Detection of Differential Unitary Space-Time Modulation in Fast Rayleigh-Fading Channels

A Thesis

Presented in Partial Fulfillment of the Requirements for

the Degree Master of Science in the

Graduate School of The Ohio State University

By

Bijoy Bhukania, B.Tech (Hons.)

* * * * *

The Ohio State University

2002

Master's Examination Committee:

Dr. Philip Schniter, Adviser

Dr. Hesham El-Gamal

Approved by

Adviser

Department of Electrical Engineering

ABSTRACT

This thesis deals with the detection of differential unitary-space time modulation (DUST) in fast-fading channels. In a fading environment, the standard single-symbol maximum likelihood (ML) detection of DUST yields an error floor in the BER-versus-SNR plot. This is due to the incorrect assumption of a constant channel over successive matrix-symbol transmissions. Assuming we know the correlation of channel fades, we design multiple-symbol differential detectors (MSD) and decision-feedback differential detectors (DFDD) which reduce this error floor inherent to the standard detector. The suggested detectors are derived under the simplifying assumption of a single channel change per matrix-symbol (i.e., a block-fading environment as opposed to continuous-fading). For the special case of diagonal space-time constellations, this block-fading assumption is no longer required.

To combat the prohibitive exponential complexity of the multiple-symbol ML detector (MSMD) we propose a linear complexity suboptimal MSD that is implemented using the Viterbi algorithm. For a further reduction of complexity, we also propose the DFDD, which is derived from the MSMD by replacing all but the latest hypothesized symbol with past decisions. The analysis of the DFDD shows the presence of a Wiener channel predictor followed by a quasi-coherent ML detector. The performance and robustness of the detectors under non-ideal conditions are analyzed via

simulations and theoretical analysis. We find that the DFDD offers a better trade-off between complexity of detection and BER performance than the MSD.

Detection of Differential Unitary Space-Time Modulation in Fast Rayleigh-Fading Channels

By

Bijoy Bhukania, M.S.

The Ohio State University, 2002

Dr. Philip Schniter, Adviser

This thesis deals with the detection of differential unitary-space time modulation (DUST) in fast-fading channels. In a fading environment, the standard single-symbol maximum likelihood (ML) detection of DUST yields an error floor in the BER-versus-SNR plot. This is due to the incorrect assumption of a constant channel over successive matrix-symbol transmissions. Assuming we know the correlation of channel fades, we design multiple-symbol differential detectors (MSD) and decision-feedback differential detectors (DFDD) which reduce this error floor inherent to the standard detector. The suggested detectors are derived under the simplifying assumption of a single channel change per matrix-symbol (i.e., a block-fading environment as opposed to continuous-fading). For the special case of diagonal space-time constellations, this block-fading assumption is no longer required.

To combat the prohibitive exponential complexity of the multiple-symbol ML detector (MSMD) we propose a linear complexity suboptimal MSD that is implemented

using the Viterbi algorithm. For a further reduction of complexity, we also propose the DFDD, which is derived from the MSMD by replacing all but the latest hypothesized symbol with past decisions. The analysis of the DFDD shows the presence of a Wiener channel predictor followed by a quasi-coherent ML detector. The performance and robustness of the detectors under non-ideal conditions are analyzed via simulations and theoretical analysis. We find that the DFDD offers a better trade-off between complexity of detection and BER performance than the MSD.

Dedicated to my family...

ACKNOWLEDGMENTS

First and foremost i would like to thank my adviser, Prof. Philip Schniter for his guidance and invaluable advice imparted throughout my stay at OSU. I also want to thank Prof. Hesham El-Gamal for being part of my Master's examinations committee. My education at OSU would have been far from complete without his insightful comments and teaching.

It has been a great experience working in the IPS laboratory at OSU. Credit for this goes to friendship and support of its affable people. I want to thank Ravi "kirin" Rajagopal for his technical support. I want to thank Defne and Emre for help and advice on research. I want to thank Jing, the computer/LaTeX wizard for his help. Saving the best for last I want to thank Siddharth, Adam, Karthik and Mike for all the fun times we had working on common course projects and for all the discussions (technical and ...) that made this MS program an unforgettably fascinating experience.

VITA

January 18, 1978	Born - Raniganj, India
2000	B.Tech (Hons.) Electronics & Electrical Communications Engineering, Indian Institute of Technology, Kharagpur
2000-2001	University Fellow, The Ohio State University.
2001-2002	Graduate Research Associate, The Ohio State University

PUBLICATIONS

Journal Publications

B. Bhukania and P. Schniter, "Multiple-symbol Detection of Differential Unitary Space-time Modulation in Fast Rayleigh Fading Channels". *Submitted to IEEE Trans. Communications*, Apr. 2002.

B. Bhukania and P. Schniter, "Decision-Feedback Detection of Differential Unitary Space-time Modulation in Fast Rayleigh Fading Channels". *Submitted to IEEE Trans. Wireless Communications*, Aug. 2002.

Conference Publications

B. Bhukania and P. Schniter, "Multiple-symbol Detection of Differential Unitary Space-time Modulation in Fast-Fading Channels with Known Correlation". *Proc. Conference on Information Sciences and Systems*, (Princeton, NJ), Mar. 2002.

B. Bhukania and P. Schniter, "Decision-Feedback Detection of Differential Unitary Space-time Modulation in Fast Rayleigh-Fading Channels". *Proc.* 40^{th} *Annual*

Allerton Conference on Communications, Control, and Computing, (Monticello, IL), Oct 2002

B. Bhukania and P. Schniter, "On the Robustness of Decision-Feedback Detection of DPSK and Differential Unitary Space-Time Modulation in Rayleigh Fading Channels". Submitted to *IEEE Wireless Communications and Networking Conference* (WCNC), 2003

FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in:

Detection & Estimation Theory, Signal Processing	Prof. Philip Schniter
Communication Theory	Prof. Hesham El-Gamal

TABLE OF CONTENTS

Appendices:

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

NOTATIONS

Notation	Definition
$\mathbf{I}_P$	Identity matrix of size $P \times P$
j	$\sqrt{-1}$ if not identified as an index
$E\{\cdot\}$	Expectation
$\mathrm{vec}(\cdot)$	Stacking of columns of matrices (e.g., $\mathbf{x}_n = \mathrm{vec}(X_n)$)
$(\cdot)^*$	Hermitian transpose (i.e., conjugate transpose)
$\otimes$	Kronecker product
$\mathrm{tr}(\cdot)$	Trace operator for square matrices
$\det(\cdot)$	Determinant
$\mathrm{Res}[\cdot]$	Residue
$\Pr(\cdot)$	Probability
$\Re(\cdot)$	Extraction of real valued component

$$\prod_{j=k_l}^{k_u} A_j = \begin{cases} A_{k_l} A_{k_l+1} \cdots A_{k_u} & k_u \geq k_l \\ \text{Identity matrix of appropriate size} & k_u < k_l \end{cases}$$

CHAPTER 1

INTRODUCTION

1.1 Motivation

The rapidly evolving global information infrastructure includes wireless communication as a key component. The major challenge posed by modern wireless channel stems from potentially rapid signal-strength attenuation and significantly random phase distortion of the signal, commonly referred to as "fading". A typical narrowband flat-fading channel can be described in the discrete baseband (i.e., match filtered and sampled) domain as

$$x_n = h_n s_n + w_n \tag{1.1}$$

where s_n is the transmitted signal taken from a finite alphabet, e.g. a QAM constellation, x_n is the received signal, h_n is a complex multiplicative fading coefficient and w_n is the additive noise at the receive antenna. In practice, h_n and s_n are both unknown at the receiver, and the receiver needs to find s_n from x_n.

In additive white Gaussian noise (AWGN) channels, where fading is absent, phase shift keying (PSK) is traditionally considered to be a power efficient modulation technique [1]. In PSK, the information is transmitted in the phase of the signal. In

the fading scenario, the phase uncertainty induced by the channel makes the detection of the PSK signal impossible unless these unknown phase distortions are corrected at the receiver*. In communication systems where the fading coefficients change slowly with time, known training symbols (also known as pilots) are multiplexed with the unknown symbols at regular intervals. These help to estimate the fading coefficients for coherent demodulation of the unknown symbols. When the fading coefficients change rapidly, accurate channel estimation entails the need for many pilots which is impractical. In these scenarios differential phase shift keying (DPSK) [1, 2] can be effectively used as it outsources the need for channel estimation.

Another form of distortion that impedes successful communication is "amplitude distortion". In an AWGN channel the bit error rate (BER) of a system employing the PSK modulation scheme decreases exponentially with the signal-to-noise ratio (SNR), while in a Rayleigh-fading channel (i.e., the fading coefficients are complex Gaussian random variables) with coherent signal detection, the BER decreases linearly with SNR. Though fading may seem as a hindrance, researchers have wisely come up with means of improving system performance by exploiting the "diversity" that fading offers. On fading channels, the trade-off between spectral efficiency and power consumption can be significantly improved by deploying multiple antennas at the transmitter and/or receiver [3–6]. In fact, in a Rayleigh flat-fading environment, a transmitter-receiver link has a theoretical capacity that increases linearly with the smaller of the number of transmit and receive antennas, provided that the complex-valued fading coefficients between all pairs of antennas are statistically independent and known at the receiver [3, 5].

*In fact, the phase of the received and transmitted signal can be shown to be independent if h_n and w_n are unknown complex Gaussian variables.

Space-time coding is a bandwidth and power efficient method of communicating over fading channels that exploits spatial diversity afforded by the use of multiple transmit antennas [7, 8]. Various space-time coding and modulation schemes have been proposed in literature for use in systems employing multiple transmit antennas. These schemes attempt to achieve diversity gains and/or increase throughput [7, 8]. While a majority of these techniques have been studied under the assumption of perfect channel state information (CSI) at the receiver, the class of unitary space-time modulation [9, 10] and differential space-time modulation [11–17] have been designed for Rayleigh flat-fading channels with no CSI. In this thesis, we focus on differential unitary space-time modulation (DUST) [11, 12, 17], which is an important subset of general differential space-time modulation schemes.

DUST can be viewed as a multiple-antenna extension of DPSK, commonly used in single antenna systems with no CSI. In DPSK, the information is encoded in the phase difference between two consecutively transmitted symbols, enabling the recovery of information using the phase difference between the corresponding two consecutive observations at the receiver. In DUST, the transmitted symbols are $M \times M$ unitary matrices, where M is the number of transmit antennas[†]. The previously transmitted matrix-symbol is pre-multiplied by the current information matrix-symbol to form the current transmitted matrix-symbol. Each received matrix-symbol provides a channel estimate (or phase reference) to the next received symbol, and thus the standard single symbol maximum-likelihood (ML) detector [11, 12], henceforth termed as the "standard detector", attempts to decode the information symbol from a pair of consecutive received matrices. There are two major drawbacks to this technique. Due to

[†]In DPSK, the transmitted symbols are unit modulus, and therefore can be considered to be 1×1 unitary matrices.

the presence of additive noise at the receive antennas, each received symbol provides a *noisy* channel estimate for the next symbol thereby increasing the noise power. This results in the well known 3dB SNR loss as opposed to a coherent detection. Also, standard detection assumes that the channel experienced by every pair of consecutively received matrices is the same. This would require a channel that is forever constant—a requirement which is never met in practice. Therefore with standard detection in continuously-fading channels, both DUST and single antenna DPSK modulation succumb to an irreducible error floor caused by failure to account for the channel variations. [18–22].

For DPSK, multiple-symbol ML detection (MSMD) [23] and prediction-based decision-feedback differential detection (DFDD) [24–26] have been proposed to improve the performance in an AWGN channel. MSMD and DFDD have been extended to the Rayleigh fading channel [21, 22, 27], where they reduce and asymptotically eliminate the error floor encountered by the standard detector.

1.2 Organization & Contributions

In this thesis, we present MSMD and DFDD for DUST in Rayleigh fast-fading channels—channels for which standard detection is unsuccessful. In deriving the new detectors, we make the simplifying assumption that the channel changes once per matrix-symbol (i.e., M channel uses), and that the receiver knows the fading correlation. However, we observe that the block fading assumption is not essential when diagonal space-time constellations [11] are used. Under these assumptions, we derive the exponential complexity optimal and the linear complexity suboptimal multiple-symbol detectors. We also derive the decision-feedback differential detector

for DUST. Performance of the detectors are analyzed via simulation and theoretical analysis. Although the detectors are derived assuming a fast block-fading channel, simulation results show that the new detectors exhibit improved performance over the standard detector in fast continuously-fading channels.

1.2.1 Multiple-Symbol Detection

In Chapter 3, single and multiple-symbol ML detectors for DUST are derived. Complexity of these multiple-symbol ML detectors is exponential in the sequence length as there does not appear to exist any Viterbi-like low-complexity implementation of the detectors. Therefore, for large sequences we propose a suboptimal detector which is implemented using the Viterbi algorithm and yields complexity linear in sequence length.

The multiple-symbol ML as well as suboptimal detection metrics derived in this thesis can be expressed as a Hermitian quadratic form of Gaussian vectors. Exact evaluation of the pair-wise word error probability (PWEP) using the characteristic function and residue theorem approach [21] is difficult as it involves computing residues of a function at repeated poles. We derive upper and lower bounds on the PWEP by perturbing the pole locations [28], and find through numerical evaluation that these bounds are very close to each other. The PWEP is then used to find approximate bit error rate (BER).

1.2.2 Decision-feedback Differential Detection

In Chapter 4, a decision-feedback detector is derived from the multiple-symbol ML detector and shown to be equivalent to a Wiener-filter based channel predictor followed by coherent ML detector. For performance and robustness analysis, exact

and Chernoff bound expressions for pairwise word error probability (PWEP) are derived for the cases of perfect and imperfect knowledge of fading correlation and SNR. An approximate expression for BER is derived from the PWEP which is verified by the simulation results.

It may be worthwhile to mention a related work [29,30], wherein DFDD is proposed for the specific case of diagonal constellations, as opposed to the general class of unitary constellations [12, 13, 17] which is our focus. Since diagonal DUST with M transmit antennas is equivalent to M decoupled single antenna DPSK systems, results for single antenna DPSK are trivially extended to diagonal DUST in [29, 30]. Unlike [29, 30], we also derive the exact PWEP and study the robustness of DFDD to imperfect parameter knowledge.

CHAPTER 2

BACKGROUND

2.1 Rayleigh Flat-Fading Channel Model

We consider a system with M transmit antennas and P receive antennas that operates in a Rayleigh flat-fading environment. Each receive antenna is coupled to each transmit antenna via a statistically independent fading coefficient. The received signal at each antenna is corrupted by an additive noise that is independent across antennas as well as the symbol periods. Consider the transmission of $M \times M$ matrices $S_n = [\mathbf{s}_{1,n}\ \mathbf{s}_{2,n}\ \ldots\ \mathbf{s}_{M,n}]^*$ in the n^{th} *matrix-symbol interval*, such that each matrix-symbol interval comprises M signaling intervals, and the elements of the k^{th} row of S_n, $\mathbf{s}^*_{k,n}$ are transmitted from the M transmit antennas in the k^{th} signaling interval within the n^{th} matrix-symbol interval. Then, collecting the corresponding received samples at the P receive antennas in the row-vector $\mathbf{x}^*_{k,n}$, we can describe the system as

$$\mathbf{x}^*_{k,n} = \sqrt{\frac{\rho}{M}}\mathbf{s}^*_{k,n} \underbrace{\begin{bmatrix} h_{k,n,1,1} & h_{k,n,1,2} & \ldots & h_{k,n,1,P} \\ h_{k,n,2,1} & h_{k,n,2,2} & \ldots & h_{k,n,2,P} \\ \vdots & \vdots & \vdots & \vdots \\ h_{k,n,M,1} & h_{k,n,M,2} & \ldots & h_{k,n,M,P} \end{bmatrix}}_{H_{k,n}} + \mathbf{w}^*_{k,n} \tag{2.1}$$

where $h_{k,n,i,j}$ is the unit-variance complex Gaussian fading coefficient between i^{th} transmit and j^{th} receive antenna during the k^{th} signaling interval within the n^{th} matrix-symbol interval, i.e., $H_{k,n}$ is the MIMO channel response matrix at the $(nM + k)^{th}$ channel use. $\mathbf{w}_{k,n}$ is a column vector containing i.i.d unit-variance proper complex Gaussian entries. The fading coefficients $h_{k,n,i,j}$ are independent with respect to i and j, but correlated over time. The transmitted signal is normalized such that ρ is the average SNR at each receive antenna.

Denoting the $M \times P$ received matrix during the n^{th} matrix-symbol interval as $X_n = [\mathbf{x}_{1,n}\ \mathbf{x}_{2,n}\ \ldots\ \mathbf{x}_{M,n}]^*$, and $W_n = [\mathbf{w}_{1,n}\ \mathbf{w}_{2,n}\ \ldots\ \mathbf{w}_{M,n}]^*$, (2.1) can be written compactly as

$$X_n = \sqrt{\frac{\rho}{M}} \begin{bmatrix} \mathbf{s}_{1,n}^* & \mathbf{0}^* & \ldots & \mathbf{0}^* \\ \mathbf{0}^* & \mathbf{s}_{2,n}^* & \ldots & \mathbf{0}^* \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}^* & \mathbf{0}^* & \ldots & \mathbf{s}_{M,n}^* \end{bmatrix} \begin{bmatrix} H_{1,n} \\ H_{2,n} \\ \vdots \\ H_{M,n} \end{bmatrix} + W_n \tag{2.2}$$

Under the assumption that the channel remains constant for one matrix-symbol interval, i.e., the channel is block-fading, we have $H_n = H_{1,n} = \cdots = H_{M,n}$, which changes (2.2) to

$$\boxed{X_n = \sqrt{\frac{\rho}{M}} S_n H_n + W_n} \tag{2.3}$$

Recall that $S_n = [\mathbf{s}_{1,n}\ \mathbf{s}_{2,n}\ \ldots\ \mathbf{s}_{M,n}]^*$. Even without the block-fading assumption, the use of *diagonal* matrices S_n implies that the continuous fading model (2.2) simplifies to

$$X_n = \sqrt{\frac{\rho}{M}} S_n H_n^{(c)} + W_n \tag{2.4}$$

where the k^{th} row of the "equivalent continuous fading channel matrix" $H_n^{(c)}$ is the k^{th} row of $H_{k,n}$, for $k = 1, \ldots, M$. If the MIMO fading process $H_{k,n}$ is independent

between antennas, then the equivalent fading process $H_n^{(c)}$ is also independent between antennas; the process $H_n^{(c)}$, however, is M-fold "faster" than $H_{k,n}$. Keeping this in mind, $H_n^{(c)}$ shall be denoted as H_n from here onwards, and (2.3) will be used as the system model. Recall that (2.3) is an approximate model when non-diagonal S_n are used in the continuous fading channel.

2.2 Single Antenna Differential Phase-Shift Keying

In this section we present a brief review of single antenna differential phase shift keying (DPSK) [1, 2]. We consider a system with 1 transmit and 1 receive antenna for our discussion in this section. Note that (2.3) is a valid system model when DPSK is used.

2.2.1 Differential Encoding

DPSK is traditionally used in systems where the signal experiences a random, unknown and slowly varying phase distortion. The information is encoded in the phase difference of two consecutive symbols. The widely used symbol set (constellation) for DPSK is

$$v_l = e^{j2\pi l/L} \quad l \in \mathcal{L} = \{0, 1, \ldots, L-1\} \tag{2.5}$$

where $L = 2^R$ and R is the data rate in terms of bits per channel use. For transmission of the information sequence $z_1, z_2, \ldots$, we select the corresponding *information symbol* sequence $v_{z_1}, v_{z_2}, \ldots$, and transmit the symbols $s_0, s_1, s_2, \ldots$ such that $s_0 = 1$ is known and does not carry any information, and $s_n = v_{z_n} s_{n-1}$.

2.2.2 Standard Single Symbol Detection of DPSK

The transmitted and received signals are related by the single antenna version of (2.3)

$$x_n = \sqrt{\rho} s_n + w_n \tag{2.6}$$

Using (2.6), $x_{n-1} = \sqrt{\rho} s_{n-1} + w_{n-1}$, $s_n = v_{z_n} s_{n-1}$, and assuming that the channel remains constant for two consecutive symbol intervals, i.e., $h_n = h_{n-1}$, we can write

$$x_n = x_{n-1} v_{z_n} + \underbrace{w_n - w_{n-1} v_{z_n}}_{\hat{w}_n} \tag{2.7}$$

where $\hat{w}_n$ is AWGN with doubled variance since $|v_{z_n}| = 1$. Observe that the problem of detection of z_n is turned into a known channel problem, and it is straightforward to show that the ML detector is

$$\hat{z}_n = \arg \max_{z_n \in \mathcal{L}} \Re[x_n^* v_{z_n} x_{n-1}] \tag{2.8}$$

which is the "standard detector" for DPSK.

2.2.3 Multiple-symbol ML Detection of DPSK

Multiple-symbol ML detection of DPSK in AWGN channel was introduced by Divsalar & Simon [23], and later extended to Rayleigh fading channel by the same authors [27], Ho [21] and Schober *et al.* [22]. Collecting $x_n, x_{n-1}, \ldots, x_{n-m}$ into a vector, from (2.6) we can write

$$\underbrace{\begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_{n-m} \end{bmatrix}}_{\mathbf{x}_n} = \sqrt{\rho} \underbrace{\begin{bmatrix} s_n & 0 & \ldots & 0 \\ 0 & s_{n-1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & s_{n-m} \end{bmatrix}}_{\mathbf{S}_n} \underbrace{\begin{bmatrix} h_n \\ h_{n-1} \\ \vdots \\ h_{n-m} \end{bmatrix}}_{\mathbf{h}_n} + \underbrace{\begin{bmatrix} w_n \\ w_{n-1} \\ \vdots \\ w_{n-m} \end{bmatrix}}_{\mathbf{w}_n} \tag{2.9}$$

The multiple-symbol ML detection, i.e., the optimal joint detection of $\{z_k\}_{k=n-m+1}^{n}$ can be accomplished by maximizing the density of $\mathbf{x}_n$ conditioned on $\{v_{z_k}\}_{k=n-m+1}^{n}$, given by [31, 32]

$$p(\mathbf{x}_n|z_n, z_{n-1}, \ldots, z_{n-m+1}) = \frac{1}{\pi^{m+1}\det(R)} e^{-\mathbf{x}_n^* R^{-1} \mathbf{x}_n} \tag{2.10}$$

where

$$R = E\{\mathbf{x}_n \mathbf{x}_n^* | z_n, z_{n-1}, \ldots, z_{n-m+1}\}$$

It can be shown that R depends on the information symbols $\{z_k\}_{k=n-m+1}^{n}$ rather than the absolute phases $\{s_k\}_{k=n-m}^{n}$. Moreover, $\det(R)$ can be shown to be independent of $\{z_k\}_{k=n-m+1}^{n}$. Thus, the ML detection rule becomes [21]

$$\{\hat{z}\}_{k=n-m+1}^{n} = \underset{z_n, \ldots, z_{n-m+1} \in \mathcal{L}}{\arg\min} \mathbf{x}_n^* R^{-1} \mathbf{x}_n \tag{2.11}$$

For performance analysis, an exact expression for the pairwise word error probability for the detector (2.11) can be derived using the residue theorem approach [21]. Our error performance analysis of MSMD of DUST in Chapter 3 is also applicable to DPSK, therefore we omit the discussion on the error performance of (2.11) here.

2.2.4 Decision-feedback Detection of DPSK

The biggest disadvantage of the MSMD is the computational complexity, which is exponential in sequence length m. Although fast algorithms for computation of the metric in (2.11) has been proposed [33], a simpler way to reduce the complexity without incurring significant performance loss is to replace all but one hypothesized symbols in (2.11) by past detected symbols, and then performing symbol by symbol detection. This scheme is called decision feedback differential detection (DFDD) [22].

Our derivation and performance analysis of DFDD in Chapter 4 is suitable for general class of differential unitary constellations, therefore can be applied to DPSK by setting number of transmit antennas $M = 1$.

2.3 Differential Unitary Space-Time Modulation

Section 2.2 shows that single antenna differential modulation uses *scalar* symbols for modulation. For a system with M transmit antennas, the constellation consists of $M \times M$ space-time matrix-symbols. The $M \times M$ matrix-symbol is transmitted from M transmit antennas in each *matrix-symbol interval*, which comprises M signaling intervals.

2.3.1 Differential Encoding

To transmit η bits per channel use in an M transmit antenna system, the required constellation size is $L = 2^{\eta M}$. Suppose we want to transmit the information symbols $z_1, z_2, \ldots \in \mathcal{L} = \{0, 1, \ldots, L-1\}$. We select the *information matrix-symbols* $V_{z_1}, V_{z_2}, \ldots, \in \mathcal{A}$, where $\mathcal{A}$ is the DUST constellation, and V_{z_n} is unitary. Then, we transmit the matrix-symbol S_n in the n^{th} matrix-symbol interval such that

$$S_n = V_{z_n} S_{n-1} \qquad n = 1, 2, \ldots \tag{2.12}$$

and $S_0 = \mathbf{I}_M$. Equation (2.12) is the *fundamental differential transmission equation* derived in [11]. Since the matrices V_{z_n} are unitary, the transmitted matrices S_n are also unitary.

2.3.2 Standard Single Symbol ML Detection of DUST

The transmitted matrix S_n and the corresponding received matrix X_n are related by (2.3). Using (2.3), $X_{n-1} = \sqrt{\frac{\rho}{M}} S_{n-1} H_{n-1} + W_{n-1}$, $S_n = V_{z_n} S_{n-1}$, and assuming that the channel remains constant for two consecutive matrix-symbol intervals (equivalently, $2M$ signaling intervals), i.e., $H_n = H_{n-1}$, we obtain

$$X_n = V_{z_n} X_{n-1} + \underbrace{W_n - V_{z_n} W_{n-1}}_{\hat{W}_n} \tag{2.13}$$

Since V_{z_n} is unitary and W_{n-1} contains i.i.d. AWGN entries that are statistically invariant to unitary transformation, $\hat{W}_n$ contains i.i.d Gaussian entries with variance twice of those in W_n. From (2.13), it is straightforward to show that the ML detector for z_n is [11]

$$\hat{z}_n = \arg \max_{z_n \in \mathcal{L}} \Re[\mathrm{tr}\{X_n^* V_{z_n} X_{n-1}\}] \tag{2.14}$$

Due to increased noise variance in $\hat{W}_n$, this detector performs at a 3dB SNR loss compared to coherent detection. Under the assumption of $H_n = H_{n-1}$, the Chernoff upper bound on P_e for the detector in (2.14) converges to zero as $\rho \to \infty$ [11]. However if the channel changes from one matrix-symbol interval to the next, i.e., if $H_n = H_{n-1} + \Delta H$, then (2.13) becomes

$$X_n = V_{z_n} X_{n-1} + \underbrace{\sqrt{\frac{\rho}{M}} S_n \Delta H}_{\tilde{X}_n} + W_n - V_{z_n} W_{n-1}$$

where the term $\tilde{X}_n$ creates additional "noise" that induces an error floor in BER curve. Note that the detector in (2.14) ignores channel variation and is therefore suboptimal in a fading environment. In Chapter 3 & 4 we derive detection rules that exploit knowledge of the autocorrelation of the time-varying channel coefficients.

2.3.3 DUST Constellations

The performance of DUST at high SNR with standard detection depends on the minimum diversity product between all possible pairs of matrices in $\mathcal{A}$ [11], defined as

$$D(\mathcal{A}) = \frac{1}{2} \min_{0 \leq j < k \leq L-1} |\det(V_j - V_k)|^{\frac{1}{M}} \tag{2.15}$$

Keeping this in mind, several constellations for DUST have been designed in [11, 12]. The special class of diagonal constellations has been designed in [11] to maximize $D(\mathcal{A})$ under the constraint that the constellation forms an Abelian group. The elements of the diagonal constellation for an M transmit antenna system are given by

$$V_k = \begin{bmatrix} e^{j\frac{2\pi}{L}u_1 k} & 0 & 0 \\ 0 & \ddots & \vdots \\ 0 & \dots & e^{j\frac{2\pi}{L}u_M k} \end{bmatrix} \quad k \in \mathcal{L} = \{0, 1, \dots, L-1\} \tag{2.16}$$

where $u_1, \dots, u_M \in \{0, 1, \dots, L-1\}$ are chosen to maximize $D(\mathcal{A})$.

For simulation and numerical examples in this thesis, we consider a system with two transmit antennas with $\eta = 1$, i.e., $L = 2^{\eta M} = 4$. The optimal diagonal constellations for this case is obtained by setting $u_1 = 1, u_2 = 1$ in (2.16) [11]. To evaluate detection performance with non-diagonal constellations, we generate non-diagonal constellation from the diagonal constellation by pre-multiplying each element with a non-diagonal unitary matrix. Because such an operation does not change $D(\mathcal{A})$, the comparison remains fair. Table 2.1 shows the constellations used in this thesis for simulation and numerical examples.

Diagonal Constellation	Non-diagonal Constellation
$V_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$V_0 = \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}$
$V_1 = \begin{bmatrix} j & 0 \\ 0 & j \end{bmatrix}$	$V_1 = \begin{bmatrix} 0.7071j & 0.7071j \\ 0.7071j & -0.7071j \end{bmatrix}$
$V_2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$	$V_2 = \begin{bmatrix} -0.7071 & -0.7071 \\ -0.7071 & 0.7071 \end{bmatrix}$
$V_3 = \begin{bmatrix} -j & 0 \\ 0 & -j \end{bmatrix}$	$V_3 = \begin{bmatrix} -0.7071j & -0.7071j \\ -0.7071j & 0.7071j \end{bmatrix}$

Table 2.1: Constellations used in this thesis for simulation and numerical examples

CHAPTER 3

MULTIPLE-SYMBOL DIFFERENTIAL DETECTION†

Multiple-symbol differential detection of DPSK has been proposed as an effective way to reduce the 3dB SNR loss incurred by one-symbol differential detection [23], as well as to enhance performance in correlated Rayleigh fading channels [21,22]. In this chapter we extend this idea to DUST and derive m-symbol ML as well as suboptimal detectors in temporally correlated fading.

3.1 m-Symbol ML Differential Detector

We collect the observation matrix-symbols $X_{n-m}, X_{n-m+1}, \ldots, X_n$ with the intention of jointly detecting $z_{n-m+1}, z_{n-m+2}, \ldots, z_n$. Denoting $\mathbf{h}_n = \mathrm{vec}(H_n)$, $\mathbf{x}_n = \mathrm{vec}(X_n)$, and $\mathbf{w}_n = \mathrm{vec}(W_n)$, we can write

$$\underbrace{\begin{bmatrix} \mathbf{x}_n \\ \vdots \\ \mathbf{x}_{n-m} \end{bmatrix}}_{\underline{\mathbf{x}}_n} = \sqrt{\frac{\rho}{M}} \underbrace{\begin{bmatrix} \mathbf{I}_P \otimes S_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_P \otimes S_{n-m} \end{bmatrix}}_{\mathcal{S}_n} \underbrace{\begin{bmatrix} \mathbf{h}_n \\ \vdots \\ \mathbf{h}_{n-m} \end{bmatrix}}_{\underline{\mathbf{h}}_n} + \begin{bmatrix} \mathbf{w}_n \\ \vdots \\ \mathbf{w}_{n-m} \end{bmatrix}$$

Assuming $\mathbf{h}_n$ and $\mathbf{w}_n$ contain zero-mean unit-variance i.i.d. Gaussian random variables, we have $E[\mathbf{h}_n \mathbf{h}_n^*] = E[\mathbf{w}_n \mathbf{w}_n^*] = \mathbf{I}_{MP}$. Furthermore, we assume that the channel is spatially white and temporally correlated, i.e., $E[\mathbf{h}_n \mathbf{h}_{n-k}^*] = \zeta_k \mathbf{I}_{MP}$. Conditioned

† The main results of this chapter also appear in the manuscript [34]

on $S_{n-m}, S_{n-m+1}, \ldots, S_n$, $\underline{\mathbf{x}}_n$ is a zero-mean Gaussian vector with auto-correlation matrix

$$\begin{aligned} R^{(m)} &= \frac{\rho}{M}\mathcal{S}_n E[\underline{\mathbf{h}}_n \underline{\mathbf{h}}_n^*]\mathcal{S}_n^* + \mathbf{I}_{(m+1)MP} \\ &= \mathcal{S}_n \left(\frac{\rho}{M} E[\underline{\mathbf{h}}_n \underline{\mathbf{h}}_n^*] + \mathbf{I}_{(m+1)MP}\right) \mathcal{S}_n^* \end{aligned} \tag{3.1}$$

Using the fact that $S_{n-m}, S_{n-m+1}, \ldots, S_n$ are unitary and $S_k = V_{z_k} S_{k-1}, k = n-m+1, \ldots, n$, $R^{(m)}$ can be expressed as

$$R^{(m)} = \begin{pmatrix} T_{0,0} & T_{0,1} & \ldots & T_{0,m} \\ T_{1,0} & T_{1,1} & \ldots & T_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ T_{m,0} & T_{m,1} & \ldots & T_{m,m} \end{pmatrix} \quad \text{where} \tag{3.2}$$

$$T_{j,i}^* = T_{i,j} = \left(\delta(i-j) + \frac{\rho}{M}\right) \zeta_{i-j}^* \, \mathbf{I}_P \otimes \prod_{k=n-i+1}^{n-j} V_{z_k}^*, \quad i \geq j$$

and $\delta(\cdot)$ is the Kronecker delta function. Note from (3.2) that the distribution of $\underline{\mathbf{x}}_n$ depends on $\{V_{z_k}\}_{k=n-m+1}^n$ rather than $\{S_k\}_{k=n-m}^n$. It can also be shown that,

$$R^{(m)-1} = -\begin{pmatrix} B_{0,0} & B_{0,1} & \ldots & B_{0,m} \\ B_{1,0} & B_{1,1} & \ldots & B_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m,0} & B_{m,1} & \ldots & B_{m,m} \end{pmatrix} \quad \text{where} \tag{3.3}$$

$$B_{j,i}^* = B_{i,j} = a_{j,i}^{(m)*} \, \mathbf{I}_P \otimes \prod_{k=n-i+1}^{n-j} V_{z_k}^*, \quad i \geq j$$

where the coefficients $\{a_{i,j}^{(m)}\}$ are given by the following lemma.

Lemma 1 *The coefficients $a_{i,j}^{(m)}$ satisfy the relation*

$$\underbrace{\begin{bmatrix} a_{0,0}^{(m)} & a_{0,1}^{(m)} & \ldots & a_{0,m}^{(m)} \\ a_{0,1}^{(m)*} & a_{1,1}^{(m)} & \ldots & a_{1,m}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{0,m}^{(m)*} & a_{1,m}^{(m)*} & \ldots & a_{m,m}^{(m)} \end{bmatrix}}_{A^{(m)}} = -\left(\mathbf{I}_{m+1} + \frac{\rho}{M}\begin{bmatrix} \zeta_0 & \zeta_1 & \ldots & \zeta_m \\ \zeta_1^* & \zeta_0 & \ldots & \zeta_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \zeta_m^* & \zeta_{m-1}^* & \ldots & \zeta_0 \end{bmatrix}\right)^{-1} \tag{3.4}$$

Proof: Using $E[\mathbf{h}_n \mathbf{h}_{n-k}^*] = \zeta_k \mathbf{I}_{MP}$ and $\underline{\mathbf{h}}_n = [\mathbf{h}_n^* \ \ldots \ \mathbf{h}_{n-m}^*]^*$, it is easy to show that

$$E[\underline{\mathbf{h}}_n \underline{\mathbf{h}}_n^*] = \underbrace{\begin{bmatrix} \zeta_0 & \zeta_1 & \cdots & \zeta_m \\ \zeta_1^* & \zeta_0 & \cdots & \zeta_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \zeta_m^* & \zeta_{m-1}^* & \cdots & \zeta_0 \end{bmatrix}}_{\Xi^{(m)}} \otimes \mathbf{I}_{MP} \tag{3.5}$$

Thus,

$$\begin{aligned} R^{(m)} &= \mathcal{S}_n \left(\frac{\rho}{M} E[\underline{\mathbf{h}}_n \underline{\mathbf{h}}_n^*] + \mathbf{I}_{(m+1)MP} \right) \mathcal{S}_n^* \\ &= \mathcal{S}_n \left(T^{(m)} \otimes \mathbf{I}_{MP} \right) \mathcal{S}_n^* \end{aligned} \tag{3.6}$$

where $T^{(m)} = \frac{\rho}{M} \Xi^{(m)} + \mathbf{I}_{m+1}$. Therefore,

$$R^{(m)^{-1}} = \mathcal{S}_n \left(T^{(m)^{-1}} \otimes \mathbf{I}_{MP} \right) \mathcal{S}_n^* \tag{3.7}$$

Expanding (3.7) and using $S_{n-i} = \left(\prod_{k=n-i+1}^{n-j} V_{z_k}^* \right) S_{n-j}, \ i > j$ we obtain (3.3) & (3.4). ■

Thus the ML detector of $\{z_k\}_{k=n-m+1}^n$ given $\{X_k\}_{k=n-m}^n$ (equivalently given $\underline{\mathbf{x}}_n$) becomes

$$\begin{aligned} \{\hat{z}_k\}_{k=n-m+1}^n &= \underset{z_n, z_{n-1}, \ldots \in \mathcal{L}}{\arg\max} \ p(\underline{\mathbf{x}}_n | V_{z_n}, V_{z_{n-1}}, \ldots, V_{z_{n-m+1}}) \\ &= \underset{z_n, z_{n-1}, \ldots \in \mathcal{L}}{\arg\max} \ e^{-\underline{\mathbf{x}}_n^* R^{(m)^{-1}} \underline{\mathbf{x}}_n} \\ &= \underset{z_n, z_{n-1}, \ldots \in \mathcal{L}}{\arg\max} \ \Re \left[\mathrm{tr} \left\{ \sum_{k=0}^{m-1} \sum_{i=0}^{m-k-1} a_{i,i+k+1}^{(m)} X_{n-i}^* \left(\prod_{j=i}^{i+k} V_{z_{n-j}} \right) X_{n-i-k-1} \right\} \right] \end{aligned} \tag{3.8}$$

Note that the multiple-symbol ML detector requires the knowledge of SNR in addition to fading correlations $\{\zeta_k\}_{k=0}^m$, because the coefficients $\{a_{j,k}^{(m)}\}$ depend on ρ. From (3.4) it can be easily seen that when channels in subsequent blocks are independent, i.e., $\zeta_k = 0 \ \forall k > 0$, then $a_{j,k}^{(m)} = 0 \ \forall j \neq k$, and hence the ML metric in (3.8) is zero, implying that differential encoding cannot be used—an intuitively satisfying

observation. Moreover, the detector (3.8) does not depend on the SNR ρ in case of a fixed channel, as proven in the following lemma.

Lemma 2 *If $\zeta_k = 1\ \forall k$, detector (3.8) becomes independent of ρ.*

Proof: To prove the lemma, we need to show that $a_{i,j}^{(m)} = a_{k,l}^{(m)} > 0\ \forall i,j,k,l \in \{0,\ldots,m\}\ \&\ i \neq j, k \neq l$, because in such a case the detector (3.8) becomes independent of the coefficients $a_{i,j}^{(m)}$, and hence ρ.

When $\zeta_k = 1\ \forall k$ the coefficients $a_{j,k}^{(m)}$ are given by (see (3.4))

$$\underbrace{\begin{bmatrix} a_{0,0}^{(m)} & \ldots & a_{0,m}^{(m)} \\ \vdots & \ddots & \vdots \\ a_{m,0}^{(m)} & \ldots & a_{m,m}^{(m)} \end{bmatrix}}_{A^{(m)}} \left(\frac{\rho}{M} \begin{bmatrix} 1 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \ldots & 1 \end{bmatrix} + \mathbf{I}_{m+1} \right) = -\mathbf{I}_{m+1}$$

Expanding the matrix equation we get the following set of equations

$$\frac{\rho}{M}\left(a_{i,i}^{(m)} + \sum_{k=0,k\neq i}^{m} a_{i,k}^{(m)}\right) + a_{i,j}^{(m)} = 0,\ i,j = 0,\ldots,m \quad i \neq j \tag{3.9}$$

$$\frac{\rho}{M}\left(a_{i,i}^{(m)} + \sum_{k=0,k\neq i}^{m} a_{k,i}^{(m)}\right) + a_{j,i}^{(m)} = 0,\ i,j = 0,\ldots,m \quad i \neq j \tag{3.10}$$

$$\left(1 + \frac{\rho}{M}\right) a_{i,i}^{(m)} + \frac{\rho}{M} \sum_{k=0,k\neq i}^{m} a_{i,k}^{(m)} = -1,\ i = 0,\ldots,m \tag{3.11}$$

It is easy to see that (3.9) implies $a_{i,j}^{(m)} = a_{i,k}^{(m)}\ \forall i,j,k = 0,\ldots,m\ \&\ j,k \neq i$. Similarly (3.10) results in $a_{j,i}^{(m)} = a_{k,i}^{(m)}\ \forall i,j,k = 0,\ldots,m\ \&\ j,k \neq i$. Together they imply that all the off-diagonal elements in A are equal, and hence real, since A is Hermitian. Subtracting (3.9) from (3.11) we get $a_{i,i}^{(m)} = a_{i,j}^{(m)} - 1\ \forall i = 0,\ldots,m$. Using this and the fact that all off-diagonal elements of A are same, the solution of (3.11) is $a_{i,j}^{(m)} = \frac{\rho}{1+(m+1)\rho} > 0\ \forall i \neq j$. ∎

3.2 Examples: 1-, 2-, and 3-symbol ML detectors

For the simulations in Section 3.5, we analyze the performance of 1-, 2-, and 3-symbol ML detectors. The m-symbol detector (3.8) can be simplified to (3.12), (3.13)

and (3.14) for $m = 1, 2$ and 3, respectively.

$$\hat{z}_n = \arg\max_{z_n \in \mathcal{L}} \Re\{\mathrm{tr}[\zeta_1 X_n^* V_{z_n} X_{n-1}]\} \tag{3.12}$$

$$\begin{aligned}\{\hat{z}_n, \hat{z}_{n-1}\} = \arg\max_{z_{n-1}, z_n \in \mathcal{L}} \Re[\mathrm{tr}\{&a_{0,1}^{(2)} X_n^* V_{z_n} X_{n-1} + a_{1,2}^{(2)} X_{n-1}^* V_{z_{n-1}} X_{n-2} \\ &+ a_{0,2}^{(2)} X_n^* V_{z_n} V_{z_{n-1}} X_{n-2}\}]\end{aligned} \tag{3.13}$$

$$\begin{aligned}\{\hat{z}_n, \hat{z}_{n-1}, \hat{z}_{n-2}\} = \arg\max_{z_n, z_{n-1}, z_{n-2} \in \mathcal{L}} \Re[\mathrm{tr}\{&a_{2,3}^{(3)} X_{n-2}^* V_{z_{n-2}} X_{n-3} \\ &+ a_{1,2}^{(3)} X_{n-1}^* V_{z_{n-1}} X_{n-2} + a_{0,1}^{(3)} X_n^* V_{z_n} X_{n-1} + a_{1,3}^{(3)} X_{n-1}^* V_{z_{n-1}} V_{z_{n-2}} X_{n-3} \\ &+ a_{0,2}^{(3)} X_n^* V_{z_n} V_{z_{n-1}} X_{n-2} + a_{0,3}^{(3)} X_n^* V_{z_n} V_{z_{n-1}} V_{z_{n-2}} X_{n-3}\}]\end{aligned} \tag{3.14}$$

While detector (3.12) is similar to detector (2.14), it exploits the known fading correlation ζ_1. When ζ_1 is real and positive, however, it does not affect the decision rule and can be removed, making (3.12) identical to (2.14). The main difference between detectors (3.13), (3.14) and one-symbol detector (3.12) is that (3.13) and (3.14) make use of additional channel parameters, and hence are expected to perform better than (3.12).

It is instructive to note that the transmitted matrix symbols S_{n-i} and S_{n-j}, for any $i < j$, are also differentially encoded: $\prod_{k=i}^{j-1} V_{z_{n-k}}$ is a valid source symbol due to the group nature of the source alphabet. Therefore all ML detection metrics are based on weighted sum of sub-metrics of the form $X_{n-i}^* \left(\prod_{k=i}^{j-1} V_{z_{n-k}}\right) X_{n-j}$.

Observe that it is not possible to write the detection metric as a sum of terms that contain strict subsets of the symbol set $\{V_{z_n}, \ldots, V_{z_{n-m+1}}\}$; there is always one term with the form $\mathrm{tr}\{X_n^* (\prod_{j=0}^{m-1} V_{z_{n-j}}) X_{n-m}\}$. Thus, maximization of the quantity in (3.8) can only be accomplished using a brute force search over all symbol combinations, yielding a complexity exponential in m. In other words, there does not appear to

exist a Viterbi-like algorithm for m-symbol ML detection that has complexity linear in m.

Though we have not derived ML detection rules for non-diagonal DUST constellations in *continuously*-fading channels, the simulation results in Section 3.5 confirm that the block-fading detectors (3.13) and (3.14) significantly outperform the standard single-symbol detector (2.14) under fast continuous fading. The coefficients $\{a_{i,j}^{(m)}\}$ used in this case would be recomputed with ζ_k defined such that $E[\mathbf{h}_{0,n}\mathbf{h}_{0,n-k}^*] = \zeta_k \mathbf{I}_{MP}$ for $\mathbf{h}_{0,n} = \mathrm{vec}(H_{0,n})$.

3.3 Suboptimal Sequence Detection

Practical applications require the detection of $N \gg 3$ symbols. Yet, as we have seen, the joint ML differential detector for N symbols has a complexity that is exponential in N. Thus, we are motivated to consider suboptimal N-symbol detection using some combination of m-symbol ML detectors for, say, $m \leq 3$.

It is instructive to note the difference between the N-symbol ML detector and any suboptimal N-symbol detector constructed from m-symbol ML detectors ($m < N$). From (3.8), we see that such suboptimal sequence detectors would use detection metrics that linearly combine terms based on subsets of the m-term set $\{V_{z_n}, \ldots, V_{z_{n-m+1}}\}$ for $n \in \{m-1, \ldots, N-1\}$. The combining coefficients might, e.g., be taken from the set $\{a_{k,l}^{(m)}\}$. In contrast, the optimal N-symbol detection metric linearly combines these terms with additional terms based on subsets of $\{V_{z_{N-1}}, \ldots, V_{z_0}\}$ that are ignored by the suboptimal detector. In addition, the N-symbol combining coefficients are $\{a_{k,l}^{(N)}\}$, which are, in general, different from $\{a_{k,l}^{(m)}\}$. Thus, the task of constructing a "good" N-symbol detector that employs (at most) m-symbol optimal detections

$(m < N)$ can be considered equivalent to the approximation of $\{a_{k,l}^{(N)}\}$ by a sparse coefficient set.

3.3.1 Non-overlapping Block Detection

A straightforward suboptimal N-symbol detector is the "non-overlapping" block detector [23], where the $(N+1)$-matrix observation sequence is divided into contiguous subsequences of length $m+1$, each of which overlaps its neighbor by one matrix-observation. m-symbol detection is then performed on each subsequence. This is equivalent to replacing the coefficients $\{a_{k,l}^{(N)}\}$ with $\{a_{k,l}^{(m)}\}$ where defined and otherwise with zeros. Simulation results show that this scheme reduces the error floor in fast-fading channel significantly. However, it is possible to further improve performance by dividing the observation sequence into blocks which overlap by more than one matrix-observation. Such schemes lead to different approximations of the coefficient set $\{a_{k,l}^{(N)}\}$.

3.3.2 Overlapping Block Detection

To detect the information sequence $\{z_n\}_{n=1}^{N}$ from the matrix-observation sequence $\{X_n\}_{n=0}^{N}$ using m-symbol detector $(m < N)$, we divide the the sequence $\{X_n\}_{n=0}^{N}$ into the sub-sequences $\mathcal{X}_{k,m} = \{X_k, X_{k+1}, \ldots, X_{k+m}\}$, $k = 0, p, 2p, \ldots, N-m$, where $p \in \{1, 2, \ldots, m\}$ and p divides $N-m$. In this case, neighboring sub-sequences overlap by $m-p+1$ matrix-observations. Note that $p = 1$ results in maximal overlap between blocks, whereas the case of $p = m$ is equivalent to non-overlapping block detection.

The information symbol sequence $\mathcal{V}_{k,m} = \{V_{z_{k+1}}, V_{z_{k+2}}, \ldots, V_{z_{k+m}}\}$ can be detected from the subsequence $\mathcal{X}_{k,m}$ by maximizing the ML detection metric in (3.8),

denoted henceforth as $L_{\text{opt}}(\mathcal{X}_{k,m}; \mathcal{V}_{k,m})$. However, optimal joint detection of $\mathcal{V}_{0,m}$ and $\mathcal{V}_{p,m}$ from $\mathcal{X}_{0,m}$ and $\mathcal{X}_{p,m}$ involves maximization of the increased complexity metric $L_{\text{opt}}(\mathcal{X}_{0,m+p}; \mathcal{V}_{0,m+p})$, because the additive noise processes in $\mathcal{X}_{0,m}$ and $\mathcal{X}_{p,m+p}$ are not mutually independent (unless $p > m$). For suboptimal detection of $\mathcal{V}_{0,N}$ from $\mathcal{X}_{0,N}$ we propose the following metric

$$L_{\text{subopt}}(\mathcal{X}_{0,N}; \mathcal{V}_{0,N}) = \sum_{k=0,p,2p,\ldots,N-m} L_{\text{opt}}(\mathcal{X}_{k,m}; \mathcal{V}_{k,m}) \tag{3.15}$$

The overlapping and non-overlapping block detection schemes are illustrated in Fig. 3.1.

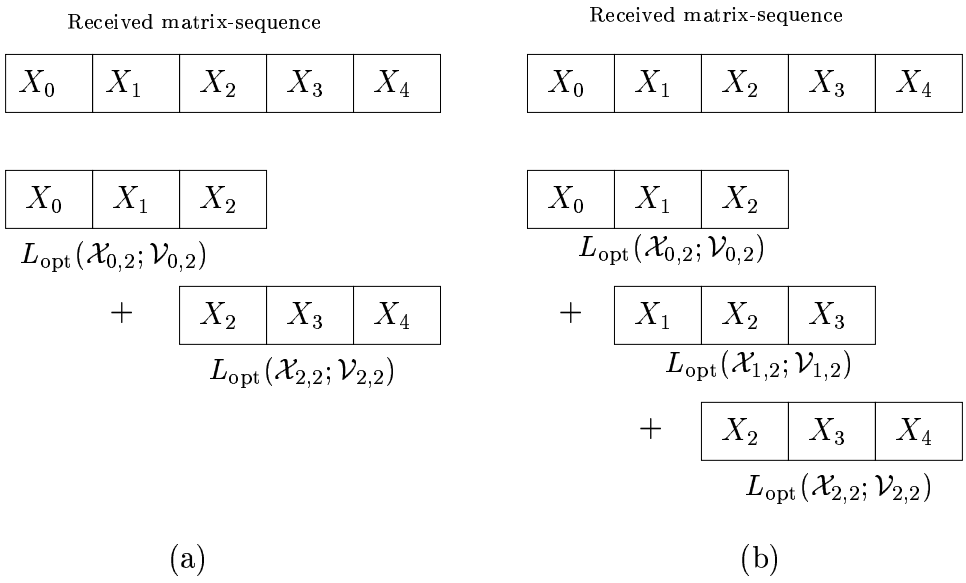

Figure 3.1: Suboptimal multiple-symbol detection with $N = 4, m = 2$ using (a) non-overlapping block detection scheme, (b) overlapping block detection scheme for $p = 1$

The suboptimal metric can be maximized using Viterbi algorithm, where at transitions indexed by $k = 0, p, 2p, \ldots, N-m$, states are defined by the hypotheses $\hat{\mathcal{V}}_{k,m-p}$,

outgoing branches by $\hat{\mathcal{V}}_{m+k-p,p-1}$, and the branch metric is given by $L_{\text{opt}}(\mathcal{X}_{k,m}; \hat{\mathcal{V}}_{k,m})$. The complexity thus becomes exponential in m and linear in N. Similar detection algorithms for single antenna DPSK can be found in [35, 36].

To evaluate the probability of error of the suboptimal detector, we will see in the next section that it will be helpful to write the suboptimal metric (3.15) in quadratic form as $L_{\text{subopt}}(\mathcal{X}_{0,N}; \mathcal{V}_{0,N}) = -\underline{\mathbf{x}}^* {R_v^{(N)}}^{-1} \underline{\mathbf{x}}$, such that the detection is performed by *minimizing* $\underline{\mathbf{x}}^* {R_v^{(N)}}^{-1} \underline{\mathbf{x}}$, where $\underline{\mathbf{x}} = [\mathbf{x}_N^*, \mathbf{x}_{N-1}^*, \ldots, \mathbf{x}_0^*]^*$. ${R_v^{(N)}}^{-1}$ is constructed as ${R^{(m)}}^{-1}$ in (3.3) but with the coefficients $\{a_{j,k}^{(N)}\}$ replaced by $\{\tilde{a}_{j,k}^{(N)}\}$, given by

$$\begin{pmatrix} \tilde{a}_{0,0}^{(N)} & \tilde{a}_{0,1}^{(N)} & \ldots & \tilde{a}_{0,N}^{(N)} \\ \tilde{a}_{1,0}^{(N)} & \tilde{a}_{1,1}^{(N)} & \ldots & \tilde{a}_{1,N}^{(N)} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{a}_{N,0}^{(N)} & \tilde{a}_{N,1}^{(N)} & \ldots & \tilde{a}_{N,N}^{(N)} \end{pmatrix} = \sum_{k=0,p,2p,\ldots,N-m} \begin{pmatrix} \boxed{\mathbf{0}_{k\times k}} & & \\ & \boxed{A_{m+1\times m+1}^{(m)}} & \\ & & \boxed{\mathbf{0}_{N-m-k\times N-m-k}} \end{pmatrix} \tag{3.16}$$

where $\mathbf{0}_{k\times k}$ denotes $k \times k$ all zero matrix and $A^{(m)}$ is defined in (3.4).

3.4 Error Performance

In this section we present the approximate BER performance of the detectors derived in previous sections. The pair-wise word error probability (PWEP), followed by approximate BER has been derived using the results from [1, 21, 28].

3.4.1 Pair-Wise Word Error Probability

The multiple-symbol detectors derived in this chapter, in general, detect the sequence $\mathbf{z} = \{z_k\}_{k=1}^N \in \mathcal{L}^N$ by minimizing the Hermitian quadratic $\underline{\mathbf{x}}^* R(\mathbf{z})^{-1} \underline{\mathbf{x}}$ over $\mathcal{L}^N$, where $\underline{\mathbf{x}} = [\mathbf{x}_N^*, \ldots, \mathbf{x}_0^*]^*$. The structure of $R(\mathbf{z})^{-1}$ is shown in (3.3), while the coefficients $\{a_{j,k}^{(N)}\}$ are computed using (3.4) and (3.16). p takes an appropriate value

in (3.16) depending on the detection method, i.e., $p = m$ for non-overlapping block detection and $p < m$ for overlapping block detection.

Given that the sequence $\mathbf{z}$ was transmitted, the detector will detect $\hat{\mathbf{z}}$, and thus make a decision error, if the random variable $Q = \underline{\mathbf{x}}^*(R(\hat{\mathbf{z}})^{-1} - R(\mathbf{z})^{-1})\underline{\mathbf{x}}$ is negative. The PWEP, denoted as $\Pr(\mathbf{z} \to \hat{\mathbf{z}})$, is therefore given by [21, 37]

$$\Pr(\mathbf{z} \to \hat{\mathbf{z}}) = \Pr(Q \leq 0) = \sum_{\substack{\text{poles } \omega = jp \\ p > 0}} \text{Res}\left[-\frac{\Phi_Q(\omega)}{\omega}\right]_{\omega = jp} \tag{3.17}$$

where $\Phi_Q(\omega) = E[e^{j\omega Q}]$ is the characteristic function of Q, given by [38]

$$\begin{aligned}\Phi_Q(\omega) &= \frac{1}{\det\left(\mathbf{I}_{(N+1)MP} - j\omega R^{(N)}(R(\hat{\mathbf{z}})^{-1} - R(\mathbf{z})^{-1})\right)} \\ &= \prod_{k=1}^{(N+1)MP} \frac{1}{1 - j\omega\lambda_k} \\ &= \prod_{k=1}^{(N+1)MP} \frac{1}{(-j\lambda_k)(\omega + j/\lambda_k)}\end{aligned} \tag{3.18}$$

and the summation is taken over the poles of $\frac{\Phi_Q(\omega)}{\omega}$ located on the upper half plane (UHP). $\{\lambda_k\}_{k=1}^{(N+1)MP}$ are eigenvalues of $R^{(N)}(R(\hat{\mathbf{z}})^{-1} - R(\mathbf{z})^{-1})$ and are real since both $R^{(N)}$ and $R(\hat{\mathbf{z}})^{-1} - R(\mathbf{z})^{-1}$ are Hermitian. Thus, the poles of $\frac{\Phi_Q(\omega)}{\omega}$ lie on the imaginary axis.

Computing the residues of $\frac{\Phi_Q(\omega)}{\omega}$ is straightforward if the poles on the UHP have a multiplicity of one. In this case, the residues are given by

$$\text{Res}\left[-\frac{\Phi_Q(\omega)}{\omega}\right]_{\omega = -j/\lambda_\ell} = -\prod_{k=1, k\neq\ell}^{(N+1)MP} \frac{1}{\lambda_k(1/\lambda_k - 1/\lambda_\ell)} \tag{3.19}$$

For the functions with repeated poles, computing the residue at a repeated pole involves higher order derivatives, which can be complicated. Since each $MP \times MP$ matrix sub-block in $R^{(N)}$, $R(\mathbf{z})^{-1}$ and $R(\hat{\mathbf{z}})^{-1}$ consists of a Kronecker product of $\mathbf{I}_P$ and

an $M \times M$ matrix, the eigenvalues $\{\lambda_k\}_{k=1}^{(N+1)MP}$ are P-fold repetition of the $(N+1)M$ eigenvalues of $R^{(N)}(R(\hat{\mathbf{z}})^{-1} - R(\mathbf{z})^{-1})|_{P=1}$. Hence poles of $\Phi_Q(\omega)$ have multiplicities equal to multiples of P.

Several methods for evaluation of (3.17), when the characteristic function has repeated poles, are available [39, 40]. We consider the method proposed in [28], where poles are perturbed by a small amount to eliminate multiplicity. This method produces a lower bound on the PWEP when all the concerned poles are moved away from the origin, and an upper bound when moved towards the origin. In our system, we replace the i^{th} repetition of $p_k = -\frac{1}{\lambda_k}, p_k > 0$ by $p_k \pm i\epsilon_k, \epsilon_k = 0.0025p_k$, then apply (3.17) and (3.19) to evaluate the PWEP. Numerical results in Section 3.5 show that the resulting lower and upper bounds on the probability are very close to each other, and hence the adopted method produces an accurate estimate of the PWEP.

3.4.2 PWEP: Imperfect Parameter Knowledge Case

The practical application of the detectors derived in this chapter is hindered by the reliance of the detectors on the knowledge of SNR and fading correlation ζ_k. So far in our analysis we assume that the detectors have perfect knowledge of these parameters. However, in practice these parameters are likely to be estimated, and therefore perfect parameter knowledge is unlikely due to estimation error. In this section we derive the PWEP expression under imperfect knowledge of these parameters.

As shown in previous sections, the detection metric of the detectors derived in this chapter can be written as a Hermitian quadratic product of a complex Gaussian vector, such that the Hermitian matrix is constructed based on the knowledge of fading correlation and SNR. Recall from Section 3.4.1 that, given that the information

sequence $\mathbf{z}$ was transmitted, the detector will detect $\hat{\mathbf{z}}$ if the random variable $Q = \underline{\mathbf{x}}^*(R(\hat{\mathbf{z}})^{-1} - R(\mathbf{z})^{-1})\underline{\mathbf{x}}$. In this case, both $R(\hat{\mathbf{z}})^{-1}$ and $R(\mathbf{z})^{-1}$ are constructed using (3.3), (3.4) and (3.16) such that (3.4) uses the *assumed/estimated* valued of ζ_k and ρ, whereas $R^{(N)} = E[\underline{\mathbf{x}}\underline{\mathbf{x}}^*]$ uses the *actual* ζ_k and ρ. Then, using the method described in Section 3.4.1 we can evaluate the PWEP for imperfect parameter knowledge case.

3.4.3 Approximate Bit Error Rate

In practice, bit error probability (BER) is more useful metric than PWEP of individual error events. To compute an approximation of the BER, or a bound on the same, we adopt the method used in [21]. We note that the constellation in [11] is symmetric, and hence the BER will be independent of the transmitted message sequence $\mathbf{z}$ if Gray labeling is used. Thus we assume $\mathbf{z} = \{0, 0, \ldots, 0\}$, i.e., $V_{z_k} = \mathbf{I}_M \ \forall k$. Noting that the message sequence $\mathbf{z} \in \mathcal{L}^N$ is encoded using ηMN bits, an upper bound on the BER can be obtained as

$$\mathbf{P}_b \leq \frac{1}{\eta MN} \sum_{k=1}^{2^{\eta MN}-1} d(\mathbf{z}, \hat{\mathbf{z}}^{(k)}) \Pr(\mathbf{z} \to \hat{\mathbf{z}}^{(k)}) \tag{3.20}$$

where $d(\cdot)$ denotes the Hamming distance between the codewords. However, for a long sequence, computing the bound (3.20) becomes prohibitively complex. In [21, 23] it has been shown that for DPSK, an accurate *approximation* of the BER is obtained by keeping only the dominant error events in (3.20). Extending these ideas to DUST, the dominant error events are defined by the sequences which maximize the following metric

$$\mu = \left| M + \sum_{k=1}^{N} \mathrm{tr}\left(S_k \hat{S}_k^*\right) \right|^2 \tag{3.21}$$

where $S_k = V_{z_k} S_{k-1},\ \hat{S}_k = V_{\hat{z}_k} \hat{S}_{k-1},\ S_0 = \hat{S}_0 = \mathbf{I}_M$.

Since $S_k = \mathbf{I}_M,\ k = 1, \ldots, N$, the sequences $\{\hat{S}_k\}_{k=1}^N$ that maximize (3.21) have the form $\{V_l, V_l, \ldots, V_l\}$, and $\{\mathbf{I}_M, \ldots, \mathbf{I}_M, V_l, \mathbf{I}_M, \ldots, \mathbf{I}_M\}$, where $l = \theta_1 = \arg\max_{k \in \mathcal{L}-\{0\}} \Re[\mathrm{tr}(V_k)]$. The corresponding information sequences $\{V_{\hat{z}_k}\}_{k=1}^M$ have the form $\{V_l, \mathbf{I}_M, \ldots, \mathbf{I}_M\}$, $\{\mathbf{I}_M, \ldots, \mathbf{I}_M, V_l, V_l^*, \mathbf{I}_M, \ldots, \mathbf{I}_M\}$ and $\{\mathbf{I}_M, \ldots, \mathbf{I}_M, V_l\}$. Note that, since $\mathcal{A}$ is a group, and hence satisfies an inverse property, $V_l \in \mathcal{A} \implies V_l^* \in \mathcal{A}$.

Although the dominant error events defined by the sequences above yield an accurate approximation of the BER for small N, additional error events need to be considered to avoid under-estimation of the BER for large N. Therefore we also consider the error events defined by sequences of the above form for $l = \theta_2$ such that $\theta_2 = \arg\max_{k \in \mathcal{L}-\{0,\theta_1\}} \Re[\mathrm{tr}(V_k)]$. For symmetric constellations, such as single antenna DPSK and multi-antenna diagonal constellations, θ_i may take multiple values. In such cases, error events for each distinct $\theta_i,\ i = 1, 2$ are considered in (3.20).

3.5 Simulations & Numerical Results

We evaluate the performance of the detectors in two types of channel: the "block fading channel" (2.3) and the "continuous fading channel" (2.2). The correlation between fading coefficients k symbols apart is given by $J_0(2\pi f_D T_s k)$ [41] in continuous fading channels, where $f_D T_s$ is the normalized Doppler frequency. In block fading channel, correlation between channel coefficients m matrix-symbols apart is given by $J_0(2\pi f_D T_s M m)$.

As shown in Chapter 2, the use of diagonal constellations in continuous-fading channels yields the same model as general constellations in block-fading channels, and hence detector performance with diagonal constellations is identical for these two channel types. Therefore, we present performance of the detectors with diagonal

and non-diagonal constellations in continuously fading channel. We consider a system with two transmit and two receive antennas with $\eta = 1$ and the constellation specified in Table 2.1

3.5.1 Simulation BER: Perfect Parameter Knowledge Case

First we present simulation results that demonstrate the performance improvement of multiple-symbol detectors over standard single-symbol detector. Detection is accomplished using the "overlapping block detection" method from Section 3.3 with $p = 1$ and based on 1-, 2-, and 3-symbol ML differential detectors (3.12), (3.13), and (3.14). Note that, for single-symbol detection, the overlapping and non-overlapping block detection schemes are identical. Results will be presented for detectors with exact knowledge of fading correlations, as well as for detectors which assume that the channel is fixed (marked by "fixed" in the figures). In the figures shown here, "1DD", "2DD", and "3DD" correspond to 1-, 2- and 3- symbol detection, respectively.

Fig. 3.2, where $f_D T_s = 0.1$, clearly illustrates the advantage of detectors which jointly detect multiple symbols *and* incorporate channel fading parameters. Note that detector (2.14), which ignores the fading correlation, succumbs to a very high error floor. Detector (3.12), designed to incorporate fading correlation into single-symbol detection, performs the same as (2.14) does, since the Rayleigh flat-fading model used in our simulations leads to $\zeta_1 > 0$. Thus, both forms of single-symbol detection perform very sub-optimally in the fading environment. The 2-symbol detector (3.13) which incorporates fading correlation exhibits considerably improved performance, although still succumbing to an error floor. Meanwhile, the 2-symbol detector which

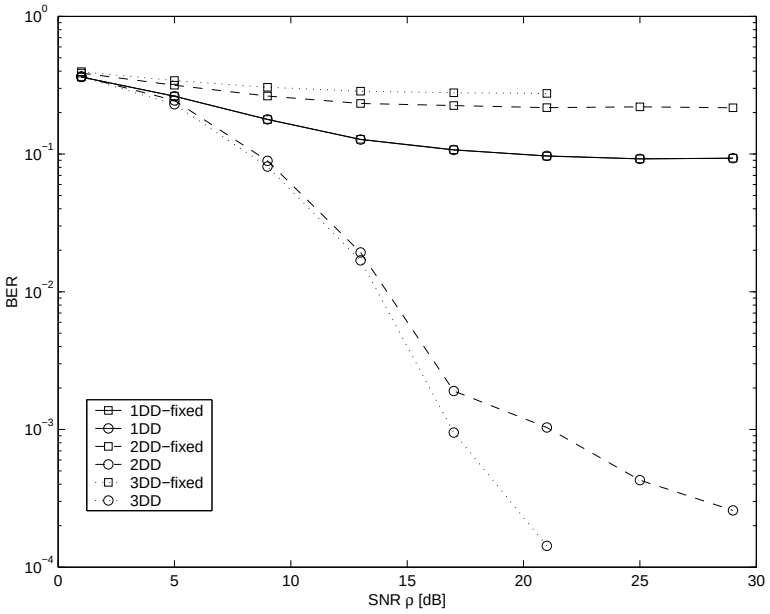

Figure 3.2: The diagonal constellation in continuous fading channel with $f_D T_s = 0.1$.

ignores fading correlation performs even worse than the one-symbol detector. Performance increases dramatically with the 3-symbol detector that incorporates fading, and decreases with the 3-symbol detector that ignores fading. As hinted by the plot, even the good 3-symbol detector will succumb to an error floor at high-enough SNR. The important point, however, is that the error floor has been pushed outside of the expected operating range.

Figure 3.3, corresponding to normalized Doppler frequencies of $f_D T_s = 0.05$, mimics the results of Fig. 3.2 but in a less pronounced fashion. Again we see the improvement associated with multiple-symbol ML differential detectors that incorporate fading correlation.

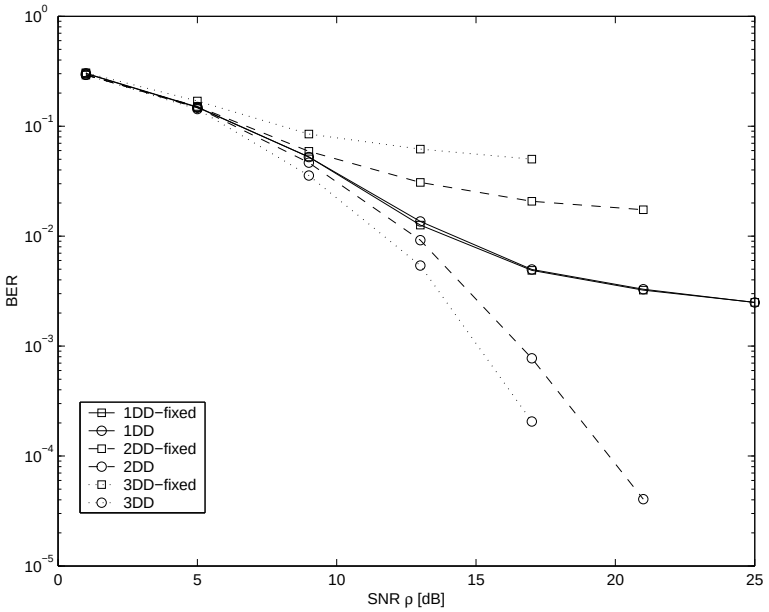

Figure 3.3: The diagonal constellation in continuous fading channel with $f_D T_s = 0.05$.

Next, the performance of the detectors has been evaluated in continuously-fading channels with the non-diagonal constellation to illustrate the performance loss due to approximation of system model (2.2) by (2.3). Recall that the non-diagonal constellation shown in Table 2.1 is generated by right multiplying the diagonal constellation by a fixed non-diagonal unitary matrix. Because such an operation does not change the product distance of the constellation [11], the comparison of diagonal to non-diagonal constellations is fair.

Figures 3.4 and 3.5 illustrate the performance of the detectors with the non-diagonal constellation in continuously-fading channels with $f_D T_s = 0.1$ and 0.05, respectively, and compare them with the performance of 3-symbol detectors with the diagonal constellation. Although a performance loss is incurred due to neglecting the

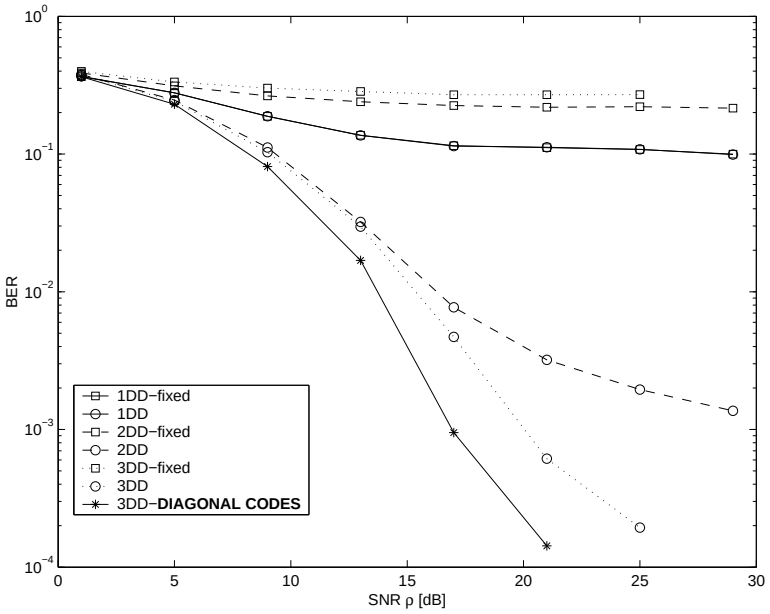

Figure 3.4: The non-diagonal constellation in continuously-fading channel with $f_D T_s = 0.1$.

channel variation within the matrix-symbol interval compared to the performance of the detectors with diagonal constellations, in both cases 3- and 2-symbol detectors that incorporate knowledge of fading correlations perform much better than the single-symbol detector.

The above observations lead us to important conclusions about the detectors described in this paper. First, multiple-symbol detection is essential to combat fading channels since the detectors (2.14) and (3.12) are often equivalent. While performing the suboptimal sequence detection, increasing the sub-block length improves the performance in terms of SNR loss and increases the robustness to continuous fading when non-diagonal constellations are used. Generally, the faster the fading, the more

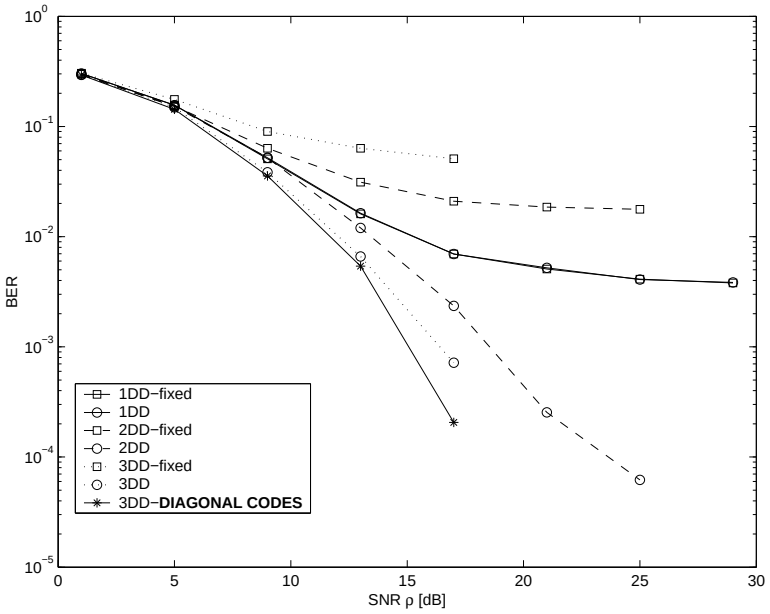

Figure 3.5: The non-diagonal constellation in continuously-fading channel with $f_D T_s = 0.05$.

symbols are required in the sub-blocks to push the error floor out of the operating range.

3.5.2 Simulation BER: Imperfect Parameter Knowledge Case

The practical application of the detectors derived in the paper is encouraged by the robustness of these detectors against imperfect knowledge of fading correlation *and* SNR ρ in continuous-fading channels, as demonstrated in the simulation results below. Diagonal constellations have been used for these simulations.

Figure 3.6 investigates the performance of the 3-symbol detector when the *actual* normalized Doppler frequency is 0.075 and the receiver has imperfect knowledge of $f_D T_s$ and SNR. As expected, the mismatch between assumed and actual Doppler

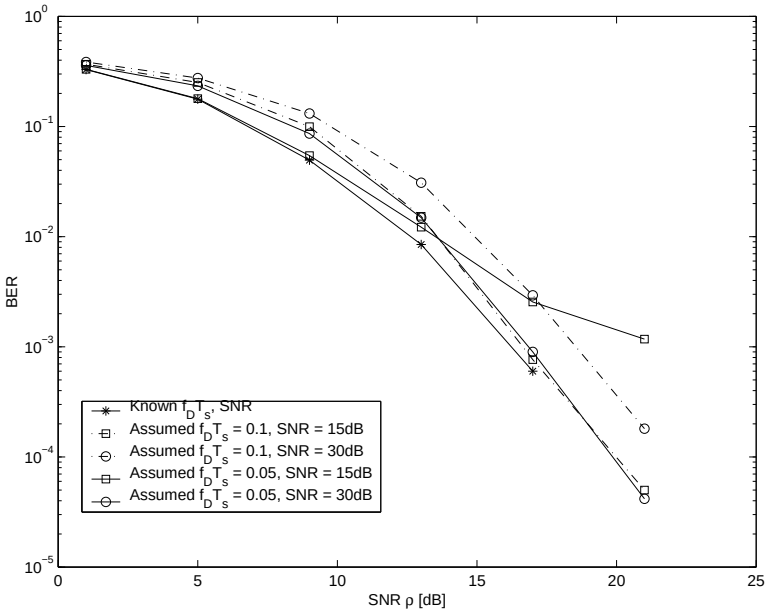

Figure 3.6: 3-symbol detector in continuously-fading channel $f_D T_s = 0.075$ (diagonal constellation)

frequency and SNR results in performance loss. An error floor appears when the receiver assumed that $f_D T_s = 0.05$ and SNR = 15dB. In this case the receiver tried to mimic the performance of the optimal receiver at 15dB SNR rather than at high SNR where error floor manifests. In all other cases, the performance loss is characterized only by SNR loss.

We can conclude from Fig. 3.6 that when the receiver underestimates the speed of the channel variation and operates assuming low SNR (15 dB), it succumbs to an error floor, whereas when high SNR (30 dB) is assumed, the performance is much better. High SNR loss characterizes the performance of the receiver which overestimates the Doppler spread and operates assuming high SNR (30 dB), while performance is much better when low SNR (15 dB) is assumed.

3.5.3 Comparison of Analytical and Simulation Results

Now we present analytical results that investigate the performance of m-symbol ML detectors for $m \gg 1$. As described in Section 3.4, approximate BER was computed assuming that the bit-errors contributed by the highly correlated sequences dominate. To establish the validity of this assumption, Fig. 3.7 compares analytical to simulated BER for 1-, 2-, and 3-symbol ML "non-overlapping" block detectors.

Figure 3.7 verifies the fact that the lower and upper bounds on the approximate BER are very close to each other. Analytical BER closely matches the simulated BER when BER is low, while analytical results over-estimate the BER for higher BER, which can be attributed to the fact that the union bound is used to approximate the BER.

Figure 3.8 demonstrates the validity of analytical BER results for 51-symbol suboptimal detector based on 2-symbol (2DD) and 3-symbol (3DD) ML detector. Note that overlapping and non-overlapping block detection schemes are identical when single-symbol ML detector is employed, and therefore the BER for single-symbol detector is omitted in Fig. 3.8. Again, we see that analytical results tend to overestimate the BER when BER is low.

Figure 3.8 shows that analytical BER tends to underestimate the simulated BER, which can be attributed to the fact that selection of dominant error events is based on PWEP, whereas contribution of an error event to overall BER depends on product of the Hamming distance and PWEP. However, the analytical results are still useful for predicting detector performance trends.

Figure 3.9 compares the analytical performance of 51-symbol ML detector and suboptimal detector based on 3-symbol sub-blocks for $p = 1, 3$ (i.e., "overlapping"

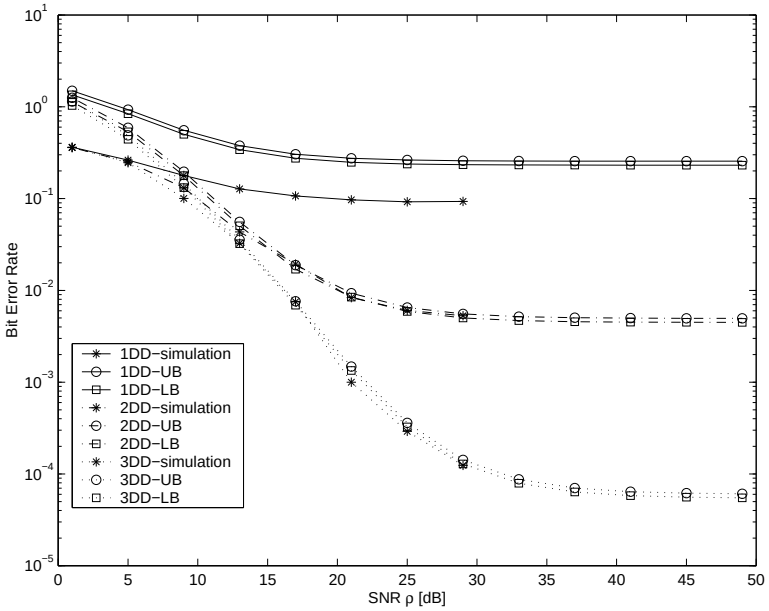

Figure 3.7: Comparison of analytical and simulated BER for $f_D T_s = 0.1$, non-overlapping block detection and diagonal constellation

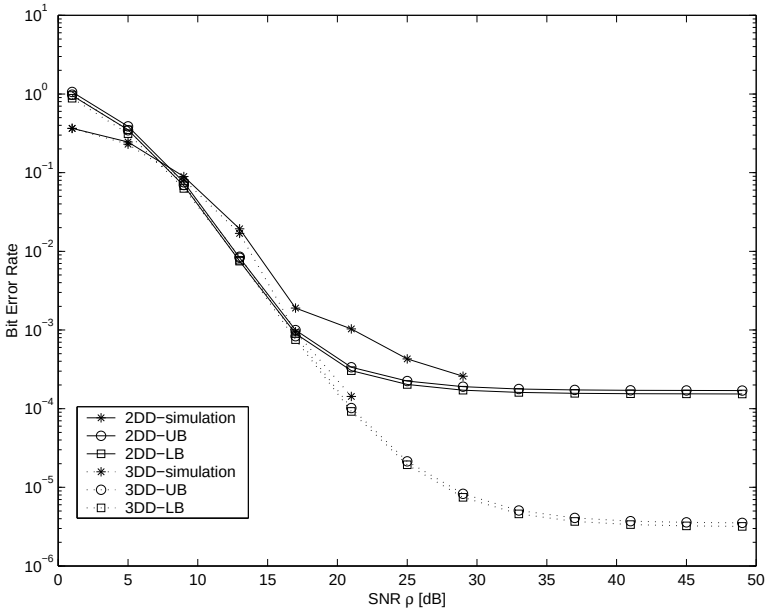

Figure 3.8: Comparison of analytical and simulated BER for $f_D T_s = 0.1$, overlapping block detection and diagonal constellation

and "non-overlapping", respectively) and $f_D T_s = 0.1$. The suboptimal "overlapping" detector ($p = 1$) provides improved performance over "non-overlapping" detector at the cost of minor increase in complexity. Both suboptimal detectors succumb to error floor whereas the optimal 51-symbol detector pushes the error floor outside the observed SNR range. The computational complexity of the optimal detector is, however, prohibitive.

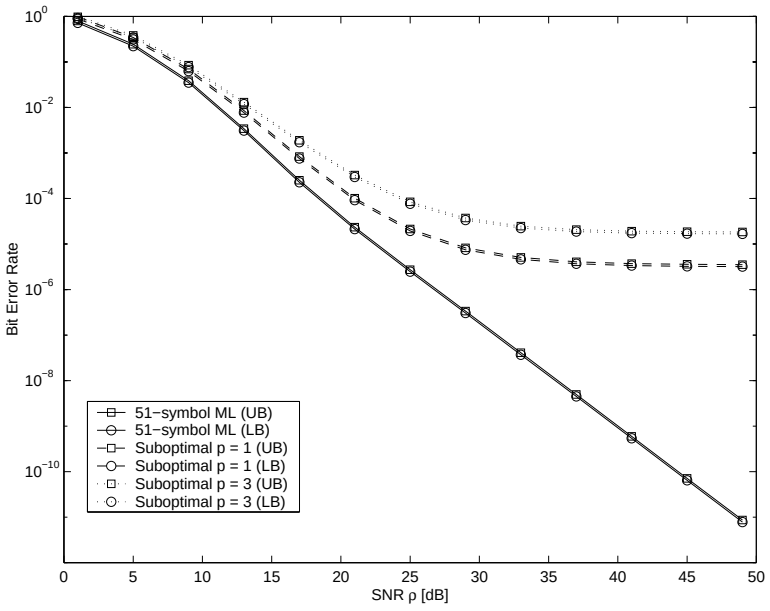

Figure 3.9: Comparison of 51-symbol ML detector and suboptimal detectors based on 3-symbol ML detector: $f_D T_s = 0.1$

Now we demonstrate the performance of 3-symbol ML detector based 51-symbol suboptimal detector for different Doppler frequencies to show the performance degradation due to Doppler spread. Observe in Fig. 3.10 that the performance degradation due to increased Doppler spread is significant at higher SNR, when the error

due to channel variation dominates that due to additive noise. Because the effective symbol-level normalized Doppler spread for DUST is $f_D T_s M$, increasing the number of transmit antennas may degrade the performance.

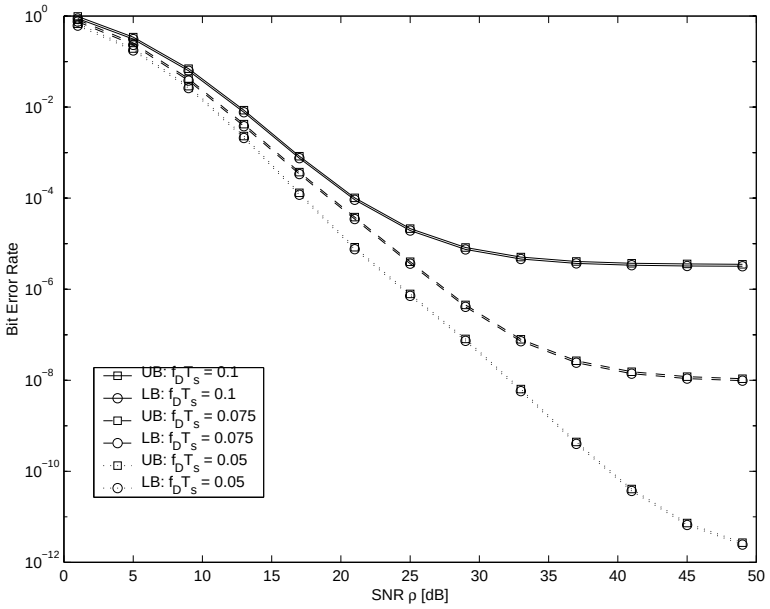

Figure 3.10: Performance of 51-symbol suboptimal detector based on 3-symbol ML detector ($p = 1$) for different Doppler frequencies

3.6 Conclusions

In this chapter, we have demonstrated, via simulation as well as theoretical error performance analysis, the efficacy of multiple-symbol ML differential detection that incorporates channel fading parameters. In fact, we have shown that multiple-symbol (versus single-symbol) detection is essential to performance enhancement in fading channels. For detection of long sequences, low-complexity suboptimal detectors have

been derived which significantly reduce the error floor. Our multiple-symbol detection rules, which assume the channels to be block fading when non-diagonal constellations are used, have been shown to improve performance in continuously-fading channels as well. In addition, they have been shown to be quite robust to imperfect knowledge of fading correlation and SNR.

CHAPTER 4

DECISION FEEDBACK DIFFERENTIAL DETECTION[†]

In this chapter, we present decision-feedback differential detectors (DFDD) for DUST in fast Rayleigh-fading channels. As before, we assume that the receiver knows the fading correlation and SNR. Under this assumption, a decision-feedback detector is derived from the multiple-symbol ML detector (3.8) and shown to be equivalent to a Wiener-filter based channel predictor followed by coherent ML detector. We derive the exact and Chernoff bound expressions for pairwise word error probability (PWEP) under the cases of perfect and imperfect knowledge of fading correlation and SNR. An approximate expression for BER is derived from the PWEP which is close agreement with the simulation results.

4.1 Decision-Feedback Differential Detection

4.1.1 DFDD from Multi-symbol ML Detection

Multiple symbol differential detection of DPSK [21, 29] and DUST [Chapter 3] has been proposed as an effective way to enhance performance in correlated Rayleigh fading. Recall from Chapter 3 that, assuming the channel to be spatially white and temporally correlated, the joint ML detector of $\{z_k\}_{k=n-m+1}^{n}$ given the observation

[†]The main results of this chapter also appear in the manuscript [42]

sequence $\{X_k\}_{k=n-m}^{n}$ is

$$\{\hat{z}_k\}_{k=n-m+1}^{n} = \underset{z_n, z_{n-1}, \cdots \in \mathcal{L}}{\arg\max} \Re \left[\mathrm{tr} \left\{ \sum_{k=0}^{m-1} \sum_{i=0}^{m-k-1} a_{i,i+k+1}^{(m)} X_{n-i}^{*} \left(\prod_{j=i}^{i+k} V_{z_{n-j}} \right) X_{n-i-k-1} \right\} \right] \tag{4.1}$$

where the combining coefficients $\{a_{i,j}^{(m)}\}$ are given by

$$\begin{bmatrix} a_{0,0}^{(m)} & a_{0,1}^{(m)} & \cdots & a_{0,m}^{(m)} \\ a_{0,1}^{(m)*} & a_{1,1}^{(m)} & \cdots & a_{1,m}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{0,m}^{(m)\,*} & a_{1,m}^{(m)\,*} & \cdots & a_{m,m}^{(m)} \end{bmatrix} = - \left(\mathbf{I}_{m+1} + \frac{\rho}{M} \underbrace{\begin{bmatrix} \zeta_0 & \zeta_1 & \cdots & \zeta_m \\ \zeta_1^{*} & \zeta_0 & \cdots & \zeta_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \zeta_m^{*} & \zeta_{m-1}^{*} & \cdots & \zeta_0 \end{bmatrix}}_{\Xi^{(m)}} \right)^{-1} \tag{4.2}$$

where ζ_k is defined such that $E[\mathbf{h}_n \mathbf{h}_{n-k}^{*}] = \zeta_k \mathbf{I}_{MP}$ for $\mathbf{h}_k = \mathrm{vec}(H_k)$. Note that $\zeta_0 = 1$ since H_k contains unit-variance entries.

The DFDD can be derived from the m-symbol ML detector (4.1) by feeding back the past decisions, i.e., replacing the hypotheses $\{z_k\}_{k=n-m+1}^{n-1}$ by the previously detected symbols $\{\hat{z}_k\}_{k=n-m+1}^{n-1}$ in the n^{th} symbol interval and maximizing the term on the right of (4.1) with respect to z_n alone. Of course, terms on the right of (4.1) that are not functions of z_n can be ignored. We denote the DFDD derived from m-symbol ML detector as the m-DFDD, given by

$$\hat{z}_n = \underset{z_n \in \mathcal{L}}{\arg\max} \Re \left[\mathrm{tr} \left\{ \sum_{k=0}^{m-1} a_{0,k+1}^{(m)} X_n^{*} \left(V_{z_n} \prod_{j=1}^{k} V_{\hat{z}_{n-j}} \right) X_{n-k-1} \right\} \right] \tag{4.3}$$

which is the generalization of the DFDD rules for DPSK [22] and diagonal constellations [29] to the general class of unitary constellations. It can be shown that when $H_n = H_{n-1}$, $a_{0,1}^{(1)}$ is real and positive (see Lemma 2). Therefore, for $m = 1$ and under the assumption that $H_n = H_{n-1}$, (4.3) reduces to the standard detector (2.14), making (2.14) a special case of the class of detectors derived in this chapter.

4.1.2 DFDD from MMSE channel prediction

Under the assumption of correct past decisions, i.e., $\hat{z}_k = z_k, k = n-m+1, \ldots, n-1$, we now derive a DFDD based on MMSE channel prediction. Although our derivations assume correct past decisions, the DFDD can be implemented in practice using symbol estimates.

In the symbol interval n we observe X_n for detection of V_{z_n}, and the observations $\{X_k\}_{k=n-m}^{n-1}$ are collected for estimation of H_n. For estimation of H_n we assume, for the moment, that $\{S_k\}_{k=n-m}^{n-1}$ is known (without error) at the receiver. We will see, however, that the resulting detector will depend on $\{V_{z_k}\}_{k=n-m+1}^{n}$ rather than $\{S_k\}_{k=n-m}^{n-1}$. Since $X_k = \sqrt{\frac{\rho}{M}} S_k H_k + W_k,\ k = n-m, \ldots, n-1$, denoting $\mathbf{h}_k = \mathrm{vec}(H_k)$, $\mathbf{x}_k = \mathrm{vec}(X_k)$, and $\mathbf{w}_k = \mathrm{vec}(W_k)$, we can write

$$\underbrace{\begin{bmatrix} \mathbf{x}_{n-1} \\ \vdots \\ \mathbf{x}_{n-m} \end{bmatrix}}_{\underline{\mathbf{x}}_{n-1}} = \sqrt{\frac{\rho}{M}} \underbrace{\begin{bmatrix} \mathbf{I}_P \otimes S_{n-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_P \otimes S_{n-m} \end{bmatrix}}_{\mathcal{S}_{n-1}} \underbrace{\begin{bmatrix} \mathbf{h}_{n-1} \\ \vdots \\ \mathbf{h}_{n-m} \end{bmatrix}}_{\underline{\mathbf{h}}_{n-1}} + \underbrace{\begin{bmatrix} \mathbf{w}_{n-1} \\ \vdots \\ \mathbf{w}_{n-m} \end{bmatrix}}_{\underline{\mathbf{w}}_{n-1}} \tag{4.4}$$

Assume that the MMSE estimate of $\mathbf{h}_n$, $\hat{\mathbf{h}}_n = \mathrm{vec}(\hat{H}_n|_{n-m}^{n-1})$ is given by $\hat{\mathbf{h}}_n = \sum_{k=1}^{m} B_k^* \mathbf{x}_{n-k} = B^* \underline{\mathbf{x}}_{n-1}$, where $B = [B_1^*\ B_2^*\ \ldots\ ; B_m^*]^*$. It can be shown straightforwardly that the mean square error $J(B) = E[\|\hat{\mathbf{h}}_n - \mathbf{h}_n\|^2]$ is minimized by selecting (for proof, see Appendix A.1)

$$B = \sqrt{\frac{\rho}{M}} \mathcal{S}_{n-1} \left((T^{(m-1)^{-1}} \mathbf{g}) \otimes \mathbf{I}_{MP} \right) \tag{4.5}$$

$$\mathbf{g} = [\zeta_1\ \zeta_2\ \ldots\ \zeta_m]^*$$

$$T^{(m)} = \mathbf{I}_{m+1} + \frac{\rho}{M} \Xi^{(m)} \tag{4.6}$$

where $\Xi^{(m)}$ is defined in (4.2) and the minimum mean-square error $J_{\min}$ is given by

$$J_{\min} = MP\left(1 - \sigma_{\hat{H}}^2\right) \tag{4.7}$$

$$\sigma^2_{\hat{H}} \;=\; \frac{\rho}{M}\mathbf{g}^* T^{(m-1)^{-1}}\mathbf{g} \tag{4.8}$$

The choice of the notation "$\sigma^2_{\hat{H}}$" will be made clear later. Now define $F^{(m-1)} = \sqrt{\frac{\rho}{M}}T^{(m-1)^{-1}}\mathbf{g}$ so that $B = \mathcal{S}_{n-1}\left(F^{(m-1)} \otimes \mathbf{I}_{MP}\right)$. Then, letting $\xi = \sqrt{\frac{M}{\rho}} + \sqrt{\frac{\rho}{M}}(1 - \sigma^2_{\hat{H}}) > 0$, we can write [43] (for proof, see Appendix A.2)

$$\begin{bmatrix} \xi\mathbf{I}_{MP} \\ \mathbf{0} \end{bmatrix} = \underbrace{\begin{bmatrix} \left(\frac{\rho}{M}+1\right)\mathbf{I}_{MP} & \frac{\rho}{M}\mathbf{g}^* \otimes \mathbf{I}_{MP} \\ \frac{\rho}{M}\mathbf{g} \otimes \mathbf{I}_{MP} & T^{(m-1)} \otimes \mathbf{I}_{MP} \end{bmatrix}}_{T^{(m)} \otimes \mathbf{I}_{MP}} \begin{bmatrix} \sqrt{\frac{M}{\rho}}\mathbf{I}_{MP} \\ -F^{(m-1)} \otimes \mathbf{I}_{MP} \end{bmatrix} \tag{4.9}$$

which implies

$$F^{(m-1)} \otimes \mathbf{I}_{MP} = \xi \begin{bmatrix} a^{(m)*}_{0,1}\mathbf{I}_{MP} \\ \vdots \\ a^{(m)\,*}_{0,m}\mathbf{I}_{MP} \end{bmatrix} \tag{4.10}$$

Therefore, the channel estimate $\hat{\mathbf{h}}_n$ is given by

$$\begin{aligned} \hat{\mathbf{h}}_n &= \left(F^{(m-1)^*} \otimes \mathbf{I}_{MP}\right)\mathcal{S}^*_{n-1}\underline{\mathbf{x}}_{n-1} \\ &= \xi \sum_{k=0}^{m-1} a^{(m)}_{0,k+1}\left(\mathbf{I}_P \otimes S^*_{n-k-1}\right)\mathbf{x}_{n-k-1} \end{aligned} \tag{4.11}$$

Defining the estimation error $\tilde{\mathbf{h}}_n = \mathbf{h}_n - \hat{\mathbf{h}}_n$, the system model (2.3) can be rewritten as

$$\mathbf{x}_n = \sqrt{\frac{\rho}{M}}\left(\mathbf{I}_P \otimes V_{z_n}\right)\left(\mathbf{I}_P \otimes S_{n-1}\right)\hat{\mathbf{h}}_n + \underbrace{\sqrt{\frac{\rho}{M}}\left(\mathbf{I}_P \otimes S_n\right)\tilde{\mathbf{h}}_n + \mathbf{w}_n}_{\tilde{\mathbf{w}}_n} \tag{4.12}$$

It is not hard to show that $E[\hat{\mathbf{h}}_n\hat{\mathbf{h}}^*_n] = \sigma^2_{\hat{H}}\mathbf{I}_{MP}$ and $E[\hat{\mathbf{h}}_n\tilde{\mathbf{h}}^*_n] = \mathbf{0}$, which implies $E[\hat{\mathbf{h}}_n\tilde{\mathbf{w}}^*_n] = \mathbf{0}$. It can also be shown that $E[\tilde{\mathbf{w}}_n\tilde{\mathbf{w}}^*_n] = \sigma^2_{\tilde{W}}\mathbf{I}_{MP}$, where $\sigma^2_{\tilde{W}} = 1 + \frac{\rho}{M}(1 - \sigma^2_{\hat{H}})$ (for proof, see Appendix A.3).

Note that computation of the estimate $\hat{\mathbf{h}}_n$ requires the knowledge of the symbols $\{S_k\}_{k=n-m}^{n-1}$ which are unknown to the receiver, whereas $\left(\mathbf{I}_P \otimes S_{n-1}\right)\hat{\mathbf{h}}_n$ depends on the previously detected (error-free) symbols $\{V_{z_k}\}_{k=n-m+1}^{n-1}$. Moreover, since the entries of

$\tilde{\mathbf{h}}_n$ are zero-mean i.i.d. Gaussian variables, multiplication by the unitary matrix does not change its distribution. Therefore, from (4.12), the ML detector of V_{z_n} for known $(\mathbf{I}_P \otimes S_{n-1})\,\hat{\mathbf{h}}_n$ is (for proof, see Appendix A.4)

$$\begin{aligned}\hat{z}_n &= \arg\max_{z_n\in\mathcal{L}} \Re\left[\mathbf{x}_n^* \left(\mathbf{I}_P \otimes V_{z_n}\right)\left(\mathbf{I}_P \otimes S_{n-1}\right)\hat{\mathbf{h}}_n\right] \\ &= \arg\max_{z_n\in\mathcal{L}} \Re\left[\mathrm{tr}\left\{\xi \sum_{k=0}^{m-1} a_{0,k+1}^{(m)} X_n^* \left(V_{z_n}\prod_{j=1}^{k} V_{z_{n-j}}\right) X_{n-k-1}\right\}\right] \qquad (4.13)\end{aligned}$$

which is identical to (4.3) under perfect past decisions since $\xi > 0$. Since perfect past decisions are not available in practice, (4.13) would be implemented using $\{\hat{z}_k\}_{k=n-m+1}^{n-1}$ in place of $\{z_k\}_{k=n-m+1}^{n-1}$, making it identical to (4.3).

Therefore, we have shown that the DFDD rules for DUST derived from multiple-symbol ML detection and from MMSE-optimal linear prediction are identical under perfect past decisions.

4.1.3 Comments

1. Though our derivations of DFDD for non-diagonal constellations assume block-fading channels, simulation results in Section 4.3 confirm that the m-DFDD for $m > 1$ derived in this paper significantly outperforms the standard single symbol detector under fast continuous-fading as well. Note that the coefficients $\{a_{i,j}^{(m)}\}$ used in the continuous-fading case would be recomputed with ζ_k defined such that $E[\mathbf{h}_{0,n}\mathbf{h}_{0,n-k}^*] = \zeta_k \mathbf{I}_{MP}$ for $\mathbf{h}_{0,n} = \mathrm{vec}(H_{0,n})$.

2. The m-symbol ML detection rule (4.1) can be re-written as (for proof, see Appendix A.5)

$$\{\hat{z}_k\}_{k=n-m+1}^{n} = \arg\max_{z_n, z_{n-1},\ldots\in\mathcal{L}} \Re\left[\mathrm{tr}\left\{\sum_{k=0}^{m-1} X_{n-k}^* V_{z_{n-k}} S_{n-k-1}(\hat{H}_{n-k}|_{n-m}^{n-k-1})\right\}\right] (4.14)$$

$$\hat{H}_{n-k}|_{n-m}^{n-k-1} = \sum_{\ell=n-m}^{n-k-1} a_{k,n-\ell}^{(m)} S_\ell^* X_\ell \tag{4.15}$$

It is easy to show that the m-symbol ML detection rule for detecting $\{z_k\}_{k=n-m+1}^{n}$ from $\{X_k\}_{k=n-m+1}^{n}$ for known $S_{n-k-1}H_{n-k}$ can be written as (4.14) but with the replacement $\hat{H}_{n-k}|_{n-m}^{n-k-1} = H_{n-k}$. In fact, if $S_{n-k-1}\hat{H}_{n-k}|_{n-m}^{n-k-1}$ are any channel estimates such that the sum of the channel estimation error and the additive noise is white, then (4.14) is still the joint ML detector. Thus we see that the multi-symbol ML detection rule has an estimator/detector structure, where the channel estimation is performed using (4.15).

Observe that in the m-symbol ML detection rule, H_{n-k} is estimated using a length-$(m-k)$ channel estimator, i.e., the channels $H_n, H_{n-1}, \ldots, H_{n-m+1}$ are estimated using length-$m, m-1, \ldots, 1$ estimators, respectively. The performance of the ML detector can be improved by increasing the estimator lengths, which can be easily accomplished by increasing m, which, however, increases the complexity exponentially. Note that the performance can still be improved without increasing m by employing equal-length estimators for H_{n-k} $\forall k$, the length being equal to m. Such equal length estimators are inherent to the "overlapping block detection" scheme of Chapter 3, which can be implemented using a Viterbi-like algorithm and which performs better than the "non-overlapping block detection" scheme.

3. From (4.11), we see that the channel estimator embedded in the m-DFDD can be described as a filter with impulse response $\{a_{0,k}^{(m)}\}_{k=1}^{m}$, input $Y_k = S_k^* X_k = \sqrt{\frac{\rho}{M}} H_k + S_k^* W_k$, and output $\hat{H}_k$. Recall that $\{H_k\}$, the response of a fading channel, is typically a lowpass random process whose bandwidth is defined by

the Doppler spread of the channel [41]. Therefore, the filter acts to attenuate the wideband additive noise $S_k^* W_k$ and predict the desired process.

The passband width of the optimal linear predictor will be commensurate with the desired-process bandwidth in the presence of noise and will shrink as the noise power increases and expand as noise power decreases. Thus the effect of under-estimating the Doppler spread can be somewhat countered by over-estimating the SNR ρ and vice versa. However, simultaneous over (or under) estimation of both Doppler spread and SNR can result in severe performance degradation. Note that the excess prediction error due to over-estimation of Doppler frequency is directly proportional to the noise power, and therefore becomes asymptotically negligible as SNR increases. On the other hand, error due to under-estimation of Doppler frequency does not decrease with increasing SNR, and so the estimation error succumbs to a floor. Note that, when the prediction filter length m is small and the receiver under-estimates the Doppler spread, the performance degradation is small since the filter has a wide "transition band". On the other hand, larger m allows a sharper transition band and hence reduces robustness to Doppler mismatch. These notions are confirmed by the numerical results in Sections 4.2 & 4.3.

4.2 Error Performance

In this section we first derive the exact PWEP and Chernoff upper bound expressions for genie-aided (i.e., perfect past decisions) DFDD, and later use them to approximate the BER. Since the DFDD requires knowledge of fading correlation and

SNR, we also derive the exact and Chernoff bound expressions for PWEP and approximate the BER under imperfect knowledge of these parameters.

4.2.1 Exact PWEP: perfect knowledge of ζ_k and ρ

Defining $\breve{\mathbf{h}}_n = \frac{1}{\sigma_{\hat{H}}}(\mathbf{I}_P \otimes S_{n-1})\hat{\mathbf{h}}_n$, $\breve{\mathbf{w}}_n = \tilde{\mathbf{w}}_n/\sigma_{\tilde{W}}$, and $\breve{\mathbf{x}}_n = \mathbf{x}_n/\sigma_{\tilde{W}}$ we can rewrite (4.12) as

$$\breve{\mathbf{x}}_n = \sqrt{\frac{\breve{\rho}}{M}}\,(\mathbf{I}_P \otimes V_{z_n})\,\breve{\mathbf{h}}_n + \breve{\mathbf{w}}_n \tag{4.16}$$

where $\breve{\rho} = \frac{\rho\sigma^2_{\hat{H}}}{\sigma^2_{\tilde{W}}}$ is the "equivalent SNR". Note that $\breve{\mathbf{h}}_n, \breve{\mathbf{w}}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{MP})$ and $E[\breve{\mathbf{h}}_n \breve{\mathbf{w}}_n^*] = \mathbf{0}$.

Given that the symbol V_1 was sent, the receiver will detect V_2, and thus make a decision error, if

$$\begin{aligned} \|\breve{\mathbf{x}}_n - \sqrt{\frac{\breve{\rho}}{M}}\,(\mathbf{I}_P \otimes V_2)\,\breve{\mathbf{h}}_n\|^2 &< \|\breve{\mathbf{x}}_n - \sqrt{\frac{\breve{\rho}}{M}}\,(\mathbf{I}_P \otimes V_1)\,\breve{\mathbf{h}}_n\|^2 \\ \iff Q = [\mathbf{y}_1^*\ \mathbf{y}_2^*] \underbrace{\begin{bmatrix} \mathbf{I}_{MP} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_{MP} \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}}_{\mathbf{y}} &< 0 \end{aligned} \tag{4.17}$$

where $\mathbf{y}_1 = \sqrt{\frac{\breve{\rho}}{M}}\,(\mathbf{I}_P \otimes (V_1 - V_2))\,\breve{\mathbf{h}}_n + \breve{\mathbf{w}}_n$ and $\mathbf{y}_2 = \breve{\mathbf{w}}_n$, and the PWEP is given by [44]

$$\Pr(V_1 \to V_2) = \Pr(Q \le 0) = \sum_{\substack{\text{poles } \omega = jp \\ p>0}} \text{Res}\left[-\frac{\Phi_Q(\omega)}{\omega}\right] \tag{4.18}$$

where the summation is taken over the poles in the upper half plane (UHP) and $\Phi_Q(\omega) = E[e^{j\omega Q}]$. The characteristic function of Q, a Hermitian quadratic of Gaussian vector, is given by [38]

$$\Phi_Q(\omega) = \frac{1}{\det(\mathbf{I}_{2MP} - j\omega R_{\mathbf{y}} K)} \tag{4.19}$$

where $R_{\mathbf{y}} = E[\mathbf{y}\mathbf{y}^*]$, given as

$$R_{\mathbf{y}} = \begin{bmatrix} \frac{\breve{\rho}}{M}\left(\mathbf{I}_P \otimes (V_1 - V_2)(V_1 - V_2)^*\right) + \mathbf{I}_{MP} & \mathbf{I}_{MP} \\ \mathbf{I}_{MP} & \mathbf{I}_{MP} \end{bmatrix} \tag{4.20}$$

Using $\det\left[\begin{smallmatrix} A & C \\ B & D \end{smallmatrix}\right] = \det(A - BD^{-1}C)\det(D)$, it can be shown that (for proof, see Appendix A.6)

$$\begin{aligned} \det(\mathbf{I}_{2MP} - j\omega R_{\mathbf{y}} K) &= \prod_{k=1}^{M} (1 - j\omega \frac{\breve{\rho}}{M}\sigma_k^2 + \omega^2 \frac{\breve{\rho}}{M}\sigma_k^2)^P \\ &= \prod_{k=1}^{M} \left(\frac{\breve{\rho}}{M}\sigma_k^2\right)^P (\omega - jp_k^+)^P (\omega - jp_k^-)^P \end{aligned} \tag{4.21}$$

where σ_k is the k^{th} singular value of $V_1 - V_2$ and

$$p_k^+ = \frac{1}{2}\left(1 + \sqrt{1 + \frac{4M}{\breve{\rho}\sigma_k^2}}\right) > 0 \tag{4.22}$$

$$p_k^- = \frac{1}{2}\left(1 - \sqrt{1 + \frac{4M}{\breve{\rho}\sigma_k^2}}\right) < 0 \tag{4.23}$$

Note that the characteristic function $\Phi_Q(\omega)$, and hence the PWEP, depend on the signal only through the singular values of $V_1 - V_2$. Since the singular values of $V_1 - V_2$ and $\mathbf{I}_M - V_2 V_1^*$ are the same, $\Pr(V_1 \to V_2) = \Pr(\mathbf{I}_M \to V_2 V_1^*)$.

Equation (4.21) implies that the poles of $\Phi_Q(\omega)$ have multiplicities equal to multiples of P. Assume that $\Phi_Q(\omega)$ has $n_Q \leq M$ distinct poles $\{\tilde{p}_k^+\}_{k=1}^{n_Q}$ in the UHP, i.e., there are n_Q distinct elements in the set $\{p_k^+\}_{k=1}^{M}$. Denoting the multiplicity of $\tilde{p}_k^+$ as Pm_k such that $\sum_{k=1}^{n_Q} m_k = M$, $\Phi_Q(\omega)$ can be expressed as

$$\Phi_Q(\omega) = \left(\prod_{k=1}^{M} \frac{M}{\breve{\rho}\sigma_k^2(\omega - jp_k^-)}\right)^P \prod_{k=1}^{n_Q} \frac{1}{(\omega - j\tilde{p}_k^+)^{Pm_k}} \tag{4.24}$$

Computation of the PWEP using (4.18) involves taking residues at $\{\tilde{p}_k^+\}_{k=1}^{n_Q}$, which can be complicated since the multiplicities of these poles are greater than 1. A simple method to evaluate the PWEP in such cases has been proposed in [28], where the

poles are perturbed by small amount to eliminate multiplicity, and the PWEP is computed by taking residues at *all* the simple poles in UHP. This method produces an lower bound on the PWEP if all the concerned poles are moved away from the origin, and upper bound when moved towards the origin. In this chapter, the i^{th} occurrence of $\tilde{p}_k^+$ is replaced by $\tilde{\tilde{p}}_{(k-1)Pm_k+i}^+ = \tilde{p}_k^+ + (i-1)\epsilon_k$, yielding the set of simple poles $\{\tilde{\tilde{p}}_k\}_{k=1}^{MP}$, and hence, the PWEP from (4.18) (for proof, see Appendix A.7)

$$\Pr(V_1 \to V_2) \gtrless \sum_{k=1}^{MP} \frac{1}{\tilde{\tilde{p}}_k^+} \left(\prod_{\ell=1}^{M} \frac{M}{\breve{\rho}\sigma_\ell^2 (\tilde{\tilde{p}}_k^+ - p_k^-)} \right)^P \prod_{\ell=1, \ell \neq k}^{MP} \frac{1}{(\tilde{\tilde{p}}_\ell^+ - \tilde{\tilde{p}}_k^+)} \tag{4.25}$$

where an upper bound is obtained by choosing $\epsilon_k = -0.0025\tilde{p}_k^+$, and a lower bound by choosing $\epsilon_k = 0.0025\tilde{p}_k^+$. Numerical results in Section 4.3 confirm that these bounds are very close to each other, and thus this method produces an accurate estimate of the PWEP.

4.2.2 Chernoff Bound on the PWEP: perfect knowledge of ζ_k and ρ

To obtain a better intuitive insight into the performance of DFDD, we now analyze the Chernoff upper bound on PWEP. Since (4.16) describes a known-channel system with equivalent SNR $\breve{\rho}$, the Chernoff bound on the PWEP is given by [9]

$$\Pr(V_1 \to V_2) \leq \frac{1}{2} \prod_{k=1}^{M} \left[1 + \frac{\breve{\rho}}{4M} \sigma_k^2 \right]^{-P} \tag{4.26}$$

Observe that the diversity advantage of the system is MP, whereas the performance is governed by the equivalent SNR $\breve{\rho}$. Recall that $\breve{\rho}$ is a function of ρ, the SNR when the channel is known perfectly (henceforth termed "coherent SNR"), the length m of DFDD, and the fading correlations ζ_k.

For numerical examples in this paper, we consider a system with $M = 2$ transmit and $P = 2$ receive antennas. The MIMO channel exhibits Rayleigh fading [41] where

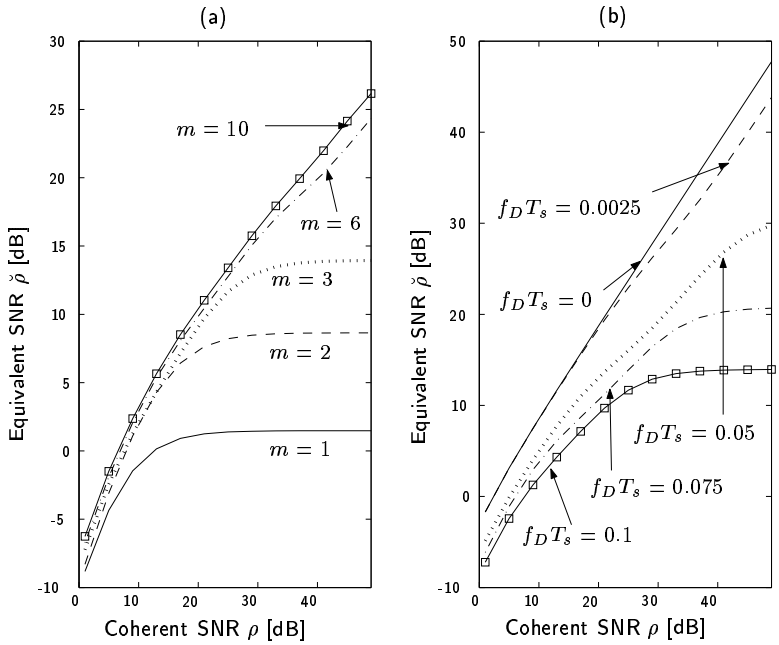

Figure 4.1: Equivalent SNR $\breve{\rho}$ vs coherent SNR ρ for (a) $f_D T_s = 0.1$, (b) $m = 3$.

the correlation between fading coefficients k matrix-symbols apart is given by $\zeta_k = J_0(2\pi f_D T_s M k)$ and where $f_D T_s$ is the normalized Doppler frequency. Observe that $f_D T_s M$ is the effective normalized Doppler frequency for $M \times M$ symbols. As we will see, the performance of the detectors degrades with increasing Doppler frequency, therefore, increasing M in a fading channel may degrade the performance.

Figure 4.1 demonstrates the variation of the equivalent SNR $\breve{\rho}$ w.r.t. the coherent SNR ρ for different $f_D T_s$ and DFDD feedback lengths m. Observe that when either $f_D T_s$ is large or m is small, $\breve{\rho}$ reaches a ceiling, implying that an error floor will appear in the BER vs. SNR curve. Fig. 4.1(a) indicates that increasing m has less pronounced effect on the performance when ρ in small, whereas the performance improvement can be significant when ρ is high. Equation (4.26) and Fig. 4.1(b) imply

that the slope of the BER vs coherent SNR ρ curve would decrease with increasing $f_D T_s$, which is in accordance with the result for DPSK reported in [22].

Figure 4.2 examines variations in $\breve{\rho}$ versus m and $f_D T_s$ for $\rho = 15$dB. Fig. 4.2 indicates that increasing m beyond, say, $m = 10$ has an insignificant effect on performance. Note also that when $f_D T_s = 0$, coherent detection performance is achieved asymptotically by DFDD. Performance loss due to increased channel variation is depicted in Fig. 4.2(b).

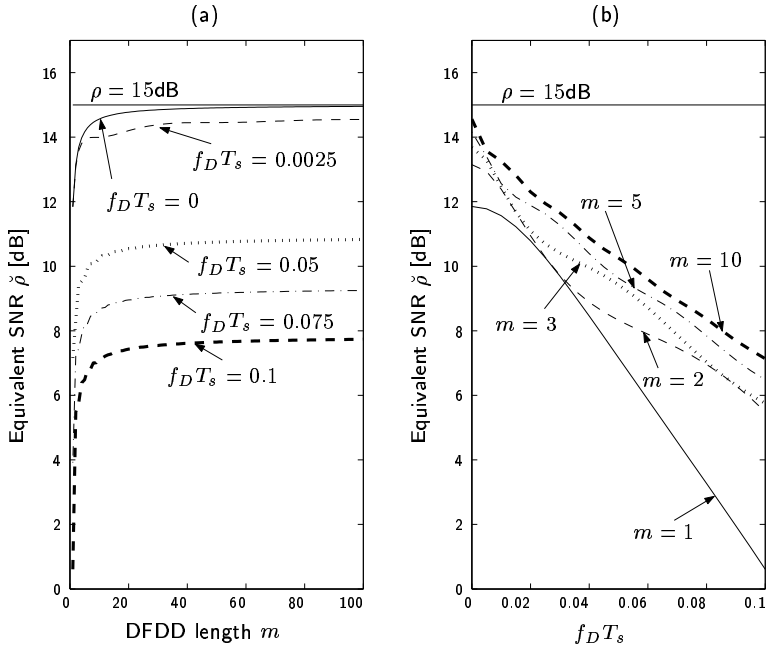

Figure 4.2: Equivalent SNR $\breve{\rho}$ vs (a) DFDD length m, (b) $f_D T_s$ for $\rho = 15$dB

4.2.3 Exact PWEP: imperfect knowledge of ζ_k and ρ

We have seen that the m-DFDD derived in this chapter requires knowledge of the "coherent SNR" ρ and the fading correlations $\{\zeta_k\}_{k=0}^{m}$. In practice, perfect knowledge

of these parameters is unlikely, especially in scenarios where channel statistics change with time. In this section, therefore, we derive exact PWEP when the receiver has imperfect knowledge of fading correlation and ρ.

Relative to the "coherent SNR" ρ, we introduce the *assumed* SNR ρ_a. Similarly, we construct the assumed matrices $\Xi_a^{(m)}, T_a^{(m)}, \mathbf{g}_a$ and B_a corresponding to $\Xi^{(m)}, T^{(m)}, \mathbf{g}$ and B defined earlier. The linear estimator is $B_a = \sqrt{\rho_a/M}\mathcal{S}_{n-1}\left((T_a^{(m-1)^{-1}}\mathbf{g}_a) \otimes \mathbf{I}_{MP}\right)$ and the channel estimate is $\hat{\tilde{\mathbf{h}}}_n = B_a^* \underline{\mathbf{x}}_{n-1}$. Using $\tilde{\tilde{\mathbf{h}}}_n = \mathbf{h}_n - \hat{\tilde{\mathbf{h}}}_n$, we can rewrite (4.12) as

$$\mathbf{x}_n = \sqrt{\frac{\rho}{M}}(\mathbf{I}_P \otimes V_{z_n})(\mathbf{I}_P \otimes S_{n-1})\hat{\tilde{\mathbf{h}}}_n + \underbrace{\sqrt{\frac{\rho}{M}}(\mathbf{I}_P \otimes S_n)\tilde{\tilde{\mathbf{h}}}_n + \mathbf{w}_n}_{\tilde{\tilde{\mathbf{w}}}_n} \tag{4.27}$$

It can be shown that (for details, see Appendix A.8)

$$E[\hat{\tilde{\mathbf{h}}}_n \hat{\tilde{\mathbf{h}}}_n^*] = \underbrace{\frac{\rho_a}{M}\left(\mathbf{g}_a^* T_a^{(m-1)^{-1}} T^{(m-1)} T_a^{(m-1)^{-1}} \mathbf{g}_a\right)}_{\hat{\tilde{\sigma}}_h^2} \mathbf{I}_{MP} \tag{4.28}$$

$$E[\tilde{\tilde{\mathbf{w}}}_n \tilde{\tilde{\mathbf{w}}}_n^*] = \underbrace{\left((\frac{\rho}{M}+1) - \frac{2\sqrt{\rho^3\rho_a}}{M^2}\Re(\mathbf{g}^* T_a^{(m-1)^{-1}}\mathbf{g}_a) + \frac{\rho}{M}\hat{\tilde{\sigma}}_h^2\right)}_{\tilde{\tilde{\sigma}}_w^2} \mathbf{I}_{MP} \tag{4.29}$$

and

$$E[\tilde{\tilde{\mathbf{w}}}_n \hat{\tilde{\mathbf{h}}}_n^*(\mathbf{I}_P \otimes S_{n-1}^*)] = \underbrace{\sqrt{\frac{\rho}{M}}\left(\frac{\sqrt{\rho\rho_a}}{M}(\mathbf{g}^* T_a^{(m-1)^{-1}}\mathbf{g}_a) - \hat{\tilde{\sigma}}_h^2\right)}_{\sigma_{wh}^2}(\mathbf{I}_P \otimes V_{z_n}) \tag{4.30}$$

Now defining $\breve{\breve{\mathbf{x}}}_n = \mathbf{x}_n/\tilde{\tilde{\sigma}}_w$, $\breve{\breve{\mathbf{h}}}_n = (\mathbf{I}_P \otimes S_{n-1})\,\hat{\tilde{\mathbf{h}}}_n/\hat{\tilde{\sigma}}_h$, $\breve{\breve{\mathbf{w}}}_n = \tilde{\tilde{\mathbf{w}}}_n/\tilde{\tilde{\sigma}}_w$, and $\breve{\breve{\rho}} = \rho\hat{\tilde{\sigma}}_h^2/\tilde{\tilde{\sigma}}_w^2$, we can write (4.27) as

$$\breve{\breve{\mathbf{x}}}_n = \sqrt{\frac{\breve{\breve{\rho}}}{M}}(\mathbf{I}_P \otimes V_{z_n})\breve{\breve{\mathbf{h}}}_n + \breve{\breve{\mathbf{w}}}_n \tag{4.31}$$

such that $E[\breve{\breve{\mathbf{h}}}_n \breve{\breve{\mathbf{h}}}_n^*] = E[\breve{\breve{\mathbf{w}}}_n \breve{\breve{\mathbf{w}}}_n^*] = \mathbf{I}_{MP}$ and $E[\breve{\breve{\mathbf{w}}}_n \breve{\breve{\mathbf{h}}}_n^*] = \frac{\sigma_{wh}^2}{\hat{\breve{\sigma}}_h \tilde{\breve{\sigma}}_w} \mathbf{I}_P \otimes V_{z_n}$. The PWEP $\Pr(V_1 \to V_2)$ in this case can be computed using the method described in Section 4.2.1 and equations (4.21)-(4.25), but with the replacements (for proof see Appendix A.9)

$$\det(\mathbf{I}_{2MP} - j\omega R_{\mathbf{y}} K) = \prod_{k=1}^{M} \left[\frac{\breve{\breve{\rho}} \sigma_k^2}{M} (1-\tau^2)(\omega - j p_k^+)(\omega - j p_k^-) \right]^P \tag{4.32}$$

$$p_k^{\pm} = \frac{1}{2} \left(1-\tau^2\right)^{-1} \left[1 + \tau \sqrt{M/\breve{\breve{\rho}}} \pm \sqrt{\left(1 + \tau \sqrt{M/\breve{\breve{\rho}}}\right)^2 + \frac{4M}{\breve{\breve{\rho}} \sigma_k^2} (1-\tau^2)} \right] \tag{4.33}$$

where σ_k is the k^{th} singular value of $V_1 - V_2$ and $\tau = \frac{\sigma_{wh}^2}{\hat{\breve{\sigma}}_h \tilde{\breve{\sigma}}_w}$. Again, the lower and upper bound on the PWEP are close to each other, producing an accurate approximation of the PWEP.

4.2.4 Chernoff Bound on the PWEP: imperfect knowledge of ζ_k and ρ

In order to further analyze the performance loss due imperfect parameter knowledge, we present the Chernoff bound on the PWEP.

Theorem 1 *The Chernoff upper bound on* $\Pr(V_1 \to V_2)$ *is given by*

$$\Pr(V_1 \to V_2) \leq \frac{1}{2} \prod_{k=1}^{M} \left[1 + \frac{\sigma_k^2}{4(1-\tau^2)} \left(\sqrt{\frac{\breve{\breve{\rho}}}{M}} + \tau \right)^2 \right]^{-P} \tag{4.34}$$

where $\tau = \frac{\sigma_{wh}^2}{\hat{\breve{\sigma}}_h \tilde{\breve{\sigma}}_w}$.

Proof: To prove this theorem, we take an approach similar to the approach in [9]. Defining $a_k = (\frac{\breve{\breve{\rho}}}{M} + \tau \sqrt{\frac{\breve{\breve{\rho}}}{M}}) \sigma_k^2$ and $b_k = \frac{\breve{\breve{\rho}} \sigma_k^2}{M} (1-\tau^2) > 0$, (4.33) can be re-written as

$$p_k^{\pm} = \frac{1}{2b_k} \left[a_k \pm \sqrt{a_k^2 + 4b_k} \right] \tag{4.35}$$

From (4.32) & (4.35) it is observed that the region of convergence of $\Phi_Q(\omega) = E[e^{j\omega Q}]$ is $\max_k(\frac{a_k}{2b_k} - \frac{1}{2b_k}\sqrt{a_k^2 + 4b_k}) < \mathrm{Im}(\omega) < \min_k(\frac{a_k}{2b_k} + \frac{1}{2b_k}\sqrt{a_k^2 + 4b_k})$. Noting that $c = \frac{a_k}{2b_k}$ is not a function of k, We evaluate the PWEP as

$$\begin{aligned}
\Pr(Q < 0) &= \int_{-\infty}^{0} f_Q(q)\, dq \\
&= \frac{1}{2\pi}\int_{-\infty}^{0}\int_{-\infty+jc}^{\infty+jc} e^{-j\omega q}\Phi_Q(\omega)\, d\omega\, dq \\
&= \frac{-1}{2\pi j}\int_{-\infty+jc}^{\infty+jc}\frac{1}{\omega}\Phi_Q(\omega)\, d\omega \\
&= \frac{-1}{2\pi j}\int_{-\infty}^{\infty}\frac{1}{\omega + jc}\Phi_Q(\omega + jc)\, d\omega \qquad (4.36)
\end{aligned}$$

where the characteristic function has been inverted by choosing the integration contour $\mathrm{Im}(\omega) = c$, which lies within the region of convergence. From (4.32) and (4.35) we get

$$\begin{aligned}
\Phi_Q(\omega + jc) &= \prod_{k=1}^{M}\left[b_k(\omega + jc - jp_k^+)(\omega + jc - p_k^-)\right]^{-P} \\
&= \prod_{k=1}^{M}\left[b_k(\omega - \frac{j}{2b_k}\sqrt{a_k^2 + 4b_k})(\omega + \frac{j}{2b_k}\sqrt{a_k^2 + 4b_k})\right]^{-P} \\
&= \prod_{k=1}^{M}\left[1 + b_k(\omega^2 + \frac{a_k^2}{4b_k^2})\right]^{-P} \qquad (4.37)
\end{aligned}$$

Since both $\Phi_Q(\omega + jc)$ and $\Pr(Q < 0)$ are real, extracting the real part in (4.36) we get

$$\begin{aligned}
\Pr(Q < 0) &= \frac{c}{2\pi}\int_{-\infty}^{\infty}\frac{1}{\omega^2 + c^2}\prod_{k=1}^{M}\left[1 + b_k(\omega^2 + \frac{a_k^2}{4b_k^2})\right]^{-P} d\omega \\
&\leq \prod_{k=1}^{M}\left[1 + \frac{a_k^2}{4b_k}\right]^{-P}\frac{c}{2\pi}\int_{-\infty}^{\infty}\frac{1}{\omega^2 + c^2}\, d\omega \\
&= \frac{1}{2}\prod_{k=1}^{M}\left[1 + \frac{a_k^2}{4b_k}\right]^{-P} \qquad (4.38)
\end{aligned}$$

By replacing $a_k = (\frac{\breve{\rho}}{M} + \tau\sqrt{\frac{\breve{\rho}}{M}})\sigma_k^2$ and $b_k = \frac{\breve{\rho}\sigma_k^2}{M}(1 - \tau^2)$ in (4.38) we obtain (4.34). ■

Under perfect parameter knowledge, i.e., $\rho_a = \rho$, $\Xi_a^{(m)} = \Xi^{(m)}$, $T_a^{(m)} = T^{(m)}$, and $\mathbf{g}_a = \mathbf{g}$, it is easy to verify that $\hat{\tilde{\sigma}}_h = \sigma_{\hat{H}}$, $\tilde{\tilde{\sigma}}_w = \sigma_{\tilde{W}}$ and $\sigma_{wh} = 0$, which implies $\tau = 0$, making (4.34) identical to (4.26). Observe that the performance in the general case is governed by $\frac{1}{1-\tau^2}(\sqrt{\frac{\breve{\breve{\rho}}}{M}} + \tau)^2$, whereas performance in the perfect knowledge case is governed by $\breve{\rho}/M$. To analyze the performance degradation due to parameter mismatch, we define the "equivalent SNR loss" α as

$$\alpha = \frac{(\sqrt{\breve{\breve{\rho}}/M} + \tau)^2}{1 - \tau^2} \frac{M}{\breve{\rho}} \tag{4.39}$$

Figure 4.3 shows the variations in α with respect to the *actual* Doppler frequency $f_D T_s$, when the actual "coherent SNR" is $\rho = 20$dB, the assumed Doppler frequency is $f_D T_s|_a = \beta f_D T_s$, and the assumed "coherent SNR" is $\rho_a = 10, 30$dB. Again, we assumed $M = 2$ transmit antennas and $\zeta_k = J_0(2\pi f_D T_s M k)$. From Fig. 4.3(a) we find that, when $\rho_a = 30\text{dB} > \rho$, $\beta < 1$, i.e., under-estimating the Doppler frequency results in lower SNR loss compared to the case of $\beta > 1$. Figure 4.3(b), where $\rho_a = 10\text{dB} < \rho$, shows the opposite trend. As predicted in Section 4.1.3 and shown in Fig. 4.3, when $\beta < 1$, performance can be improved by choosing $\rho_a > \rho$ and vice-versa, whereas the performance loss can be significant when $\rho_a < \rho, \beta < 1$, or when $\rho_a > \rho, \beta > 1$.

4.2.5 Approximate Bit Error Rate

In practice, bit error probability (BER) is a more useful metric than PWEP. We use the Chernoff bound on PWEP and the exact PWEP derived previously together with the union bound to compute approximate BER. Recalling that ηM bits are

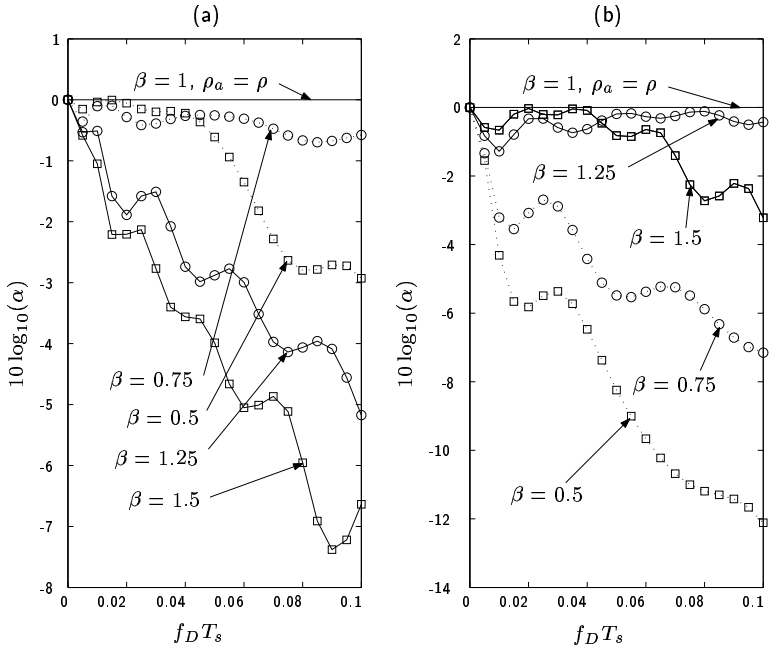

Figure 4.3: SNR loss α for $f_D T_s|_a = \beta f_D T_s$, "coherent SNR" $\rho = 20$dB and $\rho_a =$ (a) 30dB, (b) 10dB

encoded in each transmitted matrix-symbol, the BER can be written as

$$P_{\text{genie}}^{\text{Chernoff}} \approx \frac{1}{\eta M} \sum_{j=0}^{2^{\eta M}-1} \left(\sum_{k=0, k\neq j}^{2^{\eta M}-1} d(V_j, V_k) \Pr^{\text{Chernoff}}(V_j \to V_k) \right) \pi_j \tag{4.40}$$

$$P_{\text{genie}}^{(L)/(U)} \approx \frac{1}{\eta M} \sum_{j=0}^{2^{\eta M}-1} \left(\sum_{k=0, k\neq j}^{2^{\eta M}-1} d(V_j, V_k) \Pr^{(L)/(U)}(V_j \to V_k) \right) \pi_j \tag{4.41}$$

where $d(V_j, V_k)$ is the Hamming distance between the binary representations of V_j and V_k, π_j is the probability that the information symbol V_j was sent, the superscript "Chernoff" indicates that Chernoff bound on the PWEP has been used, and $(L)/(U)$ indicate that the lower and upper bound on the PWEP produced by (4.25) has been used. We will see in Section 4.3 that $P_{\text{genie}}^{(L)/(U)}$ are accurate approximations of the BER of genie-aided DFDD, whereas $P_{\text{genie}}^{\text{Chernoff}}$ is a very loose approximation. Recall

that $\Pr(V_1 \to V_2) = \Pr(\mathbf{I}_M \to V_2 V_1^*)$, that $V_1, V_2 \in \mathcal{A} \implies V_2 V_1^* \in \mathcal{A}$ since $\mathcal{A}$ is a group [11] and hence satisfies closure and inverse properties, and that $\mathcal{A}$ becomes symmetric in terms of Hamming distance if Gray mapping is used. Therefore, the right hand sides of (4.40) & (4.41) simplify to $\frac{1}{\eta^M} \sum_{j=1}^{2^{\eta M}-1} d(\mathbf{I}_M, V_j) \Pr(\mathbf{I}_M \to V_j)$ under the assumption of Gray mapping and equal prior probabilities.

For realizable m-DFDD, the influence of incorrect past-decisions has to be taken into account for $m > 1$. In [26] it has been shown that the BER of realizable DFDD is approximately twice that of genie-aided DFDD for DPSK, since every error is likely to cause another error due to error propagation. Through numerical evaluation we find that this approximation extends to DUST as well, which is in accordance with [29].

4.3 Simulations & Numerical Results

As in Chapter 3, in this chapter also we evaluate the performance of the detectors with two channel types: "block fading" (2.3) and "continuous fading" (2.2). Recall that in continuous fading the correlation between coefficients k symbols apart is given by $J_0(2\pi f_D T_s k)$ [41], while, in block-fading, the correlation between channel coefficients m matrix-symbols apart is given by $J_0(2\pi f_D T_s M m)$.

As shown in Chapter 2, the use of diagonal constellations in continuous fading yields the same model as general constellations in block fading, and hence detector performance is identical in these two cases. Therefore, we focus on detector performance using diagonal and non-diagonal constellations in continuous fading. We consider a system with two receive and two transmit antennas and use the constellations specified in Table 2.1.

4.3.1 Bit Error Rate: Perfect Parameter Knowledge Case

In this section we present the simulation as well analytical bit error rate performance of DFDD in the perfect parameter knowledge case. DFDD performance with non-diagonal constellations has also been evaluated and compared with performance in the case of diagonal constellations.

The simulated BER of genie-aided as well as realizable 1-, 2-, and 6-DFDD in continuous fading with $f_D T_s = 0.1$ and 0.05 is shown in Fig. 4.4, where the advantage of m-DFDD, $m > 1$ over standard single symbol detection is clearly illustrated. When $f_D T_s = 0.1$, 1-DFDD, which is equivalent to standard single symbol detection, succumbs to very high error floor, whereas the performance is dramatically improved when 6-DFDD is employed, as predicted by Fig. 4.1(a). Meanwhile, 2-DFDD performs much better than 1-DFDD while still succumbing to an error floor. Similar, though less pronounced, trends can be seen in Fig. 4.4(b), where $f_D T_s = 0.05$. Figure 4.4 also demonstrates the effect of error propagation via comparison of genie-aided and realizable DFDD.

Next, the performance of the detectors has been evaluated in continuous fading with a non-diagonal constellation to illustrate the performance loss due to approximation of system model (2.2) by (2.3). Recall that the non-diagonal constellation of Table 2.1 is generated by right multiplying the diagonal constellation by a fixed non-diagonal unitary matrix. Because such an operation does not change the product distance of the constellation [11], the comparison is fair.

Figure 4.5 illustrates the performance of 1-,2-, and 6-DFDD using the non-diagonal constellation in continuous fading with $f_D T_s = 0.075$ and $f_D T_s = 0.05$. Although a performance loss is incurred due to neglecting the channel variation within the

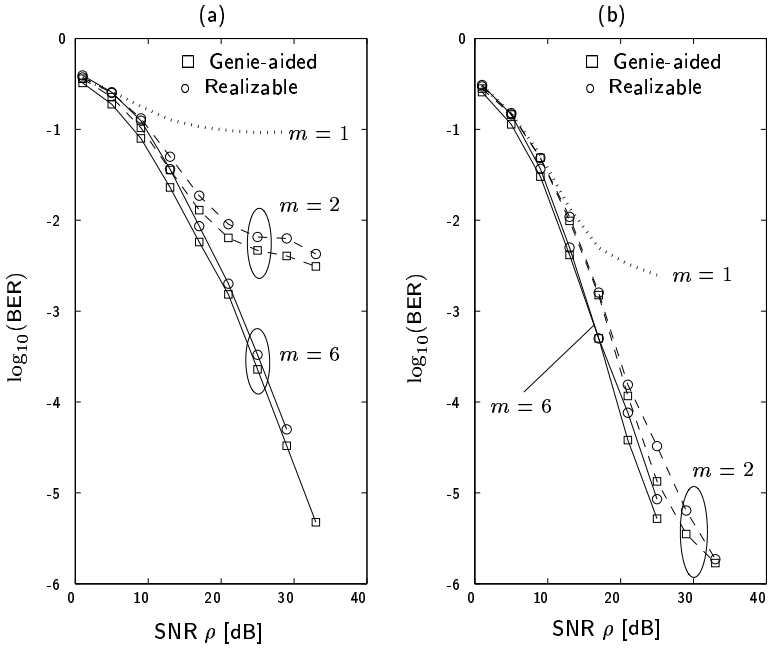

Figure 4.4: m-DFDD detection of the diagonal constellation in continuous fading with (a) $f_D T_s = 0.1$, (b) $f_D T_s = 0.05$

matrix-symbol interval, in both cases 2- and 6-DFDD perform much better than single symbol detection. Unlike the diagonal constellation case, Fig. 4.5 shows that 6-DFDD provides significant performance gain over 2-DFDD when $f_D T_s = 0.05$, whereas the gain is insignificant for higher $f_D T_s$.

Now we compare theoretical with simulated BER to verify the validity of the error analysis in Section 4.2. Since BER and PWEP expressions have been derived for genie-aided DFDD, and since the approximate BER of realizable DFDD can be obtained by multiplying the genie-aided BER by 2 [26,29], we present simulation and theoretical results for only the genie-aided case.

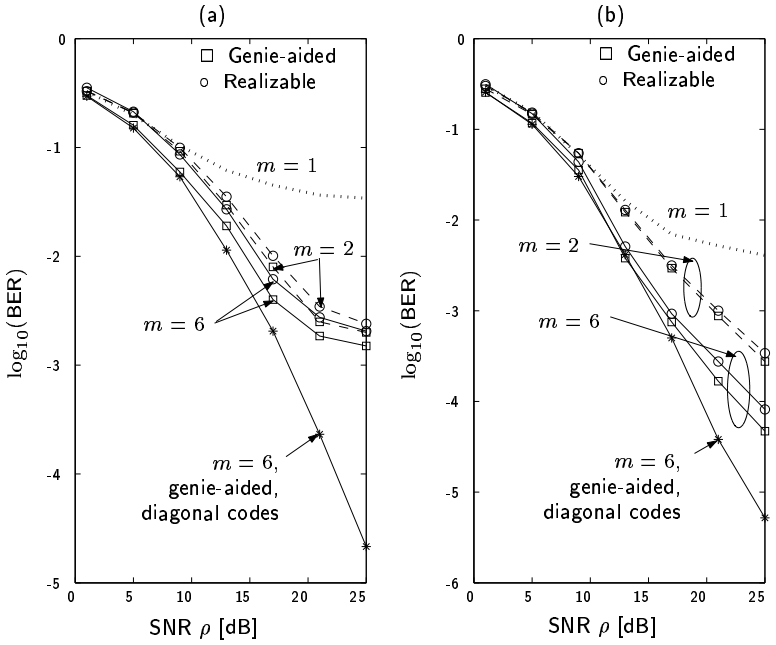

Figure 4.5: m-DFDD detection of the non-diagonal constellation in continuous fading with (a) $f_D T_s = 0.075$, (b) $f_D T_s = 0.05$

Figure 4.6 compares the theoretical and simulated BER of 2- and 6-DFDD using the diagonal constellation in continuous fading with $f_D T_s = 0.075$. Three approximations to the BER have been presented: an upper bound $P_{\text{genie}}^{\text{Chernoff}}$ from (4.40), and a lower bound $P_{\text{genie}}^{(L)}$ and an upper bound $P_{\text{genie}}^{(U)}$ from (4.41). Observe that $P_{\text{genie}}^{(L)}$ and $P_{\text{genie}}^{(U)}$ are somewhat loose at high BER since (4.41) employs the union bound. Since $P_{\text{genie}}^{(L)}$ and $P_{\text{genie}}^{(U)}$ are very close to each other, we shall only present $P_{\text{genie}}^{(U)}$ in the sequel.

4.3.2 Bit Error Rate: Imperfect Parameter Knowledge Case

In this section we present the bit error rate performance of DFDD under imperfect knowledge of fading correlation and SNR. Fig. 4.7 investigates the analytical as well as simulated performance of genie-aided 6-DFDD when the *actual* normalized Doppler

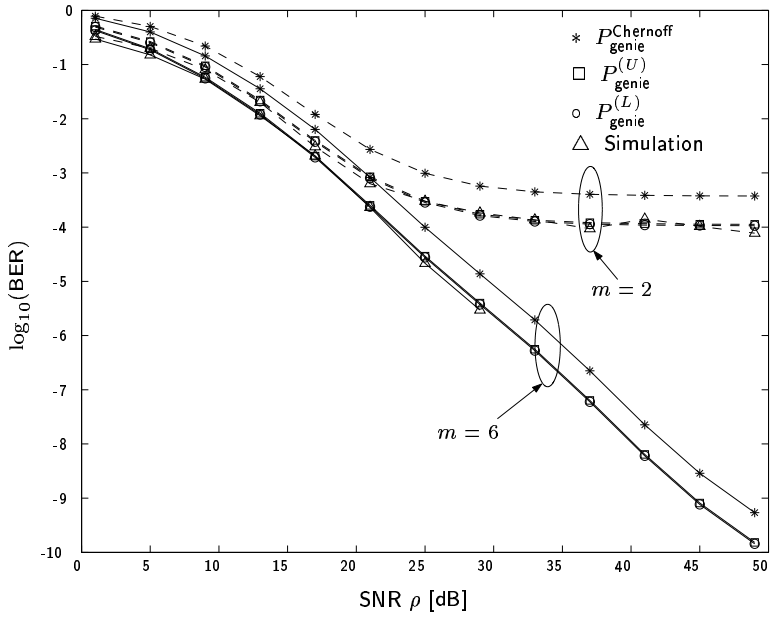

Figure 4.6: Theoretical and simulated BER of genie-aided m-DFDD: Diagonal constellation in continuous fading with $f_D T_s = 0.075$

frequency is 0.075 and the receiver has imperfect knowledge of $f_D T_s$ and SNR ρ. The performance assuming perfect knowledge of $f_D T_s$ and ρ has been compared to the performance of detectors with assumed $f_D T_s|_a = 0.075, 0.05$ and $\rho_a = 15, 30$dB. As expected, the parameter mismatch results in performance loss. Note that 6-DFDD succumbs to an error floor when it underestimates the Doppler frequency and a loss in SNR when it overestimates the Doppler frequency. For all but highest levels of BER, $P^{(U)}_{\text{genie}}$ closely matches the simulation results.

Now we analyze the relation between robustness and the DFDD length m. Figure 4.8 plots the theoretical BER $P^{(U)}_{\text{genie}}$ of genie-aided m-DFDD versus m when the *actual* normalized Doppler spread is $f_D T_s = 0.075$ and the SNR is $\rho = 20$dB. Observe that the performance loss due to under-estimation of the Doppler spread is

severe for large m, as predicted in Section 4.1.3. While over-estimation of Doppler or under/over-estimation of SNR also results in performance loss, it is less severe, and relatively constant over all values of $m > 2$. Finally, it is observed that m-DFDD is quite robust against parameter mismatch when $m = 2$.

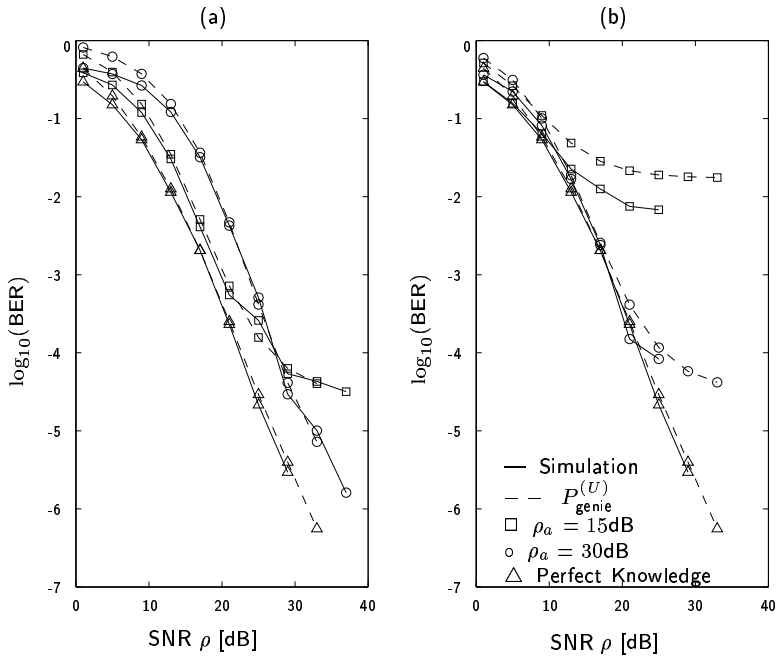

Figure 4.7: Genie-aided 6-DFDD with the diagonal constellation in continuous fading with *actual* $f_D T_s = 0.075$ and *assumed* $f_D T_s|_a =$ (a) 0.1, (b) 0.05

In this section, we have analyzed the performance of m-DFDD under various conditions and have shown its improved performance over standard single symbol detector. We have seen that the performance can be improved by increasing m when normalized Doppler and SNR are known at the receiver, while increased m increases

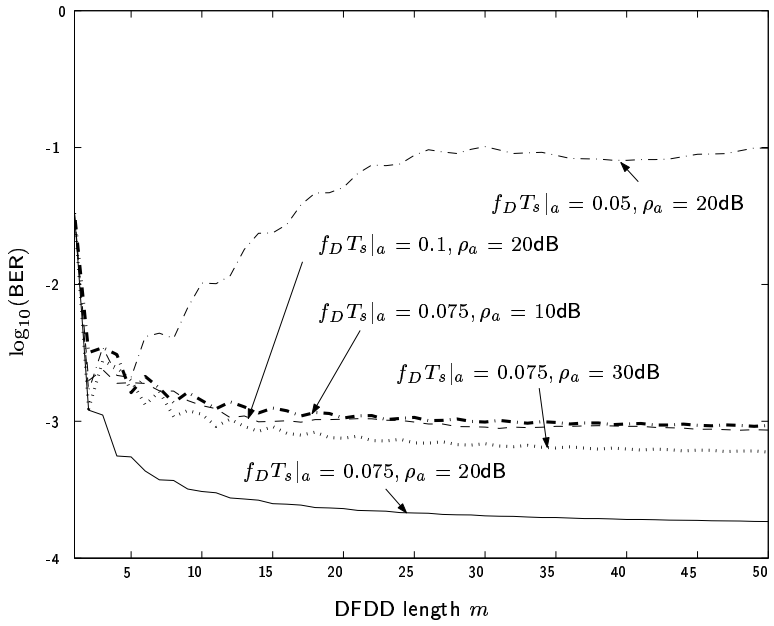

Figure 4.8: Genie-aided m-DFDD with $f_D T_s = 0.075$, $\rho = 20$dB

the vulnerability of m-DFDD towards under-estimation of Doppler frequency. Therefore, the choice of DFDD length m in practice should depend on the reliability of the Doppler frequency estimate, and of course, the complexity.

4.3.3 Comparison of Multiple-Symbol Detection and Decision-Feedback Detection

We have seen in Chapters 3 & 4 that both MSD and DFDD are capable of reducing the error floor in a fading channel. In this section, we compare the performance of DFDD with that of MSD and analyze the performance loss incurred due to replacing the hypothesized symbols with past decisions in the multiple-symbol ML detector.

Figure 4.9 compares the simulation BER performance of the realizable 3-DFDD, genie-aided 3-DFDD, non-overlapping and overlapping block detection ($p = 1$) using

3-symbol ML detector for $f_D T_s = 0.1, 0.075$ and the perfect parameter knowledge case. Observe that both realizable and genie-aided 3-DFDD exhibit similar performance as the non-overlapping block detection scheme. In fact, when $f_D T_s = 0.075$, performance of these detection methods are indistinguishable. Meanwhile, the overlapping block detection scheme demonstrates superior performance over other schemes.

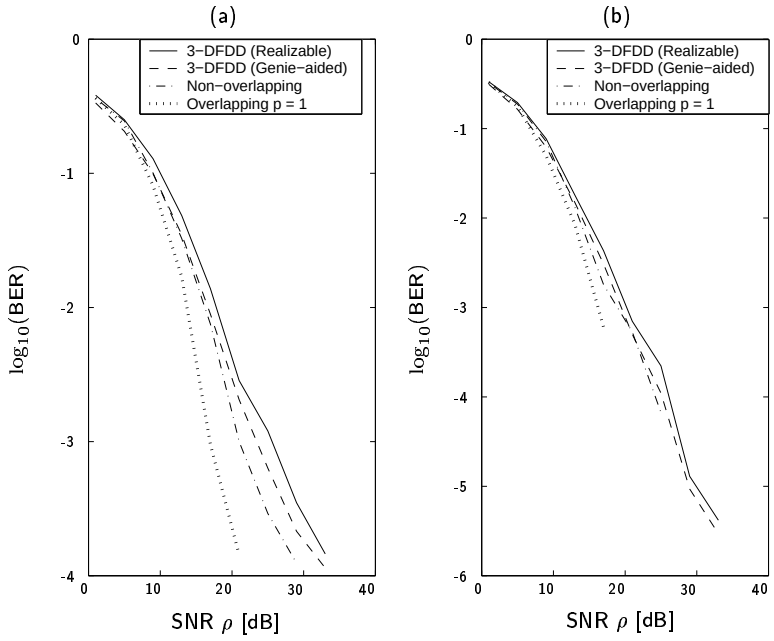

Figure 4.9: Comparison of realizable 3-DFDD, genie-aided 3-DFDD, non-overlapping and overlapping ($p = 1$) block detection based on 3-symbol ML detector for (a) $f_D T_s = 0.1$, (b) $f_D T_s = 0.075$

Figure 4.10 compares the simulation BER of MSD and DFDD under imperfect knowledge of fading correlation. Again, the performance of non-overlapping block detection scheme and 3-DFDD are identical while overlapping block-detection scheme exhibits superior performance.

While the complexity of DFDD is significantly less that the complexity of MSD, their performance are similar. In addition, the DFDD length can be increased to improve the performance without increasing the complexity significantly. Thus, DFDD deserves preference to MSD for practical implementation.

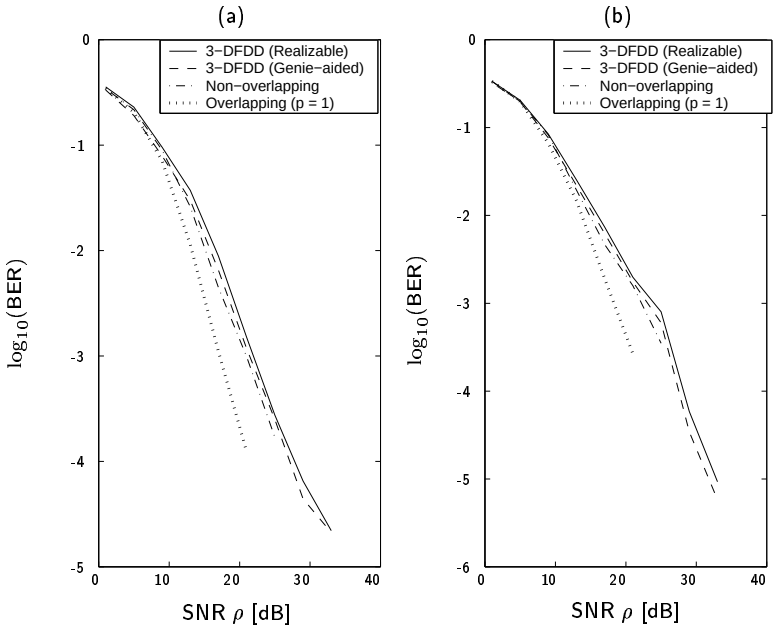

Figure 4.10: Comparison of realizable 3-DFDD, genie-aided 3-DFDD, non-overlapping and overlapping ($p = 1$) block detection based on 3-symbol ML detector for perfect SNR knowledge, *actual* $f_D T_s = 0.075$ and *assumed* (a) $f_D T_s = 0.1$, (b) $f_D T_s = 0.05$

4.4 Conclusions

In this chapter we have derived the decision-feedback differential detectors for DUST in Rayleigh fading channel. Our theoretical and simulation analysis shows that DFDD is capable of reducing the error floor in fast-fading channels. Although the DFDDs assume the channel to be block-fading when non-diagonal constellations

are used, they have been shown to improve the performance in continuous fading as well. While increasing the DFDD length improves the performance under the assumption of perfect parameter knowledge, the robustness of the DFDD to imperfect parameter knowledge has been shown to decrease with increasing the DFDD length. The performance of m-DFDD and non-overlapping block detection scheme based on m-symbol ML detector are shown to be almost identical.

CHAPTER 5

CONCLUSIONS & FUTURE WORK

In this thesis, we have presented novel detection schemes for DUST with significant performance improvement over the standard detector. The standard detector succumbs to an error floor in fading channels. The detectors presented in this thesis reduce and asymptotically eliminate the error floor.

In Chapter 3, we derive the multiple-symbol ML differential detectors and a reduced complexity multiple-symbol suboptimal detector that require the knowledge of channel fading correlation and SNR. Our theoretical and simulation analysis shows that the detectors are capable of reducing the error floor (encountered by the standard single symbol detector) in a fading channel. Simulations show that these detectors are reasonably robust against imperfect parameter knowledge.

Decision-feedback differential detection (DFDD) of DUST has been proposed in Chapter 4, wherein we have shown that the DFDD derived from the multiple-symbol ML detector comprises a Wiener-filter based channel predictor followed by coherent ML detector. Using this result, we interpret the multiple-symbol differential detectors as joint channel estimator-coherent detector. We derive theoretical BER expressions

for DFDD that are verified by the simulation results. Comparison between the performance of DFDD and MSD reveals that the increased complexity of MSD is disproportionate with the performance improvement over DFDD. Thus, DFDD should be preferred to MSD for practical implementation.

There are a number of avenues for further research.

Non-diagonal constellations in continuous-fading

This thesis assumes that the channel is block-fading when non-diagonal constellations are used, which is usually invalid in a fast-fading environment. Thus, derivation and performance evaluation of detectors for non-diagonal codes in continuous-fading is an important topic for further research.

Adaptive implementation of the detectors

Since the detectors derived in this thesis require the knowledge of channel fading correlation and SNR, they are required to be estimated and *tracked* in practice, especially in scenarios where channel statistics change with time. Leveraging the work on adaptive detectors for DPSK [45, 46] an adaptive implementation for the DFDD can be investigated.

Application to frequency-selective fading channels

We have considered flat-fading, i.e., narrowband channels for designing the detectors. Since orthogonal frequency division multiplexing (OFDM) transforms a frequency-selective channel into a decoupled set of flat-fading channels, our detectors can be combined with OFDM for use in a wideband system. Further research can conducted based on the existing results for a DPSK-OFDM system [47–51] and the work reported in this thesis.

APPENDIX A

DERIVATION DETAILS FOR Chapter 4

A.1 Proof of (4.5)

Since $\underline{\mathbf{h}}_{n-1}$ and $\underline{\mathbf{w}}_{n-1}$ contain zero-mean unit-variance i.i.d. Gaussian random variables, $E[\mathbf{h}_n \mathbf{h}_{n-k}^*] = \zeta_k \mathbf{I}_{MP}$, and $\underline{\mathbf{x}}_{n-1} = \sqrt{\frac{\rho}{M}} \mathcal{S}_{n-1} \underline{\mathbf{h}}_{n-1} + \underline{\mathbf{w}}_{n-1}$ (see (4.4)), we have

$$\begin{aligned}
E[\underline{\mathbf{x}}_{n-1} \underline{\mathbf{x}}_{n-1}^*] &= \frac{\rho}{M} \mathcal{S}_{n-1} E[\underline{\mathbf{h}}_{n-1} \underline{\mathbf{h}}_{n-1}^*] \mathcal{S}_{n-1}^* + \mathbf{I}_{mMP} \\
&= \frac{\rho}{M} \mathcal{S}_{n-1} \left(\Xi^{(m-1)} \otimes \mathbf{I}_{MP} \right) \mathcal{S}_{n-1}^* + \mathbf{I}_{mMP} \\
&= \mathcal{S}_{n-1} \left(\frac{\rho}{M} \Xi^{(m-1)} \otimes \mathbf{I}_{MP} + \mathbf{I}_{mMP} \right) \mathcal{S}_{n-1}^* \quad \text{Since } \mathcal{S}_{n-1} \text{ is unitary} \\
&= \mathcal{S}_{n-1} \left(T^{(m-1)} \otimes \mathbf{I}_{MP} \right) \mathcal{S}_{n-1}^*
\end{aligned}$$

since $T^{(m)} = \frac{\rho}{M} \Xi^{(m)} + \mathbf{I}_{m+1}$. Again, from (4.4) we have

$$\begin{aligned}
E[\underline{\mathbf{x}}_{n-1} \mathbf{h}_n^*] &= \sqrt{\rho/M} \mathcal{S}_{n-1} E[\underline{\mathbf{h}}_{n-1} \mathbf{h}_n^*] + E[\underline{\mathbf{w}}_{n-1} \mathbf{h}_n^*] \\
&= \sqrt{\rho/M} \mathcal{S}_{n-1} E[\underline{\mathbf{h}}_{n-1} \mathbf{h}_n^*] \\
&= \sqrt{\rho/M} \mathcal{S}_{n-1} (\mathbf{g} \otimes \mathbf{I}_{MP}) \quad \text{since} \quad E[\mathbf{h}_{n-k} \mathbf{h}_n^*] = \zeta_k^* \mathbf{I}_{MP}
\end{aligned}$$

Now,

$$\begin{aligned}
J(B) &= E[\|\mathbf{h}_n - \hat{\mathbf{h}}_n\|^2] \\
&= \mathrm{tr}\{E[(\mathbf{h}_n - B^* \underline{\mathbf{x}}_{n-1})(\mathbf{h}_n^* - \underline{\mathbf{x}}_{n-1}^* B)]\}
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{tr}\{E[\mathbf{h}_n\mathbf{h}_n^*] - B^*E[\underline{\mathbf{x}}_{n-1}\mathbf{h}_n^*] - E[\mathbf{h}_n\underline{\mathbf{x}}_{n-1}^*]B + B^*E[\underline{\mathbf{x}}_{n-1}\underline{\mathbf{x}}_{n-1}^*]B\} \\
&= \mathrm{tr}\{\mathbf{I}_{MP} - \sqrt{\rho/M}B^*\mathcal{S}_{n-1}(\mathbf{g}\otimes\mathbf{I}_{MP}) - \sqrt{\rho/M}(\mathbf{g}^*\otimes\mathbf{I}_{MP})\mathcal{S}_{n-1}^*B \\
&\quad + B^*\mathcal{S}_{n-1}\left(T^{(m-1)}\otimes\mathbf{I}_{MP}\right)\mathcal{S}_{n-1}^*B\} \qquad \text{(A.1)}
\end{aligned}
$$

Setting $\frac{\partial J(B)}{\partial B} = 0$, we get

$$
\begin{aligned}
0 &= 2\mathcal{S}_{n-1}\left(T^{(m-1)}\otimes\mathbf{I}_{MP}\right)\mathcal{S}_{n-1}^*B - 2\sqrt{\rho/M}\mathcal{S}_{n-1}(\mathbf{g}\otimes\mathbf{I}_{MP}) \\
\Longrightarrow\ B &= \sqrt{\rho/M}\mathcal{S}_{n-1}\left(T^{(m-1)^{-1}}\otimes\mathbf{I}_{MP}\right)\mathcal{S}_{n-1}^*\mathcal{S}_{n-1}(\mathbf{g}\otimes\mathbf{I}_{MP}) \\
&= \sqrt{\rho/M}\mathcal{S}_{n-1}\left((T^{(m-1)^{-1}}\mathbf{g})\otimes\mathbf{I}_{MP}\right)
\end{aligned}
$$

which is (4.5). Substituting it into (A.1),

$$
\begin{aligned}
J_{\min} &= \mathrm{tr}\{\mathbf{I}_{MP} - \rho/M((\mathbf{g}^*T^{(m-1)^{-1}}\mathbf{g})\otimes\mathbf{I}_{MP})\} \\
&= MP(1 - \frac{\rho}{M}\mathbf{g}^*T^{(m-1)^{-1}}\mathbf{g}) \\
&= MP(1-\sigma_{\hat{H}}^2)
\end{aligned}
$$

A.2 Proof of (4.9) & (4.10)

Recalling that $F^{(m-1)} = \sqrt{\rho/M}T^{(m-1)^{-1}}\mathbf{g}$ and $\xi = \sqrt{\frac{M}{\rho}} + \sqrt{\frac{\rho}{M}}(1-\sigma_{\hat{H}}^2) > 0$, we have the identities

$$
\begin{aligned}
\sqrt{\frac{M}{\rho}}(\frac{\rho}{M}+1)\mathbf{I}_{MP} - \frac{\rho}{M}((\mathbf{g}^*F^{(m-1)})\otimes\mathbf{I}_{MP}) &= \xi\mathbf{I}_{MP} \\
\sqrt{\frac{\rho}{M}}\mathbf{g}\otimes\mathbf{I}_{MP} - (T^{(m-1)}F^{(m-1)})\otimes\mathbf{I}_{MP} &= \mathbf{0}
\end{aligned}
$$

which imply (4.9). Now (4.9) implies

$$
\begin{aligned}
\begin{bmatrix} \sqrt{\frac{M}{\rho}}\mathbf{I}_{MP} \\ -F^{(m-1)}\otimes\mathbf{I}_{MP} \end{bmatrix} &= T^{(m)^{-1}}\otimes\mathbf{I}_{MP}\begin{bmatrix} \xi\mathbf{I}_{MP} \\ \mathbf{0} \end{bmatrix} \\
&= \left(-\begin{bmatrix} a_{0,0}^{(m)} & a_{0,1}^{(m)} & \dots & a_{0,m}^{(m)} \\ a_{0,1}^{(m)*} & a_{1,1}^{(m)} & \dots & a_{1,m}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{0,m}^{(m)*} & a_{1,m}^{(m)*} & \dots & a_{m,m}^{(m)} \end{bmatrix}\begin{bmatrix} \xi \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right)\otimes\mathbf{I}_{MP}
\end{aligned}
$$

Therefore,

$$\begin{aligned}\begin{bmatrix} \sqrt{\frac{M}{\rho}}\mathbf{I}_{MP} \\ -F^{(m-1)} \otimes \mathbf{I}_{MP} \end{bmatrix} &= -\xi \begin{bmatrix} a_{0,0}^{(m)} \\ a_{0,1}^{(m)*} \\ \vdots \\ a_{0,m}^{(m)*} \end{bmatrix} \otimes \mathbf{I}_{MP} \\ \Longrightarrow \; F^{(m-1)} &= \xi \begin{bmatrix} a_{0,1}^{(m)*} \\ \vdots \\ a_{0,m}^{(m)*} \end{bmatrix}\end{aligned}$$

which is another form of (4.10)

A.3 Details of Statistics of $\hat{\mathbf{h}}_n$, $\tilde{\mathbf{h}}_n$, and $\tilde{\mathbf{w}}_n$

Since $\hat{\mathbf{h}}_n = B^* \underline{\mathbf{x}}_{n-1}$,

$$\begin{aligned}E[\hat{\mathbf{h}}_n \hat{\mathbf{h}}_n^*] &= B^* E[\underline{\mathbf{x}}_{n-1} \underline{\mathbf{x}}_{n-1}^*] B \\ &= B^* \left[\frac{\rho}{M} \mathcal{S}_{n-1} (\Xi^{(m-1)} \otimes \mathbf{I}_{MP}) \mathcal{S}_{n-1}^* + \mathbf{I}_{mMP} \right] B \\ &= B^* \mathcal{S}_{n-1} (T^{(m-1)} \otimes \mathbf{I}_{MP}) \mathcal{S}_{n-1}^* B\end{aligned}$$

Substituting $B = \sqrt{\frac{\rho}{M}} \mathcal{S}_{n-1} \left((T^{(m-1)^{-1}} \mathbf{g}) \otimes \mathbf{I}_{MP} \right)$,

$$\begin{aligned}&E[\hat{\mathbf{h}}_n \hat{\mathbf{h}}_n^*] \\ &= \frac{\rho}{M} \left\{ \left((\mathbf{g}^* T^{(m-1)^{-1}}) \otimes \mathbf{I}_{MP} \right) \mathcal{S}_{n-1}^* \mathcal{S}_{n-1} (T^{(m-1)} \otimes \mathbf{I}_{MP}) \mathcal{S}_{n-1}^* \mathcal{S}_{n-1} \left((T^{(m-1)^{-1}} \mathbf{g}) \otimes \mathbf{I}_{MP} \right) \right\} \\ &= \frac{\rho}{M} \left\{ \left((\mathbf{g}^* T^{(m-1)^{-1}}) \otimes \mathbf{I}_{MP} \right) (T^{(m-1)} \otimes \mathbf{I}_{MP}) \left((T^{(m-1)^{-1}} \mathbf{g}) \otimes \mathbf{I}_{MP} \right) \right\} \\ &= \left(\frac{\rho}{M} \mathbf{g}^* T^{(m-1)^{-1}} \mathbf{g} \right) \otimes \mathbf{I}_{MP} \\ &= \sigma_{\hat{H}}^2 \otimes \mathbf{I}_{MP} \\ &= \sigma_{\hat{H}}^2 \mathbf{I}_{MP}\end{aligned}$$

Now,

$$\begin{aligned}E[\hat{\mathbf{h}}_n \tilde{\mathbf{h}}_n^*] &= E[\hat{\mathbf{h}}_n (\mathbf{h}_n - \hat{\mathbf{h}}_n)^*] \\ &= E[\hat{\mathbf{h}}_n \mathbf{h}_n^*] - \sigma_{\hat{H}}^2 \mathbf{I}_{MP}\end{aligned}$$

$$
\begin{aligned}
&= B^* E[\underline{\mathbf{x}}_{n-1}\mathbf{h}_n^*] - \sigma_{\hat{H}}^2 \mathbf{I}_{MP} \\
&= B^* E\left[\left(\sqrt{\rho/M}\mathcal{S}_{n-1}\underline{\mathbf{h}}_{n-1} + \underline{\mathbf{w}}_{n-1}\right)\mathbf{h}_n^*\right] - \sigma_{\hat{H}}^2 \mathbf{I}_{MP} \\
&= \sqrt{\frac{\rho}{M}} B^* \mathcal{S}_{n-1}(\mathbf{g} \otimes \mathbf{I}_{MP}) - \sigma_{\hat{H}}^2 \mathbf{I}_{MP} \\
&= \frac{\rho}{M}\left((\mathbf{g}^* T^{(m-1)^{-1}}) \otimes \mathbf{I}_{MP}\right)\mathcal{S}_{n-1}^* \mathcal{S}_{n-1}(\mathbf{g} \otimes \mathbf{I}_{MP}) - \sigma_{\hat{H}}^2 \mathbf{I}_{MP} \\
&= \left(\frac{\rho}{M}\mathbf{g}^* T^{(m-1)^{-1}}\mathbf{g}\right) \otimes \mathbf{I}_{MP} - \sigma_{\hat{H}}^2 \mathbf{I}_{MP} \\
&= \mathbf{0}
\end{aligned}
$$

Since $\tilde{\mathbf{w}}_n = \sqrt{\frac{\rho}{M}}(\mathbf{I}_P \otimes S_n)\tilde{\mathbf{h}}_n + \mathbf{w}_n$ and $\tilde{\mathbf{h}}_n \perp \mathbf{w}_n$,

$$
E[\tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^*] = \frac{\rho}{M}(\mathbf{I}_P \otimes S_n) E[\tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^*](\mathbf{I}_P \otimes S_n^*) + \mathbf{I}_{MP} \tag{A.2}
$$

Now, note that $E[\hat{\mathbf{h}}_n \tilde{\mathbf{h}}_n^*] = \mathbf{0}$ implies that $E[\hat{\mathbf{h}}_n \mathbf{h}_n^*] = \sigma_{\hat{H}}^2 \mathbf{I}_{MP}$. Therefore,

$$
\begin{aligned}
E[\tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^*] &= E[(\mathbf{h}_n - \hat{\mathbf{h}}_n)(\mathbf{h}_n^* - \hat{\mathbf{h}}_n^*)] \\
&= E[\mathbf{h}_n \mathbf{h}_n^*] + E[\hat{\mathbf{h}}_n \hat{\mathbf{h}}_n^*] - E[\mathbf{h}_n \hat{\mathbf{h}}_n^*] - E[\hat{\mathbf{h}}_n \mathbf{h}_n^*] \\
&= \mathbf{I}_{MP} + \sigma_{\hat{H}}^2 \mathbf{I}_{MP} - 2\sigma_{\hat{H}}^2 \mathbf{I}_{MP} \\
&= (1 - \sigma_{\hat{H}}^2)\mathbf{I}_{MP}
\end{aligned}
$$

Using the above equation in (A.2), we get

$$
\begin{aligned}
E[\tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^*] &= \left[1 + \frac{\rho}{M}(1 - \sigma_{\hat{H}}^2)\right]\mathbf{I}_{MP} \\
&= \sigma_{\tilde{W}}^2 \mathbf{I}_{MP}
\end{aligned}
$$

A.4 Details of (4.13)

We have from (4.12)

$$
\hat{z}_n = \arg\max_{z_n \in \mathcal{L}} \Re\left[\mathbf{x}_n^* \left(\mathbf{I}_P \otimes V_{z_n}\right)\left(\mathbf{I}_P \otimes S_{n-1}\right)\hat{\mathbf{h}}_n\right]
$$

Substituting (4.11) in the above equation, we get

$$\begin{aligned}
\hat{z}_n &= \arg\max_{z_n\in\mathcal{L}} \Re\left[\mathbf{x}_n^*(\mathbf{I}_P\otimes V_{z_n})(\mathbf{I}_P\otimes S_{n-1})\xi\sum_{k=1}^{m-1} a_{0,k+1}^{(m)}(\mathbf{I}_P\otimes S_{n-k-1}^*)\mathbf{x}_{n-k-1}\right] \\
&= \arg\max_{z_n\in\mathcal{L}} \Re\left[\xi\sum_{k=1}^{m-1} a_{0,k+1}^{(m)}\mathbf{x}_n^*(\mathbf{I}_P\otimes V_{z_n})(\mathbf{I}_P\otimes S_{n-1}S_{n-k-1}^*)\mathbf{x}_{n-k-1}\right] \\
&= \arg\max_{z_n\in\mathcal{L}} \Re\left[\mathrm{tr}\left\{\xi\sum_{k=1}^{m-1} a_{0,k+1}^{(m)}X_n^*V_{z_n}S_{n-1}S_{n-k-1}^*X_{n-k-1}\right\}\right] \\
&= \arg\max_{z_n\in\mathcal{L}} \Re\left[\mathrm{tr}\left\{\xi\sum_{k=0}^{m-1} a_{0,k+1}^{(m)}X_n^*\left(V_{z_n}\prod_{j=1}^{k}V_{z_{n-j}}\right)X_{n-k-1}\right\}\right] \qquad \text{(A.3)}
\end{aligned}$$

which is (4.13).

A.5 Proof of (4.14) & (4.15)

Substituting $\prod_{j=i}^{i+k} V_{z_{n-j}} = S_{n-i}S_{n-i-k-1}^*$ in (4.1), we can write

$$\begin{aligned}
&\sum_{k=0}^{m-1}\sum_{i=0}^{m-k-1} a_{i,i+k+1}^{(m)}X_{n-i}^*\left(\prod_{j=i}^{i+k}V_{z_{n-j}}\right)X_{n-i-k-1} \\
&\qquad= \sum_{k=0}^{m-1}\sum_{i=0}^{m-k-1} a_{i,i+k+1}^{(m)}X_{n-i}^*S_{n-i}S_{n-i-k-1}^*X_{n-i-k-1} \\
&\qquad= \sum_{i=0}^{m-1}\sum_{k=0}^{m-i-1} a_{i,i+k+1}^{(m)}X_{n-i}^*S_{n-i}S_{n-i-k-1}^*X_{n-i-k-1} \\
&\qquad= \sum_{i=0}^{m-1} X_{n-i}^*S_{n-i}\left(\sum_{k=0}^{m-i-1} a_{i,i+k+1}^{(m)}S_{n-i-k-1}^*X_{n-i-k-1}\right) \\
&\qquad= \sum_{i=0}^{m-1} X_{n-i}^*S_{n-i}\underbrace{\left(\sum_{\ell=n-m}^{n-i-1} a_{i,n-\ell}^{(m)}S_\ell^*X_\ell\right)}_{\hat{H}_{n-i}|_{n-m}^{n-i-1}} \qquad \text{(A.4)}
\end{aligned}$$

which describes (4.14) & (4.15). Note that change of order of the summations is performed in the first step, and $\ell = n-i-k-1$ is substituted in the final step.

Now, we re-write (2.3) as $X_k = \sqrt{\frac{\rho}{M}} V_{z_k} S_{k-1} H_k + W_k, k = n-m+1, \ldots, n$. Assuming that $S_{k-1}H_k$ is known to the receiver, we write

$$\underbrace{\begin{bmatrix} \mathbf{x}_n \\ \vdots \\ \mathbf{x}_{n-m+1} \end{bmatrix}}_{\underline{\mathbf{x}}_n} = \sqrt{\frac{\rho}{M}} \underbrace{\begin{bmatrix} \mathbf{I}_P \otimes V_{z_n} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I}_P \otimes V_{z_{n-m+1}} \end{bmatrix}}_{\mathcal{V}_n} \underbrace{\begin{bmatrix} (\mathbf{I}_P \otimes S_{n-1})\mathbf{h}_n \\ \vdots \\ (\mathbf{I}_P \otimes S_{n-m})\mathbf{h}_{n-m+1} \end{bmatrix}}_{\underline{\mathbf{h}}_n} + \underbrace{\begin{bmatrix} \mathbf{w}_n \\ \vdots \\ \mathbf{w}_{n-m+1} \end{bmatrix}}_{\underline{\mathbf{w}}_n}$$

The multiple-symbol ML detector is therefore given as

$$\begin{aligned}
\{\hat{z}_k\}_{k=n-m+1}^{n} &= \underset{z_n, z_{n-1}, \ldots \in \mathcal{L}}{\arg\min} \left\| \underline{\mathbf{x}}_n - \sqrt{\frac{\rho}{M}} \mathcal{V}_n \underline{\mathbf{h}}_n \right\|^2 \\
&= \underset{z_n, z_{n-1}, \ldots \in \mathcal{L}}{\arg\min} \left(\underline{\mathbf{x}}_n - \sqrt{\frac{\rho}{M}} \mathcal{V}_n \underline{\mathbf{h}}_n \right)^* \left(\underline{\mathbf{x}}_n - \sqrt{\frac{\rho}{M}} \mathcal{V}_n \underline{\mathbf{h}}_n \right) \\
&= \underset{z_n, z_{n-1}, \ldots \in \mathcal{L}}{\arg\max} \underline{\mathbf{x}}_n^* \mathcal{V}_n \underline{\mathbf{h}}_n + \underline{\mathbf{h}}_n^* \mathcal{V}_n^* \underline{\mathbf{x}}_n \\
&= \underset{z_n, z_{n-1}, \ldots \in \mathcal{L}}{\arg\max} \Re\left[\underline{\mathbf{x}}_n^* \mathcal{V}_n \underline{\mathbf{h}}_n \right] \\
&= \underset{z_n, z_{n-1}, \ldots \in \mathcal{L}}{\arg\max} \Re\left[\sum_{k=0}^{m-1} \mathbf{x}_{n-k}^* \left(\mathbf{I}_P \otimes (V_{z_{n-k}} S_{n-k-1}) \right) \mathbf{h}_{n-k} \right] \\
&= \Re\left[\mathrm{tr} \left\{ \sum_{k=0}^{m-1} X_{n-k}^* V_{z_{n-k}} S_{n-k-1} H_{n-k} \right\} \right]
\end{aligned}$$

Observe that the above derivation also applicable when $S_{k-1}H_k$ is replaced by the estimate $S_{k-1}\hat{H}_k$ such that the sum of the channel estimation error and additive noise is spatio-temporally white, yielding detector (4.14).

A.6 Proof of (4.21), (4.22) & (4.23)

Recall that

$$R_{\mathbf{y}} K = \begin{bmatrix} \frac{\breve{\rho}}{M} (\mathbf{I}_P \otimes (V_1 - V_2)(V_1 - V_2)^*) + \mathbf{I}_{MP} & -\mathbf{I}_{MP} \\ \mathbf{I}_{MP} & -\mathbf{I}_{MP} \end{bmatrix} \tag{A.5}$$

So,

$$\det(\mathbf{I}_{2MP} - j\omega R_{\mathbf{y}} K) = \det \begin{bmatrix} (1-j\omega)\mathbf{I}_{MP} - j\omega \frac{\breve{\rho}}{M} (\mathbf{I}_P \otimes (V_1 - V_2)(V_1 - V_2)^*) & j\omega \mathbf{I}_{MP} \\ -j\omega \mathbf{I}_{MP} & (1+j\omega)\mathbf{I}_{MP} \end{bmatrix} \tag{A.6}$$

Using $\det\left[\begin{smallmatrix} A & C \\ B & D \end{smallmatrix}\right] = \det(A - BD^{-1}C)\det(D)$, we have

$$\begin{aligned}
&\det(\mathbf{I}_{2MP} - j\omega R_{\mathbf{y}} K) \\
&\quad = \det\left[(1-j\omega)\mathbf{I}_{MP} - j\omega\frac{\breve{\rho}}{M}(\mathbf{I}_P \otimes (V_1 - V_2)(V_1 - V_2)^*) - \frac{\omega^2}{1+j\omega}\mathbf{I}_{MP}\right] \times \\
&\qquad\qquad \det[(1+j\omega)\mathbf{I}_{MP}] \\
&\quad = \det\left[(1+\omega^2)\mathbf{I}_{MP} - (j\omega - \omega^2)\frac{\breve{\rho}}{M}(\mathbf{I}_P \otimes (V_1 - V_2)(V_1 - V_2)^*) - \omega^2\mathbf{I}_{MP}\right] \\
&\quad = \prod_{k=1}^{M}\left[1 - j\omega\frac{\breve{\rho}}{M}\sigma_k^2 + \frac{\breve{\rho}}{M}\omega^2\sigma_2\right]^P \\
&\quad = \prod_{k=1}^{M}\left(\frac{\breve{\rho}}{M}\sigma_k^2\right)^P\left[\frac{M}{\breve{\rho}\sigma_k^2} - j\omega + \omega^2\right]^P \qquad \text{(A.7)}
\end{aligned}$$

(4.21) follows by applying $(ax^2+bx+c) = a(x - \frac{-b+\sqrt{b^2-4ac}}{2a})(x + \frac{b+\sqrt{b^2-4ac}}{2a})$ in (A.7).

A.7 Proof of (4.25)

Perturbing the multiple poles $\{\tilde{p}_k^+\}_{k=1}^{n_Q}$ by small amounts to eliminate multiplicity, we get the set of simple poles $\{\tilde{\tilde{p}}_k^+\}_{k=1}^{MP}$. Replacing $\{\tilde{p}_k^+\}_{k=1}^{n_Q}$ by $\{\tilde{\tilde{p}}_k^+\}_{k=1}^{MP}$ in (4.24), we obtain the approximate characteristic function

$$\tilde{\Phi}_Q(\omega) = \left(\prod_{\ell=1}^{M}\frac{M}{\breve{\rho}\sigma_\ell^2(\omega - jp_\ell^-)}\right)^P \prod_{\ell=1}^{MP}\frac{1}{(\omega - j\tilde{\tilde{p}}_\ell^+)} \qquad \text{(A.8)}$$

Residue of $-\tilde{\Phi}_Q(\omega)/\omega$ at $\omega = j\tilde{\tilde{p}}_k^+$ is given by

$$\begin{aligned}
\text{Res}\left[-\frac{\tilde{\Phi}_Q(\omega)}{\omega}\right]_{\omega=j\tilde{\tilde{p}}_k^+} &= \left[-\frac{\tilde{\Phi}_Q(\omega)}{\omega}(\omega - j\tilde{\tilde{p}}_k^+)\right]_{\omega=j\tilde{\tilde{p}}_k^+} \\
&= \frac{-1}{j\tilde{\tilde{p}}_k^+}\left(\prod_{\ell=1}^{M}\frac{M}{\breve{\rho}\sigma_\ell^2(j\tilde{\tilde{p}}_k^+ - jp_k^-)}\right)^P \prod_{\ell=1,\ell\neq k}^{MP}\frac{1}{(j\tilde{\tilde{p}}_k^+ - j\tilde{\tilde{p}}_\ell^+)} \\
&= \frac{-1}{j^{2MP}\tilde{\tilde{p}}_k^+}\left(\prod_{\ell=1}^{M}\frac{M}{\breve{\rho}\sigma_\ell^2(\tilde{\tilde{p}}_k^+ - p_k^-)}\right)^P \prod_{\ell=1,\ell\neq k}^{MP}\frac{1}{(\tilde{\tilde{p}}_k^+ - \tilde{\tilde{p}}_\ell^+)}
\end{aligned}$$

$$= \frac{1}{(-1)^{MP-1}\tilde{\tilde{p}}_k^+}\left(\prod_{\ell=1}^{M}\frac{M}{\breve{\rho}\sigma_\ell^2(\tilde{\tilde{p}}_k^+ - p_k^-)}\right)^P \prod_{\ell=1,\ell\neq k}^{MP}\frac{1}{(\tilde{\tilde{p}}_k^+ - \tilde{\tilde{p}}_\ell^+)}$$

$$= \frac{1}{\tilde{\tilde{p}}_k^+}\left(\prod_{\ell=1}^{M}\frac{M}{\breve{\rho}\sigma_\ell^2(\tilde{\tilde{p}}_k^+ - p_k^-)}\right)^P \prod_{\ell=1,\ell\neq k}^{MP}\frac{1}{(\tilde{\tilde{p}}_\ell^+ - \tilde{\tilde{p}}_k^+)} \qquad \text{(A.9)}$$

From (4.18) and (A.9) we obtain (4.25).

A.8 Details of (4.28), (4.29), & (4.30)

Using $B_a = \sqrt{\rho_a/M}\mathcal{S}_{n-1}\left((T_a^{(m-1)^{-1}}\mathbf{g}_a)\otimes\mathbf{I}_{MP}\right)$ and $\hat{\hat{\mathbf{h}}}_n = B_a^*\underline{\mathbf{x}}_{n-1}$ we can write

$$\begin{aligned}
E[\hat{\hat{\mathbf{h}}}_n\hat{\hat{\mathbf{h}}}_n^*] &= B_a^*E[\underline{\mathbf{x}}_{n-1}\underline{\mathbf{x}}_{n-1}^*]B_a \\
&= B_a^*[\mathcal{S}_{n-1}\left(T^{(m-1)}\otimes\mathbf{I}_{MP}\right)\mathcal{S}_{n-1}^*]B_a \\
&= \frac{\rho_a}{M}\left((\mathbf{g}_a^*T_a^{(m-1)^{-1}})\otimes\mathbf{I}_{MP}\right)\left(T^{(m-1)}\otimes\mathbf{I}_{MP}\right)\left((T_a^{(m-1)^{-1}}\mathbf{g}_a)\otimes\mathbf{I}_{MP}\right) \\
&= \frac{\rho_a}{M}\left(\mathbf{g}_a^*T_a^{(m-1)^{-1}}T^{(m-1)}T_a^{(m-1)^{-1}}\mathbf{g}_a\right)\mathbf{I}_{MP} \\
&= \hat{\hat{\sigma}}_h^2\mathbf{I}_{MP}
\end{aligned}$$

which is (4.28). Now recall that $\tilde{\tilde{\mathbf{w}}}_n = \sqrt{\frac{\rho}{M}}(\mathbf{I}_P\otimes S_n)\tilde{\tilde{\mathbf{h}}}_n + \mathbf{w}_n$ and $\tilde{\tilde{\mathbf{h}}}_n = \mathbf{h}_n - \hat{\hat{\mathbf{h}}}_n$. Therefore,

$$\begin{aligned}
E[\tilde{\tilde{\mathbf{h}}}_n\tilde{\tilde{\mathbf{h}}}_n^*] &= E[(\mathbf{h}_n - \hat{\hat{\mathbf{h}}}_n)(\mathbf{h}_n - \hat{\hat{\mathbf{h}}}_n)^*] \\
&= E[\mathbf{h}_n\mathbf{h}_n^*] - E[\mathbf{h}_n\hat{\hat{\mathbf{h}}}_n^*] - E[\hat{\hat{\mathbf{h}}}_n\mathbf{h}_n^*] + E[\hat{\hat{\mathbf{h}}}_n\hat{\hat{\mathbf{h}}}_n^*] \\
&= (1+\hat{\hat{\sigma}}_h^2)\mathbf{I}_{MP} - 2\Re\{B_a^*E[\underline{\mathbf{x}}_n\mathbf{h}_n^*]\} \\
&= (1+\hat{\hat{\sigma}}_h^2)\mathbf{I}_{MP} - 2\sqrt{\rho\rho_a/M}\Re\left\{\left((\mathbf{g}_a^*T_a^{(m-1)^{-1}})\otimes\mathbf{I}_{MP}\right)\mathcal{S}_{n-1}^*\mathcal{S}_{n-1}\mathbf{g}\right\} \qquad \text{(A.10)} \\
&= \left(1+\hat{\hat{\sigma}}_h^2 - 2\sqrt{\rho\rho_a/M}\Re\left\{\mathbf{g}_a^*T_a^{(m-1)^{-1}}\mathbf{g}\right\}\right)\mathbf{I}_{MP}
\end{aligned}$$

and

$$\begin{aligned}
E[\tilde{\tilde{\mathbf{w}}}_n\tilde{\tilde{\mathbf{w}}}_n^*] &= \frac{\rho}{M}(\mathbf{I}_P\otimes S_n)E[\tilde{\tilde{\mathbf{h}}}_n\tilde{\tilde{\mathbf{h}}}_n^*](\mathbf{I}_P\otimes S_n^*) + \mathbf{I}_{MP} \\
&= \frac{\rho}{M}\left(1+\hat{\hat{\sigma}}_h^2 - 2\sqrt{\rho\rho_a/M}\Re\left\{\mathbf{g}_a^*T_a^{(m-1)^{-1}}\mathbf{g}\right\}\right)\mathbf{I}_{MP} + \mathbf{I}_{MP}
\end{aligned}$$

$$= \left((\frac{\rho}{M} + 1) - \frac{2\sqrt{\rho^3 \rho_a}}{M^2} \Re(\mathbf{g}^* T_a^{(m-1)^{-1}} \mathbf{g}_a) + \frac{\rho}{M} \hat{\hat{\sigma}}_h^2 \right) \mathbf{I}_{MP}$$

Again,

$$\begin{aligned} E[\tilde{\tilde{\mathbf{w}}}_n \hat{\hat{\mathbf{h}}}_n^*] &= \sqrt{\frac{\rho}{M}} (\mathbf{I}_P \otimes S_n) E[\tilde{\tilde{\mathbf{h}}}_n \hat{\hat{\mathbf{h}}}_n^*] \\ &= \sqrt{\frac{\rho}{M}} (\mathbf{I}_P \otimes S_n) \left(E[\mathbf{h}_n \hat{\hat{\mathbf{h}}}_n^*] - E[\hat{\hat{\mathbf{h}}}_n \hat{\hat{\mathbf{h}}}_n^*] \right) \\ &= \sqrt{\frac{\rho}{M}} (\mathbf{I}_P \otimes S_n) \left(\sqrt{\rho \rho_a / M} \{ \mathbf{g}^* T_a^{(m-1)^{-1}} \mathbf{g}_a \} - \hat{\hat{\sigma}}_h^2 \right) \mathbf{I}_{MP} \quad \text{(see (A.10))} \end{aligned}$$

which implies (4.30).

A.9 Proof of (4.32) & (4.33)

From (4.31) it can be shown that, given that V_1 was transmitted, the receiver will detect V_2 if (4.17) is true, where $\mathbf{y}_1 = \sqrt{\frac{\breve{\rho}}{M}} (\mathbf{I}_P \otimes (V_1 - V_2)) \breve{\breve{\mathbf{h}}}_n + \breve{\breve{\mathbf{w}}}_n$ and $\mathbf{y}_2 = \breve{\breve{\mathbf{w}}}_n$. Using (4.28), (4.29), (4.30), (4.31), $\mathbf{y} = [\mathbf{y}_1^* \;\; \mathbf{y}_2^*]^*$, and defining $\tau = \frac{\sigma_{wh}^2}{\hat{\hat{\sigma}}_h \tilde{\tilde{\sigma}}_w}$, $V = V_1 V_2^*$, we can write

$$R_{\mathbf{y}} = E[\mathbf{y}\mathbf{y}^*] = \begin{bmatrix} \mathbf{I}_{MP} + (\frac{\breve{\rho}}{M} + \tau \sqrt{\frac{\breve{\rho}}{M}}) (\mathbf{I}_P \otimes (2\mathbf{I}_M - V - V^*)) & \mathbf{I}_{MP} + \tau \sqrt{\frac{\breve{\rho}}{M}} (\mathbf{I}_P \otimes (\mathbf{I}_M - V^*)) \\ \mathbf{I}_{MP} + \tau \sqrt{\frac{\breve{\rho}}{M}} (\mathbf{I}_P \otimes (\mathbf{I}_M - V)) & \mathbf{I}_{MP} \end{bmatrix} \quad \text{(A.11)}$$

Using $\det \left[\begin{smallmatrix} A & C \\ B & D \end{smallmatrix} \right] = \det(A - BD^{-1}C)\det(D)$, $\det(\mathbf{I}_P \otimes A) = (\det(A))^P$ and $2\mathbf{I}_M - V - V^* = (V_1 - V_2)(V_1 - V_2)^*$, it is straight-forward to simplify $\det(\mathbf{I}_{2MP} - j\omega R_{\mathbf{y}} K)$ to

$$\begin{aligned} &\det(\mathbf{I}_{2MP} - j\omega R_{\mathbf{y}} K) = \\ &\quad \det \Big[1 - j\omega(\breve{\rho}/M + \tau\sqrt{\breve{\rho}/M})(V_1 - V_2)(V_1 - V_2)^* + \\ &\qquad\qquad \omega^2 \breve{\rho}/M(1 - \tau^2)(V_1 - V_2)(V_1 - V_2)^* \Big]^P \end{aligned} \quad \text{(A.12)}$$

Now denoting the k^{th} singular value of $V_1 - V_2$ as σ_k^2, we have

$$\det(\mathbf{I}_{2MP} - j\omega R_{\mathbf{y}} K) = \prod_{k=1}^{M} \left[1 - j\omega(\breve{\rho}/M + \tau\sqrt{\breve{\rho}/M})\sigma_k^2 + \omega^2 \breve{\rho}/M(1 - \tau^2)\sigma_k^2 \right]^P \quad \text{(A.13)}$$

Using $(ax^2 + bx + c) = a(x - \frac{-b+\sqrt{b^2-4ac}}{2a})(x + \frac{b+\sqrt{b^2-4ac}}{2a})$ in (A.13) we get (4.32) and (4.33).

BIBLIOGRAPHY

[1] J.G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, 3rd edition, 1995.

[2] J.G. Lawton, "Investigation of digital data communication systems", Tech. Rep. UA-1420-S-1, Cornell Aeronautical Laboratory, Inc., 1961.

[3] I.E. Telatar, "Capacity of multi-antenna Gaussian channels", Tech. Rep., AT&T Bell Labs, June 1995.

[4] G.G. Raleigh and J.M. Cioffi, "Spatio-temporal coding for wireless communication", *IEEE Trans. on Communications*, vol. 46, pp. 357–366, Mar. 1998.

[5] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas", *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, Aug. 1996.

[6] G.J. Foschini and M.J.Gans, "On the limits of wireless communication in a fading environment when using multiple antennas", *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[7] A. Naguib, N. Seshadri, and A.R. Calderbank, "Increasing data rate over wireless channels", *IEEE Signal Processing Magazine*, vol. 17, pp. 76–92, May 2000.

[8] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction", *IEEE Trans. on Information Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[9] B.M. Hochwald and T.L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading", *IEEE Trans. on Information Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[10] B.M. Hochwald, T.L. Marzetta, T.L. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations", *IEEE Trans. on Information Theory*, vol. 46, no. 6, pp. 1962–1973, Sept. 2000.

[11] B.M. Hochwald and W. Sweldens, "Differential unitary space-time modulation", *IEEE Trans. on Communications*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.

[12] B.L. Hughes, "Differential space-time modulation", *IEEE Trans. on Information Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.

[13] B. Hassibi and B.M. Hochwald, "Cayley differential unitary space-time codes", *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1485–1503, June 2002.

[14] Z. Liu, G.B. Giannakis, and B.L. Hughes, "Doubly differential space-time block coding for time-selective fading channels", *IEEE Trans. on Communications*, Sept. 2000, submitted.

[15] H. Jafarkhani and V. Tarokh, "Multiple transmit antenna differential detection from generalized orthogonal designs", *IEEE Trans. on Information Theory*, vol. 47, no. 8, pp. 2626–2631, Sept. 2001.

[16] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity", *IEEE Journal on Selected Areas In Communications*, vol. 18, pp. 1169–1174, July 2000.

[17] G. Ganesan and P. Stoica, "Differential modulation using space-time block codes", *IEEE Signal Processing Letters*, vol. 9, no. 2, pp. 57–60, Feb. 2002.

[18] 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 Trans. on Communication Systems*, vol. CS-10, pp. 160–168, June 1962.

[19] C. Peel and A. Swindlehurst, "Performance of unitary space-time modulation in continuously changing channel", in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, 2001.

[20] C. Peel and A. Swindlehurst, "Performance of unitary space-time modulation in continuously changing channel", in *Proc. IEEE Intern. Conf. on Communication*, 2001.

[21] P. Ho and D. Fung, "Error performance of multiple symbol differential detection of PSK signals transmitted over correlated Rayleigh fading channels", *Proc. IEEE Intern. Conf. on Communication*, vol. 2, pp. 568 –574, 1991.

[22] R. Schober, W.H. Gerstacker and J.B. Huber, "Decision feedback differential detection of MDPSK for flat Rayleigh fading channels", *IEEE Trans. on Communications*, vol. 47, no. 7, pp. 1025–1035, July 1999.

[23] D. Divsalar and M.K. Simon, "Multiple-symbol differential detection of MPSK", *IEEE Trans. on Communications*, vol. 38, no. 3, pp. 300–308, March 1990.

[24] H. Leib and S. Pasupathy, "The phase of a vector perturbed by Gaussian noise and differentially coherent receivers", *IEEE Trans. on Information Theory*, vol. 34, no. 6, pp. 1491–1501, Nov. 1988.

[25] F. Adachi and M. Sawahashi, "Decision feedback multiple-symbol differential detection of M-ary DPSK", *Electronics Letters*, vol. 29, no. 15, pp. 1385–1387, July 1993.

[26] F. Edbauer, "Bit error rate of binary and quaternary DPSK signals with multiple differential feedback detection", *IEEE Trans. on Communications*, vol. 40, pp. 457–460, March 1992.

[27] D. Divsalar and M.K. Simon, "Maximum likelihood differential detection of uncoded and trellis coded amplitude phase modulation over AWGN and fading channel-metrics and performance", *IEEE Trans. on Communications*, vol. 42, no. 1, pp. 76–89, Jan. 1994.

[28] 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. on Information Theory*, vol. 48, no. 4, pp. 950–956, April 2002.

[29] R. Schober and H.-J. Lampe, "Noncoherent receivers for differential space-time modulation", *Proc. IEEE Global Telecommunications Conf.*, vol. 2, pp. 1127–1131, 2001.

[30] R. Schober and H.-J. Lampe, "Noncoherent receivers for differential space-time modulation", *IEEE Trans. on Communications*, vol. 50, no. 5, pp. 768–777, May 2002.

[31] X. Yu and S. Pasupathy, "Innovations-based MLSE for rayleigh fading channels", *IEEE Trans. on Communications*, vol. 43, no. 2/3/4, pp. 1534–1544, Feb./Mar./April 1995.

[32] K. Miller, *Complex Stochastic Processes*, Addison-Wesley, New York, 1974.

[33] K. M. Mackenthum Jr., "A fast algorithm for multiple-symbol differential detection of MDPSK", *IEEE Trans. on Communications*, vol. 42, pp. 1471–1474, Feb./Mar./Apr. 1994.

[34] B. Bhukania and P. Schniter, "Multiple-symbol detection of differential unitary space-time modulation in fast Rayleigh-fading channels", Submitted to *IEEE Trans. on Communications*, April 2002.

[35] F. Adachi, "Reduced state transition Viterbi differential detection of M-ary DPSK signals", *Electronics Letters*, vol. 32, no. 12, pp. 1064–1066, June 1996.

[36] G.M. Vitetta and D.P. Taylor, "Maximum likelihood detection of differentially encoded PSK signals transmitted over Rayleigh frequency-flat fading channels", *Proc. IEEE Global Telecommunications Conf.*, vol. 3, pp. 1889–1893, 1994.

[37] J.K. Cavers and P. Ho, "Analysis of the error performance of trellis-coded modulations in Rayleigh fading channels", *IEEE Trans. on Communications*, vol. 40, pp. 74 –83, Jan. 1992.

[38] G.L. Turin, "The characteristic function of Hermitian quadratic forms in complex normal variables", *Biometrika*, pp. 199–201, 1960.

[39] E. Biglieri, G. Caire and G. Taricco, "Approximating the pairwise error probability for fading channels", *Electronics Letters*, vol. 31, pp. 1625–1627, Sept. 1995.

[40] E. Biglieri, G. Caire and J. Vetura-Traveset, "Simple method for evaluating error probabilities", *Electronics Letters*, vol. 32, pp. 191–192, Feb. 1996.

[41] W.C. Jakes, *Microwave Mobile Communications*, IEEE press, Piscataway, NJ, 1993.

[42] B. Bhukania and P. Schniter, "Decision feedback detection of differential unitary space-time modulation in fast Rayleigh-fading channels", Submitted to *IEEE Trans. on Wireless Communications*, Aug. 2002.

[43] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, 3rd edition, 1996.

[44] J.G. Proakis and M. Salehi, *Communication systems engineering*, Prentice Hall, Inc., New Jersey, 1994.

[45] F. Adachi, "Adaptive differential detection using linear prediction for M-ary DPSK", *IEEE Trans. on Vehicular Technology*, vol. 47, no. 3, pp. 909–918, Aug. 1998.

[46] H. Kong and E. Shwedyk, "Detection of amplitude-phase modulated signals over frequency nonselective rayleigh fading channels with adaptive symbol-aided channel estimation", *Proc. IEEE Vehicular Technology Conference*, vol. 2, pp. 983–987, 1996.

[47] H. Rohling and T. May, "Comparison of PSK and DPSK modulation in a coded OFDM system", *Proc. IEEE Vehicular Technology Conference*, vol. 2, pp. 870–874, 1997.

[48] K.L. Baum and N.S. Nadgauda, "A comparison of differential and coherent reception for a coded OFDM system in a low C/I environment", *Proc. IEEE Global Telecommunications Conf.*, vol. 1, pp. 300–304, 1997.

[49] A. Svensson, "1 and 2 stage decision feedback coherent detectors for DQPSK in fading channels", *Proc. IEEE Vehicular Technology Conference*, vol. 2, pp. 644–648, 1995.

[50] P. Frenger and A. Svensson, "Decision-directed coherent detection in multicarrier systems on rayleigh fading channels", *IEEE Trans. on Vehicular Technology*, vol. 48, no. 2, pp. 490–498, March 1999.

[51] M. Lott, "Comparision of frequency and time domain differential modulation in an OFDM system for wireless ATM", *Proc. IEEE Vehicular Technology Conference*, vol. 2, pp. 877–883, 1999.