1}$, i.e. to the case of mismatched receivers. In this case, however the resulting analytical expressions for σ_x^2, σ_y^2 and μ_{xy} and thereby for the PEPs are less insightful.

propagation occurring typically with "real–world" DFDD by multiplying the resulting error rates with a factor of two, cf. e.g. [SGH99, SL02].

Consequently, the error–rate analysis of DFDD can be performed using the above results for MSDD, i.e. the PEP for an error event $(\boldsymbol{V}^{(l)} \to \boldsymbol{V}^{(\hat{l})})$ of genie–aided DFDD is given by

$$\mathrm{PEP}\big(\boldsymbol{V}^{(l)} \to \boldsymbol{V}^{(\hat{l})}\big) = \mathrm{PEP}_N, \tag{5.76}$$

which allows us to calculate the PEP of DFDD for general unitary DSTM, whereas the results in [SL02, DB06] are limited to cyclic DSTM.

Due to the close relation between MSDD, DFDD and CDD, the PEP for CDD can be computed by setting $N = 2$ in (5.76).

5.4.2.1.3 (Differentially) Coherent Detection In analogy to the MSDD case, it is straightforward to show that the metric difference corresponding to an error event $\big(\boldsymbol{V}^{(l)} \to \boldsymbol{V}^{(\hat{l})}\big)$ can be written as

$$\Delta_{\mathrm{coh}} = 2\sigma_n^2 \sum_{i=1}^{N_{\mathrm{T}}} \mathrm{Re}\left\{ \theta_i\big(l,\hat{l}\big) \sum_{j=1}^{N_{\mathrm{R}}} x_{i,j} y_{i,j}^* \right\}, \tag{5.77}$$

with $x_{i,j}$ and $y_{i,j}$ as elements in the ith row and jth column of

$$\boldsymbol{X} \triangleq -\frac{1}{\sigma_n}\big(\boldsymbol{Q}\big(l,\hat{l}\big)\big)^{\mathsf{H}}\boldsymbol{H} \tag{5.78}$$

$$\boldsymbol{Y} \triangleq \frac{1}{\sigma_n}\big(\boldsymbol{Q}\big(l,\hat{l}\big)\big)^{\mathsf{H}}\big(\boldsymbol{V}^{(l)}\big)^{\mathsf{H}}\boldsymbol{H}, \tag{5.79}$$

respectively, and $\boldsymbol{Q}\big(l,\hat{l}\big)$ as defined in (5.65). It can be shown in analogy to the proof given in Appendix 5.A.2 for MSDD that $\{x_{i,j}, y_{i,j}\}$ are —regardless of $\boldsymbol{Q}\big(l,\hat{l}\big)$— iid pairs of correlated zero–mean complex Gaussian random variables with [cf. (5.325)]

$$\sigma_x^2 = \frac{1}{\sigma_n^2} \tag{5.80}$$

$$\sigma_y^2 = 1 + \frac{1}{\sigma_n^2} \tag{5.81}$$

$$\mu_{xy} = -\frac{1}{\sigma_n^2}. \tag{5.82}$$

Consequently, the PEPs of (differentially) coherent detection can be computed in closed form using (5.72) with

$$v_{1|2,i} = \frac{1}{2}\left(\sqrt{1 + \frac{4\sigma_n^2}{|\Theta_i|^2}} \mp 1\right) \tag{5.83}$$

and Θ_i as defined in Section 5.4.2.1.1.

5.4.2.1.4 Comparison A comparison of (5.73) and (5.83) reveals that the $v_{1|2,j}$, $1 \leq j \leq N_\Theta$, for noncoherent, i.e. MSDD, DFDD and CDD, and coherent detection are formally very similar. As the expressions to compute the PEPs from the $v_{1|2,j}$ are identical for the different detection schemes, it is reasonable to regard ρ_n of (5.74) as a sort of *effective SNR* (ESNR) for position n, $1 \leq n \leq N$, of the MSDD observation window and ρ_N at the same time as the ESNR for DFDD which is to be compared to the true SNR $1/\sigma_n^2$ relevant for coherent detection with perfect CSI. We can therefore consider the PEPs for the different detection schemes simultaneously using the general effective SNR ρ_{eff} defined as

$$\rho_{\text{eff}} \triangleq \begin{cases} \rho_n = m_{n,n}(1+\sigma_n^2) - 1, \quad 1 \leq n \leq N, & \text{MSDD} \\ \rho_N = m_{N,N}(1+\sigma_n^2) - 1, & \text{DFDD, CDD} \\ 1/\sigma_n^2, & \text{coherent detection} \end{cases} \tag{5.84}$$

and

$$v_{1|2,i} = \frac{1}{2}\left(\sqrt{1 + \frac{4}{|\Theta_i|^2 \rho_{\text{eff}}}} \mp 1\right). \tag{5.85}$$

This holds for transmission over spatially uncorrelated MIMO channels using arbitrary DSTM constellations when the assumption of quasi–static fading is justified and for cyclic DSTM constellations under any flat–fading conditions. Similar ESNR approaches have been pursued previously in [PS03, DB06]. In particular, the ESNR ρ_{eff} for the special case of $N = 2$, i.e. CDD, can be shown to equal $\rho_{\text{eff}} = ([\boldsymbol{\Psi}^{\text{t}}]_{1,2})^2/[(1+\sigma_n^2)^2 - ([\boldsymbol{\Psi}^{\text{t}}]_{1,2})^2]$, with $\boldsymbol{\Psi}^{\text{t}}$ as defined in (5.26), which is the same as the ESNR derived in [DB06, Eq. (33)]. Furthermore, extending upon [PS03] Du et al. presented in [DB06] expressions for the ESNR of DFDD that are equivalent to those derived here. Hence, our definition of the ESNR for arbitrary positions in the MSDD observation window can be viewed as a generalization of the ESNRs of [PS03, DB06].

5.4.2.1.5 Performance for Very Slow Fading In the limit of very slow fading, i.e. when the channel remains practically constant during the transmission of N consecutive DSTM symbols, the inverse fading–plus–noise autocorrelation matrix converges to $\boldsymbol{M} = (\boldsymbol{I}_N - (N + \sigma_n^2)^{-1}\mathbf{1}_{N,N})/\sigma_n^2$. Thus, the $m_{n,n}$ can be given in closed form as

$$m_{n,n} = \frac{N + \sigma_n^2 - 1}{\sigma_n^2(N + \sigma_n^2)}, \tag{5.86}$$

i.e. the loss in SNR of MSDD and DFDD compared to differentially coherent detection with perfect CSI is [cf. (5.84)]

$$\frac{\rho_n}{1/\sigma_n^2} = \frac{N-1}{N+\sigma_n^2} \tag{5.87}$$

and we obtain the well–known result that for static fading the performance of MSDD and DFDD in the limit of $N \to \infty$ converges to that of differentially coherent detection with perfect CSI, cf. [DS90, Edb92, SL02] for PSK with MSDD, PSK with DFDD and DSTM with DFDD, respectively.

5.4.2.2 Asymptotic Performance for Very High SNR ($\sigma_n^2 \to 0$)

In this section, we consider the performance of the different detection schemes in the limit of infinite SNR. We assume for the moment that $\sigma_n^2 \to 0$ implies $\rho_{\text{eff}} \to \infty$ for any detection scheme and come back with detailed comments on this in Section 5.4.2.2.1.

In Appendix 5.A.3 we prove that the asymptotic behavior of the pairwise error probability PEP for arbitrary DSTM constellations is in the limit of $[\sigma_n^2, \rho_{\text{eff}}] \to [0, \infty]$ given by

$$\text{PEP} \asymp \rho_{\text{eff}}^{-N_{\text{T}}N_{\text{R}}} \frac{\binom{2N_{\text{T}}N_{\text{R}}-1}{N_{\text{T}}N_{\text{R}}}}{\prod_{i=1}^{N_{\text{T}}} |\theta_i(l,\hat{l})|^{2N_{\text{R}}}}, \quad [\sigma_n^2, \rho_{\text{eff}}] \to [0, \infty] \tag{5.88}$$

with ρ_{eff} as defined in (5.84) ($f(x) \asymp g(x),\ x \to \hat{x}$: "asymptotic equality" in the sense that $\lim_{x\to\hat{x}} \frac{f(x)}{g(x)} = 1$ and $\lim_{x\to\hat{x}} \frac{\mathrm{d}}{\mathrm{d}x} f(x) = \lim_{x\to\hat{x}} \frac{\mathrm{d}}{\mathrm{d}x} g(x)$).

5.4.2.2.1 (DF–)MSDD, DFDD, CDD

Validity of $\rho_{\text{eff}} \to \infty$: Eq. (5.88) is based on the assumption that $\sigma_n^2 \to 0$ implies that $\rho_{\text{eff}} = \rho_n \to \infty$. Therefore, a discussion of the behavior of ρ_{eff} in dependence of σ_n^2 is in order at this point. Here, we want to distinguish two different regions: We are interested in (i) how ρ_{eff} behaves in dependence of σ_n^2 as the latter decreases, and (ii) the limiting case of $\sigma_n^2 = 0$.

To this end, let us consider the matrix $\boldsymbol{M} = (\boldsymbol{\Psi}^{\text{t}} + \sigma_n^2 \boldsymbol{I}_N)^{-1}$. Based on the eigenvalue decomposition $\boldsymbol{\Psi}^{\text{t}} = \boldsymbol{U}_{\boldsymbol{\Psi}^{\text{t}}} \boldsymbol{\Lambda}_{\boldsymbol{\Psi}^{\text{t}}} \boldsymbol{U}_{\boldsymbol{\Psi}^{\text{t}}}^{\mathsf{H}}$ we can express $\boldsymbol{M}$ as

$$\boldsymbol{M} = \boldsymbol{U}_{\boldsymbol{\Psi}^{\text{t}}} \text{diag}\left\{ \frac{1}{\lambda_{\boldsymbol{\Psi}^{\text{t}},1} + \sigma_n^2}, \ldots, \frac{1}{\lambda_{\boldsymbol{\Psi}^{\text{t}},N} + \sigma_n^2} \right\} \boldsymbol{U}_{\boldsymbol{\Psi}^{\text{t}}}^{\mathsf{H}}. \tag{5.89}$$

Recalling the notation that the eigenvalues $\lambda_{\boldsymbol{\Psi}^{\text{t}},i}$ are returned in order of decreasing magnitude, $\boldsymbol{M}$ depends predominantly on σ_n^2 and the smallest eigenvalue $\lambda_{\boldsymbol{\Psi}^{\text{t}},N}$ of $\boldsymbol{\Psi}^{\text{t}}$ and can be approximated coarsely via[7]

$$\boldsymbol{M} \approx \frac{1}{\lambda_{\boldsymbol{\Psi}^{\text{t}},N} + \sigma_n^2} [\boldsymbol{U}_{\boldsymbol{\Psi}^{\text{t}}}]_{:,N} \left([\boldsymbol{U}_{\boldsymbol{\Psi}^{\text{t}}}]_{:,N} \right)^{\mathsf{H}}. \tag{5.90}$$

From this we can see that $\boldsymbol{M}$ and by (5.84) the effective SNR ρ_{eff} grow roughly linearly with the SNR σ_n^{-2} as long as $\sigma_n^2 > \lambda_{\boldsymbol{\Psi}^{\text{t}},N}$. On the other hand, if $\sigma_n^2 < \lambda_{\boldsymbol{\Psi}^{\text{t}},N}$ it becomes irrelevant and $\boldsymbol{M}$ converges to $(\boldsymbol{\Psi}^{\text{t}})^{-1}$ the magnitude of its elements depending mainly on the smallest

[7] The behavior of the eigenvalues of a correlation matrix will be discussed in detail in Section 5.5.2.

eigenvalue $\lambda_{\mathbf{\Psi}^{\mathrm{t}},N}$ of the fading correlation matrix $\mathbf{\Psi}^{\mathrm{t}}$. In the latter case, the ESNR approaches [cf. (5.84)]

$$\rho_\infty \stackrel{\triangle}{=} \lim_{\sigma_n^2 \to 0} \rho_{\mathrm{eff}} = \left[\left(\mathbf{\Psi}^{\mathrm{t}}\right)^{-1}\right]_{n,n} - 1. \tag{5.91}$$

For typical fading bandwidths (i.e. $B_{h,\mathrm{eff}}T \leq 0.1$) $\mathbf{\Psi}^{\mathrm{t}}$ is rather ill conditioned and $\rho_\infty = \left[\left(\mathbf{\Psi}^{\mathrm{t}}\right)^{-1}\right]_{n,n} - 1 \gg 1$ holds even for small observation window sizes N. Thus, the PEP_n are very well approximated by (5.88). This will also be confirmed in Section 5.4.5 by comparing this asymptotic result with the exact result of Section 5.4.2.1 by means of numerical examples.

Error Floor for $B_{h,\mathrm{eff}}T > 0$: From (5.88) and (5.91) it is immediately clear that MSDD entails an error floor, which for growing diagonal elements of $\left(\mathbf{\Psi}^{\mathrm{t}}\right)^{-1}$ is increasingly well approximated by

$$\lim_{\sigma_n^2 \to 0} \mathrm{PEP}_n \approx \left(\left[\left(\mathbf{\Psi}^{\mathrm{t}}\right)^{-1}\right]_{n,n} - 1\right)^{-N_{\mathrm{T}}N_{\mathrm{R}}} \frac{\binom{2N_{\mathrm{T}}N_{\mathrm{R}}-1}{N_{\mathrm{T}}N_{\mathrm{R}}}}{\prod_{i=1}^{N_{\mathrm{T}}} |\theta_i(l,\hat{l})|^{2N_{\mathrm{R}}}}. \tag{5.92}$$

Interestingly, we prove in Appendix 5.A.4 for band–limited PSDs (i.e. $\Psi^{\mathrm{t}}(f) = 0$, $B_{h,\mathrm{eff}} < |f| \leq 1/(2T)$), that are bounded in the sense that $\max_{0 \leq f \leq B_{h,\mathrm{eff}}}\{\Psi^{\mathrm{t}}(f)B_{h,\mathrm{eff}}T\} < \infty$, the asymptotic relation

$$\left[\left(\mathbf{\Psi}^{\mathrm{t}}\right)^{-1}\right]_{n,n} \asymp c \cdot (B_{h,\mathrm{eff}}T)^{-2(N-1)}, \quad B_{h,\mathrm{eff}}T \to 0, \tag{5.93}$$

where c is a constant which depends on the particular shape of the PSD. Although Clarke's fading spectrum (4.22) is not bounded in the above sense, (5.93) holds in this case as well, cf. Appendix 5.A.4. For standard fading models with PSDs that are not band–limited, such as the "Gaussian", "First–" and "Second–order Butterworth" PSDs, cf. Table 4.1, the "width" of the PSD scales effectively linearly in $B_{h,\mathrm{eff}}T$. Consequently, the same kind of behavior as for band–limited PSDs can be expected for small $B_{h,\mathrm{eff}}T$.

Since the above result also applies to DFDD for $n = N$, we infer that for $B_{h,\mathrm{eff}}T \ll 1$ and $\sigma_n^2 = 0$ the ESNRs of DFDD and MSDD decrease exponentially with N, whereas the PEP and SER error floors decrease exponentially with $NN_{\mathrm{T}}N_{\mathrm{R}}$ [cf. (5.88)]. This is a new result and explains the observations made in many papers based on simulation or numerical results, cf. e.g. [HF92, SGH99, SL02, LLK04], that increasing N rapidly reduces the error floor of DFDD and MSDD. Again, all of these results apply equally to the error–rate analysis of CDD with $N = 2$.

5.4.2.2.2 (Differentially) Coherent Detection For the sake of completeness, we point out that the asymptotic PEP of coherent detection can be obtained by letting $\rho_{\mathrm{eff}} = 1/\sigma_n^2$ in

(5.88), i.e.

$$\mathrm{PEP} \asymp (\sigma_n^2)^{N_\mathrm{T}N_\mathrm{R}} \frac{\binom{2N_\mathrm{T}N_\mathrm{R}-1}{N_\mathrm{T}N_\mathrm{R}}}{\prod_{i=1}^{N_\mathrm{T}} |\theta_i(l,\hat{l})|^{2N_\mathrm{R}}}, \quad \sigma_n^2 \to 0, \tag{5.94}$$

a result that has been derived previously in an alternative fashion in [BV01, SFG02].

We also note that —due to (5.84)— the height of the MSDD PEP error floor is equal to the PEP of coherent detection (cf. Section 5.4.1.3) with $\sigma_n^2 = 1/\left(\left[\left(\mathbf{\Psi}^\mathrm{t}\right)^{-1}\right]_{n,n} - 1\right)$ even for small ESNRs, i.e. where the approximation (5.92) is inaccurate.

5.4.2.2.3 Special Case and DSTM Design

Cyclic DSTM: It can be seen from (5.88) that for cyclic DSTM the PEP of any of the considered detection schemes associated with some error event ($\boldsymbol{V}^{(l)} \to \boldsymbol{V}^{(\hat{l})}$) is asymptotically equal to the product of the N_T corresponding PSK–PEPs of error events ($v_{i,i}^{(l)} \to v_{i,i}^{(\hat{l})}$), $1 \leq i \leq N_\mathrm{T}$, taken to the power of N_R and multiplied by a correction factor $\binom{2N_\mathrm{T}N_\mathrm{R}-1}{N_\mathrm{T}N_\mathrm{R}} \geq N_\mathrm{T}N_\mathrm{R}$. This reflects the intuitively reasonable result that a cyclic DSTM constellation achieves maximal diversity if and only if the same PSK symbol is not transmitted from the same antenna in any two different DSTM symbols.

DSTM Design: From (5.88) one can also see that the PEPs of the dominant single–symbol error events depend on $\boldsymbol{V}^{(l)}$ and $\boldsymbol{V}^{(\hat{l})}$ only through the term $\prod_{i=1}^{N_\mathrm{T}} |\theta_i(l,\hat{l})|$. The closely related "diversity product" $\frac{1}{2}\min_{1\leq l<\hat{l}\leq L}\left\{\prod_{i=1}^{N_\mathrm{T}} |\theta_i(l,\hat{l})|^{1/N_\mathrm{T}}\right\}$ constitutes the primary optimization criterion in the design of (full–rank) DSTM codes assuming CDD or coherent detection, cf. [SHHS01, HS00]. Thus, our analysis clearly shows that DSTM codes designed to yield optimal performance under CDD and coherent detection are also optimal under MSDD and DFDD. A common assumption that —to our knowledge— has not been proven before.

5.4.2.3 Asymptotic Performance for $N \to \infty$

The above description of the PEPs furthermore facilitates an elegant analysis of the asymptotic performance of MSDD and DFDD for arbitrary DSTM constellations and fading channels and for $N \to \infty$. We note that for $N \to \infty$ "edge effects" in MSDD, i.e. the increase in error probabilities in positions near the edges of the observation window mentioned in Section 2.4.2.1, are negligible and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathrm{PEP}_n = \lim_{N\to\infty} \mathrm{PEP}_{\lceil N/2 \rceil}. \tag{5.95}$$

A similar statement is true for the ESNR. Hence, we will in the following consider $\mathrm{PEP}_{\lceil N/2 \rceil}$ and PEP_N as measure of asymptotic performance of MSDD and DFDD, respectively.

Due to the relation between MSDD and linear interpolation (cf. Section 2.4.2.1) we can make use of the results of [PK88] on infinite–order interpolation. Consequently, $m_{\lceil N/2\rceil,\lceil N/2\rceil}$ converges to

$$\lim_{N\to\infty} m_{\lceil N/2\rceil,\lceil N/2\rceil} = T\int_{-1/(2T)}^{1/(2T)} \frac{1}{\Psi^{\mathrm{t}}(f)+\sigma_n^2}\,\mathrm{d}f, \tag{5.96}$$

where $\Psi^{\mathrm{t}}(f)$ [cf. (5.27)] denotes the PSD of the effective fading process. Provided that the fading PSD is band–limited, i.e. $\Psi^{\mathrm{t}}(f)=0$, for $B_{h,\mathrm{eff}} < |f| \le 1/(2T)$, we directly obtain

$$= T\int_{-B_{h,\mathrm{eff}}}^{B_{h,\mathrm{eff}}} \frac{1}{\Psi^{\mathrm{t}}(f)+\sigma_n^2}\,\mathrm{d}f \;+\; \frac{1-2B_{h,\mathrm{eff}}T}{\sigma_n^2}. \tag{5.97}$$

Similarly, it was shown in the analysis of DFDD in [SGH99] using the results of [PK88] for infinite–order prediction that

$$\lim_{N\to\infty} m_{N,N} = \exp\left(-T\int_{-1/(2T)}^{1/(2T)} \log\left(\Psi^{\mathrm{t}}(f)+\sigma_n^2\right)\,\mathrm{d}f\right), \tag{5.98}$$

and for PSDs with $\Psi^{\mathrm{t}}(f)=0$, for $B_{h,\mathrm{eff}} < |f| \le 1/(2T)$,

$$= \exp\left(-T\int_{-B_{h,\mathrm{eff}}}^{B_{h,\mathrm{eff}}} \log\left(\Psi^{\mathrm{t}}(f)+\sigma_n^2\right)\,\mathrm{d}f\right)\cdot\left(\frac{1}{\sigma_n^2}\right)^{1-2B_{h,\mathrm{eff}}T}. \tag{5.99}$$

5.4.2.3.1 Very High SNR ($\sigma_n^2 \to 0$)

$\Psi^{\mathrm{t}}(f)=0$, for $B_{h,\mathrm{eff}} \le |f| \le 1/(2T)$ with $B_{h,\mathrm{eff}} < 1/(2T)$: Assuming a general band–limited PSD with $B_{h,\mathrm{eff}} < 1/(2T)$ we obtain from (5.84), (5.97), and (5.99) the asymptotic ESNR

$$\text{MSDD:}\quad \lim_{N\to\infty} \rho_{\lceil N/2\rceil} \asymp \frac{1-2B_{h,\mathrm{eff}}T}{\sigma_n^2},\quad \sigma_n^2\to 0, \tag{5.100}$$

$$\text{DFDD:}\quad \lim_{N\to\infty} \rho_N \asymp x_{\mathrm{PSD}}\left(\sigma_n^2\right)^{-(1-2B_{h,\mathrm{eff}}T)},\quad \sigma_n^2\to 0, \tag{5.101}$$

where x_{PSD} is a constant factor that depends on the PSD, e.g. $x_{\mathrm{PSD}} = (2\pi B_{h,\mathrm{eff}}T/\mathrm{e})^{-2B_{h,\mathrm{eff}}T}$ (cf. [SGH99]) for Clarke's fading model or $x_{\mathrm{PSD}} = (2B_{h,\mathrm{eff}}T)^{-2B_{h,\mathrm{eff}}T}$ for the rectangular PSD.

While (5.101) is known from [SGH99], the MSDD result (5.100) is new and shows that, similar to DFDD, MSDD does not —in the limit of $N\to\infty$— suffer from an irreducible error floor if the PSD $\Psi^{\mathrm{t}}(f)$ of the effective fading process is zero over a non–countable set of frequencies. Thus, if the sampling theorem of the effective fading process is over–fulfilled an error floor can always be avoided regardless of the particular shape of the PSD.

$\Psi^{\mathrm{t}}(f) \neq 0$, for $0 \leq |f| \leq 1/(2T)$: If, on the other hand, $\Psi^{\mathrm{t}}(f) \neq 0$, for $0 \leq |f| \leq 1/(2T)$, e.g. if the fading process has a "Gaussian", "1st–order Butterworth" or "2nd–order Butterworth" PSD (cf. Table 4.1), we directly obtain from (5.84), (5.96) and (5.98) that

$$\text{MSDD:} \quad \lim_{N\to\infty,\sigma_n^2\to 0} \rho_{\lceil N/2\rceil} = T \int_{-1/(2T)}^{1/(2T)} \left(\Psi^{\mathrm{t}}(f)\right)^{-1} \mathrm{d}f \tag{5.102}$$

$$\text{DFDD:} \quad \lim_{N\to\infty,\sigma_n^2\to 0} \rho_{N} = \exp\left(-T \int_{-1/(2T)}^{1/(2T)} \log\left(\Psi^{\mathrm{t}}(f)\right)\, \mathrm{d}f\right). \tag{5.103}$$

From this it can be inferred that if $\Psi^{\mathrm{t}}(f) \neq 0$, for $0 \leq |f| \leq 1/(2T)$, an error floor is inevitable with MSDD and DFDD unless $\left(\Psi^{\mathrm{t}}(f)\right)^{-1}$ and $\log(\Psi^{\mathrm{t}}(f))$ are not integrable, respectively.[8]

Discussion: In summary, we can conclude that for integrable $\left(\Psi^{\mathrm{t}}(f)\right)^{-1}$ and $\log(\Psi^{\mathrm{t}}(f))$ the circumstances under which there is an irreducible error floor for MSDD are the same as for DFDD, cf. (5.99) or [SGH99, Section VI-A]. However, the height of the error floor of MSDD may be lower by several orders of magnitude than that of DFDD.

Provided that $\Psi^{\mathrm{t}}(f) = 0$ over a non–countable set of frequencies, e.g. if the sampling theorem for the fading process is over–fulfilled, we obtain the following asymptotic expressions for the PEPs of MSDD, DFDD and coherent detection by plugging $\rho_{\mathrm{eff}} = 1/\sigma_n^2$, (5.100), and (5.101) into (5.88), respectively,

$$\text{coherent: } \mathrm{PEP}\left(\boldsymbol{V}^{(l)} \to \boldsymbol{V}^{(\hat{l})}\right) \asymp \frac{\binom{2N_{\mathrm{T}}N_{\mathrm{R}}-1}{N_{\mathrm{T}}N_{\mathrm{R}}}}{\prod_{i=1}^{N_{\mathrm{T}}} |\theta_i(l,\hat{l})|^{2N_{\mathrm{R}}}} \left(\sigma_n^2\right)^{N_{\mathrm{T}}N_{\mathrm{R}}}, \quad \sigma_n^2 \to 0, \tag{5.104}$$

$$\text{MSDD:} \qquad \lim_{N\to\infty} \mathrm{PEP}_{\lceil N/2\rceil} \asymp \frac{\binom{2N_{\mathrm{T}}N_{\mathrm{R}}-1}{N_{\mathrm{T}}N_{\mathrm{R}}}}{\prod_{i=1}^{N_{\mathrm{T}}} |\theta_i(l,\hat{l})|^{2N_{\mathrm{R}}}} \left(\frac{\sigma_n^2}{1-2B_{h,\mathrm{eff}}T}\right)^{N_{\mathrm{T}}N_{\mathrm{R}}}, \quad \sigma_n^2 \to 0, \tag{5.105}$$

$$\text{DFDD, CDD:} \quad \lim_{N\to\infty} \mathrm{PEP}_{N} \asymp \frac{\binom{2N_{\mathrm{T}}N_{\mathrm{R}}-1}{N_{\mathrm{T}}N_{\mathrm{R}}}}{\prod_{i=1}^{N_{\mathrm{T}}} |\theta_i(l,\hat{l})|^{2N_{\mathrm{R}}}} x_{\mathrm{PSD}} \left(\sigma_n^2\right)^{(1-2B_{h,\mathrm{eff}}T)N_{\mathrm{T}}N_{\mathrm{R}}}, \quad \sigma_n^2 \to 0, \tag{5.106}$$

where x_{PSD} is a constant factor that depends on the PSD, e.g. $x_{\mathrm{PSD}} = (2\pi B_{h,\mathrm{eff}}T/\mathrm{e})^{-2B_{h,\mathrm{eff}}TN_{\mathrm{T}}N_{\mathrm{R}}}$ (cf. [SGH99]) for Clarke's fading model or $x_{\mathrm{PSD}} = (2B_{h,\mathrm{eff}}T)^{-2B_{h,\mathrm{eff}}TN_{\mathrm{T}}N_{\mathrm{R}}}$ for the rectangular PSD.

Based on (5.104)–(5.106) one can now easily compare the asymptotic performance of MSDD and DFDD to that of coherent detection with perfect CSI. Apparently, MSDD with infinite observation window length only suffers from a loss in *SNR* by a factor of $(1-2B_{h,\mathrm{eff}}T)$, whereas the performance of DFDD is —as noted previously in [DB06]— more severely degraded due to a loss in *diversity* by a factor of $(1-2B_{h,\mathrm{eff}}T)$, i.e. the error rate of DFDD when plotted over the SNR ultimately diverges from those of coherent and MSD detection.

[8] E.g. the PSD $\Psi_{\mathrm{c}}^{\mathrm{t}}(f) = (\alpha+1)T(2|f|T)^{\alpha}$ for $|f| \leq 1/(2T)$ and zero otherwise with $\alpha \geq 1$ [PK88] would lead to an error floor for DFDD but not for MSDD, since $\log(\Psi^{\mathrm{t}}(f))$ is integrable, whereas $\left(\Psi^{\mathrm{t}}(f)\right)^{-1}$ is not.

5.4.2.3.2 Clarke's Fading Model In this section, we evaluate (5.97) and (5.99) for finite σ_n^2 for the popular special case of Clarke's fading model, cf. (4.22). In particular, after tedious but straightforward integration we obtain for MSDD

$$\lim_{N\to\infty} m_{\lceil N/2\rceil,\lceil N/2\rceil} = \frac{1}{\sigma_n^2}\left[1+\frac{1}{\sigma_n^2}\left(\frac{1-\frac{2}{\pi}\arcsin(\pi B_{h,\text{eff}}T\sigma_n^2)}{\sqrt{1-(\pi B_{h,\text{eff}}T\sigma_n^2)^2}}-1\right)\right], \tag{5.107}$$

and for DFDD

$$\lim_{N\to\infty} m_{N,N} = \tag{5.108}$$
$$\left(\frac{1}{\sigma_n^2}\right)^{1-2B_{h,\text{eff}}T}(2\pi B_{h,\text{eff}}T)^{2B_{h,\text{eff}}T}\exp\left(\frac{\sqrt{1-(\pi B_{h,\text{eff}}T\sigma_n^2)^2}\left(1-\frac{2}{\pi}\arcsin\left(\pi B_{h,\text{eff}}T\sigma_n^2\right)\right)-1}{\sigma_n^2}\right).$$

Eqs. (5.107) and (5.108) can be combined with (5.84) and (5.85), (5.72) to obtain closed–form results for the ESNR and the PEP, respectively. We note that while $\lim_{N\to\infty} m_{N,N}$ was also considered in [SGH99, SL02, DB06], (5.108) is a new result as numerical integration was used in [SGH99, SL02, DB06] to solve (5.99).

5.4.3 Pairwise Error Probabilities for CG–MSDD

In Section 3.2, we presented an MSDD algorithm based on combinatorial geometry, so–called CG–MSDD, and showed that its complexity is exponential in the rank N_λ of the temporal correlation matrix $\mathbf{\Psi}^{\text{t}}$ of the fading channel. While this is promising, we note that especially for small observation window lengths N and moderately fast fading the rank of $\mathbf{\Psi}^{\text{t}}$ is often close to N. Hence, CG–MSDD is computationally inefficient in these cases. At the same time the number $\tilde{N}_\lambda$ of dominant eigenvalues $\lambda_{\mathbf{\Psi}^{\text{t}},i}$, $1\le i\le\tilde{N}_\lambda$, is usually significantly smaller than the rank N_λ of $\mathbf{\Psi}^{\text{t}}$.[9] It therefore appears reasonable to approximate $\mathbf{\Psi}^{\text{t}}$ using only its $\tilde{N}_\lambda$ largest eigenvalues, i.e. by means of the rank–$\tilde{N}_\lambda$ matrix

$$\mathbf{\Psi}^{\text{t}}_{\text{CG}} \triangleq \sum_{i=1}^{\tilde{N}_\lambda}\lambda_{\mathbf{\Psi}^{\text{t}},i}\left[\boldsymbol{U}_{\mathbf{\Psi}^{\text{t}}}\right]_{:,i}\left(\left[\boldsymbol{U}_{\mathbf{\Psi}^{\text{t}}}\right]_{:,i}\right)^{\mathsf{T}} \tag{5.109}$$

with $\left[\boldsymbol{U}_{\mathbf{\Psi}^{\text{t}}}\right]_{:,i}$ denoting the eigenvector of $\mathbf{\Psi}^{\text{t}}$ corresponding to the eigenvalue $\lambda_{\mathbf{\Psi}^{\text{t}},i}$, and to let CG–MSDD operate on this approximate $\mathbf{\Psi}^{\text{t}}_{\text{CG}}$ rather than the true $\mathbf{\Psi}^{\text{t}}$. I.e. we replace $\boldsymbol{Q}_{\text{c}}$ in Section 3.2.2 with its rank–$\tilde{N}_\lambda$ approximation [cf. (3.137) with (3.135)]

$$\tilde{\boldsymbol{Q}}_{\text{c}} \triangleq \bar{\boldsymbol{R}}_{\text{D}}\left[\sum_{i=1}^{\tilde{N}_\lambda}\frac{\lambda_{\mathbf{\Psi}^{\text{t}},i}}{\lambda_{\mathbf{\Psi}^{\text{t}},i}+\sigma_n^2}\left[\boldsymbol{U}_{\mathbf{\Psi}^{\text{t}}}\right]_{:,i}\left(\left[\boldsymbol{U}_{\mathbf{\Psi}^{\text{t}}}\right]_{:,i}\right)^{\mathsf{T}}\right]\bar{\boldsymbol{R}}_{\text{D}}^{\mathsf{H}}. \tag{5.110}$$

[9] The behavior of the eigenvalues of a correlation matrix will be discussed in detail in Section 5.5.2.

Clearly, this brings about a metric mismatch, which we will subsequently refer to as *rank mismatch*. In order to gain insight into its effect on performance, we consider the PEPs of CG–MSDD under rank mismatch. It should be clear now that the situation at hand is in principle the same as that of Section 5.4.1, where a metric mismatch was caused by the decoders assumption of quasi–static fading. In both cases the decision metric is generally not matched optimally to the fading process. Since furthermore the fading process and the received signal are apparently the same in both cases, we can compute the PEPs of CG–MSDD as described in Section 5.4.1, merely replacing $\boldsymbol{M}$ in (5.40) with [cf. (3.134) and (3.135)]

$$\tilde{\boldsymbol{M}} \triangleq \frac{1}{\sigma_n^2}\left(\boldsymbol{I}_N - \sum_{i=1}^{\tilde{N}_\lambda} \frac{\lambda_{\boldsymbol{\Psi}^{\mathrm{t}},i}}{\lambda_{\boldsymbol{\Psi}^{\mathrm{t}},i}+\sigma_n^2}\left[\boldsymbol{U}_{\boldsymbol{\Psi}^{\mathrm{t}}}\right]_{:,i}\left(\left[\boldsymbol{U}_{\boldsymbol{\Psi}^{\mathrm{t}}}\right]_{:,i}\right)^{\mathsf{T}}\right). \tag{5.111}$$

5.4.4 Approximation of Symbol–Error Rate

An exact computation of the symbol–error rates (SER) is usually not feasible due to the complex structure of the decision regions. Thus, having derived expressions for the PEPs of various detection schemes, we will in this section present approximations for the corresponding SERs, that are based on truncated union bounds over the PEPs of the dominant error events and therefore amenable to computationally very efficient evaluation. These approximations turn out to be quite tight.

5.4.4.1 (S–)MSDD and DF–MSDD

In Section 2.4.2.2 we argued based on the relationship between MSDD and linear MMSE interpolation (cf. Section 2.4.2.1) that decisions in different positions of the observation window should be unequally reliable. Therefore, we will first consider the SERs of the individual positions of the observation window and present approximations for the average SERs of (S–)MSDD and DF–MSDD later on.

In general the symbol–error rate SER_n for data symbols $\boldsymbol{V}_n$ in position n of the observation window can be upper bounded using the union bound averaged over all L^{N-1} relevant transmit sequences:

$$\mathrm{SER}_n \leq \frac{1}{L^{N-1}} \sum_{\forall \bar{\boldsymbol{S}}} \sum_{\forall \hat{\bar{\boldsymbol{S}}},\, \hat{\boldsymbol{V}}_n \neq \boldsymbol{V}_n} \mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big), \tag{5.112}$$

where the L^{N-1} $\bar{\boldsymbol{S}}$ and the $L^{N-1}-L$ $\hat{\bar{\boldsymbol{S}}}$ with $\hat{\boldsymbol{V}}_n \neq \boldsymbol{V}_n$ are formed from $\boldsymbol{V}_\nu$ and $\hat{\boldsymbol{V}}_\nu$, $1 \leq \nu \leq N-1$, respectively, as in (3.8).

Evaluation of (5.112) is computationally tractable only for small constellations and observation window sizes N. In order to obtain a simpler approximation for SER_n we examine the

PEP of (5.47) for the general case of the spatially correlated CFC more closely. Based on the definition

$$\boldsymbol{P} \triangleq \bar{\boldsymbol{S}}_{\mathrm{D}}^{\mathsf{H}} \bar{\boldsymbol{S}}_{\mathrm{D,c}} \tag{5.113}$$

it is shown in Appendix 5.A.5 that the $\mathrm{PEP}\big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\big)$ depends on $\bar{\boldsymbol{S}}$ and $\hat{\bar{\boldsymbol{S}}}$ only through the matrix

$$\boldsymbol{T}_{\mathrm{c}} \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \left(\boldsymbol{P}^{\mathsf{H}} \left(\bar{\boldsymbol{S}}_{\mathrm{D}}^{\mathsf{H}} \hat{\bar{\boldsymbol{S}}}_{\mathrm{D}} \left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}} \right) \hat{\bar{\boldsymbol{S}}}_{\mathrm{D}}^{\mathsf{H}} \bar{\boldsymbol{S}}_{\mathrm{D}} - \left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}} \right) \right) \boldsymbol{P} \right) \right), \tag{5.114}$$

with $\boldsymbol{T}_{\mathrm{c}}$ and $\boldsymbol{M}$ as defined in (5.37) and (5.25), respectively.

5.4.4.1.1 Group Constellations For L being a power of two, cyclic and dicyclic codes fully represent full–rank unitary group constellations, cf. [Hug03]. In both cases $\boldsymbol{P}$ is a sparse matrix with a single one in each row. Whereas $\boldsymbol{P}$ is constant and independent of $\bar{\boldsymbol{S}}_{\mathrm{D}}$ for cyclic constellations, there are 2^{N-1} different matrices $\boldsymbol{P}$ in case of dicyclic constellations accounting for the two different types of matrices appearing in the dicyclic–DSTM signal constellation, cf. Section 2.1.1.2. This means that the set of L^{N-1} matrices $\bar{\boldsymbol{S}}$ can be partitioned into $N_{\boldsymbol{P}}$ equivalence classes $\bar{\boldsymbol{\mathcal{S}}}_{\boldsymbol{P}}^{k}$ with respect to $\boldsymbol{P}$, $1 \leq k \leq N_{\boldsymbol{P}}$, where $N_{\boldsymbol{P}} = 1$ and $N_{\boldsymbol{P}} = 2^{N-1}$ for cyclic and dicyclic constellations, respectively. Furthermore, since $\bar{\boldsymbol{S}}_{\mathrm{D}}^{\mathsf{H}} \hat{\bar{\boldsymbol{S}}}_{\mathrm{D}} = \mathrm{diag}\big\{ \boldsymbol{S}_1^{\mathsf{H}} \hat{\boldsymbol{S}}_1, \ldots, \boldsymbol{S}_N^{\mathsf{H}} \hat{\boldsymbol{S}}_N \big\}$ with $\boldsymbol{S}_n^{\mathsf{H}} \hat{\boldsymbol{S}}_n \in \boldsymbol{\mathcal{V}}$, $1 \leq n \leq N$, for group constellations, the inner sum of (5.112) is independent of which class representative $\bar{\boldsymbol{S}} \in \bar{\boldsymbol{\mathcal{S}}}_{\boldsymbol{P}}^{k}$ is chosen. Forming a set $\bar{\boldsymbol{\mathcal{S}}}$ of $N_{\boldsymbol{P}}$ representatives one for each of the $N_{\boldsymbol{P}}$ different equivalence classes $\bar{\boldsymbol{\mathcal{S}}}_{\boldsymbol{P}}^{k}$, $1 \leq k \leq N_{\boldsymbol{P}}$, we can thus limit the outer summation to only $N_{\boldsymbol{P}} \ll L^{N-1}$ terms ($N_{\boldsymbol{P}} = 1$ or $N_{\boldsymbol{P}} = 2^{N-1}$).

To reduce the number of terms of the inner sum in (5.112), we propose to take only the dominating instead of all $L^{N-1} - L$ relevant error events into account. As motivated at the beginning of Section 5.4.2, we consider single–transmit–symbol errors as dominating error events, i.e. error events, where $\hat{\bar{\boldsymbol{S}}}$ differs from $\bar{\boldsymbol{S}}$ in only one DSTM symbol $\hat{\boldsymbol{S}}_n \neq \boldsymbol{S}_n$. In particular, we further restrict our attention to error events ($\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}$) that dominate the asymptotic performance of MSDD for $\sigma_n^2 \to 0$. Using the results of Section 5.4.2.2 on the behavior of the PEPs under spatially uncorrelated fading in very high SNR [cf. (5.88)] these are given by

$$\hat{\bar{\boldsymbol{\mathcal{S}}}}_{n,l} \triangleq \left\{ \hat{\bar{\boldsymbol{S}}} = \left[\boldsymbol{S}_1^{\mathsf{H}}, \ldots, \boldsymbol{S}_{n-1}^{\mathsf{H}}, \hat{\boldsymbol{S}}_n^{\mathsf{H}}, \boldsymbol{S}_{n+1}^{\mathsf{H}}, \ldots, \boldsymbol{S}_N^{\mathsf{H}} \right]^{\mathsf{H}} \middle| \hat{\boldsymbol{S}}_n = \big(\boldsymbol{V}^{(\hat{l})} \big)^{\mathsf{H}} \boldsymbol{V}^{(l)} \boldsymbol{S}_n,\ \hat{l} \in \hat{\mathcal{L}}_l \right\}, \tag{5.115}$$

with

$$\hat{\mathcal{L}}_l \triangleq \left\{ \hat{l} = \underset{\tilde{l} \in \{1,\ldots,L\} \setminus \{l\}}{\mathrm{argmin}} \left\{ \prod_{i=1}^{N_{\mathrm{T}}} \left| \theta_i\big(l, \tilde{l}\big) \right| \right\} \right\} \tag{5.116}$$

and $\theta_i\big(l, \tilde{l}\big)$ as eigenvalues from the eigenvalue decomposition of $\big(\boldsymbol{V}^{(\hat{l})} \big)^{\mathsf{H}} \boldsymbol{V}^{(l)} - \boldsymbol{I}_{N_{\mathrm{T}}}$, cf. (5.65).

With this, and the observation that due to differential encoding a data–symbol error occurs in position n if either the transmit–symbol in position n or in position $n+1$ is detected erroneously we can approximate SER_n by

$$\mathrm{SER}_n \approx \frac{1}{N_{\boldsymbol{P}}} \sum_{\forall \bar{\boldsymbol{S}} \in \bar{\boldsymbol{\mathcal{S}}}} \left(\sum_{\forall \hat{\bar{\boldsymbol{S}}} \in \hat{\bar{\boldsymbol{\mathcal{S}}}}_{n,l}} \mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big) + \sum_{\forall \hat{\bar{\boldsymbol{S}}} \in \hat{\bar{\boldsymbol{\mathcal{S}}}}_{n+1,l}} \mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big) \right) , \quad 1 \leq n \leq N-1, \tag{5.117}$$

where the subscript l in the inner sums in governed by the choice of $\bar{\boldsymbol{S}}$ as lth element, $1 \leq l \leq N_{\boldsymbol{P}}$, of the set $\bar{\boldsymbol{\mathcal{S}}}$.

Consequently, the SER of general S–MSDD (cf. Section 2.4.2.2), which returns decisions on symbols $\boldsymbol{V}_n$, $\lfloor (N-N')/2 \rfloor + 1 \leq n \leq N - \lceil (N-N')/2 \rceil$, can be approximated via

$$\mathrm{SER} \approx \frac{1}{N_{\boldsymbol{P}} N'} \sum_{\forall \bar{\boldsymbol{S}} \in \bar{\boldsymbol{\mathcal{S}}}} \sum_{n=\left\lfloor \frac{N-N'}{2} \right\rfloor + 1}^{N - \left\lceil \frac{N-N'}{2} \right\rceil} \left[\sum_{\forall \hat{\bar{\boldsymbol{S}}} \in \hat{\bar{\boldsymbol{\mathcal{S}}}}_{n,l}} \mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big) + \sum_{\forall \hat{\bar{\boldsymbol{S}}} \in \hat{\bar{\boldsymbol{\mathcal{S}}}}_{n+1,l}} \mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big) \right] . \tag{5.118}$$

Similarly, we approximate the SER of DF–MSDD (cf. Section 2.4.4), which returns decisions on symbols $\boldsymbol{V}_n$, $N-1-\kappa_{\mathrm{U}}^{\mathrm{DF}} \leq n \leq N-1-\kappa_{\mathrm{L}}^{\mathrm{DF}}$, using

$$\mathrm{SER}_{\text{DF–MSDD}} \approx \frac{1}{N_{\boldsymbol{P}}(\kappa_{\mathrm{U}}^{\mathrm{DF}} - \kappa_{\mathrm{L}}^{\mathrm{DF}} + 1)} \sum_{\forall \bar{\boldsymbol{S}} \in \bar{\boldsymbol{\mathcal{S}}}} \sum_{n=N-1-\kappa_{\mathrm{U}}^{\mathrm{DF}}}^{N-1-\kappa_{\mathrm{L}}^{\mathrm{DF}}} \left[\sum_{\forall \hat{\bar{\boldsymbol{S}}} \in \hat{\bar{\boldsymbol{\mathcal{S}}}}_{n,l}} \mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big) + \sum_{\forall \hat{\bar{\boldsymbol{S}}} \in \hat{\bar{\boldsymbol{\mathcal{S}}}}_{n+1,l}} \mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big) \right] , \tag{5.119}$$

i.e. under the assumption of perfect feedback.

5.4.4.1.2 Non–Group Constellations For non–group constellations such as orthorgonal and Cayley codes a simplification of (5.112) based on equivalence classes with respect to $\boldsymbol{P}$ is not possible. Furthermore —contrary to the case of DSTM from group codes— it is not guaranteed for non–group DSTM codes that the combination $\boldsymbol{\mathcal{H}}_n \triangleq \{\boldsymbol{S}_{n-1}, \hat{\boldsymbol{S}}_{n-1}, \boldsymbol{S}_n, \hat{\boldsymbol{S}}_n, \boldsymbol{S}_{n+1}, \hat{\boldsymbol{S}}_{n+1}\}$ with $\hat{\boldsymbol{S}}_{n-1} = \boldsymbol{S}_{n-1}$, $\hat{\boldsymbol{S}}_n = \hat{\boldsymbol{V}}_n^{\mathsf{H}} \boldsymbol{V}_n \big(\boldsymbol{V}^{(\hat{l})}\big)^{\mathsf{H}} \boldsymbol{V}^{(l)} \boldsymbol{S}_n$, $\hat{\boldsymbol{S}}_{n+1} = \boldsymbol{S}_{n+1}$, and $\hat{\boldsymbol{V}}_n = \boldsymbol{V}^{(\hat{l})}$, $\hat{l} \in \hat{\mathcal{L}}_l$ [cf. (5.116)] is admissible, i.e. that there is a $\hat{\boldsymbol{V}}_{n-1} \in \boldsymbol{\mathcal{V}}$ such that $\hat{\boldsymbol{S}}_n = \hat{\boldsymbol{V}}_{n-1} \boldsymbol{S}_{n-1}$. Hence, the minimization criterion in (5.116) is not *necessarily* an appropriate indicator for the dominant error events for every realization of $\bar{\boldsymbol{S}}$. However, we found from numerical evaluations that the SER of non–group codes is approximated quite accurately using (5.117) and (5.118) with L candidates $\bar{\boldsymbol{S}} = \big[\boldsymbol{I}_{N_{\mathrm{T}}}, (\boldsymbol{V}^{(l)})^{\mathsf{H}}, \boldsymbol{I}_{N_{\mathrm{T}}}, (\boldsymbol{V}^{(l)})^{\mathsf{H}}, \ldots\big]^{\mathsf{H}}$, $1 \leq l \leq L$, for which $\boldsymbol{\mathcal{H}}_n$ is admissible, and with $\hat{\bar{\boldsymbol{S}}}$ from the corresponding set $\hat{\bar{\boldsymbol{\mathcal{S}}}}_{n,l}$, $1 \leq n \leq N-1$, as defined in (5.115) with (5.116).

5.4.4.2 DFDD

For a fair comparison with the SER approximation of MSDD, we approximate the SER of DFDD by forming a truncated union bound over the set $\hat{\bar{\boldsymbol{\mathcal{S}}}}_{N,l}$ [cf. (5.115)] of dominant error events of DFDD with perfect feedback and averaging over $\boldsymbol{V}^{(l)} \in \boldsymbol{\mathcal{V}}$. Contrary to DF–MSDD, error propagation is more pronounced in DFDD and accounted for by a factor of two. I.e. we consider [SL02]

$$\mathrm{SER} \approx \frac{2}{L} \sum_{l=1}^{L} \sum_{\forall \hat{l} \in \hat{\mathcal{L}}_l} \mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big), \tag{5.120}$$

where the PEPs are computed using the expressions for MSDD with

$$\bar{\boldsymbol{S}} = \big[\boldsymbol{S}_1^{\mathsf{H}}, \ldots, \boldsymbol{S}_{N-1}^{\mathsf{H}}, (\boldsymbol{V}^{(l)}\boldsymbol{S}_{N-1})^{\mathsf{H}}\big]^{\mathsf{H}}, \quad 1 \le l \le L, \tag{5.121}$$

$$\hat{\bar{\boldsymbol{S}}} = \big[\boldsymbol{S}_1^{\mathsf{H}}, \ldots, \boldsymbol{S}_{N-1}^{\mathsf{H}}, (\boldsymbol{V}^{(\hat{l})}\boldsymbol{S}_{N-1})^{\mathsf{H}}\big]^{\mathsf{H}}, \quad \hat{l} \in \hat{\mathcal{L}}_l, \tag{5.122}$$

and $\hat{\mathcal{L}}_l$ as defined in (5.116). Furthermore, averaging with respect to $\boldsymbol{V}^{(l)} \in \boldsymbol{\mathcal{V}}$ can be omitted for group DSTM codes.

5.4.4.3 CDD

Due to the close relationship between CDD, MSDD and DFDD, the SER of CDD can be approximated using

$$\mathrm{SER} \approx \frac{1}{L} \sum_{l=1}^{L} \sum_{\forall \hat{l} \in \hat{\mathcal{L}}_l} \mathrm{PEP}\Big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\Big), \tag{5.123}$$

with the PEPs computed using the expressions for MSDD with

$$\bar{\boldsymbol{S}} = \big[\boldsymbol{S}_1^{\mathsf{H}}, (\boldsymbol{V}^{(l)}\boldsymbol{S}_1)^{\mathsf{H}}\big]^{\mathsf{H}}, \quad 1 \le l \le L, \tag{5.124}$$

$$\hat{\bar{\boldsymbol{S}}} = \big[\boldsymbol{S}_1^{\mathsf{H}}, (\boldsymbol{V}^{(\hat{l})}\boldsymbol{S}_1)^{\mathsf{H}}\big]^{\mathsf{H}}, \quad \hat{l} \in \hat{\mathcal{L}}_l, \tag{5.125}$$

and $\hat{\mathcal{L}}_l$ as defined in (5.116). Note that contrary to DFDD with $N > 2$ there is no error propagation here. Again, averaging with respect to $\boldsymbol{V}^{(l)} \in \boldsymbol{\mathcal{V}}$ can be omitted for group DSTM codes.

5.4.4.4 (Differentially) Coherent Detection

For a fair comparison with the above SER approximations, we approximate the SER of (differentially) coherent detection via

$$\mathrm{SER} \approx \frac{x_{\mathrm{e}}}{L} \sum_{l=1}^{L} \sum_{\forall \hat{l} \in \hat{\mathcal{L}}_l} \mathrm{PEP}\big(\boldsymbol{V}^{(l)} \to \boldsymbol{V}^{(\hat{l})}\big), \tag{5.126}$$

with $x_{\mathrm{e}} = 1$ and $x_{\mathrm{e}} = 2$, for coherent and differentially coherent detection, respectively, $\hat{\mathcal{L}}_l$ as defined in (5.116), and averaging with respect to $\boldsymbol{V}^{(l)} \in \boldsymbol{\mathcal{V}}$ can be omitted for group DSTM codes.

5.4.5 Numerical Results

In this section, we complement our mathematical performance analysis using a number of illustrative numerical examples. At this, we concentrate on three major aspects. In Section 5.4.5.1 we compare ML MSDD with reference schemes such as conventional differential detection (CDD, $N = 2$), decision–feedback differential detection (DFDD) and (differentially) coherent detection with perfect channel state information (CSI), cf. Sections 2.4.1, 2.4.3 and 2.4.5, respectively. In the second part presented in Section 5.4.5.2, we turn to investigating the impact of the different (suboptimal) implementations of MSDD, e.g. the effect of the use of the Fano–type metric (cf. Section 3.1.2.3), the efficient symbol–search algorithms of Section 3.1.4, or rank mismatch in CG–MSDD. The performance of the various MSDD implementations as a function of the maximum admissible complexity per decoded symbol will follow in Section 5.5.3.

Unless explicitly stated otherwise, we present results obtained from the analytical expressions derived above and only resort to simulations when analytical expressions are not available or when the tightness of an approximation is to be illustrated. We refer to them as "analytical" and "simulation" results, respectively. Also, unless explicitly stated otherwise, the results are based on the frequently considered fading model due to Clarke [Cla68].

5.4.5.1 ML MSDD

In this section, we consider the performance of ML MSDD when solved by means of some optimal algorithm (e.g. MSDSD–FS) in comparison with the benchmark detection schemes.

Most of the results in this section will be presented in terms of the effective normalized fading bandwidth $B_{h,\mathrm{eff}}T$ [cf. (5.28)] and effective SNR ρ_{eff} [cf. (5.84)], because it allows for general statements regardless of system parameters such as numbers of antennas, transmission rate or individual DSTM constellations. Furthermore, we illustrate the tightness of the SER approximations given in Section 5.4.4 and present some results on the effects of spatial correlation in the MIMO channel.

5.4.5.1.1 Spatially Uncorrelated Channel

ESNR vs. Position n for Different N: Fig. 5.1 shows the effective SNR $\rho_{\mathrm{eff}} = \rho_n$ in the individual positions of the MSDD observation window for an effective fading bandwidth $B_{h,\mathrm{eff}}T = 0.1$, an SNR of $10\log_{10}(\sigma_n^{-2}) = 40\,\mathrm{dB}$ and different values of N. For comparison,

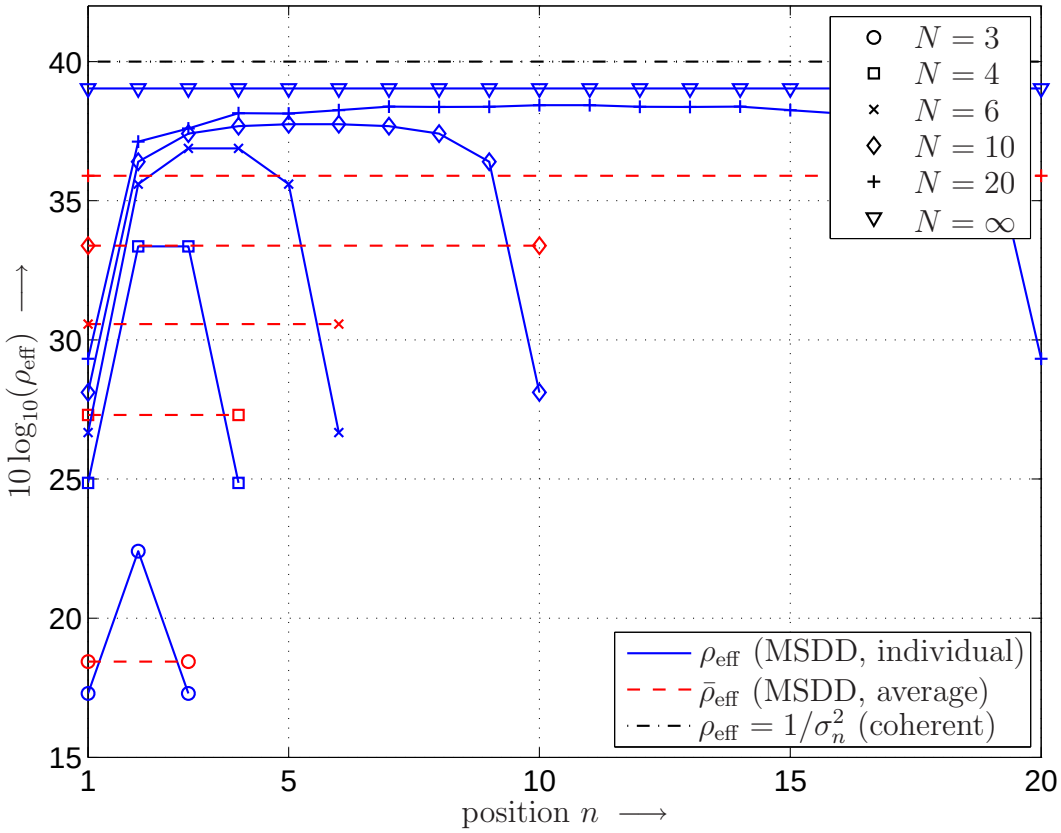

Figure 5.1: ESNR ρ_{eff} vs. position n of observation window for different values of N, $B_{h,\mathrm{eff}}T = 0.1$ and $10\log_{10}(\sigma_n^{-2}) = 40\,\mathrm{dB}$.

we included the true SNR $1/\sigma_n^2$, i.e. the effective SNR for coherent detection with perfect CSI and an "average" effective SNR $\bar{\rho}_{\mathrm{eff}}$ of MSDD as dash–dotted and dashed lines, respectively. Since (i) the average transmit–symbol error rate of MSDD is given by the arithmetic mean $1/N\sum_{n=1}^{N}\mathrm{PEP}_n$ of the error rates PEP_n in the individual positions of the observation window and (ii) the PEPs are asymptotically proportional to $\rho_{\mathrm{eff}}^{-N_{\mathrm{T}}N_{\mathrm{R}}}$ [cf. (5.88)], it is easy to see that $1/N\sum_{n=1}^{N}\mathrm{PEP}_n$ is asymptotically proportional to $\bar{\rho}_{\mathrm{eff}}^{-N_{\mathrm{T}}N_{\mathrm{R}}}$ with

$$\bar{\rho}_{\mathrm{eff}} \triangleq \sqrt[N_{\mathrm{T}}N_{\mathrm{R}}]{N\left(\sum_{n=1}^{N}\rho_{\mathrm{eff}}^{-N_{\mathrm{T}}N_{\mathrm{R}}}\right)^{-1}}. \tag{5.127}$$

Therefore, we included $\bar{\rho}_{\mathrm{eff}}$ with $N_{\mathrm{T}}N_{\mathrm{R}} = 1$ in the figure to reflect the average performance of MSDD, noting that $\bar{\rho}_{\mathrm{eff}}$ decreases if $N_{\mathrm{T}}N_{\mathrm{R}}$ increases.

First, it can be seen that the effective SNR in the center of the observation window is higher than at the edges, as was to be expected based on the relationship between MSDD and linear interpolation (cf. Section 2.4.2.1). More specifically, the individual effective SNRs ρ_n of MSDD are almost identical for $2 \leq n \leq N-1$, but deteriorate significantly at the

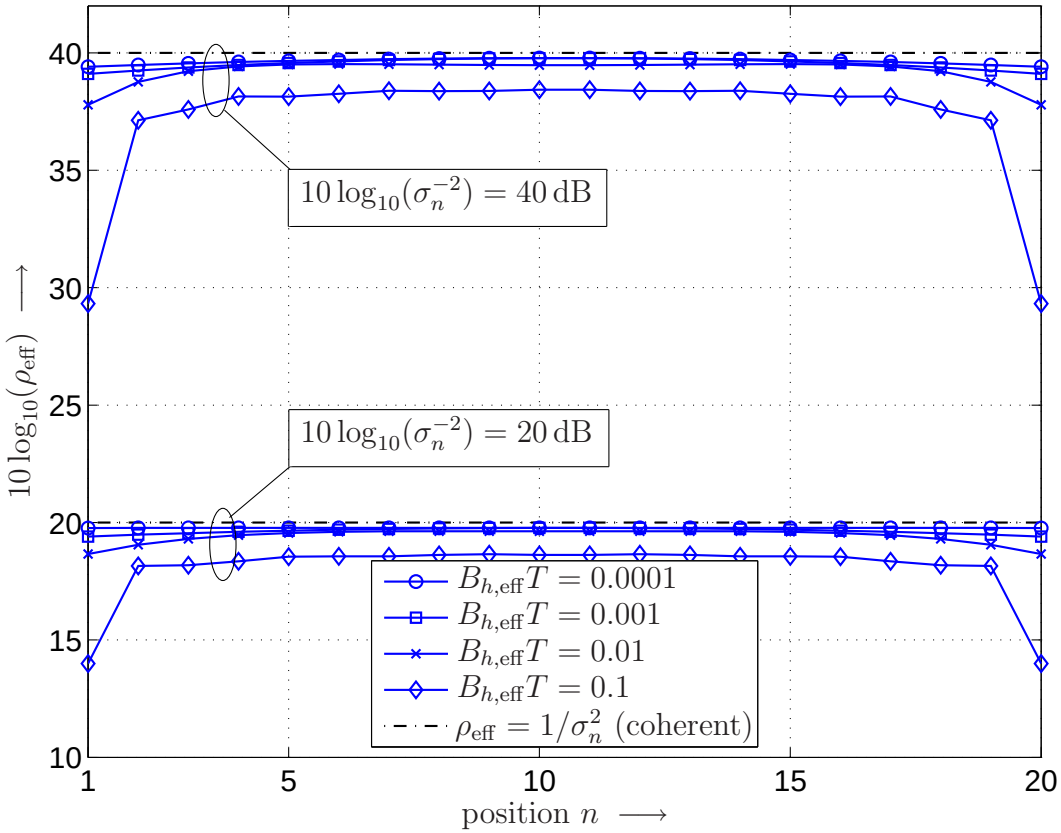

Figure 5.2: ESNR ρ_{eff} vs. position n of observation window for $N = 20$, different fading bandwidths $B_{h,\text{eff}}T$ and SNRs.

very edges of the observation window, i.e. for $n = 1$ and $n = N$, with differences of up to 10 dB when comparing non–edge and edge positions. As the detection error rates are by (5.88) asymptotically proportional to $\rho_{\text{eff}}^{-N_{\text{T}}N_{\text{R}}}$ the performance of MSDD is dominated by the relatively poor reliability of the decisions at the edges of the observation window. This is a strong argument for the use of S–MSDD with $N' = N - 3$, i.e. MSDD where the two decisions at the very edges of the observation window are discarded. With respect to DF–MSDD, this suggests to deploy DF–MSDD with some $\kappa_{\text{U}}^{\text{DF}} \geq 1$ and $\kappa_{\text{L}}^{\text{DF}} = 1$, such that $\kappa_{\text{U}}^{\text{DF}}$ decisions are returned in positions $N - \kappa_{\text{U}}^{\text{DF}} - 1, \ldots, N - 2$ of the observation window, i.e. on $\boldsymbol{V}_{N-1-\kappa} = \boldsymbol{V}[k - \kappa]$, $\kappa_{\text{U}}^{\text{DF}} \leq \kappa \leq 1$.

Second, Fig. 5.1 illustrates how the performance of MSDD with growing observation window size approaches that of coherent detection with perfect CSI, as the number N of samples, based on which the implicit channel estimation is performed, increases.

Finally, one can observe that —even in this rapid effective fading scenario— the use of moderate observation window sizes suffices to achieve close to optimal noncoherent detection performance, as the loss in power efficiency compared to the limiting case of $N \to \infty$ is 1.3 dB

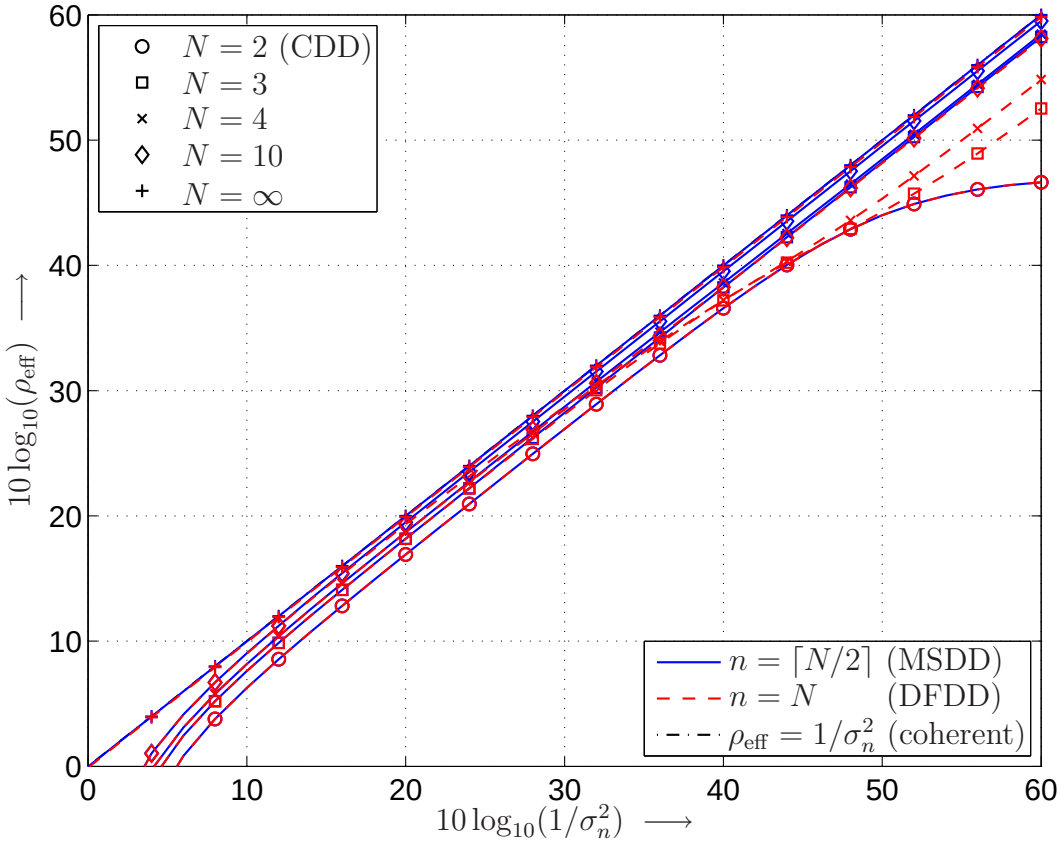

Figure 5.3: ESNR ρ_{eff} vs. the SNR $1/\sigma_n^2$ for $B_{h,\text{eff}}T = 0.001$, different values of N and $n \in \{\lceil N/2 \rceil, N\}$.

and 0.7 dB for $N = 10$ and $N = 20$ in center positions of the observation window, respectively.

ESNR vs. Position n for Different $B_{h,\text{eff}}T$: Fig. 5.2 shows the ESNR of the individual positions n of the MSDD observation window for different values of $B_{h,\text{eff}}T$ and σ_n^2 for $N = 20$. For comparison, we included the true SNR (dash–dotted line). Note that the ESNR of DFDD is equal to that of MSDD for $n = N$. One can clearly observe how —as $B_{h,\text{eff}}T$ increases— the ESNR deteriorates significantly at the edges of the observation window compared to its center, hinting at the large gains in power efficiency achievable with MSDD compared to DFDD (cf. also Fig. 5.11 for symbol error rates). For the non–edge positions $2 \leq n \leq N-2$ on the other hand there is only a noteworthy loss in power efficiency compared to coherent detection with perfect CSI for normalized effective fading bandwidths of the order of $B_{h,\text{eff}}T = 0.1$. Hence, we will in the following concentrate on comparing the ESNR for $n = \lceil N/2 \rceil$ and $n = N$ associating them with the performance of (S–)MSDD and DFDD, respectively.

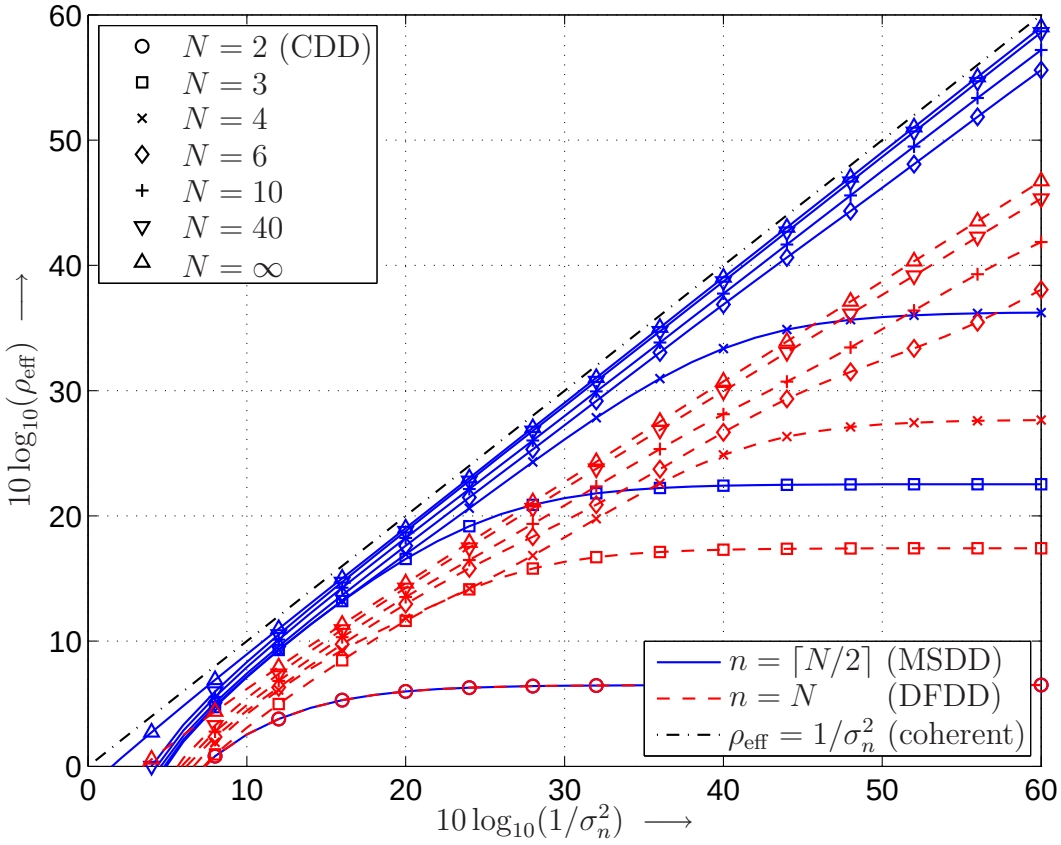

Figure 5.4: ESNR ρ_{eff} vs. the SNR $1/\sigma_n^2$ for $B_{h,\text{eff}}T = 0.1$, different values of N and $n \in \{\lceil N/2 \rceil, N\}$.

ESNR vs. SNR: Figs. 5.3 and 5.4 illustrate the loss in SNR of MSDD ($n = \lceil N/2 \rceil$, solid lines) and DFDD ($n = N$, dashed lines) with different values of N for $B_{h,\text{eff}}T = 0.001$ and $B_{h,\text{eff}}T = 0.1$, respectively.

In the case of $B_{h,\text{eff}}T = 0.001$ even very small observation windows such as $N = 3$ are sufficient for MSDD to achieve nearly optimum power efficiency and full diversity, whereas DFDD requires an observation window of size $N = 10$ to achieve the same performance and the effective SNR of CDD saturates at approximately 46 dB.

For $B_{h,\text{eff}}T = 0.1$ on the other hand the ESNR of CDD saturates well below 10 dB and even $N = 4$ leads to a saturation of ESNR in the considered SNR range for both MSDD and DFDD. Still, MSDD significantly outperforms DFDD for any observation window length $N > 2$. One can also clearly observe, how MSDD with $N \geq 10$ practically provides full diversity, while DFDD —even in the limit of $N \to \infty$— suffers from a significant diversity loss and achieves a maximum asymptotic slope of $1 - 2B_{h,\text{eff}}T = 0.8$ [cf. (5.106)].

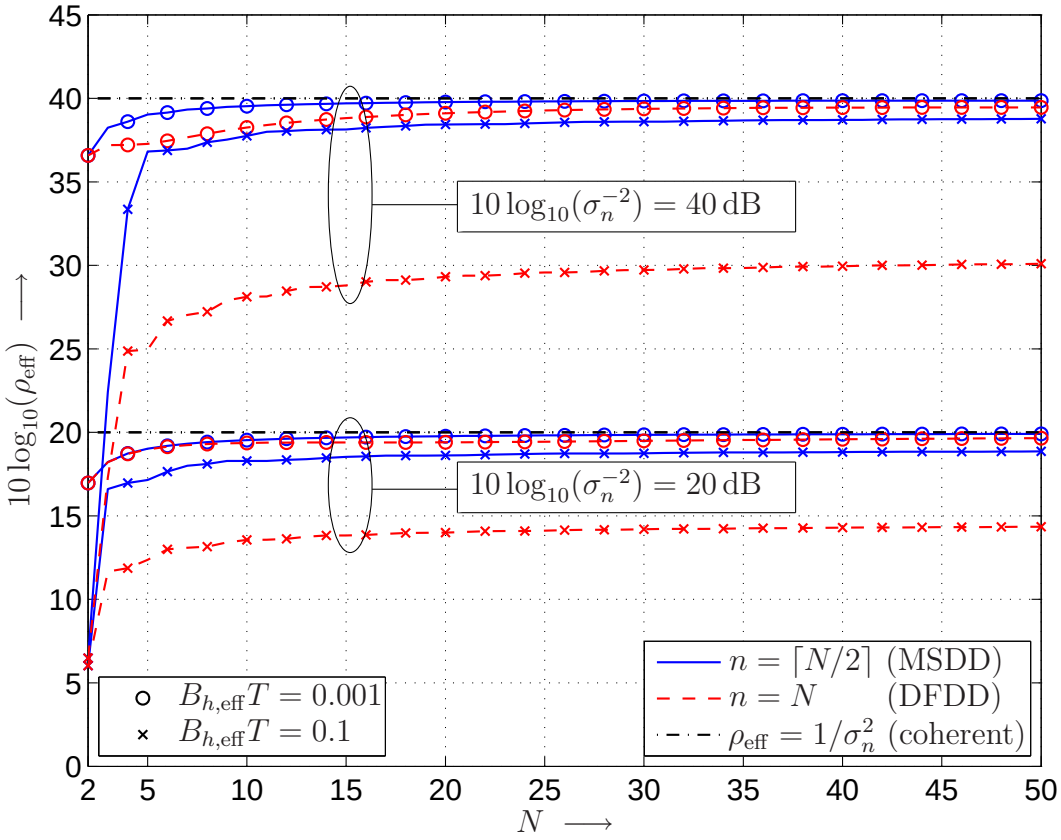

Figure 5.5: ESNR ρ_{eff} vs. the length N of the observation window for different fading bandwidths $B_{h,\text{eff}}T$ and SNRs $1/\sigma_n^2$, and $n \in \{\lceil N/2 \rceil, N\}$.

ESNR vs. Observation Window Length N: Fig. 5.5 shows the ESNR as a function of the observation window length N for $n \in \{\lceil N/2 \rceil, N\}$, different values of $B_{h,\text{eff}}T$ and SNRs.

In this figure, one can observe the rapid convergence of the ESNR as N increases, and that a significant gap even for large N remains only for DFDD in fast–fading environments, i.e. for $B_{h,\text{eff}}T$ of the order of 0.1.

ESNR vs. $B_{h,\text{eff}}T$ for $\sigma_n^2 = 0$: Figs. 5.6 and 5.7 show the effective SNR ρ_{eff} for $\sigma_n^2 = 0$ as a function of $B_{h,\text{eff}}T$ for $n \in \{\lceil N/2 \rceil, N\}$, different values of N and Clarke's and the Gaussian–PSD fading models, respectively. We compare ML MSDD [(5.91), solid and dashed lines] with the asymptotic result (5.262) for the suboptimal decoder whose coefficients are computed according to (5.253) (dash–dotted and dotted lines).

One can observe (i) that —as expected from our analysis in Section 5.4.2.2.1— the asymptotic slopes of the curves for both the optimal ML MSDD and the suboptimal (5.262) decoder are $-2(N-1)$ on the adopted double–logarithmic scale, (ii) that the curves for (5.262) are asymptotic (for $B_{h,\text{eff}}T \to 0$) lower bounds for the optimal curves, (iii) that positions around

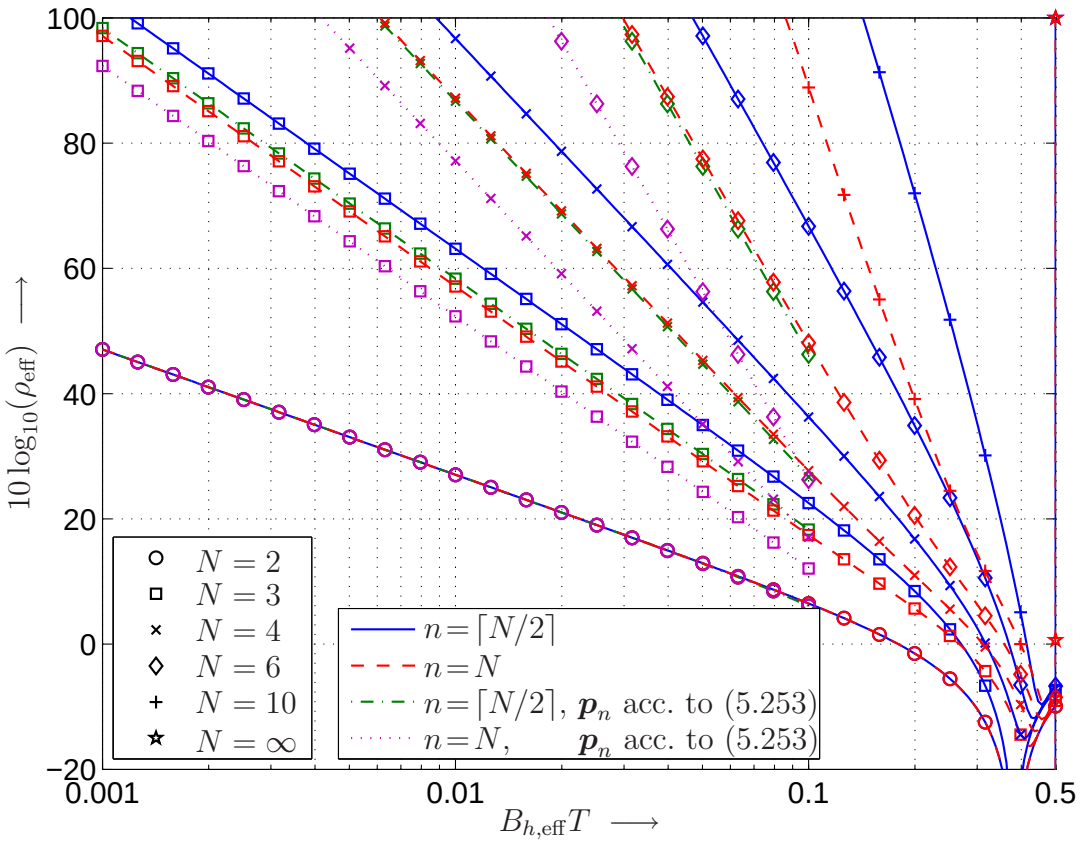

Figure 5.6: ESNR ρ_{eff} vs. $B_{h,\text{eff}}T$ for $\sigma_n^2 = 0$, Clarke's fading model, different values of N and $n \in \{\lceil N/2\rceil, N\}$.

the center of the observation window are also for $\sigma_n^2 = 0$ more reliably detected than those at the edges, and (iv) that $\lim_{[N,\sigma_n^2]\to[\infty,0]} \rho_{\text{eff}} \to \infty$ for $B_{h,\text{eff}}T < 0.5$ regardless of n.

Furthermore, Fig. 5.7 for ML MSDD and the Gaussian–PSD fading model supports our conjecture made in Section 5.4.2.2.1 that even for fading models with PSDs that are not band–limited the same kind of behavior, i.e. $\rho_{\text{eff}} \propto (B_{h,\text{eff}}T)^{-2(N-1)}$, is obtained.

PEP vs. SNR for Finite N: Fig. 5.8 compares the PEPs for one of the dominant error events for $n = \lceil N/2\rceil$ (solid lines), $n = N$ (dashed lines), coherent detection (dash–dotted lines) and the corresponding asymptotes according to (5.88) (dotted lines) for finite N. We adopted a cyclic (B)DSTM [HS00, Table I] with $N_{\text{T}} = 3$, $R = 1$, $N_{\text{R}} = 1$, $B_{h,\text{eff}}T = 0.1$ and $N \in \{4, 6, 10\}$. The horizontal and vertical lines mark the asymptotic ESNR $\lim_{\sigma_n^2\to 0} \rho_{\text{eff}} = \left[(\boldsymbol{\Psi}^{\text{t}})^{-1}\right]_{n,n} - 1$ and the corresponding error floor (5.92) for $N \in \{3, 4\}$, respectively.

This numerical example nicely confirms our analytical findings in Section 5.4.2.2 regarding the asymptotic performance of the detection schemes for $\sigma_n^2 \to 0$. In particular, one can observe the accuracy of the approximation of the error floor in (5.92). Only for relatively high error

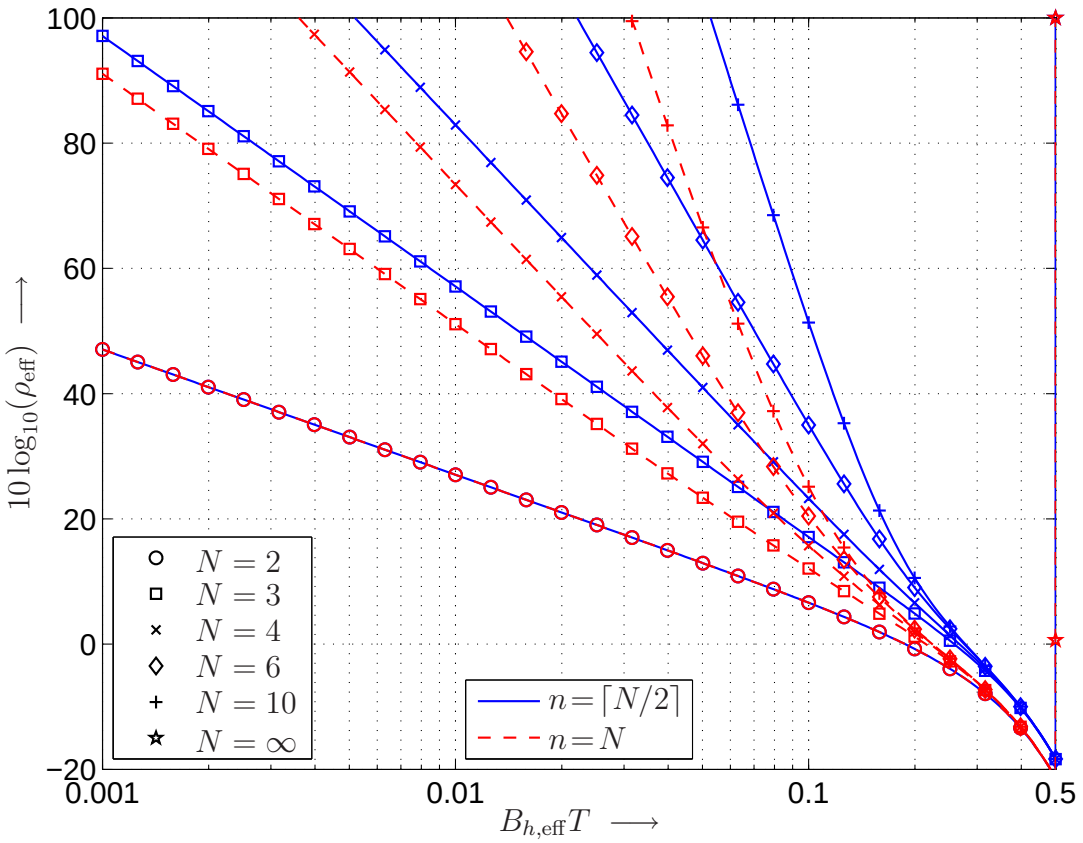

Figure 5.7: ESNR ρ_{eff} vs. $B_{h,\text{eff}}T$ for $\sigma_n^2 = 0$, the Gaussian–PSD fading model, different values of N and $n \in \{\lceil N/2\rceil, N\}$.

floors there is a small deviation, which is due to the fact that ρ_∞ is finite, while the asymptotic result (5.92) was derived based on the assumption that $\sigma_n^2 \to 0$ implies $\rho_\infty \to \infty$. This figure also shows that even for $B_{h,\text{eff}}T = 0.1$ quite small values of N are sufficient to lower the error floor below the PEP–region of interest.

PEP vs. SNR for $N \to \infty$: Fig. 5.9 compares the PEPs for one of the dominant error events for $n = \lceil N/2\rceil$ (solid lines), $n = N$ (dashed lines), coherent detection (dash–dotted lines), and the corresponding asymptotes according to (5.88) (dotted lines) for various $B_{h,\text{eff}}T$ and $N \to \infty$. As exemplary parameters we chose cyclic (B)DSTM [HS00, Table I] with $N_{\text{T}} = 5$, $R = 1$, and $N_{\text{R}} = 1$.

Fig. 5.9 clearly shows that for large $B_{h,\text{eff}}T$ even in the limit of $N \to \infty$ the PEP of DFDD ($n = N$) is severely affected by the diversity loss of $(1-2B_{h,\text{eff}}T)$, whereas the SNR loss entailed by MSDD ($n = \lceil N/2\rceil$) has a far less detrimental effect on the PEP.

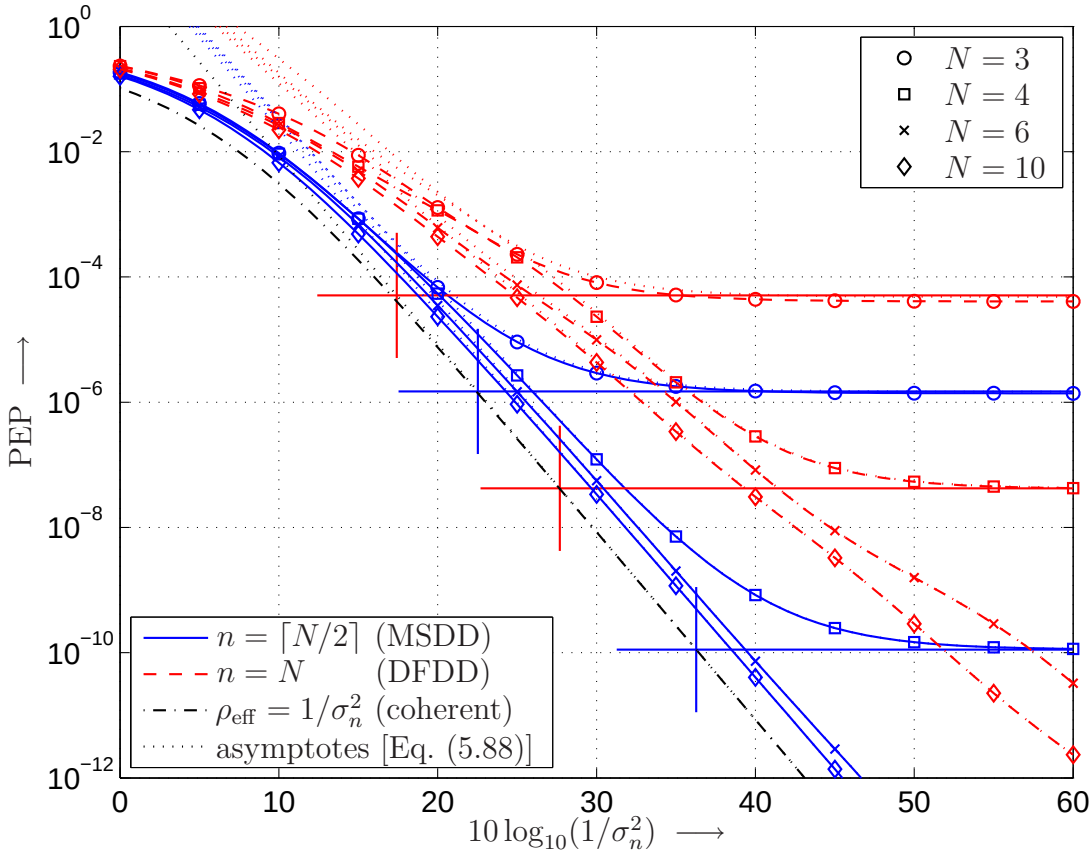

Figure 5.8: PEPs of dominant error event vs. SNR for cyclic (B)DSTM with $N_{\mathrm{T}} = 3$, $R = 1$, $N_{\mathrm{R}} = 1$, $B_{h,\mathrm{eff}}T = 0.1$, $N \in \{3, 4, 6, 10\}$ and $n \in \{N/2, N\}$. For comparison: coherent detection with perfect CSI and asymptotes for $\sigma_n^2 \to 0$. Horizontal lines: approximation of error floor according to (5.92).

5.4.5.1.2 Symbol–Error Rate Approximation Having presented numerous results on the various detection schemes using the general framework of the effective SNR, we will in the following illustrate the tightness of the symbol–error rate (SER) approximations of Section 5.4.4 using a few numerical examples.

Required SNR for SER $= 10^{-5}$ **vs. Position** n **for Different** N**:** Fig. 5.10 shows the SNR required by ML MSDD to achieve, respectively, $\mathrm{SER}_n = 10^{-5}$ (solid lines) and an average error rate $\mathrm{SER} = \frac{1}{N-1}\sum_{n=1}^{N-1}\mathrm{SER}_n = 10^{-5}$ (dashed lines) as function of the position n, $1 \le n \le N-1$, for different window sizes N. As an example, we consider cyclic (B)DSTM with $R = 1$, $N_{\mathrm{T}} = 3$, $N_{\mathrm{R}} = 1$. Also included is the SER for differentially coherent detection assuming perfect CSI (dash–dotted line).

First, we observe a good agreement between the SER approximation from (5.117) and the simulated SER. Second, it can be seen that the individual error rates SER_n are almost identical

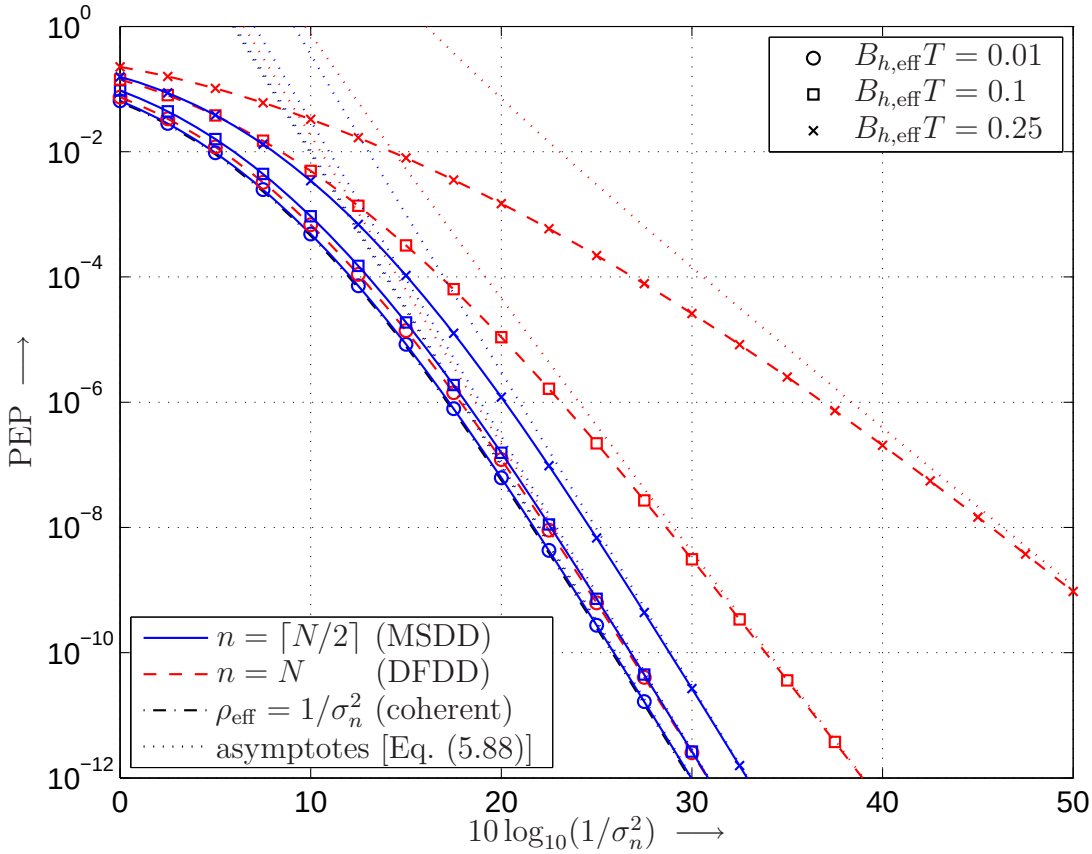

Figure 5.9: PEPs of dominant error event vs. SNR for cyclic (B)DSTM with $N_\mathrm{T} = 5$, $R = 1$, $N_\mathrm{R} = 1$ for $N \to \infty$. For comparison: coherent detection with perfect CSI and asymptotes for $\sigma_n^2 \to 0$.

for symbols $\boldsymbol{V}_n$ and $2 \leq n \leq N-2$, but significantly deteriorate for symbols $\boldsymbol{V}_n$ at the edges of the observation window, i.e. $n = 1$ and $n = N-1$. More specifically, there are differences of $5 - 8\,\mathrm{dB}$ in power efficiency when comparing non–edge and edge positions. This complies with the observations made when considering a similar plot in terms of ESNR in Fig. 5.1.

Required SNR for SER $= 10^{-5}$ **vs.** $B_{h,\mathrm{eff}}T$**, Cyclic (B)DSTM:** Fig. 5.11 compares the various detectors in terms of the SNR required to achieve SER $= 10^{-5}$ as function of $B_{h,\mathrm{eff}}T$ for cyclic (B)DSTM with $R = 1$, $N_\mathrm{T} = 3$, $N_\mathrm{R} = 1$. For MSDD and DFDD an observation window length of $N = 10$ was used. Both analytical results according to Section 5.4.4 (lines) and simulation results (markers) are plotted. As reference curves, the SNR for MSDD and a very–slow–fading model ($B_{h,\mathrm{eff}}T \to 0$) and for (differentially) coherent detection with perfect CSI are shown, as well.

First, we note the good match of the SER approximation from Section 5.4.4 and the simulated SER. Only for decision–feedback MSDD (DF–MSDD) with $\kappa_\mathrm{U}^\mathrm{DF} = \kappa_\mathrm{L}^\mathrm{DF} = 1$ and subset

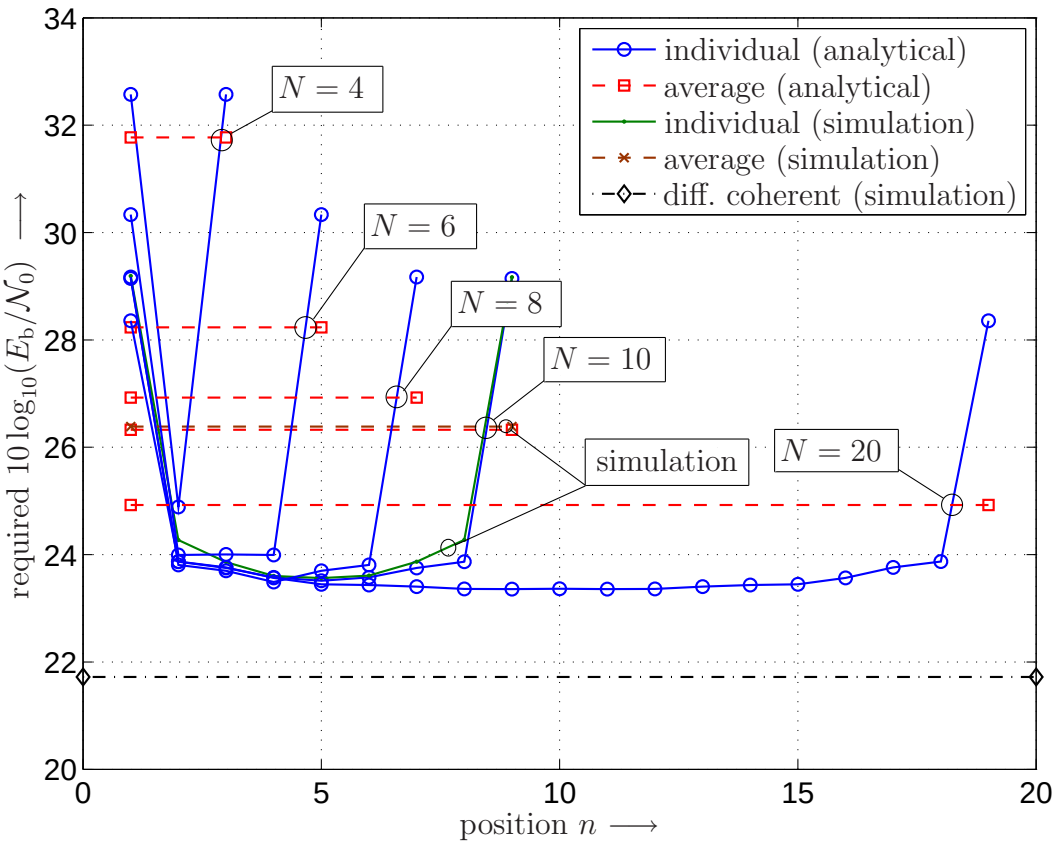

Figure 5.10: Required $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ to achieve SER $= 10^{-5}$ for position n in observation window of ML MSDD. Parameters: Cyclic (B)DSTM constellation, $N_\mathrm{T} = 3$, $N_\mathrm{R} = 1$, $R = 1$, and $B_{h,\mathrm{eff}}T = 0.09$.

MSDD (S–MSDD) with $N' = 7$ there is a noteworthy deviation in fast–fading scenarios as apparently decisions close to the edges of the observation window are somewhat less reliable than predicted by SER approximation from Section 5.4.4, cf. also Fig. 5.10. In the case of DF–MSDD this deviation is increased by the fact that we assume perfect feedback in our SER approximation.

Second, we observe that CDD suffers from a relatively high error floor already in moderately fast fading with $B_{h,\mathrm{eff}}T \geq 0.018$. While DFDD significantly improves power efficiency compared to CDD, it is still outperformed by MSDD by about $2-6$ dB depending on the fading bandwidth.

Finally, it can be seen that DF–MSDD and S–MSDD significantly improve performance in the fast–fading regime. Almost the entire gain in power efficiency achievable with S–MSDD is already accomplished with $N' = 7$, which corresponds to a moderate complexity increase by a factor of 1.29 compared to MSDD with $N = 10$.

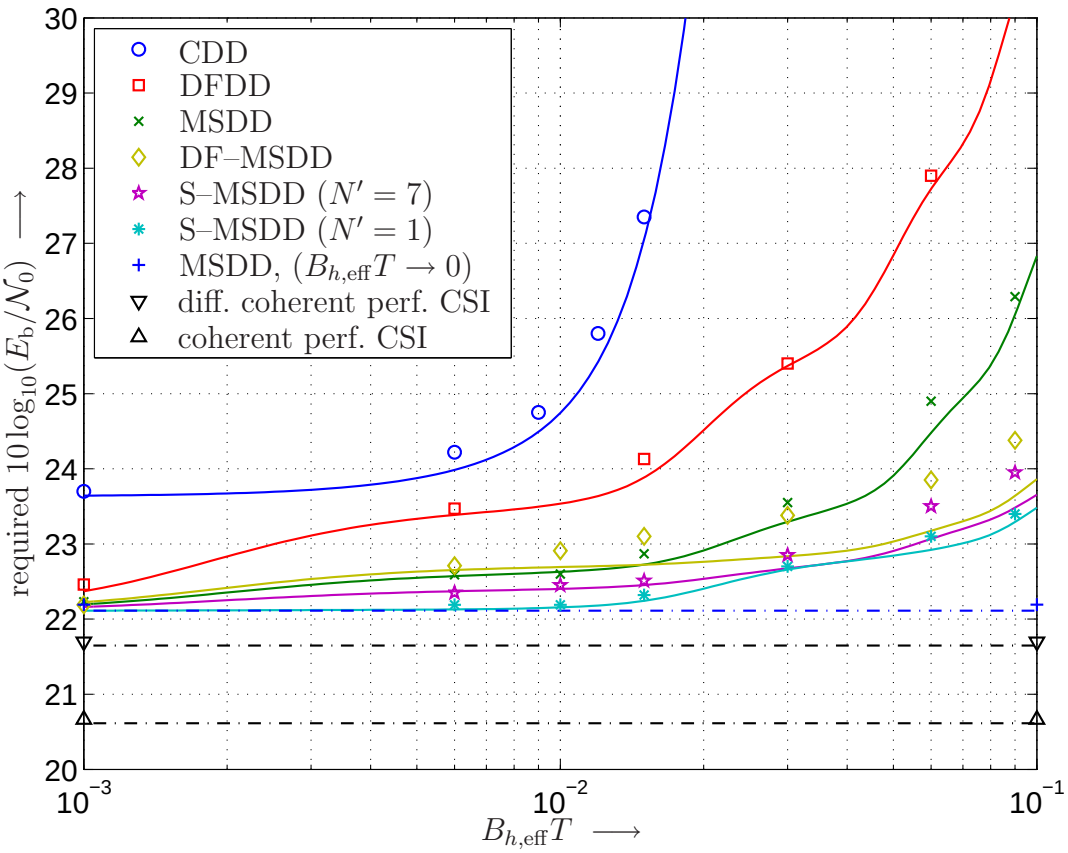

Figure 5.11: Comparison of various decoders, MSDD and DFDD with $N = 10$. Required $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ to achieve SER $= 10^{-5}$ vs. $B_{h,\mathrm{eff}}T$. Cyclic constellation with $N_\mathrm{T} = 3$, $N_\mathrm{R} = 1$, $R = 1$. Lines: analytical results, Markers: simulation results.

Required SNR for SER $= 10^{-5}$ **vs.** B_hT**, Orthogonal DSTM:** Cyclic DSTM as considered above plays a special role for noncoherent detection, because the detector's QSFC assumption, i.e. the assumption of the channel being constant during the transmission of each DSTM symbol, does not lead to a performance degradation even in fast–fading scenarios (cf. Section 5.2.2 for more details). For non–diagonal DSTM constellations, such as orthogonal DSTM constellation, this leads to a metric mismatch, which ultimately limits the performance of detectors in high SNR regimes.

Fig. 5.12 shows the performance of DSTM with orthogonal DSTM with $R = 1$, $N_\mathrm{T} = 2$, $N_\mathrm{R} = 1$ in terms of the SNR required to achieve SER $= 10^{-5}$ vs. B_hT. MSDD is compared with DFDD and CDD. For MSDD and DFDD an observation window length of $N = 10$ was used. Analytical results (see Section 5.4.4, lines) and simulation results (markers) are plotted. In order to illustrate the degradation due to the discrepancy between the QSFC model as assumed by the detectors and the CFC model, that serves as basis for our results, analytical and simulation results for a channel that actually follows the QSFC model are also included as

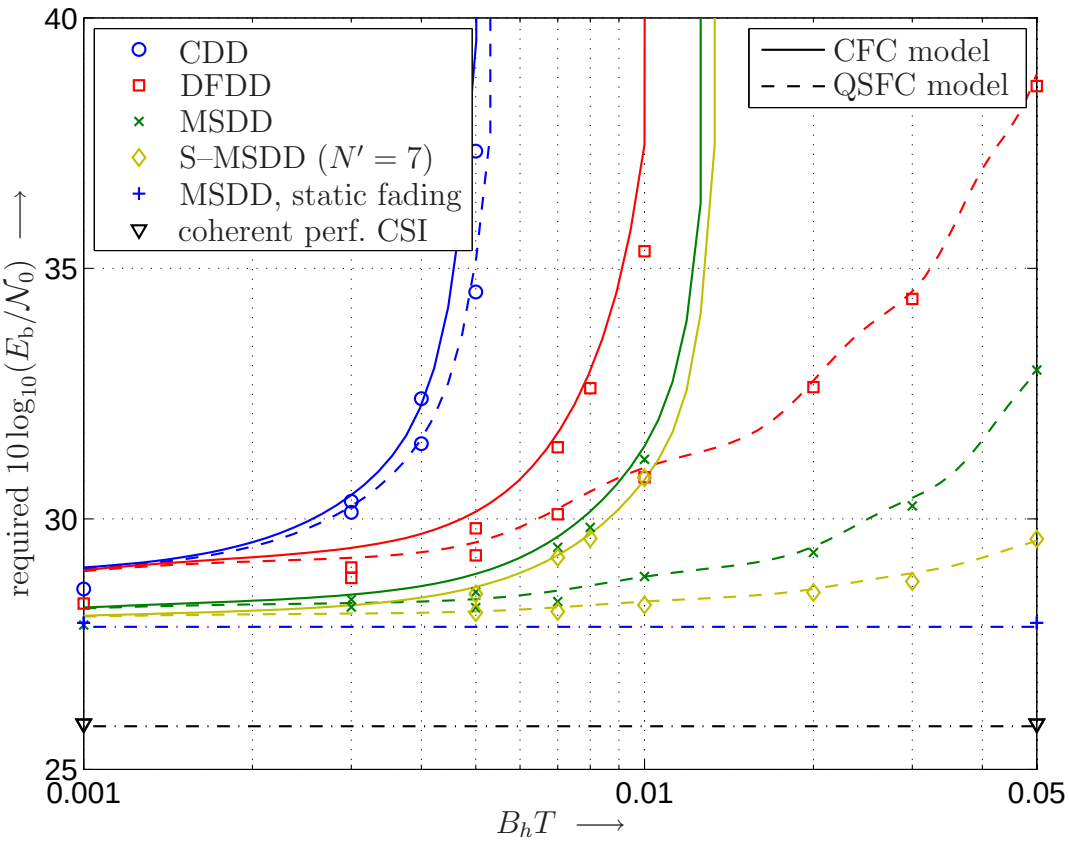

Figure 5.12: Comparison of various decoders, MSDD and DFDD with $N = 10$. Required $10\log_{10}(E_b/\mathcal{N}_0)$ to achieve SER $= 10^{-5}$ vs. B_hT. Orthogonal design with $N_T = 2$, $N_R = 1$, and $R = 1$. CFC model (solid lines) and QSFC model (dashed lines). Lines: analytical results, Markers: simulation results.

dashed lines in the figure.

First, we observe that SERs from the approximation in Section 5.4.4 and simulated SERs closely match also for this non–group constellation. MSDD and DFDD can cope with much faster fading than CDD. However, it can be seen that the performance of the detectors is limited by channel variations during the transmission of one ST symbol, which are not accounted for in the decision metrics, as the noncoherent detectors are based on the QSFC model. Clearly, the impact of this metric mismatch on the reliability of decisions in the individual positions of the observation window is the same. This explains, why the use of S–MSDD (or DF–MSDD, not shown) does not lead to the expected significant gains in rapid fading scenarios.

5.4.5.1.3 Spatially Correlated Channel To illustrate the impact of spatial correlation in the MIMO channel [cf. (4.8) with (4.9) and (4.10)], we use the simple exponential correlation model [CRTP03]

$$\psi^{\mathrm{Tx}}[i,j] = \mathrm{e}^{-\alpha\left|\left|\boldsymbol{x}_i^{\mathrm{Tx}}-\boldsymbol{x}_j^{\mathrm{Tx}}\right|\right|}, \tag{5.128}$$

$$\psi^{\mathrm{Rx}}[i,j] = \mathrm{e}^{-\alpha\left|\left|\boldsymbol{x}_i^{\mathrm{Rx}}-\boldsymbol{x}_j^{\mathrm{Rx}}\right|\right|}, \tag{5.129}$$

where $\boldsymbol{x}_i^{\mathrm{Tx}}$ and $\boldsymbol{x}_j^{\mathrm{Rx}}$ denote the position vectors of transmit antenna i and receive antenna j in multiples of the carrier wavelength, respectively, and $\alpha \in \mathbb{R}^+$. In accordance with the commonly acknowledged fact that spatial subchannels from two different transmit antennas to the same receive antenna or from one transmit antenna to two different receive antenna are practically uncorrelated if the elements of the antenna arrays are separated by at least half of the carrier wavelength (cf. e.g. [SFGK00]), we choose the parameter α of this model as $\alpha = 0.8$. While other array geometries may at times be more appropriate, we restrict our attention to *linear equispaced* antenna arrays.

Comparison of Receivers Aware / Unaware of Spatial Correlation: Fig. 5.13 compares the power efficiency of the various detection algorithms for cyclic DSTM with $N_{\mathrm{T}} = 3$, $R = 1$, $N_{\mathrm{R}} = 1$ and a linear equispaced transmit–antenna array with normalized interelement spacing $\left|\left|\boldsymbol{x}_i^{\mathrm{Tx}} - \boldsymbol{x}_{i+1}^{\mathrm{Tx}}\right|\right| = 0.01$ such that $\psi^{\mathrm{Tx}}[i, i+1] = 0.992$, $i \in \{1, 2\}$, i.e. in the presence of strong spatial correlation in the MIMO channel. The normalized fading bandwidth is chosen as $B_h T = 0.03$ and $N = 10$. Fig. 5.13 shows analytical results for receivers that are unaware (solid lines) and aware (dashed lines) of the spatial correlation.[10] Simulation results for the receivers that are unaware of the spatial correlation are depicted as dash–dotted lines. For comparison, we also included analytical results for the SERs in the case of a spatially uncorrelated channel (dotted lines).

First, Fig. 5.13 shows that our SER approximation of Section 5.4.4 is quite tight also in the case of a spatially correlated MIMO channel. We further can observe that the performance degradation due to the detectors' ignorance of spatial correlation is particularly pronounced for CDD, whereas for DFDD and MSDD only a moderate loss occurs. For S–MSDD ($N' = 7$) and DF–MSDD (not shown) the receiver unaware of spatial correlation achieves practically the same power efficiency as the one taking spatial correlation into account, which means that the ignorance of spatial correlation has a particularly strong impact on the reliability of decisions at the very edges of the observation window. Note, that in the case of coherent detection with

[10] "Unaware of spatial correlation" means that the receivers assume that spatial correlation is not present. All noncoherent detectors considered in this work belong to this class of algorithms. "Aware of spatial correlation" means that the receivers take the spatial correlation into account, i.e. use matrix coefficients $\tilde{\boldsymbol{C}}_{i,j}$ instead of scalar coefficients $c_{i,j}$ as discussed in Section 5.3.1.

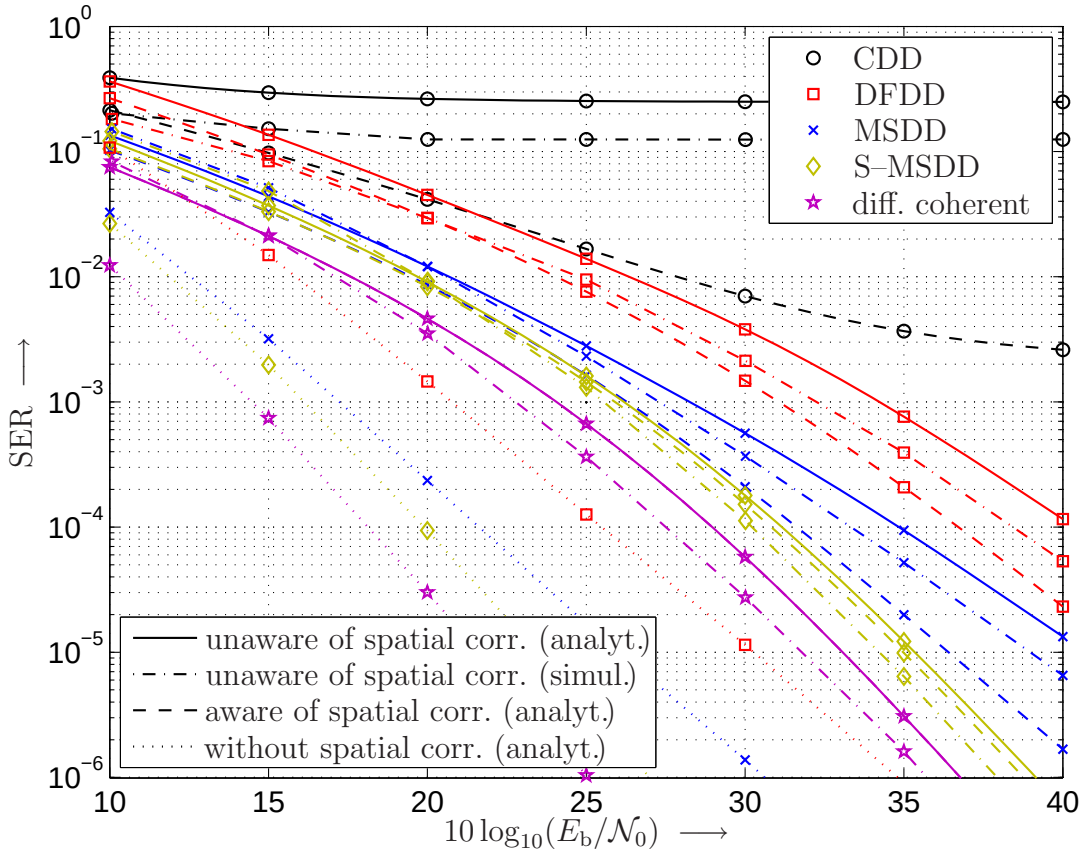

Figure 5.13: SER of the various detectors vs. SNR for linear equispaced transmit–antenna array with $\left|\left|\boldsymbol{x}_i^{\mathrm{Tx}} - \boldsymbol{x}_{i+1}^{\mathrm{Tx}}\right|\right| = 0.01$, cyclic DSTM with $N_{\mathrm{T}} = 3$, $R = 1$, $N_{\mathrm{R}} = 1$, $N = 10$, $B_h T = 0.03$. Dotted lines with markers: simulation results for receivers unaware of spatial correlation.

perfect CSI the detector is naturally aware of spatial correlation. Hence, the dashed and the solid lines coincide in this case.

Furthermore, a comparison with the results for a spatially uncorrelated channel (dotted lines) reveals, that the strong spatial correlation hardly influences the asymptotic diversity order achieved by the detection schemes. Instead it mainly causes a (significant) loss in SNR.

Finally, it is worth mentioning that BDSTM helps to avoid detrimental effects of spatial correlation on the performance of the detectors, since different elements of a DSTM symbol are transmitted in modulation interval that are usually separated by several modulation intervals. Therefore, the temporal variations of the fading channel decorrelate the respective channel coefficients.

Required SNR vs. Interantenna Spacing: Fig. 5.14 shows analytical results for the SNR required to achieve SER $= 10^{-5}$ vs. the interantenna spacing $\left|\left|\boldsymbol{x}_i^{\mathrm{Tx}} - \boldsymbol{x}_{i+1}^{\mathrm{Tx}}\right|\right|$ of a linear equispaced transmit–antenna array of various numbers N_{T} of antennas. In particular, we

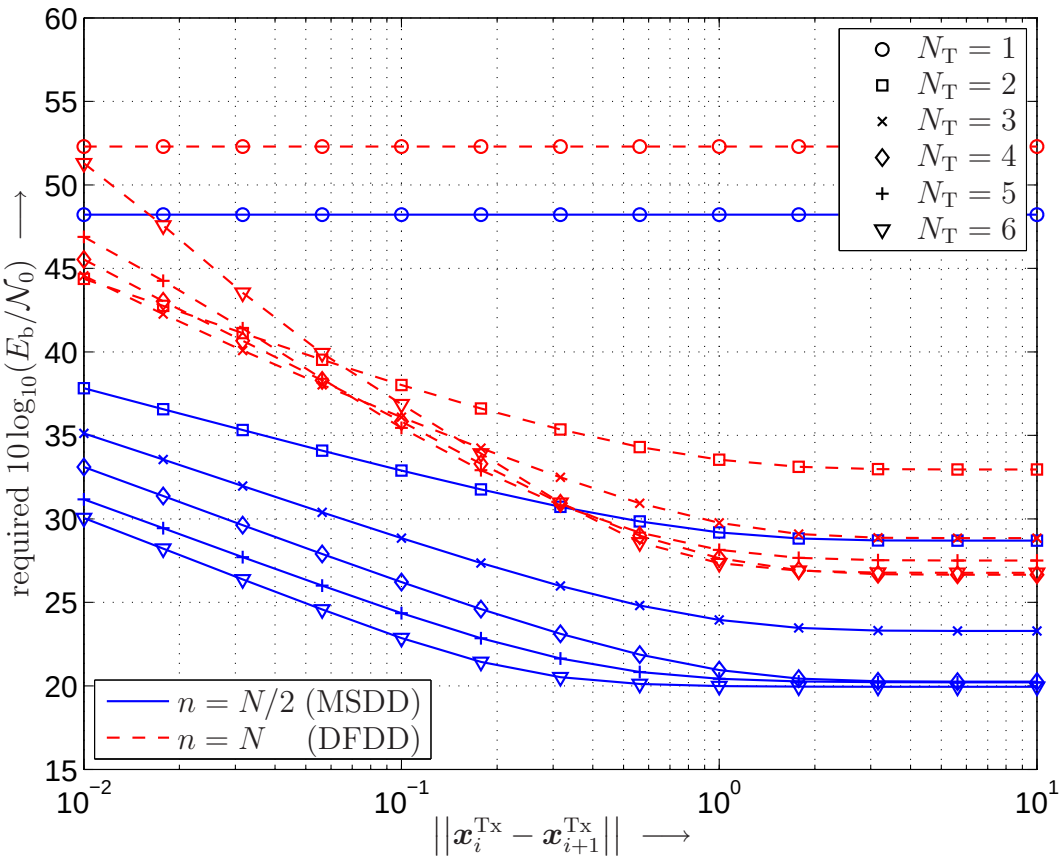

Figure 5.14: Required SNR to achieve $\mathrm{SER}_n = 10^{-5}$ vs. interantenna spacing $\left|\left|\boldsymbol{x}_i^{\mathrm{Tx}} - \boldsymbol{x}_{i+1}^{\mathrm{Tx}}\right|\right|$ for cyclic DSTM with $R = 1$, $N_{\mathrm{R}} = 1$, $N = 10$, $B_h T = 0.03$ and linear equispaced transmit–antenna array.

plotted the required SNR for center and edge positions of an MSDD observation window of length N, when the receiver is unaware of the spatial correlation, for the example of cyclic DSTM with $R = 1$, $N_{\mathrm{R}} = 1$, $B_h T = 0.03$.

One can observe that there is consistent improvement in the reliability of decisions in the center of the observation window with increasing N_{T} for spatially dense transmit–antenna arrays due to the increased diversity. Note, that this holds despite the fact that the effective fading bandwidth $B_{h,\mathrm{eff}} T$ increases as N_{T} grows. Even for very dense arrays, when the interantenna spacing is only 1% of the carrier wavelength, our model predicts significant gains in power efficiency, when using a larger number of transmit antennas.

Interestingly, matters at the edges of the observation window are fundamentally different for spatially dense antenna arrays. Here, the spatial correlation seriously impairs the reliability of the decoder output. Especially for large values of N_{T} the loss in power efficiency of edge positions (DFDD) compared to center position (S–MSDD and DF–MSDD) is tremendous. Thus S–MSDD and DF–MSDD are contrary to DFDD or regular MSDD very robust against

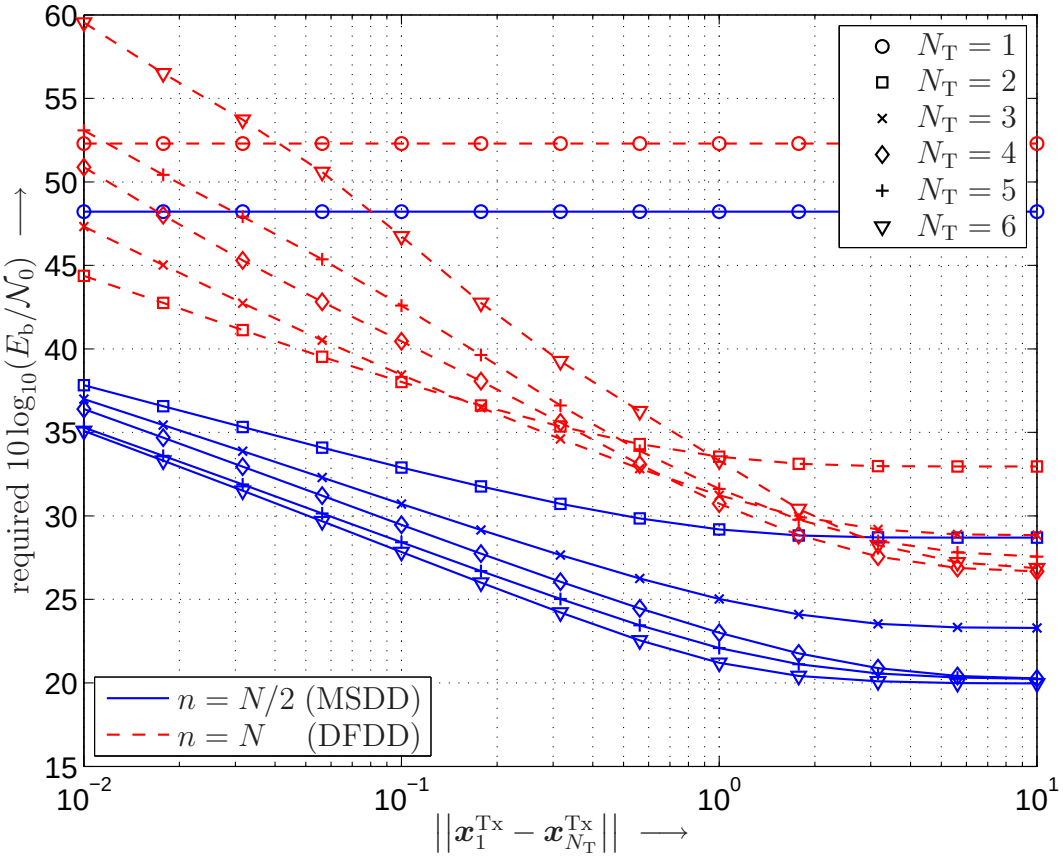

Figure 5.15: Required SNR to achieve $\mathrm{SER}_n = 10^{-5}$ vs. length $\left|\left|\boldsymbol{x}_1^{\mathrm{Tx}} - \boldsymbol{x}_{N_\mathrm{T}}^{\mathrm{Tx}}\right|\right|$ of linear equispaced transmit–antenna array for cyclic DSTM with $R = 1$, $N_\mathrm{R} = 1$, $N = 10$, $B_h T = 0.03$.

impairment due to spatial correlation in the MIMO channel.

Required SNR vs. Length of Antenna Array: Fig. 5.15 shows analytical results for the SNR required to achieve SER $= 10^{-5}$ vs. the physical length $\left|\left|\boldsymbol{x}_1^{\mathrm{Tx}} - \boldsymbol{x}_{N_\mathrm{T}}^{\mathrm{Tx}}\right|\right|$ of a linear equispaced transmit–antenna array of various numbers N_T of antennas. As in Fig. 5.14, we plotted the required SNR for center and edge positions of an MSDD observation window of length $N = 10$, when the receiver is unaware of the spatial correlation, for the example of cyclic DSTM with $R = 1$, $N_\mathrm{R} = 1$, $B_h T = 0.03$.

The different options for the outer decoder, i.e. for the (usually) $(N-1)$–dimensional tree–search decoding problem, are listed in Table 5.2. This figure shows, that while matters in center positions, i.e. for S–MSDD and DF–MSDD are not fundamentally different from the scenario considered in Fig. 5.14, the performance at edge positions, i.e. for DFDD and also regular MSDD, deteriorates even more significantly. Here, antenna arrays of larger numbers of antennas are advantageous only for physically relatively long antenna arrays. For small array lengths $\left|\left|\boldsymbol{x}_1^{\mathrm{Tx}} - \boldsymbol{x}_{N_\mathrm{T}}^{\mathrm{Tx}}\right|\right|$ it would —for DFDD— even be better to use a single antenna only.

Variants of MSDD		
Name:	Mnemonic:	Section:
Subset (MSDD)	S	2.4.2.2
Decision–feedback (MSDD)	DF	2.4.4

Table 5.1: Summary of variants of MSDD.

Outer MSDD decoders (Section 3.1.3)		
Name:	Mnemonic:	Section:
Agrell–SpD	MSDSD	3.1.3.1
Fano–algorithm	Fano–MSDD	3.1.3.2

Table 5.2: Summary of outer decoders for MSDD.

Inner MSDD decoders (Section 3.1.4)			
Name:	Mnemonic:	Section:	Restrictions:
Full search	FS	3.1.4	—
Lattice decoder	LD	3.1.4.1.2	only for (di–)cyclic DSTM
Bound intersect detect	BID	3.1.4.1.3	only for (di–)cyclic DSTM
Linear approximation	LA	3.1.4.4	only for Cayley Codes

Table 5.3: Summary of inner decoders for MSDD.

MSDD metrics (Section 3.1.2)		
Name:	Mnemonic:	Section:
Maximum–likelihood	ML	3.1.2.2
Fano–type	FM	3.1.2.3

Table 5.4: Summary of metrics for tree–search MSDD.

5.4.5.2 Implementation–Dependent Performance Analysis

In the following, we present some results based on our analysis and results obtained via Monte–Carlo simulation to illustrate the difference between the performance of the various optimal and suboptimal detection algorithms introduced in Chapters 2 and 3.

For the sake of clarity, let us briefly summarize the various options of implementation. The variants of MSDD proposed in this work are listed in Table 5.1 along with their respective mnemonics and the section where they were introduced. For inner decoding, i.e. for the search for the (next–)best extension in the outer tree–search decoder, we consider the options summarized in Table 5.3. Recall, that for the orthogonal DSTM MSDD can be cast into the form of a single $2(N-1)$–dimensional tree–search decoding problem with respect to PSK symbols, which can therefore be solved as in the single antenna case, cf. [LSPW05] or Section 3.1.4.1.1. Finally, the possible metrics for the outer MSDD decoder are given in Table 5.4. Based on Tables 5.1–5.4 individual implementations of MSDD are identified using the following code:

[Variant]–[Outer decoder]–[Inner decoder]–[Metric]

This way DF–MSDSD–LD–FM refers to an implementation of decision–feedback MSDD using the Agrell–SpD as outer tree–search decoder, the lattice–decoder–based symbol search and the Fano–type metric. In the case of Fano–MSDD the Fano–type metric is used always, since —as discussed in Section 3.1.3.2— the combination with the ML metric would lead to a computationally quite inefficient decoder.

As benchmark algorithms, we also consider CDD ($N=2$) and DFDD and their LD–based implementations proposed in [CSZ01] and [SL02], respectively. Unless specified otherwise, the CFC model from Section 5.2.1 is applied for simulations.

Comparison of Inner Decoders: Fig. 5.16, exemplarily shows simulation results for the performance with LD–based symbol search for cyclic (B)DSTM with $N_{\mathrm{T}}=3$, $R=1$, $N_{\mathrm{R}}=1$, $B_{h,\mathrm{eff}}T=0.09$ and $N=10$. A–SpD–based MSDD, i.e. MSDSD, its variants S–MSDSD and DF–MSDSD ($\kappa_{\mathrm{U}}^{\mathrm{DF}}=\kappa_{\mathrm{L}}^{\mathrm{DF}}=1$), DFDD and CDD are considered, and suboptimal LD (dashed lines) is compared with optimal FS or BID–based symbol search, i.e. ML MSDD, (solid lines). It can be seen that the cosine–approximation applied in LD causes only small performance degradations in the order of $0.2-0.3\,\mathrm{dB}$ compared to BID or FS symbol search regardless of the particular detector.

Comparison of MSDD Metrics: Fig. 5.17 compares the power efficiency of MSDSD and its variants S–MSDSD ($N'=7$) and DF–MSDSD ($\kappa_{\mathrm{U}}^{\mathrm{DF}}=\kappa_{\mathrm{L}}^{\mathrm{DF}}=1$) with Fano–type metric (MSDSD–FM) to optimal MSDSD and Fano–MSDD. As example, we consider cyclic (B)DSTM with $N_{\mathrm{T}}=3$, $R=2$, $N_{\mathrm{R}}=1$ and $N=10$. Since $L=64$ in this scenario, FS inner decoding is

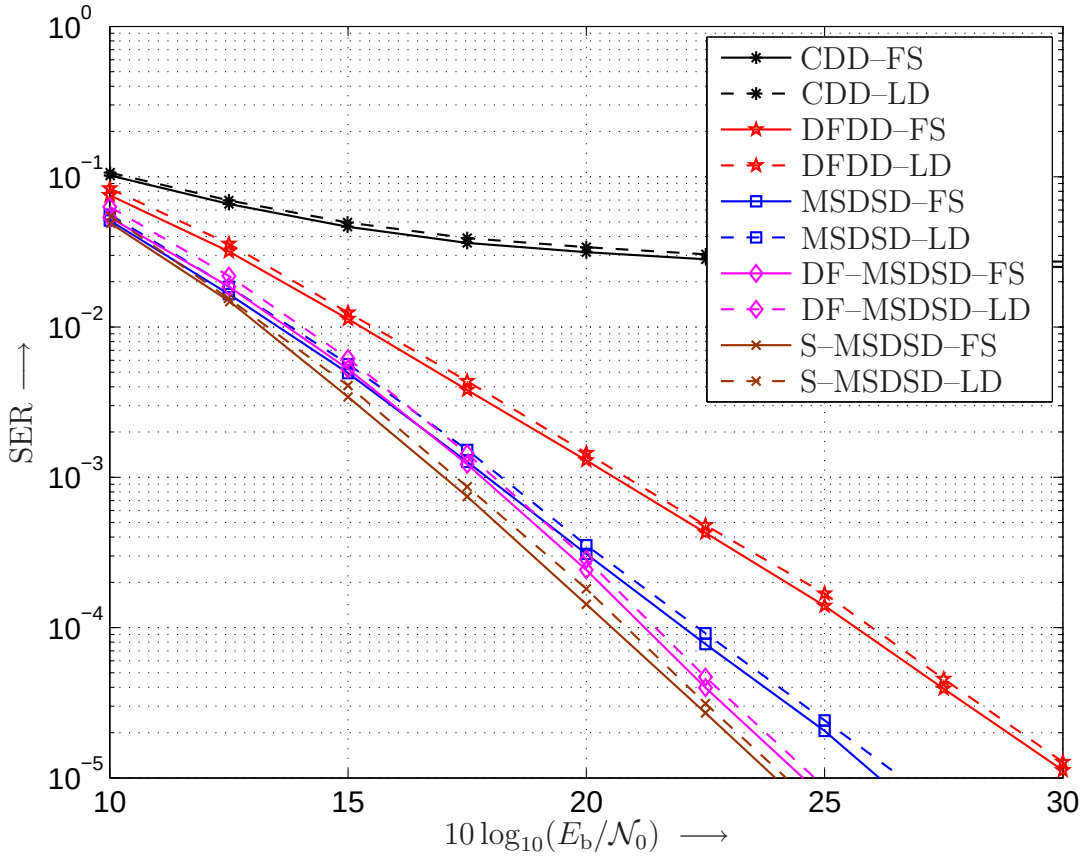

Figure 5.16: Comparison of FS and LD–based inner decoding. SER vs. $10\log_{10}(E_b/\mathcal{N}_0)$ for MSDSD with $N = 10$, S–MSDSD with $N = 10$ and $N' = 7$, DFDD, and CDD. Cyclic (B)DSTM with $N_T = 3$, $N_R = 1$, $R = 1$ and $B_{h,\text{eff}}T = 0.09$. Solid lines: FS, BID, Dashed lines: LD.

computationally complex and we consider LD–based symbol search in all cases. The respective curves for CDD–LD and DFDD–LD are also included for comparison. MSDSD–LD–FM clearly outperforms CDD–LD, which suffer from a very high error floor in this relatively fast fading scenario, and DFDD–LD. We further observe that the deployment of the suboptimal Fano–type metric results in relatively small performance losses of about $0.5 - 1.0\,\text{dB}$. Interestingly, the performance loss due to the Fano–type metric is almost negligible for DF–MSDSD–LD, which must be a result of the very low dimension (two) of the tree–search decoding problem in this case. This way, DF–MSDSD–LD–FM achieves almost the same performance as S–MSDSD–LD–FM. Also —as expected— MSDSD–LD–FM shows a very similar performance as Fano–MSDD–LD.

Finally, a comparison between the results for MSDSD–LD and MSDSD–BID and the respective lines in Fig. 5.16 reveals that the performance loss due to the cosine–approximation in LD–based symbol search is even smaller for larger signal constellations. This is because the difference between PSK–components of the DSTM symbols that lie next to each other on

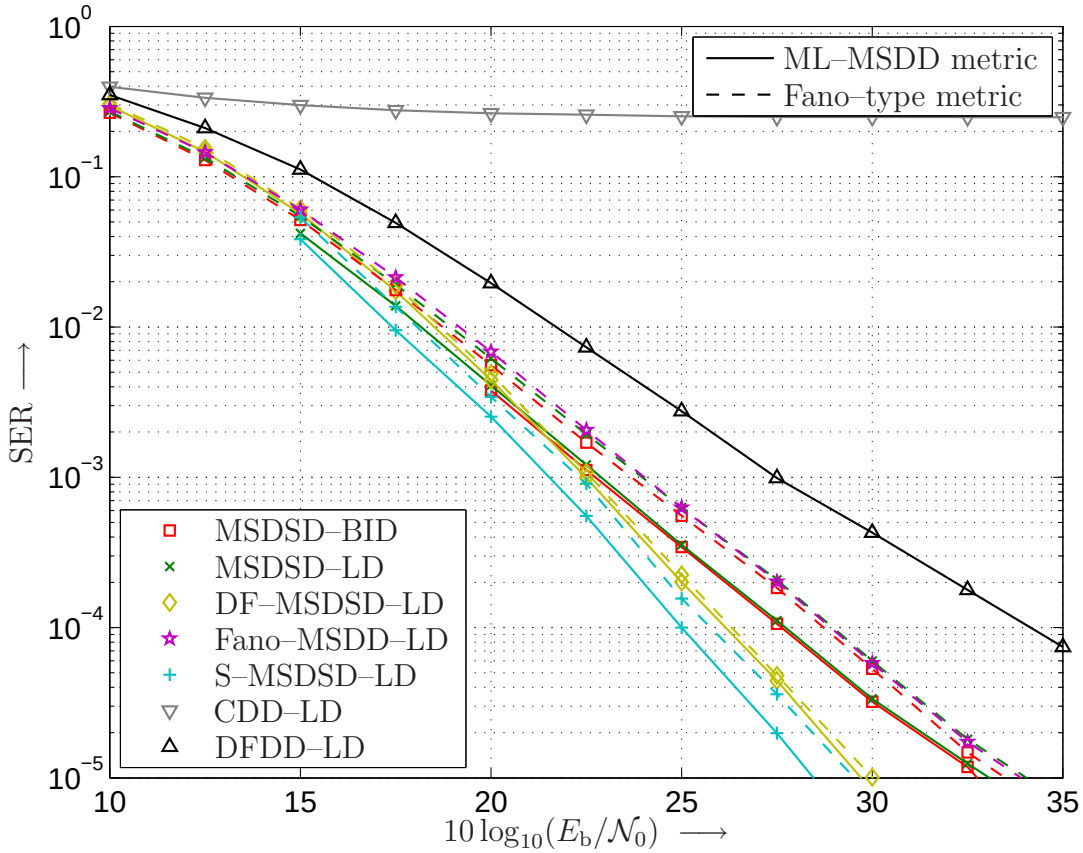

Figure 5.17: Comparison of various decoders with $N = 10$ (CDD with $N = 2$) and LD–based inner decoding. SER vs. $10\log_{10}(E_b/\mathcal{N}_0)$ for cyclic (B)DSTM with $N_T = 3$, $N_R = 1$, $R = 2$, and $B_{h,\text{eff}}T = 0.09$.

the unit–circle and thereby the inaccuracy of the cosine–approximation are reduced for larger constellations.

Error Floor of CG–MSDD with Rank Mismatch: Recall, that we suggested to use a rank–reduced approximation of the $(N \times N)$–dimensional autocorrelation matrix $\mathbf{\Psi}^{(t)}$ of the fading process based on its $\tilde{N}_\lambda$ largest eigenvalues, cf. Section 5.4.3. We referred to the resulting metric mismatch as "rank mismatch". In Fig. 5.18, we consider the performance of CG–MSDD with rank mismatch. In particular, we are interested in the resulting error floor for this non–ML–MSDD. For this purpose, the PEP obtained from evaluation of (5.47) for the dominant error event, i.e. a single–transmit–symbol error at the edge of the observation window, and $\sigma_n^2 \to 0$ is plotted as function of the normalized fading bandwidth B_hT. Fading according to Clarke's (left subplot) and the Gaussian–PSD (right subplot) model are assumed, cf. Table 4.1. Exemplarily, we consider BPSK and observation window lengths $N = 10$ and $N = 20$.

We observe that the error floor is monotonically lowered with increasing $\tilde{N}_\lambda$ and especially

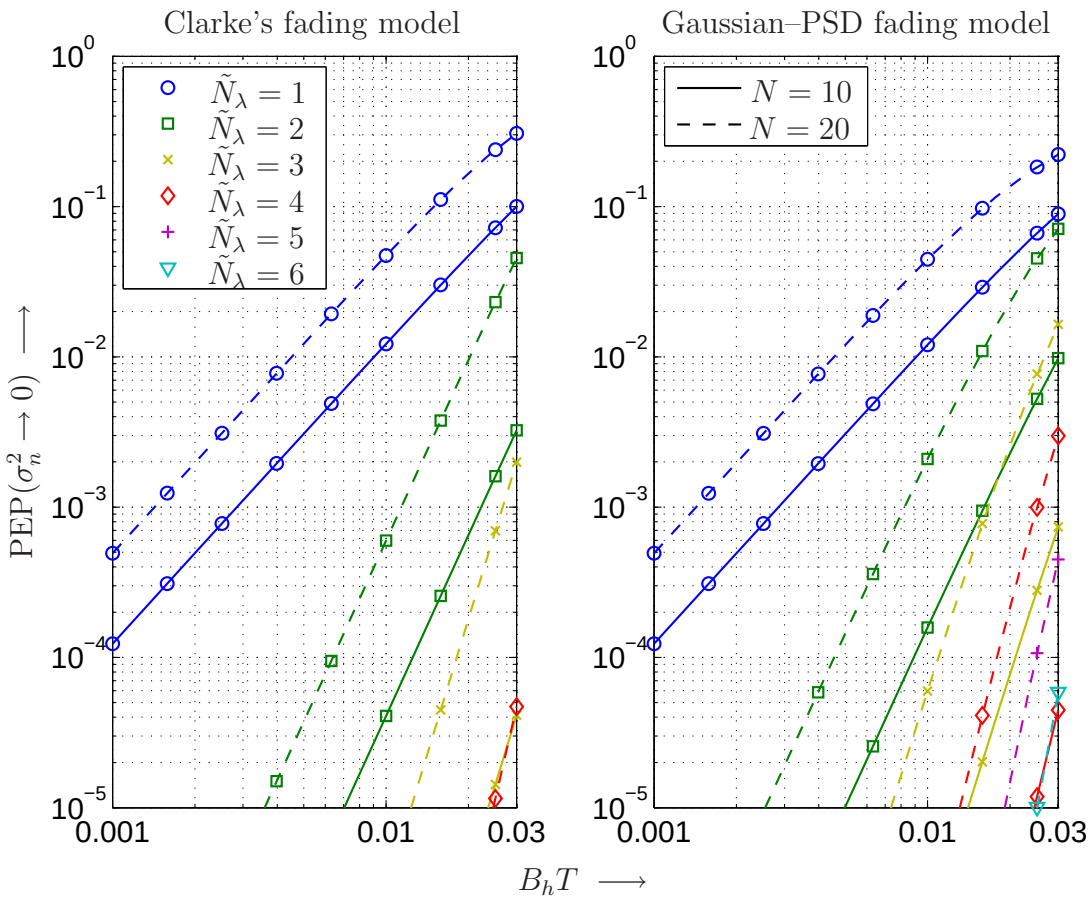

Figure 5.18: Pairwise error probability of dominant error event vs. $B_h T$ for BPSK, very high SNR ($\sigma_n^2 \to 0$) and different values of $\tilde{N}_\lambda$ and N.

that reductions by orders of magnitude are achieved with relatively small values $\tilde{N}_\lambda \ll N$. Clearly, the error floor increases with the fading bandwidth for fixed $\tilde{N}_\lambda$ as the rank mismatch becomes more pronounced. It is also interesting to note that the error floor with $N = 20$ is higher than that with $N = 10$, as the number of dominant eigenvalues of the corresponding $(N \times N)$–dimensional autocorrelation matrix $\mathbf{\Psi}^{\mathrm{t}}$ is lower in the latter case. Therefore, using a smaller N may in the case of CG–MSDD be not only advantageous from a complexity point of view, but also with respect to performance. Since the error floor with respect to BER follows closely the error floor considering the dominant PEP, the PEP–based analysis may serve as quick guideline for choosing an appropriate value for $\tilde{N}_\lambda$ given N, $B_h T$, and the target error rate for which near ML–MSDD performance is desired.

Performance of CG–MSDD with Rank Mismatch: Figs. 5.19 and 5.20 show for $N = 10$ and $N = 20$, respectively, the SERs from an approximation as in Section 5.4.4 (solid) and simulations (dashed) vs. the SNR $10 \log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ for CG–MSDD with rank mismatch and

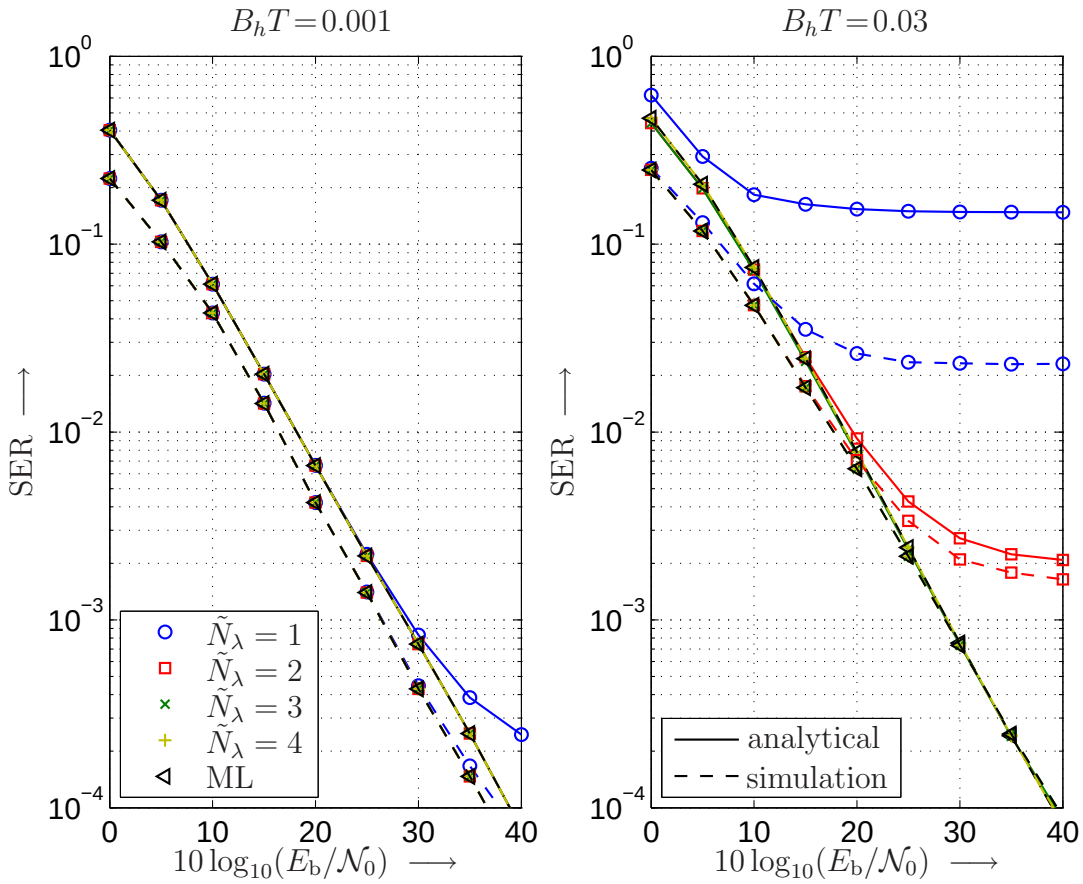

Figure 5.19: SER of CG–MSDD vs. SNR $10\log_{10}(E_b/\mathcal{N}_0)$ for BPSK and Clarke's fading model with $N = 10$, and different values of $\tilde{N}_\lambda$ and B_hT. Solid lines: analytical results. Dotted lines with markers: simulation results. For comparison: ML–MSDD.

for ML–MSDD. Contrary to the other MSDD detection schemes further error events besides the single–transmit–symbol error events, which are mainly caused by AWGN, have to be taken into account to get a reasonably tight approximation of the SER of CG–MSDD with rank mismatch. The reason is that the rank–reduced approximation of $\mathbf{\Psi}^{\mathrm{t}}$ corresponds to an underestimation of the fading bandwidth. This in turn leads to an increased number of errors caused by rapid channel variations. These error events, which correspond to single–*data*–symbol errors, are therefore included in the SER approximation as well. Clarke's fading model is considered, with different values of B_hT. It can be seen that the performance of CG–MSDD with $\tilde{N}_\lambda < N$ is very close to that of ML–MSDD until the error floor due to the rank approximation of $\mathbf{\Psi}^{\mathrm{t}}$ kicks in. For example, to achieve ML–MSDD performance down to SER $= 10^{-4}$, $\tilde{N}_\lambda = [2, 4]$ and $\tilde{N}_\lambda = [3, 4]$ are required for $B_hT = [0.001, 0.03]$ and $N = 10$ and $N = 20$, respectively. These results strongly motivate the application of CG–MSDD based on the rank–reduced fading model, while keeping in mind that with CG–MSDD a larger N does not necessarily lead to

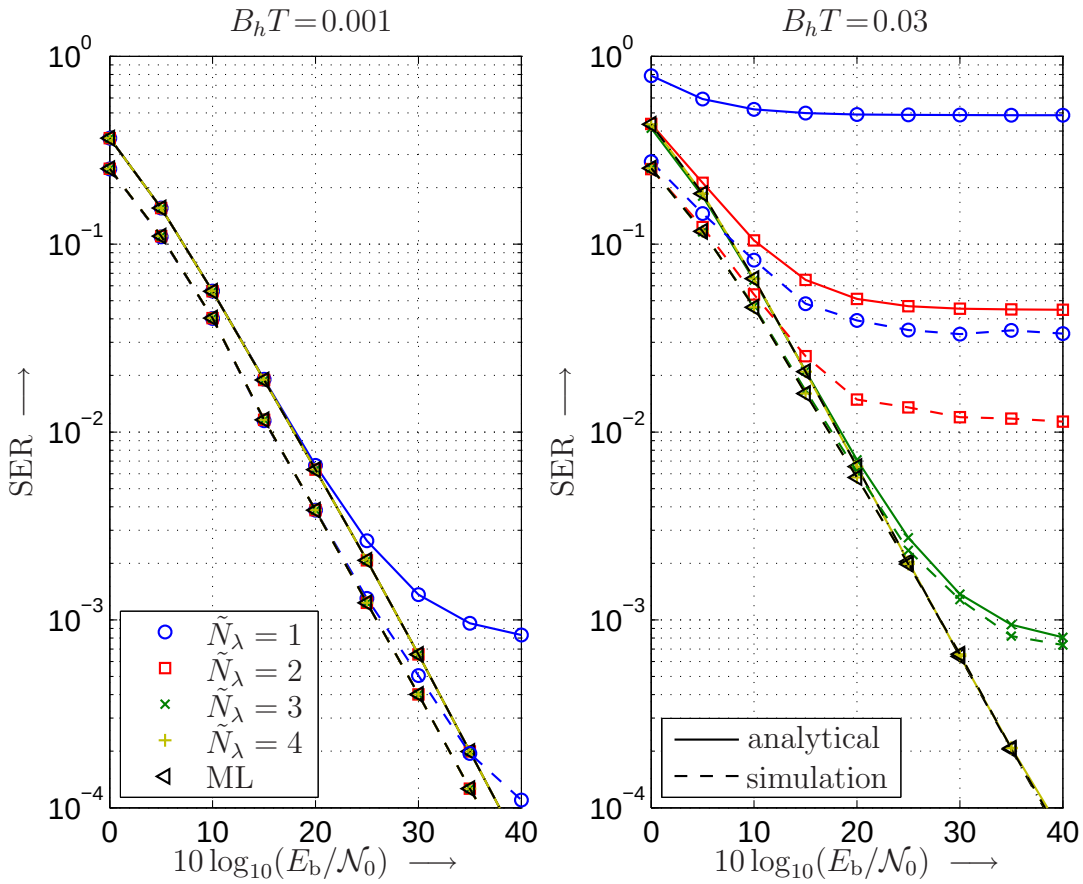

Figure 5.20: SER of CG–MSDD vs. SNR $10\log_{10}(E_b/\mathcal{N}_0)$ for BPSK and Clarke's fading model with $N = 20$, and different values of $\tilde{N}_\lambda$ and B_hT. Solid lines: analytical results. Dotted lines with markers: simulation results. For comparison: ML–MSDD.

improved performance. We also observe that analytical and simulated SER results match quite well, except for small values of $\tilde{N}_\lambda$, where the rank mismatch causes a high error floor and the analytical approximation overestimates the height of the error floor. This is because the abovementioned single–data–symbol error events are due to their connection with the fading process highly correlated. However, predictions for the $\tilde{N}_\lambda$ required to lower the error floor below SER $= 10^{-4}$ from the PEP analysis (see Fig. 5.18) are quite accurate.

5.5 Complexity Analysis

In the previous section, we analyzed the performance of the various detection schemes. We showed that MSDD and its variants can achieve a power efficiency very close to that of coherent detection with perfect channel state information, if the observation window length N is chosen sufficiently large. Since the complexity of the brute–force approach to MSDD is exponential in N, the question that remains to be answered is what the complexity of the alleged computationally efficient implementations proposed in this work is.

As it is customary in the literature (cf. e.g. [HV02, JO05a, FFL05, JZ99, Chapter 6]), we consider the average number C^{blo} of candidates examined per decoder use, i.e. in the case of MSDD per *block* of $(N-1)$ decoded data symbol $\boldsymbol{V}[k]$, as fundamental measure of decoder complexity in the complexity analysis of the various detection algorithms. Note, that in the case of tree–search decoding, this also includes all partial candidates $\tilde{\tilde{\boldsymbol{S}}}_n = \left[\tilde{\boldsymbol{S}}_n^{\mathsf{H}}, \ldots, \tilde{\boldsymbol{S}}_N^{\mathsf{H}}\right]^{\mathsf{H}}$, $1 \leq n \leq N-1$, examined in the process of tree–search decoding. The average number C^{sym} of examined candidates $\tilde{\tilde{\boldsymbol{S}}}_n$ per decoded *symbol*, which is the relevant quantity when comparing algorithms that return different numbers of decisions per decoder use, is trivially derived from C^{blo} by division with the number of decisions returned per decoder use, i.e. $(N-1)$ for MSDD, N' for S–MSDD and $\kappa_{\mathrm{U}}^{\mathrm{DF}} - \kappa_{\mathrm{L}}^{\mathrm{DF}} + 1$ for DF–MSDD. While other quantities such as the average number of floating–point operations (flops) per decoded symbol may be more accurate in describing decoder complexity, because the complexity of considering a candidate may vary from algorithm to algorithm, the latter depends very much upon implementation raising questions such as "How many flops are required to perform a certain algebraic computation?", which are of no interest at this point. On the other hand, C^{blo} as introduced here is uniquely specified by the properties of the algorithms as described above and does not depend on programming skills or the like. In case one is actually interested in e.g. the average number of flops one must merely weigh the number of candidates with the number of flops required in the consideration of each candidate.

In this section, we first provide analytical expressions for the average complexity of MSDD based on sphere decoding and combinatorial geometry in Sections 5.5.1 and 5.5.2, respectively. These are then evaluated in Section 5.5.3 —augmented by simulation results where an analytical evaluation appears to be intractable. We will reveal a number of interesting characteristics of the complexity of the MSDD–based detection schemes, showing that a power efficiency close to that of coherent detection with perfect CSI is attainable at *average and maximal* complexity comparable to that of simple CDD.

5.5.1 Tree–Search MSDD

In this section, we provide a complexity analysis of tree–search–based MSDD. In the case of tree–search decoding, the average total complexity per decoder use is given by the sum

$$C^{\mathrm{blo}} = \sum_{n=1}^{N-1} C_n^{\mathrm{blo}}, \tag{5.130}$$

where C_n^{blo} denotes the average number of partial candidates $\tilde{\tilde{\boldsymbol{S}}}_n$ consisting of $\tilde{n} = N - n + 1$ DSTM symbols, $1 \leq n \leq N-1$, that are examined by the tree–search decoder per block of $(N-1)$ decoded data symbols.

For reasons of analytical tractability, we consider the spatially uncorrelated QSFC model and ML MSDD based on the Fincke–Pohst refinement of the SpD (cf. Page 26), referring to it as FP–MSDSD and discuss implications of the following results on other tree–search–based MSDD algorithms later.

FP–MSDSD examines all (partial) candidate sequences $\tilde{\tilde{\boldsymbol{S}}}_n$, $1 \leq n \leq N-1$, that "lie inside the sphere of radius ρ", i.e. for which [cf. (3.20)]

$$d_n\Big(\tilde{\tilde{\boldsymbol{S}}}_n\Big) = \sum_{i=n}^{N} \Big|\Big| \breve{\boldsymbol{R}}_{i,i}^{\mathsf{H}} \tilde{\boldsymbol{S}}_i + \boldsymbol{X}_i \Big|\Big|^2 \leq \rho \tag{5.131}$$

holds. Contrary to the Agrell–SpD the FP–SpD does *not* updated ρ via $\rho := d_1\big(\tilde{\tilde{\boldsymbol{S}}}\big)$ when a sequence $\tilde{\tilde{\boldsymbol{S}}}$ is found with $d_1\big(\tilde{\tilde{\boldsymbol{S}}}\big) \leq \rho$ (see also [HV02, JO05a]). It is therefore bound to using a finite initial search radius ρ, which has to be chosen as small as possible to minimize decoder complexity, but at the same time sufficiently large, to guarantee that the ML estimate $\hat{\tilde{\tilde{\boldsymbol{S}}}}$ is found inside the sphere with high probability thereby avoiding frequent repetition of the search with an increased search radius. Note that keeping the search radius ρ *fixed* is not a complexity–optimal variant in practice, but it (i) renders a complexity analysis feasible and (ii) serves as an upper bound for decoders which update the search radius during decoding.

Consequently, the quantity C_n^{blo} denotes the average number of candidates $\tilde{\tilde{\boldsymbol{S}}}_n$, $1 \leq n \leq N-1$, that fulfill (5.131) for an appropriately chosen ρ.

In Section 5.5.1.1 we derive a lower bound for C^{blo}, which shows that the complexity of FP–MSDSD is exponential in the length of the observation window N regardless of the SNR and the temporal channel correlation. In Section 5.5.1.2 we present expressions for exact computation of C^{blo}. In Section 5.5.1.3 we derive an analytical expression for the asymptotic complexity of FP–MSDSD for the block–fading channel and group DSTM codes and motivate why these results should also extend to non–group DSTM codes.

5.5.1.1 Lower Bound on Complexity

In [JO05a] a simple lower bound was given for the computational complexity of FP–SpD when applied to coherent MIMO detection. In order to derive a corresponding bound for the decoding problem at hand, we first recall that ρ should be chosen according to

$$\rho = (1+\varepsilon_\rho)\mathcal{E}\big\{d_1\big(\bar{\boldsymbol{S}}\big)\,\big|\bar{\boldsymbol{S}}\big\} = (1+\varepsilon_\rho)NN_{\mathrm{T}}N_{\mathrm{R}}, \qquad \varepsilon_\rho > 0, \tag{5.132}$$

with some arbitrarily small but positive constant ε_ρ, to assure that —in the limit of $N \to \infty$— the probability of finding $\bar{\boldsymbol{S}}$ inside the sphere tends to one, whereas for $\rho \leq (1-\varepsilon_\rho)NN_{\mathrm{T}}N_{\mathrm{R}}$ the probability of finding $\bar{\boldsymbol{S}}$ inside the sphere would eventually tend to zero as N grows, cf. (3.41) and (3.42).

Let us introduce the search depth

$$\delta^{\mathrm{s}}\Big(\tilde{\bar{\boldsymbol{S}}}\Big) \triangleq \sup\Big\{k \in \mathbb{Z} \;\Big|\; k \in [1, N-1], d_{N-k}\Big(\tilde{\bar{\boldsymbol{S}}}_{N-k}\Big) \leq \rho\Big\}. \tag{5.133}$$

It indicates the depth k up to which a particular path corresponding to $\tilde{\bar{\boldsymbol{S}}}_{N-k}$ is pursued in the decoding tree, i.e. the maximal depth k, where $d_{N-k}\big(\tilde{\bar{\boldsymbol{S}}}_{N-k}\big) \leq \rho$ still holds.

Let us further introduce an indicator function $I_k\big(\tilde{\bar{\boldsymbol{S}}}_{N-k}\big)$ that equals one if $d_{N-k}\big(\tilde{\bar{\boldsymbol{S}}}_{N-k}\big) \leq \rho$ and zero otherwise, i.e.

$$I_k\Big(\tilde{\bar{\boldsymbol{S}}}_{N-k}\Big) = \begin{cases} 1 & \text{if } d_{N-k}\big(\tilde{\bar{\boldsymbol{S}}}_{N-k}\big) \leq \rho \\ 0 & \text{if } d_{N-k}\big(\tilde{\bar{\boldsymbol{S}}}_{N-k}\big) > \rho \end{cases}, \quad 1 \leq k \leq N-1. \tag{5.134}$$

In analogy to [JO05a, Lemma 1] we can sum over all possible candidates using the indicator function and write for the *instantaneous* complexity C^{inst}, i.e. the total number of examined candidates excluding the root for a particular received sequence

$$C^{\mathrm{inst}} = \sum_{k=1}^{N-1} \sum_{\forall \tilde{\bar{\boldsymbol{S}}}_{N-k}} I_k\Big(\tilde{\bar{\boldsymbol{S}}}_{N-k}\Big) \tag{5.135}$$

$$= \sum_{k=1}^{N-1} L^{-(N-k-1)} \sum_{\forall \tilde{\bar{\boldsymbol{S}}}} I_k\Big(\tilde{\bar{\boldsymbol{S}}}_{N-k}\Big), \tag{5.136}$$

where the second line follows from the fact that due to the change $\forall \tilde{\bar{\boldsymbol{S}}}_{N-k} \to \forall \tilde{\bar{\boldsymbol{S}}}$ each addend of the second sum in (5.135) is accounted for L^{N-k-1} times. Interchanging the sums we can write

$$= L^{-(N-1)} \sum_{\forall \tilde{\bar{\boldsymbol{S}}}} \sum_{k=1}^{N-1} L^k I_k\Big(\tilde{\bar{\boldsymbol{S}}}_{N-k}\Big), \tag{5.137}$$

with the definitions (5.133) and (5.134) for the search depth and indicator functions, respectively,

$$= L^{-(N-1)} \sum_{\forall \tilde{\tilde{\boldsymbol{S}}}} \sum_{k=1}^{\delta^{\mathrm{s}}\left(\tilde{\tilde{\boldsymbol{S}}}\right)} L^k \tag{5.138}$$

and with the definitions of the finite geometric series and expectation

$$= L^{-(N-1)} \sum_{\forall \tilde{\tilde{\boldsymbol{S}}}} \frac{L^{\delta^{\mathrm{s}}\left(\tilde{\tilde{\boldsymbol{S}}}\right)+1} - L}{L-1} \tag{5.139}$$

$$= \frac{\mathcal{E}_{\tilde{\tilde{\boldsymbol{S}}}}\left\{L^{\delta^{\mathrm{s}}\left(\tilde{\tilde{\boldsymbol{S}}}\right)+1}\right\} - L}{L-1}, \tag{5.140}$$

where the expectation is taken with respect to $\tilde{\tilde{\boldsymbol{S}}}$ only.

Averaging C^{inst} over all received sequences, i.e. over all $\bar{\boldsymbol{S}}$, $\bar{\boldsymbol{H}}$ and $\bar{\boldsymbol{N}}$, we obtain

$$C^{\mathrm{blo}} \triangleq \mathcal{E}_{\bar{\boldsymbol{S}},\bar{\boldsymbol{H}},\bar{\boldsymbol{N}}}\left\{C^{\mathrm{inst}}\right\} \tag{5.141}$$

$$= \left(\mathcal{E}_{\tilde{\tilde{\boldsymbol{S}}},\bar{\boldsymbol{S}},\bar{\boldsymbol{H}},\bar{\boldsymbol{N}}}\left\{L^{\delta^{\mathrm{s}}\left(\tilde{\tilde{\boldsymbol{S}}}\right)}\right\} - 1\right) \frac{L}{L-1} \tag{5.142}$$

for the average total number of examined candidates $\tilde{\tilde{\boldsymbol{S}}}_n$, $1 \leq n \leq N-1$, examined by the sphere decoder. Applying Jensen's inequality (cf. e.g. [CT91]) to the convex exponential function, one can see that the expected number C^{blo} of candidates inside the sphere is lower bounded by

$$C^{\mathrm{blo}} \geq \left(L^{\mathcal{E}_{\tilde{\tilde{\boldsymbol{S}}},\bar{\boldsymbol{S}},\bar{\boldsymbol{H}},\bar{\boldsymbol{N}}}\{\delta^{\mathrm{s}}\left(\tilde{\tilde{\boldsymbol{S}}}\right)\}} - 1\right) \frac{L}{L-1}. \tag{5.143}$$

Thus in order to prove that the lower bound (5.143) is exponential in N, we need to show that the expected search depth $\mathcal{E}\left\{\delta^{\mathrm{s}}\left(\tilde{\tilde{\boldsymbol{S}}}\right)\right\}$ grows asymptotically linearly in N. While we indicated the random variables, with respect to which expectation is taken in the above expressions, using subscripts, we will in the following drop the subscripts noting that all expectations are taken with respect to $\tilde{\tilde{\boldsymbol{S}}}$, $\bar{\boldsymbol{S}}$, $\bar{\boldsymbol{H}}$ and $\bar{\boldsymbol{N}}$.

In analogy to [JO05a] we compute a lower bound on $\mathcal{E}\left\{\delta^{\mathrm{s}}\left(\tilde{\tilde{\boldsymbol{S}}}\right)\right\}$ by considering $\Pr\left(\delta^{\mathrm{s}}\left(\tilde{\tilde{\boldsymbol{S}}}\right) \geq k\right)$, which by (5.133) and Markov's inequality (cf. e.g. [CT91]) can be written as

$$\Pr\left(\delta^{\mathrm{s}}\left(\tilde{\tilde{\boldsymbol{S}}}\right) \geq k\right) = 1 - \Pr\left(d_{N-k}\left(\tilde{\tilde{\boldsymbol{S}}}_{N-k}\right) > \rho\right) \tag{5.144}$$

$$\geq 1 - \frac{\mathcal{E}\left\{d_{N-k}\left(\tilde{\tilde{\boldsymbol{S}}}_{N-k}\right)\right\}}{\rho}. \tag{5.145}$$

With the definition of $d_{N-k}\big(\tilde{\tilde{\boldsymbol{S}}}_{N-k}\big)$ from (3.20) and $||\boldsymbol{X}||^2 = \mathrm{tr}\big\{\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}}\big\}$ we can write

$$\begin{aligned}\mathcal{E}\Big\{d_{N-k}\big(\tilde{\tilde{\boldsymbol{S}}}_{N-k}\big)\Big\} &= \sum_{i=N-k}^{N}\sum_{j_1=i}^{N}\sum_{j_2=i}^{N}\mathcal{E}\Big\{\mathrm{tr}\Big\{\Big(c_{i,j_1}^{*}\boldsymbol{H}_{j_1}^{\mathsf{H}}\boldsymbol{S}_{j_1}^{\mathsf{H}}\tilde{\boldsymbol{S}}_{j_1} + c_{i,j_1}^{*}\boldsymbol{N}_{j_1}^{\mathsf{H}}\tilde{\boldsymbol{S}}_{j_1}\Big)\cdot \\ &\qquad\qquad \Big(c_{i,j_2}\tilde{\boldsymbol{S}}_{j_2}^{\mathsf{H}}\boldsymbol{S}_{j_2}\boldsymbol{H}_{j_2} + c_{i,j_2}\tilde{\boldsymbol{S}}_{j_2}^{\mathsf{H}}\boldsymbol{N}_{j_2}\Big)\Big\}\Big\} \\ &= \sum_{i=N-k}^{N}\sum_{j=i}^{N}|c_{i,j}|^2\,N_{\mathrm{T}}N_{\mathrm{R}}(1+\sigma_n^2), \end{aligned} \tag{5.146}$$

where we exploited (i) that $\boldsymbol{S}_i$, $\boldsymbol{H}_i$ and $\boldsymbol{N}_i$ are mutually independent, (ii) the power constraint $\mathcal{E}\big\{\sum_{j=1}^{N_{\mathrm{T}}}|s_{i,j}[k]|^2\big\} = 1$, and (iii) $\mathcal{E}\{\boldsymbol{S}[k]\} = \boldsymbol{0}_{N_{\mathrm{T}},N_{\mathrm{T}}}$ such that

$$\mathcal{E}\Big\{\mathrm{tr}\Big\{\boldsymbol{H}_{j_1}^{\mathsf{H}}\boldsymbol{S}_{j_1}^{\mathsf{H}}\tilde{\boldsymbol{S}}_{j_1}\tilde{\boldsymbol{S}}_{j_2}^{\mathsf{H}}\boldsymbol{S}_{j_2}\boldsymbol{H}_{j_2}\Big\}\Big\} = \begin{cases}\mathcal{E}\big\{||\boldsymbol{H}_{j_1}||^2\big\} = N_{\mathrm{T}}N_{\mathrm{R}} & \text{if } j_1 = j_2\\ 0 & \text{if } j_1 \neq j_2\end{cases} \tag{5.147}$$

$$\mathcal{E}\Big\{\boldsymbol{N}_{j_1}^{\mathsf{H}}\tilde{\boldsymbol{S}}_{j_1}\tilde{\boldsymbol{S}}_{j_2}^{\mathsf{H}}\boldsymbol{N}_{j_2}\Big\} = \begin{cases}\mathcal{E}\big\{||\boldsymbol{N}_{j_1}||^2\big\} = N_{\mathrm{T}}N_{\mathrm{R}}\sigma_n^2 & \text{if } j_1 = j_2\\ 0 & \text{if } j_1 \neq j_2\end{cases} \tag{5.148}$$

$$\mathcal{E}\Big\{\boldsymbol{H}_{j_1}^{\mathsf{H}}\boldsymbol{S}_{j_1}^{\mathsf{H}}\tilde{\boldsymbol{S}}_{j_1}\tilde{\boldsymbol{S}}_{j_2}^{\mathsf{H}}\boldsymbol{N}_{j_2}\Big\} = 0\,. \tag{5.149}$$

Recall that $c_{i,j}$, $i \leq j \leq N$, are the coefficients of the $(N-i)$th order linear backward MMSE prediction error filter for the fading–plus–noise process normalized by the corresponding standard deviation of the prediction error, cf. Section 3.1.2.2 for details. As the noise process is assumed temporally uncorrelated with variance σ_n^2 we obtain $\sum_{j=i}^{N}|c_{i,j}|^2 \leq \sigma_n^{-2}$, $1 \leq i \leq N$, where equality holds only for static fading and $(N-i) \to \infty$, cf. also Section 5.4.2.1.5. Plugging this into (5.146) we obtain the upper bound

$$\mathcal{E}\Big\{d_{N-k}\big(\tilde{\tilde{\boldsymbol{S}}}_{N-k}\big)\Big\} \leq \frac{1+\sigma_n^2}{\sigma_n^2}N_{\mathrm{T}}N_{\mathrm{R}}(k+1), \tag{5.150}$$

which by (5.145), with the definitions

$$\beta \triangleq \frac{1+\sigma_n^2}{\sigma_n^2}N_{\mathrm{T}}N_{\mathrm{R}} \tag{5.151}$$

$$K \triangleq \lfloor \rho/\beta \rfloor \tag{5.152}$$

and $\delta^{\mathrm{s}}\big(\tilde{\tilde{\boldsymbol{S}}}\big)$ being an integer in turn leads to

$$\begin{aligned}\Pr\Big(\delta^{\mathrm{s}}\Big(\tilde{\tilde{\boldsymbol{S}}}\Big) \geq k\Big) &\geq 1 - \frac{\beta}{\rho}(k+1) &(5.153)\\ &= 1 - \frac{k+1}{K}. &(5.154)\end{aligned}$$

Let us further —quite arbitrarily— introduce a K–ary random variable ν, which is uniformly distributed over the set $\{-1, 0, 1, \ldots, K-2\}$. With this, we can continue from (5.154) writing

for $0 \leq k \leq K-2$

$$\Pr\Big(\delta^{\mathrm{s}}\Big(\tilde{\tilde{\boldsymbol{S}}}\Big) \geq k\Big) \;\geq\; 1 - \Pr(\nu < k) \tag{5.155}$$

$$\;=\; \Pr(\nu \geq k)\,. \tag{5.156}$$

Note that this also holds for $K-1 \leq k \leq N$ as in this case (5.156) trivially states that $\Pr\big(\delta^{\mathrm{s}}\big(\tilde{\tilde{\boldsymbol{S}}}\big) \geq k\big) \geq 0$. Consequently, $\mathcal{E}\big\{\delta^{\mathrm{s}}\big(\tilde{\tilde{\boldsymbol{S}}}\big)\big\}$ can be bounded via

$$\mathcal{E}\Big\{\delta^{\mathrm{s}}\Big(\tilde{\tilde{\boldsymbol{S}}}\Big)\Big\} \;\geq\; \mathcal{E}\{\nu\} = \sum_{\nu=-1}^{K-2} \frac{\nu}{K} = \frac{K-3}{2} \tag{5.157}$$

$$\;>\; \frac{\rho}{2\beta} - 2 = \gamma N - 2, \tag{5.158}$$

where we used $K = \lfloor \rho/\beta \rfloor > \rho/\beta - 1$, $\rho = (1+\varepsilon_\rho) N N_{\mathrm{T}} N_{\mathrm{R}}$ and defined

$$\gamma \triangleq \frac{\sigma_n^2(1+\varepsilon_\rho)}{2(1+\sigma_n^2)}. \tag{5.159}$$

Clearly, by (5.158) $\mathcal{E}\big\{\delta^{\mathrm{s}}\big(\tilde{\tilde{\boldsymbol{S}}}\big)\big\}$ asymptotically grows at least linearly in N. Plugging (5.158) into (5.143) we obtain a lower bound for the average complexity

$$C^{\mathrm{blo}} \geq \frac{L^{N\gamma-1} - L}{L-1}. \tag{5.160}$$

This bound shows that the complexity of FP–MSDSD with an appropriately chosen radius $\rho = (1+\varepsilon_\rho) N N_{\mathrm{T}} N_{\mathrm{R}}$, $\varepsilon_\rho > 0$, is exponential in N regardless of the noise variance or the bandwidth of the fading process. It thus represents a non–trivial extension of [JO05a, Theorem 2, Ineq. (29)] to the case of MIMO MSDD considered here.

Since the bound (5.160) turns out to be quite loose (cf. Section 5.5.3) for observation window lengths of interest, we consider the problem of exact computation of the complexity of FP–MSDSD in the next section.

5.5.1.2 Computing the Exact Complexity

In order to compute the complexity of FP–MSDSD we proceed as follows. In Section 5.5.1.2.1 we derive an expression for the probability $\Pr\big(d_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big) \leq \rho \mid \bar{\boldsymbol{S}}_n\big)$ that a particular candidate $\tilde{\tilde{\boldsymbol{S}}}_n$, $1 \leq n \leq N-1$, lies inside the sphere of radius ρ given that a particular $\bar{\boldsymbol{S}}_n$ was transmitted. In Section 5.5.1.2.2 we determine how to appropriately choose ρ such that the transmitted sequence $\bar{\boldsymbol{S}}$ lies inside the sphere with a desired probability p^{des}. Finally, in Section 5.5.1.2.3 we provide an expression for the complexity C^{blo} and we devise simplifications for efficient evaluation of this expression for special fading–channel scenarios.

5.5.1.2.1 Calculation of $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \leq \rho \mid \bar{\boldsymbol{S}}_n\big)$ To compute the probability that a particular $\tilde{\bar{\boldsymbol{S}}}_n$ lies in the sphere of radius ρ given that a particular $\bar{\boldsymbol{S}}_n$ was transmitted we use the same methods applied in the performance analysis in Section 5.4.1. It is convenient to vectorize the channel model. With $\tilde{n} = N - n + 1$ [cf. (3.4)] and the definitions of $\tilde{n}N_{\mathrm{T}}N_{\mathrm{R}}$–dimensional vectors

$$\bar{\boldsymbol{r}}_n \triangleq \mathrm{vec}\left\{\left[\boldsymbol{R}_n^{\mathsf{H}}, \ldots, \boldsymbol{R}_N^{\mathsf{H}}\right]^{\mathsf{H}}\right\} \tag{5.161}$$

$$\bar{\boldsymbol{h}}_n \triangleq \mathrm{vec}\left\{\left[\boldsymbol{H}_n^{\mathsf{H}}, \ldots, \boldsymbol{H}_N^{\mathsf{H}}\right]^{\mathsf{H}}\right\} \tag{5.162}$$

$$\bar{\boldsymbol{n}}_n \triangleq \mathrm{vec}\left\{\left[\boldsymbol{N}_n^{\mathsf{H}}, \ldots, \boldsymbol{N}_N^{\mathsf{H}}\right]^{\mathsf{H}}\right\} \tag{5.163}$$

and $(\tilde{n}N_{\mathrm{T}} \times \tilde{n}N_{\mathrm{T}})$–dimensional

$$\bar{\boldsymbol{S}}_{\mathrm{D},n} \triangleq \mathrm{diag}\big\{\boldsymbol{S}_n, \ldots, \boldsymbol{S}_N\big\}, \tag{5.164}$$

the received signal corresponding to the last $\tilde{n}$ transmit symbols of the observation window can be expressed as

$$\bar{\boldsymbol{r}}_n = \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \bar{\boldsymbol{S}}_{\mathrm{D},n}\right)\bar{\boldsymbol{h}}_n + \bar{\boldsymbol{n}}_n\,, \quad 1 \leq n \leq N. \tag{5.165}$$

The correlation matrix of the received vector $\bar{\boldsymbol{r}}_n$ conditioned on $\bar{\boldsymbol{S}}_n$ is given by

$$\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}|\bar{\boldsymbol{S}},n} \triangleq \mathcal{E}\big\{\bar{\boldsymbol{r}}_n\bar{\boldsymbol{r}}_n^{\mathsf{H}}|\bar{\boldsymbol{S}}_n\big\} = \boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{\Psi}_{\boldsymbol{r}\boldsymbol{r}|\bar{\boldsymbol{S}},n}, \tag{5.166}$$

where

$$\boldsymbol{\Psi}_{\boldsymbol{r}\boldsymbol{r}|\bar{\boldsymbol{S}},n} \triangleq \bar{\boldsymbol{S}}_{\mathrm{D},n}\left(\left(\boldsymbol{\Psi}_n^{\mathrm{t}} + \sigma_n^2\boldsymbol{I}_{\tilde{n}}\right) \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\bar{\boldsymbol{S}}_{\mathrm{D},n}^{\mathsf{H}} \tag{5.167}$$

$$\boldsymbol{\Psi}_n^{\mathrm{t}} \triangleq \mathrm{toeplitz}\big\{\psi_{1,1}^{\mathrm{t}}, \psi_{2,1}^{\mathrm{t}}, \ldots, \psi_{\tilde{n},1}^{\mathrm{t}}\big\} \tag{5.168}$$

and $\psi_{i,j}^{\mathrm{t}}$ as element in the ith row and jth column of $\boldsymbol{\Psi}^{\mathrm{t}}$ [cf. (5.26)] describing the temporal correlation of the effective fading process.

With (5.165) the difference η_n between the MSDD metric $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$ of a candidate $\tilde{\bar{\boldsymbol{S}}}_n$ and the radius ρ can be written as

$$\eta_n\left(\tilde{\bar{\boldsymbol{S}}}_n\right) \triangleq d_n\left(\tilde{\bar{\boldsymbol{S}}}_n\right) - \rho \tag{5.169}$$

$$= \bar{\boldsymbol{r}}_n^{\mathsf{H}}\left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{F}_n\right)\bar{\boldsymbol{r}}_n - \rho\,, \quad 1 \leq n \leq N-1\,, \tag{5.170}$$

where

$$\boldsymbol{F}_n \triangleq \tilde{\bar{\boldsymbol{S}}}_{\mathrm{D},n}\left(\left(\boldsymbol{\Psi}_n^{\mathrm{t}} + \sigma_n^2\boldsymbol{I}_{\tilde{n}}\right)^{-1} \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\tilde{\bar{\boldsymbol{S}}}_{\mathrm{D},n}^{\mathsf{H}}. \tag{5.171}$$

We observe that $\eta_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$ includes a Hermitian quadratic form in zero–mean complex Gaussian random variables with correlation matrix $\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}|\bar{\boldsymbol{S}},n}$ [cf. (5.166)]. Hence, its characteristic function $\Phi_{\eta_n}(v) \triangleq \mathcal{E}\big\{\mathrm{e}^{\mathrm{j}v\eta_n}\big\}$ is given by [SBS66]

$$\begin{aligned}
\Phi_{\eta_n}(v) &= \mathrm{e}^{-\mathrm{j}v\rho}\det\big\{\boldsymbol{I}_{\tilde{n}N_\mathrm{T}} - \mathrm{j}v\boldsymbol{\Psi}_{\boldsymbol{rr}|\bar{\boldsymbol{S}},n}\boldsymbol{F}_n\big\}^{-N_\mathrm{R}} && (5.172)\\
&= \mathrm{e}^{-\mathrm{j}v\rho}\prod_{i=1}^{\tilde{n}N_\mathrm{T}}\Big(1 - \mathrm{j}v\lambda_{\boldsymbol{\Psi}_{\boldsymbol{rr}|\bar{\boldsymbol{S}},n}\boldsymbol{F}_n,i}\Big)^{-N_\mathrm{R}} && (5.173)\\
&= \mathrm{e}^{-\mathrm{j}v\rho}\prod_{i=1}^{\tilde{n}N_\mathrm{T}}\left(\frac{-\mathrm{j}v_{\eta_n,i}}{v - \mathrm{j}v_{\eta_n,i}}\right)^{N_\mathrm{R}} && (5.174)
\end{aligned}$$

with poles $\mathrm{j}v_{\eta_n,i}$, where

$$v_{\eta_n,i} \triangleq -\frac{1}{\lambda_{\boldsymbol{\Psi}_{\boldsymbol{rr}|\bar{\boldsymbol{S}},n}\boldsymbol{F}_n,i}}, \quad 1 \le i \le \tilde{n}N_\mathrm{T}, \tag{5.175}$$

$\lambda_{\boldsymbol{\Psi}_{\boldsymbol{rr}|\bar{\boldsymbol{S}},n}\boldsymbol{F}_n,i}$ denoting the ith eigenvalue of $\boldsymbol{\Psi}_{\boldsymbol{rr}|\bar{\boldsymbol{S}},n}\boldsymbol{F}_n$ and $\boldsymbol{\Psi}_{\boldsymbol{rr}|\bar{\boldsymbol{S}},n}$ and $\boldsymbol{F}_n$ as defined in (5.167) and (5.171), respectively. Using $\Phi_{\eta_n}(v)$, the probability that a candidate $\tilde{\bar{\boldsymbol{S}}}_n$ lies inside the sphere of radius ρ given that $\bar{\boldsymbol{S}}_n$ was transmitted, can be expressed as

$$\begin{aligned}
\Pr\Big(d_n\Big(\tilde{\bar{\boldsymbol{S}}}_n\Big) \le \rho \,\Big|\, \bar{\boldsymbol{S}}_n\Big) &= \Pr\Big(\eta_n\Big(\tilde{\bar{\boldsymbol{S}}}_n\Big) \le 0 \,\Big|\, \bar{\boldsymbol{S}}_n\Big) && (5.176)\\
&= -\frac{1}{2\pi\mathrm{j}}\int_{-\infty+\mathrm{j}c}^{\infty+\mathrm{j}c}\frac{\Phi_{\eta_n}(v)}{v}\,\mathrm{d}v && (5.177)\\
&= -\sum_{i=1}^{N_{\eta_n,v}}\operatorname*{Res}_{v=\mathrm{j}v^+_{\eta_n,i}}\left\{\frac{\Phi_{\eta_n}(v)}{v}\right\}, && (5.178)
\end{aligned}$$

with $\mathrm{j}v^+_{\eta_n,i}$, $1 \le i \le N_{\eta_n,v} \le \tilde{n}N_\mathrm{T}$, denoting the $N_{\eta_n,v}$ poles of $\Phi_{\eta_n}(v)$ [cf. (5.174)] that lie in the upper half of the complex v–plane. The constant $c \in \mathbb{R}$ that must satisfy $0 \le c \le \min_{1\le i\le N_{\eta_n,v}}\big(\mathrm{Re}\big\{v^+_{\eta_n,i}\big\}\big)$ is introduced to move the path of integration away from the pole at $v = 0$.

Consequently, the same methods as deployed in the performance analysis in Section 5.4.1 can be used here to compute $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \le \rho \mid \bar{\boldsymbol{S}}_n\big)$, i.e. e.g. through numerical integration of (5.177) [cf. (5.50) and (5.51)] or analytically applying (5.52) to (5.178) with (5.174) and (5.175).

5.5.1.2.2 ρ such that $\Pr\big(d_1\big(\bar{\boldsymbol{S}}\big) \le \rho \mid \bar{\boldsymbol{S}}_n\big) = p^{\mathrm{des}}$ As pointed out earlier it is important to properly choose the radius ρ. On the one hand, choosing ρ too small results in a high decoder complexity, because the decoder is likely to not find a candidate $\hat{\bar{\boldsymbol{S}}}$ inside the sphere, i.e. with $d_1\big(\bar{\boldsymbol{S}}\big) \le \rho$, and will subsequently have to repeat the search with an increased ρ. On the other hand, choosing ρ too large will lead to a high probability of the decoder finding an $\hat{\bar{\boldsymbol{S}}}$ inside the sphere, but at the same time to a high computational complexity, since many other candidates lie inside the sphere as well. Hence, the objective must be to find an intermediate value for ρ,

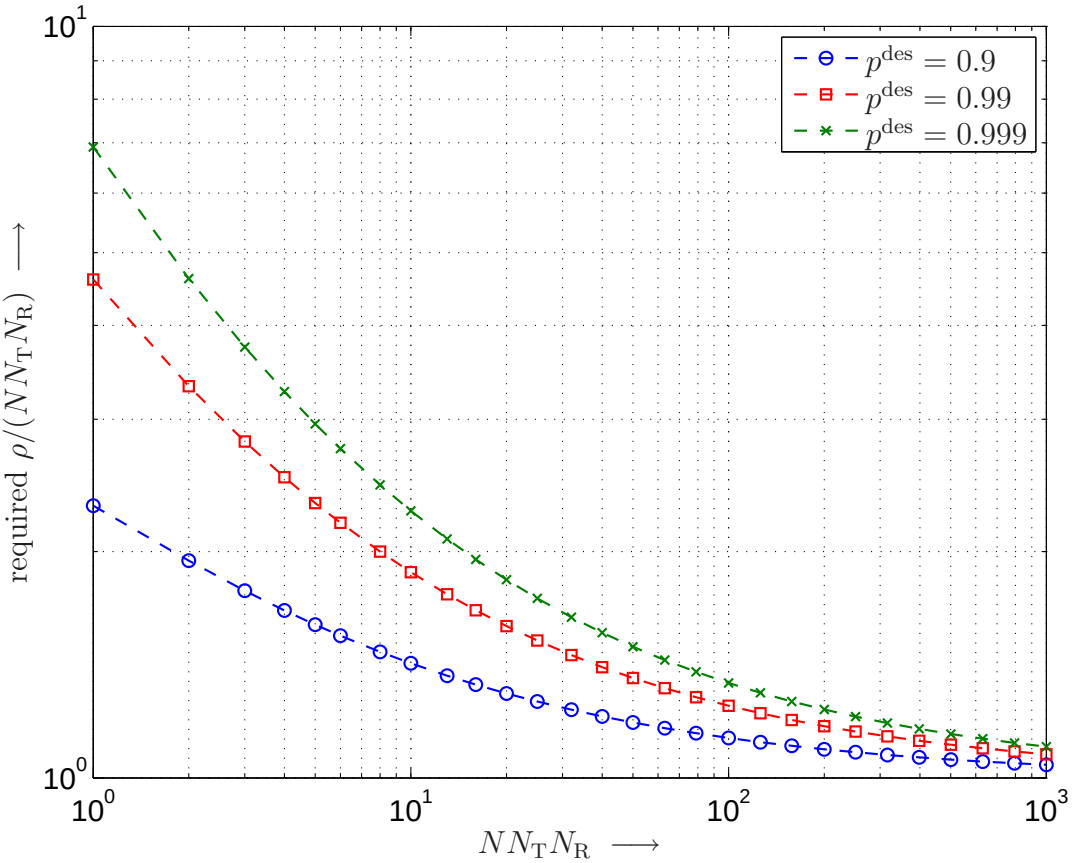

Figure 5.21: Required normalized radius $\rho/(NN_{\mathrm{T}}N_{\mathrm{R}})$ versus $NN_{\mathrm{T}}N_{\mathrm{R}}$ such that $\bar{\boldsymbol{S}}$ lies inside the sphere with probability p^{des}.

such that the decoder succeeds in finding a candidate with high probability p^{des}, but in doing so does not examine too many other candidates.

For $N \to \infty$ this problem has a simple solution, as ρ should be chosen as $\rho := (1+\varepsilon_\rho)NN_{\mathrm{T}}N_{\mathrm{R}}$ with an arbitrarily small positive constant ε_ρ, cf. (3.41) and (3.42).

For the proper choice of the radius ρ in case of finite N we consider $\tilde{\bar{\boldsymbol{S}}} = \bar{\boldsymbol{S}}$ and the corresponding metric $d_1(\bar{\boldsymbol{S}})$. As $d_1(\bar{\boldsymbol{S}})$ is $\chi^2(NN_{\mathrm{T}}N_{\mathrm{R}}, 2NN_{\mathrm{T}}N_{\mathrm{R}})$ distributed in this case (cf. Section 3.1.2.2), we can use well–known results from stochastics (cf. e.g. [Pro00]) and obtain from (5.176)

$$\Pr\bigl(d_1(\bar{\boldsymbol{S}}) \le \rho \,\big|\, \bar{\boldsymbol{S}}\bigr) = \frac{\gamma(NN_{\mathrm{T}}N_{\mathrm{R}}, \rho)}{\Gamma(NN_{\mathrm{T}}N_{\mathrm{R}})} \tag{5.179}$$

($\gamma(A,x) \stackrel{\triangle}{=} \int_0^x y^{A-1}\mathrm{e}^{-y}\,\mathrm{d}y$: (lower) incomplete Gamma function). Using (5.179) we can compute the radius ρ such that $\bar{\boldsymbol{S}}$ is found inside the sphere with a given probability p^{des}. Note that this result depends neither on the SNR nor on the temporal correlation of the fading channel which is very convenient for decoder design. It also depends on the system parameters N, N_{T},

N_R only through the product $NN_\mathrm{T}N_\mathrm{R}$ of the three quantities.

For illustration Fig. 5.21 shows the normalized radius $\rho/(NN_\mathrm{T}N_\mathrm{R})$ required for the true $\bar{\boldsymbol{S}}$ to lie inside the sphere with a given probability p^des as a function of the product $NN_\mathrm{T}N_\mathrm{R}$. It illustrates, how the required value of ρ converges from above to $NN_\mathrm{T}N_\mathrm{R}$ in the sense that $\lim_{NN_\mathrm{T}N_\mathrm{R}\to\infty} \rho/(NN_\mathrm{T}N_\mathrm{R}) = 1$ for any fixed $p^\mathrm{des} < 1$.

5.5.1.2.3 Complexity Given ρ such that $\Pr\big(d_1\big(\bar{\boldsymbol{S}}\big) \le \rho \mid \bar{\boldsymbol{S}}\big) = p^\mathrm{des}$ according to (5.179) for some desired p^des and the probabilities $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \le \rho \mid \bar{\boldsymbol{S}}_n\big) = \Pr\big(\eta_n(\tilde{\bar{\boldsymbol{S}}}_n) \le 0 \mid \bar{\boldsymbol{S}}_n\big)$ from (5.176), the expected number $C_n^\mathrm{blo}\big(\bar{\boldsymbol{S}}_n\big)$ of candidates $\tilde{\bar{\boldsymbol{S}}}_n$, $1 \le n \le N-1$, examined by the FP–MSDSD provided that $\bar{\boldsymbol{S}}_n$ was transmitted is given by

$$C_n^\mathrm{blo}\big(\bar{\boldsymbol{S}}_n\big) = \sum_{\forall \tilde{\bar{\boldsymbol{S}}}_n} \Pr\Big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \le \rho \,\Big|\, \bar{\boldsymbol{S}}_n\Big)\ , \quad 1 \le n \le N-1. \tag{5.180}$$

Thus, the overall complexity given $\bar{\boldsymbol{S}}$ was transmitted reads

$$C^\mathrm{blo}\big(\bar{\boldsymbol{S}}\big) = \sum_{n=1}^{N-1} \sum_{\forall \tilde{\bar{\boldsymbol{S}}}_n} \Pr\Big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \le \rho \,\Big|\, \bar{\boldsymbol{S}}_n\Big)\,. \tag{5.181}$$

The complexity C^blo is finally obtained from averaging $C^\mathrm{blo}\big(\bar{\boldsymbol{S}}\big)$ over all $\bar{\boldsymbol{S}}$.

For DSTM with group codes, however, we find from inserting (5.167) and (5.171) into (5.172) that $\Phi_{\eta_n}(v)$ depends on $\bar{\boldsymbol{S}}_n$ and $\tilde{\bar{\boldsymbol{S}}}_n$ only through

$$\bar{\boldsymbol{Z}}_n \stackrel{\triangle}{=} \big[\boldsymbol{Z}_n^\mathsf{H}, \ldots, \boldsymbol{Z}_{N-1}^\mathsf{H}\big]^\mathsf{H} \tag{5.182}$$

with

$$\boldsymbol{Z}_i \stackrel{\triangle}{=} \tilde{\boldsymbol{S}}_i^\mathsf{H} \boldsymbol{S}_i. \tag{5.183}$$

Consequently, the set of all $\bar{\boldsymbol{Z}}_n$ is equal to the set of all $\big[\tilde{\boldsymbol{S}}_n^\mathsf{H}, \ldots, \tilde{\boldsymbol{S}}_N^\mathsf{H}\big]^\mathsf{H}$ regardless of $\bar{\boldsymbol{S}}_n$ and averaging over $\bar{\boldsymbol{S}}$ is thus not required, i.e.

$$C^\mathrm{blo}\big(\bar{\boldsymbol{S}}\big) = C^\mathrm{blo}. \tag{5.184}$$

5.5.1.2.4 Special Fading Scenarios Evaluation of (5.181) can be computationally expensive for arbitrary fading channels and large (non–group) constellations and observation window sizes N. Therefore, it is interesting to consider DSTM with group constellations in two particular fading scenarios, (i) block fading and (ii) an artifical scenario which we refer to as "semi–block fading" to approximate very slow and slow fading, respectively. In these cases it is possible to determine sets of sequences $\tilde{\bar{\boldsymbol{S}}}_n$ that are equivalent with respect to $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \le \rho \mid \bar{\boldsymbol{S}}_n\big)$, such that only a relatively small number of set representatives need to be considered in (5.180).

Block Fading: For this classical special case, the channel is assumed to be constant over the entire observation window, i.e.

$$\bar{\boldsymbol{H}} = \mathbf{1}_{N,1} \otimes \boldsymbol{H} \tag{5.185}$$

and $\boldsymbol{\Psi}^{\mathsf{t}} = \mathbf{1}_{N,N}$. Using the identity $(\mathbf{1}_{N,N} + x\boldsymbol{I}_N)^{-1} = (\boldsymbol{I}_N - (N+x)^{-1}\mathbf{1}_{N,N})/x$ the MSDD metric of (3.20) reduces to

$$d_n\Big(\tilde{\bar{\boldsymbol{S}}}_n\Big) = \frac{1}{\sigma_n^2}\left(\big|\big|\bar{\boldsymbol{R}}_n\big|\big|^2 - \frac{1}{\tilde{n}+\sigma_n^2}\,||\boldsymbol{\Xi}_n\boldsymbol{H} + \boldsymbol{N}'_n||^2\right), \tag{5.186}$$

where

$$\boldsymbol{N}'_n \;\stackrel{\triangle}{=}\; \tilde{\bar{\boldsymbol{S}}}_n^{\mathsf{H}}\bar{\boldsymbol{N}}_n \tag{5.187}$$

$$\boldsymbol{\Xi}_n \;\stackrel{\triangle}{=}\; \boldsymbol{I}_{N_{\mathrm{T}}} + \sum_{i=n}^{N-1}\boldsymbol{Z}_i \tag{5.188}$$

with $\boldsymbol{Z}_i$ as defined in (5.183) were introduced. As the distributions of $\boldsymbol{H}$ and $\boldsymbol{N}_i$ are invariant to unitary transformations, $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \le \rho \mid \bar{\boldsymbol{S}}_n\big)$ depends on the true and the candidate transmit signal only through the singular values $\zeta_{\boldsymbol{\Xi}_n,m}$ of $\boldsymbol{\Xi}_n$, $1 \le m \le N_{\mathrm{T}}$, $1 \le n \le N-1$.

While, for general DSTM group codes, it does not appear to be feasible to give closed–form expression for sets of sequences $\tilde{\bar{\boldsymbol{S}}}_n$ that lead to the same $\zeta_{\boldsymbol{\Xi}_n,m}$, equivalence sets with respect to $\boldsymbol{\Xi}_n$ can be determined as follows. $\boldsymbol{\Xi}_n$ is the sum of an identity matrix and $\tilde{n}-1$ random matrices $\boldsymbol{Z}_i$ uniformly iid over the set $\boldsymbol{\mathcal{V}}$, i.e. its support is given by the Cartesian product of $(N-n)$ sets $\boldsymbol{\mathcal{V}}$ and shifted by $\boldsymbol{I}_{N_{\mathrm{T}}}$. Formally, we can therefore determine the distribution $f_{\boldsymbol{\Xi}_n}(\boldsymbol{\Xi})$ of $\boldsymbol{\Xi}_n$ by $(N-n)$–fold convolution, which leads to

$$f_{\boldsymbol{\Xi}_n}(\boldsymbol{\Xi}) = \sum_{\nu_1=0}^{\tilde{n}-1}\cdots\sum_{\nu_l=0}^{\tilde{n}-1-\sum_{i=1}^{l-1}\nu_i}\cdots\sum_{\nu_{L-1}=0}^{\tilde{n}-1-\sum_{i=1}^{L-2}\nu_i}\left(\prod_{l=1}^{L-1}\binom{\tilde{n}-1-\sum_{i=1}^{l-1}\nu_i}{\nu_l}\right)\cdot\delta_{\mathrm{M}}\left[\boldsymbol{\Xi} - \left(\boldsymbol{I}_{N_{\mathrm{T}}} + \sum_{l=1}^{L}\nu_l\boldsymbol{V}^{(l)}\right)\right], \tag{5.189}$$

where $\nu_L \stackrel{\triangle}{=} \tilde{n} - 1 - \sum_{i=1}^{L-1}\nu_i$ ($\delta_{\mathrm{M}}[\boldsymbol{X}-\boldsymbol{Y}]$: generalized Kronecker–$\delta$ with respect to complex–valued matrix arguments, which equals one if $\boldsymbol{X} = \boldsymbol{Y}$ and zero if $\boldsymbol{X} \neq \boldsymbol{Y}$). Every summand in (5.189) represents a set $\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}$, $\boldsymbol{q} \stackrel{\triangle}{=} [\nu_1,\ldots,\nu_L]^{\mathsf{T}}/(\tilde{n}-1)$, of sequences $\bar{\boldsymbol{Z}}_n$ that consist of ν_1 symbols $\boldsymbol{V}^{(1)}$, ν_2 symbols $\boldsymbol{V}^{(2)}$, etc. in arbitrary order, i.e. that are equivalent with respect to $\boldsymbol{\Xi}_n$. The product of binomial coefficients in (5.189) gives the number of possible ways to arrange those symbols, i.e. the cardinality $|\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}|$ of each set $\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}$.

Thus, the expected number C_n^{blo} of candidates $\tilde{\bar{\boldsymbol{S}}}_n$ inside the sphere can be computed by

$$C_n^{\mathrm{blo}} = \sum_{\bar{\boldsymbol{Z}}_n \in \boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)},\forall \boldsymbol{q}} |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}|\,\Pr\Big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \le \rho \,\Big|\, \bar{\boldsymbol{S}}_n\Big), \tag{5.190}$$

with $\tilde{\bar{\boldsymbol{S}}}_n = \mathrm{diag}\{\boldsymbol{Z}_n^{\mathsf{H}}, \ldots, \boldsymbol{Z}_{N-1}^{\mathsf{H}}, \boldsymbol{I}_{N_\mathrm{T}}\}\,\bar{\boldsymbol{S}}_n$, some fixed $\bar{\boldsymbol{S}}_n$ and only one representative per set $\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}$. This approach of computing C_n^{blo} is similar to that of [HV02] in that equivalence sets of candidate sequences are used to compute the expected complexity. However, the criterion with respect to which equivalence sets are formed is different from that in [HV02] and thus the methods used in [HV02] are not applicable here.

LPSK: For the special case of single–antenna transmission, i.e. L–ary phase–shift keying (LPSK), the metric $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$ depends on the candidate $\tilde{\bar{\boldsymbol{S}}}_n$ and the true transmit signal $\bar{\boldsymbol{S}}_n$ only through the quantity [cf. (5.186) and (5.188)]

$$\zeta_{\Xi_n,1} = \left| 1 + \sum_{i=n}^{N-1} \boldsymbol{Z}_i \right|^2 . \tag{5.191}$$

Again, the exact distribution of $\zeta_{\Xi_n,1}$ does not appear to be analytically tractable for arbitrary L, whereas the distribution of Ξ_n is again of the form of (5.189). For $L = 2$ (BPSK) and $L = 4$ (QPSK), on the other hand, solutions for the distribution of $\zeta_{\Xi_n,1}$ can be given in closed form.

In particular, we obtain in the case of BPSK

$$f_{\zeta_1}(\zeta) = \begin{cases} \displaystyle\sum_{\nu=0}^{\lceil(\tilde{n}-1)/2\rceil-1} \binom{\tilde{n}}{\nu} \delta[\zeta-(\tilde{n}-2\nu)] + \binom{\tilde{n}-1}{\lceil\frac{\tilde{n}-1}{2}\rceil} \delta\left[\zeta-\left(\tilde{n}-2\left\lceil\frac{\tilde{n}-1}{2}\right\rceil\right)\right] & \tilde{n} \text{ even} \\ \displaystyle\sum_{\nu=0}^{\lceil(\tilde{n}-1)/2\rceil} \binom{\tilde{n}}{\nu} \delta[\zeta-(\tilde{n}-2\nu)] & \tilde{n} \text{ odd} \end{cases} \tag{5.192}$$

($\delta[i-j]$: Kronecker–δ function which returns 1 if $i = j$ and 0 otherwise), i.e. we can compute the average complexity C_n^{blo} based on the evaluation of $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \le \rho \mid \bar{\boldsymbol{S}}_n\big)$ for $\lceil(\tilde{n}+1)/2\rceil+1$ representatives $\tilde{\bar{\boldsymbol{S}}}$ for which $\bar{\boldsymbol{Z}}_n$ consists of $(\tilde{n}-1-\nu)$ entries 1 and ν entries -1, $0 \le \nu \le \tilde{n}-1$. This means that the complexity of computing C_n^{blo} is only linear in $(N-n)$ for BPSK.

In the case of QPSK one finds that with respect to $\zeta_{\Xi_n,1}$ there are $\frac{1}{2}(\lceil(\tilde{n}+1)/2\rceil+1)(\lceil(\tilde{n}+1)/2\rceil+2)$ equivalence sets $\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}$, with $\boldsymbol{q} = [\tilde{n}-1-\nu, \nu-\mu, \mu, 0]^{\mathsf{T}}/(\tilde{n}-1)$, $0 \le \nu \le \lceil(\tilde{n}-1)/2\rceil$, $0 \le \mu \le \nu$, whose representatives $\bar{\boldsymbol{Z}}_n$ consist of $(\tilde{n}-1-\nu)$ entries 1, μ entries -1 and $(\nu-\mu)$ entries j. The cardinalities of these sets read

$$\left|\boldsymbol{\mathcal{T}}^{(n)}_{[\tilde{n}-1-\nu,\nu-\mu,\mu,0]^{\mathsf{T}}/(\tilde{n}-1)}\right| = \binom{\tilde{n}}{\nu} \cdot \binom{\tilde{n}}{\mu} \cdot \begin{cases} \frac{1}{4} & \text{if } \mathrm{Re}\{\Xi_n\} = \mathrm{Im}\{\Xi_n\} = 0 \\ 1 & \text{elseif } \mathrm{Re}\{\Xi_n\} = \mathrm{Im}\{\Xi_n\} \neq 0 \text{ or } \mathrm{Im}\{\Xi_n\} = 0 \\ 2 & \text{otherwise} \end{cases}, \tag{5.193}$$

$0 \le \nu \le \lceil(\tilde{n}-1)/2\rceil$, $0 \le \mu \le \nu$. This means that the complexity of computing C_n^{blo} is only quadratic in $(N-n)$ for QPSK.

Semi–Block Fading: This artificial model is supposed to approximate slow and moderately fast fading while allowing to determine equivalence sets to reduce the complexity of assessing the complexity of FP–MSDSD. Here, the channel is constant for the transmission of $N/N_{\mathrm{SB}} \in \mathbb{N}$ successive symbols $\boldsymbol{S}[k]$ and the correlation among the N_{SB} channel states is described by the $(N_{\mathrm{SB}} \times N_{\mathrm{SB}})$–dimensional correlation matrix $\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}$ such that

$$\boldsymbol{\Psi}^{\mathrm{t}} = \boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}} \otimes \mathbf{1}_{N/N_{\mathrm{SB}},N/N_{\mathrm{SB}}}.^{11} \tag{5.194}$$

Based on the eigenvalue decomposition

$$\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}} = \boldsymbol{U}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} \boldsymbol{\Lambda}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} \boldsymbol{U}^{\mathsf{H}}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} \tag{5.195}$$

$\boldsymbol{\Psi}^{\mathrm{t}}$ has rank N_{SB} and can be expressed as

$$\boldsymbol{\Psi}^{\mathrm{t}} = (\boldsymbol{U}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} \otimes \mathbf{1}_{N/N_{\mathrm{SB}},1}) \boldsymbol{\Lambda}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} (\boldsymbol{U}^{\mathsf{H}}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} \otimes \mathbf{1}_{1,N/N_{\mathrm{SB}}}). \tag{5.196}$$

Consequently, the channel matrix $\bar{\boldsymbol{H}}$ [cf. (5.21)] can —based on the Karhunen–Loève expansion— be written as

$$\bar{\boldsymbol{H}} = \left(\left(\left(\boldsymbol{U}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} \otimes \mathbf{1}_{N/N_{\mathrm{SB}},1} \right) \sqrt{\boldsymbol{\Lambda}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}}} \right) \otimes \boldsymbol{I}_{N_{\mathrm{T}}} \right) \bar{\boldsymbol{W}} \tag{5.197}$$

where $\bar{\boldsymbol{W}}$ denotes an $(N_{\mathrm{SB}}N_{\mathrm{T}} \times N_{\mathrm{R}})$–dimensional matrix of iid $\mathcal{N}_{\mathrm{c}}(0,1)$ entries. Furthermore, we can apply Woodbury's identity for matrix inversion (cf. e.g. [HJ85]) to obtain

$$\boldsymbol{M} = \frac{1}{\sigma_n^2} \left(\boldsymbol{I}_N - (\boldsymbol{U}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} \otimes \mathbf{1}_{N/N_{\mathrm{SB}},1}) \boldsymbol{\Lambda}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} \left(\boldsymbol{\Lambda}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{SB}}} \right)^{-1} (\boldsymbol{U}^{\mathsf{H}}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} \otimes \mathbf{1}_{1,N/N_{\mathrm{SB}}}) \right). \tag{5.198}$$

Based on the assumption that the non–zero eigenvalues of $\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}$ are distinct, such that $\boldsymbol{\Lambda}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}}$ and thereby $\boldsymbol{\Lambda}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} \left(\boldsymbol{\Lambda}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}} + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{SB}}} \right)^{-1}$ are diagonal matrices we can then express the ML–MSDD path metric as

$$d_n\left(\tilde{\bar{\boldsymbol{S}}}_n\right) = \tag{5.199}$$
$$\frac{1}{\sigma_n^2} \left[\left|\left| \bar{\boldsymbol{R}}_n \right|\right|^2 - \sum_{i=1}^{N_{\mathrm{SB}}} \frac{\lambda_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}},i}}{\lambda_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}},i} + \sigma_n^2} \sum_{j=1}^{N_{\mathrm{R}}} \left|\left| \sum_{l=1}^{N_{\mathrm{SB}}} \sqrt{\lambda_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}},l}} \left(\sum_{k=\lceil n/N_{\mathrm{SB}} \rceil}^{N_{\mathrm{SB}}} u^*_{k,i} u_{n,l} \boldsymbol{\Xi}_{k,n} \right) \boldsymbol{w}_{j,l} + \boldsymbol{n}_{i,j} \right|\right|^2 \right],$$

where $u_{i,j}$ and $\lambda_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}},i}$ are the elements of $\boldsymbol{U}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}}$ as introduced in (5.195), $\boldsymbol{w}_{j,l}$ and $\boldsymbol{n}_{i,j}$ are mutually independent N_{T}–dimensional vectors of iid $\mathcal{N}_{\mathrm{c}}(0,1)$ and $\mathcal{N}_{\mathrm{c}}(0,\sigma_n^2)$ distributed

[11] For an appropriate choice of N_{SB} and $\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}$ one can consider the eigenvalues of the autocorrelation matrix $\boldsymbol{\Psi}^{\mathrm{t}}$ of the fading process to be approximated by the semi–block fading model and determine N_{SB} as the smallest number that (i) is greater than the number of dominant eigenvalues of $\boldsymbol{\Psi}^{\mathrm{t}}$ and (ii) satisfies $N/N_{\mathrm{SB}} \in \mathbb{N}$. $\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}}$ should then be chosen as $\boldsymbol{\Psi}^{\mathrm{t}}_{\mathrm{SB}} = \mathrm{toeplitz}\{\psi^{\mathrm{t}}[0], \psi^{\mathrm{t}}[xN/N_{\mathrm{SB}}], \ldots, \psi^{\mathrm{t}}[xN(N_{\mathrm{SB}}-1)/N_{\mathrm{SB}}]\}$, where $x = N_{\mathrm{T}}$ and $x = 1$ for DSTM and BDSTM, respectively.

random variables, respectively, and

$$\boldsymbol{\Xi}_{k,n} \triangleq \sum_{i=\max(n,(k-1)N/N_{\mathrm{SB}}+1)}^{kN/N_{\mathrm{SB}}} \boldsymbol{Z}_i \,, \quad 1 \leq k \leq N_{\mathrm{SB}} - 1 \tag{5.200}$$

$$\boldsymbol{\Xi}_{N_{\mathrm{SB}},n} \triangleq \boldsymbol{I}_{N_{\mathrm{T}}} + \sum_{i=\max(n,N-N/N_{\mathrm{SB}}+1)}^{N-1} \boldsymbol{Z}_i, \tag{5.201}$$

with $\boldsymbol{Z}_i$ as defined in (5.183). Consequently, sequences $\bar{\boldsymbol{Z}}_n$ for which all $\boldsymbol{\Xi}_{k,n}$, $1 \leq k \leq N_{\mathrm{SB}}$, are equal lead to the same metric $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$ and therefore lie inside the sphere with the same probability. Representatives

$$\bar{\boldsymbol{Z}}_{k,n} \triangleq \Big[\boldsymbol{Z}^{\mathsf{H}}_{\max(n,(k-1)N/N_{\mathrm{SB}}+1)}, \ldots, \boldsymbol{Z}^{\mathsf{H}}_{kN/N_{\mathrm{SB}}}\Big]^{\mathsf{H}}, \quad 1 \leq k \leq N_{\mathrm{SB}} - 1 \tag{5.202}$$

$$\bar{\boldsymbol{Z}}_{N_{\mathrm{SB}},n} \triangleq \Big[\boldsymbol{Z}^{\mathsf{H}}_{\max(n,N-N/N_{\mathrm{SB}}+1)}, \ldots, \boldsymbol{Z}^{\mathsf{H}}_{N-1}\Big]^{\mathsf{H}} \tag{5.203}$$

and cardinalities of equivalence sets with respect to $\boldsymbol{\Xi}_{k,n}$, $1 \leq k \leq N_{\mathrm{SB}}$, are obtained in the same way as for block fading [cf. (5.189)].

In order to determine representatives for and cardinalities of equivalence sets with respect to $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$ one has to form the Cartesian product of the above equivalence sets, i.e. form all possible concatenations $\bar{\boldsymbol{Z}}_n = \big[\bar{\boldsymbol{Z}}^{\mathsf{H}}_{1,n}, \ldots, \bar{\boldsymbol{Z}}^{\mathsf{H}}_{N_{\mathrm{SB}},n}\big]^{\mathsf{H}}$ of representatives $\bar{\boldsymbol{Z}}_{k,n}$, $1 \leq k \leq N_{\mathrm{SB}}$, and multiply the cardinalities of the respective constituent equivalence sets. Thus, choosing a larger N_{SB} increases the complexity of the computation and in the limiting case of $N_{\mathrm{SB}} = N$ this leads naturally to the result that all "equivalence sets" are of size one, which means that all L^{N-n} relevant $\tilde{\bar{\boldsymbol{S}}}_n$ have to be considered individually.

5.5.1.3 Asymptotic Complexity Analysis

The complexity analysis presented in the previous sections is —when considering arbitrary constellations and fading scenarios— useful for exact computation of the complexity of FP–MSDSD for moderate observation window lengths N and constellation sizes L. In this section, we derive an expression C_{as} for the asymptotic complexity C^{blo} of FP–MSDSD for $N \to \infty$ that is tight in the sense that

$$\lim_{N\to\infty} \frac{C^{\mathrm{blo}}}{C_{\mathrm{as}}} = 1 \tag{5.204}$$

and provides interesting insights into the dependence of the complexity on system parameters. Although we do not have stringent proof our numerical results strongly suggest that the complexity of FP–MSDSD grows with increasing memory of the fading process (see Section 5.5.3). This conjecture is also supported by an interpretation of the result derived in the following. For this reason, and for analytical tractability, we consider the block–fading channel case. While we

only give an explicit proof for DSTM with group codes, it should become clear in the following that this result can be expected to also hold for non–group codes, as it hinges on the fact that $\boldsymbol{\Xi}_n/\tilde{n}$ in the limit of $\tilde{n} \to \infty$ tends to zero for almost all pairs of sequences $\tilde{\bar{\boldsymbol{S}}}_n$ and $\bar{\boldsymbol{S}}_n$ (see below).

Derivation: The derivation of C_{as} consists of two major steps. In the first step, we consider $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$ and determine, which terms can be neglected in the limit of $N \to \infty$, thereby obtaining a significantly simplified expression for $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$. In the second step, we make use of the method of types (cf. e.g. [CT91]) to group the $\tilde{\bar{\boldsymbol{S}}}_n$ into "type classes". Based on this, we determine a so–called "strongly typical set" [CT91] of $\tilde{\bar{\boldsymbol{S}}}_n$ that lie inside the sphere with the same probability. A proof that the contribution to the average complexity of the $\tilde{\bar{\boldsymbol{S}}}_n$ that do not belong to the strongly typical set is negligible concludes the derivation.

Using (5.183), (5.185)–(5.188), and $\bar{\boldsymbol{S}}_n^{\mathsf{H}}\bar{\boldsymbol{S}}_n = \tilde{n}\boldsymbol{I}_{N_{\mathrm{T}}}$ with $\tilde{n} = N - n + 1$, we can write[12]

$$
\begin{aligned}
d_n\Big(\tilde{\bar{\boldsymbol{S}}}_n\Big)\,\frac{\sigma_n^2}{\tilde{n}} &= \frac{1}{\tilde{n}}\left|\left|\bar{\boldsymbol{R}}_n\right|\right|^2 - \frac{1}{\tilde{n}(\tilde{n}+\sigma_n^2)}\left|\left|\boldsymbol{\Xi}_n\boldsymbol{H} + \tilde{\bar{\boldsymbol{S}}}_n^{\mathsf{H}}\bar{\boldsymbol{N}}_n\right|\right|^2 && (5.205)\\
&= ||\boldsymbol{H}||^2 + \frac{1}{\tilde{n}}\left|\left|\bar{\boldsymbol{N}}_n\right|\right|^2 + \frac{2}{\tilde{n}}\mathrm{Re}\left\{\mathrm{tr}\Big\{\boldsymbol{H}\bar{\boldsymbol{N}}_n^{\mathsf{H}}\bar{\boldsymbol{S}}_n\Big\}\right\} && (5.206)\\
&\quad -\frac{1}{\tilde{n}(\tilde{n}+\sigma_n^2)}\left(||\boldsymbol{\Xi}_n\boldsymbol{H}||^2 + 2\mathrm{Re}\left\{\mathrm{tr}\Big\{\boldsymbol{\Xi}_n\boldsymbol{H}\bar{\boldsymbol{N}}_n^{\mathsf{H}}\tilde{\bar{\boldsymbol{S}}}_n\Big\}\right\} + \left|\left|\tilde{\bar{\boldsymbol{S}}}_n^{\mathsf{H}}\bar{\boldsymbol{N}}_n\right|\right|^2\right).
\end{aligned}
$$

Later in this section, we will show that terms C_n^{blo} with n close to N are in the limit of $N \to \infty$ negligible in C^{blo}. Let us therefore discuss the asymptotic behavior of some of the terms in (5.206) for $\tilde{n} \to \infty$.

- $\left|\left|\bar{\boldsymbol{N}}_n\right|\right|^2/\tilde{n}$: As the elements of $\bar{\boldsymbol{N}}_n$ are iid $\mathcal{N}_{\mathrm{c}}\big(0,\sigma_n^2\big)$ random variables, $\left|\left|\bar{\boldsymbol{N}}_n\right|\right|^2/\tilde{n}$ is a $\chi^2\big(N_{\mathrm{T}}N_{\mathrm{R}}\sigma_n^4/\tilde{n}, 2\tilde{n}N_{\mathrm{T}}N_{\mathrm{R}}\big)$ random variable. As its variance for $\tilde{n} \to \infty$ tends to zero this term converges in distribution to its mean $N_{\mathrm{T}}N_{\mathrm{R}}\sigma_n^2$.

- $\left|\left|\tilde{\bar{\boldsymbol{S}}}_n^{\mathsf{H}}\bar{\boldsymbol{N}}_n\right|\right|^2/(\tilde{n}(\tilde{n}+\sigma_n^2))$: The elements of $\tilde{\bar{\boldsymbol{S}}}_n^{\mathsf{H}}\bar{\boldsymbol{N}}_n/\tilde{n}$ are $\mathcal{N}_{\mathrm{c}}\big(0,\sigma_n^2/\tilde{n}\big)$ random variables, as the $\tilde{\boldsymbol{S}}_n$ are unitary matrices. Thus $\left|\left|\tilde{\bar{\boldsymbol{S}}}_n^{\mathsf{H}}\bar{\boldsymbol{N}}_n\right|\right|^2/(\tilde{n}(\tilde{n}+\sigma_n^2)) \to 0$ in the limit of $\tilde{n} \to \infty$.

- Based on similar arguments, it is easy to see that the two remaining terms $2/\tilde{n}\,\mathrm{Re}\big\{\mathrm{tr}\big\{\bar{\boldsymbol{S}}_n\boldsymbol{H}\bar{\boldsymbol{N}}_n^{\mathsf{H}}\big\}\big\}$ and $2/(\tilde{n}(\tilde{n}+\sigma_n^2))\,\mathrm{Re}\big\{\mathrm{tr}\big\{\boldsymbol{\Xi}_n\boldsymbol{H}\bar{\boldsymbol{N}}_n^{\mathsf{H}}\tilde{\bar{\boldsymbol{S}}}_n\big\}\big\}$ can be neglected for $\tilde{n} \to \infty$, as well.

In summary, we obtain

$$
\lim_{\tilde{n}\to\infty}\Pr\Big(d_n\Big(\tilde{\bar{\boldsymbol{S}}}_n\Big) \le \rho \,\Big|\, \bar{\boldsymbol{S}}_n\Big) = \lim_{\tilde{n}\to\infty}\Pr\left(||\boldsymbol{H}||^2 - \frac{1}{\tilde{n}^2}||\boldsymbol{\Xi}_n\boldsymbol{H}||^2 \le \sigma_n^2\left(\frac{\rho}{\tilde{n}} - N_{\mathrm{T}}N_{\mathrm{R}}\right)\right). \quad (5.207)
$$

[12]We consider the normalized version $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)\frac{\sigma_n^2}{\tilde{n}}$ of the metric $d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big)$ because this clarifies the notion of what contributions to the metric can be neglected.

Based on (5.207) and the "method of types" [CT91], we prove in Appendix 5.A.6 that for DSTM with group constellations the asymptotic approximation

$$C_{\mathrm{as}} = \frac{\gamma\left(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2\left(\rho/N - N_{\mathrm{T}}N_{\mathrm{R}}\right)\right)}{\Gamma(N_{\mathrm{T}}N_{\mathrm{R}})} \cdot \frac{L}{L-1} \cdot L^{N-1} \tag{5.208}$$

is tight in the sense of (5.204).

Approximation: Recalling that a reasonable choice of the sphere radius is $\rho = (1 + \varepsilon_\rho)NN_{\mathrm{T}}N_{\mathrm{R}}$, with a small positive constant ε_ρ, we see that for not too large σ_n^2 the second argument of the incomplete Gamma function is small and we can approximate C_{as} using (5.296) by

$$C_{\mathrm{as}} \approx \frac{\left(\sigma_n^2\left(\rho/N - N_{\mathrm{T}}N_{\mathrm{R}}\right)\right)^{N_{\mathrm{T}}N_{\mathrm{R}}}}{(N_{\mathrm{T}}N_{\mathrm{R}})!} \cdot \frac{L}{L-1} \cdot L^{N-1} \tag{5.209}$$

$$= \frac{\left(\sigma_n^2\varepsilon_\rho N_{\mathrm{T}}N_{\mathrm{R}}\right)^{N_{\mathrm{T}}N_{\mathrm{R}}}}{(N_{\mathrm{T}}N_{\mathrm{R}})!} \cdot \frac{L}{L-1} \cdot L^{N-1}\,. \tag{5.210}$$

Discussion: Interestingly, (5.204) together with (5.208) and (5.210) states that the asymptotic computational complexity of FP–MSDSD (i) differs fundamentally from that of FP–SpD applied to coherent MIMO detection, cf. [JO05a, Section VII] in that the exponential rate of growth of complexity is independent of the SNR and (ii) is equal to that of brute–force MSDD which examines L^{N-1} candidates. This can be explained intuitively by looking back at the above derivation, from which we have that for large N and $\rho = (1 + \varepsilon_\rho)NN_{\mathrm{T}}N_{\mathrm{R}}$, $d_1\big(\tilde{\tilde{\boldsymbol{S}}}\big) - \rho = N(||\boldsymbol{H}||^2/\sigma_n^2 - \varepsilon_\rho N_{\mathrm{T}}N_{\mathrm{R}})$ for all but a negligible number of pairs $\big\{\tilde{\tilde{\boldsymbol{S}}}_n, \bar{\boldsymbol{S}}_n\big\}$. This means that for $||\boldsymbol{H}||^2$ smaller than $\varepsilon_\rho N_{\mathrm{T}}N_{\mathrm{R}}\sigma_n^2$ practically all $L^N/(L-1)$ (partial) candidate vectors lie inside the sphere and are thus examined by the FP–MSDSD algorithm. The probability for this event is given by the first factor in (5.208), which for $\rho = (1 + \varepsilon_\rho)NN_{\mathrm{T}}N_{\mathrm{R}}$ is independent of N and by (5.296) approximately proportional to $\sigma_n^{2N_{\mathrm{T}}N_{\mathrm{R}}}$, cf. (5.210). This means that the average complexity of FP–MSDSD for the block–fading channel is dominated by the worst–case complexity, whereas for FP–SpD applied to coherent MIMO detection the instantaneous complexity converges in distribution to the average complexity, cf. [JO05a].

Clearly, in continuous fading the probability of all matrices $\boldsymbol{H}_n$ containing only small entries reduces as N grows. Hence, we may —based on the above arguments— speculate that block fading represents the worst case fading scenario as far as detection complexity is concerned.

Non–Group Codes: The above result hinges on the fact that $\boldsymbol{\Xi}_n/\tilde{n}$ [cf. (5.188)] in the limit of $\tilde{n} \to \infty$ converges in distribution, i.e. for all but a negligible number of pairs $\big\{\tilde{\tilde{\boldsymbol{S}}}_n, \bar{\boldsymbol{S}}_n\big\}$, to $\boldsymbol{0}_{N_{\mathrm{T}},N_{\mathrm{T}}}$. Clearly, in the case of non–group codes a stringent proof based on the method of types is not feasible, as the support of the $\boldsymbol{Z}_i$ is usually not a finite set of elements. Still, it is not

difficult to show that the sequences $\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots$ and $\tilde{\boldsymbol{S}}_1, \tilde{\boldsymbol{S}}_2, \ldots$ can both be interpreted as output of a memoryless random source generating zero–mean unitary matrices. Thus, one can see that by the law of large numbers $\boldsymbol{\Xi}_n/\tilde{n}$ [cf. (5.188)] converges in distribution to $\mathbf{0}_{N_\mathrm{T},N_\mathrm{T}}$, even for non–group codes. So while we did not give an explicit proof that the rate of this convergence is sufficiently fast to justify the disregard of pairs $\left\{\tilde{\bar{\boldsymbol{S}}}_n, \bar{\boldsymbol{S}}_n\right\}$ for which $\boldsymbol{\Xi}_n/\tilde{n}$ deviates significantly from $\mathbf{0}_{N_\mathrm{T},N_\mathrm{T}}$ it can be expected that the asymptotic behavior of the complexity for non–group codes does not differ fundamentally from that for group codes. This was also confirmed by means of numerical examples.

5.5.2 MSDD Based on Combinatorial Geometry

In Section 3.2 we derived CG–MSDD assuming that the channel autocorrelation matrix $\boldsymbol{\Psi}^\mathrm{t}$ has a reduced rank N_λ. We also found that its computational complexity is of the order $O((\log_2(L)N)^{2N_\lambda N_\mathrm{T} N_\mathrm{R}})$, $L = \{2, 4\}$, which means it would be *polynomial* in N *if* N_λ was independent of N. We note that the latter assumption has also been made in [MAKA07], where the related MAPSqD algorithm was proposed, cf. Section 3.2.2.4. To shed light on its validity, we study the relation between N_λ and N considering the eigenvalue distribution of $\boldsymbol{\Psi}^\mathrm{t}$ for asymptotically large N. In fact, we will give insightful exact expressions for the complexity of CG–MSDD and show that for various popular fading models N_λ grows linearly with N. This clarifies that the complexity of combinatorial–geometry based MSDD, i.e. CG–MSDD and the algorithm of [MAKA07], is effectively exponential in N for these fading models.

Recall that the number of examined candidates $\tilde{\bar{\boldsymbol{S}}}$ is equal to the number of cells in the central arrangement, as there is an one–to–one correspondence between each cell and a $\tilde{\bar{\boldsymbol{S}}}$. Due to the symmetry of the central arrangement and the fact that $\tilde{\boldsymbol{S}}_N$ can be fixed as phase reference, the value of C^blo equals half the number of cells in the central arrangement. Thus, provided that $\boldsymbol{\Psi}^\mathrm{t}$ has N_λ non–zero eigenvalues $\lambda_{\boldsymbol{\Psi}^\mathrm{t},n}$, C^blo is given by [cf. (3.129)]

$$C^\mathrm{blo} = \sum_{n=0}^{2N_\lambda N_\mathrm{T} N_\mathrm{R}-1} \binom{Nx-1}{n}, \quad x = \begin{cases} 1, & \text{BPSK} \\ 2, & \text{QPSK} \end{cases}. \tag{5.211}$$

As already established before, C^blo is polynomial in N if N_λ is fixed.

The questions that need to be answered now are (i) how many eigenvalues need to be considered to have (near) ML–MSDD performance and (ii) what is the asymptotic behavior of C^blo. We will in the following answer the second question and come back to the first one with comments later on.

5.5.2.1 Asymptotic Complexity

Provided that the discrete–time fading process is such that its PSD

$$\Psi^{\mathrm{t}}(f) \stackrel{\triangle}{=} \sum_{k=-\infty}^{\infty} \psi^{\mathrm{t}}[k]\mathrm{e}^{-\mathrm{j}2\pi fTk} = \lim_{N\to\infty} \sum_{k=-N+1}^{N-1} \psi^{\mathrm{t}}[k]\mathrm{e}^{-\mathrm{j}2\pi fTk} \tag{5.212}$$

exists and that $|\Psi^{\mathrm{t}}(f)|$ is measurable in the sense of Lebesgue the eigenvalues of the autocorrelation (Toeplitz) matrix $\boldsymbol{\Psi}^{\mathrm{t}}$ converge to those of the corresponding circulant matrix $\mathrm{circ}\left\{\boldsymbol{\Psi}^{\mathrm{t}}_{:,1}\right\}$, [GS58, Section 5.2], see also [Gra71] for a tutorial treatment ($\mathrm{circ}\{\boldsymbol{x}\}$: circulant matrix with vector $\boldsymbol{x}$ as first column). It can further be shown that these eigenvalues converge to the spectrum of the corresponding autocorrelation function as the dimensions of the matrix grow to infinity [GS58, Gra71], i.e. the sequence

$$\beta_n \stackrel{\triangle}{=} \Psi^{\mathrm{t}}(n/(NT)), \quad 1 \le n \le N, \tag{5.213}$$

and the sequence $\lambda_{\boldsymbol{\Psi}^{\mathrm{t}},n}$, $1 \le n \le N$, of eigenvalues of $\boldsymbol{\Psi}^{\mathrm{t}}$ (sorted in order of decreasing magnitude) are for $N \to \infty$ asymptotically equally distributed. Since

$$\Psi^{\mathrm{t}}(f) = \frac{1}{T} \sum_{k=-\infty}^{\infty} \Psi^{\mathrm{t}}_{\mathrm{c}}(f - k/T), \tag{5.214}$$

where $\Psi^{\mathrm{t}}_{\mathrm{c}}(f)$ is the PSD of the continuous–time fading process, we conclude from (5.213) and (5.214) that if $\Psi^{\mathrm{t}}_{\mathrm{c}}(f)$ is non–zero over a continuous set of frequencies, the number of non–zero eigenvalues N_λ grows linearly with N, i.e. complexity of CG–MSDD is exponential in N.

Remark: Note that if repetition transmit diversity is employed and the different transmit antennas operate in an alternating fashion (compare regular cyclic DSTM versus BDSTM as discussed in Section 5.3.2) the sequence of eigenvalues $\lambda_{\boldsymbol{\Psi}^{\mathrm{t}},n}$ and the sequence $\Psi^{\mathrm{t}}(n/(NTN_{\mathrm{T}}))$, $1 \le n \le N$, are asymptotically equally distributed.

5.5.2.2 Further Discussion and Examples

For fading processes with effective normalized fading bandwidth $B_{h,\mathrm{eff}}T$ there are asymptotically

$$N_\lambda = 2\lfloor NB_{h,\mathrm{eff}}T \rfloor + 1 \tag{5.215}$$

non–zero eigenvalues and C^{blo} is of the order

$$O((\log_2(L)N)^{2(2\lfloor NB_{h,\mathrm{eff}}T\rfloor+1)N_{\mathrm{T}}N_{\mathrm{R}}-1}). \tag{5.216}$$

Besides fading models, such as the "rectangular" and Clarke's model, that have strictly band–limited PSDs $\Psi^{\mathrm{t}}_{\mathrm{c}}(f)$, there are a number of fading models, such as the "Gaussian", the "1st–order Butterworth" and the "2nd–order Butterworth" models, whose PSDs are not band–limited, cf. Table 4.1 for a list of these common fading models. Consequently, all eigenvalues

of $\mathbf{\Psi}^{\mathrm{t}}$ are non–zero, i.e. $\lambda_{\mathbf{\Psi}^{\mathrm{t}},n} > 0$, $1 \leq n \leq N$, and nothing would be gained in terms of complexity if *optimal* ML–MSDD was desired. However, the latter fading models exhibit quite a strong concentration of spectral power at low frequencies, i.e.

$$\frac{1}{N}\sum_{n=1}^{\tilde{N}_\lambda} \lambda_n = 1 - \varepsilon_\lambda, \tag{5.217}$$

with a small positive constant ε_λ and $\tilde{N}_\lambda \ll N$. Consequently, $\mathbf{\Psi}^{\mathrm{t}}$ is, depending on ε_λ, well approximated using only its largest *(dominant)* $\tilde{N}_\lambda$ eigenvalues $\lambda_{\mathbf{\Psi}^{\mathrm{t}},n}$, and thus near ML–MSDD performance can be achieved in this case as well. (See Section 5.4 for a performance analysis of CG–MSDD with $\tilde{N}_\lambda < N_\lambda$.)

Table 5.5 contains the (approximate) asymptotic eigenvalue distributions for the above–mentioned commonly used fading models, "approximate" in the sense that we used $\Psi^{\mathrm{t}}(f) \approx \frac{1}{T}\Psi^{\mathrm{t}}_{\mathrm{c}}(f)$ for $|f| \leq \tilde{N}_\lambda/(NT)$ also for non–band–limited PSDs. We also listed $\tilde{N}_\lambda$ as a function of ε_λ [cf. (5.217)], and the number of relevant eigenvalues $\lambda_{\mathbf{\Psi}^{\mathrm{t}},n}$ if $\varepsilon_\lambda = 10^{-4}$ is desired.[14]

Interestingly, we see from (5.211) in conjunction with Table 5.5 that the complexity of CG–MSDD is *independent of the SNR*, which is in stark contrast to tree–search based MSDD. Instead the complexity exponent is in all cases directly proportional to the normalized Doppler spread $B_{h,\mathrm{eff}}T$ and the observation window length N. This means that while the complexity of CG–MSDD is polynomial in N if the number of (dominant) eigenvalues $\lambda_{\mathbf{\Psi}^{\mathrm{t}},n}$ is fixed, the complexity of (near–ML) CG–MSDD for all these standard fading scenarios is in fact exponential in N, as the number N_λ / $\tilde{N}_\lambda$ of non–zero / dominant eigenvalues $\lambda_{\mathbf{\Psi}^{\mathrm{t}},n}$ asymptotically grows linearly in N. Only for certain pathological fading models such as the block–fading model, where the channel remains constant and $\mathbf{\Psi}^{\mathrm{t}}$ has only a single non–zero eigenvalue regardless of N, the complexity of CG–MSDD is truly polynomial in N.

For further illustration see Figs. 5.22 and 5.23, where we exemplarily plotted the eigenvalues $\lambda_{\mathbf{\Psi}^{\mathrm{t}},n}$ sorted in order of decreasing magnitude for $N = 1000$ and different values of $B_{h,\mathrm{eff}}T$ for Clarke's and the Gaussian–PSD fading model, respectively. One can observe that for finite window length N, a slightly larger number $N_\lambda > 2\lfloor NB_{h,\mathrm{eff}}T \rfloor + 1$ of eigenvalues than suggested by the asymptotic results in Table 5.5 has to be taken into account to achieve near ML–MSDD performance.

[13] Only $\varepsilon_\lambda(\tilde{N}_\lambda)$ can be given in closed form: $\varepsilon_\lambda = 1 - \frac{1}{\pi}\big(\frac{1}{2}\log\left(\frac{x^2+x/\sqrt{2}+1/4}{x^2-x/\sqrt{2}+1/4}\right) + \arctan\left(2\sqrt{2}x+1\right) + \arctan\left(2\sqrt{2}x-1\right)\big)$ with $x \triangleq \frac{\tilde{N}_\lambda - 1}{2NB_{h,\mathrm{eff}}T}$.

[14] We have verified that this is sufficient to achieve ML–MSDD performance for error rates down to at least 10^{-4}.

Asymptotic eigenvalue distribution	$\tilde{N}_\lambda(\varepsilon_\lambda)$ [see (5.217)] ($\tilde{N}_\lambda(\varepsilon_\lambda = 10^{-4})$)
Rectangular model:	
$\lambda_{\mathbf{\Psi}^{\mathrm{t}},n} = \begin{cases} \dfrac{1}{2B_hT}, & 1 \le n \le \tilde{N}_\lambda \\ 0, & \text{otherwise} \end{cases}$	$2\lfloor NB_{h,\mathrm{eff}}T \rfloor + 1\,(= N_\lambda)$ ($2\lfloor NB_{h,\mathrm{eff}}T \rfloor + 1$)
Clarke's model (land–mobile model):	
$\lambda_{\mathbf{\Psi}^{\mathrm{t}},n} = \begin{cases} \dfrac{N}{\pi\sqrt{(NB_{h,\mathrm{eff}}T)^2 - \lfloor n/2 \rfloor^2}}, & 1 \le n \le \tilde{N}_\lambda \\ 0, & \text{otherwise} \end{cases}$ (5.218)	$2\lfloor NB_{h,\mathrm{eff}}T \rfloor + 1\,(= N_\lambda)$ ($2\lfloor NB_{h,\mathrm{eff}}T \rfloor + 1$)
Gaussian–PSD model:	
$\lambda_{\mathbf{\Psi}^{\mathrm{t}},n} \approx \begin{cases} \dfrac{\exp\left(-\left(\frac{\lfloor n/2 \rfloor}{NB_{h,\mathrm{eff}}T}\right)^2\right)}{\sqrt{\pi}B_hT}, & 1 \le n \le \tilde{N}_\lambda \\ 0 & \text{otherwise} \end{cases}$ (5.219)	$2\left\lfloor NB_{h,\mathrm{eff}}T\,\mathrm{erfc}^{-1}(\varepsilon_\lambda) \right\rfloor + 1$ ($2\lfloor 2.75NB_{h,\mathrm{eff}}T \rfloor + 1$)
First–order Butterworth PSD model:	
$\lambda_{\mathbf{\Psi}^{\mathrm{t}},n} \approx \begin{cases} \dfrac{N^2B_{h,\mathrm{eff}}T}{\pi\left((NB_{h,\mathrm{eff}}T)^2 + \lfloor n/2 \rfloor^2\right)}, & 1 \le n \le \tilde{N}_\lambda \\ 0 & \text{otherwise} \end{cases}$	$2\left\lfloor NB_{h,\mathrm{eff}}T\tan\left(\frac{(1-\varepsilon_\lambda)\pi}{2}\right) \right\rfloor + 1$ ($2\lfloor 6366NB_{h,\mathrm{eff}}T \rfloor + 1$)
Second–order Butterworth PSD model:	
$\lambda_{\mathbf{\Psi}^{\mathrm{t}},n} \approx \begin{cases} \dfrac{2\sqrt{2}(B_{h,\mathrm{eff}}T)^3N^4}{\pi\left((NB_{h,\mathrm{eff}}T)^4 + 16\lfloor n/2 \rfloor^4\right)}, & 1 \le n \le \tilde{N}_\lambda \\ 0 & \text{otherwise} \end{cases}$	—[13] ($2\lfloor 7.21NB_{h,\mathrm{eff}}T \rfloor + 1$)

Table 5.5: Asymptotic distributions of the eigenvalues $\lambda_{\mathbf{\Psi}^{\mathrm{t}},n}$ for various fading models, cf. Table 4.1 and number $\tilde{N}_\lambda$ of "dominant" eigenvalues for approximation of $\mathbf{\Psi}^{\mathrm{t}}$ in CG–MSDD (in brackets the value of $\tilde{N}_\lambda$ when $\varepsilon_\lambda = 10^{-4}$).

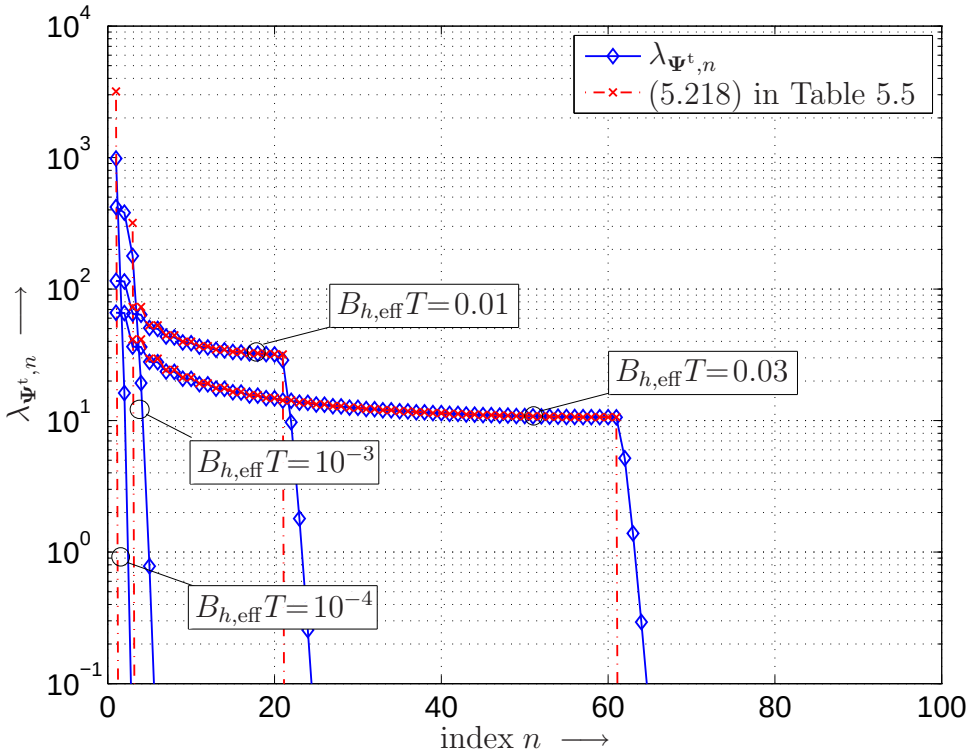

Figure 5.22: Eigenvalues $\lambda_{\mathbf{\Psi}^t,n}$ sorted according to their magnitude for Clarke's fading model, $N = 1000$ and different values of $B_{h,\mathrm{eff}}T$.

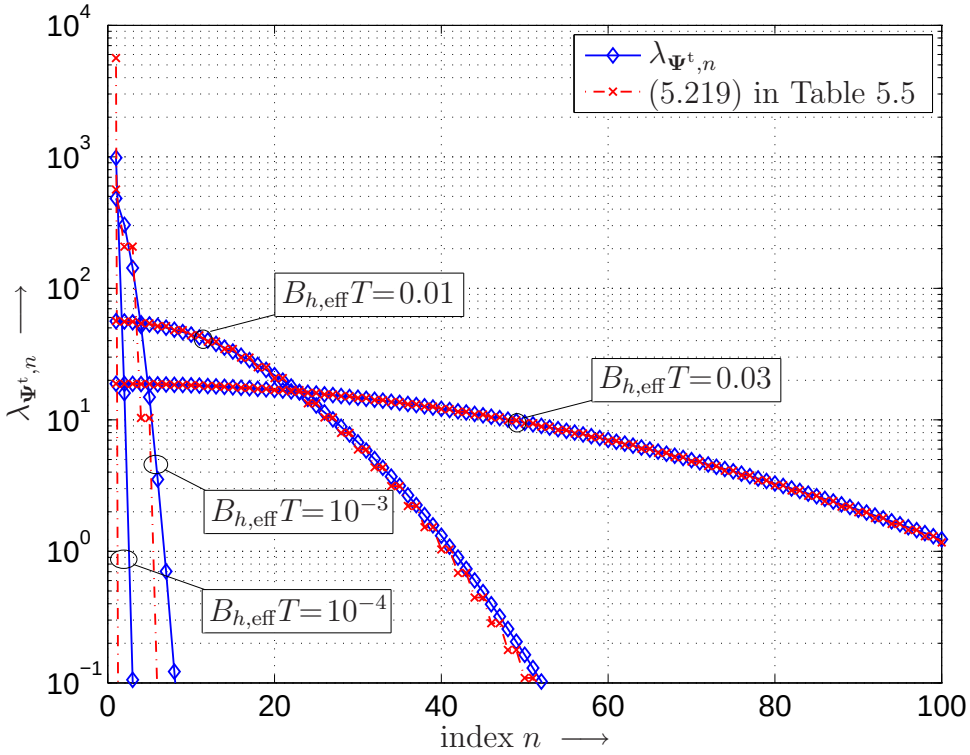

Figure 5.23: Eigenvalues $\lambda_{\mathbf{\Psi}^t,n}$ sorted according to their magnitude for the Gaussian PSD, $N = 1000$ and different values of $B_{h,\mathrm{eff}}T$.

5.5.3 Numerical Results

In this section, we provide numerical results to augment our analytical results from Sections 5.5.1 and 5.5.2. At this, our focus lies on tree–search methods. Here, we will first present some results to illustrate our analytical findings from Section 5.5.1, before we move on to simulation results for the complexity of the more involved tree–search methods such as MSDSD and Fano–MSDD, for which closed–form expressions to describe their complexity appear inaccessible. This is concluded by a comparison of the various MSDD implementations with respect to achievable performance when the instantaneous complexity per decoded symbol is limited, i.e. if the search may be truncated.

In the second part of this section, we compare the computational complexity of tree–search MSDD with that of CG–MSDD.

Unless explicitly stated otherwise, we consider a spatially uncorrelated MIMO fading channel with temporal correlation according to Clarke's fading model.

5.5.3.1 Tree–Search MSDD

C^{blo} **vs. SNR:** Fig. 5.24 compares the exact complexity of FP–MSDSD with $N = 10$ computed as described in Section 5.5.1.2 for different fading scenarios as a function of the SNR. In particular, DQPSK transmission ($N_{\mathrm{T}} = N_{\mathrm{R}} = 1$) over block, semi–block with $N_{\mathrm{SB}} = 2$, and continuous fading channels is considered. In the semi–block fading model, where in general the channel remains constant over $NN_{\mathrm{T}}/N_{\mathrm{SB}}$ modulation intervals, the off–diagonal element $\psi^{\mathrm{t}}_{\mathrm{SB},1,2}$ of (2×2)–dimensional $\mathbf{\Psi}^{\mathrm{t}}_{\mathrm{SB}}$ describing the correlation between the channel coefficients of the two fading blocks is chosen as $\mathrm{J}_0(2\pi 0.01N/2) = 0.975$ to approximate continuous fading with normalized fading bandwidth $B_hT = 0.01$. The radius ρ is chosen such that $\Pr\big(d_1\big(\bar{\boldsymbol{S}}\big) \leq \rho \mid \bar{\boldsymbol{S}}\big) = 0.99$, i.e. the correct transmit sequence is found inside the sphere with 99 % probability. Simulation results for $10\log_{10}(E_{\mathrm{b}}/\mathcal{N}_0) \geq 15\,\mathrm{dB}$ (not shown in the figure) coincide with those obtained from our analysis.

It can be seen that the number of candidate signals inside the sphere rapidly decreases as the SNR increases and converges to the lowest possible complexity for *successful* decoding $C^{\mathrm{blo}}_{\mathrm{min}}(N) \triangleq N - 1$. Except for the high–SNR region the semi–block fading model serves as good approximation for the continuous fading model. We further find our intuitive reasoning at the end of Section 5.5.1.3 confirmed, as the complexity of FP–MSDSD is largest for the case of block fading and reduces with decreasing memory of the fading process.

C^{blo} **vs.** N**:** In Fig. 5.25 we consider the complexity of FP–MSDSD for dicyclic DSTM with $N_{\mathrm{T}} = 4$, $R = 0.5$ and $N_{\mathrm{R}} = 1$ when transmitting over a block–fading and a QS–fading channel with normalized fading bandwidth $B_hT = 0.01$ at various SNR values. The different types of

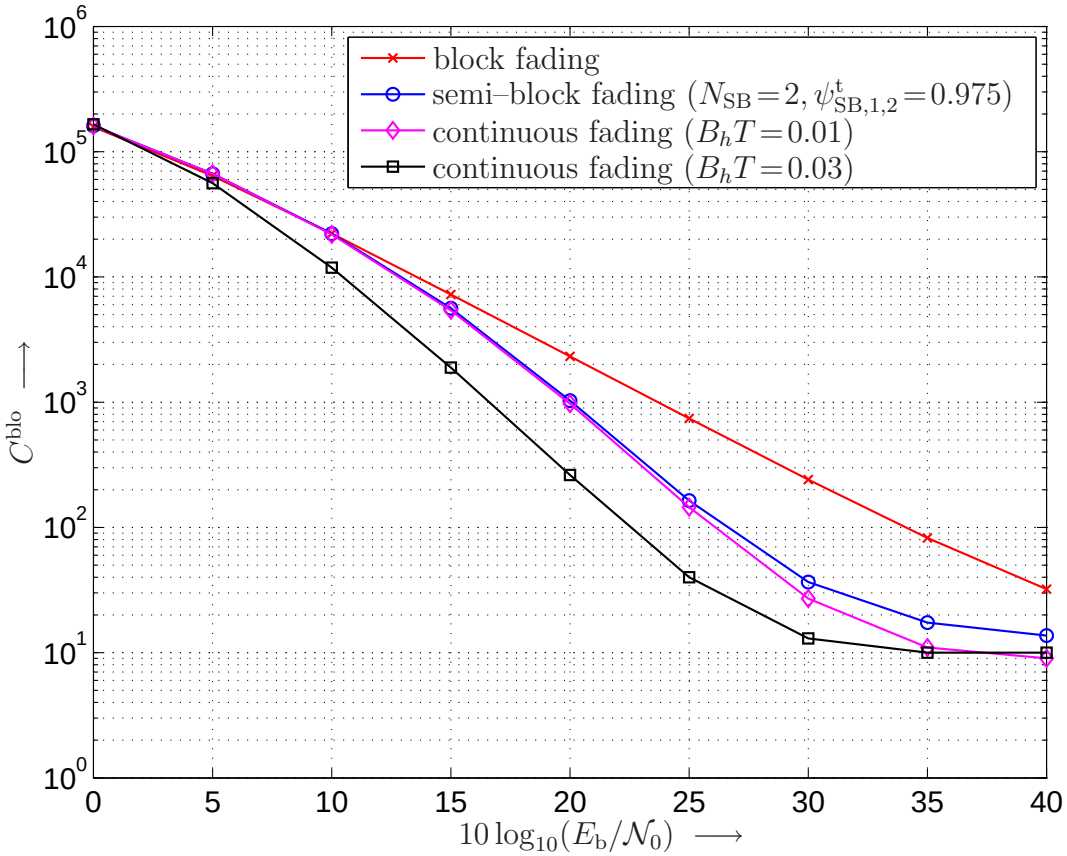

Figure 5.24: Complexity C^{blo} of FP–MSDSD vs. SNR for different fading scenarios. Parameters: DQPSK, $N_{\text{T}} = N_{\text{R}} = 1$, $N = 10$, ρ such that $\Pr\big(d_1(\bar{\boldsymbol{S}}) \leq \rho \mid \bar{\boldsymbol{S}}\big) = 0.99$.

curves in Fig. 5.25 are (i) the exact complexity for the block–fading (solid lines) and the QS–fading channel with $B_hT = 0.01$ (dashed lines), (ii) the lowest possible complexity $C^{\text{blo}}_{\text{min}}(N) = N - 1$ (dash–dotted line), and (iii) the complexity of brute–force ML MSDD $C^{\text{blo}}_{\text{bf}}(N) \triangleq L^{N-1}$ (dash–dotted line). One can observe that while the complexity for the block–fading channel is lower for small values of N and low SNR, it grows at a far higher rate towards larger N than for the QS fading channel. It can be suspected that long backtrackings to levels close to the root of the search tree occur more frequently for the block–fading channel with strong dependencies between all channel coefficients considered in the observation window, while for the QS–fading channel the dependences are limited due to the non–zero bandwidth, i.e. finite memory, of the fading process. Comparisons with $C^{\text{blo}}_{\text{bf}}$ and $C^{\text{blo}}_{\text{min}}$ illustrate the complexity savings feasible compared to brute–force ML MSDD and the increase in complexity compared to a DFE–type decoder.

The lower complexity bound (cf. Section 5.5.1.1) and the asymptotic complexity (cf. Section 5.5.1.3) are illustrated and compared in Fig. 5.26 for the example of cyclic codes with $N_{\text{T}} = 3$, $R = 1$, $N_{\text{R}} = 1$, and a block–fading channel. Again, the search radius ρ is adjusted

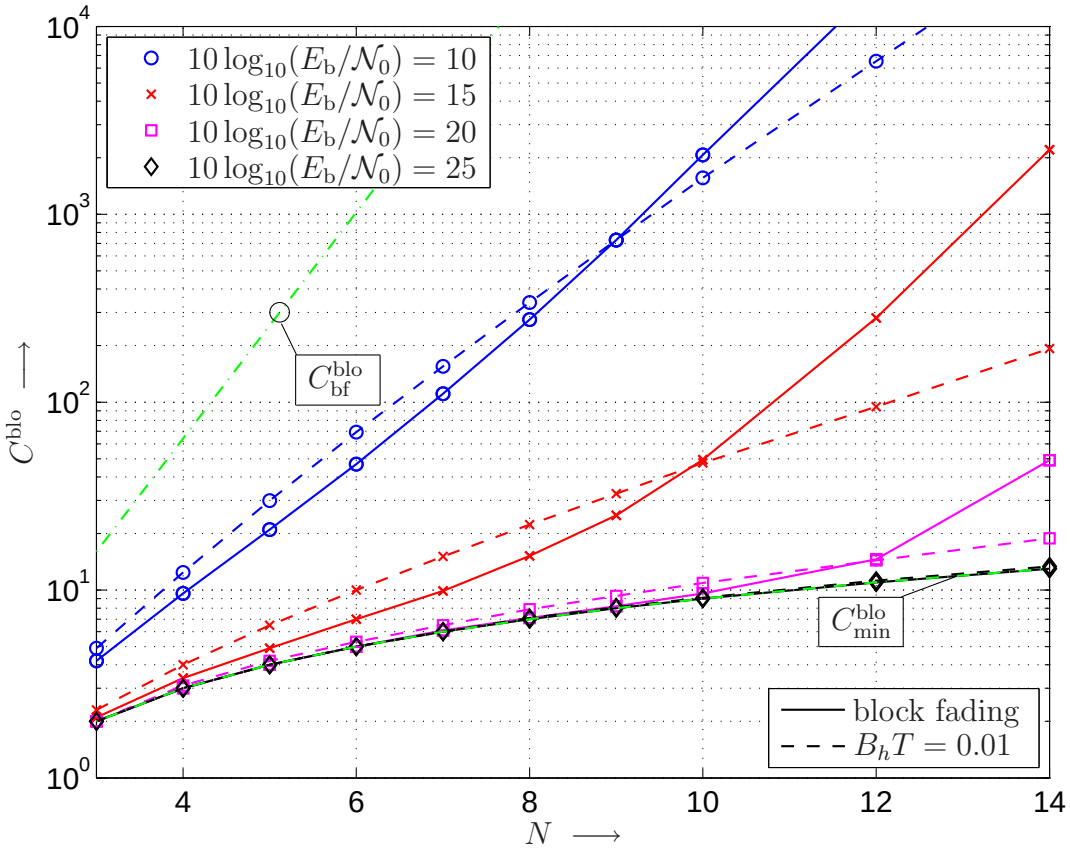

Figure 5.25: Exact complexity C^{blo} for QS fading with $B_hT = 0.01$ (dashed lines) vs. block fading (solid lines) over length of observation window N. For comparison: Complexity of brute–force ML MSDD ($C_{\mathrm{bf}}^{\mathrm{blo}} = L^{N-1}$) and of DFE–type decoder ($C_{\mathrm{min}}^{\mathrm{blo}} = N - 1$). Parameters: Dicyclic DSTM with $N_{\mathrm{T}} = 4$, $R = 0.5$, $N_{\mathrm{R}} = 1$, $10\log_{10}(E_{\mathrm{b}}/\mathcal{N}_0) = [10,\ 15,\ 20,\ 25]$, ρ such that $\Pr\big(d_1(\bar{\boldsymbol{S}}) \leq \rho \mid \bar{\boldsymbol{S}}\big) = 0.99$.

such that $\Pr\big(d_1(\bar{\boldsymbol{S}}) \leq \rho \mid \bar{\boldsymbol{S}}\big) = 0.99$. In particular, the different types of curves are (i) the exact complexity (Section 5.5.1.2, solid lines), (ii) the lower bound (Section 5.5.1.1, dotted lines), (iii) the approximation (5.209) for C_{as} (Section 5.5.1.3, dashed lines), (iv) $C_{\mathrm{min}}^{\mathrm{blo}}(N) = N - 1$ (dash–dotted line), and (v) $C_{\mathrm{bf}}^{\mathrm{blo}}(N) = L^{N-1}$ (dash–dotted line). It can be observed that the complexity bound derived in Section 5.5.1.1 is quite loose for $N \leq 30$. This is similar to the bound found in [JO05a, Section VI] for sphere decoding in MIMO detection with CSI. The exponential growth of complexity with increasing N is clearly visible from the exact results for C^{blo}. Also, there is good agreement between the asymptotic approximation and the exact complexity of FP–MSDSD for almost any value of N and SNR as long as the complexity exceeds $C_{\mathrm{min}}^{\mathrm{blo}}$. Consequently, $\max(C_{\mathrm{as}}, C_{\mathrm{min}}^{\mathrm{blo}})$ with C_{as} according to (5.209) represents a very easy to evaluate yet reasonably accurate approximation of the complexity of FP–MSDSD. Furthermore, one can observe the reduction of complexity in the exponential region by a factor

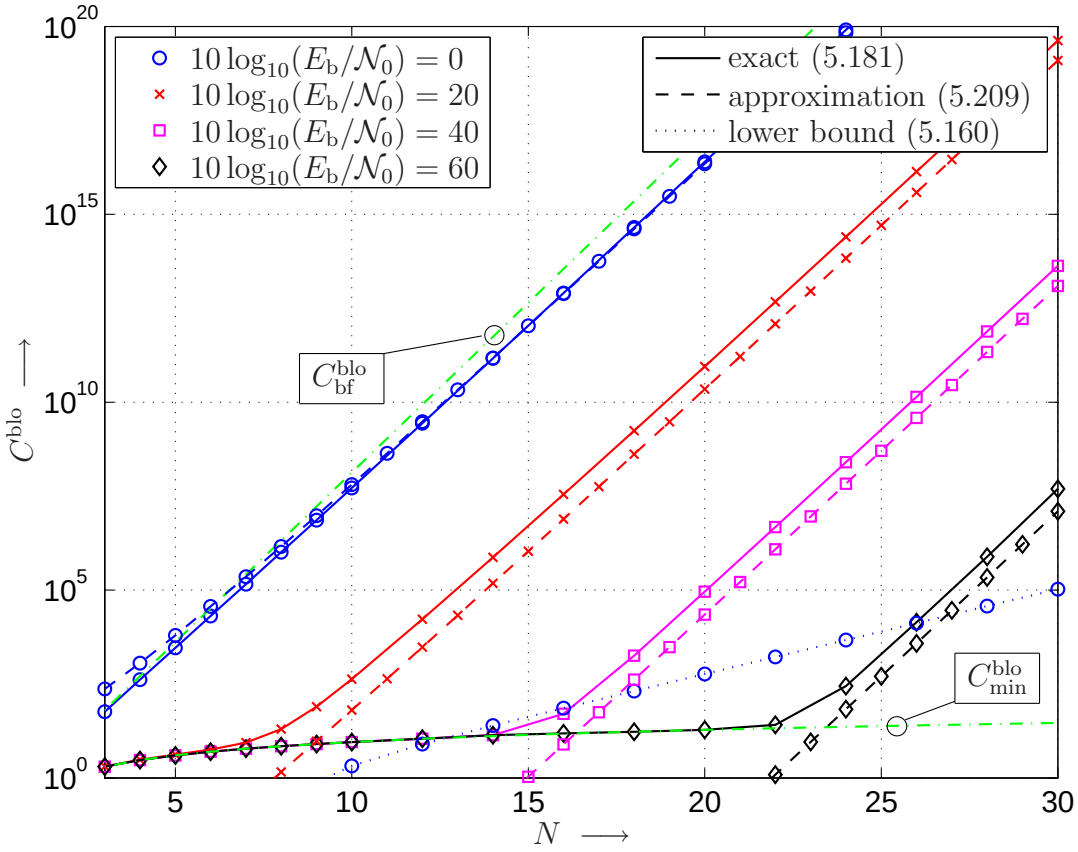

Figure 5.26: Complexity vs. observation window size N. Parameters: Cyclic DSTM with $N_{\text{T}} = 3$, $R = 1$, $N_{\text{R}} = 1$, static fading, $10\log_{10}(E_{\text{b}}/\mathcal{N}_0) = [0,\ 20,\ 40,\ 60]$, ρ such that $\Pr\big(d_1(\bar{\boldsymbol{S}}) \leq \rho \mid \bar{\boldsymbol{S}}\big) = 0.99$.

of 10^3 per 10 dB difference in SNR, as was to be expected from (5.209) for this example with $N_{\text{T}} = 3$ and $N_{\text{R}} = 1$. Fig. 5.26 also again illustrates the tremendous complexity savings with FP–MSDSD compared to brute–force ML MSDD ($C_{\text{bf}}^{\text{blo}}$).

C^{blo} vs. $N_{\text{T}} = N_{\text{R}}$ and SNR: Fig. 5.27 compares the complexity of FP–MSDSD for different rate–1 DSTM codes transmitted over a block–fading channel with equal numbers of transmit and receive antennas. It shows (on a logarithmic scale) the approximate complexity $\max(C_{\text{as}}, C_{\text{min}}^{\text{blo}})$ with C_{as} according to (5.209) over the SNR and the number of transmit and receive antennas $N_{\text{T}} = N_{\text{R}}$ for observation windows of lengths $N = 10,\ 20,\ 30$. Note that according to the power normalization of $\boldsymbol{S}[k]$ and the definition (5.7) of the SNR the total average received power is independent of N_{T} and scales linearly with N_{R}. Interestingly, it can be seen that an increase in the number of antennas at both ends of the channel not only increases reliability of transmission but may also be beneficial in terms of detection complexity. The minimum number of antennas required at both ends of the channel to push C^{blo} towards

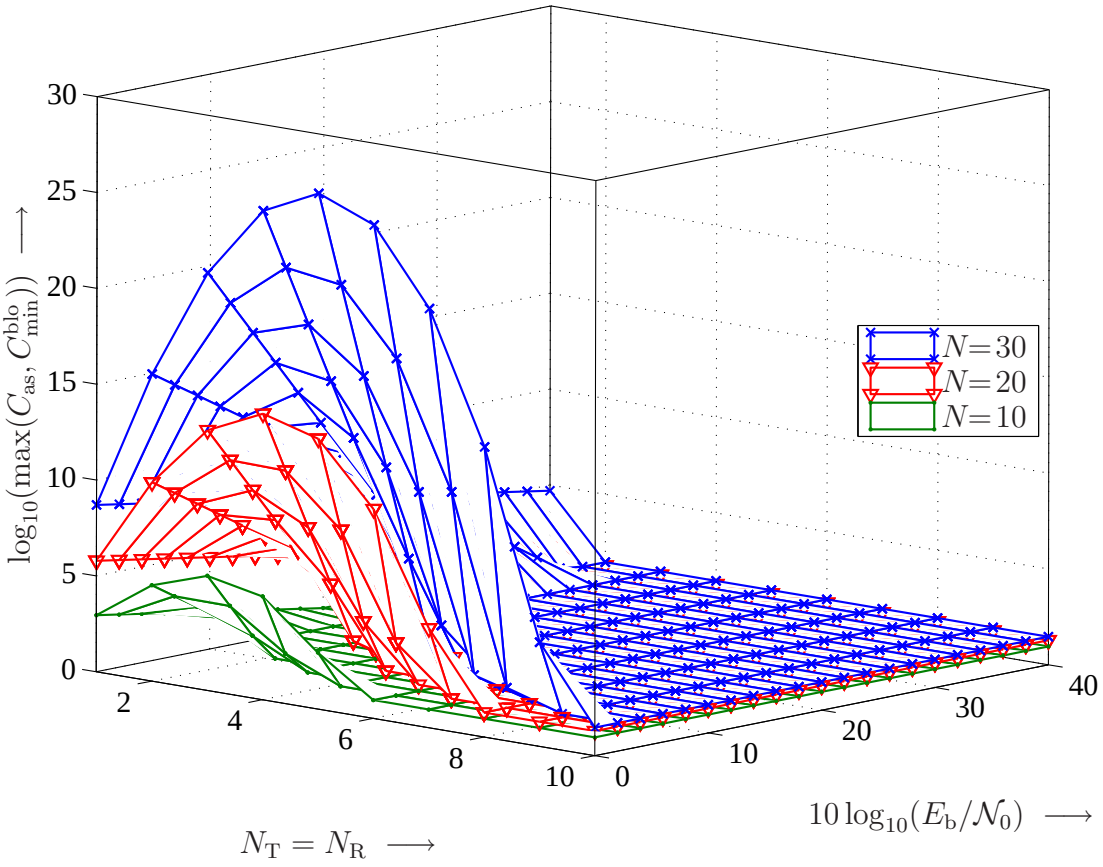

Figure 5.27: Complexity $\max(C_{\mathrm{as}}, C_{\mathrm{min}}^{\mathrm{blo}})$ with C_{as} according to (5.209) vs. SNR and $N_{\mathrm{T}} = N_{\mathrm{R}}$ for cyclic DSTM over block–fading channel for different values of N. Parameters: $R = 1$ and ρ such that $\Pr\big(d_1(\bar{\boldsymbol{S}}) \leq \rho \mid \bar{\boldsymbol{S}}\big) = 0.99$.

$C_{\mathrm{min}}^{\mathrm{blo}}$ strongly depends on N and the SNR and it grows towards larger N and lower SNR. This behavior can be understood based on the discussion at the end of Section 5.5.1.3, where it was pointed out that the exponential increase in the average complexity C^{blo} of FP–MSDSD is dominated by the worst–case complexity if *all* $N_{\mathrm{T}}N_{\mathrm{R}}$ coefficients of the MIMO channel happen to be small compared to the power of the AWGN. From basic stochastics, it is clear that the probability of this event decreases (exponentially) in $N_{\mathrm{T}}N_{\mathrm{R}}$, since the channel coefficients are uncorrelated. Therefore, the point, where the worst–case complexity becomes dominant, moves towards lower SNR, i.e. higher power of the AWGN, as N_{T} and / or N_{R} increase.

While FP–MSDSD is particularly interesting from an analytical point of view, it is in many cases not optimal in the sense of minimal decoder complexity. In the following, we will therefore direct our attention at the computationally more efficient MSDD algorithms developed in this work. For a fair comparison with the benchmark decoders CDD and DFDD, we subsequently consider the average number C^{sym} of examined (partial) candidates per decoded *symbol*, which

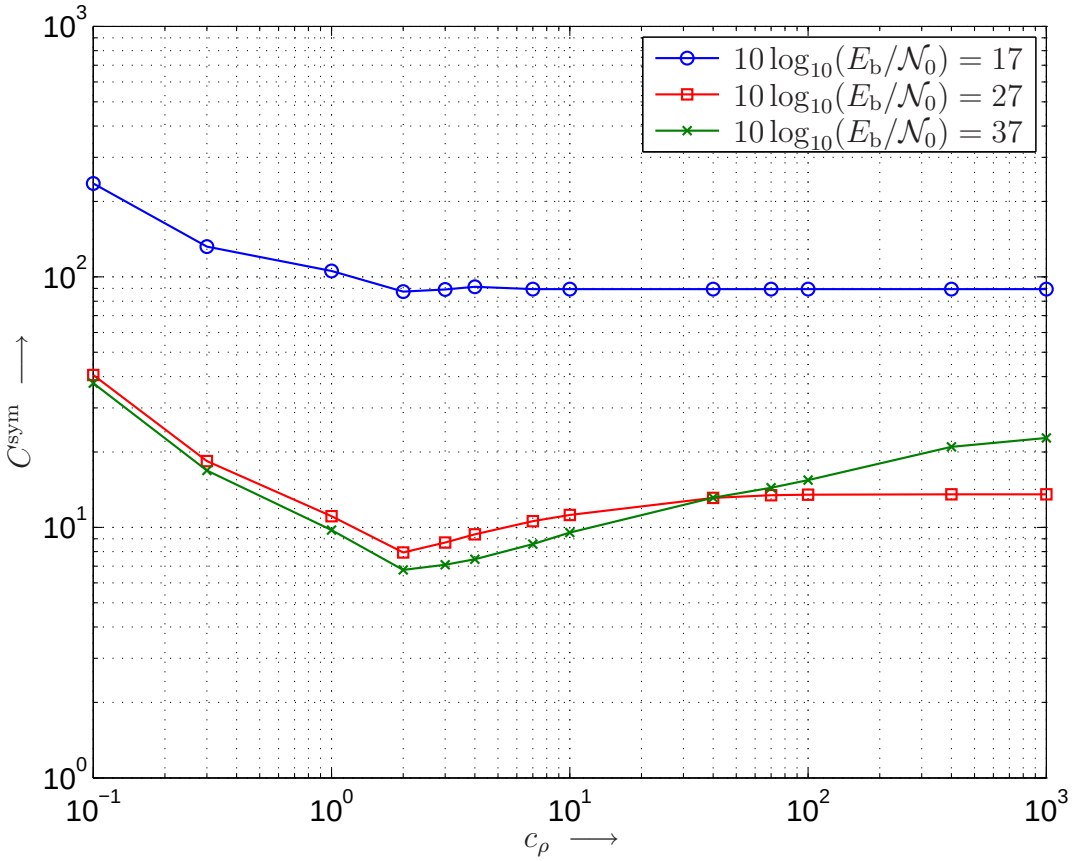

Figure 5.28: Complexity of MSDSD–LD with $N = 10$ vs. parameter c_ρ (initial radius $\rho_{\mathrm{init}} = NN_{\mathrm{T}}N_{\mathrm{R}}c_\rho$). Cyclic (B)DSTM with $N_{\mathrm{T}} = 3$, $N_{\mathrm{R}} = 1$, $R = 2$, and $B_{h,\mathrm{eff}}T = 0.09$.

for S–MSDD is given by C^{blo}/N' and for DF–MSDD by $C^{\mathrm{blo}}/(\kappa_{\mathrm{U}}^{\mathrm{DF}} - \kappa_{\mathrm{L}}^{\mathrm{DF}} + 1)$.

Initial Radius: To study the effect of the initial radius $\rho_{\mathrm{init}} = NN_{\mathrm{T}}N_{\mathrm{R}}c_\rho$ on the complexity of regular MSDSD, i.e. MSDSD based on the A–SpD (cf. Section 3.1.3.1), we exemplarily consider MSDSD–LD (the results are very similar for other MSDSD algorithms) for cyclic (B)DSTM with $N_{\mathrm{T}} = 3$, $R = 2$, $N_{\mathrm{R}} = 1$ and $B_{h,\mathrm{eff}}T = 0.09$. Fig. 5.28 depicts the complexity C^{sym} for different SNRs as a function of the constant c_ρ introduced in (3.39). While complexity increases if the initial radius is chosen too small, substantial complexity savings of up to 70 % compared to $\rho_{\mathrm{init}} \to \infty$ can be obtained by choosing an optimal finite start radius. It is interesting to note that the optimal choice of the parameter $c_\rho \approx 2$ is roughly independent of the observation window length N (not shown in the figure), whereas the probability $\Pr\big(d_1(\bar{\boldsymbol{S}}_1) \leq c_\rho NN_{\mathrm{T}}N_{\mathrm{R}} \mid \bar{\boldsymbol{S}}\big)$ of the true $\bar{\boldsymbol{S}}$ lying inside the sphere of radius ρ_{init} increases with N.[15] It can also be observed that, for this relatively fast fading example, the complexity of

[15]This follows from evaluation of $\Pr\big(d_1(\bar{\boldsymbol{S}}_1) \leq c_\rho NN_{\mathrm{T}}N_{\mathrm{R}} \mid \bar{\boldsymbol{S}}\big)$, cf. (3.116).

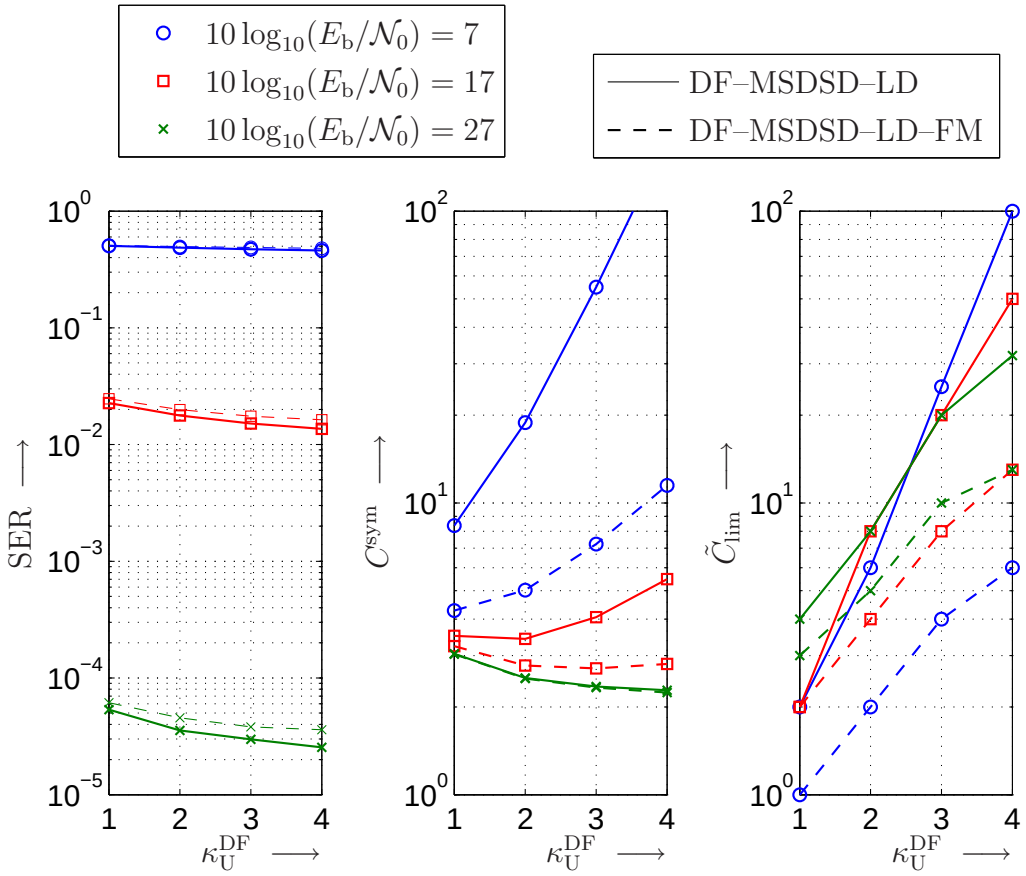

Figure 5.29: Results for DF–MSDSD–LD with $N = 10$, $c_\rho = 2$, $\kappa_{\mathrm{L}}^{\mathrm{DF}} = 1$ and different values of $\kappa_{\mathrm{U}}^{\mathrm{DF}}$. Left: SER, center: average complexity C^{sym}, right: minimal allowed limiting complexity $\tilde{C}_{\mathrm{lim}}$ (see text). Cyclic (B)DSTM with $N_{\mathrm{T}} = 3$, $N_{\mathrm{R}} = 1$, $R = 2$, and $B_{h,\mathrm{eff}}T = 0.09$. With ML–MSDD metric (solid) and with Fano–type metric (dashed).

MSDSD with infinite initial radius may increase for larger SNR. This is due to the fact that the radius after the first tentative decision $\hat{\tilde{\boldsymbol{S}}}$ may be much larger than the metric for the ML decision for high SNR. A finite start radius, on the other hand, leads to the more intuitive result that the decoding complexity decreases monotonically with the SNR. All of the following results assume the optimal finite start radius with

$$c_\rho = 2. \tag{5.220}$$

Results for DF–MSDSD: Before we compare the different MSDD algorithms, we take a look at appropriate choices for the parameters $\kappa_{\mathrm{U}}^{\mathrm{DF}}$ and $\kappa_{\mathrm{L}}^{\mathrm{DF}}$ of DF–MSDSD (cf. Section 2.4.4). The evaluation of the effective SNR and the SER in the individual positions of the MSDD observation window (cf. Figs. 5.2 and 5.10, respectively), revealed that the reliability of the decisions is roughly independent of the position in the observation window, with the exception

of the edge positions in fast–fading scenarios. Since the complexity (normalized by the depth of the tree) of a tree–search decoder typically is an increasing function of the tree depth, it is therefore reasonable to choose $\kappa_{\mathrm{L}}^{\mathrm{DF}} = 1$, i.e. regardless of the choice of $\kappa_{\mathrm{U}}^{\mathrm{DF}}$, such that only the decision on the last symbol in the observation window is discarded.

In Fig. 5.29 we therefore plotted the results as a function of $\kappa_{\mathrm{U}}^{\mathrm{DF}}$, which for $\kappa_{\mathrm{L}}^{\mathrm{DF}} = 1$ is equal to the number of decisions returned per decoded block. As exemplary system parameters, we consider cyclic (B)DSTM with $N_{\mathrm{T}} = 3$, $R = 2$, DF–MSDSD with $N = 10$, $c_\rho = 2$, with regular ML–MSDD metric (solid lines) and with Fano–type metric (dashed lines), $B_{h,\mathrm{eff}}T = 0.09$, various values of $E_{\mathrm{b}}/\mathcal{N}_0$ and $N_{\mathrm{R}} = 1$. In particular, we consider (i) the SER (left), (ii) the average number C^{sym} of examined (partial) candidates per decoded symbol (center), and (iii) a quantity $\tilde{C}_{\mathrm{lim}}$ defined as the minimal number of examined (partial) candidates per decoded symbol, after which the tree–search can be terminated without noteworthy impact on the SER.[16]

The SER results in the left subplot show that the performance improves slightly if $\kappa_{\mathrm{U}}^{\mathrm{DF}}$ is increased, which is (i) due to the fact that the reliability of the decisions improves (slightly) from the second–to–last position towards the center of the observation window (cf. Fig. 5.10), and (ii) because the effect of error propagation is reduced if fewer previous decisions are fed back into the DF–MSDD metric.

The results on C^{sym} show that for low SNR, the average complexity of DF–MSDSD shows the behavior typical for tree–search algorithms, i.e. the complexity is an increasing function of the tree depth $\kappa_{\mathrm{U}}^{\mathrm{DF}} + 1$. In high SNR (cf. lines for $10\log_{10}(E_{\mathrm{b}}/\mathcal{N}_0) = 27\,\mathrm{dB}$), on the other hand, the average complexity of DF–MSDSD even decreases with growing $\kappa_{\mathrm{U}}^{\mathrm{DF}}$. This is because in the considered range of parameters the average complexity per decoder use is close to its minimum $2\kappa_{\mathrm{U}}^{\mathrm{DF}} + 1$. As $\kappa_{\mathrm{U}}^{\mathrm{DF}}$ decisions are returned per decoder use, the minimal complexity per decoded symbol of (unlimited) DF–MSDSD is given by $2 + 1/\kappa_{\mathrm{U}}^{\mathrm{DF}}$.

Finally, the results on $\tilde{C}_{\mathrm{lim}}$ clearly show that small values of $\kappa_{\mathrm{U}}^{\mathrm{DF}}$ are preferable. Interestingly, while the average complexity C^{sym} grows towards lower SNR, we can see that the tree–search can be terminated after only 2 and even after only 1 examined candidate(s) per decoded symbol for DF–MSDSD operating based on the ML–MSDD or the Fano–type metric, respectively, without noteworthy impact on the power efficiency. I.e. while DF–MSDSD with unlimited search would require 8.4 (ML metric) and 4.3 (Fano–type metric) metric computations per decoded symbol at $10\log_{10}(E_{\mathrm{b}}/\mathcal{N}_0) = 7\,\mathrm{dB}$ an implementation where the average complexity is equal to the maximal complexity of 2 (ML metric) and 1 (Fano–type metric) would achieve practically the same power efficiency. Towards higher SNR, $\tilde{C}_{\mathrm{lim}}$ grows rapidly with $\kappa_{\mathrm{U}}^{\mathrm{DF}}$. However, for $\kappa_{\mathrm{U}}^{\mathrm{DF}} = 1$ it is still practically equal to the average complexity. Hence, for practical implementation, where

[16] More specifically, $\tilde{C}_{\mathrm{lim}}$ is determined such that the SER is increased by no more than a factor of 1.1 compared to the unlimited search.

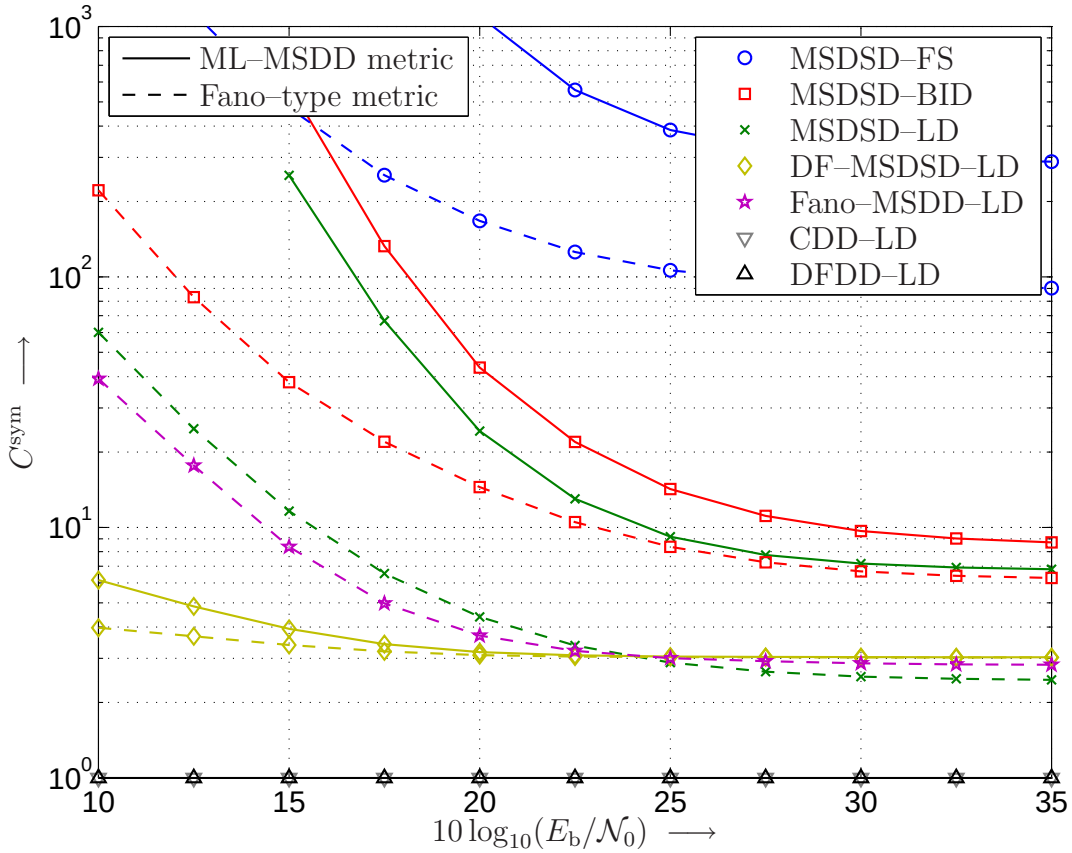

Figure 5.30: Complexity vs. $10\log_{10}(E_{\mathrm{b}}/\mathcal{N}_0)$ for various detection algorithms with $N = 10$ and optimal $c_\rho = 2$. Cyclic (B)DSTM with $N_{\mathrm{T}} = 3$, $N_{\mathrm{R}} = 1$, $R = 2$, and $B_{h,\mathrm{eff}}T = 0.09$. Solid lines: without Fano–type metric, Dashed lines: with Fano–type metric.

maximal complexity plays a very important role, $\kappa_{\mathrm{U}}^{\mathrm{DF}} = 1$ is the appropriate choice. In the following, we consider

$$\kappa_{\mathrm{U}}^{\mathrm{DF}} = \kappa_{\mathrm{L}}^{\mathrm{DF}} = 1 \tag{5.221}$$

in our comparison of the various detection schemes, i.e. the tree searched in DF–MSDSD is of depth two.

Average Complexity: Fig. 5.30 compares the different possibilities to reduce the complexity of MSDSD, i.e. deployment of ML–MSDD (solid lines) or Fano–type metric (dashed lines) and different implementations for outer and inner decoders. As example, we again consider cyclic (B)DSTM with $N_{\mathrm{T}} = 3$, $N_{\mathrm{R}} = 1$, $R = 2$, $B_{h,\mathrm{eff}}T = 0.09$, and $N = 10$. For comparison, we also included the complexities of CDD–LD and DFDD–LD, which are constant one. The corresponding performance results can be found in Fig. 5.17. It can be seen that the complexity of all implementations decreases rapidly with increasing SNR, since the

search quickly terminates for small enough noise. The variants MSDSD–BID and especially MSDSD–LD with improved symbol–search strategies operate at significantly reduced average complexity compared to MSDSD–FS. Furthermore, applying the Fano–type metric speeds up the search especially in the lower SNR range, as was anticipated in Section 3.1.2.3. Combining Fano–type metric with inner LD, i.e. MSDSD–LD–FM, and Fano–MSDD–LD yield tremendous complexity reductions by approximately two decimal powers compared to MSDSD–FS. This gain comes at the expense of a loss in power efficiency of less than 1 dB (cf. Fig. 5.17). However, due to the large size of the signal constellation ($L = 64$) and the depth of the search tree, the average complexity of MSDSD–LD–FM and Fano–MSDD–LD is still quite high in low–SNR regimes. DF–MSDSD–LD on the other hand, when based on the ML–MSDD and on the Fano–type metric on average requires no more than approximately 4 and 6 metric computations per decoded symbol, respectively, even in very low SNR, where the SER already lies well above 10^{-1}. The reason for this lies in the fact that the depth of the search tree is only two. At higher SNR, the average complexity of DF–MSDSD is slightly increased compared to the other MSDD implementations included in this comparison. This is because —contrary to the other implementations— for DF–MSDSD the depth of the search tree is larger (by one) than the step size, in which the observation window slides forward from decoder use to decoder use.

Taking into account that the performance of DF–MSDD is even better than that of regular ML–MSDD (cf. Fig. 5.17) DF–MSDSD(–LD–FM) clearly has the most appealing *average*–complexity–to–performance tradeoff of all algorithms considered in this work.

Performance with Limited Maximal Complexity: For practical implementation the *maximal* decoder complexity is of great importance. Therefore, in Figs. 5.31–5.33 we consider the performance that can be achieved by the different MSDD implementations if the maximal number of examined (partial) candidates per decoded symbol is limited to a finite C_{lim}, i.e. if the tree–search process in the outer decoder is terminated after $(N-1)C_{\mathrm{lim}}$ (MSDSD, Fano–MSDD) and $(\kappa_{\mathrm{U}}^{\mathrm{DF}} - \kappa_{\mathrm{L}}^{\mathrm{DF}} + 1)C_{\mathrm{lim}} = C_{\mathrm{lim}}$ (DF–MSDSD) examined candidates. In all cases, we again considered the example of cyclic (B)DSTM with $N_{\mathrm{T}} = 3$, $R = 2$, $N_{\mathrm{R}} = 1$, $N = 10$ and $B_{h,\mathrm{eff}}T \in \{0.003, 0.09\}$. The short vertical lines with markers indicate the average complexity C^{sym} of the detection schemes.

Fig. 5.31 shows results for MSDSD with FS, LD and BID and for DF–MSDSD with LD inner decoding, when the ML–MSDD metric and $c_\rho = \infty$, i.e. $\rho_{\mathrm{init}} \to \infty$, are used. One can observe that for relatively slow fading (cf. left subplot with $B_{h,\mathrm{eff}}T = 0.003$) all algorithms find very good first estimates and only minor gains are achieved if the decoder is granted more decoding complexity. For MSDSD–FS and MSDSD–BID a very steep cliff occurs at $C_{\mathrm{lim}} = L = 64$, because with $\rho_{\mathrm{init}} \to \infty$ MSDSD–BID like MSDSD–FS examines all candidate branches emanating from a considered node before the radius is updated for the first time. In faster fading

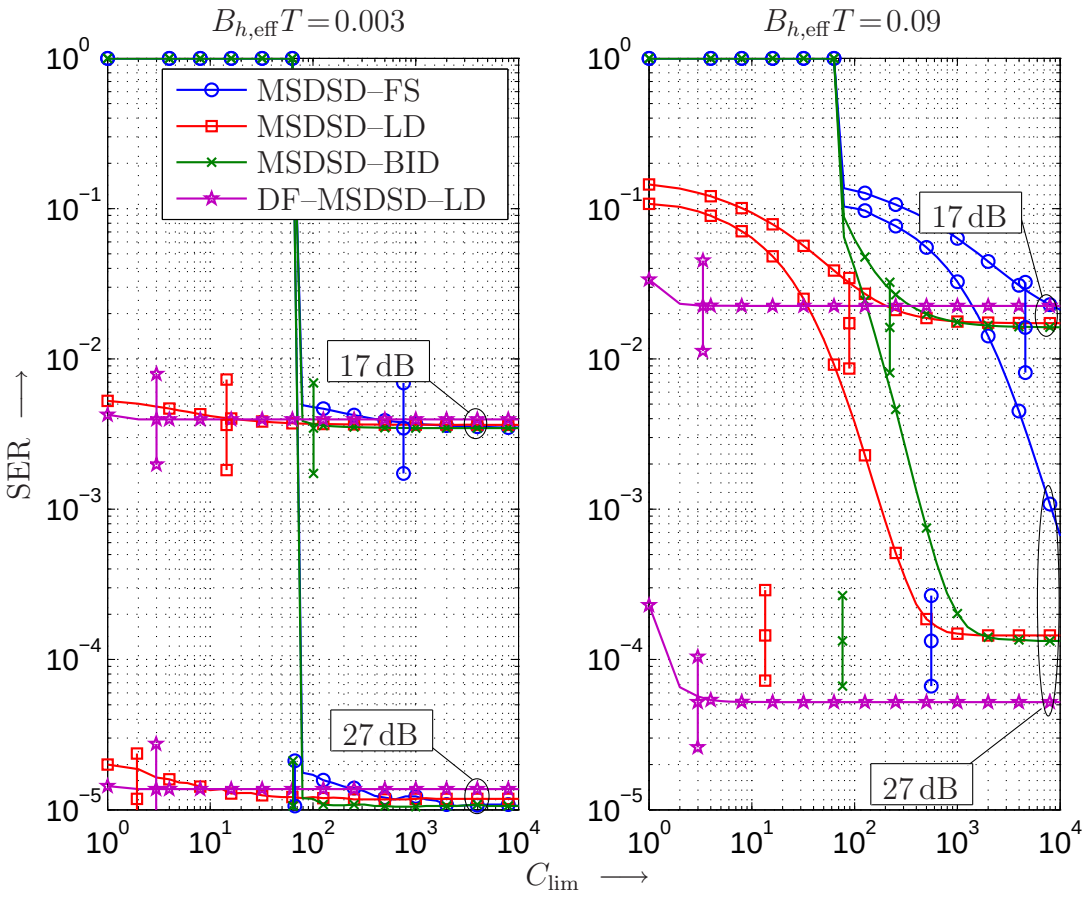

Figure 5.31: Performance of MSDSD with various symbol–search strategies vs. maximal allowed complexity per decoded symbol for different SNRs $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ and fading bandwidths $B_{h,\mathrm{eff}}T$. Parameters: Cyclic (B)DSTM with $N_\mathrm{T} = 3$, $R = 2$, $N_\mathrm{R} = 1$, $N = 10$, $c_\rho \to \infty$, ML–MSDD metric. Vertical lines: average complexity C^{sym}.

(cf. right subplot with $B_{h,\mathrm{eff}}T = 0.09$) the first estimate obtained by the (regular) MSDSD is far from optimal. Here, both MSDSD–LD and MSDSD–BID do not suffer a noteworthy performance degradation if the maximal complexity is limited to $C_{\mathrm{lim}} = 10^3 \ldots 10^4$, with higher values for higher SNR. For MSDSD–FS the behavior is very similar to that of MSDSD–LD when C_{lim} is multiplied by a factor of 40. For all three symbol–search algorithms the required maximal complexity is higher by orders of magnitude than the average complexity C^{sym}, which is considered very inconvenient for practical implementation. DF–MSDSD–LD on the other hand in all cases converges very quickly to its final solution. After only one examined candidate per decoded symbol, DF–MSDSD–LD achieves practically the same power efficiency as regular MSDSD. Further improvements are also realized quite quickly, so that the complexity can be limited to values $C_{\mathrm{lim}} \approx C^{\mathrm{sym}}$, which also lies well below the average complexity of the other MSDSD algorithms considered here.

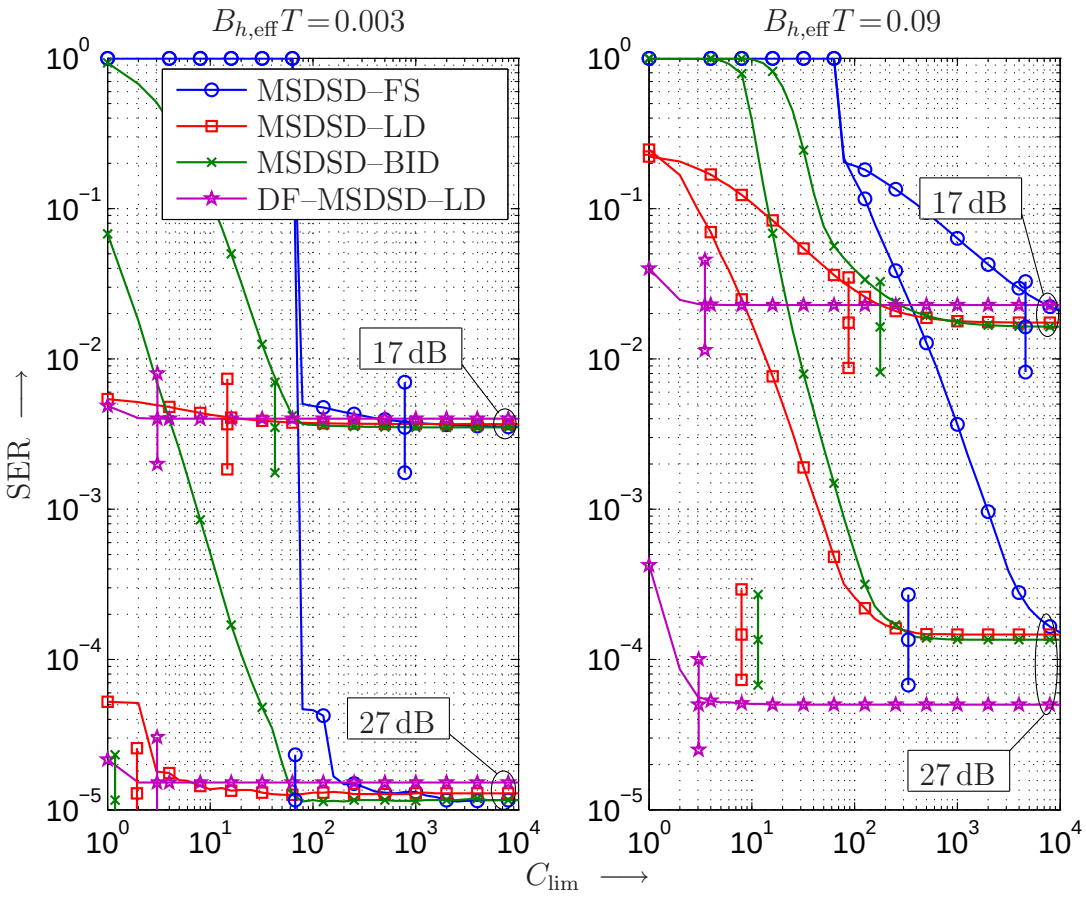

Figure 5.32: Performance of MSDSD with various symbol–search strategies vs. maximal allowed complexity per decoded symbol for different SNRs $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ and fading bandwidths $B_{h,\mathrm{eff}}T$. Parameters: Cyclic (B)DSTM with $N_\mathrm{T} = 3$, $R = 2$, $N_\mathrm{R} = 1$, $N = 10$, $c_\rho = 2$, ML–MSDD metric. Vertical lines: average complexity C^{sym}.

Fig. 5.32 shows results for the same transmission scenario, but with $c_\rho = 2$, i.e. a finite initial sphere radius $\rho := \rho_{\mathrm{init}} = 2NN_\mathrm{T}N_\mathrm{R}$. Here, the convergence of the tree–search in regular MSDSD is accelerated significantly in rapid fading (right subplot). Now limitations of maximal complexity to $C_{\mathrm{lim}} = 300 \ldots 600$ are feasible for MSDSD–LD and MSDSD–BID. In the case of DF–MSDSD–LD the performance–complexity tradeoff for $c_\rho \to \infty$ was already so good that the use of a finite threshold actually slightly increases the SER when limiting the complexity per symbol to very small values. The same effect can be observed for MSDSD–FS and MSDSD–LD in the case of very slow fading (left subplot).

If, in addition the Fano–type metric is used, the speed of convergence of all algorithms is increased further, as can be seen in Fig. 5.33. However, this improvement is only minor, showing that the importance of the Fano–type metric lies mainly in the reduction of the average complexity when the maximal complexity is unlimited. For comparison, we also included Fano–

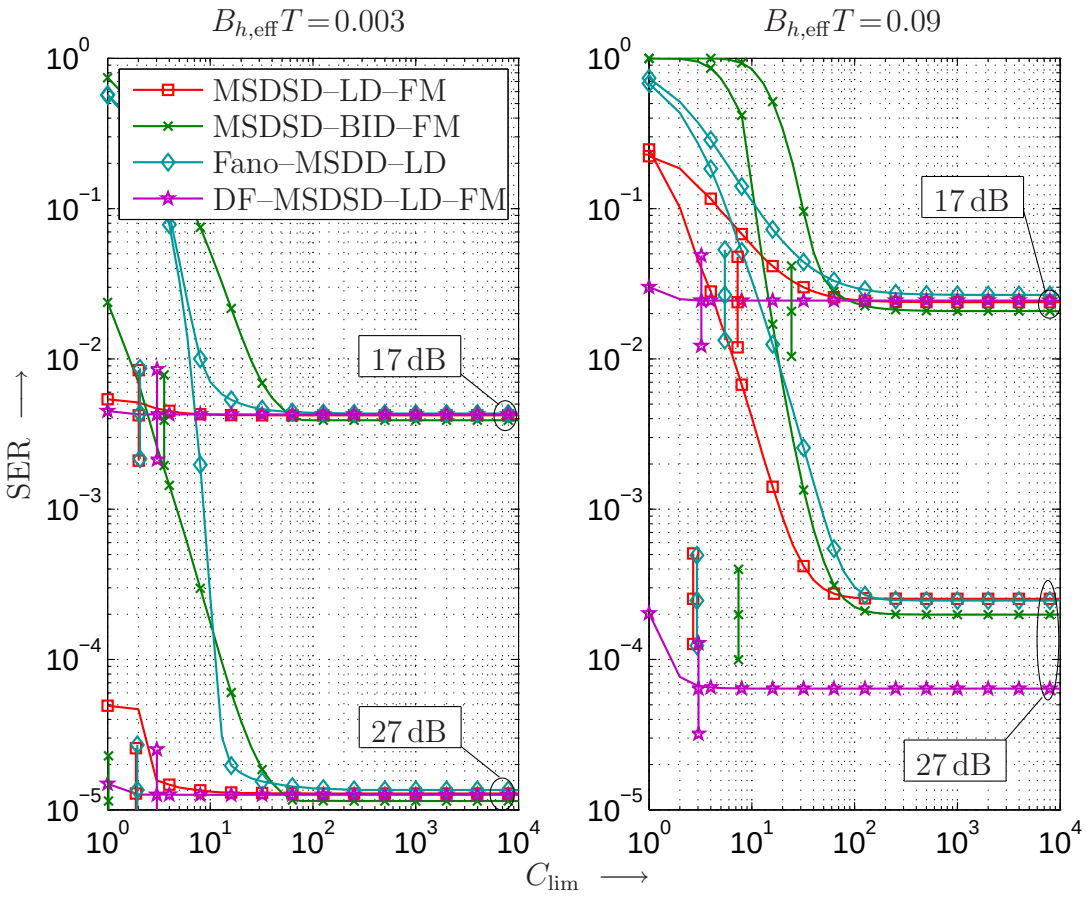

Figure 5.33: Performance of various MSDD implementations vs. maximal allowed complexity per decoded symbol for different SNRs $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ and fading bandwidths $B_{h,\mathrm{eff}}T$. Parameters: Cyclic (B)DSTM with $N_\mathrm{T} = 3$, $R = 2$, $N_\mathrm{R} = 1$, $N = 10$, $c_\rho = 2$, Fano–type metric. Vertical lines: average complexity C^{sym}.

MSDD–LD. The SER when the maximal complexity is limited consistently lies above that of MSDSD–LD–FM. Especially in slow–fading scenarios (left subplot) Fano–MSDD–LD requires a significantly larger value of C_{lim} than MSDSD–LD–FM, although the average complexities are practically identical. Interestingly, the SERs of MSDSD–LD–FM, MSDSD–BID–FM and Fano–MSDD–LD all converge around $C_{\mathrm{lim}} \approx 100$. While the average complexities of these algorithms are now of the same order of magnitude as that of DF–MSDSD–LD, the latter is still superior by far with respect to required C_{lim}.

Having clearly identified DF–MSDSD–LD as most promising algorithm for power–efficient tree–search–based noncoherent detection we want to investigate the influence of the length of the observation window on the performance–complexity tradeoff. To this end Fig. 5.34 shows the SER for DF–MSDSD–LD–FM with $c_\rho = 2$ and $N \in \{6, 10\}$ as a function of the maximally allowed complexity C_{lim} per decoded symbol for the example of cyclic (B)DSTM with $N_\mathrm{T} = 3$,

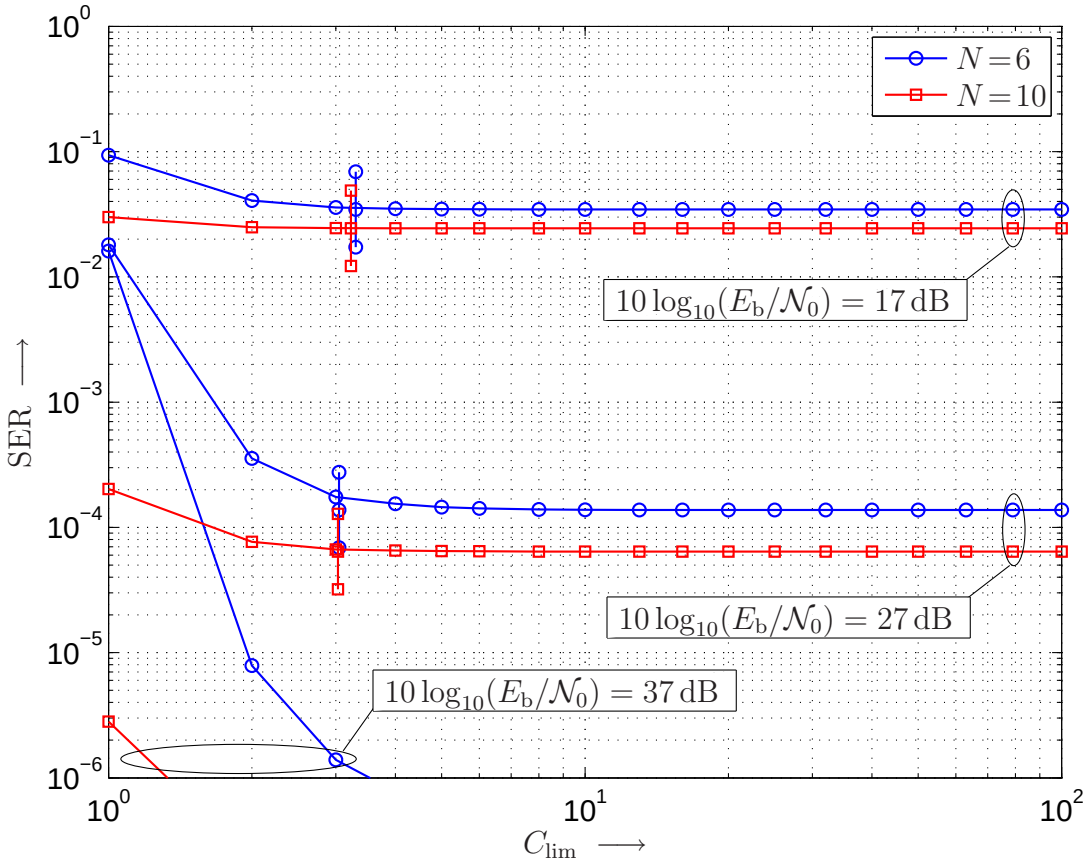

Figure 5.34: Performance of DF–MSDSD–LD–FM vs. maximal allowed complexity per decoded symbol for different SNRs $10\log_{10}(E_b/\mathcal{N}_0)$ and N. Parameters: Cyclic (B)DSTM with $N_T = 3$, $R = 2$, $N_R = 1$, $B_{h,\mathrm{eff}}T = 0.09$, $c_\rho = 2$. Vertical lines: average complexity C^{sym}.

$R = 2$, $B_{h,\mathrm{eff}}T = 0.09$, $N_R = 1$.

One can observe that a larger N not only improves the power efficiency of the unlimited tree search, but also leads to better performance for any value of C_{lim}. The reason for this is that the reference symbols $\boldsymbol{Y}_{\mathrm{DF},n}$, $N-1 \leq n \leq N$, [cf (3.61)] computed from the $N-3$ feedback symbols facilitate the inherent channel estimation on the remaining two positions of the observation window. Despite common perception regarding MSDD we here observe the effect that increasing the length N of the observation window not only leads to better performance, but also to lower complexity, as both C^{sym} and the required C_{lim} are (slightly) lower.

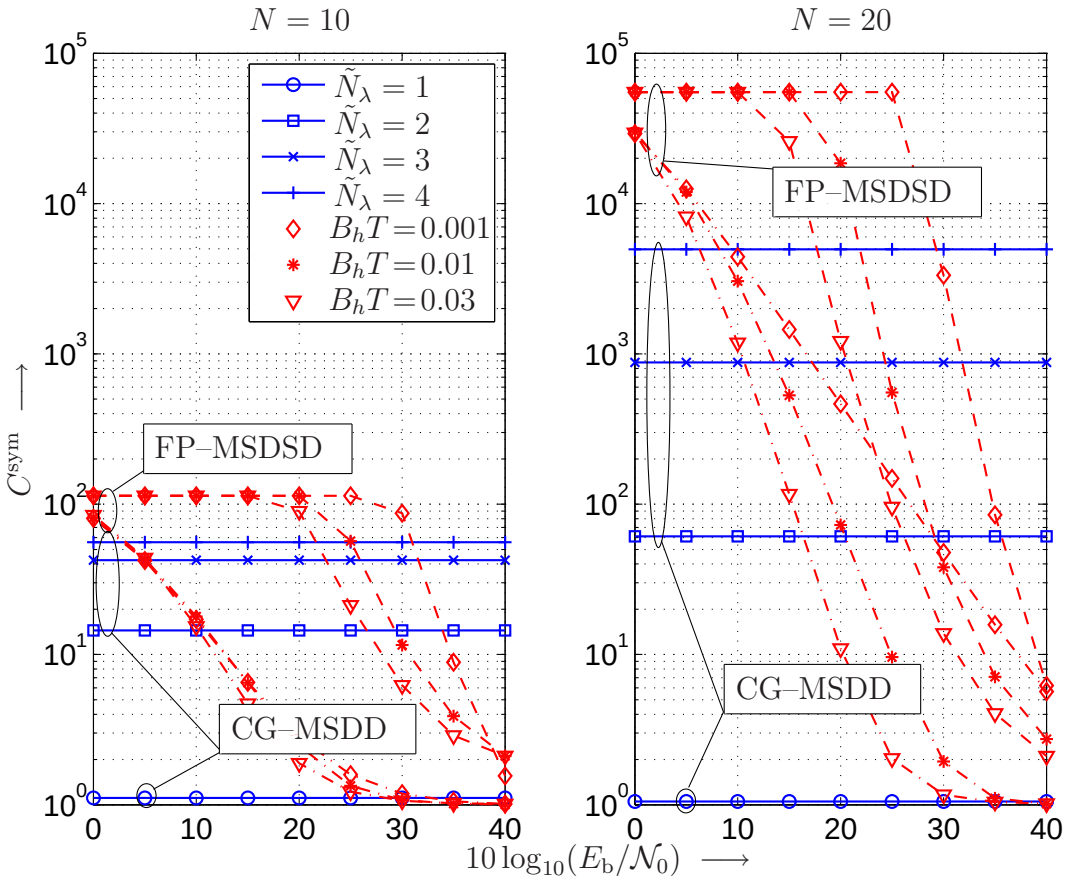

Figure 5.35: Complexity of CG–MSDD vs. SNR $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ for $N = 10$ (left) and $N = 20$ (right), different values of $\tilde{N}_\lambda$ and DBPSK. For comparison: average (dash–dotted) and 0.1%–complexity (dashed) of FP–MSDSD for Clarke's fading model and different values of B_hT.

5.5.3.2 CG–MSDD

Having illustrated the complexity of the various tree–search–based MSDD implementations in the previous section, we now direct our attention at MSDD based on combinatorial geometry, so–called CG–MSDD and compare its performance–complexity tradeoff to that of tree–search–based MSDD.

Complexity vs. SNR: Figures 5.35 and 5.36 compare the computational complexity of CG–MSDD according to (5.211) with those of FP–MSDSD and MSDSD (based on the A–SpD), respectively, as function of SNR for DBPSK, $N \in \{10, 20\}$. In particular, the average complexity C^{sym} and the complexity that is exceeded by MSDSD in 0.1% of all blocks are plotted for FP–MSDSD (Fig. 5.35) and MSDSD (Fig. 5.36). We observe that, if average complexity of MSDSD is considered, CG–MSDD is superior only for relatively low SNR and slow fading, where small values of $\tilde{N}_\lambda$ are sufficient to achieve low SERs (see Figs. 5.19 and 5.20). However,

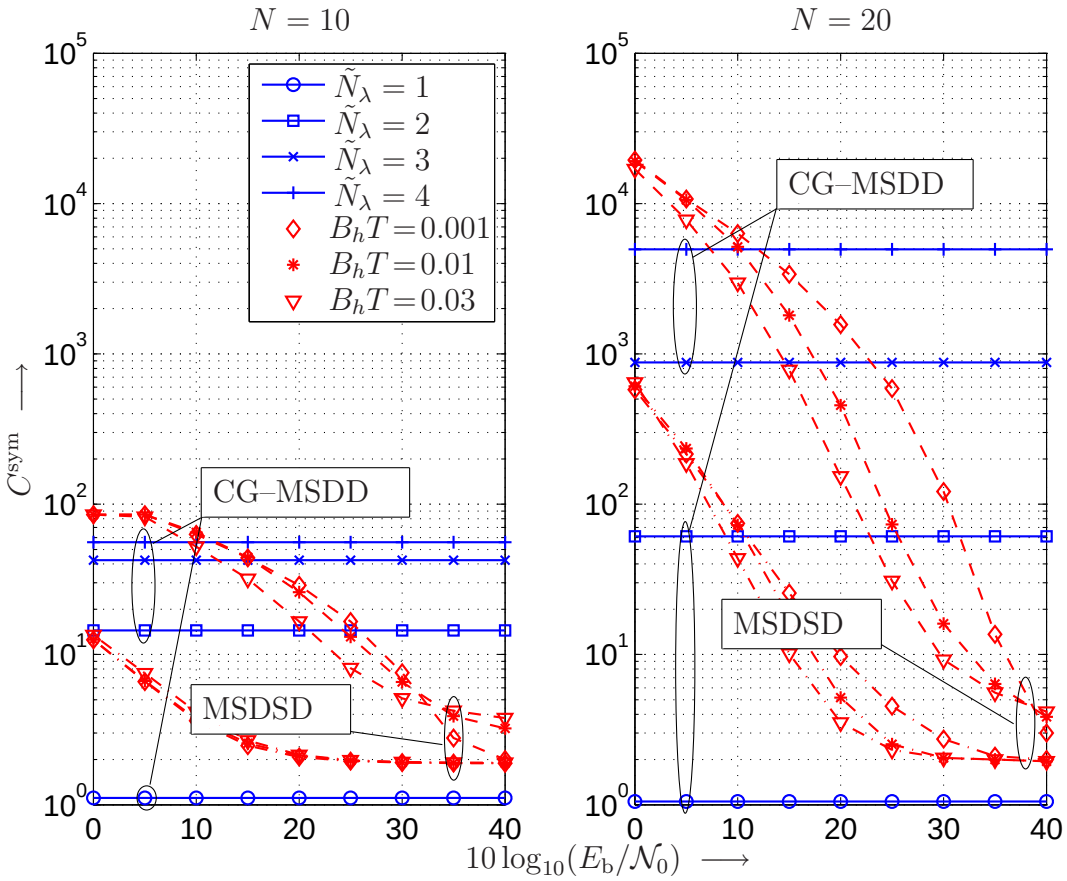

Figure 5.36: Complexity of CG–MSDD vs. SNR $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ for $N = 10$ (left) and $N = 20$ (right), different values of $\tilde{N}_\lambda$ and DBPSK. For comparison: average (dash–dotted) and 0.1%–complexity (dashed) of MSDSD for Clarke's fading model and different values of B_hT.

when the comparison is made with respect to the 0.1%–complexity of MSDSD, CG–MSDD turns out to be a far more attractive alternative, as e.g. the 0.1%–complexity of FP–MSDSD attains its maximum of $\sum_{n=1}^{N-1} L^n$ for SNRs up to $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0) = 10\,\mathrm{dB}$, $15\,\mathrm{dB}$, and $25\,\mathrm{dB}$ for $B_hT = 0.03$, 0.01, and 0.001, respectively. Such a comparison is particularly relevant for practical implementations where the variable complexity and thus processing delay inherent to sphere decoding is a problem.

Simulations for DF–MSDSD (not shown in the figure) revealed that in this transmission scenario each run of DF–MSDSD with $\kappa_\mathrm{U}^{\mathrm{DF}} = \kappa_\mathrm{L}^{\mathrm{DF}} = 1$ and $N \in \{10, 20\}$ can be terminated without any performance penalty after only two examined candidates per decoded symbol, i.e. both C^{sym} and the 0.1%–complexity do not exceed two in this case. This clearly illustrates the superiority of DF–MSDSD in cases where $\tilde{N}_\lambda > 1$ is required for CG–MSDD.

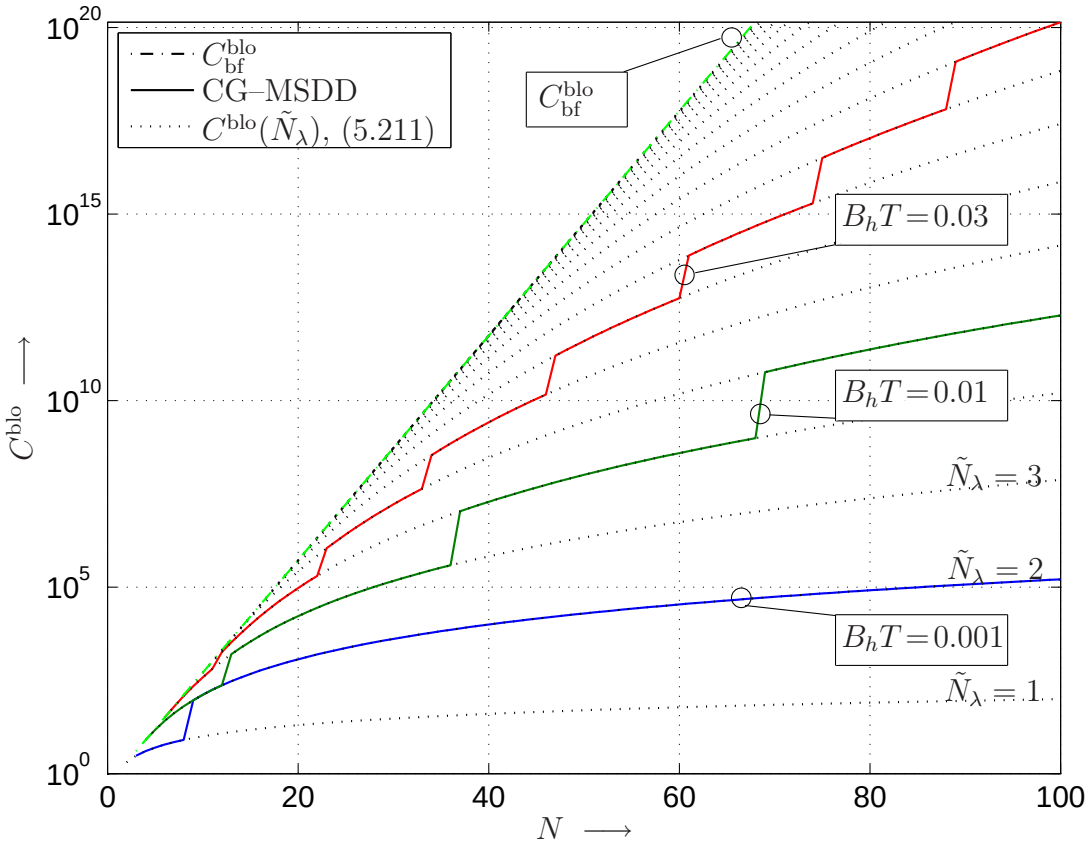

Figure 5.37: Asymptotic complexity of CG–MSDD for DBPSK, Clarke's fading model and different values of B_hT with $\tilde{N}_\lambda$ such that PEP of dominant error event lies below 10^{-4} (solid lines). For comparison: complexity of brute–force MSDD $C_{\text{bf}}^{\text{blo}} = L^{N-1}$ (dash–dotted), and $C(\tilde{N}_\lambda)$ according to (5.211) for fixed $\tilde{N}_\lambda = 1, 2, 3, \ldots$ (dotted).

Asymptotic Complexity vs. N: Finally, we consider the asymptotic complexity of CG–MSDD as N grows large. Figure 5.37 shows the average complexity as function of N for CG–MSDD with rank mismatch, where $\tilde{N}_\lambda$ is chosen such that the error floor of the PEP for the dominant error event lies below 10^{-4}. Also plotted is the complexity $C_{\text{bf}}^{\text{blo}}$ for brute–force MSDD. It can be seen that the required value of $\tilde{N}_\lambda$ —depending on the fading bandwidth B_hT— increases with N, which means that complexity is exponential in N. On the other hand, this exponential growth is much smaller than that for brute–force MSDD or FP–MSDSD in the block–fading channel. Thus, CG–MSDD has an asymptotic advantage over FP–MSDSD and it can be expected that the same holds when compared to MSDSD.

5.A Derivations and Proofs

5.A.1 Derivation of Eq. (5.10)

In this appendix, we show that $\psi_{hh}[\kappa, i_1, i_2, j_1, j_2] = \psi^{\mathrm{Tx}}[i_1, i_2]\psi^{\mathrm{Rx}}[j_1, j_2]\psi^{\mathrm{t}}[\kappa]$. Using (5.8) and introducing for notational simplicity $\boldsymbol{A} \triangleq \sqrt{\boldsymbol{\Psi}^{\mathrm{Tx}}}$ and $\boldsymbol{B} \triangleq \sqrt{\boldsymbol{\Psi}^{\mathrm{Rx}^*}}$ we can write:

$$\begin{aligned}
\psi_{hh}[\kappa, i_1, i_2, j_1, j_2] &\triangleq \mathcal{E}\big\{h_{i_1,j_2}[k+\kappa]h^*_{i_2,j_2}[k]\big\} && (5.222)\\
&= \mathcal{E}\big\{\boldsymbol{A}_{i_1,:}\boldsymbol{W}[k+\kappa](\boldsymbol{B}_{j_1,:})^{\mathsf{H}}\boldsymbol{B}_{j_2,:}\boldsymbol{W}^{\mathsf{H}}[k](\boldsymbol{A}_{i_2,:})^{\mathsf{H}}\big\} && (5.223)\\
&= \boldsymbol{A}_{i_1,:}\sum_{x=1}^{N_{\mathrm{R}}}\sum_{y=1}^{N_{\mathrm{R}}} b^*_{j_1,x}b_{j_2,y}\mathcal{E}\big\{\boldsymbol{W}_{:,x}[k+\kappa]\boldsymbol{W}^{\mathsf{H}}_{:,y}[k]\big\}\,(\boldsymbol{A}_{i_2,:})^{\mathsf{H}}. && (5.224)
\end{aligned}$$

Exploiting the fact that the elements of $\boldsymbol{W}[k]$ are iid with temporal correlation $\psi^{\mathrm{t}}[\kappa]$:

$$\begin{aligned}
&= \boldsymbol{A}_{i_1,:}(\boldsymbol{A}_{i_2,:})^{\mathsf{H}}\sum_{x=1}^{N_{\mathrm{R}}} b^*_{j_1,x}b_{j_2,x}\psi^{\mathrm{t}}[\kappa] && (5.225)\\
&= \boldsymbol{A}_{i_1,:}(\boldsymbol{A}_{i_2,:})^{\mathsf{H}}\boldsymbol{B}^*_{j_1,:}(\boldsymbol{B}_{j_2,:})^{\mathsf{T}}\psi^{\mathrm{t}}[\kappa] && (5.226)\\
&\triangleq \psi^{\mathrm{Tx}}[i_1, i_2]\psi^{\mathrm{Rx}}[j_1, j_2]\psi^{\mathrm{t}}[\kappa] && (5.227)
\end{aligned}$$

with $\psi^{\mathrm{Tx}}[i, j]$ and $\psi^{\mathrm{Rx}}[i, j]$ denoting the elements in the ith row and jth column of $\boldsymbol{\Psi}^{\mathrm{Tx}} = \boldsymbol{A}\boldsymbol{A}^{\mathsf{H}}$ and $\boldsymbol{\Psi}^{\mathrm{Rx}} = (\boldsymbol{B}\boldsymbol{B}^{\mathsf{H}})^{\mathsf{T}}$, respectively.

5.A.2 Derivation of Eqs. (5.69)–(5.71)

In order to prove that $\{x_{i,j}, y_{i,j}\}$ are iid pairs of correlated zero–mean complex Gaussian random variables with variances and covariance as given in (5.69)–(5.71) we recall that $x_{i,j}$ and $y_{i,j}$ are the elements in the ith row and jth column of $\sum_{\mu=1,\mu\neq n}^{N} m^*_{\mu,n}\boldsymbol{E}_{n,\mu}$ and $\boldsymbol{F}_n$ as defined in (5.64) and (5.63), respectively. Consequently, we have

$$\begin{aligned}
[x_{1,1}, x_{2,1}, \ldots, x_{N,1}, x_{1,2}, \ldots, x_{N,N}]^{\mathsf{H}} &= \mathrm{vec}\left\{\sum_{\mu=1,\mu\neq n}^{N} m^*_{\mu,n}\boldsymbol{E}_{n,\mu}\right\} && (5.228)\\
[y_{1,1}, y_{2,1}, \ldots, y_{N,1}, y_{1,2}, \ldots, y_{N,N}]^{\mathsf{H}} &= \mathrm{vec}\left\{\boldsymbol{F}_n\right\}. && (5.229)
\end{aligned}$$

Due to the definitions of $\boldsymbol{E}_{n,\mu}$ and $\boldsymbol{F}_n$ in (5.64) and (5.63) it is clear that $x_{i,j}$ and $y_{i,j}$ are zero–mean complex Gaussian random variables.

5.A.2.1 Variance σ_x^2

To show that the $x_{i,j}$ are mutually independent with equal variance we write with (5.64)

$$\mathcal{E}\left\{ \operatorname{vec}\left\{ \sum_{\mu=1,\mu\neq n}^{N} m_{\mu,n}^{*} \boldsymbol{E}_{n,\mu} \right\} \left[\operatorname{vec}\left\{ \sum_{\mu=1,\mu\neq n}^{N} m_{\mu,n}^{*} \boldsymbol{E}_{n,\mu} \right\} \right]^{\mathsf{H}} \right\} \tag{5.230}$$

$$= \sum_{\mu_1=1,\mu_1\neq n}^{N} m_{\mu_1,n}^{*} \sum_{\mu_2=1,\mu_2\neq n}^{N} m_{\mu_2,n} \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \left(\left(\boldsymbol{Q}\left(l,\hat{l}\right) \right)^{\mathsf{H}} \boldsymbol{S}_n \boldsymbol{S}_{\mu_1}^{\mathsf{H}} \right) \right) \tag{5.231}$$

$$\cdot\, \mathcal{E}\left\{ \operatorname{vec}\left\{ \boldsymbol{R}_{\mu_1} \right\} \operatorname{vec}\left\{ \boldsymbol{R}_{\mu_2} \right\}^{\mathsf{H}} \right\} \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \left(\boldsymbol{S}_{\mu_2} \boldsymbol{S}_n^{\mathsf{H}} \boldsymbol{Q}\left(l,\hat{l}\right) \right) \right).$$

Since the channel coefficients are spatially uncorrelated so are the elements of $\boldsymbol{R}_\mu$. Exploiting that both $\boldsymbol{S}_\mu$, $1 \leq \mu \leq N$, and $\boldsymbol{Q}\left(l,\hat{l}\right)$ are unitary, and introducing the nth unit vector $\boldsymbol{e}_n$ we can write

$$= \boldsymbol{I}_{N_{\mathrm{T}}N_{\mathrm{R}}} \sum_{\mu_1=1,\mu_1\neq n}^{N} m_{\mu_1,n}^{*} \sum_{\mu_2=1,\mu_2\neq n}^{N} m_{\mu_2,n} \left(\psi_{\mu_1,\mu_2}^{\mathrm{t}} + \delta[\mu_1-\mu_2]\,\sigma_n^2 \right) \tag{5.232}$$

$$= \boldsymbol{I}_{N_{\mathrm{T}}N_{\mathrm{R}}} \left(\boldsymbol{M}_{:,n} - \boldsymbol{e}_n m_{n,n} \right)^{\mathsf{H}} \left(\boldsymbol{\Psi}^{\mathrm{t}} + \sigma_n^2 \boldsymbol{I}_N \right) \left(\boldsymbol{M}_{:,n} - \boldsymbol{e}_n m_{n,n} \right). \tag{5.233}$$

Recalling that $\boldsymbol{M} = \left(\boldsymbol{\Psi}^{\mathrm{t}} + \sigma_n^2 \boldsymbol{I}_N \right)^{-1}$ [cf. (5.25)] we obtain

$$= \boldsymbol{I}_{N_{\mathrm{T}}N_{\mathrm{R}}} \left(\boldsymbol{M}_{:,n} - \boldsymbol{e}_n m_{n,n} \right)^{\mathsf{H}} \left(\boldsymbol{e}_n - m_{n,n} \left(\boldsymbol{\Psi}_{:,n}^{\mathrm{t}} + \sigma_n^2 \boldsymbol{e}_n \right) \right) \tag{5.234}$$

$$= \boldsymbol{I}_{N_{\mathrm{T}}N_{\mathrm{R}}} m_{n,n} \left(m_{n,n} \left(\psi_{n,n}^{\mathrm{t}} + \sigma_n^2 \right) - 1 \right), \tag{5.235}$$

($\boldsymbol{e}_n$: nth unit vector, $\delta[i-j]$: Kronecker–δ function which returns 1 if $i=j$ and 0 otherwise). Thus, with $\psi_{n,n}^{\mathrm{t}} = 1$ and (5.228) we obtain the desired result that the $x_{i,j}$ are iid with variance $\sigma_x^2 = m_{n,n}(m_{n,n}(1+\sigma_n^2)-1)$.

5.A.2.2 Variance σ_y^2

With (5.229), (5.63), the absence of spatial correlation in the channel and unitary $\boldsymbol{Q}\left(l,\hat{l}\right)$ and $\boldsymbol{S}_n$ we quickly find that

$$\mathcal{E}\left\{ \boldsymbol{y}\boldsymbol{y}^{\mathsf{H}} \right\} = \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \left(\boldsymbol{Q}\left(l,\hat{l}\right) \right)^{\mathsf{H}} \right) \mathcal{E}\left\{ \operatorname{vec}\left\{ \boldsymbol{R}_n \right\} \operatorname{vec}\left\{ \boldsymbol{R}_n \right\}^{\mathsf{H}} \right\} \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{Q}\left(l,\hat{l}\right) \right) \tag{5.236}$$

$$= \boldsymbol{I}_{N_{\mathrm{T}}N_{\mathrm{R}}} (1+\sigma_n^2), \tag{5.237}$$

proving that the $y_{i,j}$ are mutually independent with equal variance $\sigma_y^2 = 1+\sigma_n^2$.

5.A.2.3 Covariance μ_{xy}

Similar to the above proofs, we can write

$$\mathcal{E}\{\boldsymbol{x}\boldsymbol{y}^{\mathsf{H}}\} = \sum_{\mu=1,\mu\neq n}^{N} m_{\mu,n}^{*} \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \left(\left(\boldsymbol{Q}\left(l,\hat{l}\right)\right)^{\mathsf{H}} \boldsymbol{S}_n \boldsymbol{S}_{\mu}^{\mathsf{H}}\right)\right) \mathcal{E}\left\{\mathrm{vec}\left\{\boldsymbol{R}_{\mu}\right\} \mathrm{vec}\left\{\boldsymbol{R}_n\right\}^{\mathsf{H}}\right\} \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{Q}\left(l,\hat{l}\right)\right) \tag{5.238}$$

$$= \boldsymbol{I}_{N_{\mathrm{T}}N_{\mathrm{R}}} \left(\boldsymbol{\Psi}_{:,n}^{\mathrm{t}} + \sigma_n^2 \boldsymbol{e}_n\right)^{*} \left(\boldsymbol{M}_{n,:} - \boldsymbol{e}_n m_{n,n}\right)^{*} \tag{5.239}$$

$$= \boldsymbol{I}_{N_{\mathrm{T}}N_{\mathrm{R}}} \left(1 - m_{n,n}\left(1+\sigma_n^2\right)\right), \tag{5.240}$$

i.e. $\mu_{xy} = 1 - m_{n,n}\left(1+\sigma_n^2\right)$, which concludes the proof.

5.A.3 Derivation of Eq. (5.88)

It follows from (5.85) and $\sqrt{1+1/x} \asymp 1+1/(2x),\ x \to \infty$, that in the limit of infinite (effective) SNR

$$v_{1,i} \asymp \frac{1}{|\Theta_i|^2 \rho_{\mathrm{eff}}}, \quad [\sigma_n^2, \rho_{\mathrm{eff}}] \to [0,\infty] \tag{5.241}$$

$$v_{2,i} \asymp 1, \quad [\sigma_n^2, \rho_{\mathrm{eff}}] \to [0,\infty]. \tag{5.242}$$

Hence, the characteristic function $\Phi_{\Delta}(v)$ [cf. (5.331)] of Δ has one pole at $v = \mathrm{j}$ of multiplicity $N_{\mathrm{T}}N_{\mathrm{R}}$ and N_{Θ} poles at $v = -\mathrm{j}/(|\Theta_i|^2\rho_{\mathrm{eff}})$ of multiplicities l_i, $1 \le i \le N_{\Theta}$, and the expression for the PEP simplifies to

$$\mathrm{PEP} \asymp -\operatorname*{Res}_{v=\mathrm{j}} \left\{ \frac{1}{v} \frac{1}{(v-\mathrm{j})^{N_{\mathrm{T}}N_{\mathrm{R}}}} \prod_{i=1}^{N_{\Theta}} \left(\frac{v_{1,i}}{v+\mathrm{j}v_{1,i}}\right)^{l_i} \right\}, \quad [\sigma_n^2, \rho_{\mathrm{eff}}] \to [0,\infty]. \tag{5.243}$$

The remainder of the proof is done in analogy to that of the general error–probability result presented in Appendix 5.B based on repeated application of (5.333)–(5.334) with definitions

$$P(v) \triangleq \frac{1}{(v-\mathrm{j})^{N_{\mathrm{T}}N_{\mathrm{R}}}} \prod_{i=1}^{N_{\Theta}} \left(\frac{v_{1,i}}{v+\mathrm{j}v_{1,i}}\right)^{l_i} \tag{5.244}$$

$$Q^{(n)}(v) \triangleq \prod_{i=n}^{N_{\Theta}} \left(\frac{v_{1,i}}{v+\mathrm{j}v_{1,i}}\right)^{l_i}, \quad 1 \le n \le N_{\Theta}-1, \tag{5.245}$$

and with $\prod_{i=1}^{N_{\Theta}} |\Theta_i|^{l_i} = \prod_{i=1}^{N_{\mathrm{T}}} |\theta_i\left(l,\hat{l}\right)|$ leading directly to (5.88).

5.A.4 Derivation of Eq. (5.93)

In order to prove (5.93) we assume a suboptimal interpolation filter, prove that it fulfills (5.93) and show that the optimal filter can not do better than (5.93).

Let e_n denote the interpolation error in the interpolation of $\big[\boldsymbol{H}_n\big]_{i,j}$ from $\big[\boldsymbol{H}_\nu\big]_{i,j}$ using a filter with coefficients $\boldsymbol{p}_n = [p_1, \ldots, p_{n-1}, p_{n+1}, \ldots, p_N]^{\mathsf{T}}$, i.e.

$$e_n \stackrel{\triangle}{=} \big[\boldsymbol{H}_n\big]_{i,j} - \sum_{\nu=1,\nu\neq n}^{N} p_\nu \big[\boldsymbol{H}_\nu\big]_{i,j}. \tag{5.246}$$

Then, it is well known from fundamentals of stochastic processes (cf. e.g. [Kay98]) that the interpolation–error variance is given by

$$\sigma_{\mathrm{i},n}^2 \stackrel{\triangle}{=} \mathcal{E}\big\{|e_n|^2\big\} = T \int_{-1/(2T)}^{1/(2T)} \Psi^{\mathrm{t}}(f) P(f)\,\mathrm{d}f, \tag{5.247}$$

where $\Psi^{\mathrm{t}}(f)$ denotes the PSD of the effective discrete–time fading process [cf. (5.27)] and

$$P(f) \stackrel{\triangle}{=} \left| 1 - \sum_{\nu=1,\nu\neq n}^{N} p_\nu \mathrm{e}^{\mathrm{j}2\pi f T(n-\nu)} \right|^2 \tag{5.248}$$

is the power transfer function of the filter $\boldsymbol{p}$.

Recall that we are interested in the asymptotic behavior for very slow fading, i.e. fading PSDs with very small bandwidths. Hence, in order to achieve a small interpolation error variance $\sigma_{\mathrm{i},n}^2$ [cf. (5.247)] it appears reasonable to choose the interpolation coefficients such that the minimal power of f in (a series expansion of) $P(f)$ is maximized. Therefore, we use the series expansion $\exp(x) = \sum_{i=0}^{\infty} x^i/i!$ of the exponential function and write

$$\begin{aligned} P(f) &= \left| 1 - \sum_{\nu=1,\nu\neq n}^{N} \sum_{i=0}^{\infty} p_\nu \frac{(\mathrm{j}2\pi f T(n-\nu))^i}{i!} \right|^2 && (5.249)\\ &= \left| P_{N-2}(f) - \sum_{\nu=1,\nu\neq n}^{N} \sum_{i=N-1}^{\infty} p_\nu \frac{(\mathrm{j}2\pi f T(n-\nu))^i}{i!} \right|^2, && (5.250) \end{aligned}$$

where

$$P_{N-2}(f) \stackrel{\triangle}{=} 1 - \sum_{\nu=1,\nu\neq n}^{N} \sum_{i=0}^{N-2} p_\nu \frac{(\mathrm{j}2\pi f T(n-\nu))^i}{i!}. \tag{5.251}$$

Following our above goal, we choose the coefficients in $\boldsymbol{p}_n$ such that

$$P_{N-2}(f) = 0 \quad \forall f. \tag{5.252}$$

Since $P_{N-2}(f)$ is a polynomial in f of degree $N-2$, this can be achieved by choosing $\boldsymbol{p}_n$ such that all $N-1$ coefficients of this polynomial are zero. I.e. $\boldsymbol{p}_n$ is determined as solution of the

$((N-1)\times(N-1))$–dimensional system of equations

$$\begin{bmatrix} 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ n-1 & n-2 & \dots & 1 & -1 & \dots & n-N \\ (n-1)^2 & (n-2)^2 & \dots & 1 & (-1)^2 & \dots & (n-N)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (n-1)^{N-2} & (n-2)^{N-2} & \dots & 1 & (-1)^{N-2} & \dots & (n-N)^{N-2} \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n-1} \\ p_{n+1} \\ \vdots \\ p_N \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \tag{5.253}$$

Plugging this into (5.250) it is straightforward to see that in the limit of $fT \to 0$

$$P(f) \asymp c_0 \cdot (fT)^{2(N-1)}, \quad fT \to 0, \tag{5.254}$$

where

$$c_0 \triangleq \left| \frac{2\pi}{\Gamma(N)} \sum_{\nu=1,\nu\neq n}^{N} p_\nu (n-\nu)^{N-1} \right|^2 . \tag{5.255}$$

Provided that $\Psi^{\mathrm{t}}(f) = 0$, for $B_{h,\mathrm{eff}} < |f| \leq 1/(2T)$ we therefore obtain

$$\sigma^2_{\mathrm{i},n} \asymp c_0 T \int_{-B_{h,\mathrm{eff}}}^{B_{h,\mathrm{eff}}} \Psi^{\mathrm{t}}(f) \cdot (fT)^{2(N-1)} \,\mathrm{d}f, \quad B_{h,\mathrm{eff}} T \to 0, \tag{5.256}$$

which for PSDs that are bounded in the sense that

$$\max_{0\leq f\leq B_{h,\mathrm{eff}}} \{\Psi^{\mathrm{t,n}}(f)\} < \infty, \quad \text{with} \quad \Psi^{\mathrm{t,n}}(f) \triangleq \Psi^{\mathrm{t}}(f) B_{h,\mathrm{eff}} T \tag{5.257}$$

is trivially upper bounded by

$$\frac{2c_0}{2N-1} \max_{0\leq f\leq B_{h,\mathrm{eff}}} \{\Psi^{\mathrm{t,n}}(f)\} (B_{h,\mathrm{eff}} T)^{2(N-1)}. \tag{5.258}$$

This result does apparently not apply to the interesting special case of Clarke's fading model, as the magnitude of its PSD is not bounded in the sense of (5.257), cf. (4.22). Here however, the exact solution can be given in closed form as

$$\begin{aligned} T\int_{-B_{h,\mathrm{eff}}}^{B_{h,\mathrm{eff}}} \Psi^{\mathrm{t}}(f) \cdot (fT)^{2(N-1)} \,\mathrm{d}f &= \frac{2T^{2(N-1)}}{\pi} \int_0^{B_{h,\mathrm{eff}}} \frac{f^{2(N-1)}}{\sqrt{B^2_{h,\mathrm{eff}} - f^2}} \,\mathrm{d}f & (5.259) \\ &= \frac{2T^{2(N-1)}}{(2N-1)\pi} B^{2(N-1)}_{h,\mathrm{eff}} \, {}_2\mathrm{F}_1\left(N-\frac{1}{2}; \frac{1}{2}; N+\frac{1}{2}; 1\right), & (5.260) \end{aligned}$$

where ${}_2\mathrm{F}_1(a,b,c,z)$ denotes the hypergeometric function, cf. e.g. [AS72, Ch. 15]. With Gauss' hypergeometric theorem ${}_2\mathrm{F}_1(a;b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}$, $\sigma^2_{\mathrm{i},n}$ is for Clarke's fading model asymptotically given by

$$\sigma^2_{\mathrm{i},n} \asymp \frac{2c_0\langle N\rangle_{1/2}}{(2N-1)\sqrt{\pi}} \cdot (B_{h,\mathrm{eff}}T)^{2(N-1)}, \quad B_{h,\mathrm{eff}}T \to 0. \tag{5.261}$$

As these results apply to a suboptimal filter, whose coefficients $\boldsymbol{p}_n$ are chosen according to (5.253), it is clear —by definition— that linear MMSE interpolation can not lead to higher interpolation–error variances. At the same time, as the number of degrees of freedom in the design of the filter is limited by the number $N-1$ of coefficients p_ν, there can not be a filter that achieves $\sigma^2_{\mathrm{i},n} \asymp c \cdot (B_{h,\mathrm{eff}}T)^x,\ B_{h,\mathrm{eff}}T \to 0$ with $x > 2(N-1)$.

Due to the relationship between MSDD and linear MMSE interpolation (cf. Section 2.4.2.1) we therefore obtain in the limit of $B_{h,\mathrm{eff}}T \to 0$

$$\left[\left(\boldsymbol{\Psi}^{\mathrm{t}}\right)^{-1}\right]_{n,n} \asymp c \cdot (B_{h,\mathrm{eff}}T)^{-2(N-1)}, \quad B_{h,\mathrm{eff}}T \to 0, \tag{5.262}$$

with some constant c, which concludes our proof of (5.93).

5.A.5 Derivation of Eq. (5.114)

Plugging the definitions of $\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}|\bar{\boldsymbol{S}}}$ and $\boldsymbol{F}$ from (5.35) and (5.40), respectively, into expression (5.44) for the characteristic function $\Phi_\Delta(v)$ and making use of the identities $\det\{\boldsymbol{I}_a + \boldsymbol{AB}\} = \det\{\boldsymbol{I}_b + \boldsymbol{BA}\}$ and $(\boldsymbol{I}_N \otimes \boldsymbol{A})(\boldsymbol{I}_N \otimes \boldsymbol{B}) = \boldsymbol{I}_N \otimes (\boldsymbol{AB})$, where a and b denote the numbers of rows of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, and we can write

$$\begin{aligned}
\Phi_\Delta(v) &= \det\left\{\boldsymbol{I}_{NN_{\mathrm{T}}N_{\mathrm{R}}} - \mathrm{j}v\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}|\bar{\boldsymbol{S}}}(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{F})\right\}^{-1} && (5.263)\\
&= \det\Big\{\boldsymbol{I}_{NN_{\mathrm{T}}N_{\mathrm{R}}} - \mathrm{j}v\left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \bar{\boldsymbol{S}}_{\mathrm{D,c}}\right)\boldsymbol{T}_{\mathrm{c}}\left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \bar{\boldsymbol{S}}^{\mathsf{H}}_{\mathrm{D,c}}\right) && (5.264)\\
&\qquad \cdot \Big[\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \left(\hat{\bar{\boldsymbol{S}}}_{\mathrm{D}}\left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\hat{\bar{\boldsymbol{S}}}^{\mathsf{H}}_{\mathrm{D}} - \bar{\boldsymbol{S}}_{\mathrm{D}}\left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\bar{\boldsymbol{S}}^{\mathsf{H}}_{\mathrm{D}}\right)\Big]\Big\}^{-1}\\
&= \det\Big\{\boldsymbol{I}_{NN_{\mathrm{T}}N_{\mathrm{R}}} - \mathrm{j}v\boldsymbol{T}_{\mathrm{c}}\Big[\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \Big(\bar{\boldsymbol{S}}^{\mathsf{H}}_{\mathrm{D,c}}\hat{\bar{\boldsymbol{S}}}_{\mathrm{D}}\left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\hat{\bar{\boldsymbol{S}}}^{\mathsf{H}}_{\mathrm{D}}\bar{\boldsymbol{S}}_{\mathrm{D,c}}\\
&\qquad -\bar{\boldsymbol{S}}^{\mathsf{H}}_{\mathrm{D,c}}\bar{\boldsymbol{S}}_{\mathrm{D}}\left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\bar{\boldsymbol{S}}^{\mathsf{H}}_{\mathrm{D}}\bar{\boldsymbol{S}}_{\mathrm{D,c}}\Big)\Big]\Big\}^{-1}. && (5.265)
\end{aligned}$$

If we further introduce

$$\boldsymbol{P} \stackrel{\triangle}{=} \bar{\boldsymbol{S}}^{\mathsf{H}}_{\mathrm{D}}\bar{\boldsymbol{S}}_{\mathrm{D,c}} \tag{5.266}$$

and exploit that $\bar{\boldsymbol{S}}_{\mathrm{D}}$ is unitary, we obtain

$$\Phi_\Delta(v) = \det\Big\{\boldsymbol{I}_{NN_{\mathrm{T}}N_{\mathrm{R}}} - \mathrm{j}v\boldsymbol{T}_{\mathrm{c}}\Big[\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \Big(\boldsymbol{P}^{\mathsf{H}}\bar{\boldsymbol{S}}^{\mathsf{H}}_{\mathrm{D}}\hat{\bar{\boldsymbol{S}}}_{\mathrm{D}}\left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\hat{\bar{\boldsymbol{S}}}^{\mathsf{H}}_{\mathrm{D}}\bar{\boldsymbol{S}}_{\mathrm{D}}\boldsymbol{P} - \boldsymbol{P}^{\mathsf{H}}\left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}}\right)\boldsymbol{P}\Big)\Big]\Big\}^{-1}. \tag{5.267}$$

From this, one can see that the characteristic function $\Phi_\Delta(v)$ of the metric difference between $\bar{\boldsymbol{S}}$ and $\hat{\bar{\boldsymbol{S}}}$ and through (5.46) also the $\mathrm{PEP}\big(\bar{\boldsymbol{S}} \to \hat{\bar{\boldsymbol{S}}}\big)$ depends on $\bar{\boldsymbol{S}}$ and $\hat{\bar{\boldsymbol{S}}}$ only through the matrix

$$\boldsymbol{T}_{\mathrm{c}} \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \left(\boldsymbol{P}^{\mathsf{H}} \left(\bar{\boldsymbol{S}}_{\mathrm{D}}^{\mathsf{H}} \hat{\bar{\boldsymbol{S}}}_{\mathrm{D}} \left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}} \right) \hat{\bar{\boldsymbol{S}}}_{\mathrm{D}}^{\mathsf{H}} \bar{\boldsymbol{S}}_{\mathrm{D}} - \left(\boldsymbol{M} \otimes \boldsymbol{I}_{N_{\mathrm{T}}} \right) \right) \boldsymbol{P} \right) \right) \tag{5.268}$$

as stated in (5.114).

5.A.6 Proof of Eq. (5.208)

In this appendix, we prove (5.208) by showing that for DSTM group codes

$$1 \leq \lim_{N \to \infty} \frac{C}{C_{\mathrm{as}}} \leq 1 + \varepsilon_{\mathrm{as}}, \tag{5.269}$$

with C_{as} given in (5.208) and some arbitrarily small positive constant $\varepsilon_{\mathrm{as}}$. At this, it is intuitively convenient, to regard $\bar{\boldsymbol{Z}}_n$ as defined in (5.182) as a sequence of $\tilde{n}-1$ random matrices uniformly iid over the signal constellation $\boldsymbol{\mathcal{V}}$ and $C_n^{\mathrm{blo}}/L^{\tilde{n}-1}$ as the expectation of $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \leq \rho \mid \bar{\boldsymbol{S}}_n\big)$ with respect to $\bar{\boldsymbol{Z}}_n$, i.e.

$$C_n^{\mathrm{blo}} = \mathcal{E}_{\bar{\boldsymbol{Z}}_n} \Big\{ \Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \leq \rho \mid \bar{\boldsymbol{S}}_n\big) \Big\} L^{\tilde{n}-1}. \tag{5.270}$$

For the proof, we make use of the method of "types" [CT91].

A type $\boldsymbol{q}_{\bar{\boldsymbol{Z}}_n}$ (or empirical probability distribution) of $\bar{\boldsymbol{Z}}_n \in \boldsymbol{\mathcal{V}}^{\tilde{n}-1}$ is defined as an L–dimensional vector whose lth component is given by $q_{\bar{\boldsymbol{Z}}_n,l} = P(\boldsymbol{V}^{(l)}|\bar{\boldsymbol{Z}}_n)/(\tilde{n}-1)$, where $P(\boldsymbol{V}^{(l)}|\bar{\boldsymbol{Z}}_n)$ denotes the number of occurrences of the symbol $\boldsymbol{V}^{(l)}$ as submatrices in $\bar{\boldsymbol{Z}}_n$. The set of all $\bar{\boldsymbol{Z}}_n$ of a particular type $\boldsymbol{q}_{\bar{\boldsymbol{Z}}_n} = \boldsymbol{q}$ is referred to as "type class" and denoted by

$$\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)} \stackrel{\triangle}{=} \left\{ \bar{\boldsymbol{Z}}_n \in \boldsymbol{\mathcal{V}}^{\tilde{n}-1} \;\middle|\; \boldsymbol{q}_{\bar{\boldsymbol{Z}}_n} = \boldsymbol{q} \right\}. \tag{5.271}$$

Let us further define the "strongly typical set" [CT91]

$$\boldsymbol{\mathcal{A}}_{\mu_{\mathrm{as}}}^{(n)} \stackrel{\triangle}{=} \left\{ \bar{\boldsymbol{Z}}_n \in \boldsymbol{\mathcal{V}}^{\tilde{n}-1} \;\middle|\; \left| q_{\bar{\boldsymbol{Z}}_n,l} - p_l \right| < \frac{\mu_{\mathrm{as}}}{L}, l \in \{1, \ldots, L\} \right\}, \tag{5.272}$$

where $\boldsymbol{p} = \boldsymbol{1}_{L,1}/L$ denotes the true distribution of the elements of $\boldsymbol{\mathcal{V}}$ (or "typical type" of $\bar{\boldsymbol{Z}}_n$) and μ_{as} is an usually arbitrarily small positive constant which specifies how "typical" a $\bar{\boldsymbol{Z}}_n$ has to be to be admitted to the strongly typical set. Further, let

$$\boldsymbol{\mathcal{Q}}^{(n)} \stackrel{\triangle}{=} \left\{ \boldsymbol{q}_{\bar{\boldsymbol{Z}}_n} \;\middle|\; \bar{\boldsymbol{Z}}_n \in \boldsymbol{\mathcal{V}}^{\tilde{n}-1} \right\}, \tag{5.273}$$

$$\boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)} \stackrel{\triangle}{=} \left\{ \boldsymbol{q}_{\bar{\boldsymbol{Z}}_n} \;\middle|\; \bar{\boldsymbol{Z}}_n \in \boldsymbol{\mathcal{A}}_{\mu_{\mathrm{as}}}^{(n)} \right\}, \tag{5.274}$$

denote the set of all types $\boldsymbol{q}_{\bar{\boldsymbol{Z}}_n}$ and the set of the different types of $\bar{\boldsymbol{Z}}_n$ that belong to the strongly typical set $\boldsymbol{\mathcal{A}}_{\mu_{\mathrm{as}}}^{(n)}$, respectively. Finally, let us with (5.207) introduce for convenience

$$\alpha_{\boldsymbol{q}}^{(n)} \stackrel{\triangle}{=} \Pr\left(||\boldsymbol{H}||^2 - \frac{1}{\tilde{n}^2} ||\boldsymbol{\Xi}_n \boldsymbol{H}||^2 \leq \sigma_n^2 (\rho/\tilde{n} - N_{\mathrm{T}} N_{\mathrm{R}}) \right) \quad \text{with} \quad \bar{\boldsymbol{Z}}_n \in \boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}, \tag{5.275}$$

which due to $||\boldsymbol{\Xi}_n \boldsymbol{H}||^2 \leq ||\boldsymbol{\Xi}_n||^2 \cdot ||\boldsymbol{H}||^2$ can be upper bounded by

$$\alpha_{\boldsymbol{q}}^{(n)} \leq \Pr\left(||\boldsymbol{H}||^2 \leq \frac{\sigma_n^2(\rho/\tilde{n} - N_{\mathrm{T}}N_{\mathrm{R}})}{1 - ||\boldsymbol{\Xi}_n||^2/\tilde{n}^2} \right). \tag{5.276}$$

For $\boldsymbol{\Xi}_n$ as defined in (5.188) we have for any $\boldsymbol{q}$,

$$\lim_{\tilde{n}\to\infty} \boldsymbol{\Xi}_n/\tilde{n} = \lim_{\tilde{n}\to\infty} \frac{1}{\tilde{n}} \left(\boldsymbol{I}_{N_{\mathrm{T}}} + \sum_{l=1}^{L} q_l(\tilde{n}-1)\boldsymbol{V}^{(l)} \right) = \sum_{l=1}^{L} (q_l - p_l)\boldsymbol{V}^{(l)}, \tag{5.277}$$

where we used $\sum_{l=1}^{L} \boldsymbol{V}^{(l)} = \boldsymbol{0}$, and in particular $\lim_{\tilde{n}\to\infty} \boldsymbol{\Xi}_n/\tilde{n} = \boldsymbol{0}$ for $\boldsymbol{q} = \boldsymbol{p}$. Plugging this into (5.276) and recalling that the norm $||\boldsymbol{H}||^2$ is a $\chi^2\big(N_{\mathrm{T}}N_{\mathrm{R}}, 2N_{\mathrm{T}}N_{\mathrm{R}}\big)$ random variable, as the elements of $\boldsymbol{H}$ are iid $\mathcal{N}_{\mathrm{c}}\big(0,1\big)$ random variables, we obtain

$$\begin{aligned} \lim_{\tilde{n}\to\infty} \frac{\gamma\left(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2(\rho/\tilde{n} - N_{\mathrm{T}}N_{\mathrm{R}})\right)}{\Gamma(N_{\mathrm{T}}N_{\mathrm{R}})} &= \lim_{\tilde{n}\to\infty} \alpha_{\boldsymbol{p}}^{(n)} \\ &\leq \lim_{\tilde{n}\to\infty} \alpha_{\boldsymbol{q}}^{(n)} \leq \lim_{\tilde{n}\to\infty} \frac{\gamma\left(N_{\mathrm{T}}N_{\mathrm{R}}, \frac{\sigma_n^2(\rho/\tilde{n} - N_{\mathrm{T}}N_{\mathrm{R}})}{1 - \left|\left|\sum_{l=1}^{L}(q_l - p_l)\boldsymbol{V}^{(l)}\right|\right|^2}\right)}{\Gamma(N_{\mathrm{T}}N_{\mathrm{R}})}. \end{aligned} \tag{5.278}$$

With this and (5.208) we can write

$$\lim_{N\to\infty} \frac{C^{\mathrm{blo}}}{C_{\mathrm{as}}} = \frac{L-1}{L} \lim_{N\to\infty} \frac{C^{\mathrm{blo}}}{\alpha_{\boldsymbol{p}}^{(1)} L^{N-1}} \tag{5.279}$$

and we will in the following prove (5.204) with C_{as} as defined in (5.208) by deriving bounds on $\lim_{N\to\infty} C^{\mathrm{blo}}/(\alpha_{\boldsymbol{p}}^{(1)} L^{N-1})$.

Remark: Note that the "typical" sequences $\bar{\boldsymbol{Z}}_n$ are not those few that represent strongly correlated pairs $\big\{\tilde{\bar{\boldsymbol{S}}}_n, \bar{\boldsymbol{S}}_n\big\}$, i.e. for which $||\boldsymbol{\Xi}_n||$ is large, but those that correspond to the vast majority of $\tilde{\bar{\boldsymbol{S}}}_n$ that are not correlated with $\bar{\boldsymbol{S}}_n$. Since MSDD is essentially an *estimator–correlator* structure [Kai60] it is intuitively reasonable that these candidates are unlikely to become the ML solution, i.e. to lie inside a sphere of given radius with higher probability than other "untypical" $\tilde{\bar{\boldsymbol{S}}}_n$ that are strongly correlated with $\bar{\boldsymbol{S}}_n$. This fact is expressed by (5.278).

Lower Bound: Using the definition of types we can write

$$\begin{aligned}
\liminf_{N\to\infty} \frac{C^{\mathrm{blo}}}{\alpha_{\boldsymbol{p}}^{(1)} L^{N-1}} &= \liminf_{N\to\infty} \frac{\sum_{n=1}^{N-1} \sum_{\bar{\boldsymbol{Z}}_n \in \boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}, \forall \boldsymbol{q} \in \boldsymbol{\mathcal{Q}}^{(n)}} |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}| \Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \leq \rho \mid \bar{\boldsymbol{S}}_n\big)}{\alpha_{\boldsymbol{p}}^{(1)} L^{N-1}} && (5.280)\\
&\geq \liminf_{N\to\infty} \frac{\sum_{n=1}^{N-1} \sum_{\bar{\boldsymbol{Z}}_n \in \boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}, \forall \boldsymbol{q} \in \boldsymbol{\mathcal{Q}}^{(n)}} |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}| \Pr\big(d_1\big(\tilde{\bar{\boldsymbol{S}}}\big) \leq \rho \mid \bar{\boldsymbol{S}}\big)}{\alpha_{\boldsymbol{p}}^{(1)} L^{N-1}} && (5.281)\\
&\geq \liminf_{N\to\infty} \frac{\sum_{n=1}^{N-1} \sum_{\forall \boldsymbol{q} \in \boldsymbol{\mathcal{Q}}^{(n)}} |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}| \alpha_{\boldsymbol{q}}^{(1)}}{\alpha_{\boldsymbol{p}}^{(1)} L^{N-1}} && (5.282)\\
&\geq \liminf_{N\to\infty} \frac{\sum_{n=1}^{N-1} \sum_{\forall \boldsymbol{q} \in \boldsymbol{\mathcal{Q}}^{(n)}} |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}|}{L^{N-1}} && (5.283)\\
&= \liminf_{N\to\infty} \sum_{n=1}^{N-1} L^{-n+1} && (5.284)\\
&= \frac{L}{L-1} \stackrel{\triangle}{=} B_{\mathrm{L}} \,. && (5.285)
\end{aligned}$$

where we used (i) the fact that $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \leq \rho \mid \bar{\boldsymbol{S}}_n\big) \geq \Pr\big(d_1\big(\tilde{\bar{\boldsymbol{S}}}\big) \leq \rho \mid \bar{\boldsymbol{S}}\big)$, $\forall \tilde{\bar{\boldsymbol{S}}}, \bar{\boldsymbol{S}} \in \boldsymbol{\mathcal{V}}^N$ and the corresponding $\tilde{\bar{\boldsymbol{S}}}_n$, $\bar{\boldsymbol{S}}_n$, $1 \leq n \leq N-1$ to obtain (5.281), (ii) the relations (5.207), (5.275) and (5.276) to get to (5.282), (iii) (5.278) for writing (5.283), (iv) the identity $\sum_{\forall \boldsymbol{q} \in \boldsymbol{\mathcal{Q}}^{(n)}} |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}| = L^{N-n}$ in (5.284), and the fact that the limit is (5.284) is well–known to exist.

Upper Bound: We distinguish between $\tilde{n}$ for $1 \leq n \leq N/2$ where (5.207) is valid and for $N/2+1 \leq n \leq N-1$, where we use a simple upper bound on $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \leq \rho \mid \bar{\boldsymbol{S}}_n\big)$. At this, the choice of the value $N/2$ is rather arbitrary.

For $1 \leq n \leq N/2$, i.e. $N \to \infty$ implies $\tilde{n} \to \infty$, we can write with (5.275)

$$\begin{aligned}
\limsup_{N\to\infty} \frac{C_n^{\mathrm{blo}}}{\alpha_{\boldsymbol{p}}^{(1)} L^{N-1}} &= \limsup_{N\to\infty} \sum_{\forall \boldsymbol{q} \in \boldsymbol{\mathcal{Q}}^{(n)}} \frac{\alpha_{\boldsymbol{q}}^{(n)}}{\alpha_{\boldsymbol{p}}^{(1)}} |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}| L^{-N+1} && (5.286)\\
&= \limsup_{N\to\infty} \left[\sum_{\forall \boldsymbol{q} \in \boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}} \frac{\alpha_{\boldsymbol{q}}^{(n)}}{\alpha_{\boldsymbol{p}}^{(1)}} |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}| \right] \cdot \left[1 + \frac{\sum_{\forall \boldsymbol{q} \notin \boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}} \alpha_{\boldsymbol{q}}^{(n)} |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}|}{\sum_{\forall \boldsymbol{q} \in \boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}} \alpha_{\boldsymbol{q}}^{(n)} |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}|} \right] L^{-N+1} && (5.287)\\
&\leq \limsup_{N\to\infty} \left| \boldsymbol{\mathcal{A}}_{\mu_{\mathrm{as}}}^{(n)} \right| \max_{\boldsymbol{q} \in \boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}} \left\{ \frac{\alpha_{\boldsymbol{q}}^{(n)}}{\alpha_{\boldsymbol{p}}^{(1)}} \right\} \left[1 + \frac{\sum_{\forall \boldsymbol{q} \notin \boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}} \alpha_{\boldsymbol{q}}^{(n)} 2^{(\tilde{n}-1)H(\boldsymbol{q})}}{\alpha_{\boldsymbol{p}}^{(n)} \tilde{n}^{-L} 2^{(\tilde{n}-1)H(\boldsymbol{p})}} \right] L^{-N+1} && (5.288)\\
&\leq \limsup_{N\to\infty} \max_{\boldsymbol{q} \in \boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}} \left\{ \frac{\alpha_{\boldsymbol{q}}^{(n)}}{\alpha_{\boldsymbol{p}}^{(1)}} \right\} \left[1 + \tilde{n}^L \sum_{\forall \boldsymbol{q} \notin \boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}} \frac{\alpha_{\boldsymbol{q}}^{(n)}}{\alpha_{\boldsymbol{p}}^{(n)}} 2^{-(\tilde{n}-1)(H(\boldsymbol{p})-H(\boldsymbol{q}))} \right] L^{-n+1} && (5.289)\\
&= \limsup_{N\to\infty} \max_{\boldsymbol{q} \in \boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}} \left\{ \frac{\alpha_{\boldsymbol{q}}^{(n)}}{\alpha_{\boldsymbol{p}}^{(1)}} \right\} L^{-n+1}, && (5.290)
\end{aligned}$$

where we used [CT91, Theorem 12.1.3], which states that the size of a type class $\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}$ is bounded by $\tilde{n}^{-L} 2^{(\tilde{n}-1)H(\boldsymbol{q})} \leq |\boldsymbol{\mathcal{T}}_{\boldsymbol{q}}^{(n)}| \leq 2^{(\tilde{n}-1)H(\boldsymbol{q})}$ with the entropy $H(\boldsymbol{q}) \stackrel{\triangle}{=} -\sum_{l=1}^{L} q_l \log_2(q_l)$ of

a memoryless source whose distribution is given by the type or empirical probability distribution $\boldsymbol{q}$, to get to (5.288). To obtain (5.289) we used the trivial upper bound $|\boldsymbol{\mathcal{A}}_{\mu_{\mathrm{as}}}^{(n)}| \leq L^{\tilde{n}-1}$. The second addend in the squared brackets in (5.289) tends to zero as $\tilde{n} \to \infty$, because (i) $1 < \alpha_{\boldsymbol{q}}^{(n)}/\alpha_{\boldsymbol{p}}^{(n)} < 1/\alpha_{\boldsymbol{p}}^{(n)}$ [cf. (5.278)], (ii) $H(\boldsymbol{q}) < H(\boldsymbol{p})$, $\forall \boldsymbol{q} \notin \boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}$, $1 \leq n \leq N-1$, for any arbitrarily small positive constant μ_{as}, and (iii) the number of different types $\boldsymbol{q}$ is according to [CT91, Theorem 12.1.1] at most polynomial in $\tilde{n}$. With (5.278) we therefore obtain

$$\limsup_{N\to\infty} \frac{C_n^{\mathrm{blo}}}{\alpha_{\boldsymbol{p}}^{(1)} L^{N-1}} \leq \limsup_{N\to\infty} \frac{\gamma\left(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2(\rho/\tilde{n} - N_{\mathrm{T}}N_{\mathrm{R}})\beta^{(n)}\right)}{\gamma\left(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2(\rho/N - N_{\mathrm{T}}N_{\mathrm{R}})\right)} \cdot L^{-n+1}, \quad 1 \leq n \leq N/2, \tag{5.291}$$

where

$$\beta^{(n)} \stackrel{\triangle}{=} \max_{\boldsymbol{q}\in\boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}} \left\{ \left[1 - \left\| \sum_{l=1}^{L} (q_l - p_l) \boldsymbol{V}^{(l)} \right\|^2 \right]^{-1} \right\}. \tag{5.292}$$

Note that, since $|q_l - p_l| < \mu_{\mathrm{as}}/L$, $\forall \boldsymbol{q} \in \boldsymbol{\mathcal{Q}}_{\mu_{\mathrm{as}}}^{(n)}$, $\beta^{(n)}$ can by choice of μ_{as} be brought arbitrarily close to 1, regardless of $\tilde{n}$.

For $N/2+1 \leq n \leq N-1$ we use the trivial upper bound $\Pr\big(d_n\big(\tilde{\bar{\boldsymbol{S}}}_n\big) \leq \rho \mid \bar{\boldsymbol{S}}_n\big) \leq 1$, $\forall \bar{\boldsymbol{Z}}_n \in \boldsymbol{\mathcal{V}}^{\tilde{n}-1}$, i.e.

$$\limsup_{N\to\infty} \frac{C_n^{\mathrm{blo}}}{\alpha_{\boldsymbol{p}}^{(1)} L^{N-1}} \leq \limsup_{N\to\infty} \frac{\Gamma(N_{\mathrm{T}}N_{\mathrm{R}})}{\gamma(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2(\rho/N - N_{\mathrm{T}}N_{\mathrm{R}}))} \cdot L^{-n+1}, \quad N/2+1 \leq n \leq N-1. \tag{5.293}$$

We thus have

$$\limsup_{N\to\infty} \frac{C^{\mathrm{blo}}}{\alpha_{\boldsymbol{p}}^{(1)} L^{N-1}} \leq \limsup_{N\to\infty} B_{\mathrm{U}}(N)\,, \tag{5.294}$$

where

$$B_{\mathrm{U}}(N) \stackrel{\triangle}{=} \underbrace{\sum_{n=1}^{N/2} \frac{\gamma(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2(\rho/\tilde{n} - N_{\mathrm{T}}N_{\mathrm{R}})\beta^{(n)})}{\gamma(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2(\rho/N - N_{\mathrm{T}}N_{\mathrm{R}}))} L^{-n+1}}_{\stackrel{\triangle}{=} B_{\mathrm{U},1}(N)} + \underbrace{\sum_{n=N/2+1}^{N-1} \frac{\Gamma(N_{\mathrm{T}}N_{\mathrm{R}})}{\gamma(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2(\rho/N - N_{\mathrm{T}}N_{\mathrm{R}}))} L^{-n+1}}_{\stackrel{\triangle}{=} B_{\mathrm{U},2}(N)}. \tag{5.295}$$

Let us first consider the term $B_{\mathrm{U},1}(N)$. Using the series expansion of the incomplete Gamma function

$$\gamma(N, x) = \frac{x^N}{N} \left(1 + \sum_{k=1}^{\infty} \frac{N+1}{N+k+1} \frac{(-x)^k}{k!} \right) \tag{5.296}$$

we obtain with $\beta^{(n)} \geq 1$

$$\frac{\gamma(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2(\rho/\tilde{n} - N_{\mathrm{T}}N_{\mathrm{R}})\beta^{(n)})}{\gamma(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2(\rho/N - N_{\mathrm{T}}N_{\mathrm{R}}))} \leq \left(\frac{(\rho/\tilde{n} - N_{\mathrm{T}}N_{\mathrm{R}})\beta^{(n)}}{\rho/N - N_{\mathrm{T}}N_{\mathrm{R}}} \right)^{N_{\mathrm{T}}N_{\mathrm{R}}}, \tag{5.297}$$

which is increasingly tight for small arguments of the incomplete Gamma function. With (5.297), $\rho = (1+\varepsilon_\rho)NN_{\mathrm{T}}N_{\mathrm{R}}$ and

$$\varepsilon_{\mathrm{as}} \stackrel{\triangle}{=} \max_{n\in\{1,\ldots,N/2\}} \{\beta^{(n)}\} - 1 \tag{5.298}$$

we can upper bound $B_{\mathrm{U},1}(N)$ via

$$\begin{aligned}
\frac{B_{\mathrm{U},1}(N)}{(1+\varepsilon_{\mathrm{as}})^{N_{\mathrm{T}}N_{\mathrm{R}}}} &\leq \sum_{n=1}^{N/2} \left(\frac{\rho/\tilde{n} - N_{\mathrm{T}}N_{\mathrm{R}}}{\rho/N - N_{\mathrm{T}}N_{\mathrm{R}}}\right)^{N_{\mathrm{T}}N_{\mathrm{R}}} L^{-n+1} && (5.299)\\
&= \sum_{n=1}^{N/2} \left(1 + \sum_{k=1}^{N_{\mathrm{T}}N_{\mathrm{R}}} \binom{N_{\mathrm{T}}N_{\mathrm{R}}}{k} \cdot \left(\frac{\rho(n-1)}{(\rho - NN_{\mathrm{T}}N_{\mathrm{R}})\tilde{n}}\right)^k\right) L^{-n+1} && (5.300)\\
&= \frac{L(1-L^{-N/2})}{L-1} + \sum_{k=1}^{N_{\mathrm{T}}N_{\mathrm{R}}} \binom{N_{\mathrm{T}}N_{\mathrm{R}}}{k} \cdot \left(1+\frac{1}{\varepsilon_\rho}\right)^k \sum_{n=1}^{N/2} \left(\frac{n-1}{N-n+1}\right)^k L^{-n+1}\,. && (5.301)
\end{aligned}$$

In the limit $N \to \infty$ the first term in (5.301) converges to $L/(L-1)$ while the second vanishes, as

$$\lim_{N\to\infty} \sum_{n=1}^{N/2-1} \left(\frac{n}{N-n}\right)^k L^{-n} \leq \lim_{N\to\infty} \frac{1}{(N/2+1)^k} \left(\frac{L}{L-1}\right)^{k+1} \sum_{i=1}^{k} c_i^{(k)} L^{-i} \;=\; 0,\; k \in \mathbb{N}, \tag{5.302}$$

where we used (cf. Appendix 5.A.7)

$$\sum_{n=1}^{\infty} n^k x^n = (1-x)^{-(k+1)} \sum_{i=1}^{k} c_i^{(k)} x^i, \quad |x| < 1,\; k \in \mathbb{N}, \tag{5.303}$$

with coefficients $c_i^{(k)}$ that can be computed recursively via

$$\begin{aligned}
c_1^{(k)} &= c_k^{(k)} = 1 && (5.304)\\
c_i^{(k)} &= i c_i^{(k-1)} + (k-i+1) c_{i-1}^{(k-1)}. && (5.305)
\end{aligned}$$

With $\rho = (1+\varepsilon_\rho)NN_{\mathrm{T}}N_{\mathrm{R}}$ the second addend $B_{\mathrm{U},2}(N)$ of the upper bound (5.295) can be rewritten as

$$B_{\mathrm{U},2}(N) = \frac{\Gamma(N_{\mathrm{T}}N_{\mathrm{R}})}{\gamma(N_{\mathrm{T}}N_{\mathrm{R}}, \sigma_n^2 \varepsilon_\rho N_{\mathrm{T}}N_{\mathrm{R}})} \cdot \frac{L^{-N/2+1} - L^{-N+2}}{L-1}\,. \tag{5.306}$$

Since the first factor in (5.306) is independent of N, $B_{\mathrm{U},2}(N)$ becomes zero for $N \to \infty$.

Hence, the upper bound (5.294) converges as

$$\limsup_{N\to\infty} B_{\mathrm{U}}(N) = \limsup_{N\to\infty} B_{\mathrm{U},1}(N) + \limsup_{N\to\infty} B_{\mathrm{U},2}(N) = \frac{L}{L-1}(1+\varepsilon_{\mathrm{as}})^{N_{\mathrm{T}}N_{\mathrm{R}}}\,, \tag{5.307}$$

which together with the lower bound (5.285) and $\varepsilon_{\mathrm{as}}$ arbitrarily small establishes (5.204) as desired.

5.A.7 Proof of Eq. (5.303)

In this appendix, we prove that

$$\sum_{n=1}^{\infty} n^k x^n = (1-x)^{-(k+1)} \sum_{i=1}^{k} c_i^{(k)} x^i, \quad |x| < 1,\ k \in \mathbb{N} \tag{5.308}$$

with coefficients $c_i^{(k)}$ that can be computed recursively via

$$c_1^{(k)} = c_1^{(k-1)} = 1 \tag{5.309}$$

$$c_k^{(k)} = c_k^{(k-1)} = 1 \tag{5.310}$$

$$c_i^{(k)} = i c_i^{(k-1)} + (k-i+1) c_{i-1}^{(k-1)}. \tag{5.311}$$

To this end we deploy the technique of complete induction, i.e. we show that (5.308) holds for $k = 1$, $k = 2$ and that it holds for k if it holds for $k-1$. At this, we repeatedly make use of the relations

$$\sum_{n=1}^{\infty} n^k x^n = x \frac{\mathrm{d}}{\mathrm{d}x} \sum_{n=1}^{\infty} n^{k-1} x^n. \tag{5.312}$$

For $k = 1$ we can write

$$\sum_{n=1}^{\infty} n x^n = x \frac{\mathrm{d}}{\mathrm{d}x} \sum_{n=1}^{\infty} x^n \tag{5.313}$$

$$= x \frac{\mathrm{d}}{\mathrm{d}x} \frac{x}{1-x} \tag{5.314}$$

$$= \frac{x}{(1-x)^2}, \tag{5.315}$$

which satisfies (5.308)–(5.311).

For $k = 2$ we obtain using the result for $k = 1$

$$\sum_{n=1}^{\infty} n^2 x^n = x \frac{\mathrm{d}}{\mathrm{d}x} \sum_{n=1}^{\infty} n x^n \tag{5.316}$$

$$= x \frac{\mathrm{d}}{\mathrm{d}x} \frac{x}{(1-x)^2} \tag{5.317}$$

$$= \frac{x + x^2}{(1-x)^3} \tag{5.318}$$

which also satisfies (5.308)–(5.311).

Finally, for arbitrary k we can write

$$\sum_{n=1}^{\infty} n^k x^n = x\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n=1}^{\infty} n^{k-1}x^n \tag{5.319}$$

$$= x\frac{\mathrm{d}}{\mathrm{d}x}\left[(1-x)^{-k}\sum_{i=1}^{k-1} c_i^{(k-1)}x^i\right] \tag{5.320}$$

$$= x(1-x)^{-2k}\left[(1-x)^k\sum_{i=1}^{k-1} c_i^{(k-1)}ix^{i-1} + (1-x)^{k-1}\sum_{i=1}^{k-1} c_i^{(k-1)}x^i k\right] \tag{5.321}$$

$$= x(1-x)^{-(k+1)}\left[\sum_{i=1}^{k-1} c_i^{(k-1)}ix^{i-1} - \sum_{i=1}^{k-1} c_i^{(k-1)}ix^i + \sum_{i=1}^{k-1} c_i^{(k-1)}kx^i\right] \tag{5.322}$$

$$= (1-x)^{-(k+1)}\left[xc_1^{(k-1)} + x^k c_{k-1}^{(k-1)} + \sum_{i=2}^{k-1}\left(ic_i^{(k-1)} + (k-i+1)c_{i-1}^{(k-1)}\right)x^i\right]. \tag{5.323}$$

A comparison with (5.308)–(5.311) shows that (5.308)–(5.311) hold for any k if (5.308)–(5.311) hold for $k-1$ as assumed in (5.320). This concludes our proof.

5.B Error–Probability Result for Multichannel Signaling

In multichannel communication systems the metric difference between a pair of transmit symbols can often be expressed as a special case of the quadratic form

$$\Delta \triangleq \sum_{i=1}^{N_\Theta} \mathrm{Re}\left\{\Theta_i \sum_{j=1}^{l_i} x_{i,j}y_{i,j}^*\right\}, \tag{5.324}$$

where $\{x_{i,j}, y_{i,j}\}$, $1 \le j \le l_i$, $1 \le i \le N_\Theta$, are iid pairs of correlated zero–mean complex Gaussian random variables with

$$\mathcal{E}\left\{\begin{bmatrix} x_{i,j} \\ y_{i,j}\end{bmatrix} \cdot \begin{bmatrix} x_{i,j}^* & y_{i,j}^*\end{bmatrix}\right\} \triangleq \begin{bmatrix} \sigma_x^2 & \mu_{xy} \\ \mu_{xy}^* & \sigma_y^2\end{bmatrix} \tag{5.325}$$

and Θ_i, $1 \le i \le N_\Theta$, are *distinct* constants usually related in some way to the difference between two elements of the signal constellation (see below). The l_i, $1 \le i \le N_\Theta$, may or may not be equal.

Theorem: Given a random variable Δ as defined in (5.324) the probability that $\Delta < 0$ can be given in closed form as

$$\Pr(\Delta<0) = \sum_{i=1}^{N_\Theta}(-1)^{\sum_{\substack{j=1\\ j\neq i}}^{N_\Theta} l_j} \sum_{k_{1,0}=0}^{l_i-1}\sum_{k_{1,1}=0}^{k_{1,0}} \cdots \sum_{k_{1,N_\Theta}=0}^{k_{1,N_\Theta-1}}\sum_{k_{2,1}=0}^{k_{2,0}} \cdots \sum_{k_{2,i-1}=0}^{k_{1,i-2}}\sum_{k_{2,i+1}=0}^{k_{1,i-1}} \cdots \sum_{k_{2,N_\Theta-1}=0}^{k_{1,N_\Theta-2}} \tag{5.326}$$

$$v_{2,i}^{k_{1,0}} \frac{\binom{l_i+k_{1,i-1}-k_{1,i}-1}{l_i-1} v_{1,i}^{l_i}}{(v_{2,i}+v_{1,i})^{l_i+k_{1,i-1}-k_{1,i}}} \prod_{\substack{j=1\\ j\neq i}}^{N_\Theta} \frac{\binom{l_j+k_{1,j-1}-k_{1,j}-1}{l_j-1}\binom{l_j+k_{2,j-1}-k_{2,j}-1}{l_j-1}(v_{1,j}v_{2,j})^{l_j}}{(v_{2,i}+v_{1,j})^{l_j+k_{1,j-1}-k_{1,j}}(v_{2,i}-v_{2,j})^{l_j+k_{2,j-1}-k_{2,j}}},$$

where

$$v_{1,i} \triangleq \sqrt{w_i^2 + \frac{1}{|\Theta_i|^2(\sigma_x^2\sigma_y^2 - |\mu_{xy}|^2)}} - w_i \tag{5.327}$$

$$v_{2,i} \triangleq \sqrt{w_i^2 + \frac{1}{|\Theta_i|^2(\sigma_x^2\sigma_y^2 - |\mu_{xy}|^2)}} + w_i \tag{5.328}$$

$$w_i \triangleq \frac{\mathrm{Re}\left\{\mu_{xy}\Theta_i^*\right\}}{|\Theta_i|^2(\sigma_x^2\sigma_y^2 - |\mu_{xy}|^2)} \tag{5.329}$$

and $k_{2,N_\Theta} \equiv 0$, $k_{2,0} \equiv k_{1,N_\Theta}$, $k_{2,i} \equiv k_{2,i-1}$.

Proof: The probability $\Pr(\Delta < 0)$ can be computed via (cf. e.g. [Pro00])

$$\Pr(\Delta < 0) = -\sum_{i=1}^{N_\Theta} \operatorname*{Res}_{v=\mathrm{j}v_i^+} \left\{ \frac{\Phi_\Delta(v)}{v} \right\} \tag{5.330}$$

where $\Phi_\Delta(v) \triangleq \mathcal{E}\left\{\mathrm{e}^{\mathrm{j}v\Delta}\right\}$ and $\mathrm{j}v_i^+$, $1 \le i \le N_\Theta$, denote the characteristic function of Δ and the poles of $\Phi_\Delta(v)$ on the positive imaginary axis of the complex v–plane, respectively. As Δ is a sum of iid contributions $\delta_{i,j} \triangleq \mathrm{Re}\left\{\Theta_i x_{i,j} y_{i,j}^*\right\}$, $1 \le i \le N_\Theta$, $1 \le j \le l_i$, in pairs $\{x_{i,j}, y_{i,j}\}$ of correlated $\mathcal{N}_\mathrm{c}(0, \sigma_x^2)$ and $\mathcal{N}_\mathrm{c}\big(0, \sigma_y^2\big)$ with $\mu_{xy} \triangleq \mathcal{E}\left\{x_{i,j} y_{i,j}^*\right\}$, we can write

$$\Phi_\Delta(v) = \prod_{i=1}^{N_\Theta} \frac{(v_{1,i}v_{2,i})^{l_i}}{(v + \mathrm{j}v_{1,i})^{l_i}(v - \mathrm{j}v_{2,i})^{l_i}}, \tag{5.331}$$

with $v_{1|2,i}$ and w_i as defined in (5.327)–(5.329), respectively. Note, that both $v_{1,i}$ and $v_{2,i}$ are positive regardless of $\mathrm{Re}\left\{\mu_{xy}\Theta_i^*\right\}$ as $|\mu_{xy}|^2 \le \sigma_x^2\sigma_y^2$. Hence, poles $\mathrm{j}v_{2,i}$ and $-\mathrm{j}v_{1,i}$ lie on the positive and negative imaginary axis, respectively, and the residues in (5.330) correspond to poles $\mathrm{j}v_i^+ = \mathrm{j}v_{2,i}$.

The proof essentially makes use of the following relations:

- First, if $\tilde{v}$ is a pole of multiplicity l of a function $f(v)$, then the residue corresponding to that pole can be computed via

$$\operatorname*{Res}_{v=\tilde{v}}\{f(v)\} = \lim_{v\to\tilde{v}} \frac{1}{\Gamma(l)} \left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^{l-1} \left[(v-\tilde{v})^l f(v)\right]. \tag{5.332}$$

- Second, the Leibniz rule also known as generalized product rule states that

$$\left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^x [f(v)g(v)] = \sum_{k=0}^{x} \binom{x}{k} \left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^{x-k} f(v) \left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^k g(v). \tag{5.333}$$

- Third, the following simple derivative is used repeatedly

$$\left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^x (v-\tilde{v})^{-l} = (-1)^x \langle l \rangle_x \, (v-\tilde{v})^{-l-x}, \tag{5.334}$$

where $\langle l \rangle_x \triangleq \Gamma(l+x)/\Gamma(l)$ denotes the so–called Pochhammer symbol.

Since $\Phi_\Delta(v)$ is a rational function in v the proof boils down to computing the residues via

$$\operatorname*{Res}_{v=\mathrm{j}v_{2,i}}\left\{\frac{\Phi_\Delta(v)}{v}\right\} = R(\mathrm{j}v_{2,i})\prod_{j=1}^{N_\Theta}(v_{1,j}v_{2,j})^{l_j}. \tag{5.335}$$

where

$$R(v) \triangleq \left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^{l_i-1}\left(\frac{1}{v}\,P(v)\right) \tag{5.336}$$

$$P(v) \triangleq \left[(v+\mathrm{j}v_{1,i})^{l_i}\prod_{\substack{j=1\\ j\neq i}}^{N_\Theta}(v+\mathrm{j}v_{1,j})^{l_j}(v-\mathrm{j}v_{2,j})^{l_j}\right]^{-1}. \tag{5.337}$$

Using the Leibniz rule (5.333) and (5.334) $R(v)$ can be written as

$$R(v) = \sum_{k_{1,0}=0}^{l_i-1}\binom{l_i-1}{k_{1,0}}\frac{(-1)^{l_i-1-k_{1,0}}\Gamma(l_i-k_{1,0})}{v^{l_i-k_{1,0}}}\left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^{k_{1,0}}P(v). \tag{5.338}$$

Assuming for the moment that $1 < i < N_\Theta$, we successively chop factors off $R(v)$ defining a sequence of functions

$$\begin{aligned}
Q^{(1,1)}(v) &\triangleq P(v)(v+\mathrm{j}v_{1,1})^{l_1}, && (5.339)\\
Q^{(1,2)}(v) &\triangleq Q^{(1,1)}(v)(v+\mathrm{j}v_{1,2})^{l_2}, && (5.340)\\
&\vdots\\
Q^{(1,N_\Theta)}(v) &\triangleq Q^{(1,N_\Theta-1)}(v)(v+\mathrm{j}v_{1,N_\Theta})^{l_{N_\Theta}}, && (5.341)\\
Q^{(2,1)}(v) &\triangleq Q^{(1,N_\Theta)}(v)(v-\mathrm{j}v_{2,1})^{l_1}, && (5.342)\\
Q^{(2,2)}(v) &\triangleq Q^{(2,1)}(v)(v-\mathrm{j}v_{2,2})^{l_2}, && (5.343)\\
&\vdots\\
Q^{(2,i-1)}(v) &\triangleq Q^{(2,i-2)}(v)(v-\mathrm{j}v_{2,i-1})^{l_{i-1}}, && (5.344)\\
Q^{(2,i+1)}(v) &\triangleq Q^{(2,i-1)}(v)(v-\mathrm{j}v_{2,i+1})^{l_{i+1}}, && (5.345)\\
Q^{(2,i+2)}(v) &\triangleq Q^{(2,i+1)}(v)(v-\mathrm{j}v_{2,i+1})^{l_{i+1}}, && (5.346)\\
&\vdots\\
Q^{(2,N_\Theta-1)}(v) &\triangleq Q^{(2,N_\Theta-2)}(v)(v-\mathrm{j}v_{2,N_\Theta-1})^{l_{N_\Theta-1}} = \frac{1}{(v-\mathrm{j}v_{2,N_\Theta})^{l_{N_\Theta}}}, && (5.347)
\end{aligned}$$

and repeatedly apply (5.333) and (5.334) to compute e.g.

$$\begin{aligned}
\left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^{k_{1,0}}P(v) &= \left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^{k_{1,0}}\left(\frac{1}{(v+\mathrm{j}v_{1,1})^{l_1}}Q^{(1,1)}(v)\right) && (5.348)\\
&= \sum_{k_{1,1}=0}^{k_{1,0}}\binom{k_{1,0}}{k_{1,1}}\frac{(-1)^{k_{1,0}-k_{1,1}}\langle l_1\rangle_{k_{1,0}-k_{1,1}}}{(v+\mathrm{j}v_{1,1})^{l_1+k_{1,0}-k_{1,1}}}\left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^{k_{1,1}}Q^{(1,1)}(v) && (5.349)
\end{aligned}$$

as in (5.338). Having done this for $Q^{(1,1)}(v)$ through $Q^{(2,N_\Theta-1)}(v)$ we eventually obtain

$$R(v) = (-1)^{l_i-1} \sum_{k_{1,0}=0}^{l_i-1} \sum_{k_{1,1}=0}^{k_{1,0}} \cdots \sum_{k_{1,N_\Theta}=0}^{k_{1,N_\Theta-1}} \sum_{k_{2,1}=0}^{k_{1,N_\Theta}} \sum_{k_{2,2}=0}^{k_{2,1}} \cdots \sum_{k_{2,i-1}=0}^{k_{2,i-2}} \sum_{k_{2,i+1}=0}^{k_{2,i-1}} \cdots \sum_{k_{2,N_\Theta-1}=0}^{k_{2,N_\Theta-2}} \tag{5.350}$$
$$\frac{\frac{\Gamma(l_i-k_{1,0})}{\Gamma(l_i)}\binom{l_i-1}{k_{1,0}}}{v^{l_i-k_{1,0}}} \prod_{j=1}^{N_\Theta} \frac{\binom{k_{1,j-1}}{k_{1,j}}\langle l_j\rangle_{k_{1,j-1}-k_{1,j}}}{(v+\mathrm{j}v_{1,j})^{l_j+k_{1,j-1}-k_{1,j}}} \prod_{\substack{j=1\\ j\neq i}}^{N_\Theta-1} \frac{\binom{k_{2,j-1}}{k_{2,j}}\langle l_j\rangle_{k_{2,j-1}-k_{2,j}}}{(v-\mathrm{j}v_{2,j})^{l_j+k_{2,j-1}-k_{2,j}}} \langle l_{N_\Theta}\rangle_{k_{2,N_\Theta-1}}$$

where we introduced indeces $k_{2,0} \equiv k_{1,N_\Theta}$ and $k_{2,i} \equiv k_{2,i-1}$ to simplify the notation. Combining the various Γ–functions, binomial coefficients and Pochhammer symbols, introducing $k_{2,N_\Theta} \equiv 0$ for notational brevity, and plugging the results into (5.335) and (5.330) leads to (5.326).

The proof that (5.326) also holds for $i \in \{1, N_\Theta\}$ follows in complete analogy to the above proof for $1 < i < N_\Theta$ and is omitted for brevity. ■

We would like to point out that this general result extends upon related ones that can be found in the literature, e.g. [SA01, Pro00, SA98], as the coefficients Θ_i, $1 \leq i \leq N_\Theta$, are distinct and may also have distinct "multiplicities" l_i. It is therefore possible to use (5.326) to e.g. analyze the performance of DSTM, whereas the results in [SA01, Pro00, SA98] are only applicable to DPSK.

5.C Block Differential Space–Time Modulation

In [DB06] Du et al. proposed to permute the individual PSK elements of blocks of *cyclic* DSTM transmit symbols $\boldsymbol{S}[k]$ prior to transmission in order to avoid the increase in effective fading bandwidth incured by regular DSTM.

More specifically, at the transmitter side, DSTM transmit symbols are grouped in blocks of L^{BDSTM} symbols $\boldsymbol{S}[k+\kappa]$, $0 \leq \kappa \leq L^{\mathrm{BDSTM}}-1$. Instead of transmitting $\boldsymbol{S}[k]$ followed by $\boldsymbol{S}[k+1]$ and so on in a row–by–row fashion, i.e. due to the diagonal structure of $\boldsymbol{S}[k]$ using alternating transmit antennas, they suggested to first transmit $s_{1,1}[k+\kappa]$ from antenna 1 during modulation intervals $kN_\mathrm{T}+\kappa+1$, $0 \leq \kappa \leq L^{\mathrm{BDSTM}}-1$, followed by $s_{2,2}[k+\kappa]$ from antenna 2 during modulation intervals $kN_\mathrm{T}+\kappa+1+L^{\mathrm{BDSTM}}$, $0 \leq \kappa \leq L^{\mathrm{BDSTM}}-1$, and so on. After transmission of $s_{N_\mathrm{T},N_\mathrm{T}}[k+\kappa]$ from antenna N_T during modulation intervals $kN_\mathrm{T}+\kappa+1+L^{\mathrm{BDSTM}}(N_\mathrm{T}-1)$, $0 \leq \kappa \leq L^{\mathrm{BDSTM}}-1$, the next block of L^{BDSTM} DSTM symbols is processed in the same fashion.

This is advantageous in two ways:

- The artificial channel decorrelation incured by regular DSTM is avoided, as symbols $s_{i,i}[k]$ and $s_{i,i}[k+1]$ that are successive in the direction of differential encoding are transmitted from the same antenna in two successive modulation intervals.

- In the case of spatially correlated channels BDSTM leads to a decorrelation of subchannels, over which the different elements of the same $\boldsymbol{S}[k]$ are transmitted, due to the temporal fading. Provided that L^{BDSTM} is chosen sufficiently large, full transmit diversity order N_{T} can still be achieved even in a spatially fully correlated channel.

Chapter 6

Multiple–Symbol Differential Detection for Frequency–Selective Channels

Orthogonal frequency division multiplexing (OFDM) is a popular modulation technique for transmission over frequency–selective channels and has been adopted in several standards such as, e.g. for digital audio and video broadcasting, wireless local and metropolitan area networks, and ultra–wideband radio. OFDM decomposes a both time– and frequency–selective channel into a number of parallel time–selective but frequency–nonselective channels, which facilitates low–complexity channel equalization, cf. e.g. [WE71, Bin90, BKS92]. The combination of OFDM with antenna arrays at both transmitter and receiver, which is usually referred to as MIMO–OFDM, has recently received considerable attention as a means to improve bandwidth and/or power efficiency, cf. e.g. [SBM$^+$04] for an overview.

Especially in strongly frequency–selective channels and mobile environments with relatively fast changes of channel conditions, the problem of acquiring accurate channel state information is aggravated compared to the frequency–nonselective case considered in Chapter 5. In such scenarios, differential encoding and detection without the need for CSI becomes an even more attractive alternative.

In this chapter, we therefore consider the application of the computationally efficient implementations of MSDD and its variants presented in Chapters 2 and 3 to MIMO–OFDM. In addition, we propose a new signal–allocation scheme for MIMO–OFDM based on cyclic DSTM constellations, that leads to improved power efficiencies of the noncoherent detectors compared to existing schemes. A detailed outline of this chapter is given in the next section following the discussion of related literature.

6.1 Related Literature and Chapter Outline

Differential encoding in frequency direction and CDD, which relies on negligible variations of the channel frequency response at adjacent subchannels, is considered in e.g. [ER93]. However, as for the frequency–nonselective channel CDD suffers from a high error floor already in moderately frequency–selective channels. Luise et al. [LRV98] devised an improved detector, which overcomes the error–floor problem by exploiting the subchannel frequency correlation and using the Viterbi algorithm with predictor–based branch metrics. Subsequently, we refer to detection algorithms that exploit correlation only in time or frequency direction as "1–dimensional" (1D) detection schemes. Since, in general, OFDM subchannels are correlated in both time and frequency direction, a number of 2–dimensional (2D) detection schemes exploiting time and frequency correlation have also been proposed in the literature. In [FS99], the authors present a coherent detection scheme for differential encoding in time using a decision–aided 2D channel estimation algorithm, i.e. no pilots are needed. Haas et al. [HK03] proposed to take "detours" in the 2D neighborhood (in time and frequency) of an actual differential encoding step such that phase errors along individual steps are minimized. This approach was shown to yield moderate gains over CDD for differential encoding in time direction. Chang and Su [CS04] proposed a 2D regression model for the channel and applied a branch–and–bound technique to perform 1D data detection without CSI in time direction for a subset of subchannels and a subsequent decision–aided channel estimation with 2D interpolation. Cui et al. [CT04b] investigated both decision–feedback and branch–and–bound data detection methods exploiting subchannel correlations in frequency direction only.

All the mentioned work considered transmission with only a single transmit antenna. For MIMO–OFDM, which employs antenna arrays at the transmitter, the concept of space–frequency or space–time–frequency coding has attracted considerable attention in the literature, cf. e.g. [LXG02, BBP03, SSL05]. Space–frequency codes for noncoherent detection have been designed and analyzed by Borgmann et al. [BB05]. While utilizing full space and frequency (multipath) diversity of the frequency–selective MIMO channel, detection complexity is exponential in the number of OFDM subchannels for fixed rate. Another approach pursued in e.g. [LS01, DADSC02, WY02, LHC04, SL05, CLL05, HSL05c] is to employ differential modulation with matrix constellations developed for differential space–time modulation (DSTM) and single–carrier MIMO transmission (e.g. [TJ00, Hug00a, HS00]). In [LS01, WY02, LHC04, SL05, HSL05c], diagonal constellations ([Hug00a, HS00]) are considered and, similar to space–frequency codes for coherent detection with CSI, signal components of the DSTM code matrices are allocated to different transmit antennas and OFDM subchannels to accomplish diversity gains. This is referred to as differential modulation diversity (DMD) in [LS01], which also

introduced DFDD into OFDM to avoid the error floor of CDD.[1] Signal allocation can be further optimized for the case of known channel power–delay profile (PDP) at the transmitter as discussed in [LHC04, SL05, HSL05c] (see also [SSL05]). Whereas diversity and optimal coding advantage are considered in [LHC04, HSL05c], performance improvement of CDD is achieved in [SL05] by creating a smooth logical channel for differential encoding in frequency direction. A comparison of optimal and (pseudo–)random signal allocation can be found in [HSL05c]. Extending the work in [DADSC02], Chiang et al. [CLL05] studied orthogonal constellations ([TJ00]) in conjunction with CDD, DFDD, and prediction–based Viterbi decoding for MIMO–OFDM. New matrix constellations tailored for differential MIMO–OFDM under the assumption of time–invariant channels were designed by Ma et al. [MTL05] and by Li [Li03].

In this chapter, we study MIMO–OFDM transmission with differential encoding and we consider the general case in which differential encoding can be performed in time or in frequency direction. To exploit frequency diversity while minimizing transmission delay and detrimental effects of temporal channel variations on performance, all components of a matrix signal are allocated within *one* OFDM symbol, which is similar to [LS01, LHC04, SL05, HSL05c] and unlike the allocation in [WY02, MTL05]. We refer to the resulting scheme as differential space–frequency modulation (DSFM), a term coined by Su and Liu [SL05] in analogy to DSTM. As [LS01, WY02, LHC04, SL05, HSL05c], we consider diagonal constellations. Cyclic (diagonal) constellations are particularly suited for DSFM, since (i) no minimal channel coherence bandwidth is required, which would be the case for non–diagonal constellation (cf. e.g. [CLL05]), and (ii) spatial and spectral channels can be traded freely and thus, different numbers of transmit antennas can be accommodated with the same constellation, without changing bandwidth efficiency.

In section 6.4, we devise a novel signal–allocation scheme for DSFM, which unifies DMD of [LS01] and ideas from BDSTM proposed for frequency–flat MIMO fading channels in [LWZD04] (cf. also Appendix 5.C) and which is applicable for differential encoding in either time or frequency. Under the reasonable assumption that the PDP is *not* known at the transmitter, it optimally combines the goals of (i) minimally correlated subchannels used for different components of one space–frequency matrix signal, (ii) maximally correlated subchannels experienced by matrix signals successive in differential encoding, and (iii) maximal correlation between subchannels comprised by the 2D observation window (see below) for detection without CSI. Thereby, *full* space–frequency diversity and *optimal* coding gain are not our primary concern, as (i) this usually requires some knowledge of the PDP at the transmitter, (ii) signal constellations may become extremely large, which increases detection complexity, and (iii) asymptotic performance gains through increasing diversity beyond a certain point are not visible for practically

[1] The assignment of signal components to subcarriers is also often referred to as subcarrier grouping [LXG01].

relevant error rates.

Different from the aforementioned literature, we consider ML MSDD for DSFM. Inspired by 2D channel estimation schemes in e.g. [HKR97, LCS98, CS02, SBM+04], we propose ML MSDD with a 2D "observation window", which, different from previous work for noncoherent MIMO–OFDM, is capable of exploiting channel correlation in both time and frequency direction. We develop a representation of the detection problem that is amenable to the tree–search methods developed in Chapter 3. This is done in Section 6.5.

Performance and complexity of the proposed transmission scheme are analyzed by means of analytical methods and numerical examples in Sections 6.6 and 6.7, respectively.

6.2 Time– and Frequency–Selective Channel Model

In the case of a time and frequency–selective channel, the continuous–time input delay–spread function describing the spatial subchannel between transmit antenna i and receive antenna j is given by

$$\tilde{h}_{i,j}(t,\tau) = \sum_{l=1}^{L_h} h_{i,j}^{(l)}(t)\delta(\tau-\tau_l) \tag{6.1}$$

where the L_h propagation paths are assumed to be iid and specified by their complex amplitudes $h_{i,j}^{(l)}(t)$ and delays τ_l, $1 \le l \le L_h$, cf. Chapter 4. As shown in Chapter 4 the discrete–time received signal at antenna j and discrete time kT reads

$$r_j[k'] = \sum_{i=1}^{N_\mathrm{T}} \sum_{k=0}^{\infty} x_i[k] \sum_{l=1}^{L_h} h_{i,j}^{(l)}(kT+\tau_1) h_\mathrm{G}((k'-k)T+\tau_1-\tau_l) + n_j[k'], \tag{6.2}$$

where $h_\mathrm{G}(t)$ combines the effects of transmit pulse shaping and receiver matched filtering [cf. (4.15)] and $n_j[k']$ represents AWGN.

6.3 OFDM Transmission

For a brief recapitulation of MIMO–OFDM we consider Fig. 6.1, which illustrates the MIMO–OFDM transmission system with N_T transmit and N_R receive antennas in the ECB domain. In a way that is to be specified in Section 6.4 information is mapped to symbols $x_i^\mathrm{f}[m,k]$, $1 \le i \le N_\mathrm{T}$, $1 \le m \le D$, $k \in \mathbb{N}$. Following an inverse discrete Fourier transform (IDFT) of blocks of D symbols $x_i^\mathrm{f}[m,k]$, $1 \le m \le D$, resulting in blocks of D symbols $x_i^\mathrm{p}[m,k]$, $1 \le m \le D$, the D_g symbols $x_i^\mathrm{p}[m,k]$, $D-D_\mathrm{g}+1 \le m \le D$, are added as so–called "cyclic prefix" prior to parallel–to–serial (P/S) conversion to symbols $x_i[k']$. After pulse shaping with $h_\mathrm{T}(t)$ these symbols are transmitted successively at rate $1/T$ over the frequency–selective subchannel $\tilde{h}_{i,j}(t,\tau)$ to be re-

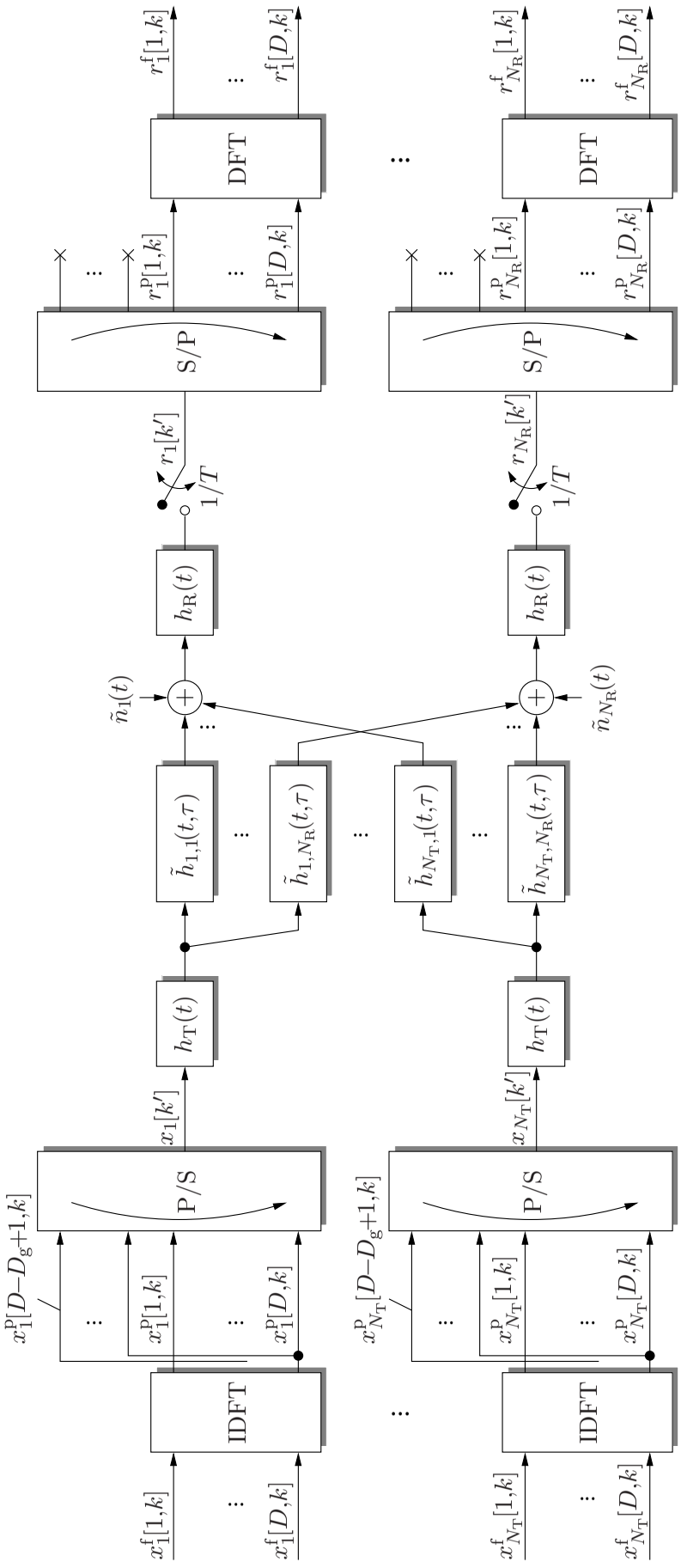

Figure 6.1: MIMO–OFDM transmission system.

ceived by the jth receive antenna. Following matched filtering with $h_{\mathrm{R}}(t)$ [cf. (4.2)] and sampling at rate $1/T$ the resulting discrete time signal $r_j[k']$ is serial–to–parallel (S/P) converted and the received symbols corresponding to the cyclic prefix are discarded leaving blocks of D received symbols $r_j^{\mathrm{p}}[m,k]$, $1 \le m \le D$. These are fed into a discrete Fourier transform (DFT) yielding symbols $r_j^{\mathrm{f}}[m,k]$, $1 \le m \le D$.

In order to clearly show the inner workings of OFDM, we progress through Fig. 6.1 from the center to its edges building upon (6.2). Based on the cyclic prefix insertion and the P/S conversion at the transmitter such that $x_i[(k-1)(D+D_{\mathrm{g}})+m] = x_i^{\mathrm{p}}[m,k]$ the received signal $r_j[k']$ reads

$$\begin{aligned} r_j[k'] &= \sum_{i=1}^{N_{\mathrm{T}}} \sum_{\tilde{k}=0}^{\infty} \sum_{\tilde{m}=-D_{\mathrm{g}}+1}^{D} x_i^{\mathrm{p}}[\mathrm{mod}(\tilde{m},D),\tilde{k}] \sum_{l=1}^{L_h} h_{i,j}^{(l)}(k'T+\tau_1) \\ &\quad h_{\mathrm{G}}(k'T+\tau_1-\tau_l-((\tilde{k}-1)(D+D_{\mathrm{g}})+\tilde{m}+D_{\mathrm{g}}-1)T)+n_j[k']. \end{aligned} \tag{6.3}$$

With the S/P conversion and the cyclic prefix removal at the receiver and $T_{\mathrm{f}} = (D+D_{\mathrm{g}})T$ we obtain

$$\begin{aligned} r_j^{\mathrm{p}}[m,k] &= r_j[(k-1)(D+D_{\mathrm{g}})+m+D_{\mathrm{g}}-1] \qquad (6.4) \\ &= \sum_{i=1}^{N_{\mathrm{T}}} \sum_{\tilde{k}=0}^{\infty} \sum_{\tilde{m}=-D_{\mathrm{g}}+1}^{D} x_i^{\mathrm{p}}[\mathrm{mod}(\tilde{m},D),\tilde{k}] \sum_{l=1}^{L_h} h_{i,j}^{(l)}((k-1)T_{\mathrm{f}}+(m+D_{\mathrm{g}}-1)T+\tau_1) \\ &\quad h_{\mathrm{G}}((k-\tilde{k})T_{\mathrm{f}}+(m-\tilde{m})T+\tau_1-\tau_l)+n_j[m,k], \qquad (6.5) \end{aligned}$$

where $n_j[m,k] \triangleq n_j[(k-1)(D+D_{\mathrm{g}})+m+D_{\mathrm{g}}-1]$. If the cyclic prefix is chosen properly such that $t_{h_{\mathrm{G}}}+\tau_{L_h}-\tau_1 < D_{\mathrm{g}}T$, where $t_{h_{\mathrm{G}}}$ denotes the period of time, where $h_{\mathrm{G}}(t)$ deviates noticeably form zero, blocks $[x_i^{\mathrm{p}}[1,k],\ldots,x_i^{\mathrm{p}}[D,k]]$ and $[x_i^{\mathrm{p}}[1,k+1],\ldots,x_i^{\mathrm{p}}[D,k+1]]$ can be avoided effectively. Consequently, we can discard the sum over $\tilde{k}$ from (6.5) fixing $\tilde{k}=k$ which yields

$$\begin{aligned} r_j^{\mathrm{p}}[m,k] &= \sum_{i=1}^{N_{\mathrm{T}}} \sum_{\tilde{m}=-D_{\mathrm{g}}+1}^{D} x_i^{\mathrm{p}}[\mathrm{mod}(\tilde{m},D),k] \\ &\quad \sum_{l=1}^{L_h} h_{i,j}^{(l)}((k-1)T_{\mathrm{f}}+(m+D_{\mathrm{g}}-1)T+\tau_1) h_{\mathrm{G}}((m-\tilde{m})T+\tau_1-\tau_l)+n_j[m,k], \end{aligned} \tag{6.6}$$

i.e. blocks of D received symbols $r_j^{\mathrm{p}}[m,k]$, $1 \le m \le D$, can be processed independently at the receiver.

For OFDM to function optimally it is required that the channel does not change significantly during the transmission of an OFDM frame, i.e. $\tilde{h}_{i,j}(t,\tau) \approx \tilde{h}_{i,j}(t+\Delta t,\tau)$, $0 \le \Delta t \le T_{\mathrm{f}}$, $T_{\mathrm{f}} \triangleq (D+D_{\mathrm{g}})T$. In this case we can apply the simplification

$$h_{i,j}^{(l)}((k-1)T_{\mathrm{f}}+(m+D_{\mathrm{g}}-1)T+\tau_1) = h_{i,j}^{(l)}(kT_{\mathrm{f}}+\tau_1) \tag{6.7}$$

in (6.6) and see that the influence of the dispersive channel on the transmit signal becomes that of a *cyclic* convolution, i.e. we can write[2]

$$\begin{bmatrix} r_j^{\mathrm{p}}[1,k] \\ \vdots \\ r_j^{\mathrm{p}}[D,k] \end{bmatrix} = \sum_{i=1}^{N_{\mathrm{T}}} \boldsymbol{H}_{i,j}^{\mathrm{circ}} \begin{bmatrix} x_i^{\mathrm{p}}[1,k] \\ \vdots \\ x_i^{\mathrm{p}}[D,k] \end{bmatrix} + \begin{bmatrix} n_j[1,k] \\ \vdots \\ n_j[D,k] \end{bmatrix}, \tag{6.8}$$

with the circulant matrices

$$\boldsymbol{H}_{i,j}^{\mathrm{circ}} \triangleq \left[\sum_{l=1}^{L_h} h_{i,j}^{(l)}(kT_{\mathrm{f}}+\tau_1) h_{\mathrm{G}}(\mathrm{mod}(m-\tilde{m},D)\,T+\tau_1-\tau_l) \right]_{\substack{m=1,\ldots,D \\ \tilde{m}=1,\ldots,D}}. \tag{6.9}$$

As the DFT matrix $\boldsymbol{D} \triangleq D^{-1/2}\left[\exp\left(\mathrm{j}\frac{2\pi}{D}mn\right)\right]_{\substack{m=1,\ldots,D \\ n=1,\ldots,D}}$ is the modal matrix of any circulant matrix, cf. e.g. [Gra71], we obtain

$$\begin{bmatrix} r_j^{\mathrm{f}}[1,k] \\ \vdots \\ r_j^{\mathrm{f}}[D,k] \end{bmatrix} = \sum_{i=1}^{N_{\mathrm{T}}} \boldsymbol{D}\boldsymbol{H}_{i,j}^{\mathrm{circ}}\boldsymbol{D}^{\mathsf{H}} \begin{bmatrix} x_i^{\mathrm{f}}[1,k] \\ \vdots \\ x_i^{\mathrm{f}}[D,k] \end{bmatrix} + \boldsymbol{D} \begin{bmatrix} n_j[1,k] \\ \vdots \\ n_j[D,k] \end{bmatrix} \tag{6.10}$$

$$= \sum_{i=1}^{N_{\mathrm{T}}} \begin{bmatrix} \lambda_{\boldsymbol{H}_{i,j}^{\mathrm{circ}},1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \lambda_{\boldsymbol{H}_{i,j}^{\mathrm{circ}},D} \end{bmatrix} \begin{bmatrix} x_i^{\mathrm{f}}[1,k] \\ \vdots \\ x_i^{\mathrm{f}}[D,k] \end{bmatrix} + \begin{bmatrix} n_j'[1,k] \\ \vdots \\ n_j'[D,k] \end{bmatrix}, \tag{6.11}$$

where the $n_j'[m,k]$ have the same distribution as the $n_j[m,k]$ as $\boldsymbol{D}$ is unitary, and the eigenvalues $\lambda_{\boldsymbol{H}_{i,j}^{\mathrm{circ}},m}$, $1 \le m \le D$, are equal to the DFT of the sequence

$$\sum_{l=1}^{L_h} h_{i,j}^{(l)}(kT_{\mathrm{f}}+\tau_1) h_{\mathrm{G}}(mT+\tau_1-\tau_l) = \tilde{h}_{i,j}(kT_{\mathrm{f}}+\tau_1, t+\tau_1) * h_{\mathrm{G}}(t)\Big|_{t=mT}, \quad 0 \le m \le D-1. \tag{6.12}$$

Since the frequency response $\left|\mathcal{F}\{h_{\mathrm{G}}(t)\}(f)\right|$ of $h_{\mathrm{G}}(t)$ is practically flat in the frequency range of interest, we can neglect it and obtain the eigenvalues $\lambda_{\boldsymbol{H}_{i,j}^{\mathrm{circ}},m}$, $1 \le m \le D$, as spectrum of the input delay–spread function $\tilde{h}_{i,j}(t,\tau)$ [cf. (4.7)] and write

$$r_j^{\mathrm{f}}[m,k] = \sum_{i=1}^{N_{\mathrm{T}}} g_{i,j}[m,k] x_i^{\mathrm{f}}[m,k] + n_j'[m,k], \tag{6.13}$$

[2]If the coefficients $h_{i,j}^{(l)}(t)$ change too rapidly, the modeling as cyclic convolution incurs a mismatch, which can be included in the equivalent channel model as an additional AWGN term, cf. [RS95]. However, for $B_h T_{\mathrm{f}} \le 0.01$ the variance of this additional noise term is sufficiently low to be negligible in transmission scenarios of interest and we shall restrict our attention to this case.

where

$$g_{i,j}[m,k] \triangleq c_{\mathrm{n}}\, \mathcal{F}\left\{\tilde{h}_{i,j}(kT_{\mathrm{f}}+\tau_1, t+\tau_1)\right\}(f)\Big|_{f=m\Delta f} = \sum_{l=1}^{L_h} c_{\mathrm{n}} h_{i,j}^{(l)}(kT_{\mathrm{f}}+\tau_1)\mathrm{e}^{-\mathrm{j}2\pi m\Delta f(\tau_l-\tau_1)},\ ^{3} \tag{6.14}$$

$\Delta f \triangleq 1/(DT)$ and the constant c_{n} contains all influence of $h_{\mathrm{G}}(t)$ compensating for channel attenuation such that $\mathcal{E}\{|g_{i,j}[m,k]|^2\} = 1$. In consequence, the overall discrete–time channel model is again non–amplifying and the average power of the lth normalized channel coefficient $c_{\mathrm{n}}h_{i,j}^{(l)}(t)$ is given by

$$\sigma_l^2 \triangleq \frac{\mathcal{E}_t\left\{\left|h_{i,j}^{(l)}(t)\right|^2\right\}}{\sum\limits_{\mu=1}^{L_h}\mathcal{E}_t\left\{\left|h_{i,j}^{(\mu)}(t)\right|^2\right\}}. \tag{6.15}$$

Apparently, the symbols $x_i^{\mathrm{f}}[m,k]$, $1 \le m \le D$, are transmitted over D *parallel* discrete–time *frequency–nonselective* $N_{\mathrm{T}} \times N_{\mathrm{R}}$ MIMO channels as illustrated in Fig. 6.2. As the channel coefficient of the mth effective (narrowband) frequency–nonselective channel is given by the spectrum of the (wideband) frequency–selective channel at frequency $f = m\Delta f$, OFDM is often referred to as "discrete multitone" (DMT) or "multicarrier system" and the D parallel subchannels as "subcarriers" spaced at frequency increments —or "intercarrier spacing"— Δf.

For our considerations in this chapter, (6.13) shall serve as channel model. Again, the correlation of the channels is of interest for what follows. It is shown in Appendix 6.A.1 that the autocorrelation of $g_{i,j}[m,k]$ is separable into a spectral, a temporal, and two spatial correlation functions

$$\begin{aligned}\psi_{gg}[\mu,\kappa,i_1,j_1,i_2,j_2] &\triangleq \mathcal{E}\{g_{i_1,j_1}[m+\mu,k+\kappa]g_{i_2,j_2}^*[m,k]\} &(6.16)\\ &= \psi^{\mathrm{f}}[\mu]\tilde{\psi}^{\mathrm{t}}[\kappa]\psi^{\mathrm{Tx}}[i_1,i_2]\psi^{\mathrm{Rx}}[j_1,j_2], &(6.17)\end{aligned}$$

where the spectral correlation is determined by the PDP of the frequency–selective channel as

$$\psi^{\mathrm{f}}[\mu] \triangleq \sum_{l=1}^{L_h}\sigma_l^2\mathrm{e}^{-\mathrm{j}2\pi\mu\Delta f(\tau_l-\tau_1)}\ , \tag{6.18}$$

and the temporal correlation is given by

$$\tilde{\psi}^{\mathrm{t}}[\kappa] \triangleq \psi^{\mathrm{t}}(\kappa T_{\mathrm{f}}). \tag{6.19}$$

with $\psi^{\mathrm{t}}(t)$ as defined in (4.14). The spatial correlation coefficients $\psi^{\mathrm{Tx}}[i,j]$ and $\psi^{\mathrm{Rx}}[i,j]$ are the elements in the ith row and jth column of $\mathbf{\Psi}^{\mathrm{Tx}}$ and $\mathbf{\Psi}^{\mathrm{Rx}}$ [cf. (4.9) and (4.10)], respectively.

[3]The Fourier transform is formed with respect to the second argument of $\tilde{h}_{i,j}(t,\tau)$.

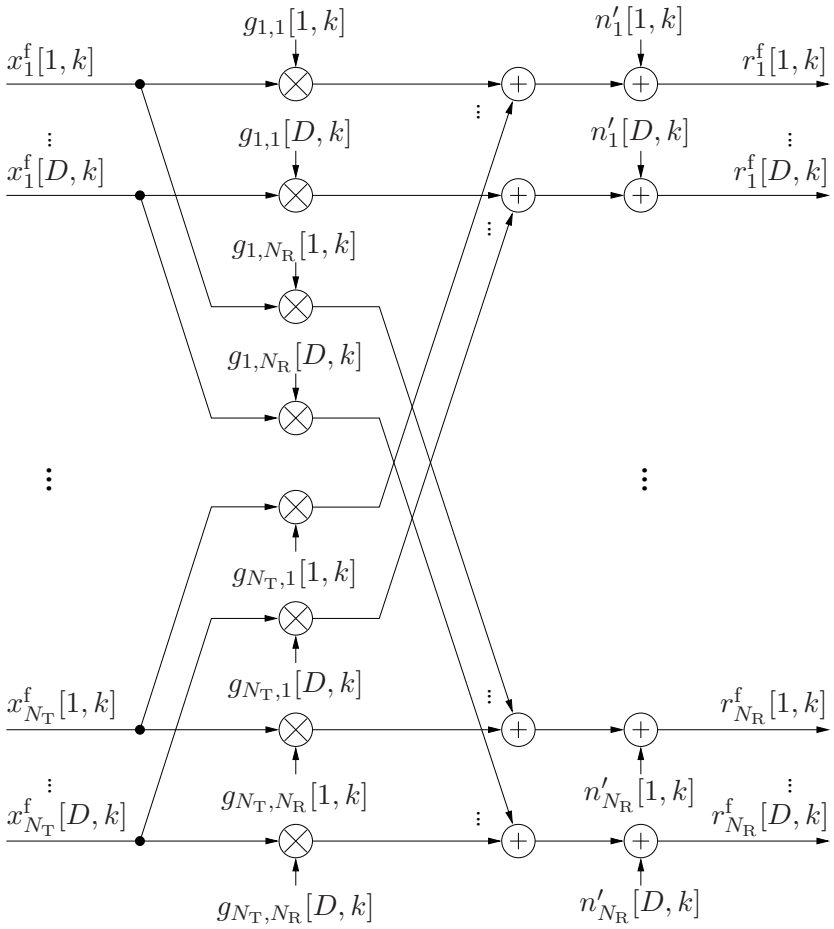

Figure 6.2: Discrete–time model of MIMO–OFDM transmission system.

6.4 Differential Space–Frequency Modulation (DSFM)

In this section, we describe DSFM to achieve spatial and spectral diversity and facilitate non-coherent detection (discussed in Section 6.5) based on cyclic DSTM constellations (cf. Section 2.1.1.1). To this end, we briefly present the two fundamental options for differential encoding in time and frequency direction in Section 6.4.1. The proposed signal–allocation (SA) scheme, i.e. the allocation of elements of the diagonal signal matrix to OFDM subchannels and transmit antennas, is developed in Section 6.4.2.

Here, we consider the general case, where cyclic DSTM codes of $(N_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional matrices with

$$N_{\mathrm{S}} = N_{\mathrm{T}} N_{\mathrm{B}}, \quad N_{\mathrm{B}} \in \mathbb{N} \tag{6.20}$$

are used for transmission from the N_{T} transmit antennas and allocate the individual components of each matrix symbol to different antennas and subcarriers such that we do not only exploit spatial diversity but also spectral diversity.

6.4.1 Differential Encoding

We denote diagonal data matrices chosen from $\boldsymbol{\mathcal{V}}$ [cf. (2.2)] by $\boldsymbol{V}[m,k]$, where m and k are discrete–frequency and –time indices as described in detail in Section 6.4.2. In DSFM, diagonal transmit matrices $\boldsymbol{S}[m,k]$ can be obtained either by differential encoding in time direction, i.e. time–differential SFM (T–DSFM), via

$$\boldsymbol{S}[m,k+1] = \boldsymbol{V}[m,k]\boldsymbol{S}[m,k], \quad \boldsymbol{S}[m,1] = \boldsymbol{I}_{N_\mathrm{S}} \; . \tag{6.21}$$

or by differential encoding in frequency direction, i.e. frequency–differential SFM (F–DSFM), via

$$\boldsymbol{S}[m+1,k] = \boldsymbol{V}[m,k]\boldsymbol{S}[m,k], \quad \boldsymbol{S}[1,k] = \boldsymbol{I}_{N_\mathrm{S}}. \tag{6.22}$$

F–DSFM is preferable for burst transmission or if detection delay is to be minimized. T–DSFM is advantageous for continuous transmission since (i) the share of reference symbols $\boldsymbol{S}[m,1]$ tends to zero with increasing transmission time and (ii) OFDM is usually employed for channels that exhibit significant frequency but moderate time selectivity.

6.4.2 Signal Allocation (SA) Scheme

There are several possibilities to use the three dimensions space, frequency, and time to transmit the two–dimensional matrices $\boldsymbol{S}[m,k]$. In general, frequency and time are interchangeable, but to minimize the transmission delay, space–frequency modulation, i.e. DSFM, is considered in this paper, and all elements of $\boldsymbol{S}[m,k]$ are allocated to subcarriers of the kth OFDM frame. Hence, assuming D active subcarriers and D/N_S being integer, the D/N_S symbols $[\boldsymbol{S}[1,k], \boldsymbol{S}[2,k], \ldots, \boldsymbol{S}[D/N_\mathrm{S},k]]$ are transmitted during the kth OFDM frame using N_T transmit antennas. In order to exploit *full spatial* diversity, the elements of each $\boldsymbol{S}[m,k]$ are transmitted from all N_T antennas over $N_\mathrm{B} = N_\mathrm{S}/N_\mathrm{T}$ subcarriers per antenna. To *maximize spectral* diversity under the constraint that the PDP is not known at the transmitter, these N_B subcarriers should be spread uniformly over the used frequency band, i.e. spaced by D/N_B subcarriers. This is similar to the subchannel grouping proposed for space–time–frequency coding in [LXG02] also adopted in [LS01, LHC04, MTL05, Li03].

Finally, differential encoding needs to be taken into account. In the case of T–DSFM (see (6.21)), the proposed scheme ensures that the time between receiving two consecutive transmit symbols $\boldsymbol{S}[m,k]$ and $\boldsymbol{S}[m,k+1]$ is the minimum of one OFDM frame. As long as $s_{i,i}[m,k]$ is assigned to the same OFDM subcarrier and antenna as $s_{i,i}[m,k+1]$, the "effective" fading bandwidth attains its minimum of $B_h T_\mathrm{f}$. In the case of F–DSFM [cf. (6.22)], the spectral correlation between the subcarriers allocated to $s_{i,i}[m,k]$ and $s_{i,i}[m+1,k]$ should be maximized

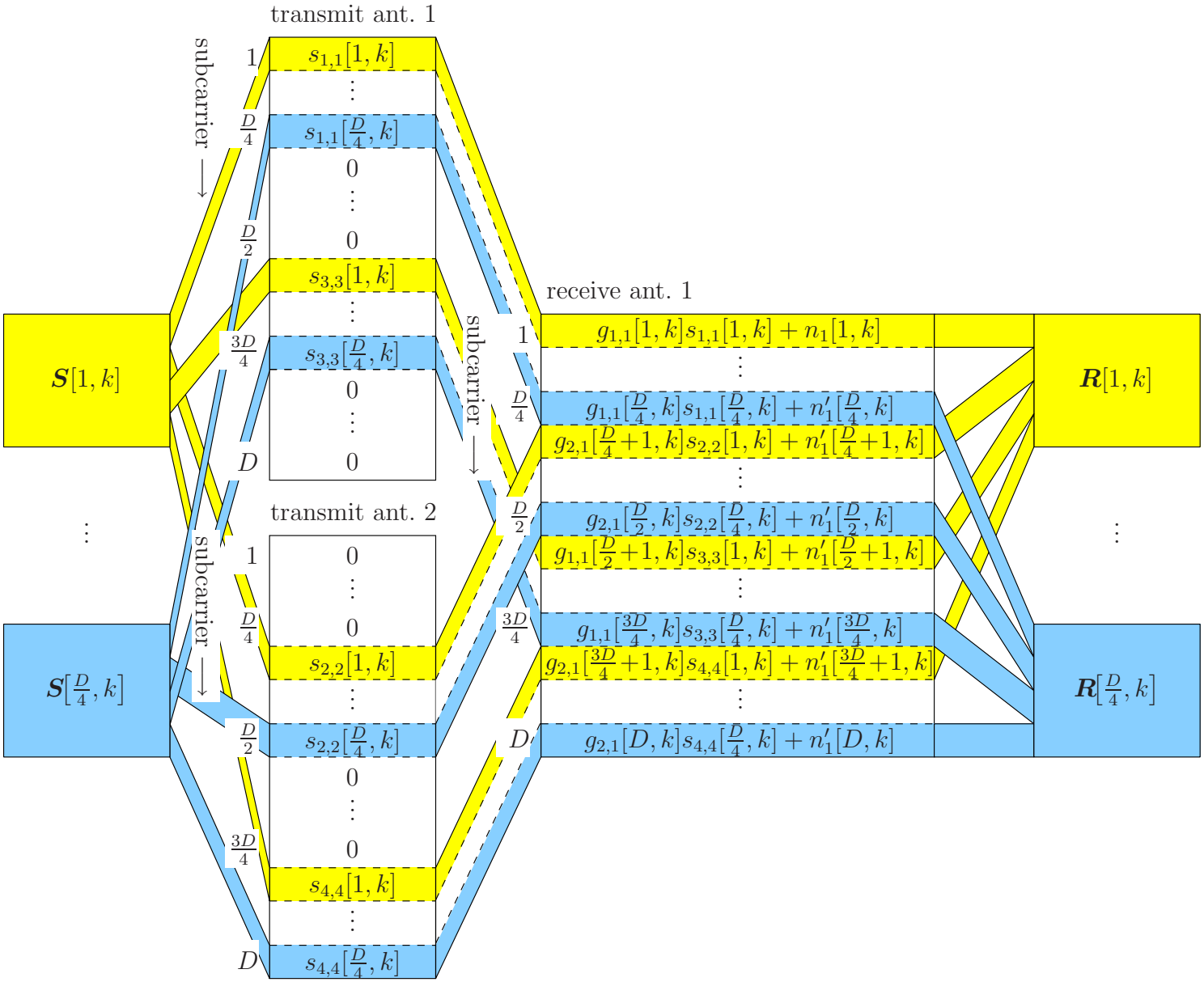

Figure 6.3: Illustration of transmission of one OFDM frame for example of $N_{\mathrm{T}} = 2$, $N_{\mathrm{B}} = 2$ and $N_{\mathrm{R}} = 1$.

for all $1 \leq i \leq N_{\mathrm{S}}$. As, different from [SL05], the channel PDP is assumed to be unknown at the transmitter, the most reasonable choice is to transmit $[s_{i,i}[1,k], \ldots, s_{i,i}[D/N_{\mathrm{S}},k]]$ over a set of D/N_{S} contiguous subcarriers. This can be regarded as an application of the idea underlying BDSTM (cf. Appendix 5.C) to DSFM.

In summary, according to the proposed SA scheme the diagonal elements $s_{i,i}[m,k]$, $1 \leq i \leq N_{\mathrm{S}}$, of $\boldsymbol{S}[m,k]$, $1 \leq m \leq D/N_{\mathrm{S}}$, $k \in \mathbb{N}$, are transmitted from antenna $[\mathrm{mod}(i-1,N_{\mathrm{T}})+1]$ over subcarrier $[(i-1)D/N_{\mathrm{S}}+m]$ of the kth OFDM frame, while all other antennas do not transmit over this subcarrier. In the notation of Section 6.3 we consequently have

$$x^{\mathrm{f}}_{\mathrm{mod}(i-1,N_{\mathrm{T}})+1}[(i-1)D/N_{\mathrm{S}}+m,k] = s_{i,i}[m,k]. \tag{6.23}$$

Fig. 6.3 illustrates the SA scheme for the example of $N_{\mathrm{T}}=2$ and $N_{\mathrm{B}}=2$, $N_{\mathrm{S}}=N_{\mathrm{T}}N_{\mathrm{B}}=4$, and $N_{\mathrm{R}}=1$.

Compared to related SA schemes based on diagonal–matrix signals this scheme is advantageous in that neither the effective fading bandwidth (DMD of [LS01]) nor the spectral spacing between subcarriers allocated to $s_{i,i}[m,k]$ and $s_{i,i}[m+1,k]$ ([WY02, HSL05b, MTL05]) are increased beyond their respective minima B_hT_{f} and 1.

We note that this SA scheme is used both for T– and F–DSFM and that DR and $(D-1)R$ bits per OFDM frame can be transmitted regardless of the parameter N_{B} with T–DSFM and F–DSFM, respectively.

6.5 MSDD with 2–Dimensional Observation Window

In MSDD, N received symbols are jointly processed. Usually, the observation window extends in the direction of differential encoding, i.e. time for single–carrier transmission as considered in Chapter 5. A straightforward adaption of MSDD to OFDM–based transmission is to apply a 1D observation window which extends either in time or in frequency direction depending on the direction of differential encoding. Such an approach has been chosen for DFDD and Viterbi detection in [LS01, CLL05].

Inspired by 2D channel estimation schemes for OFDM and MIMO–OFDM, cf. e.g. [HKR97, LCS98, CS02, SBM+04], in this section we present MSDD with a 2D observation window. We first derive the ML metric for 2D MSDD in Section 6.5.1 and then transform it into a form amenable to the tree–search decoding methods of Chapter 3 in Section 6.5.2. Implementation details and variants of 2D MSDD are then discussed in Section 6.5.3.

6.5.1 Maximum–Likelihood Metric for 2D MSDD

From the channel model (6.14) and according to the SA scheme devised in Section 6.4.2, we define the $(N_{\mathrm{S}} \times N_{\mathrm{R}})$–dimensional MIMO–OFDM channel matrix

$$\boldsymbol{G}[m,k] \triangleq \Big[g_{\mathrm{mod}(i-1,N_{\mathrm{T}})+1,j}[(i-1)D/N_{\mathrm{S}}+m,k] \Big]_{\substack{i=1,\ldots,N_{\mathrm{S}} \\ j=1,\ldots,N_{\mathrm{R}}}}, \quad 1 \le m \le D/N_{\mathrm{S}},\ k \in \mathbb{N}. \tag{6.24}$$

The $(N_{\mathrm{S}} \times N_{\mathrm{R}})$–dimensional receive matrix $\boldsymbol{R}[m,k]$ corresponding to transmission of a space–frequency symbol $\boldsymbol{S}[m,k]$ is given by

$$\boldsymbol{R}[m,k] = \boldsymbol{S}[m,k]\boldsymbol{G}[m,k] + \boldsymbol{N}[m,k], \tag{6.25}$$

where $\boldsymbol{N}[m,k]$ is an $N_{\mathrm{S}} \times N_{\mathrm{R}}$ AWGN matrix. For illustration see Fig. 6.3. We note that (6.25) applies to both F–DSFM and T–DSFM.

The 2D MSDD observation window extends over

$$N = N_{\mathrm{f}} \cdot N_{\mathrm{t}} \tag{6.26}$$

received matrices $\boldsymbol{R}[m_i,k_j]$, $1 \le i \le N_{\mathrm{f}}$, $1 \le j \le N_{\mathrm{t}}$, i.e. N_{f} and N_{t} are the dimensions of the window in frequency and time, respectively. For T–DSFM we need $k_{j+1} = k_j + 1$, $1 \le j \le N_{\mathrm{t}} - 1$, and for F–DSFM we have $m_{i+1} = m_i + 1$, $1 \le i \le N_{\mathrm{f}} - 1$, to allow for detection of the data symbols $\boldsymbol{V}[m,k]$ according to (6.21) and (6.22), respectively. There are no restrictions on the indices m_i for T–DSFM and k_j for F–DSFM and appropriate choices will be discussed in Section 6.5.3.2. The corresponding received symbols can be collected in an $(N_{\mathrm{f}}N_{\mathrm{S}} \times N_{\mathrm{t}}N_{\mathrm{R}})$–dimensional matrix[4]

$$\boldsymbol{R}^{\square} \triangleq \Big[\boldsymbol{R}[m_i,k_j] \Big]_{\substack{i=1,\ldots,N_{\mathrm{f}} \\ j=1,\ldots,N_{\mathrm{t}}}}, \tag{6.27}$$

where "$\square$" emphasizes the rectangular extension of the 2D observation window in time and frequency. In order to formulate the ML–MSDD decision rule it is convenient to block–vectorize the received matrix $\boldsymbol{R}^{\square}$ and therewith the transmission model. To this end, we define the $(N_{\mathrm{f}}N_{\mathrm{t}}N_{\mathrm{S}} \times N_{\mathrm{R}})$–dimensional matrices

$$\boldsymbol{X}^{\parallel} \triangleq \big[\boldsymbol{X}^{\mathsf{H}}[m_1,k_1],\ \boldsymbol{X}^{\mathsf{H}}[m_2,k_1],\ \ldots,\ \boldsymbol{X}^{\mathsf{H}}[m_{N_{\mathrm{f}}},k_{N_{\mathrm{t}}}] \big]^{\mathsf{H}}, \quad \boldsymbol{X} \in \{\boldsymbol{R},\boldsymbol{G},\boldsymbol{N}\}, \tag{6.28}$$

the $(N_{\mathrm{f}}N_{\mathrm{t}}N_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional matrix

$$\boldsymbol{S}^{\parallel} \triangleq \big[\boldsymbol{S}^{\mathsf{H}}[m_1,k_1],\ \boldsymbol{S}^{\mathsf{H}}[m_2,k_1],\ \ldots,\ \boldsymbol{S}^{\mathsf{H}}[m_{N_{\mathrm{f}}},k_{N_{\mathrm{t}}}] \big]^{\mathsf{H}}, \tag{6.29}$$

[4]For the sake of readability, we omit the dependence of 2D MSDD block matrices of index sets $\{m_1,\ldots,m_{N_{\mathrm{f}}}\}$ and $\{k_1,\ldots,k_{N_{\mathrm{t}}}\}$.

and its $(N_{\mathrm{f}}N_{\mathrm{t}}N_{\mathrm{S}} \times N_{\mathrm{f}}N_{\mathrm{t}}N_{\mathrm{S}})$–dimensional block–diagonal version

$$\boldsymbol{S}_{\mathrm{D}}^{\parallel} \triangleq \mathrm{diag}\{\boldsymbol{S}[m_1,k_1],\boldsymbol{S}[m_2,k_1],\ldots,\boldsymbol{S}[m_{N_{\mathrm{f}}},k_{N_{\mathrm{t}}}]\} \ , \tag{6.30}$$

where "$\parallel$" indicates that matrices $\boldsymbol{X}[m_i,k_j]$ are stacked first according to the frequency index m_i and then according to the time index k_j (another possibility is discussed below). We can then rewrite (6.25) to match the 2D–MSDD observation window as

$$\boldsymbol{R}^{\parallel} = \boldsymbol{S}_{\mathrm{D}}^{\parallel}\boldsymbol{G}^{\parallel} + \boldsymbol{N}^{\parallel} \ . \tag{6.31}$$

Since given $\boldsymbol{S}^{\parallel}$, $\boldsymbol{R}^{\parallel}$ is zero–mean complex Gaussian distributed with autocorrelation matrix

$$\boldsymbol{\Psi}_{\boldsymbol{RR}|\boldsymbol{S}}^{\parallel} \triangleq \mathcal{E}\left\{ \boldsymbol{R}^{\parallel}\left(\boldsymbol{R}^{\parallel}\right)^{\mathsf{H}} \middle| \boldsymbol{S}^{\parallel} \right\} = \boldsymbol{S}_{\mathrm{D}}^{\parallel}\mathcal{E}\left\{\boldsymbol{G}^{\parallel}\left(\boldsymbol{G}^{\parallel}\right)^{\mathsf{H}}\right\}\left(\boldsymbol{S}_{\mathrm{D}}^{\parallel}\right)^{\mathsf{H}} + \mathcal{E}\left\{\boldsymbol{N}^{\parallel}\left(\boldsymbol{N}^{\parallel}\right)^{\mathsf{H}}\right\} \ , \tag{6.32}$$

the 2D ML–MSDD decision rule reads

$$\hat{\boldsymbol{S}}^{\parallel} = \underset{\boldsymbol{S}^{\parallel}}{\mathrm{argmin}}\left\{ \mathrm{tr}\left\{ \left(\boldsymbol{R}^{\parallel}\right)^{\mathsf{H}}\left(\boldsymbol{\Psi}_{\boldsymbol{RR}|\boldsymbol{S}}^{\parallel}\right)^{-1}\boldsymbol{R}^{\parallel} \right\} \right. . \tag{6.33}$$

It should be clear from the above that this decision rule applies to both T–DSFM and F–DSFM. From $\hat{\boldsymbol{S}}^{\parallel}$ the $N_{\mathrm{f}}(N_{\mathrm{t}}-1)$ and $(N_{\mathrm{f}}-1)N_{\mathrm{t}}$ data–matrix estimates $\hat{\boldsymbol{V}}[m,k]$ are determined using (6.21) and (6.22), respectively.

6.5.2 Tree–Search Algorithms for 2D MSDD

For a representation of (6.33) amenable to the application of tree–search algorithms as developed in Section 3.1, we make the assumption that the channel coefficients collected in $\boldsymbol{G}[m,k]$ defined in (6.24) are mutually uncorrelated. We note that this is a fairly mild assumption, since it only requires that (i) N_{B} subchannels spaced by $(D/N_{\mathrm{B}})\Delta f$ are uncorrelated, i.e. the channel coherence bandwidth is significantly less than $(D/N_{\mathrm{B}})\Delta f$, and (ii) the minimal distance between different elements of the transmit or receive antenna array is greater than half of the wavelength [SFGK00]. The effect of subcarrier and spatial correlation on performance and detection complexity will be discussed in Section 6.7. Applying these assumptions, it is shown in Appendix 6.A.2 that the autocorrelation matrix of $\boldsymbol{G}^{\parallel}$ can be written as

$$\mathcal{E}\left\{\boldsymbol{G}^{\parallel}\left(\boldsymbol{G}^{\parallel}\right)^{\mathsf{H}}\right\} = N_{\mathrm{R}}\tilde{\boldsymbol{\Psi}}^{\mathrm{t}} \otimes \boldsymbol{\Psi}^{\mathrm{f}} \otimes \boldsymbol{I}_{N_{\mathrm{S}}} \triangleq N_{\mathrm{R}}\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\parallel} \otimes \boldsymbol{I}_{N_{\mathrm{S}}} \ , \tag{6.34}$$

with time and frequency correlation matrices $\tilde{\boldsymbol{\Psi}}^{\mathrm{t}}$ and $\boldsymbol{\Psi}^{\mathrm{f}}$ as defined in (6.73) and (6.74), respectively. Taking further into account that the noise autocorrelation matrix is given by

$$\mathcal{E}\left\{\boldsymbol{N}^{\parallel}\left(\boldsymbol{N}^{\parallel}\right)^{\mathsf{H}}\right\} = \sigma_n^2 N_{\mathrm{R}}\boldsymbol{I}_{NN_{\mathrm{S}}} \ , \tag{6.35}$$

one can see that the detection problem is of the same form as the one considered in Section 3, and we can solve it by means of tree–search decoding based on

$$\hat{\boldsymbol{S}}^{\shortparallel} = \underset{\boldsymbol{S}^{\shortparallel}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \left\| \sum_{j=i}^{N} \breve{\boldsymbol{R}}_{i,j}^{\mathsf{H}} \boldsymbol{S}_j \right\|^2 \right\}, \tag{6.36}$$

where in analogy to the generic model used in Chapter 3 $\breve{\boldsymbol{R}}_{i,j} = c_{i,j}^{\shortparallel} \boldsymbol{R}_j$ and $\boldsymbol{R}_j$ and $\boldsymbol{S}_j$ denote the jth $(N_{\mathrm{S}} \times N_{\mathrm{R}})$–dimensional and $(N_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional submatrices of $\boldsymbol{R}^{\shortparallel}$ and $\boldsymbol{S}^{\shortparallel}$, respectively. The coefficients $c_{i,j}^{\shortparallel}$ are the elements of the upper triangular matrix $\boldsymbol{C}^{\shortparallel}$ defined as Cholesky factor of $\left(\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\shortparallel} + \sigma_n^2 \boldsymbol{I}_N\right)^{-1}$.

6.5.3 Implementation Aspects

Based on (6.36) all tree–search methods devised in Section 3.1 including Fano–type metric can be readily applied here. Apart from this, some further "fine–tuning" due to the 2D structure of the observation window of MSDD for MIMO–OFDM is possible and discussed in the following.

6.5.3.1 Sorting

In Section 6.5.1 we transformed the 2D into an equivalent 1D MSDD problem by stacking the block–columns of $\boldsymbol{R}^{\square}$, i.e. sorting received matrices $\boldsymbol{R}[m_i, k_j]$ first according to frequency index m_i and then according to time index k_j. We could have also sorted the other way, i.e. concatenating the block–rows of $\boldsymbol{R}^{\square}$, yielding

$$\boldsymbol{R}^{=} \triangleq \left[\boldsymbol{R}^{\mathsf{H}}[m_1, k_1], \boldsymbol{R}^{\mathsf{H}}[m_1, k_2], \ldots, \boldsymbol{R}^{\mathsf{H}}[m_{N_{\mathrm{f}}}, k_{N_{\mathrm{t}}}]\right]^{\mathsf{H}}. \tag{6.37}$$

With the corresponding definitions of $\boldsymbol{S}^{=}$, $\boldsymbol{G}^{=}$, and $\boldsymbol{N}^{=}$, this again leads to (6.36), but with $\boldsymbol{R}_j$ and $\boldsymbol{S}_j$ denoting the jth $(N_{\mathrm{S}} \times N_{\mathrm{R}})$–dimensional and $(N_{\mathrm{S}} \times N_{\mathrm{S}})$–dimensional submatrices of $\boldsymbol{R}^{=}$ and $\boldsymbol{S}^{=}$, respectively, and correlation matrix

$$\boldsymbol{\Psi}_{\boldsymbol{GG}}^{=} \triangleq \boldsymbol{\Psi}^{\mathrm{f}} \otimes \tilde{\boldsymbol{\Psi}}^{\mathrm{t}} \tag{6.38}$$

instead of $\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\shortparallel}$. More generally, any block–vectorization of the matrices in $\boldsymbol{R}^{\square}$ is possible, and the respective 2D ML–MSDD minimization problems will yield the same solution.

Computational complexity of the tree–search algorithms applied to perform 2D MSDD, on the other hand, may well depend on the sorting order. A tree–search algorithm is most conveniently imagined as climbing up branches of the code tree as long as they appear promising, thereby searching for the best branch of full length. Assume that at some point the decoder makes a false tentative decision and from there on travels along an erroneous branch. It will likely realize this mistake when it reaches a symbol that was transmitted over a channel

strongly correlated with that where the error took place. Hence, in order to minimize the time spent traversing erroneous paths and therewith complexity, sorting should be done such that symbols transmitted over strongly correlated channels are arranged close to each other. We note that this guideline for sorting is similar to the "sparsity index" criterion introduced in [MGDC06, Section III-B]. It also suggests that detection based on $\boldsymbol{R}^{\parallel}$ or $\boldsymbol{R}^{=}$ is preferable over arbitrarily stacked matrices, since spectral ($\boldsymbol{R}^{\parallel}$) and temporal ($\boldsymbol{R}^{=}$) channel correlations are best exploited (see also discussion below regarding observation window construction). We found that for relevant channel and detection parameters (see Section 6.7), differences in complexity when using $\boldsymbol{R}^{\parallel}$ and $\boldsymbol{R}^{=}$ are negligible, and we therefore solely consider 2D MSDD based on $\boldsymbol{R}^{\parallel}$ in the following.

6.5.3.2 Observation Window Construction (OWC)

Let us introduce the variables p, $1 \leq p \leq D/(N_{\mathrm{S}}N_{\mathrm{f}})$, and $q \in \mathbb{N}$ to identify the position of the $N_{\mathrm{f}} \times N_{\mathrm{t}}$ rectangular observation window in the frequency–time plane. The corresponding sets of frequency and time indices are denoted as $\mathcal{M}^{(p)} \triangleq \left\{m_1^{(p)}, m_2^{(p)}, \ldots, m_{N_{\mathrm{f}}}^{(p)}\right\} \subset \{1, 2, \ldots, D/N_{\mathrm{S}}\}$ and $\mathcal{K}^{(q)} \triangleq \left\{k_1^{(q)}, k_2^{(q)}, \ldots, k_{N_{\mathrm{t}}}^{(q)}\right\} \subset \mathbb{N}$. As stated in Section 6.5.1, the index sets in the direction of differential encoding, i.e. $\mathcal{M}^{(p)}$ for F–DSFM and $\mathcal{K}^{(q)}$ for T–DSFM, must be sets of consecutive integers. Furthermore, adjacent observation windows need to overlap by one index, i.e. $m_1^{(p+1)} = m_{N_{\mathrm{f}}}^{(p)}$ for F–DSFM and $k_1^{(q+1)} = k_{N_{\mathrm{t}}}^{(q)}$ for T–DSFM.

What is not immediately clear is how the index sets perpendicular to the direction of differential encoding, i.e. $\mathcal{M}^{(p)}$ for T–DSFM and $\mathcal{K}^{(q)}$ for F–DSFM, should be chosen. We note that for T–DSFM the sets $\mathcal{M}^{(p)}$ are disjoint and that the union $\cup_{p=1}^{D/(N_{\mathrm{S}}N_{\mathrm{f}})}\mathcal{M}^{(p)}$ is the set of integers from 1 to D/N_{S}. Similarly, $\mathcal{K}^{(q)}$ are disjoint and $\cup_{q\in\mathbb{N}}\mathcal{K}^{(q)} = \mathbb{N}$ in case of F–DSFM. While an optimization with respect to error rate appears to be intractable, it is clear due to the relation between MSDD and linear MMSE interpolation (cf. Section 2.4.2.1) that the channel matrices $\boldsymbol{G}\left[m_i^{(p)}, k_j^{(q)}\right]$ captured in the observation window should be correlated to the largest possible degree.

In the case of F–DSFM, minimizing detection delay would suggest to chose $\mathcal{K}^{(q)}$ as sets of consecutive integers as well. In practically relevant scenarios, this also coincides with the choice for maximal correlation. Fig. 6.4 shows the temporal correlation $\tilde{\psi}^{\mathrm{t}}[\kappa]$ (see Eq. (6.19)) for Clarke's fading model with normalized fading bandwidths up to $B_h T_{\mathrm{f}} = 0.01$. It can be seen that for reasonable values of N_{t}, contiguous sequences of indices $k_j^{(q)}$, i.e. $k_{j+1}^{(q)} = k_j^{(q)} + 1$, $1 \leq j \leq N_{\mathrm{t}} - 1$, maximize correlation and are thus optimal.

The same is true considering $\mathcal{M}^{(p)}$ for T–DSFM and certain commonly used channel models such as the typical urban (TU) model [Eur01] and the model with exponential PDP with decay parameter $d = 1$ ($\tau_l = lT$, $\sigma_l^2 \propto \mathrm{e}^{-l/d}$, $l \in \mathbb{N}$). This can be seen from Fig. 6.5, where the spectral

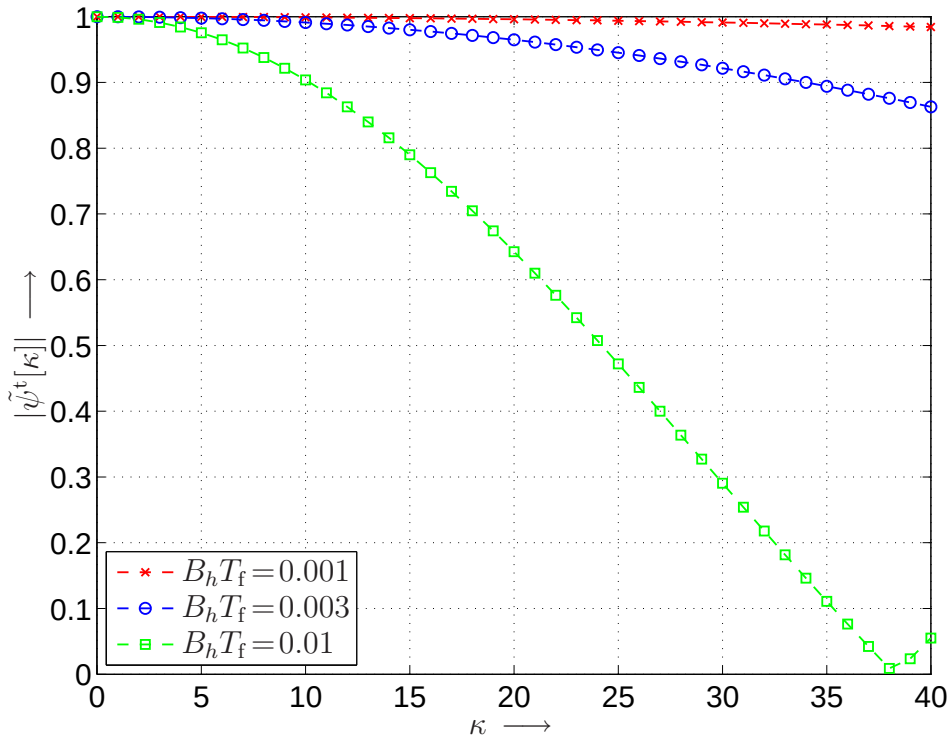

Figure 6.4: Magnitude of temporal correlation function for Clarke's fading model with various values of $B_h T_{\mathrm{f}}$.

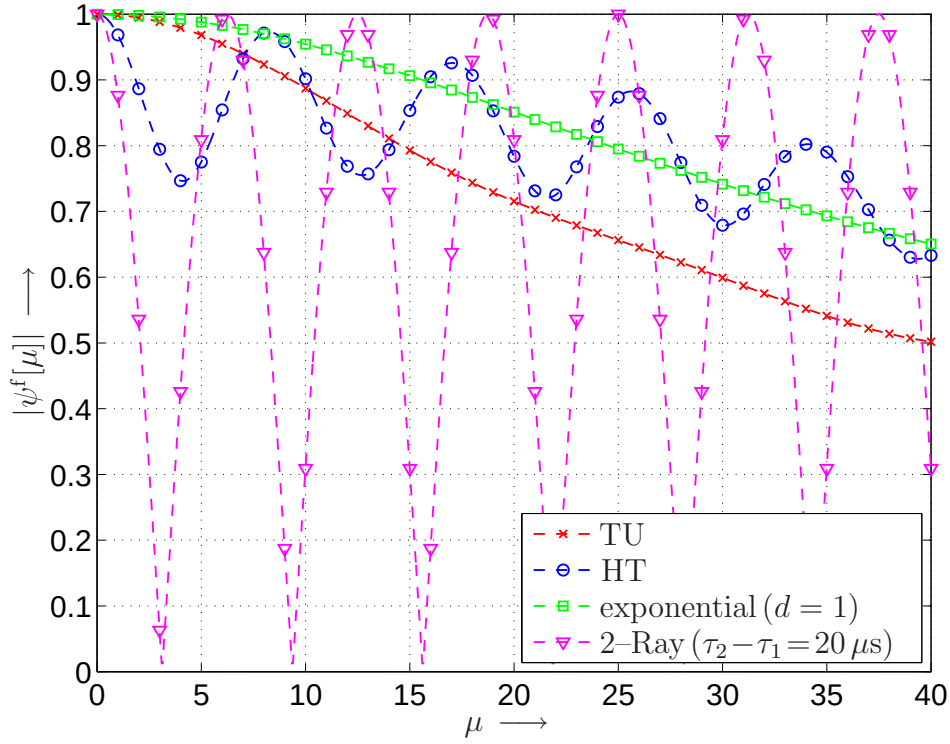

Figure 6.5: Magnitude of spectral correlation function for various PDPs, $D = 192$, and $\Delta f = 8\,\mathrm{kHz}$.

correlation $\psi^{\mathrm{f}}[\mu]$ [cf. (6.18)] is plotted assuming $D = 192$ subcarriers and a subcarrier frequency spacing of $\Delta f = 8\,\mathrm{kHz}$ as used for digital audio broadcasting (DAB) systems [Eur95]. Fig. 6.5 also shows $\psi^{\mathrm{f}}[\mu]$ for the hilly terrain (HT) channel model [Eur01] and the two–ray (2–Ray) channel with $\tau_2 - \tau_1 = 20\,\mu\mathrm{s}$. For these channels we observe a "strongly non–concave" behavior of the spectral correlation function, which suggests that partitioning the D/N_{S} frequency indices into disjoint sets $\mathcal{M}^{(p)}$ of non–consecutive integers is likely to improve the performance of T–DSFM with 2D MSDD. A meaningful figure of merit for a set $\mathcal{M}^{(p)}$ is the sum of spectral correlations within the observation window,

$$\mu\left(\mathcal{M}^{(p)}\right) \triangleq \sum_{i=2}^{N_{\mathrm{f}}} \sum_{l=1}^{i-1} \left| \psi^{\mathrm{f}} \left[m_i^{(p)} - m_l^{(p)} \right] \right| , \tag{6.39}$$

which should be maximized for each p. We therefore propose the following greedy–type algorithm to construct the sets $\mathcal{M}^{(p)}$.

Initialize $\mathcal{I}^{(0)} = \{1, \ldots, D/N_{\mathrm{S}}\}$

for $p = 1, \ldots, D/(N_{\mathrm{S}} N_{\mathrm{f}})$

Pick $m_1^{(p)} \in \mathcal{I}^{(p-1)}$

$$\left\{ m_2^{(p)}, \ldots, m_{N_{\mathrm{f}}}^{(p)} \right\} = \underset{\left\{ m_2^{(p)}, \ldots, m_{N_{\mathrm{f}}}^{(p)} \right\} \subseteq \mathcal{I}^{(p-1)} \setminus \left\{ m_1^{(p)} \right\}}{\operatorname{argmax}} \left\{ \mu\left(m_1^{(p)}, \ldots, m_{N_{\mathrm{f}}}^{(p)} \right) \right\}$$

$$\mathcal{M}^{(p)} = \left\{ m_1^{(p)}, \ldots, m_{N_{\mathrm{f}}}^{(p)} \right\}$$

$$\mathcal{I}^{(p)} = \mathcal{I}^{(p-1)} \setminus \mathcal{M}^{(p)}$$

end

Remark: While the problem at hand is a kind of "traveling–salesman problem" [LLKS85] whose criterion of optimality over *all* sets $\mathcal{M}^{(p)}$ is difficult to grasp, this greedy algorithm succeeds nicely in achieving near optimal average SER performance. The reason is that due to the strong temporal correlation in the channel only very high spectral correlation helps to improve the reliability of the decisions in the observation window. Hence, the greedy approach of finding as many sets $\mathcal{M}^{(p)}$ as possible with maximum μ and more or less "patching together the leftovers" in a small number of sets is a very good solution with respect to average SER.

6.6 Performance Analysis

In this section, we derive expressions to analyze the SER performance of DSFM with 2D ML–MSDD as presented above, and we illustrate their usefulness to guide the design of the appropriate transmission parameters.

The course of action in deriving the SER approximations is the same as the one used for the analysis in the case of DSTM transmission over the frequency–nonselective channel, cf. Section 5.4. I.e. first the pairwise error probability $\mathrm{PEP}\big(\boldsymbol{S}^{\|} \to \hat{\boldsymbol{S}}^{\|}\big)$ that $\hat{\boldsymbol{S}}^{\|}$ is detected while $\boldsymbol{S}^{\|}$ was transmitted is considered. In the second step, we identify the dominant error events and sum the corresponding PEPs to tightly approximate the average SER.

6.6.1 Pairwise Error Probabilities for Spatially Correlated Channel

6.6.1.1 (DF–)MSDD

Let us write the matrix channel model (6.31) in a vector format as

$$\bar{\boldsymbol{r}}' \triangleq \mathrm{vec}\left\{\boldsymbol{R}^{\|}\right\} = \left(\boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{S}_\mathrm{D}^{\|}\right) \mathrm{vec}\left\{\boldsymbol{G}^{\|}\right\} + \mathrm{vec}\left\{\boldsymbol{N}^{\|}\right\} \ . \tag{6.40}$$

With the help of Appendix 6.A.2 it is easy to see that the $(N_\mathrm{R}N_\mathrm{f}N_\mathrm{t}N_\mathrm{S} \times N_\mathrm{R}N_\mathrm{f}N_\mathrm{t}N_\mathrm{S})$–dimensional autocorrelation matrix $\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}'\bar{\boldsymbol{r}}'|\boldsymbol{S}^{\|}}$ of $\bar{\boldsymbol{r}}'$ conditioned on $\boldsymbol{S}^{\|}$ is given by

$$\begin{aligned} \boldsymbol{\Psi}_{\bar{\boldsymbol{r}}'\bar{\boldsymbol{r}}'|\boldsymbol{S}^{\|}} &\triangleq \mathcal{E}\Big\{\bar{\boldsymbol{r}}'(\bar{\boldsymbol{r}}')^{\mathsf{H}} \,\Big|\, \boldsymbol{S}^{\|}\Big\} & (6.41)\\ &= \left(\boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{S}_\mathrm{D}^{\|}\right) \cdot \left(\tilde{\boldsymbol{\Psi}}_{\boldsymbol{GG}}^{\|} + \sigma_n^2 \boldsymbol{I}_{N_\mathrm{R}NN_\mathrm{S}}\right) \cdot \left(\boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{S}_\mathrm{D}^{\|}\right)^{\mathsf{H}}, & (6.42) \end{aligned}$$

with $\tilde{\boldsymbol{\Psi}}_{\boldsymbol{GG}}^{\|}$ from (6.77).

Introducing further the $(N \times N)$–dimensional inverse correlation matrix of the channel–plus–noise process as assumed by the receiver

$$\boldsymbol{M}' \triangleq \left(\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\|} + \sigma_n^2 \boldsymbol{I}_N\right)^{-1} \ , \tag{6.43}$$

with $\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\|}$ according to (6.34) and the $(N_\mathrm{f}N_\mathrm{t}N_\mathrm{S} \times N_\mathrm{f}N_\mathrm{t}N_\mathrm{S})$–dimensional difference matrix

$$\boldsymbol{F}' \triangleq \hat{\boldsymbol{S}}_\mathrm{D}^{\|} \left(\boldsymbol{M}' \otimes \boldsymbol{I}_{N_\mathrm{S}}\right) \left(\hat{\boldsymbol{S}}_\mathrm{D}^{\|}\right)^{\mathsf{H}} - \boldsymbol{S}_\mathrm{D}^{\|} \left(\boldsymbol{M}' \otimes \boldsymbol{I}_{N_\mathrm{S}}\right) \left(\boldsymbol{S}_\mathrm{D}^{\|}\right)^{\mathsf{H}} \ , \tag{6.44}$$

we can express the difference Δ' between the 2D ML–MSDD metrics [see (6.33)] for $\hat{\boldsymbol{S}}^{\|}$ and $\boldsymbol{S}^{\|}$ as

$$\Delta' = (\bar{\boldsymbol{r}}')^{\mathsf{H}} \left(\boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{F}'\right) \bar{\boldsymbol{r}}' \ . \tag{6.45}$$

Apparently, the metric difference Δ' is —as in Section 5.4.1— a Hermitian quadratic form in zero–mean complex Gaussian random variables. Hence its characteristic function $\Phi_{\Delta'}(v)$ reads

$$\Phi_{\Delta'}(v) = \prod_{i=1}^{NN_\mathrm{S}N_\mathrm{R}} \left(1 - \mathrm{j}v\lambda_{\boldsymbol{\Psi}_{\bar{\boldsymbol{r}}'\bar{\boldsymbol{r}}'|\boldsymbol{S}^{\|}}\left(\boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{F}'\right),i}\right)^{-1} \ , \tag{6.46}$$

and the PEP can be computed using the methods discussed in Section 5.4.1.

It is important to note that while the correlation matrix $\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\|}$ assuming uncorrelated elements in $\boldsymbol{G}[m,k]$ is used in the decision rule (6.33), the true channel correlation matrix $\tilde{\boldsymbol{\Psi}}_{\boldsymbol{GG}}^{\|}$ is applied in (6.46). Thus, the effect of metric mismatch due to the decoder's assumption of no correlation is captured by the PEP analysis.

6.6.1.2 CDD and DFDD

Due to the close relation between CDD, DFDD and MSDD, the expressions derived above for the PEPs of MSDD extend to CDD ($N = 2$) and DFDD as well.

6.6.1.3 (Differentially) Coherent Detection

In analogy to our considerations on the PEPs of coherent detection for the flat–fading channel in Section 5.4.1.3 we can compute the PEPs of coherent detection based on the evaluation of the characteristic function

$$\begin{aligned} \Phi_{\Delta'_{\mathrm{coh}}}(v) &= \det\left\{ \boldsymbol{I}_{N_{\mathrm{S}}N_{\mathrm{R}}} - \mathrm{j}v\boldsymbol{\Psi}_{\bar{\boldsymbol{z}}'\bar{\boldsymbol{z}}'|\boldsymbol{S}}\boldsymbol{F}'_{\mathrm{coh}} \right\}^{-1} && (6.47) \\ &= \prod_{j=1}^{N_{\mathrm{S}}N_{\mathrm{R}}} \left(1 - \mathrm{j}v\lambda_{\boldsymbol{\Psi}_{\bar{\boldsymbol{z}}'\bar{\boldsymbol{z}}'|\boldsymbol{S}}\boldsymbol{F}'_{\mathrm{coh}},j}\right)^{-1}, && (6.48) \end{aligned}$$

of the Hermitian quadratic form

$$\Delta'_{\mathrm{coh}} = (\bar{\boldsymbol{z}}')^{\mathsf{H}} \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{F}'_{\mathrm{coh}}\right) \bar{\boldsymbol{z}}', \tag{6.49}$$

with

$$\bar{\boldsymbol{z}}' \triangleq \begin{bmatrix} \mathrm{vec}\left\{\boldsymbol{G}[m,k]\right\} \\ \mathrm{vec}\left\{\boldsymbol{R}[m,k]\right\} \end{bmatrix} \tag{6.50}$$

$$\boldsymbol{F}'_{\mathrm{coh}} \triangleq \begin{bmatrix} \boldsymbol{0}_{N_{\mathrm{S}},N_{\mathrm{S}}} & \left(\boldsymbol{S}[m,k] - \hat{\boldsymbol{S}}[m,k]\right)^{\mathsf{H}} \\ \boldsymbol{S}[m,k] - \hat{\boldsymbol{S}}[m,k] & \boldsymbol{0}_{N_{\mathrm{S}},N_{\mathrm{S}}} \end{bmatrix} \tag{6.51}$$

and

$$\begin{aligned} \boldsymbol{\Psi}_{\bar{\boldsymbol{z}}'\bar{\boldsymbol{z}}'|\boldsymbol{S}} &\triangleq \mathcal{E}\left\{ \bar{\boldsymbol{z}}'(\bar{\boldsymbol{z}}')^{\mathsf{H}} \mid \boldsymbol{S}[m,k] \right\} && (6.52) \\ &= \begin{bmatrix} \boldsymbol{\Psi}_{\boldsymbol{GG}} & \boldsymbol{\Psi}_{\boldsymbol{GG}} \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{S}^{\mathsf{H}}[m,k]\right) \\ \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{S}[m,k]\right) \boldsymbol{\Psi}_{\boldsymbol{GG}} & \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{S}[m,k]\right) \boldsymbol{\Psi}_{\boldsymbol{GG}} \left(\boldsymbol{I}_{N_{\mathrm{R}}} \otimes \boldsymbol{S}^{\mathsf{H}}[m,k]\right) + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{S}}N_{\mathrm{R}}} \end{bmatrix}. \end{aligned}$$

Using the results of Appendix 6.A.2.1 it is straightforward to show that

$$\begin{aligned} \boldsymbol{\Psi}_{\boldsymbol{GG}} &\triangleq \mathcal{E}\Big\{ \mathrm{vec}\left\{\boldsymbol{G}[m,k]\right\} \mathrm{vec}\left\{\boldsymbol{G}[m,k]\right\}^{\mathsf{H}} \Big\} && (6.53) \\ &= \boldsymbol{\Psi}^{\mathrm{Rx}} \otimes \left(\tilde{\boldsymbol{\Psi}}^{\mathrm{f}}[0] \circ \left(\boldsymbol{1}_{N_{\mathrm{B}},N_{\mathrm{B}}} \otimes \boldsymbol{\Psi}^{\mathrm{Tx}} \right) \right) && (6.54) \end{aligned}$$

($\boldsymbol{A} \circ \boldsymbol{B}$: Hadamard (elementwise) product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$) with $\boldsymbol{\Psi}^{\mathrm{Tx}}$, $\boldsymbol{\Psi}^{\mathrm{Rx}}$ and $\tilde{\boldsymbol{\Psi}}^{\mathrm{f}}[m]$ as defined in (4.9), (4.10) and (6.71), respectively

6.6.2 Pairwise Error Probabilities for Spatially Uncorrelated Channel

From Eq. (6.25) one can see that the $N_\mathrm{S} \times N_\mathrm{R}$ equivalent MIMO channel over which matrix symbols $\boldsymbol{S}[m,k]$ are transmitted is accurately described by a QSFC–type channel model, which is due to the fact that the $\boldsymbol{S}[m,k]$ are diagonal matrices.

Absence of "spatial correlation" in the $N_\mathrm{S} \times N_\mathrm{R}$ equivalent MIMO channel is obtained, if

$$\boldsymbol{\Psi}^{\mathrm{Rx}} = \boldsymbol{I}_{N_\mathrm{R}} \tag{6.55}$$

and either

- $N_\mathrm{S} = N_\mathrm{T}$ and $\boldsymbol{\Psi}^{\mathrm{Tx}} = \boldsymbol{I}_{N_\mathrm{T}}$ or
- subcarriers allocated to the different elements $s_{i,i}[m,k]$, $1 \leq i \leq N_\mathrm{S}$, of an $\boldsymbol{S}[m,k]$ are sufficiently decorrelated [cf. (6.77)].

In this scenario, we have

$$\tilde{\boldsymbol{\Psi}}^{\parallel}_{\boldsymbol{GG}} = \boldsymbol{I}_{N_\mathrm{R}} \otimes \boldsymbol{\Psi}^{\parallel}_{\boldsymbol{GG}} \otimes \boldsymbol{I}_{N_\mathrm{S}} \tag{6.56}$$

and we can make immediate use of the results obtained in Section 5.4.2 for the spatially uncorrelated quasi–static flat–fading channel by replacing $\bar{\boldsymbol{H}}$ as used in Section 5.4.2 with $\boldsymbol{G}^{\parallel}$. More specifically, all results for the PEPs of the various detection schemes found there in dependence of the fading correlation matrix including the effective SNR are directly applicable.

6.6.2.1 MSDD, DFDD, CDD

In the case of noncoherent detection the results of Section 5.4.2 can be transfered to this transmission scenario by replacing $\boldsymbol{\Psi}^{\mathrm{t}}$ as used in Section 5.4.2 with $\boldsymbol{\Psi}^{\parallel}_{\boldsymbol{GG}}$.

One interesting result that is inherent to OFDM–based transmission is the following. Recall that we showed in Section 5.4.2.2 that for transmission over flat–fading channels the noncoherent detectors all encounter an error floor at high SNR, whose level is directly proportional to $(B_{h,\mathrm{eff}}T)^{2N_\mathrm{T}N_\mathrm{R}(N-1)}$, cf. (5.92) with (5.93). This clearly carries over to noncoherent detection with 1D observation window for T–DSFM. For the interesting special case of an 1D observation window for F–DSFM, i.e. in frequency direction, a similar and intuitively reasonable result can be obtained in dependence of the normalized delay spread of the channel, i.e. the quantity $(\tau_{L_h} - \tau_1)/(DT)$. As shown in Appendix 6.A.3, the PEP error floor in position n of the observation window of the noncoherent detectors is in this case given by

$$\lim_{\sigma_n^2 \to 0} \mathrm{PEP}_n \approx \left(\left[\left(\boldsymbol{\Psi}^{\parallel}_{\boldsymbol{GG}} \right)^{-1} \right]_{n,n} - 1 \right)^{-N_\mathrm{T}N_\mathrm{R}} \frac{\binom{2N_\mathrm{T}N_\mathrm{R}-1}{N_\mathrm{T}N_\mathrm{R}}}{\prod_{i=1}^{N_\mathrm{T}} |\theta_i(l,\hat{l})|^{2N_\mathrm{R}}}, \tag{6.57}$$

where the diagonal elements $\left[(\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\parallel})^{-1}\right]_{n,n}$ of the L_h inverse of the fading correlation matrix $\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\parallel}$ are asymptotically given by

$$\left[\left(\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\parallel}\right)^{-1}\right]_{n,n} \asymp c' \cdot \left(\frac{\tau_{L_h}-\tau_1}{DT}\right)^{-2(N-1)}, \quad \frac{\tau_{L_h}-\tau_1}{DT} \to 0, \tag{6.58}$$

with some constant c' that depends on the shape of the power delay profile, i.e. the relative position $(\tau_l-\tau_1)/(\tau_{L_h}-\tau_1)$ and average power σ_l^2 of the individual channel tabs.

Recall (cf. Appendix 5.A.4) that the accuracy of the PEP approximation (6.57) improves rapidly as the diagonal elements of $\left(\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\parallel}\right)^{-1}$ grow beyond one, i.e. if the delay spread $(\tau_{L_h}-\tau_1)$ becomes small compared to DT.

6.6.2.2 (Differentially) Coherent Detection

Obtaining expressions for the exact PEPs of coherent detection for DSFM from those for the flat–fading channel is equally simple. All that needs to be done is to replace $\boldsymbol{H}$ in (5.78) and (5.79) with $\boldsymbol{G}[m,k]$ [cf. (6.24)].

6.6.3 Approximation of Symbol–Error Rate

Having determined an expression for the PEP between any pair of transmit sequences we use a truncated union bound over the dominant error events to approximate the average SER. As argued in Section 5.4 for the frequency–nonselective channel, the dominant error events of MSDD are single transmit–symbol error events $\left(\boldsymbol{S}[m_i,k_j] \to \hat{\boldsymbol{S}}[m_i,k_j]\right)$ for which

$$\prod_{\nu=1}^{N_{\mathrm{S}}} \left|\lambda_{\hat{\boldsymbol{S}}^{\mathsf{H}}[m_i,k_j]\boldsymbol{S}[m_i,k_j],\nu}\right| \tag{6.59}$$

is minimized. Accordingly, we define sets $\hat{\boldsymbol{\mathcal{S}}}_{i,j}^{\parallel}$ of all matrices $\hat{\boldsymbol{S}}^{\parallel} = \big[\boldsymbol{S}^{\mathsf{T}}[m_1,k_1], \ldots, \hat{\boldsymbol{S}}^{\mathsf{T}}[m_i,k_j], \ldots, \boldsymbol{S}^{\mathsf{T}}[m_{N_{\mathrm{f}}},k_{N_{\mathrm{t}}}]\big]^{\mathsf{T}}$ that differ from $\boldsymbol{S}^{\parallel}$ only in $\hat{\boldsymbol{S}}[m_i,k_j] \neq \boldsymbol{S}[m_i,k_j]$ and that minimize (6.59). Taking into account that a data–symbol error occurs whenever the preceding or succeeding transmit symbol is erroneously detected, the SER for the data–symbol $\boldsymbol{V}[m_i,k_j]$ in the ith row and jth column of the observation window can be approximated by

$$\mathrm{SER}_{i,j}^{\mathrm{TD}} = \sum_{\hat{\boldsymbol{S}}^{\parallel}\in\hat{\boldsymbol{\mathcal{S}}}_{i,j}^{\parallel}} \mathrm{PEP}\Big(\boldsymbol{S}^{\parallel} \to \hat{\boldsymbol{S}}^{\parallel}\Big) + \sum_{\hat{\boldsymbol{S}}^{\parallel}\in\hat{\boldsymbol{\mathcal{S}}}_{i,j+1}^{\parallel}} \mathrm{PEP}\Big(\boldsymbol{S}^{\parallel} \to \hat{\boldsymbol{S}}^{\parallel}\Big), \quad \begin{array}{l} 1 \le i \le N_{\mathrm{f}}, \\ 1 \le j \le N_{\mathrm{t}}-1, \end{array} \tag{6.60}$$

for T–DSFM and by

$$\mathrm{SER}_{i,j}^{\mathrm{FD}} = \sum_{\hat{\boldsymbol{S}}^{\parallel}\in\hat{\boldsymbol{\mathcal{S}}}_{i,j}^{\parallel}} \mathrm{PEP}\Big(\boldsymbol{S}^{\parallel} \to \hat{\boldsymbol{S}}^{\parallel}\Big) + \sum_{\hat{\boldsymbol{S}}^{\parallel}\in\hat{\boldsymbol{\mathcal{S}}}_{i+1,j}^{\parallel}} \mathrm{PEP}\Big(\boldsymbol{S}^{\parallel} \to \hat{\boldsymbol{S}}^{\parallel}\Big), \quad \begin{array}{l} 1 \le i \le N_{\mathrm{f}}-1, \\ 1 \le j \le N_{\mathrm{t}}. \end{array} \tag{6.61}$$

for F–DSFM. Finally, the average SER over all symbols comprised in the 2D observation window is obtained from

$$\text{T–DSFM:} \quad \text{SER}^{\text{TD}} = \frac{1}{N_{\text{f}}(N_{\text{t}}-1)} \sum_{i=1}^{N_{\text{f}}} \sum_{j=1}^{N_{\text{t}}-1} \text{SER}_{i,j}^{\text{TD}} \,, \tag{6.62}$$

$$\text{F–DSFM:} \quad \text{SER}^{\text{FD}} = \frac{1}{(N_{\text{f}}-1)N_{\text{t}}} \sum_{i=1}^{N_{\text{f}}-1} \sum_{j=1}^{N_{\text{t}}} \text{SER}_{i,j}^{\text{FD}} \,. \tag{6.63}$$

Note that averaging with respect to $\boldsymbol{S}^{\|}$ is not required as the matrix signal constellation underlying our DSFM scheme are cyclic DSTM codes, cf. Section 5.4.4.1.1.

The SERs for DFDD, CDD and (differentially) coherent detection follow in complete analogy to Sections 5.4.4.2, 5.4.4.3 and 5.4.4.4, respectively.

6.7 Results and Discussion

In this section, we present results obtained from our analysis and simulation results to illustrate the performance of MIMO–OFDM with the proposed DSFM SA scheme and fast 2D MSDD. We exemplarily consider the digital–audio–broadcasting (DAB) system OFDM parameters $D = 192$ and $\Delta f = 8\,\text{kHz}$ [Eur95]. As test channels, we consider channels with HT and TU PDPs taken from [Eur01] and the exponential PDP with decay parameter $d = 1$ ($\tau_l = lT$, $\sigma_l^2 \propto \text{e}^{-l/d}$, $l \in \mathbb{N}$). The coefficients $[c_1, \ldots, c_{N_{\text{S}}}]$ for the diagonal constellations [cf. (2.2)] are again taken from [HS00, Table I]. To limit the number of parameters, we fix the data rate to $R = 1$ bit per OFDM subchannel use and assume a single receive antenna, i.e. $N_{\text{R}} = 1$. For implementation of fast 2D MSDD we consider MSDSD–FS, which achieves ML–MSDD performance, and DF–MSDSD–LD–FM, which typically achieves comparable performance at a significantly lower complexity especially in low–SNR regimes. As benchmark detectors we again consider conventional differential detection (CDD) ($N = 2$) and decision–feedback differential detection (DFDD) with 1D observation window of length N, their respective lattice–decoder (LD) based counterparts CDD–LD and DFDD–LD, and coherent detection with perfect CSI at the receiver. In the case of F–DSFM, DFDD was implemented making use of $N-1$ pilot symbols $\boldsymbol{S}[m,k]$, $1 \leq m \leq N-1$, $k \in \mathbb{N}$, at the beginning of each OFDM frame in order to avoid a high error floor due to (i) a growing observation window at the start of DFDD and (ii) subsequent excessive error propagation. The corresponding rate–loss is accounted for in $E_{\text{b}}/\mathcal{N}_0$. Unless specified otherwise, we assume a spatially uncorrelated channel, i.e. $\boldsymbol{\Psi}^{\text{Tx}} = \boldsymbol{I}_{N_{\text{T}}}$.

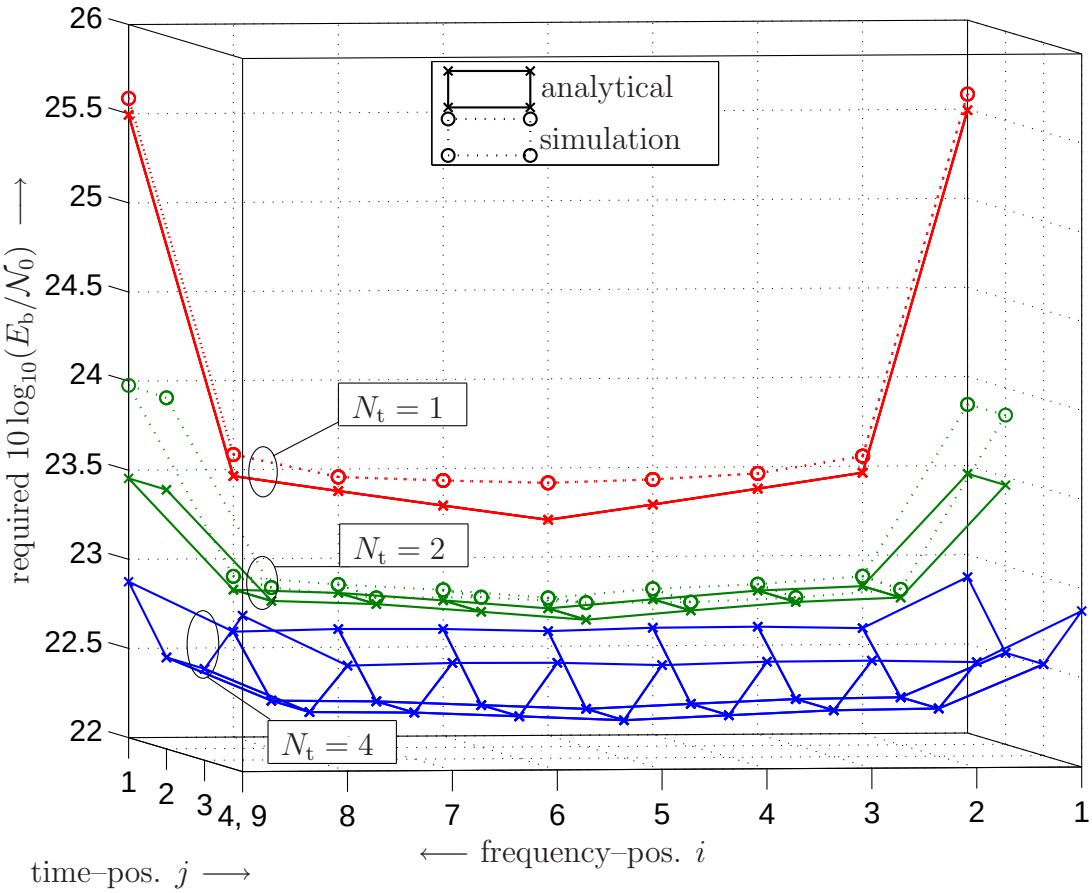

Figure 6.6: Required $10\log_{10}(E_b/\mathcal{N}_0)$ to achieve SER $= 10^{-5}$ for individual positions of 2D observation window of MSDD. Parameters: F–DSFM, $N_T = 3$, $N_B = 1$, HT PDP, $B_h T_f = 0.01$, $N_f = 10$ and $N_t = [1, 2, 4]$.

6.7.1 SER Performance

Required SNR for SER $= 10^{-5}$ **vs. Position** $[i, j]$**:** For illustration purposes, we evaluate $\text{SER}_{i,j}^{\text{FD}}$ from (6.61) for F–DSFM with the following system parameters: DSFM with $N_T = 3$, $N_B = 1$, $N_R = 1$ and $R = 1$, and a time– and frequency–selective channel with HT PDP [Eur01] and $B_h T_f = 0.01$. The 2D–MSDD observation window extends over $N_f = 10$ subcarriers and different numbers N_t of OFDM frames. Fig. 6.6 shows the SNR in terms of $10\log_{10}(E_b/\mathcal{N}_0)$ required to achieve individual error rates of $\text{SER}_{i,j}^{\text{FD}} = 10^{-5}$ as function of the position $[i, j]$, $1 \le i \le N_f - 1$, $1 \le j \le N_t$, for $N_t = [1, 2, 4]$. Simulation results for $N_t = 1$ and $N_t = 2$ are included, as well. MSDD is implemented using MSDSD with FS inner decoding (MSDSD–FS).

First, we note that simulation results closely match the results obtained from (6.61), which nicely confirms the accuracy of the proposed SER approximation. We further observe a kind of "hammock behavior" with larger values for the required SNR towards the edges of the 2D observation window. This was to be expected given the relation between MSDD and linear

SA	$[N_\mathrm{f}, N_\mathrm{t}]$	$N_\mathrm{T}=2$	$N_\mathrm{T}=3$	$N_\mathrm{T}=4$	$N_\mathrm{T}=6$
[LS01]	$[10,1]$	30.7	26.0	23.4	20.7
	$[5,2]$	*35.0	*29.2	—	—
	$[20,1]$	29.3	24.0	20.8	18.1
	$[10,2]$	28.8	23.3	20.1	17.2
[WY02]	$[10,1]$	29.8	24.0	20.4	17.4
	$[5,2]$	29.2	23.4	19.9	17.1
	$[20,1]$	28.7	23.1	19.7	16.8
	$[10,2]$	28.6	22.9	19.5	16.7
Section 6.4.2	$[10,1]$	29.8	24.0	20.4	17.4
	$[5,2]$	28.9	23.1	19.7	16.8
	$[20,1]$	28.7	23.1	19.7	16.8
	$[10,2]$	28.4	22.8	19.4	16.5
coherent:		27.4	21.9	18.5	15.8

Table 6.1: Required $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ to achieve SER $= 10^{-5}$ using MSDD and various SA schemes. Parameters: F–DSFM, $N_\mathrm{S} = N_\mathrm{T}$, $N_\mathrm{B} = 1$, HT PDP, $B_h T_\mathrm{f} = 0.003$. For comparison: coherent detection with perfect CSI. Numerical results from (6.63), figures with "*" are simulation results and "—" indicates an error floor above SER $= 10^{-5}$.

MMSE interpolation (cf. Section 2.4.2.1). However, as N_t increases this effect diminishes, which means that the application of the 2D window is highly beneficial for achieving uniform individual error rates. It can also be seen that increasing N_t leads to significant improvements in power efficiency, i.e. lowers the required SNR, of the detector even in this relatively fast fading scenario. Quite remarkably, for $N_\mathrm{t} = 4$ the gap to coherent detection with perfect CSI (not shown in the figure) is only approximately 0.6 dB on average and 0.4 dB in the center positions of the observation window.

F–DSFM and ML MSDD: To illustrate the benefits of the SA scheme presented in Section 6.4.2 and the application of a 2D observation window we consider the example of F–DSFM with different values of $N_\mathrm{S} = N_\mathrm{T}$, i.e. $N_\mathrm{B} = 1$, a channel with HT PDP and $B_h T_\mathrm{f} = 0.003$, and MSDD with different observation window parameters $[N_\mathrm{f}, N_\mathrm{t}]$. MSDD is implemented using MSDSD–FS. Table 6.1 compares the SNR required to achieve SER $= 10^{-5}$ for (i) the proposed SA scheme, (ii) a scheme, where $s_{i,i}[m,k]$ and $s_{i,i}[m+1,k]$ are separated by N_T subcarriers while the effective fading bandwidth is not increased, e.g. [LS01], and (iii) a scheme, where $s_{i,i}[m,k]$ and $s_{i,i}[m+1,k]$ are allocated to neighboring subcarriers but the effective fading

bandwidth is increased by a factor N_{T}, e.g. [WY02, MTL05]. SA schemes such as [HSL05b]–[SL05] that exploit knowledge of the PDP at the transmitter are not included in the comparison. The SA variant proposed in [HSL05c, Section V] is not considered either, as the pseudo–random subcarrier assignment seriously complicates MSDD (and DFDD) due to the fact that $\mathbf{\Psi}^{\mathrm{f}}$ changes as the observation window slides across the received data. The respective SNR values for coherent detection with perfect CSI are also included for comparison. The results are obtained from the SER approximation (6.63), except for the values for SA according to [LS01] and $[N_{\mathrm{f}}, N_{\mathrm{t}}] = [5, 2]$, which are simulation results ("—" indicates the occurrence of an error floor above SER $= 10^{-5}$). Here, $N_{\mathrm{f}} = 5$ is not sufficient for tracking the channel variations in frequency direction and performance is significantly degraded by multiple–symbol errors due to fading. These are not accounted for in (6.63), which therefore is too optimistic in this case.

The new SA method consistently achieves the best power efficiency. The gains over the scheme of [LS01] are more pronounced than those when comparing with [WY02, MTL05], since the correlations of the channel gains captured within the observation window are lowest for the scheme of [LS01] (see also Figs. 6.4 and 6.5). In particular, in the case of 1D MSDD the SA schemes of [WY02, MTL05] achieve the same power efficiency as our SA scheme, as the increase of the effective fading bandwidth has no effect here. With larger total window size $N = N_{\mathrm{f}} N_{\mathrm{t}}$ the gains are smaller as the performance approaches that of coherent detection. Similarly, comparing 2D MSDD and 1D MSDD, the largest gains are achieved for relatively small N. For example, for the new SA scheme 2D MSDD with $[N_{\mathrm{f}}, N_{\mathrm{t}}] = [5, 2]$ improves performance by $0.6 - 0.9\,\mathrm{dB}$ over 1D MSDD with $N_{\mathrm{f}} = N = 10$. 2D MSDD with $[N_{\mathrm{f}}, N_{\mathrm{t}}] = [5, 2]$ even achieves practically the same power efficiency as 1D MSDD with $N_{\mathrm{f}} = N = 20$, i.e. double the window size, which corresponds to significant savings in decoder complexity especially in low SNR (see Section 6.7.2). We finally observe that the new SA scheme combined with 2D MSDD with $[N_{\mathrm{f}}, N_{\mathrm{t}}] = [10, 2]$ reduces the performance gap to coherent detection to no more than 1 dB.

T–DSFM and ML MSDD: Next, we consider T–DSFM and (i) compare 1D and 2D ML–MSDD, (ii) illustrate the performance improvements due to the observation window construction (OWC) proposed in Section 6.5.3.2, and (iii) further demonstrate the benefits of the new SA scheme. Fig. 6.7 shows the SNR required to achieve SER $= 10^{-5}$ as a function of the normalized fading bandwidth $B_h T_{\mathrm{f}}$ for various channel PDPs. $N_{\mathrm{T}} = 3$ transmit antennas and $N_{\mathrm{B}} = 1$ are assumed. 1D MSDD with $N_{\mathrm{t}} = N = 5$ and $N_{\mathrm{t}} = N = 10$ and 2D MSDD with $[N_{\mathrm{f}}, N_{\mathrm{t}}] = [2, 5]$ are compared. Note that for the considered case of $N_{\mathrm{B}} = 1$ the results for $N_{\mathrm{t}} = N$ are independent of the PDP. The curves for the SA schemes of [WY02, MTL05] and $N_{\mathrm{f}} = 1$ are also included, as well as the curve for coherent detection with perfect CSI. Again, the results are obtained from (6.62).

From the results for the 2–Ray channel we observe that extending the observation win-

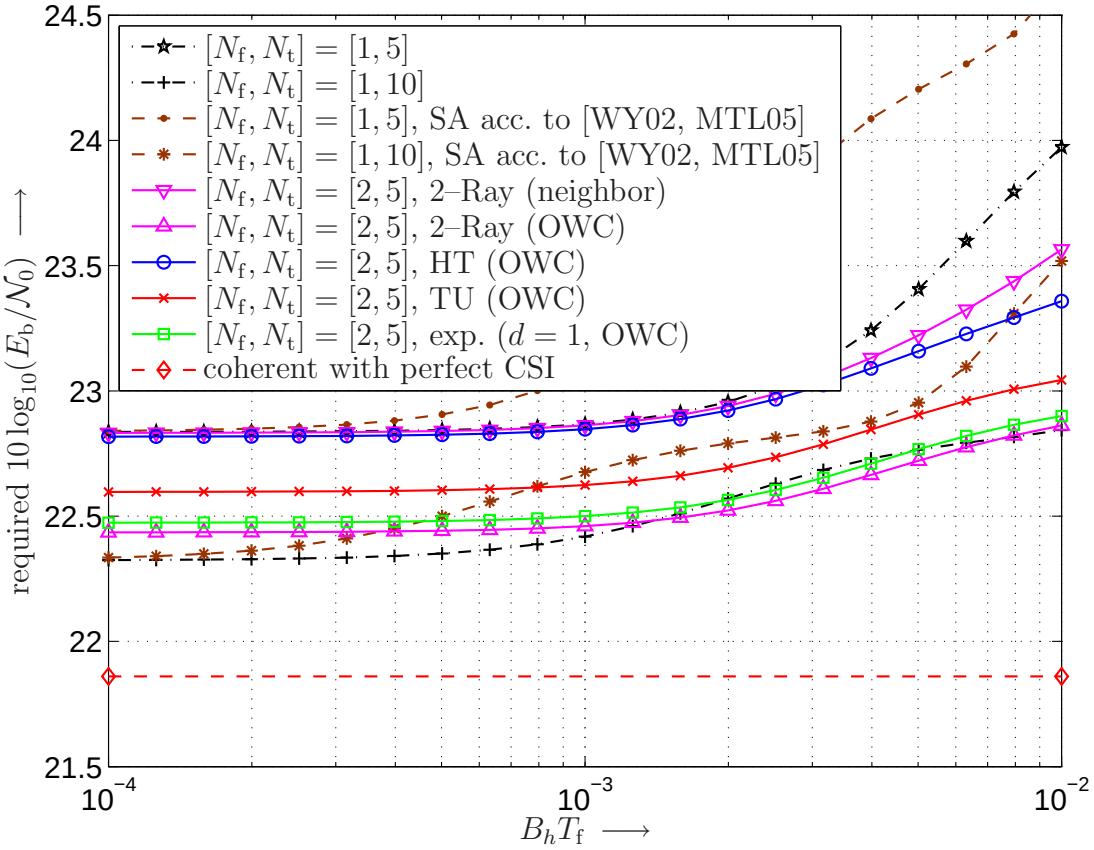

Figure 6.7: Required $10\log_{10}(E_b/\mathcal{N}_0)$ to achieve SER $= 10^{-5}$ vs. B_hT_f for various PDPs and observation windows of MSDD. Parameters: T–DSFM, $N_T = 3$, $N_B = 1$. For comparison: results for SA scheme of [WY02, MTL05] and coherent detection with perfect CSI.

dow over $N_f = 2$ *neighboring* subcarriers is only beneficial for relatively fast fading with $B_hT_f \geq 0.002$. In contrast to this, application of the proposed OWC scheme leads to consistent improvements when applying $[N_f, N_t] = [2, 5]$ compared to $[N_f, N_t] = [1, 5]$ for all fading bandwidths. Extension of the observation window in frequency direction is also beneficial for the TU and exponential PDP. For the case of the HT PDP, where the maximal channel frequency correlation is smaller than for 2–Ray, TU, and exponential PDP (see Fig. 6.5), 2D MSDD with $[N_f, N_t] = [2, 5]$ noticeably outperforms 1D MSDD with $[N_f, N_t] = [1, 5]$ only for $B_hT_f \geq 0.002$.

Comparing the respective curves for 2D MSDD and $[N_f, N_t] = [2, 5]$ with that for 1D MSDD and $N = N_t = 10$, we conclude that it is not *per se* clear that 2D MSDD is advantageous over 1D MSDD for identical total window size $N = N_fN_t$ (see also Table 6.1 for DMD [LS01]). In particular, N_f and N_t are not simply interchangeable for fixed N and the same channel correlation, but the direction of differential encoding needs to be taken into account too. Whereas increasing the observation window in direction of differential encoding lowers the error of pre-

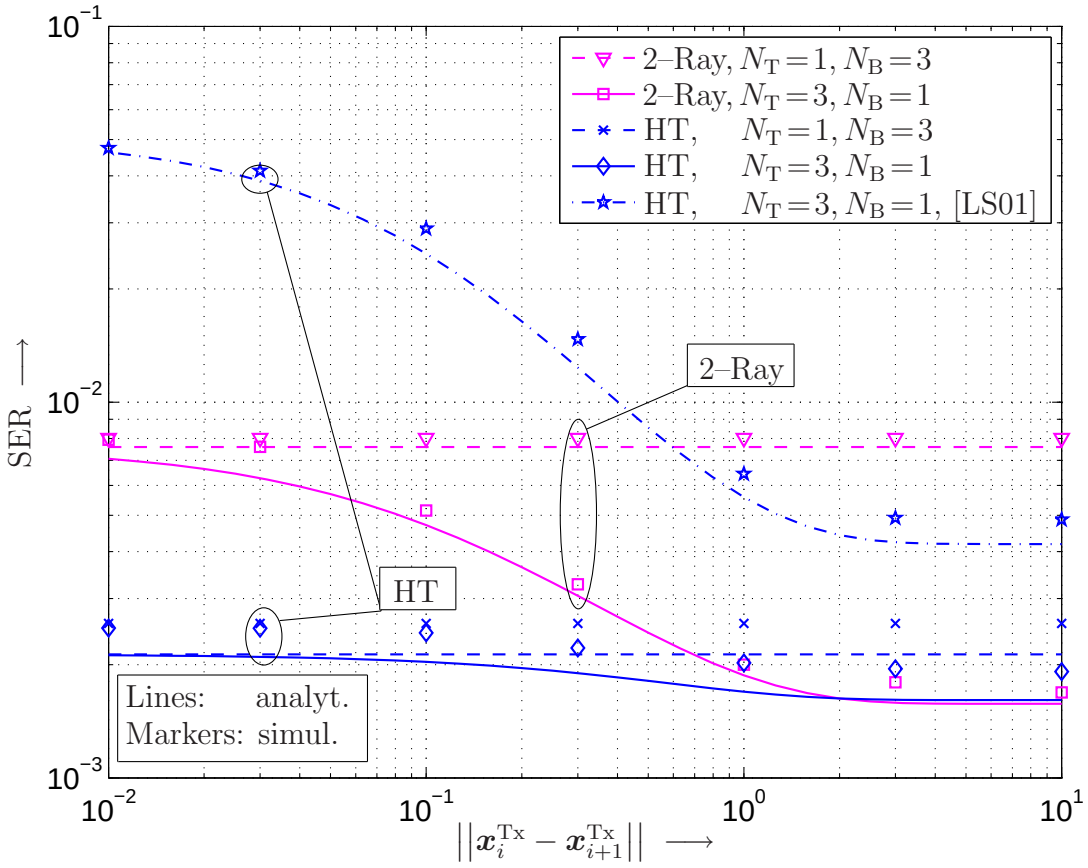

Figure 6.8: SER at $10\log_{10}(E_b/\mathcal{N}_0) = 15\,\mathrm{dB}$ vs. antenna spacing for a linear equispaced transmit–antenna array. SA scheme acc. to Section 6.4.2 for $[N_T, N_B] = [1, 3]$ and $[N_T, N_B] = [3, 1]$ and 2D MSDD with $[N_f, N_t] = [5, 2]$. Also shown: SA acc. to [LS01] for $[N_T, N_B] = [3, 1]$ and 1D MSDD with $[N_f, N_t] = [10, 1]$. Parameters: F–DSFM, 2–Ray ($\tau_2 - \tau_1 = 20\,\mu\mathrm{s}$) and HT PDP, $B_h T_f = 0.001$.

dicting the channel variation, increasing the other window dimension improves the suppression of estimation noise prior to prediction. This also emphasizes the usefulness of the SER approximations developed in Section 6.6 to guide the choice of the parameters $[N_f, N_t]$ for different channel scenarios and DSFM parameters.

A comparison of the respective results for the different SA schemes demonstrates the superior performance of the propose SA method. Finally, we note that both 1D and 2D MSDD with $N = 10$ achieve a performance within $0.4 - 1.0\,\mathrm{dB}$ of that for detection with perfect CSI even for relatively fast fading channels.

Spatial vs. Spectral Diversity: Fig. 6.8 presents analytical (lines) and simulation (markers) results for the SER of F–DSFM with ML–MSDD at $10\log_{10}(E_b/\mathcal{N}_0) = 15\,\mathrm{dB}$ as a function of the normalized inter–transmit–antenna spacing $\left|\left|\boldsymbol{x}_i^{\mathrm{Tx}} - \boldsymbol{x}_{i+1}^{\mathrm{Tx}}\right|\right|$, $1 \le i \le N_T - 1$, when as-

suming a linear equispaced transmit antenna array. 2–Ray ($\tau_2 - \tau_1 = 20\,\mu$s) and HT PDPs are considered. We compare SA according to Section 6.4.2 with $N_\mathrm{B} = N_\mathrm{S} = 3$ (dashed lines) and $N_\mathrm{T} = N_\mathrm{S} = 3$ (solid lines) (ignore the dash–dotted line for the moment), employing 2D MSDD with $[N_\mathrm{f}, N_\mathrm{t}] = [5, 2]$ in both cases.

As can be seen, the performance for $N_\mathrm{T} = N_\mathrm{S}$ converges to that for $N_\mathrm{T} = 1$ as the antenna spacing tends to zero. This is because for the new SA scheme the assignment of elements $s_{i,i}[m, k]$ to subcarriers does not depend on N_B. Hence, increasing N_T while keeping N_S fixed can only improve performance by reducing subchannel correlation, i.e. correlation among elements within $\boldsymbol{G}[m, k]$ (see Fig. 6.3). Since for the 2–Ray channel $N_\mathrm{S} = 3 > L_h = 2$, diversity order, which is $\min\{N_\mathrm{S}N_\mathrm{R}, N_\mathrm{T}N_\mathrm{R}L_h\}$ (cf. e.g. [HSL05c]), can be increased from two for $N_\mathrm{B} = N_\mathrm{S} = 3$ to three for $N_\mathrm{T} = N_\mathrm{S} = 3$, and relatively large performance gains are observed with increasing antenna spacing. For the HT PDP, on the other hand, the same diversity order of three is accomplished in both cases and thus, performance is largely independent of the antenna spacing.

To further illustrate the advantage of the proposed SA scheme, Fig. 6.8 also includes the SER curve for DMD [LS01] with $N_\mathrm{T} = N_\mathrm{S} = 3$ and the HT channel (dash–dotted line). We note that for $N_\mathrm{B} = N_\mathrm{S}$ DMD and the new SA scheme are identical. Here it is advantageous to set $[N_\mathrm{f}, N_\mathrm{t}] = [10, 1]$ due to the increased spectral spacing of symbols $s_{i,i}[m, k]$ and $s_{i,i}[m+1, k]$. Different from the proposed SA scheme, the performance with DMD significantly deteriorates for larger spatial correlation, as $s_{i,i}[m, k]$, $1 \leq i \leq N_\mathrm{T}$, are transmitted over neighboring subcarriers. Due to the increased spectral spacing of symbols captured in the observation window, the SER for $N_\mathrm{T} > 1$ is higher than that for $N_\mathrm{T} = 1$ even if there is no spatial correlation.

F–DSFM and Different Detection Algorithms: We now compare MSDSD–FS, which accomplishes ML–MSDD, and DF–MSDSD–LD–FM for 2D MSDD with observation window $[N_\mathrm{f}, N_\mathrm{t}] = [5, 2]$. For DF–MSDSD–LD–FM we insert previous decisions on the symbols $\boldsymbol{V}[m_i, k_1]$, $1 \leq i \leq N_\mathrm{f}$, in the first column of the observation window and optimize the metric only over the symbols $\boldsymbol{V}[m_i, k_2]$, $1 \leq i \leq N_\mathrm{f}$, in the second column of the observation window. This way, the depth of the search tree is cut in half, which leads to significant savings in computational complexity in low–SNR regimes (see below). CDD, DFDD(–LD) with 1D windows of size $N_\mathrm{f} = N = 2$ and $N_\mathrm{f} = N = 10$, respectively, and coherent detection with perfect CSI are considered as benchmark detectors. The SER is shown as function of the SNR for F–DSFM with $N_\mathrm{T} = 3$ and $N_\mathrm{B} = 1$ and TU, exponential ($d = 1$), and HT PDP in Figs. 6.9 a)–c), respectively. Lines and markers represent analytical and simulation results, respectively.

We observe that MSDSD–FS outperforms DFDD by about $1 - 4\,$dB and approaches the performance of coherent detection with perfect CSI within about 1 dB for all channel scenar-

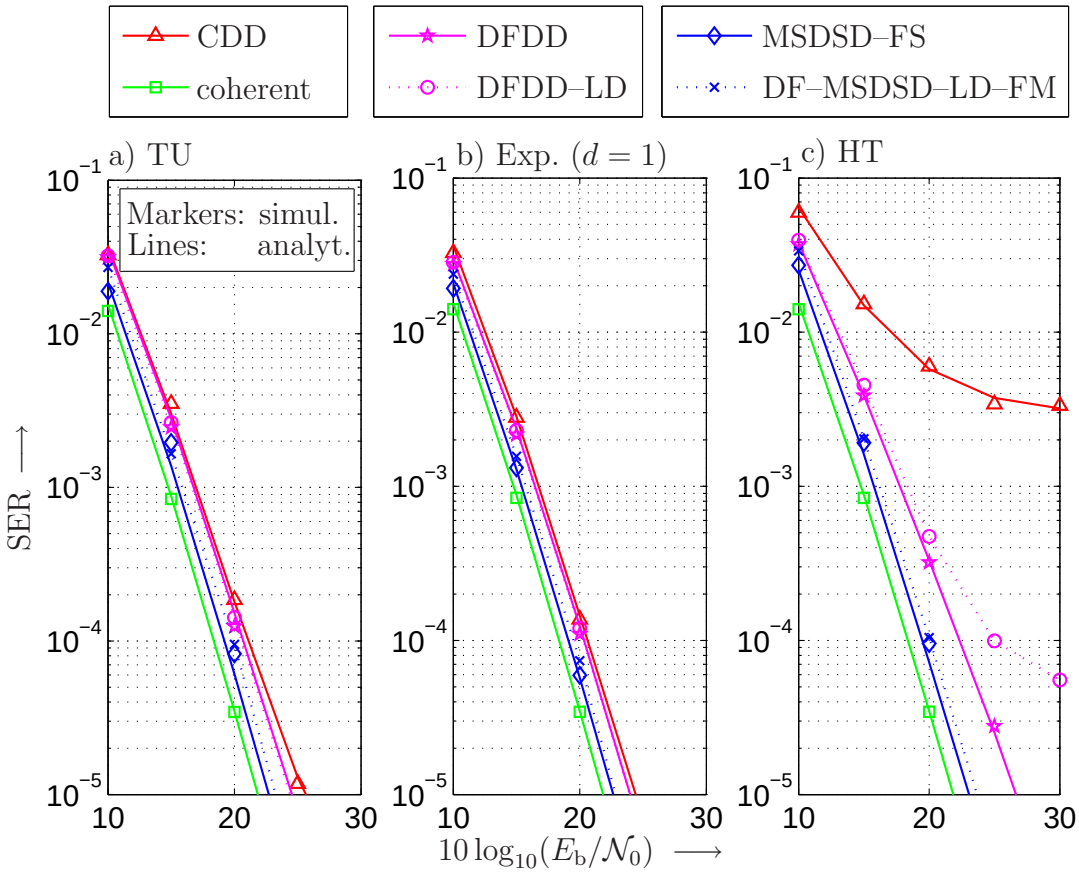

Figure 6.9: SER vs. $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ for TU, exponential ($d = 1$), and HT PDPs. F–DSFM and MSDSD and DF–MSDSD–LD–FM with $[N_\mathrm{f}, N_\mathrm{t}] = [5, 2]$, CDD and DLD, DFDD and DFDD–LD with $N = 10$ and coherent detection with perfect CSI. Parameters: $N_\mathrm{T} = 3$, $N_\mathrm{B} = 1$, and $B_h T_\mathrm{f} = 0.001$.

ios and in the SER–range of interest. One can also see that DF–MSDSD–LD–FM achieves practically the samer performance as the far more complex MSDSD–FS and provides consistent improvements over DFDD and CDD. This is especially true in highly dispersive channels, cf. Fig. 6.9 c), where CDD and computationally efficient DFDD–LD even encounter relatively high error floors. It should be noted that these gains are due to both the application of MSDD instead of DFDD and the use of a 2D as opposed to an 1D observation window.

6.7.2 Complexity

The computational complexity of MSDD, i.e. the number of candidates that need to be examined by MSDSD or its variants essentially hinges on the SNR and on the depth of the search tree. This means, that the gains of 2D MSDD over 1D MSDD with equal window size N as observed especially for F–DSFM come at practically no computational cost. It also means,

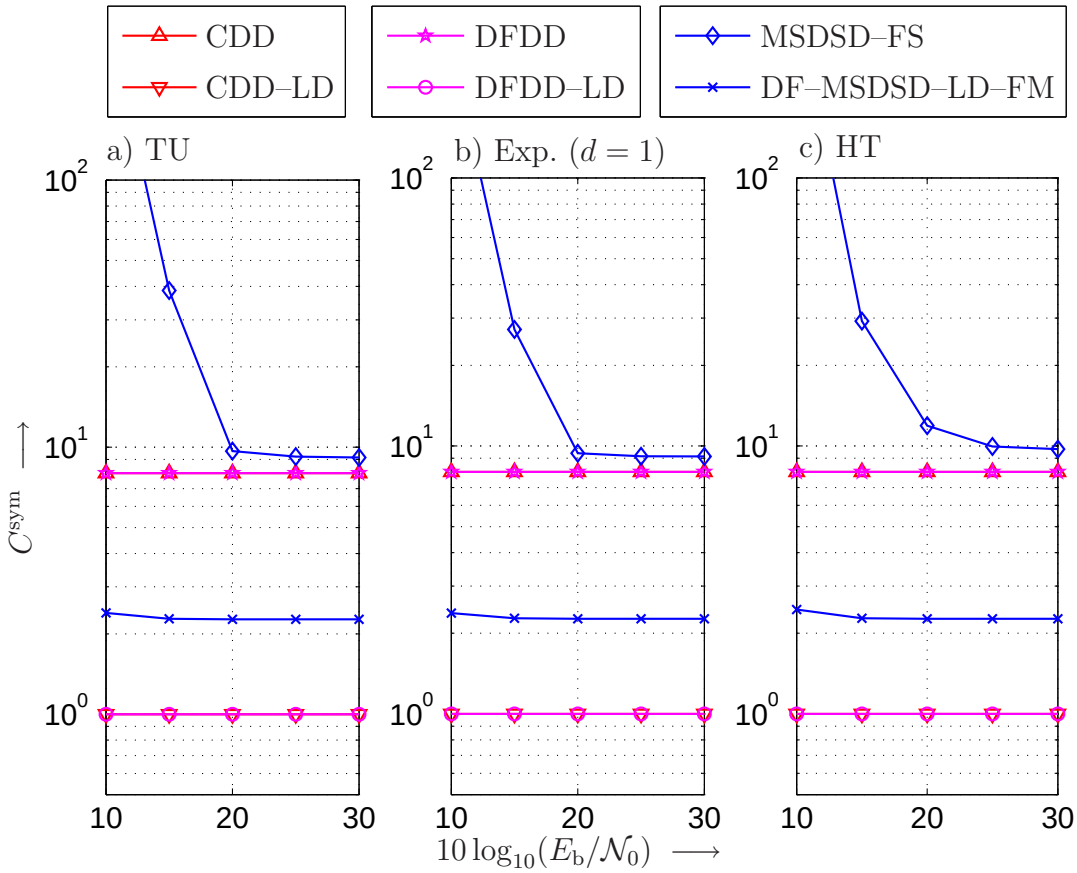

Figure 6.10: Average computational complexity vs. $10\log_{10}(E_\mathrm{b}/\mathcal{N}_0)$ for various detection algorithms and PDPs. Parameters: F–DSFM, $N_\mathrm{T} = 3$, $N_\mathrm{B} = 1$, $B_h T_\mathrm{f} = 0.001$. MSDSD–FS and DF–MSDSD–LD–FM with $[N_\mathrm{f}, N_\mathrm{t}] = [5, 2]$, DFDD and DFDD–LD with $N_\mathrm{f} = N = 10$, CDD and DLD with $N_\mathrm{f} = N = 2$.

that we can desist from presenting numerous numerical examples illustrating the complexity of tree–search–based MSDD algorithms when applied to MIMO–OFDM. We therefore restrict ourselves to exemplarily presenting complexity results corresponding to the performance comparison shown in Fig. 6.9.

F–DSFM and Different Detection Algorithms: In Fig. 6.10 computational complexities of MSDSD–FS, DF–MSDSD–LD–FM, DFDD, DFDD–LD, CDD, and CDD–LD in terms of average number C^{sym} of examined candidates per decoded symbol are compared for the same system and channel parameters as in Fig. 6.9. While the computational complexity of MSDSD–FS is quite high at low SNR, DF–MSDSD–LD–FM, which achieves practically the same power efficiency as MSDSD–FS, succeeds in keeping average complexity practically constant over the entire range of relevant SNR. In particular, the complexity of DF–MSDSD–LD–FM is very well comparable to that of DFDD–LD and CDD–LD, which is quite remarkable considering

the performance advantages of DF–MSDSD–LD–FM over DFDD–LD and the small gap in performance compared to coherent detection with perfect CSI (see Fig. 6.9). Furthermore, it can be seen from a comparison of subplots a)–c) that the PDP hardly influences the decoder complexity.

6.A Derivations and Proofs

6.A.1 Derivation of Eq. (6.16)

Starting from (6.14) and introducing for notational simplicity $\boldsymbol{A} \triangleq \sqrt{\boldsymbol{\Psi}^{\mathrm{Tx}}}$ and $\boldsymbol{B} \triangleq \sqrt{\boldsymbol{\Psi}^{\mathrm{Rx}}}^{*}$ we arrive at the product form for $\psi_{gg}[\mu, \kappa, i_1, j_1, i_2, j_2]$ in (6.19) as follows.

$$\psi_{gg}[\mu,\kappa,i_1,j_1,i_2,j_2] = \mathcal{E}\big\{g_{i_1,j_1}[m+\mu,k+\kappa]g^{*}_{i_2,j_2}[m,k]\big\} =$$

$$= \mathcal{E}\left\{\sum_{l_1=1}^{L_h}\sum_{l_2=1}^{L_h} c_{\mathrm{n}} h^{(l_1)}_{i_1,j_1}((k+\kappa)T_{\mathrm{f}}+\tau_1)\Big(c_{\mathrm{n}} h^{(l_2)}_{i_2,j_2}(kT_{\mathrm{f}}+\tau_1)\Big)^{*} \mathrm{e}^{-\mathrm{j}2\pi\Delta f((m+\mu)(\tau_{l_1}-\tau_1)-m(\tau_{l_2}-\tau_1))}\right\} \quad (6.64)$$

$$= \sum_{l=1}^{L_h} \mathcal{E}\Big\{c_{\mathrm{n}}^2 h^{(l)}_{i_1,j_1}((k+\kappa)T_{\mathrm{f}}+\tau_1)\Big(h^{(l)}_{i_2,j_2}(kT_{\mathrm{f}}+\tau_1)\Big)^{*}\Big\}\mathrm{e}^{-\mathrm{j}2\pi\Delta f\mu(\tau_l-\tau_1)} \quad (6.65)$$

$$= \sum_{l=1}^{L_h} \mathcal{E}\Big\{c_{\mathrm{n}}^2 \boldsymbol{A}_{i_1,:}\boldsymbol{W}^{(l)}((k+\kappa)T_{\mathrm{f}}+\tau_1)(\boldsymbol{B}_{j_1,:})^{\mathsf{H}}\boldsymbol{B}_{j_2,:}\Big(\boldsymbol{W}^{(l)}(kT_{\mathrm{f}}+\tau_1)\Big)^{\mathsf{H}}(\boldsymbol{A}_{i_2,:})^{\mathsf{H}}\Big\}\mathrm{e}^{-\mathrm{j}2\pi\Delta f\mu(\tau_l-\tau_1)} \quad (6.66)$$

$$= \sum_{l=1}^{L_h} \boldsymbol{A}_{i_1,:}\sum_{x=1}^{N_{\mathrm{R}}}\sum_{y=1}^{N_{\mathrm{R}}} b^{*}_{j_1,x}b_{j_2,y}\mathcal{E}\Big\{c_{\mathrm{n}}^2 \boldsymbol{W}^{(l)}_{:,x}((k+\kappa)T_{\mathrm{f}}+\tau_1)\Big(\boldsymbol{W}^{(l)}_{:,y}(kT_{\mathrm{f}}+\tau_1)\Big)^{\mathsf{H}}\Big\}(\boldsymbol{A}_{i_2,:})^{\mathsf{H}}\mathrm{e}^{-\mathrm{j}2\pi\Delta f\mu(\tau_l-\tau_1)} \quad (6.67)$$

$$= \boldsymbol{A}_{i_1,:}(\boldsymbol{A}_{i_2,:})^{\mathsf{H}}\sum_{x=1}^{N_{\mathrm{R}}} b^{*}_{j_1,x}b_{j_2,x}\psi^{\mathrm{t}}(T_{\mathrm{f}}\kappa)\sum_{l=1}^{L_h}\sigma_l^2\mathrm{e}^{-\mathrm{j}2\pi\Delta f\mu(\tau_l-\tau_1)} \quad (6.68)$$

$$= \boldsymbol{A}_{i_1,:}(\boldsymbol{A}_{i_2,:})^{\mathsf{H}}\boldsymbol{B}^{*}_{j_1,:}(\boldsymbol{B}_{j_2,:})^{\mathsf{T}}\psi^{\mathrm{t}}(T_{\mathrm{f}}\kappa)\sum_{l=1}^{L_h}\sigma_l^2\mathrm{e}^{-\mathrm{j}2\pi\Delta f\mu(\tau_l-\tau_1)} \quad (6.69)$$

$$\triangleq \psi^{\mathrm{Tx}}[i_1,i_2]\psi^{\mathrm{Rx}}[j_1,j_2]\tilde{\psi}^{\mathrm{t}}[\kappa]\psi^{\mathrm{f}}[\mu] \quad (6.70)$$

where we used the absence of correlation between channel tabs $h^{(l_1)}_{i,j}(t)$ and $h^{(l_2)}_{i,j}(t)$, $l_1 \neq l_2$, in (6.65), the spatial correlation model (4.8) in (6.66), the fact that the elements of $\boldsymbol{W}^{(l)}(t)$ are iid, the temporal correlation model (4.14) and the definition of σ_l^2 (6.15) in (6.68), and $\boldsymbol{\Psi}^{\mathrm{Tx}} = \boldsymbol{A}\boldsymbol{A}^{\mathsf{H}}$ and $\boldsymbol{\Psi}^{\mathrm{Rx}} = (\boldsymbol{B}\boldsymbol{B}^{\mathsf{H}})^{\mathsf{T}}$ and the definitions (4.9) and (4.10) in (6.70).

6.A.2 Derivation of Eq. (6.34) and Eq. (6.41)

6.A.2.1 Expression for $\tilde{\boldsymbol{\Psi}}^{\parallel}_{\boldsymbol{GG}}$

It is useful to introduce the following definitions.

$$\tilde{\boldsymbol{\Psi}}^{\mathrm{f}}[m] \triangleq \left[\psi^{\mathrm{f}}[m+(i-j)D/N_{\mathrm{S}}]\right]_{\substack{i=1,\dots,N_{\mathrm{S}}\\ j=1,\dots,N_{\mathrm{S}}}} \tag{6.71}$$

$$\tilde{\boldsymbol{\Psi}}^{\mathrm{f}} \triangleq \left[\tilde{\boldsymbol{\Psi}}^{\mathrm{f}}[m_i-m_j]\right]_{\substack{i=1,\dots,N_{\mathrm{f}}\\ j=1,\dots,N_{\mathrm{f}}}} \tag{6.72}$$

$$\tilde{\boldsymbol{\Psi}}^{\mathrm{t}} \triangleq \left[\tilde{\psi}^{\mathrm{t}}[k_i-k_j]\right]_{\substack{i=1,\dots,N_{\mathrm{t}}\\ j=1,\dots,N_{\mathrm{t}}}} \tag{6.73}$$

$$\boldsymbol{\Psi}^{\mathrm{f}} \triangleq \left[\psi^{\mathrm{f}}[m_i-m_j]\right]_{\substack{i=1,\dots,N_{\mathrm{f}}\\ j=1,\dots,N_{\mathrm{f}}}} \tag{6.74}$$

Recall from Section 6.5.1 that the channel matrix $\boldsymbol{G}[m,k]$ introduced in (6.24) contains $g_{\mathrm{mod}(i-1,N_{\mathrm{T}})+1,j}\left[(i-1)\frac{D}{N_{\mathrm{S}}}+m,k\right]$, $1 \le m \le D/N_{\mathrm{S}}$, $k \in \mathbb{N}$, in row i, $1 \le i \le N_{\mathrm{S}}$, and column j, $1 \le j \le N_{\mathrm{R}}$. With (6.70) it is then straightforward to verify that the cross correlation between any pair of vectors $\boldsymbol{G}_{:,i}[m_x,k_x]$ and $\boldsymbol{G}_{:,j}[m_y,k_y]$ reads

$$\mathcal{E}\left\{\boldsymbol{G}_{:,i}[m_x,k_x]\left(\boldsymbol{G}_{:,j}[m_y,k_y]\right)^{\mathsf{H}}\right\} = \psi^{\mathrm{Rx}}[i,j]\tilde{\psi}^{\mathrm{t}}[k_x-k_y]\tilde{\boldsymbol{\Psi}}^{\mathrm{f}}[m_x-m_y] \circ \left(\mathbf{1}_{N_{\mathrm{B}},N_{\mathrm{B}}} \otimes \boldsymbol{\Psi}^{\mathrm{Tx}}\right) \tag{6.75}$$

($\boldsymbol{A} \circ \boldsymbol{B}$: Hadamard (elementwise) product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$). Extending dimensions we can further write for the correlation between a pair of columns of $\boldsymbol{G}^{\parallel}$

$$\mathcal{E}\left\{\boldsymbol{G}^{\parallel}_{:,i}\left(\boldsymbol{G}^{\parallel}_{:,j}\right)^{\mathsf{H}}\right\} = \psi^{\mathrm{Rx}}[i,j]\tilde{\boldsymbol{\Psi}}^{\mathrm{t}} \otimes \left(\tilde{\boldsymbol{\Psi}}^{\mathrm{f}} \circ \left(\mathbf{1}_{N_{\mathrm{B}}N_{\mathrm{f}}} \otimes \boldsymbol{\Psi}^{\mathrm{Tx}}\right)\right) . \tag{6.76}$$

Finally, for the vectorized channel model as used in (6.40) we obtain

$$\tilde{\boldsymbol{\Psi}}^{\parallel}_{\boldsymbol{GG}} \triangleq \mathcal{E}\left\{\mathrm{vec}\left\{\boldsymbol{G}^{\parallel}\right\}\mathrm{vec}\left\{\boldsymbol{G}^{\parallel}\right\}^{\mathsf{H}}\right\} = \boldsymbol{\Psi}^{\mathrm{Rx}} \otimes \tilde{\boldsymbol{\Psi}}^{\mathrm{t}} \otimes \left(\tilde{\boldsymbol{\Psi}}^{\mathrm{f}} \circ \left(\mathbf{1}_{N_{\mathrm{B}}N_{\mathrm{f}}} \otimes \boldsymbol{\Psi}^{\mathrm{Tx}}\right)\right) . \tag{6.77}$$

6.A.2.2 Expression for $\boldsymbol{\Psi}^{\parallel}_{\boldsymbol{GG}}$

If there is no "transmit correlation", i.e. if there is no correlation between the elements of $\boldsymbol{G}[m,k]$, as it was assumed in the derivation of tree–search algorithms for 2D MSDD in Section 6.5.2, (6.76) reduces to

$$\mathcal{E}\left\{\boldsymbol{G}^{\parallel}_{:,i}\left(\boldsymbol{G}^{\parallel}_{:,j}\right)^{\mathsf{H}}\right\} = \psi^{\mathrm{Rx}}[i,j]\tilde{\boldsymbol{\Psi}}^{\mathrm{t}} \otimes \boldsymbol{\Psi}^{\mathrm{f}} \otimes \boldsymbol{I}_{N_{\mathrm{S}}} . \tag{6.78}$$

With this one immediately finds that

$$\mathcal{E}\left\{\boldsymbol{G}^{\parallel}\left(\boldsymbol{G}^{\parallel}\right)^{\mathsf{H}}\right\} = \sum_{i=1}^{N_{\mathrm{R}}}\mathcal{E}\left\{\boldsymbol{G}^{\parallel}_{:,i}\left(\boldsymbol{G}^{\parallel}_{:,i}\right)^{\mathsf{H}}\right\} = N_{\mathrm{R}}\tilde{\boldsymbol{\Psi}}^{\mathrm{t}} \otimes \boldsymbol{\Psi}^{\mathrm{f}} \otimes \boldsymbol{I}_{N_{\mathrm{S}}} \triangleq N_{\mathrm{R}}\boldsymbol{\Psi}^{\parallel}_{\boldsymbol{GG}} \otimes \boldsymbol{I}_{N_{\mathrm{S}}} , \tag{6.79}$$

which is the expression for the correlation matrix $\boldsymbol{\Psi}^{\parallel}_{\boldsymbol{GG}}$ used in Section 6.5.2.

6.A.3 Derivation of Eq. (6.58)

In order to prove (6.58) we first note that in the case of an 1D observation window in frequency direction

$$\boldsymbol{\Psi}^{\parallel}_{\boldsymbol{GG}} = \boldsymbol{\Psi}^{\mathrm{f}} \tag{6.80}$$

holds. Second, we once more make use of the relationship between MSDD and linear MMSE interpolation (cf. Section 2.4.2.1) and compute the nth diagonal element of $(N\times N)$–dimensional $\left(\boldsymbol{\Psi}^{\parallel}_{\boldsymbol{GG}}\right)^{-1}$ using

$$\left[\left(\boldsymbol{\Psi}^{\parallel}_{\boldsymbol{GG}}\right)^{-1}\right]_{n,n} = \frac{1}{\sigma^2_{\mathrm{i},n}}, \tag{6.81}$$

where $\sigma^2_{\mathrm{i},n}$ denotes the error variance of a linear filter interpolating the nth out of N samples of a random process, whose correlation function based on (6.80) is given by $\psi^{\mathrm{f}}[k]$ [cf. (6.18)], from the remaining $N-1$ samples.

We proceed as in the flat–fading case (cf. Appendix 5.A.4), i.e. we consider a suboptimal interpolation filter with coefficients $\boldsymbol{p}_n = [p_1, \ldots, p_{n-1}, p_{n+1}, \ldots, p_N]^{\mathsf{T}}$ and prove that it satisfies (6.58). As in Appendix 5.A.4, we can then argue that an optimal filter can not lead to a better result than (6.58).

As pointed out in Appendix 5.A.4 $\sigma^2_{\mathrm{i},n}$ can be computed through integration of the PSD of the interpolation error, i.e.

$$\sigma^2_{\mathrm{i},n} = T\int_{-1/(2T)}^{1/(2T)} \Psi^{\mathrm{f}}(f)P(f)\,\mathrm{d}f, \tag{6.82}$$

where

$$P(f) \stackrel{\triangle}{=} \left|1 - \sum_{\nu=1,\nu\neq n}^{N} p_\nu \mathrm{e}^{\mathrm{j}2\pi fT(n-\nu)}\right|^2 \tag{6.83}$$

denotes the power transfer function of the interpolation filter with coefficients $\boldsymbol{p}$ and the PSD $\Psi^{\mathrm{f}}(f)$ of the fading process to be interpolated is given by

$$\begin{aligned}
\Psi^{\mathrm{f}}(f) &\stackrel{\triangle}{=} \mathcal{F}_*\left\{\psi^{\mathrm{f}}[k]\right\}(f) && (6.84)\\
&= \sum_{l=1}^{L_h} \sigma_l^2 \mathcal{F}_*\left\{\mathrm{e}^{-\mathrm{j}2\pi k(\tau_l-\tau_1)/(DT)}\right\}(f) && (6.85)\\
&= \sum_{l=1}^{L_h} \sigma_l^2 \sum_{k=-\infty}^{\infty} \frac{1}{T}\delta\left(f - \frac{1}{T}\left(k - \frac{\tau_l-\tau_1}{DT}\right)\right) && (6.86)
\end{aligned}$$

($\mathcal{F}_* \{x[k]\}(f) \stackrel{\triangle}{=} \sum_{k=-\infty}^{\infty} x[k]\exp(-\mathrm{j}2\pi fTk)$: discrete–time Fourier transform). Plugging (6.86) into (6.82) we obtain for the interpolation error variance

$$\sigma_{\mathrm{i},n}^2 = \int_{-1/(2T)}^{1/(2T)} \sum_{l=1}^{L_h} \sigma_l^2 \sum_{k=-\infty}^{\infty} \delta\left(f - \frac{1}{T}\left(k - \frac{\tau_l - \tau_1}{DT}\right)\right) P(f)\,\mathrm{d}f \tag{6.87}$$

$$= \int_{-1/(2T)}^{1/(2T)} \sum_{l=1}^{L_h} \sigma_l^2 \delta\left(f - -\frac{1}{T}\mathrm{mod}^*\left(\frac{\tau_l - \tau_1}{DT}, 1\right)\right) P(f)\,\mathrm{d}f \tag{6.88}$$

$$= \sum_{l=1}^{L_h} \sigma_l^2 P\left(-\mathrm{mod}^*\left(\frac{\tau_l - \tau_1}{DT}, 1\right)\right). \tag{6.89}$$

If the filter coefficients are chosen according to (5.253) $P(f)$ satisfies (cf. Appendix 5.A.4)

$$P(f) \asymp c_0 \cdot (fT)^{2(N-1)}, \quad fT \to 0, \tag{6.90}$$

with some constant c_0. In consequence, we obtain for the asymptotic behavior of $\sigma_{\mathrm{i},n}^2$

$$\sigma_{\mathrm{i},n}^2 \asymp c_0 \sum_{l=1}^{L_h} \sigma_l^2 \left(\mathrm{mod}^*\left(\frac{\tau_l - \tau_1}{DT}, 1\right)\right)^{2(N-1)}, \quad \frac{\tau_{L_h} - \tau_1}{DT} \to 0, \tag{6.91}$$

and since $(\tau_{L_h} - \tau_1)/(DT) \to 0$ implies $\tau_{L_h} - \tau_1 \leq TD/2$

$$\sigma_{\mathrm{i},n}^2 \asymp c_0 \sum_{l=1}^{L_h} \sigma_l^2 \left(\frac{\tau_l - \tau_1}{DT}\right)^{2(N-1)}, \quad \frac{\tau_{L_h} - \tau_1}{DT} \to 0 \tag{6.92}$$

$$= c_0 \sum_{l=1}^{L_h} \sigma_l^2 \left(\frac{\tau_l - \tau_1}{\tau_{L_h} - \tau_1}\right)^{2(N-1)} \cdot \left(\frac{\tau_{L_h} - \tau_1}{DT}\right)^{2(N-1)}. \tag{6.93}$$

Note that

$$c_0 \sum_{l=1}^{L_h} \sigma_l^2 \left(\frac{\tau_l - \tau_1}{\tau_{L_h} - \tau_1}\right)^{2(N-1)} \tag{6.94}$$

is invariant to scaling of the delay spread if the shape of the PDP, i.e. the relative position $(\tau_l - \tau_1)/(\tau_{L_h} - \tau_1)$ and average power σ_l^2 of the individual channel tabs, remains unaltered.

As these results apply to a suboptimal filter, whose coefficients $\boldsymbol{p}_n$ are chosen according to (5.253), it is clear —by definition— that linear MMSE interpolation can not lead to higher interpolation–error variances. At the same time, as the number of degrees of freedom in the design of the filter is limited by the number $N-1$ of coefficients p_ν, there can not be a filter that achieves $\sigma_{\mathrm{i},n}^2 \asymp c' \cdot ((\tau_{L_h} - \tau_1)/(DT))^x$, $(\tau_{L_h} - \tau_1)/(DT) \to 0$ with $x > 2(N-1)$.

Due to the relationship between MSDD and linear MMSE interpolation (cf. Section 2.4.2.1) we therefore obtain in the limit of $(\tau_{L_h} - \tau_1)/(DT) \to 0$

$$\left[\left(\boldsymbol{\Psi}_{\boldsymbol{GG}}^{\parallel}\right)^{-1}\right]_{n,n} \asymp c' \cdot \left(\frac{\tau_{L_h} - \tau_1}{DT}\right)^{-2(N-1)}, \quad \frac{\tau_{L_h} - \tau_1}{DT} \to 0, \tag{6.95}$$

with some constant c', which concludes our proof of (6.58).

Chapter 7

Conclusions

In this thesis, noncoherent receivers for transmission over multiple–input multiple–output (MIMO) fading channels have been studied. In particular, we considered transmission over highly time– and / or frequency–selective channels, where coherent schemes become bandwidth inefficient due to an increasing overhead in pilot symbols.

In principle multiple–symbol differential detection (MSDD), where N received symbols are processed simultaneously to estimate (at most) $N-1$ differentially encoded data symbols, is the method of choice to achieve high power efficiency in adverse fading scenarios without the need for transmission of pilot symbols. While the performance of MSDD improves with increasing block length N, it very quickly becomes computationally intractable, as it involves finding the best out of L^{N-1}, where L is the size of the signal constellation.

Because existing practicable noncoherent detection schemes such as conventional differential detection (CDD, $N = 2$) or decision–feedback differential detection (DFDD) fail to accomplish satisfactory performance in challenging fading scenarios, the objective of this work has been to devise new noncoherent detection schemes, that provide highly reliable output at manageable receiver complexity.

To this end, we have considered two different approaches to efficiently solve the multi–dimensional search problem involved in MSDD, namely

- tree–search (TS) decoding, whose roots lie in lattice theory and sequential decoding of convolutional codes, and
- methods from combinatorial geometry, that have been deployed previously in convex quadratic maximization.

An in–depth analysis of the new schemes with respect to complexity and achievable power efficiency has been carried out.

The main results of our investigations can be summarized as follows:

- The first part of this work was concerned with the *development of practicable* noncoherent detection schemes that achieve high power efficiencies even in adverse fading channel environments.

  At this, our primary focus was on methods from TS decoding. Generalizing upon our earlier work on single–input single–output (SISO) systems a representation of the MSDD metric for a generic differential space–time modulation (DSTM) based MIMO transmission system has been developed, that is amenable to TS decoding in an $(N-1)$–dimensional tree. In TS decoding, the $(N-1)$–dimensional optimization problem is solved by means of a search in a tree of depth $N-1$, whose L^{N-1} leaves correspond to the possible solutions. Here, nodes in the tree are associated with metrics and the efficiency of these algorithms originates from the fact, that —given these metrics— the algorithm can discard entire subtrees, i.e. large numbers of candidates, simultaneously based on the examination of metrics of intermediate nodes.

  Based on a detailed comparison and classification of the numerous algorithms, that have been presented both in the lattice–decoding and the sequential–decoding literature, two promising algorithms —one from the lattice–decoding and one from the sequential–decoding literature— have been selected and applied to MSDD. The resulting implementations were referred to as multiple–symbol differential sphere detection (MSDSD) and Fano–MSDD. While this direct application has led to tremendous savings in terms of computational complexity compared to brute–force MSDD, which examines all L^{N-1} relevant candidate sequences, in high signal–to–noise ratio (SNR) regimes, the complexity still grew rapidly towards lower SNR. The reason for this effect lies in the fact, that the metric of nodes is a monotonously increasing function of the search depth, which forces the decoder to examine a great number of branches near the root of the tree. In order to overcome this, we have introduced a Fano–type path–length bias into the MSDD metric, which allows for fair comparisons between paths of different lengths in the tree and has lead to further substantial complexity savings especially in low–SNR regimes.

  An issue of great importance in TS decoding is the candidate enumeration, i.e. the order in which branches emanating from a particular node are examined. Exploiting a connection between candidate enumeration in TS–based MSDD and conventional differential detection (CDD, $N=2$), we have devised particularly efficient nested decoder structures for MSDD consisting of one outer TS decoder and $(N-1)$ inner TS decoders for candidate enumeration.

  Such receivers allowed for noncoherent detection with power efficiencies close to that of idealized coherent detection with perfect channel state information (CSI) at an *average* complexity very well comparable to that of existing, far less power–efficient noncoherent

detectors such as CDD and decision–feedback differential detection (DFDD).

A drawback of these decoders is that their instantaneous complexity is a random variable. Therefore, despite their low average complexity, a much larger maximal complexity must be allowed to achieve the desired high reliability also in adverse fading scenarios. Inspired by DFDD and a newly found connection between MSDD and linear minimum–mean squared error (MMSE) interpolation a new hybrid noncoherent detection scheme has been devised. While using a large observation window the dimension of the search space is reduced significantly based on the introduction of a number of feedback symbols into the MSDD metric. This scheme, which we refer to as decision–feedback MSDD (DF–MSDD), achieves at times even better performance than conventional MSDD, but at a complexity that is (i) practically independent of channel conditions, (ii) comparable to that of regular TS–based MSDD in high SNR, and most importantly from a practical point of view (iii) where the maximal required complexity and average complexity are practically identical.

As an alternative to the above TS–based approaches, we considered MSDD based on combinatorial geometry (CG), socalled CG–MSDD, where the detection problem is cast into a parameter space whose dimension is usually much smaller than the observation window length N. While algorithms exist, that solve the detection problem in time polynomial in N if the dimension of the parameter space is fixed, we were able to show that the dimension of the parameter space is proportional to the bandwidth of the fading process and must asymptotically grow linearly in N to guarantee close–to–optimal performance. Thus, the complexity of CG–MSDD is effectively exponential in N. Still, this detector is an interesting alternative to the above TS–based algorithms, as its complexity is (i) quite low in slow fading scenarios compared to MSDSD and (ii) practically independent of both the SNR and the instantaneous channel state.

- Furthermore, analytical investigations regarding both the complexity of the novel detection schemes and the achievable performance, have been in the focus of our work.

  First, we have derived closed–form expressions for the exact computational complexity of MSDD when implemented using either (i) the Fincke–Pohst refinement of MSDSD (FP–MSDSD) or (ii) CG–MSDD. Evaluation of these expressions facilitates the comparison of the different approaches and reasonable choices for the observation window length N without resorting to simulations. In addition, expressions for the respective asymptotic complexities have revealed that the complexities of both approaches are in principle exponential in N. Interestingly, a comparison of the respective expressions has shown that CG–MSDD is computationally highly efficient in slow–fading scenarios, whereas (FP–)MSDSD is to be preferred in rapid–fading scenarios.

Second, we have developed a detailed analytical performance study for both correlated and uncorrelated flat–fading MIMO channels. While we have resorted to standard methods from the analysis of Hermitian quadratic forms in the case of spatially correlated MIMO channels, we devised an entirely new performance analysis for the interesting special case of uncorrelated MIMO channels, that —contrary to existing literature— requires neither numerical integration nor eigenvalue decomposition of channel dependent matrices. It furthermore, allows for a number of interesting new insights into the interconnections between the performance of the different coherent and noncoherent detection schemes. These include (i) the definition of an effective SNR, which allows a unified and elegant treatment of coherent detection, CDD, DFDD, and MSDD, and (ii) insightful closed–form expression for the asymptotic performance in the limits of infinite observation window lengths and / or infinite SNR.

- Finally, we considered transmission based on orthogonal frequency division multiplex (OFDM) over frequency–selective MIMO channels. Inspired by previous work presented in the literature, we have devised a novel differential space–frequency modulation (DSFM) scheme, which makes use of spatial and / or spectral (multipath) diversity and is particularly suited for MIMO–OFDM and power–efficient, low–delay MSDD.

  We have further proposed the application of MSDD with a two–dimensional observation window, which exploits channel correlations in both time and frequency direction for detection. Based on a transformation of the 2D–MSDD problem into the form of 1D MSDD, the same efficient TS algorithms as considered for single–carrier MIMO systems were directly applicable. We have presented an analytical approximation of the SER that allows us to quickly and accurately assess the performance of DSFM with 2D MSDD. Numerical and simulation results confirmed considerable performance improvements due to (i) the new SA scheme, (ii) the use of MSDD as compared to DFDD or CDD and (iii) the application of 2D MSDD as compared to 1D MSDD. Employing TS algorithms such as MSDSD, Fano–MSDD or DF–MSDSD to implement 2D MSDD, these gains entail only moderate increases in computational complexity.

In our opinion there are a number of further interesting research topics for continuing work on low–complexity power–efficient noncoherent detection for rapid fading MIMO channels. Especially, detection in the presence of non–Gaussian interference and iterative receiver structures for coded transmission should be promising research topics.

Appendix A

List of Important Symbols and Mnemonics

In this appendix, we list important functions and symbols and frequently used mnemonics along with the page where they are introduced.

Functions and Operators

Symbol:	*Meaning:*	*Page:*
$f(x) \asymp g(x),\ x \to \hat{x}$	asymptotic equality	94
$\boldsymbol{x}^{\mathsf{T}}$, $\boldsymbol{X}^{\mathsf{T}}$	transpose of vector $\boldsymbol{x}$, matrix $\boldsymbol{X}$	13
$\boldsymbol{x}^{\mathsf{H}}$, $\boldsymbol{X}^{\mathsf{H}}$	Hermitian transpose of vector $\boldsymbol{x}$, matrix $\boldsymbol{X}$	11
$\boldsymbol{x}^{*}$, $\boldsymbol{X}^{*}$	Elementwise complex conjugate of vector $\boldsymbol{x}$, matrix $\boldsymbol{X}$	9
$\mathrm{angle}(x)$	argument of a complex number x	53
$\mathrm{argmax}_{x\in\mathcal{X}}\{f(x)\}$	returns $x \in \mathcal{X}$ that maximizes $f(x)$	12
$\mathrm{argmin}_{x\in\mathcal{X}}\{f(x)\}$	returns $x \in \mathcal{X}$ that minimizes $f(x)$	14
$\det\{\boldsymbol{X}\}$	determinant of a matrix $\boldsymbol{X}$	13
$\mathrm{diag}\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N\}$	$(NK \times NL)$ block–diagonal matrix with $(K \times L)$ $\boldsymbol{X}_n$ on main block–diagonal	7
$\delta(x-y)$	Dirac–δ function	74
e^{x}, $\exp(x)$	exponential function	7
$\mathcal{E}_x\{f(x)\}$	expectation of $f(x)$ with respect to random variable x	10
$\Gamma(x)$	complete Gamma function	88
$\gamma(A, x)$	(lower) incomplete Gamma function	136
$\circ$	Hadamard product	202
$\otimes$	Kronecker product	11

$\mathrm{mod}(x, L)$	modulo function	55
$\mathrm{mod}^*(x, L)$	modified modulo function	55
$p(x)$	probability density function	11
$p(x \,\vert\, y)$	conditional probability density function	11
$\Pr(x)$	probability	43
$\Pr(x \,\vert\, y)$	conditional probability	43
$\Phi_x(v)$	characteristic function of x	87
$\mathrm{Re}\{\boldsymbol{X}\}$, $\mathrm{Im}\{\boldsymbol{X}\}$	real and imaginary part	12
$\mathrm{Res}_{x=\hat{x}}\{f(x)\}$	residue of $f(x)$ corresponding to pole at $x = \hat{x}$	87
$\mathrm{sgn}(x)$	sign function	65
$\mathrm{toeplitz}\{x_1, \ldots, x_N\}$	$(N \times N)$ Hermitian symmetric Toeplitz matrix with $[x_1, \ldots, x_N]^{\mathsf{T}}$ in the first column	13
$\mathrm{tr}\{\boldsymbol{X}\}$	trace of a matrix	11
$\mathrm{vec}\{\boldsymbol{X}\}$	vector operator	75
$\|\|\boldsymbol{X}\|\|$	Frobenius norm of a matrix (or vector)	20
$\lceil x \rceil$	smallest integer larger than $x \in \mathbb{R}$	31
$\lfloor x \rceil$	round function (integer closest to $x \in \mathbb{R}$)	54
$:=$	assignment	14
$\stackrel{\triangle}{=}$	definition	7

Auxiliary Signs

Symbol:	*Meaning:*	*Page:*
$\hat{\cdot}$	decoder output symbol / sequence	14
$\tilde{\cdot}$	candidate symbol / sequence	14

Constants

Symbol:	*Meaning:*	*Page:*
j	imaginary unit: $\mathrm{j} \stackrel{\triangle}{=} \sqrt{-1}$	
π	the number pi: $\pi = 3.14159265358979\ldots$	
e	Euler number: $\mathrm{e} = 2.71828182845905\ldots$	
$\boldsymbol{e}_n$	nth unit vector	167
$\boldsymbol{I}_M$	$M \times M$ identity matrix	8
$\mathbf{1}_{M,N}$	$M \times N$ all–ones matrix	12

$\mathbf{0}_{M,N}$	$M \times N$ all–zeros matrix	8

Symbols

Scalars

Symbol:	*Meaning:*	*Page:*
$B_h T$	single–sided normalized fading bandwidth	74
$B_{h,\mathrm{eff}} T$	single–sided effective normalized fading bandwidth	84
C_{as}	asymptotic complexity of FP–MSDSD	141
C^{blo}	measure of decoder complexity per decoder run	128
C_n^{blo}	measure of decoder complexity per decoder run	129
C^{sym}	measure of decoder complexity per decoded symbol	128
c_i	coefficients of (di–)cyclic DSTM constellations	7
Δ_{F}	Stepsize of the Fano algorithm	31
E_{b}	average received energy per bit	80
E_{s}	average received energy per symbol	80
$\kappa_{\mathrm{L}}^{\mathrm{DF}}$	parameter of DF–MSDD	20
$\kappa_{\mathrm{U}}^{\mathrm{DF}}$	parameter of DF–MSDD	19
L	Cardinality of signal constellation	7
d_{b}	Metric of parent of current path	31
d_{c}	Metric of current path	31
d_{f}	Metric of child of the current path	31
$\delta_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$	ML MSDD branch metric	41
$d_n\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$	ML MSDD path metric	41
$\delta_{\mathrm{F},n}\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$	Fano–type MSDD branch metric	44
$d_{\mathrm{F},n}\big(\tilde{\tilde{\boldsymbol{S}}}_n\big)$	Fano–type MSDD path metric	44
$\lambda_{\boldsymbol{X},i}$	ith eigenvalue of $\boldsymbol{X}$	87
N	length of observation window	12
N'	number of decision per block returned by S–MSDD	17
$\mathcal{N}_0$	double–sided PSD of ECB noise process	73
N_{B}	parameter in DSFM	191
N_λ	number of non–zero eigenvalues of fading correlation matrix	22
$\tilde{N}_\lambda$	number of dominant eigenvalues of fading correlation matrix	99
N_{R}	number of receive antennas	10
N_{S}	dimension of unitary data and transmit symbols	7

N_{T}	number of transmit antennas	11
$\tilde{n}$	number of DSTM symbols in $\bar{\boldsymbol{S}}_n$	39
p^{des}	desired probability of $d_1\big(\bar{\boldsymbol{S}}\big) \leq \rho$ of FP–MSDSD	136
$\Psi_{\mathrm{c}}^{\mathrm{t}}(f)$	PSD of effective continuous–time fading process	84
$\Psi^{\mathrm{t}}(f)$	PSD of effective discrete–time fading process	84
R	rate in bit per channel use	80
ρ	threshold of (outer MSDD) tree–search decoder	25
ρ_{init}	initial threshold	37
ρ_{eff}	effective SNR	93
σ_n^2	variance of AWGN	11
$\zeta_{\boldsymbol{X},i}$	ith singular value of $\boldsymbol{X}$	138
T	modulation interval (1/sampling rate)	72

Matrices

Symbol:	*Meaning:*	*Page:*
$\boldsymbol{C}$	Cholesky factor of $\boldsymbol{M}$	40
$\boldsymbol{H}_n = \boldsymbol{H}[k-N+n]$	$(N_{\mathrm{S}} \times N_{\mathrm{R}})$ channel matrix	82
$\bar{\boldsymbol{H}}[k]$	$(NN_{\mathrm{T}} \times N_{\mathrm{R}})$ channel matrix	82
$\boldsymbol{G}_n = \boldsymbol{G}[k-N+n]$	$(N_{\mathrm{S}} \times N_{\mathrm{R}})$ generic channel matrix	10
$\bar{\boldsymbol{G}}[k]$	$(NN_{\mathrm{S}} \times N_{\mathrm{R}})$ generic channel matrix	13
$\boldsymbol{\Lambda}_{\boldsymbol{X}}$	contains eigenvalues of $\boldsymbol{X}$ on main diagonal	68
$\boldsymbol{M}$	$(N \times N)$ inverse fading–plus–noise acm	14
$\boldsymbol{N}_n = \boldsymbol{N}[k-N+n]$	$(N_{\mathrm{S}} \times N_{\mathrm{R}})$ AWGN matrix	10
$\bar{\boldsymbol{N}}[k]$	$(NN_{\mathrm{S}} \times N_{\mathrm{R}})$ AWGN matrix	13
$\boldsymbol{\Psi}_{\boldsymbol{XX}}$	autocorrelation matrix of a matrix $\boldsymbol{X}$	13
$\boldsymbol{\Psi}_{\boldsymbol{XX}\vert\boldsymbol{Y}}$	autocorrelation matrix of a matrix $\boldsymbol{X}$ conditioned on $\boldsymbol{Y}$	13
$\boldsymbol{\Psi}^{\mathrm{t}}$	$(N \times N)$–dimensional temporal ACM of fading process	84
$\boldsymbol{R}_n = \boldsymbol{R}[k-N+n]$	$(N_{\mathrm{S}} \times N_{\mathrm{R}})$ received matrix	10
$\bar{\boldsymbol{R}}[k]$	$(NN_{\mathrm{S}} \times N_{\mathrm{R}})$ received matrix	13
$\breve{\boldsymbol{R}}_{i,j}$		40
$\boldsymbol{S}_n = \boldsymbol{S}[k-N+n]$	$(N_{\mathrm{S}} \times N_{\mathrm{S}})$ transmit matrix	10
$\bar{\boldsymbol{S}}[k]$	$(NN_{\mathrm{S}} \times N_{\mathrm{S}})$ transmit matrix	12
$\bar{\boldsymbol{S}}_n = \big[\boldsymbol{S}_n^{\mathsf{H}}, \ldots, \boldsymbol{S}_N^{\mathsf{H}}\big]^{\mathsf{H}}$	$((N-n+1)N_{\mathrm{S}} \times N_{\mathrm{R}})$ transmit matrix	39
$\bar{\boldsymbol{S}}_{\mathrm{D}}[k]$	$(NN_{\mathrm{S}} \times NN_{\mathrm{S}})$ block–diagonal transmit matrix	13
$\boldsymbol{U}_{\boldsymbol{X}}$	$(N \times \mathrm{rank}\{\boldsymbol{X}\})$–dimensional unitary matrix whose columns	68

	are given by the eigenvectors of $(N \times N)$–dimensional $\boldsymbol{X}$	
$\boldsymbol{V}_n = \boldsymbol{V}[k-N+1+n]$	$(N_\mathrm{S} \times N_\mathrm{S})$ data matrix	10
$\bar{\boldsymbol{V}}[k]$	$((N-1)N_\mathrm{S} \times N_\mathrm{S})$ data matrix	12
$\bar{\boldsymbol{V}}_n = \left[\boldsymbol{V}_n^\mathsf{H}, \ldots, \boldsymbol{V}_{N-1}^\mathsf{H}\right]^\mathsf{H}$	$((N-n)N_\mathrm{S} \times N_\mathrm{R})$ data matrix	39
$\boldsymbol{V}^{(l)}$	lth element of the signal constellation	7
$\boldsymbol{X}_n$	reference symbol in MSDSD	40

Sets

Symbol:	*Meaning:*	*Page:*
$\boldsymbol{\mathcal{V}}$	(DSTM) signal constellation	7
$\mathbb{C}$	complex numbers	53
$\mathbb{R}$	real numbers	25
$\mathbb{Z}$	integer numbers	25

Distributions

Symbol:	*Meaning:*	*Page:*
$\mathcal{N}_\mathrm{c}(m, \sigma^2)$	circ. symm. complex Gaussian with mean m and variance σ^2	10
$\chi^2(\sigma^2, K)$	central χ^2 distribution with variance σ^2 and K degrees of freedom	42

Abbreviations

Mnemonic:	*Meaning:*	*Page:*
AWGN	additive white Gaussian noise	11
A–SpD	Agrell sphere decoder	33
BDSTM	block–DSTM	84
BID	bound intersect detect	58
BFS	best first search	26
CDD	conventional differential detection	7
CFC	continuous fading channel	81
CG–MSDD	combinatorial–geometry MSDD	63
CSI	(instantaneous) channel state information	9
DFDD	decision–feedback differential detection	12
DFE	decision–feedback equalization	26
DF–MSDD	decision–feedback MSDD	19
DFS	depth first search	26
DMD	differential modulation diversity	184
DPSK	differential phase–shift keying	8
DSFM	differential space–frequency modulation	185
DSTM	differential space–time modulation	6
ECB	equivalent complex baseband	6
F–DSFM	frequency–DSFM	192
Fano–MSDD	MSDD based on Fano algorithm with Fano–type metric	47
FP–SpD	Fincke–Pohst sphere decoder	26
FS	full search	51
HT	hilly terrain (PDP)	200
iid	independent identically distributed	10
LD	lattice decoder	55
MFS	metric first search	27
MIMO	multiple–input multiple–output	6
ML	maximum likelihood	11
MMSE	minimum mean–squared error	15
MSDD	multiple–symbol differential detection	12
S–MSDD	subset multiple–symbol differential detection	17
MSDSD	MSDD based on A–SpD with ML metric	46
MSDSD–FM	MSDD based on A–SpD with Fano–type metric	46
OFDM	orthogonal frequency division multiplexing	6

OWC	observation window construction	198
PDF	probability density function	11
PDP	power–delay profile	185
PEP	pairwise error probability	85
PSD	power spectral density	73
QSFC	quasi–static fading channel	82
SA	signal allocation	191
SpD	sphere decoder	22
SISO	single–input single–output	18
SE	Schnorr–Euchner (candidate enumeration strategy)	46
SER	symbol–error rate	21
SNR	signal–to–noise–power ratio	7
T–DSFM	time–DSFM	192
TU	typical urban (PDP)	198

Bibliography

[Ada96] F. Adachi. Reduced State Transition Viterbi Differential Detection of M–ary DPSK Signals. *IEE Electronics Letters*, 32(12):1064–1066, June 1996.

[Ada98] F. Adachi. Adaptive Differential Detection Using Linear Prediction for M–ary DPSK. *IEEE Transactions on Vehicular Technology*, 47(3):909–918, August 1998.

[AEVZ02] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger. Closest Point Search in Lattices. *IEEE Trans. Inform. Theory*, 48(8):2201–2214, August 2002.

[AH77] J. B. Anderson and C.-W. P. Ho. Architecture and Construction of a Hardware Sequential Encoder for Speech. *IEEE Trans. Commun.*, 25(7):703–707, July 1977.

[Ala98] S. M. Alamouti. A Simple Transmitter Diversity Scheme for Wireless Communications. *IEEE J. Select. Areas in Commun.*, 16(7):1451–1458, October 1998.

[AM84] J. B. Anderson and S. Mohan. Sequential Coding Algorithms: A Survey and Cost Analysis. *IEEE Trans. Commun.*, 32(2):169–176, February 1984.

[ARU00] D. Agrawal, T. J. Richardson, and R. Urbanke. Multiple–Antenna Signal Constellations for Fading Channels. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, page 365, Sorrento, Italy, June 2000.

[ARU01] D. Agrawal, T. J. Richardson, and R. Urbanke. Multiple–Antenna Signal Constellations for Fading Channels. *IEEE Trans. Inform. Theory*, 47(6):2618–2626, September 2001.

[AS72] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, New York, 9th edition, 1972.

[AS89] S. T. Andersson and N. A. B. Svensson. Noncoherent Detection of Convolutionally Encoded Continuous Phase Modulation. *IEEE J. Select. Areas in Commun.*, 7(9):1402–1414, December 1989.

[AS93] F. Adachi and M. Sawahashi. Decision Feedback Multiple–Symbol Differential Detection of M–ary DPSK. *IEE Electronics Letters*, 29(15):1385–1387, July 1993.

[AS95] F. Adachi and M. Sawahashi. Decision Feedback Differential Phase Detection of M–ary DPSK Signals. *IEEE Transactions on Vehicular Technology*, 44(2):203–210, May 1995.

[Bab86] L. Babai. On Lovász' Lattice Reduction and the Nearest Lattice Point Problem. *Combinatorica*, 6(1):1–13, 1986.

[Bar87] M. J. Barret. Error Probability for Optimal and Suboptimal Quadratic Receivers in Rapid Rayleigh Fading Channels. *IEEE J. Select. Areas in Commun.*, 5(4):302–304, February 1987.

[Bau92] D. Bauer. Einheitliche Beschreibung Sequentieller Detektionsverfahren für gestörte Impulsinterferenzbehaftete Datenübertragungssysteme. *Archiv für Elektronik und Übertragungstechnik (International Journal of Electronics)*, 46(1):39–53, 1992. (In German).

[BB05] M. Borgmann and H. Bölcskei. Noncoherent Space–Frequency Coded MIMO–OFDM. *IEEE J. Select. Areas in Commun.*, 23:1799–1810, September 2005.

[BBP03] H. Bölcskei, M. Borgmann, and A. J. Paulraj. Impact of Propagation Environment on the Performance of Space–Frequency Coded MIMO–OFDM. *IEEE J. Select. Areas in Commun.*, 21:427–439, April 2003.

[BCTV98] E. Biglieri, G. Caire, G. Taricco, and J. Ventura–Traveset. Computing Error Probabilities over Fading Channels: A Unified Approach. *European Transactions on Telecommunications*, 9:15–25, 1998.

[BGBF03] J. J. Boutros, N. Gresset, L. Brunel, and M. Fossorier. Soft–Input Soft–Output Lattice Sphere Decoder for Linear Channels. In *Proc. of IEEE Global Telecommunications Conference (Globecom)*, volume 3, pages 1583–1587, San Francisco, CA, November 2003.

[Bin90] J. A. C. Bingham. Multicarrier Modulation for Data Transmission: An Idea whose Time has come. *IEEE Communications Magazine*, 28(5):5–14, May 1990.

[BKS92] J. Benndorf, D. L. Korobkow, and A. G. Schwanja. Matched Signals in the Presence of Intersymbol Interference. *Archiv für Elektronik und Übertragungstechnik (International Journal of Electronics)*, 46(6):409–414, 1992.

[BN62] P. A. Bello and B. D. Nelin. The Influence of Fading Spectrum on the Binary Error Probabilities of Incoherent and Differentially Coherent Matched Filter Receivers. *IRE Transactions on Communication Systems*, 10:160–168, June 1962.

[BS02] B. Bhukania and P. Schniter. Multiple–Symbol Detection of Differential Unitary Space–Time Modulation in Fast–Fading Channels with Known Correlation. In

Conference on Information Sciences and Systems (CISS), Princeton University, March 2002.

[Buc43] R. C. Buck. Partition of Space. *American Mathematical Monthly*, 50(9):541–544, November 1943.

[BV01] M. Brehler and M. K. Varanasi. Asymptotic Error Probability Analysis of Quadratic Receivers in Rayleigh–Fading Channels With Application to a Unified Analysis of Coherent and Noncoherent Space–Time Receivers. *IEEE Trans. Inform. Theory*, 47(6):2383–2399, September 2001.

[BZRF06] S. Bittner, E. Zimmermann, W. Rave, and G. Fettweis. List Sequential MIMO Detection: Noise Bias Term and Partial Path Augmentation. In *Proc. of IEEE International Conference on Communications (ICC)*, volume 3, pages 1300–1305, Istanbul, Turkey, June 2006.

[Cav91] J. K. Cavers. An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels. *IEEE Transactions on Vehicular Technology*, 40(4):686–693, November 1991.

[Cav92] J. K. Cavers. On the Validity of the Slow and Moderate Fading Models for Matched Filter Detection of Rayleigh Fading Signals. *Canadian J. Elect. & Comp. Eng.*, 17(4):183–189, October 1992.

[CH92] J. K. Cavers and P. Ho. Analysis of the Error Performance of Trellis–Coded Modulations in Rayleigh–Fading Channels. *IEEE Trans. Commun.*, 40(1):74–83, January 1992.

[Cla68] R. H. Clarke. A Statistical Theory of Mobile–Radio Reception. *The Bell Systems Technical Journal*, 47:957–1000, July 1968.

[CLL05] P.-H. Chiang, D.-B. Lin, and H.-J. Li. Performance of 2IMO Differentially Transmit Diversity Block Coded OFDM Systems in Doubly Selective Channels. In *Proc. of IEEE Global Telecommunications Conference (Globecom)*, pages 3768–3773, Saint Louis, November 2005.

[CMH00] C. Cozzo, D. N. Mohile, and B. L. Hughes. Extension to the Theory of Differential Space–Time Modulation. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, page 285, Sorrento, Italy, June 2000.

[Coh93] H. Cohen. *A Course in Computational Algebraic Number Theory*. Springer–Verlag, Berlin, Germany, 1993.

[CR99] G. Colavolpe and R. Raheli. Noncoherent Sequence Detection. *IEEE Trans. Commun.*, 47(9):1376–1385, September 1999.

[CRTP03] N. Chiurtu, B. Rimoldi, E. Telatar, and V. Pauli. Impact of Correlation and Coupling on the Capacity of MIMO Systems. In *Proc. of the 3rd IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 154–157, Darmstadt, December 2003.

[CS99] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*. Springer–Verlag, New York, NY, 3rd edition, 1999.

[CS02] M.-X. Chang and Y. T. Su. Model–Based Channel Estimation for OFDM Signals in Rayleigh Fading. *IEEE Trans. Commun.*, 50(4):540–544, April 2002.

[CS04] M.-X. Chang and Y. T. Su. Blind and Semiblind Detections of OFDM Signals in Fading Channels. *IEEE Trans. Commun.*, 52(5):744–754, May 2004.

[CSZ01] K. L. Clarkson, W. Sweldens, and A. Zheng. Fast Multiple Antenna Differential Decoding. *IEEE Trans. Commun.*, 49(2):253–261, February 2001.

[CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.

[CT04a] T. Cui and C. Tellambura. An Efficient Generalized Sphere Decoder for Rank–Deficient MIMO Systems. In *Proc. of IEEE Vehicular Technology Conference (VTC)*, volume 5, pages 3689–3693, Los Angeles, CA, September 2004.

[CT04b] T. Cui and C. Tellambura. Joint Channel Estimation and Data Detection for OFDM Systems via Sphere Decoding. In *Proc. of IEEE Global Telecommunications Conference (Globecom)*, Dallas, TX, November 2004.

[CT05a] T. Cui and C. Tellambura. Bound–Intersection Detection for Multiple–Symbol Differential Unitary Space–Time Modulation. *IEEE Trans. Commun.*, 53(12):2114–2123, December 2005.

[CT05b] T. Cui and C. Tellambura. Multiple–Symbol Differential Unitary Space–Time Demodulation with Reduced–Complexity. In *Proc. of IEEE International Conference on Communications (ICC)*, Seoul, May 2005.

[CT06] T. Cui and C. Tellambura. Multiple–Symbol Differential Detection for Single–Antenna and Multiple–Antenna Systems over Ricean–Fading Channels. In *Proc. of IEEE International Conference on Communications (ICC)*, volume 3, pages 1439–1444, Istanbul, Turkey, June 2006.

[DADSC02] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank. Differential Space–Time Coding for Frequency–Selective Channels. *IEEE Communications Letters*, 6:253–255, June 2002.

[DB06] Z. Du and N. C. Beaulieu. Decision–Feedback Detection for Block Differential Space–Time Modulation. *IEEE Trans. Commun.*, 54(5):900–910, May 2006.

[DEC03] O. Damen, H. El Gamal, and G. Caire. On Maximum Likelihood Detection and the Search for the Closest Lattice Point. *IEEE Trans. Inform. Theory*, 49(10):2389–2402, October 2003.

[Die75] U. Dieter. How to Calculate Shortest Vectors in a Lattice. *Mathematics of Computation*, 29(131):827–833, July 1975.

[DS90] D. Divsalar and M. K. Simon. Multiple–Symbol Differential Detection of MPSK. *IEEE Trans. Commun.*, 38(3):300–308, March 1990.

[DS94] D. Divsalar and M. K. Simon. Maximum–Likelihood Differential Detection of Uncoded and Trellis Coded Amplitude Phase Modulation over AWGN and Fading Channels—Metrics and Performance. *IEEE Trans. Commun.*, 42(1):76–89, January 1994.

[DSS90] D. Divsalar, M. K. Simon, and M. Shahshahani. The Performance of Trellis–Coded MDPSK with Multiple Symbol Detection. *IEEE Trans. Commun.*, 38(9):1391–1403, September 1990.

[Edb92] F. Edbauer. Bit Error Rate of Binary and Quarternary DPSK Signals with Multiple Differential Feedback Detection. *IEEE Trans. Commun.*, 40(3):457–460, March 1992.

[Ede87] H. Edelsbrunner. *Algorithms of Combinatorial Geometry*. Springer–Verlag, Berlin, 1987.

[EOS86] H. Edelsbrunner, J. O'Rourke, and R. Seidel. Constructing Arrangements of Lines and Hyperplanes with Applications. *SIAM Journal on Computing*, 15(2):341–363, May 1986.

[ER93] V. Engels and H. Rohling. Multilevel Differential Modulation Techniques (64-DAPSK) for Multicarrier Transmission Systems. *European Transactions on Telecommunications*, 6:633–640, Nov.-Dec. 1993.

[Eur95] European Telecommunications Standards Institute (ETSI). Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to Mobile, Portable and Fixed Receivers. ETSI ETS 300 401, February 1995.

[Eur01] European Telecommunications Standards Institute (ETSI). Digital Cellular Telecommunications System (Phase 2+); Radio Transmission and Reception; (3GPP TS 05.05 version 8.9.0 Release 1999). ETSI TS 100 910, April 2001.

[Fan63] R. M. Fano. A Heuristic Discussion of Probabilistic Decoding. *IEEE Trans. Inform. Theory*, 9:64–74, April 1963.

[FF] K. Fukuda and J.-A. Ferrez. Implementations of LP–Based Reverse Search Algorithms for the Zonotope Construction and the Fixed–Rank Convex Quadratic Maximization in Binary Variables Using the ZRAM and the cddlib Libraries. http://www.cs.mcgill.ca/∼fukuda/download/mink/RS_TOPE020713.tar.gz.

[FFL] J.-A. Ferrez, K. Fukuda, and Th. M. Liebling. Solutions to Random Instances of the 01QP obtained by the Parallel Zonotope Construction Code rs_tope.c. http://www.cs.mcgill.ca/∼fukuda/download/cutzono_solutions.tar.gz.

[FFL05] J.-A. Ferrez, K. Fukuda, and Th. M. Liebling. Solving the Fixed Rank Convex Quadratic Maximization in Binary Variables by a Parallel Zonotope Construction Algorithm. *European Journal of Operation Research*, 166(1):35–50, 2005.

[FP85] U. Fincke and M. Pohst. Improved Methods for Calculating Vectors of Short Length in a Lattice, Including a Complexity Analysis. *Mathematics of Computation*, 44:463–471, April 1985.

[FS99] P. K. Frenger and N. A. B. Svensson. Decision–Directed Coherent Detection in Multicarrier Systems on Rayleigh Fading Channels. *IEEE Transactions on Vehicular Technology*, 48(2):490–498, March 1999.

[Gal74] R. G. Gallager. Tree Encoding for Symmetric Sources with a Distortion Measure. *IEEE Trans. Inform. Theory*, 20(1):65–76, January 1974.

[Gei73] J. M. Geist. Search Properties of Some Sequential Decoding Algorithms. *IEEE Trans. Inform. Theory*, 19(4):519–526, April 1973.

[GN04] Z. Guo and P. Nilsson. Reduced Complexity Schnorr–Euchner Decoding Algorithms For MIMO Systems. *IEEE Trans. Commun.*, 8(5):286–288, May 2004.

[Gra71] R. M. Gray. Toeplitz and Circulant Matrices: A Review. *Stanford Electron. Lab., Tech. Rep 6502-1*, June 1971.

[GS58] U. Grenander and G. Szegö. *Toeplitz Forms and Their Applications*. University of California Press, Berkley and Los Angeles, 1958.

[GS02] G. Ganesan and P. Stoica. Differential Modulation Using Space–Time Block Codes. *IEEE Signal Processing Letters*, 9:57–60, February 2002.

[GvL96] G. H. Golub and C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, USA, 3rd edition, 1996.

[Hay96] S. Haykin. *Adaptive Filter Theory*. Prentice–Hall, Englewood Cliffs, NJ, third edition, 1996.

[Her50] C. Hermite. Extraits de Lettres à M. Jacobi sur Différents Objets de la Théorie des Nombres. *J. Reine und Angewandte Mathematik*, 40(3-4):261–315, 1850. (in French).

[HF92] P. Ho and D. Fung. Error Performance of Multiple–Symbol Differential Detection of PSK Signals Transmitted over Correlated Rayleigh Fading Channels. *IEEE Trans. Commun.*, 40:25–29, October 1992.

[HH02] B. Hassibi and B. M. Hochwald. Cayley Differential Unitary Space–Time Codes. *IEEE Trans. Inform. Theory*, 48(6):1485–1503, June 2002.

[HHSS00] B. Hassibi, B. M. Hochwald, A. Shokrollahi, and W. Sweldens. Codes for Differential Signaling with Many Antennas. In *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pages 23–24, Chicago, IL, September 2000.

[HJ85] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, United Kingdom, 1985.

[HJ91] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, United Kingdom, 1991.

[HK03] E. Haas and S. Kaiser. Two–Dimensional Differential Demodulation for OFDM. *IEEE Trans. Commun.*, 51(4):580–586, April 2003.

[HKR97] P. Höher, S. Kaiser, and P. Robertson. Two–Dimensional Pilot–Symbol–Aided Channel Estimation By Wiener Filtering. *Proc. of IEEE Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, pages 1845–1848, April 1997.

[HM00] B. M. Hochwald and T. L. Marzetta. Unitary Space–Time Modulation for Multiple–Antenna Communications in Rayleigh Flat Fading. *IEEE Trans. Inform. Theory*, 46:543–564, March 2000.

[HM01] B. M. Hochwald and T. L. Marzetta. Space–Time Autocoding. *IEEE Trans. Inform. Theory*, 47(7):2761–2781, November 2001.

[HMR+00] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke. Systematic Design of Unitary Space–Time Constellations. *IEEE Trans. Inform. Theory*, 46:1962–1973, September 2000.

[HS00] B. M. Hochwald and W. Sweldens. Differential Unitary Space–Time Modulation. *IEEE Trans. Commun.*, 48(12):2041–2052, December 2000.

[HSL05a] T. Himsoon, W. Su, and K. J. R. Liu. Differential Unitary Space–Time Signal Design Using Matrix Rotation Structure. *IEEE Signal Processing Letters*, 12(1):45–48, January 2005.

[HSL05b] T. Himsoon, W. Su, and K. J. R. Liu. Multiband Differential Modulation for UWB Communication Systems. In *Proc. of IEEE Global Telecommunications Conference (Globecom)*, November 2005.

[HSL05c] T. Himsoon, W. Su, and K. J. R. Liu. Single–Block Differential Transmit Scheme for Frequency Selective MIMO–OFDM Systems. In *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pages 532–537, March 2005.

[Hub97] J. B. Huber. *Systemtheoretische Grundlagen der Modellierung von Mobilfunkkanälen (in German)*. Technical Report, University of Erlangen–Nürnberg, 1997.

[Hug99a] B. L. Hughes. Differential Space–Time Modulation. In *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pages 145–149, New Orleans, LA, September 1999.

[Hug99b] B. L. Hughes. Space–Time Coding without Channel Estimation. In *Conference on Information Sciences and Systems (CISS)*, Johns Hopkins University, Baltimore, MD, March 1999.

[Hug00a] B. L. Hughes. Differential Space–Time Modulation. *IEEE Trans. Inform. Theory*, 46(7):2567–2578, November 2000.

[Hug00b] B. L. Hughes. Further Results on Differential Space–Time Modulation. In *Proc. of IEEE Sensor Array and Multichannel Signal Processing Workshop*, Cambridge, MA, March 2000.

[Hug03] B. L. Hughes. Optimal Space–Time Constellations from Groups. *IEEE Trans. Inform. Theory*, 49(2):401–410, February 2003.

[HV02] B. Hassibi and H. Vikalo. On the Expected Complexity of Integer Least–Squares Problems. In *Proc. of IEEE Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, pages 1497–1500, May 2002.

[HV05] B. Hassibi and H. Vikalo. On the Sphere Decoding Algorithm I: Expected Complexity. *IEEE Transactions on Signal Processing*, 53(8):2806–2818, August 2005.

[JBM75] F. Jelinek, L. R. Bahl, and R. L. Mercer. Design of a Linguistic Statistical Decoder for the Recognition of Continuous Speech. *IEEE Trans. Inform. Theory*, 21(3):250–256, May 1975.

[Jel69] F. Jelinek. A Fast Sequential Decoding Algorithm Using a Stack. *IBM J. Res. and Dev.*, 13:675–685, November 1969.

[JO05a] J. Jaldén and B. Ottersten. On the Complexity of Sphere Decoding in Digital Communications. *IEEE Transactions on Signal Processing*, 53(4):1474–1484, April 2005.

[JO05b] J. Jaldén and B. Ottersten. Parallel Implementation of a Soft Output Sphere Decoder. In *Proc. 39th Asilomar Conference on Signals, Systems, and Computers*, pages 581–585, Pacific Grove, CA, October 2005.

[Jou93] A. Joux. *La Réduction de Réseaux en Cryptographie*. PhD thesis, École Polytechnique, Palaiseau, 1993. (in French).

[JZ99] R. Johannesson and K.Sh. Zigangirov. *Fundamentals of Convolutional Coding*. IEEE Press, Piscataway, N.J., Febr. 1999.

[Kai60] T. Kailath. Correlation Detection of Signals Perturbed by a Random Channel. *IRE Transactions on Information Theory*, 6:361–366, June 1960.

[Kay98] S. M. Kay. *Fundamentals of Statistical Signal Processing: Detection Theory*, volume II. Prentice–Hall, Upper Saddle River, NJ, USA, 1998.

[KHR05] J. H. Kim, P. K. M. Ho, and M. L. B. Riediger. Suboptimal Multiple–Symbol Differential Detection of MPSK with Diversity Reception. *IEE Proceedings–Communications*, 152(4):469–475, August 2005.

[KL94] R. Knopp and H. Leib. M–ary Phase Coding for the Noncoherent AWGN Channel. *IEEE Trans. Inform. Theory*, 40(6):1968–1984, November 1994.

[Kuh06] C. Kuhn. *Detection, Decoding and Estimation with a List–Sequential (LISS) Algorithm*. PhD thesis, Technische Universität München, Munich, Germany, May 2006.

[LCS98] Y. Li, L. J. Cimini, and N. R. Sollenberger. Robust Channel Estimation for OFDM Systems with Rapid Dispersive Fading Channels. *IEEE Trans. Commun.*, 46(7):902–915, July 1998.

[LHC04] S. Liu, G. Huang, and J.-W. Chong. Performance of Differential Unitary Space–Time Modulation over MIMO–OFDM Systems. In *Proc. of International Conference on Communications, Circuits and Systems*, pages 140–144, Chengdu, June 2004.

[Li03] H. Li. Differential Space–Time–Frequency Modulation Over Frequency–Selective Fading Channels. *IEEE Communications Letters*, 7(8):349–351, August 2003.

[LLK04] C. Ling, K. H. Li, and A. C. Kot. On Decision–Feedback Detection of Differential Space–Time Modulation in Continuous Fading. *IEEE Trans. Commun.*, 52(10):1613–1617, October 2004.

[LLKS85] E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys. *The Traveling Salesman Problem*. John Wiley & Sons, Chichester, 1985.

[LLL82] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring Polynomials with Rational Coefficients. *Mathematische Annalen*, pages 515–534, 1982.

[LLLL01a] J. Liu, J. Li, H. Li, and E. G. Larsson. Differential Space–Time Modulation for Interference Suppression. *IEEE Transactions on Signal Processing*, 49:1786–1795, August 2001.

[LLLL01b] J. Liu, J. Li, H. Li, and E. G. Larsson. High–Rate Differential Space–Time Modulation for Interference Suppression. In *IEEE 3rd Workshop on Signal Processing Advances in Wireless Communications (SPWAC)*, pages 283–286, March 2001.

[LM90] J. H. Lodge and M. L. Moher. Maximum Likelihood Sequence Estimation of CPM Signals Transmitted Over Rayleigh Flat–Fading Channel. *IEEE Trans. Commun.*, 38:787–794, June 1990.

[LMLK05] C. Ling, W. H. Mow, K. H. Li, and A. C. Kot. Multiple–Antenna Differential Lattice Decoding. *IEEE J. Select. Areas in Commun.*, 23:1821–1829, September 2005.

[LP88] H. Leib and S. Pasupathy. The Phase of a Vector Pertubed by Gaussian Noise and Differentially Coherent Receivers. *IEEE Trans. Inform. Theory*, 34(6):1491–1501, November 1988.

[LP91] H. Leib and S. Pasupathy. Optimal Noncoherent Block Demodulation of Differential Phase Shift Keying (DPSK). *Archiv für Elektronik und Übertragungstechnik*, 45(5):299–305, 1991.

[LP92] H. Leib and S. Pasupathy. Noncoherent Block Demodulation of MSK with Inherent and Enhanced Encoding. *IEEE Trans. Commun.*, 40(9):1430–1441, November 1992.

[LRV98] M. Luise, R. Reggiannini, and G. M. Vitetta. Blind Equalization/Detection for OFDM Signals over Frequency–Selective Channels. *IEEE J. Select. Areas in Commun.*, 16:1568–1578, October 1998.

[LS01] L. Lampe and R. Schober. Differential Modulation Diversity for OFDM. In *Proc. of 6th International OFDM–Workshop (InOWo)*, pages 19–1–19–6, Hamburg, September 2001.

[LSPW04] L. Lampe, R. Schober, V. Pauli, and C. Windpassinger. Multiple–Symbol Differential Sphere Decoding. *Proc. of IEEE International Conference on Communications (ICC)*, 2004.

[LSPW05] L. Lampe, R. Schober, V. Pauli, and C. Windpassinger. Multiple–Symbol Differential Sphere Decoding. *IEEE Trans. Commun.*, 53:1981–1985, December 2005.

[LW66] E. L. Lawler and D. E. Wood. Branch and Bound Methods: A Survey. *Operations Research*, 14:699–719, December 1966.

[LW90] Y. M. Lam and P. H. Wittke. Frequency–Hopped Spread–Spectrum Transmission with Bandwidth–Efficient Modulations and Simplified Noncoherent Sequence Estimation. *IEEE Trans. Commun.*, 38:2184–2196, December 1990.

[LWZD04] S. Lv, G. Wei, J. Zhu, and Z. Du. Differential Unitary Space–Time Modulation in Fast Fading Channel. In *Proc. of IEEE Vehicular Technology Conference (VTC)*, Los Angeles, September 2004.

[LX00] X.-B. Liang and X.-G. Xia. Some Unitary Signal Constellations for Differential Space–Time Modulation. In *Proc. 34th Asilomar Conference on Signals, Systems, and Computers*, pages 716–720, Pacific Grove, CA, October 2000.

[LX01] X.-B. Liang and X.-G. Xia. A Class of Two by Two Unitary Signal Constellations for Differential Space–Time Modulation. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, page 81, Washington, D.C., June 2001.

[LX02] X.-B. Liang and X.-G. Xia. Unitary Signal Constellations for Differential Space–Time Modulation with two Transmit Antennas: Parametric Codes, Optimal Designs, and Bounds. *IEEE Trans. Inform. Theory*, 48(8):2291–2322, August 2002.

[LXG01] Z. Liu, Y. Xin, and G. B. Giannakis. Linear Constellation–Precoding for OFDM with Maximum Multipath Diversity and Coding Gain. In *Proc. 35th Asilomar Conference on Signals, Systems, and Computers*, pages 1445–1449, Pacific Grove, November 2001.

[LXG02] Z. Liu, Y. Xin, and G. B. Giannakis. Space–Time–Frequency Coded OFDM over Frequency–Selective Fading Channels. *IEEE Transactions on Signal Processing*, 50:2465–2476, October 2002.

[LXW+05] A. Li, W. Xu, Y. Wang, Z. Zhou, and J. Wang. A Faster ML Sphere Decoder with Competing Branches. In *Proc. of IEEE Vehicular Technology Conference (VTC)*, volume 1, pages 438–441, Stockholm, May 2005.

[MA91] S. Mohan and J. B. Anderson. *Source and Channel Coding: An Algorithmic Approach*. Kluwer Academic Publishers, Norwell, MA, 1 edition, 1991.

[Mac94] K. M. Mackenthun, Jr. A Fast Algorithm for Multiple–Symbol Differential Detection of MPSK. *IEEE Trans. Commun.*, 42(2/3/4):1471–1474, February/March/April 1994.

[MAKA07] I. Motedayen-Aval, A. Krishnamoorthy, and A. Anastasopoulos. Optimal Joint Detection/Estimation in Fading Channels with Polynomial Complexity. *IEEE Trans. Inform. Theory*, 53(1):209–223, January 2007.

[Mar99] T. L. Marzetta. Blast Training: Estimating Channel Characteristics for High Capacity Space–Time Wireless. In *Annual Allerton Conference on Communications, Control, and Computing*, pages 958–966, Monticello, IL, USA, September 1999.

[Mas72] J. L. Massey. Variable–Length Codes and the Fano Metric. *IEEE Trans. Inform. Theory*, 18(1):196–198, January 1972.

[MB84] J. P. McGeehan and A. J. Bateman. Phase–Locked Transparent Tone–in–Band (TTIB). *IEEE Trans. Commun.*, 32:81–87, January 1984.

[MBA81] J. W. Modestino, V. Bhaskaran, and J. B. Anderson. Tree Encoding of Images in the Presence of Channel Errors. *IEEE Trans. Inform. Theory*, 27(6):677–697, November 1981.

[MC06] G. Manglani and A. K. Chaturvedi. Application of Computational Geometry to Multiuser Detection in CDMA. *IEEE Trans. Commun.*, 54(2):204–207, February 2006.

[MD97] U. Mengali and A. N. D'Andrea. *Synchronization Techniques for Digital Receivers*. Plenum Press, New York, 1997.

[MF90] D. Makrakis and K. Feher. Optimal Noncoherent Detection of PSK Signals. *IEE Electronics Letters*, 26(6):398–400, March 1990.

[MGDC06] A. D. Murugan, H. El Gamal, M. O. Damen, and G. Caire. A Unified Framework for Tree Search Decoding: Rediscovering the Sequential Decoder. *IEEE Trans. Inform. Theory*, 52(3):933–953, March 2006.

[MHH00] T. L. Marzetta, B. M. Hochwald, and B. Hassibi. Space–Time Autocoding: Arbitrarily Reliable Communication in a Single Fading Interval. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, page 313, Sorrento, Italy, June 2000.

[MHH02] T. L. Marzetta, B. Hassibi, and B. M. Hochwald. Structured Unitary Space–Time Autocoding Constellations. *IEEE Trans. Inform. Theory*, 48(4):942–950, April 2002.

[Mil74] K. Miller. *Complex Stochastic Processes*. Addison–Wesley, New York, 1974.

[Min11] H. Minkowski. *Gesammelte Abhandlungen*, volume 1. Teubner, Leipzig, 1911. (in German).

[ML89] M. L. Moher and J. H. Lodge. TCMP—A Modulation and Coding Strategy for Rician Fading Channels. *IEEE J. Select. Areas in Commun.*, 7(9):1347–1355, December 1989.

[MMB94] D. Makrakis, P. T. Mathiopoulos, and D. P. Bouras. Optimal Decoding of Coded PSK and QAM Signals in Correlated Fast Fading Channels and AWGN: A Combined Envelope, Multiple Differential and Coherent Detection Approach. *IEEE Trans. Commun.*, 42(1):63–74, January 1994.

[MTL05] Q. Ma, C. Tepedelenioğlu, and Z. Liu. Differential Space–Time–Frequency Coded OFDM With Maximum Multipath Diversity. *IEEE Transactions on Wireless Communications*, 4(5):2232–2243, September 2005.

[PH05a] P. Pun and P. Ho. Fano Space–Time Multiple Symbol Differential Detectors. In *Proc. of IEEE Global Telecommunications Conference (Globecom)*, pages 3169–3174, Saint Louis, November 2005.

[PH05b] P. Pun and P. Ho. The Performance of Fano–Multiple Symbol Differential Detection. In *Proc. of IEEE International Conference on Communications (ICC)*, pages 2516–2521, Seoul, May 2005.

[PK88] B. Picinbono and J.-M. Kerilis. Some Properties of Prediction and Interpolation Errors. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(4):525–531, April 1988.

[PL05] V. Pauli and L. Lampe. Multiple–Symbol Differential Sphere Decoding for Unitary Space–Time Modulation. In *Proc. of IEEE Global Telecommunications Conference (Globecom)*, pages 1630–1635, Saint Louis, November 2005.

[PL06] V. Pauli and L. Lampe. On the Complexity of Sphere Decoding for MSDD. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 932–936, Seattle, July 2006.

[PL07a] V. Pauli and L. Lampe. On the Complexity of Sphere Decoding for Differential Detection. *IEEE Trans. Inform. Theory*, 53(4):1595–1603, April 2007.

[PL07b] V. Pauli and L. Lampe. Tree–Search Multiple–Symbol Differential Decoding for Unitary Space–Time Modulation. *Acc. for publ.: IEEE Trans. Commun., preprint available from http://www.ece.ubc.ca/∼lampe/publicat.html*, accepted January 2007.

[PLH07a] V. Pauli, L. Lampe, and J. Huber. Differential Space–Frequency Modulation and 2D–Detection for MIMO–OFDM. In *Acc. for publ.: International Conference on Communications (ICC)*, Glasgow, accepted January 2007.

[PLH07b] V. Pauli, L. Lampe, and J. Huber. Differential Space–Frequency Modulation and Fast 2D–Multiple–Symbol Differential Detection for MIMO–OFDM. *Acc. for publ.: IEEE Transactions on Vehicular Technology, preprint available from http://www.ece.ubc.ca/∼lampe/publicat.html*, accepted March 2007.

[PLS06] V. Pauli, L. Lampe, and R. Schober. "Turbo DPSK" Using Soft Multiple–Symbol Differential Sphere Decoding. *IEEE Trans. Inform. Theory*, 52(4):1385–1398, April 2006.

[PLSF06] V. Pauli, L. Lampe, R. Schober, and K. Fukuda. Multiple–Symbol Differential Detection based on Combinatorial Geometry. *In Revision: IEEE Trans. Commun., preprint available from http://www.ece.ubc.ca/∼lampe/publicat.html*, submitted September 2006.

[PLSF07] V. Pauli, L. Lampe, R. Schober, and K. Fukuda. Multiple–Symbol Differential Detection based on Combinatorial Geometry. In *Acc. for publ.: International Conference on Communications (ICC)*, Glasgow, accepted January 2007.

[Poh81] M. Pohst. On the Computation of Lattice Vectors of Minimal Length, Successive Minima and Reduced Bases with Applications. *ACM, Special Interest Group on Symbolic and Algebraic Manipulation (SIGSAM) Bulletin*, 15(1):37–44, February 1981.

[Pro00] J. G. Proakis. *Digital Communications*. McGraw–Hill, New York, fourth edition, 2000.

[PS03] C. B. Peel and A. L. Swindlehurst. Effective SNR for Space–Time Modulation over a Time–Varying Rician Channel. *IEEE Trans. Commun.*, 52(1):17–23, January 2003.

[PSL06] V. Pauli, R. Schober, and L. Lampe. A Unified Performance Analysis Framework for Differential Detection in MIMO Rayleigh Fading Channels. *Submitted to: IEEE Trans. Commun.*, submitted December 2006.

[PTV02] W. H. Press, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, United Kingdom, 2002.

[Rap96a] D. Raphaeli. Noncoherent Coded Modulation. *IEEE Trans. Commun.*, 44:172–183, February 1996.

[Rap96b] T. S. Rappaport. *Wireless Communications*. Prentice–Hall, New Jersey, 1996.

[RS95] M. Russell and G. L. S. Stüber. Interchannel Interference Analysis of OFDM in a Mobile Environment. In *Proc. of IEEE Vehicular Technology Conference (VTC)*, pages 820–824, Chicago, IL, USA, July 1995.

[SA98] M. K. Simon and M.-S. Alouini. A Unified Approach to the Probability of Error for Noncoherent and Differentially Coherent Modulations Over Generalized Fading Channels. *IEEE Trans. Commun.*, 46(12):1625–1638, December 1998.

[SA01] M. K. Simon and M.-S. Alouini. Multiple Symbol Differential Detection with Diversity Reception. *IEEE Trans. Commun.*, 49(8):1312–1319, August 2001.

[SA05] M. K. Simon and M. S. Alouini. *Digital Communication over Fading Channels*. John Wiley & Sons, New York, 2nd edition, 2005.

[SBM+04] G. L. Stüber, J. R. Barry, S. W. McLaughlin, Y. Li, M. A. Ingram, and T. G. Pratt. Broadband MIMO–OFDM Wireless Communications. *Proceedings of the IEEE*, 92(2):271–294, February 2004.

[SBS66] M. Schwartz, W. Bennett, and S. Stein. *Communication Systems and Techniques*. McGraw–Hill, New York, 1966.

[SE94] C. P. Schnorr and M. Euchner. Lattice Basis Reduction: Improved Practical Algorithms and Solving Subset Sum Problems. *Mathematical Programming*, 66:181–191, 1994.

[SFG02] S. Siwamogsatham, M. P. Fitz, and J. H. Grimm. A New View of Performance Analysis of Transmit Diversity Schemes in Correlated Rayleigh Fading. *IEEE Trans. Inform. Theory*, 48(4):950–956, April 2002.

[SFGK00] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn. Fading Correlation and Its Effects on the Capacity of Multielement Antenna Systems. *IEEE Trans. Commun.*, 48(3):502–513, March 2000.

[SG00] R. Schober and W. H. Gerstacker. Decision–Feedback Differential Detection Based on Linear Prediction for MDPSK Signals Transmitted over Ricean Fading Channels. *IEEE J. Select. Areas in Commun.*, 18:391–402, March 2000.

[SGH99] R. Schober, W. H. Gerstacker, and J. B. Huber. Decision–Feedback Differential Detection of MDPSK for Flat Rayleigh Fading Channels. *IEEE Trans. Commun.*, 47(7):1025–1035, July 1999.

[SGH01] R. Schober, W. H. Gerstacker, and J. B. Huber. Decision–Feedback Differential Detection Based on Linear Prediction for 16DAPSK Signals Transmitted over Flat Ricean Fading Channels. *IEEE Trans. Commun.*, 49(8):1339–1342, August 2001.

[SHHS01] A. Shokrollahi, B. Hassibi, B. M. Hochwald, and W. Sweldens. Representation Theory for High–Rate Multiple–Antenna Code Design. *IEEE Trans. Inform. Theory*, 47(6):2335–2367, September 2001.

[SHL94] M. K. Simon, S. M. Hinedi, and W. C. Lindsey. *Digital Communication Techniques: Signal Design and Detection*. Prentice–Hall, Englewood Cliffs, NJ, USA, 1994.

[Sho01a] A. Shokrollahi. Design of Unitary Space–Time Codes from Representation of $SU(2)$. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, page 241, Washington, D.C., June 2001.

[Sho01b] A. Shokrollahi. Group Characters and Unitary Space–Time Codes. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, page 107, Washington, D.C., June 2001.

[SL02] R. Schober and L. Lampe. Noncoherent Receivers for Differential Space–Time Modulation. *IEEE Trans. Commun.*, 50(5):768–777, May 2002.

[SL05] W. Su and K. J. R. Liu. Differential Space–Frequency Modulation via Smooth Logical Channel for Broadband Wireless Communications. *IEEE Trans. Commun.*, 53(12):2024–2028, December 2005.

[SSL05] W. Su, Z. Safar, and K. J. R. Liu. Full–Rate Full–Diversity Space–Frequency Codes with Optimum Coding Advantage. *IEEE Trans. Inform. Theory*, 51:229–249, January 2005.

[Ste87] S. Stein. Fading Channel Issues in System Engineering. *IEEE J. Select. Areas in Commun.*, 5(2):68–89, February 1987.

[Ste99] R. Steele. *Mobile Radio Communications*. John Wiley & Sons, New York, 2nd edition, 1999.

[Sve94] A. Svensson. Coherent Detection Based on Linear Prediction and Decision Feedback for QDPSK. *IEE Electronics Letters*, 30(20):1642–1643, 1994.

[SVH06] M. Stojnic, H. Vikalo, and B. Hassibi. Further Results on Speeding up the Sphere Decoder. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 4, pages 549–552, Toulouse, May 2006.

[Swe01] W. Sweldens. Fast Block Noncoherent Decoding. *IEEE Communications Letters*, 5(4):132–134, April 2001.

[SY03] X. Shao and J. Yuan. A New Differential Space–Time Block Coding Scheme. *IEEE Communications Letters*, 7(9):437–439, sep 2003.

[TC01] M. Tao and R. S. Cheng. Differential Space–Time Block Codes. In *Proc. of IEEE Global Telecommunications Conference (Globecom)*, pages 1098–1102, San Antonio, TX, November 2001.

[TJ00] V. Tarokh and H. Jafarkhani. A Differential Detection Scheme for Transmit Diversity. *IEEE J. Select. Areas in Commun.*, 18(7):1168–1174, July 2000.

[TJC99] V. Tarokh, H. Jafarkhani, and A. R. Calderbank. Space–Time Block Codes from Orthogonal Designs. *IEEE Trans. Inform. Theory*, 45(5):1456–1467, July 1999.

[Tre71] H. L. Van Trees. *Detection, Estimation, and Modulation Theory*, volume III. John Wiley & Sons, New York, NY, USA, 1971.

[VB99] E. Viterbo and J. Boutros. A Universal Lattice Code Decoder for Fading Channels. *IEEE Trans. Inform. Theory*, 45:1639–1643, July 1999.

[VCBT97] J. Ventura–Traveset, G. Caire, E. Biglieri, and G. Taricco. Impact of Diversity Reception on Fading Channels with Coded Modulation—Part II: Differential Block Detection. *IEEE Trans. Commun.*, 45(6):676–686, June 1997.

[Vin84] A. J. H. Vinck. A Low Complexity Stack Decoder for a Class of Binary Rate $(n-1)/n$ Convolutional Codes. *IEEE Trans. Commun.*, 32(4):476–479, April 1984.

[Vit67] A. J. Viterbi. Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. *IEEE Trans. Commun.*, 13(2):260–269, April 1967.

[Vor09] G. Voronoï. Nouvelles Applications des Paramètres Continus à la Théorie des Formes Quadratiques. *J. Reine und Angewandte Mathematik*, 133:97–178, 1909. Also, 134:198–287, 1908; and 136:67–181, 1909 (in French).

[VT95a] G. M. Vitetta and D. P. Taylor. Maximum Likelihood Decoding of Uncoded and Coded PSK Signal Sequences Transmitted over Rayleigh Flat–Fading Channels. *IEEE Trans. Commun.*, 43(11):2750–2758, November 1995.

[VT95b] G. M. Vitetta and D. P. Taylor. Viterbi Decoding of Differentially Encoded PSK Signals Transmitted over Rayleigh Frequency–Flat Fading Channels. *IEEE Trans. Commun.*, 43(2/3/4):1256–1259, February/March/April 1995.

[vzGG99] J. von zur Gathen and J. Gerhard. *Modern Computer Algebra*. Cambridge University Press, Cambridge, UK, 1999.

[WE71] S. B. Weinstein and P. M. Ebert. Data Transmission by Frequency–Division Multiplexing Using the Discrete Fourier Transform. *IEEE Communications Magazine*, 19(5):628–634, October 1971.

[WFM89] S. G. Wilson, J. Freebersyser, and C. Marshall. Multi–Symbol Detection of M–DPSK. In *Proc. of IEEE Global Telecommunications Conference (Globecom)*, pages 47.3.1–47.3.6, Dallas, November 1989.

[Win04] C. Windpassinger. *Detection and Precoding for Multiple Input Multiple Output Channels*. PhD thesis, Universität Erlangen–Nürnberg, Erlangen, Germany, January 2004.

[Woz57] J. M. Wozencraft. Sequential Decoding for Reliable Communication. In *IRE Convention Record*, pages 11–25, 1957.

[WWM05] J. Wang, X. Wang, and M. Madihian. Design of Minimum Error–Rate Cayley Differential Unitary Space–Time Codes. *IEEE J. Select. Areas in Commun.*, 23(9):1779–1787, September 2005.

[WY02] J. Wang and K. Yao. Differential Unitary Space–Time–Frequency Coding for MIMO OFDM Systems. In *Proc. of 36th Asilomar Conf. on Signals, Systems and Computers*, volume 2, pages 1867–1871, Pacific Grove, November 2002.

[XWZW04] W. Xu, Y. Wang, Z. Zhou, and J. Wang. A Computationally Efficient Exact ML Sphere Decoder. In *Proc. of IEEE Global Telecommunications Conference (Globecom)*, pages 2594–2598, Dallas, TX, November 2004.

[YL95] R. J. Young and J. H. Lodge. Detection of CPM Signals in Fast Rayleigh Flat–Fading Using Adaptive Channel Estimation. *IEEE Transactions on Vehicular Technology*, pages 338–347, May 1995.

[YP95] X. Yu and S. Pasupathy. Innovations–Based MLSE for Rayleigh Fading Channels. *IEEE Trans. Commun.*, 43(2/3/4):1534–1544, February/March/April 1995.

[Zas75] T. Zaslavsky. *Facing up to Arrangements: Face–Count Formulas for Partitions of Space by Hyperplanes*. American Mathematical Society, Providence, 1975.

[ZF06] E. Zimmermann and G. Fettweis. Improved Length Term Calculation and MMSE Extension for LISS MIMO Detection. In *Proc. of IEEE Information Theory Workshop (ITW)*, Chengdu, China, October 2006.

[Zie94] G. M. Ziegler. *Lectures on Polytopes*. Graduate Texts in Mathematics. Springer–Verlag, 1994.

[Zig66] K. Zigangirov. Some Sequential Decoding Procedures. *Probl. Peredachi Inf.*, 2:13–25, 1966.