

Communication over Doubly Selective Channels:
Efficient Equalization and Max-Diversity Precoding

Dissertation

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Sung Jun Hwang, B.S., M.S.

Graduate Program in Electrical and Computer Engineering

The Ohio State University

2010

Dissertation Committee:

Philip Schniter, Advisor

Hesham El-Gamal

Lee C. Potter

© Copyright by

Sung Jun Hwang

2010

ABSTRACT

We consider the problem of practical communication over a doubly selective (DS) channel, i.e., a time and frequency selective channel. The problem is approached in two different ways: coherent communication and noncoherent communication, and for each communication scheme we propose practical and near-optimal equalizers and maximum-diversity precoders. Toward these ends, we adopt 1) basis expansion (BE) modeling of the channel, which allows for an efficient and unified way of describing a DS channel in both time and frequency domain; and 2) tree-search algorithms (TSAs), which facilitate near-optimal performance with low complexity.

For practical coherent communication, we focus on the pulse-shaped (PS) multi-carrier modulation (MCM), where controlled inter-symbol-interference (ISI) and inter-carrier-interference (ICI) can be leveraged for computationally efficient receiver structures. Then, we propose a novel channel adaptive TSA with a novel fast minimum mean-squared error (MMSE) generalized decision-feedback equalizer (GDFE) preprocessing, and a rank-reduced channel estimation by using the BE channel model. Also, a new finding about optimality of MMSE-GDFE preprocessing is presented, which states that under constant modulus constellation the minimum distance property is preserved by the MMSE-GDFE preprocessing.

Then, two practically realizable noncoherent equalization schemes are proposed: a sequential algorithm and a Bayesian expectation maximization (EM)-based algorithm. The sequential algorithm is derived from the optimal noncoherent metric, and made practical by a fast algorithm and a TSA to evaluate and search over the metric. The Bayesian EM-based noncoherent algorithm is derived from optimal maximum *a posteriori* (MAP) estimation of the BE parameters, and efficiently implemented via iteration between soft coherent equalizer and soft channel estimator. Efficient operations are accomplished using fast algorithms whose overall complexities grow linearly in the block size and quadratically in the number of BE parameters. Also, we demonstrate that the noncoherent equalization can be readily applied to the communication problem in a highly spread underwater acoustic channel (UAC).

Finally, we establish maximum-diversity conditions for each affine and linear precoder, which imply that under some mild channel assumptions almost any random affine (linear) precoder facilitates the maximum-diversity noncoherent (coherent) reception.

dedicated to Noory, Jion and Andy

ACKNOWLEDGMENTS

I thank my advisor Prof. Phil Schniter for guiding me all the way through to finish this work, and for his unbelievable generosity, patience and support. He taught me how to think, how to write, and most importantly how to enjoy life doing research. This five year period under his guidance will be remembered as one of most fruitful and yet entertaining days of my life. Many thanks also go to my committee members, Prof. Hesham El-Gamal, Prof. Lee Potter and Prof. Andrea Serrani. Prof. El-Gamal led me to the world of sequential decoding and taught me to see the problems with big pictures.

I thank my IPS colleagues for making our lab a place filled with not only academic discussions and problem solvings but also fun of cross-cultural entertainments. I thank Jeri McMichael for all her effort to make all the complicated, lab-related things smooth.

Finally, I would like to thank my family for unmeasurable love and support. I'm tremendously indebted to Noory for following me to U.S. Especially, for her heart-warming care and inspiring encouragement.

I cannot possibly acknowledge quite enough all the people who has given to me an opportunity for whatever entertaining during the stay in Columbus. Again, I'll be always missing all the days of my stay in the OSU.

VITA

January 13, 1970 Born - Pusan, Republic of Korea

1993 B.S. Electronics,
Yonsei University, Republic of Korea

1995 M.S. Electronics,
Yonsei University, Republic of Korea

1995-2003 Engineer,
Samsung Electronics Inc., Republic of
Korea

2004-present Graduate Student,
The Ohio State University.

PUBLICATIONS

- S.-J. Hwang and P. Schniter, “Soft Noncoherent Equalization Schemes for Doubly Selective Channels using Tree Search and Basis Expansion,” *to be submitted Jan. 2010*
- P. Schniter, S.-J. Hwang, S. Das, and A. P. Kannu, “Equalization of Time-Varying Channels,” *Wireless Communications over Rapidly Time-Varying Channels*, (F. Hlawatsch and G. Matz, eds.), Elsevier Academics Press, 2009.
- S.-J. Hwang and P. Schniter, “Efficient Communication over Highly Spread Underwater Acoustic Channels,” *IEEE Journal on Selected Areas in Communication*, vol. 26, no. 9, pp. 1674-1683, Dec. 2008.
- S.-J. Hwang and P. Schniter, “Efficient Sequence Detection of Multi-Carrier Transmissions over Doubly Selective Channels,” *EURASIP Journal on Applied Signal Processing special issue on Reliable Communications over Rapidly Time-Varying Channels*, Article ID 93638, 2006.

- S.-J. Hwang and P. Schniter, “EM-Based Soft Noncoherent Equalization of Doubly Selective Channel using Tree Search and Basis Expansion,” *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, (Perugia, Italy), June 2009.
- S.-J. Hwang and P. Schniter, “Fast Noncoherent Decoding of Block Transmission Over Doubly Dispersive Channels,” *Proc. Asilomar Conf. on Signals, Systems, and Computers (IEEE co-sponsored)*, (Pacific Grove, CA), Nov. 2007.
- S.-J. Hwang and P. Schniter, “Efficient Communication over Highly Spread Underwater Acoustic Channels,” *Proc. The Second ACM International Workshop on Underwater Networks (IEEE co-sponsored)*, (Montreal, Quebec), Sept. 2007.
- S.-J. Hwang and P. Schniter, “Maximum Diversity Affine Precoding for the Noncoherent Doubly Dispersive Channel,” *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, (Helsinki, Finland), June 2007.
- S.-J. Hwang and P. Schniter, “Near-Optimal Noncoherent Sequence Detection for Doubly Dispersive Channels,” *Proc. Asilomar Conf. on Signals, Systems, and Computers (IEEE co-sponsored)*, (Pacific Grove, CA), Nov. 2006. (Invited.)
- S.-J. Hwang and P. Schniter, “Efficient Sequence Detection of Multi-Carrier Transmissions Over Doubly Dispersive Channels,” *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, (Cannes, France), July 2006.
- S.-J. Hwang and P. Schniter, “On the Optimality of MMSE-GDFE Pre-Processed Sphere Decoding,” *Proc. Allerton Conf. on Communication, Control, and Computing (IEEE co-sponsored)*, (Monticello, IL), Oct. 2005.

FIELDS OF STUDY

Major Field: Electric and Computer Engineering

Studies in:

Wireless Comm. and Signal Processing	Prof. Phil Schniter
Wireless Comm. and Signal Processing	Prof. Hesham El-Gamal
Wireless Comm. and Signal Processing	Prof. Lee C. Potter

TABLE OF CONTENTS

	Page
Abstract	ii
Dedication	iv
Acknowledgments	v
Vita	vi
List of Tables	xii
List of Figures	xiii
Chapters:	
1. Introduction	1
1.1 Motivation	3
1.2 Contributions and Outline	4
1.3 Notation	6
2. Problem Statement and System Model	9
2.1 Coherent Equalization for Multi-Carrier Systems	10
2.2 Noncoherent Equalization	11
2.3 Maximum-Diversity Precoding	13
2.4 System Model	14
2.4.1 Single-Carrier Modulation/Demodulation	16
2.4.2 Multi-Carrier Modulation/Demodulation	16
2.4.3 Real-Valued System Model for Tree Search	17
2.4.4 Basis Expansion Modeling of DS Channel	18

3.	Tree Search Algorithms	23
3.1	Preprocessing	25
3.2	Tree Search Algorithms	28
3.2.1	TSAs for MCM Application	28
3.3	Optimality of MMSE-GDFE Preprocessed Estimates	33
3.3.1	Numerical Experiments	37
4.	Near-Optimal Coherent Equalization for PS-MCM	42
4.1	Proposed PS-MCM Coherent Equalization	43
4.1.1	TSA Preprocessing	43
4.1.2	Channel-Adaptive T-algorithm	46
4.1.3	Channel Estimation	49
4.2	Numerical Results	52
4.2.1	FER Performance	54
4.2.2	Complexity	60
4.2.3	Effect of Imperfect Channel Estimates	62
5.	Near-Optimal Noncoherent Equalization	65
5.1	Basis Expansion Modeling	67
5.2	Sequential Noncoherent Equalization	69
5.2.1	Optimum Soft Noncoherent Equalization	69
5.2.2	Practical Sequential Soft Noncoherent Equalization	71
5.3	Equalization via the Bayesian EM Algorithm	72
5.3.1	Exact Bayesian-EM Soft Equalization	72
5.3.2	Practical EM-Based Soft Noncoherent Equalization	75
5.3.3	Refinement of the Prior	76
5.4	Implementation Details	76
5.4.1	Choice of $\tilde{\boldsymbol{\theta}}^{(j)}$ and \mathbf{R}_θ	76
5.4.2	Pilot and Guard Patterns	78
5.4.3	Complexity	80
5.5	Numerical Results	81
5.5.1	Effect of System Parameters	82
5.5.2	Performance Comparison	86
5.5.3	Robustness to Statistical Mismatch	87
5.5.4	Multi-carrier Transmission	91

6.	Application to UAC Communication	95
6.1	System Model	100
6.1.1	Modifications for Noncoherent Soft Equalization	100
6.2	Noncoherent Soft Equalization	102
6.3	Tracking Sparsity	103
6.3.1	Active-Tap Identification	103
6.3.2	Residual Tap Compensation	106
6.4	Numerical Results	106
6.4.1	Setup	107
6.4.2	BER versus Number of Pilot Clusters K	109
6.4.3	BER versus SNR	109
7.	Maximum-Diversity Precoding	113
7.1	System Model	114
7.2	Diversity-Order Analysis	115
7.2.1	Noncoherent Receiver	115
7.2.2	Coherent Receiver	120
7.3	Numerical Examples	121
8.	Conclusions	123
8.1	Summary of the Work	123
8.2	Possible Future Research	126
Appendices:		
A.	Fourier Basis Representation of MCM channel coefficients	128
A.1	A Sparse Frequency-Domain BEM	130
A.2	Covariance of the BE Parameter	131
B.	Derivations for Noncoherent Equalization Algorithms	133
B.1	A Fast Recursive Update for $\mu(\mathbf{x}_n)$ (Table 5.1)	133
B.1.1	Recursion for $\mu_1(\mathbf{x}_n)$	133
B.1.2	Recursion for $\mu_2(\mathbf{x}_n)$	135
B.2	Derivation of (5.12)	136

C.	Proofs for Chapter 7	137
C.1	Proof of Lemma 6	137
C.2	Proof of Lemma 8	140
	Bibliography	142

LIST OF TABLES

Table	Page
1.1 List of abbreviations	8
4.1 Fast Cholesky factorization of quasi-banded \mathbf{A}	45
5.1 Fast recursion for $\mu(\mathbf{x}_n)$	70
5.2 N -step computation of $\hat{\boldsymbol{\theta}}[i+1]$	76
5.3 Summary of computational complexity	81

LIST OF FIGURES

Figure	Page
1.1 Time-varying channel occurs by mobility of transmitter, reflector and/or receiver.	2
2.1 Effective degrees of variation N_b versus normalized single-sided Doppler spread $f_D T_s$ for Jakes' channel at different block lengths N	21
3.1 Channel matrices associated with MCM: (a) "quasi-banded" channel matrix, (b) "V-shaped" channel matrix.	25
3.2 Illustration of $\boldsymbol{\rho} = \tilde{\mathbf{R}}\bar{\mathbf{s}} + \mathbf{n}$ for V-shaped $\tilde{\mathbf{R}}$. The PAM symbol \bar{s}_{L-2D-1} does not affect $\{\rho_0, \dots, \rho_{L-4D-2}\}$	31
3.3 A comparison of the per-frame average search complexity of Schnorr-Euchner (SE) SpD with two forms of preprocessing: ZF-DFE versus MMSE-GDFE, both with greedy ordering. A BPSK system of dimension $M = N = 32$ was employed and $\gamma = \text{SNR}$ was used in MMSE-GDFE.	39
3.4 Frame error rate of the ML detector (via ZF-DFE preprocessed SpD) versus the MMSE-GDFE preprocessed SpD for a BPSK system under several channel dimensions (M, N) and several values of MMSE-GDFE parameter γ	40
3.5 Frame error rate of the ML detector (via ZF-DFE preprocessed SpD) versus the MMSE-GDFE preprocessed SpD for a 4-PAM system with $M = N = 8$	41
4.1 Illustration of path evolution in the T-algorithm when $Q = 2$ and $L = 4$. The circled points denote the minimum path metrics, the crossed points denote the discarded path metrics, and the bold line denotes the true path. Note that, in this example, $\mathcal{M}(\bar{\mathbf{s}}_*^2) < \mathcal{M}(\bar{\mathbf{s}}_7^2)$	47

4.2	ML and MFB performance of several MCM schemes using global ICI (“full H”) or local ICI (“D=6”) at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$.	57
4.3	MMSE-DFE performance of several MCM schemes using global ICI (“full H”) or local ICI (“D=6”) at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$.	58
4.4	Performance of several TSAs on doubly selective PS-MCM with perfect knowledge of local ICI (i.e., $D = 6$) at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$.	59
4.5	Number of real MAC operations per frame for doubly selective PS-MCM reception at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$. Viterbi complexity above assumes inactive edge subcarriers (as in [1]); for active subcarriers, the tail-biting version would require $L^{5.10}$ MACs/frame.	61
4.6	Performance of several TSAs operating on doubly selective PS-MCM with rank-reduced pilot-aided estimates of local-ICI (i.e., $D = 6$) at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$, using pilot spacing $P = 2$.	63
4.7	Performance of several TSAs operating on doubly selective PS-MCM with rank-reduced pilot-aided estimates of local-ICI (i.e., $D = 6$) at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$, using pilot spacing $P = 3$.	64
5.1	Turbo receiver with soft noncoherent equalizer.	68
5.2	For $N = 32$, $N_H = 3$, and $N_p = 8$, illustration of a) $\mathbf{H}^{(j)}$ support, b) single-carrier pilot pattern, and c) multi-carrier pilot pattern with $P = 4$ and $L = 2$.	80
5.3	Coded BER vs. E_b/N_o for ncT-BE with M-algorithm parameter $M \in \{16, 32, 64, 128\}$.	83
5.4	BER vs. E_b/N_o for turbo iterations $\in \{1, 4, 8\}$. ncT-BE and (cT+sBE) ³ schemes were tested.	84
5.5	ncT-BE algorithm’s BER vs. number of pilots $\in \{1, 3, 6, 9, 12\}$.	85
5.6	BER vs. E_b/N_o for various equalization schemes under $f_D T_s = 0.002$.	88
5.7	BER vs. E_b/N_o for various equalization schemes under very high Doppler with $f_D T_s = 0.005$.	89

5.8	Required E_b/N_o to achieve 10^{-2} BER vs. true $f_D T_s$ for $f_D T_s$ -matched and fixed- $f_D T_s$ (i.e. mismatched) reception.	90
5.9	BER vs. E_b/N_o for various frequency-domain equalization schemes under $f_D T_s = 0.002$ using $L = 1$	93
5.10	BER vs. E_b/N_o for various frequency-domain equalization schemes under $f_D T_s = 0.002$ using $L = 4$	94
6.1	Receiver structure.	102
6.2	Illustration of pilot pattern with $N = 32$, $P = 4$, $K = 2$, $N_p = 8$, and $D = 1$. The columns represent $\mathcal{C}^{(j-P+2)}, \dots, \mathcal{C}^{(j+2P-3)}$, respectively.	104
6.3	BER versus number of pilot clusters for the “sparse” channel at $E_b/N_o = 9$ dB.	110
6.4	BER versus SNR for the “perfectly sparse” channel. The noncoherent soft equalizer is compared to coherent soft equalizer with genie-estimated $\boldsymbol{\theta}^{(j)}$ for both true and estimated $\mathcal{L}^{(j)}$	111
6.5	BER versus SNR for the “sparse” channel. The noncoherent soft equalizer is compared to coherent soft equalizer with genie-estimated $\boldsymbol{\theta}^{(j)}$ for both true and estimated $\mathcal{L}^{(j)}$	112
7.1	Average PWEF versus SNR for $N = 8$, $N_h = N_b = 2$, and various N_s . The dashed line confirms the asymptotic slope of -4	122

CHAPTER 1

INTRODUCTION

Over the past several years, wireless telecommunications have significantly changed many aspects of our world. Just ten years ago, mobile telecommunication services were, for the most part, dedicated to voice communication. Today, however, we rely on them not only for voice communication, but also for data communication. Consider, for example, email and information retrieval services, as well as the many special-purpose services for entertainment, e-commerce, education, and health-care. Former U.S. vice president Al Gore credited “the emergence of mobile platforms as the catalyst behind the new communications revolution” [2]. Furthermore, he identified the wireless industry as “a bright spot in an otherwise grim global economy” [2]. Therefore, it is not an overstatement to say that research on wireless telecommunication system-design is of critical importance to our nation and to the world, both now and in the foreseeable future.

With the wireless revolution comes an ever-growing demand for higher data-rate and higher mobility. Wireless users want access to everything, everywhere, and anytime. Thus, wireless devices must send and receive high-rate information streams through *quickly time-varying channels*. (Here and in the sequel, we use “channel” to denote the signal propagation medium.) Roughly speaking, signal reflections from

structures like buildings, mountains, and vehicles cause *selectivity* in the receive signal. (See Fig. 1.1.) Reflections caused by physical objects between the transmitter and the receiver create dispersion in time and these reflected signals are sometimes summed in a destructive way, causing so called multipath fading. In this case, energy from each symbol disperses and spills out to the adjacent symbols, which we call inter-symbol-interference (ISI). This channel behaves as a filter, whose frequency response has frequency selectivity. Thus, it is also called *frequency selective* channel. When the transmitter, the reflector, and/or the receiver are in motion, the reflected signals travel through different channels at each time instance. Thus, the received signal undergoes time-varying, i.e., time-selective, channel effect. In the frequency domain, the received signal's spectrum suffers dispersion from a phenomenon called Doppler spreading. Wireless channels that are both time and frequency selective are called doubly selective (DS).

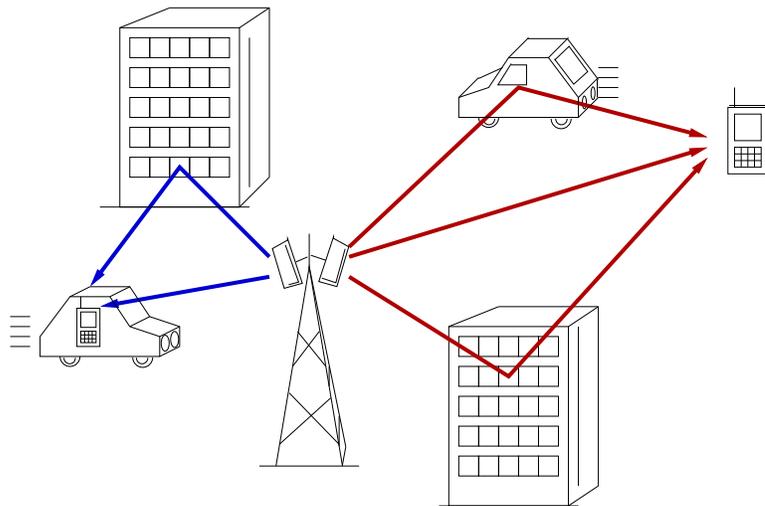


Figure 1.1: Time-varying channel occurs by mobility of transmitter, reflector and/or receiver.

1.1 Motivation

In conventional wireless telecommunication systems, the principal effect of a DS channel is a reduction in data rate. For example, 3rd Generation (3G) cellular service can provide a rate of 384 kbit/sec at or below pedestrian speeds, but only 128 kbit/sec in a moving car [3]. The reduction in data rate results primarily from the channel's increasing degrees of variation as indicated by Zheng and Tse in [4]. In practice, the DS channel's high degrees of variation yields difficulties in estimation of the channel parameters. Compared to the singly selective channels, DS channel offers much more channel coefficients to be estimated—involving hundreds or thousands of parameters, since in each time instant the channel coefficients are different and to be tracked. Therefore, our first goal is to efficiently model the DS channel with as small number of parameters as possible.

Even if the channel coefficients were perfectly estimated, the mitigation of the channel effect, i.e., *equalization* of DS channel itself can be computationally intense. One possible equalization method is to extend the efficient equalization for the frequency selective channels to the DS channels. It is well known that one-tap equalizer for cyclic-prefixed (CP)-orthogonal frequency division multiplexing (OFDM) is simplest and effective for the frequency selective channels. Thus, many multi-carrier modulation (MCM) schemes modified from the CP-OFDM have been proposed. For example, one can intentionally distort transmit signals (e.g., by pulse-shaping) so that effective frequency response of the channel has a certain favorable tendency, e.g., small number of non-zero taps in the frequency response.

Not only the efficient equalization and the channel estimation at the receiver, but also the effective transmission of the signal plays an important role in improving

the error performance of the communication system. For example, if the channel is randomly time-varying, one can spread transmitted signal on a long time period and combine multiple received signals at the receiver to increase chances of correct detection of the signal. Interestingly, it is known that such strategy for highly varying channels facilitates better performance than slowly varying channel, as long as the channel coefficients are correctly tracked. Thus, we focus on characterization of the DS channel’s degrees of variation and design of transmission scheme that guarantees the maximum exploitation of channel’s degrees of variation.

1.2 Contributions and Outline

In this dissertation, we propose

- practical, yet near-optimal implementations of DS channel equalizers for each *coherent* receiver and *noncoherent* receiver,¹ which adopt generic tree search algorithms (TSAs) and efficient channel modeling via basis expansion (BE).
- a novel channel adaptive TSA that shows near-optimal performance with low complexity,

and we establish

- *maximum-diversity* conditions to maximally exploit the DS channel’s degrees of variation for each coherent receiver and noncoherent receiver and the fact that *almost any* random affine (linear) precoding facilitates maximum-diversity noncoherent (coherent) detection.

¹We will define “coherent” and “noncoherent” receivers in Chapter 2.

More specifically, in Chapter 2, we present an efficient modeling of DS channel via BE, which leads to a unified system model. With some proper modulation and demodulation schemes, the unified system model is expressed as a cyclic banded channel matrix. In particular, we adopt PS-MCM by Schniter and Das in [5] to make the frequency-domain channel matrix banded.

In Chapter 3, generic preprocessing and search strategies for TSAs are discussed and their application to PS-MCM is presented. Then, a new finding about the optimality of minimum mean-squared error (MMSE)- general decision feedback equalizer (GDFE) preprocessing for TSA is established. In particular, the MMSE-GDFE preprocessing has been previously known as a sub-optimal preprocessing in the sense of preservation of the distance between the constellation points, thereby degrading the performance of the minimum-distance detector. When applied to the constant-modulus constellation, however, it preserves the distance so that it facilitates the optimal minimum-distance detection.

In Chapter 4, optimal coherent equalization is discussed for PS-MCM application. Then, a novel low-complexity MMSE-GDFE preprocessing and a novel low-complexity channel-adaptive TSA algorithm are proposed, which together achieve a near-optimal performance. Also, a low complexity channel estimation algorithm based on BE is presented.

In Chapter 5, optimal noncoherent equalization based on a BE modeling and a MAP noncoherent metric is discussed. Then, a near-optimal implementation via a novel sequential algorithm is presented, which embeds generic TSA to search for best sequences over the noncoherent MAP metric. Also, a low-complexity noncoherent

equalizer based on Bayesian expectation maximization (EM) is proposed, which iterates between the coherent TSA proposed in Chapter 4 and a BE channel estimator.

In Chapter 6, the sequential noncoherent equalization scheme proposed in Chapter 5 is applied to the underwater acoustic channel (UAC), which is characterized by a sparse but long spread in impulse response. The sparsity is modeled and exploited to further reduce the complexity of the equalization.

In Chapter 7, we characterize a maximum-diversity conditions for each coherent detector and noncoherent detector by using BE channel modeling. Then, we establish the fact that almost any random affine (linear) precoding facilitates maximum-diversity reception for noncoherent (coherent) detector under mild channel condition.

In Chapter 8, we conclude by summarizing our work and providing future research possibilities.

1.3 Notation

We use $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ to denote conjugate, transpose and Hermitian transpose, respectively. We write the k^{th} entry of vector \mathbf{x} as $[\mathbf{x}]_k$, and the $(k, l)^{th}$ entry of matrix \mathbf{A} as $[\mathbf{A}]_{k,l}$. We denote the determinant by $\det(\cdot)$, and the null space of matrix \mathbf{A} by $\mathcal{N}(\mathbf{A})$. We denote $M \times 1$ zero-valued column vector by $\mathbf{0}_M$, and the $M \times N$ zero-valued matrix by $\mathbf{0}_{M \times N}$. The $N \times N$ identity matrix is denoted by \mathbf{I}_N , and the proper complex Gaussian distribution with mean vector \mathbf{m} and covariance matrix \mathbf{C} is denoted by $\mathcal{CN}(\mathbf{m}, \mathbf{C})$. For vector norms, we use $\|\mathbf{x}\|^2 \triangleq \mathbf{x}^H \mathbf{x}$ and $\|\mathbf{x}\|_{\mathbf{A}}^2 \triangleq \mathbf{x}^H \mathbf{A} \mathbf{x}$, where \mathbf{A} is positive semi-definite. We use \odot to denote the element-wise product of matrices, $\mathcal{D}_d(\mathbf{x})$ to denote the diagonal matrix constructed from the d^{th} cyclic downshift of vector \mathbf{x} , and $\mathcal{D}(\mathbf{x})$ is shorthand for $\mathcal{D}_0(\mathbf{x})$. Expectation is denoted by $\mathbb{E}\{\cdot\}$,

the Kronecker delta by δ_l , and the modulo- N operation by $\langle \cdot \rangle_N$. Real part of complex number x is denoted by $\Re\{x\}$ and imaginary part $\Im\{x\}$. We use \mathbb{R} to denote the real field, \mathbb{C} the complex field, and \mathbb{Z} the integers. We abbreviate “with probability one” as “w.p.1”.

We summarize in Table 1.1 the abbreviations used in the dissertation.

Table 1.1: List of abbreviations

AWGN	Additive White Gaussian Noise
BE	Basis Expansion
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CP	Cyclic Prefix
CWGN	Circular White Gaussian Noise
DFT	Discrete Fourier Transform
DFE	Decision Feedback Equalizer
DS	Doubly Selective
DPP	Delay-Power Profile
EM	Expectation Maximization
FER	Frame Error Rate
i.i.d.	independent and identically distributed
ICI	Inter-Carrier Interference
ISI	Inter-Symbol Interference
LLR	Log Likelihood Ratio
LMMSE	Linear Minimum Mean-Squared Error
MAC	Multiplication and Accumulation
MAP	Maximum a Posteriori
MCM	Multi-Carrier Modulation
MD	Minimum Distance
MIMO	Multi Input Multi Output
ML	Maximum Likelihood
MMSE	Minimum Mean-Squared Error
OFDM	Orthogonal Frequency Division Multiplexing
PAM	Pulse Amplitude Modulation
PS	Pulse-Shaped
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
SD	Sequence Detection
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SpD	Sphere Decoder
SVD	Singular Value Decomposition
TSA	Tree Search Algorithm
ZF	Zero Forcing
ZP	Zero-Padded

CHAPTER 2

PROBLEM STATEMENT AND SYSTEM MODEL

As discussed in Chapter 1, to reliably recover the transmitted information corrupted by the DS channel and thus to increase the throughput of the system, a judicious procedure to undo the effect of DS channel, i.e., equalization, needs to be carried out first. Generally speaking, there are two designs of equalizers, i.e., coherent equalizers and noncoherent equalizers. *Coherent* equalizers are assumed to be informed of the channel state parameters explicitly. The typical examples of the channel state parameters are time-domain channel impulse response coefficients or frequency domain inter-carrier interference coefficients. In order for the coherent equalizer to operate, the channel state information should be fully provided by the channel estimator. *Noncoherent* equalizers, however, perform the task without explicit state knowledge of the channel but with either channel statistics, e.g., mean and correlation, or channel parameters, e.g., maximum delay of the impulse response. Note that training-based, blind, and semi-blind schemes all fall under the category of noncoherent equalization. Similarly, the term “joint channel/symbol estimation” sometimes refers to noncoherent equalization, even though explicit channel estimates are not strictly needed for data decoding.

2.1 Coherent Equalization for Multi-Carrier Systems

In what we will refer to as the multicarrier modulation (MCM) approach [6], linear modulation waveforms are designed to yield a “simple” interference response—in order to ease the equalization task—without explicitly considering the achievable FER performance. The vast majority of DS-channel communication schemes fit into this category, e.g., cyclic-prefix (CP) orthogonal frequency-division multiplexing (OFDM) [7], zero-padded (ZP) OFDM [8], and Strohmer and Beaver’s “optimal” OFDM [9]. For example, CP-OFDM and ZP-OFDM were originally designed for time-selective—rather than doubly selective—channels, and are capable of totally suppressing inter-symbol interference (ISI). When used in DS channels, however, CP-OFDM and ZP-OFDM succumb to significant inter-carrier interference (ICI) which greatly complicates the optimal (i.e., maximum likelihood) equalization. In response, more sophisticated MCM schemes have been proposed based on smooth ISI/ICI-minimizing pulses. Though these “pulse-shaped” MCM schemes succumb to less ICI than their ZP-OFDM and CP-OFDM counterparts, their ISI/ICI responses are, in general, still too complicated for practical maximum-likelihood (ML) sequence estimation based equalizers.

Due to the impracticality of optimal equalizer in DS-channel MCM, several methods of reduced-complexity reception have been proposed. These schemes are typically based on the combination of *ISI/ICI truncation* with *suboptimal sequence estimation*. By ISI/ICI truncation, we mean that only the “significant” ICI/ISI coefficients are estimated at the receiver and used for the sequence estimation. Examples of suboptimal sequence estimation include linear detection (e.g., [10–12]), DF detection

(e.g., [13–15]), iterative/turbo detection (e.g., [5, 16, 17]), and approximate-ML detection (e.g., [1, 18–20]). We identify that the pulse-shaped (PS)-MCM by Schniter and Das [5, 16] is suitable for the practical near-optimal coherent equalization in that PS-MCM near-perfectly suppresses all but a small number of ISI/ICI coefficients. Then we conclude that judicious design of a DS channel communication system comes down to

- designing a near-optimal (i.e., near-ML) equalization algorithm which leverages the structure of significant-ISI/ICI for complexity reduction.

In Chapter 4 we’ll present a near-optimal coherent equalization algorithm based on a novel tree search algorithm (TSA) tailored to the PS-MCM applications.

2.2 Noncoherent Equalization

For the noncoherent equalization, we consider the problem of decoding a data sequence transmitted over an *unknown* or *partially known* DS channel, such as a time-varying ISI channel or a frequency-varying ICI channel, whose statistics are known.

Optimal noncoherent equalization requires evaluating a noncoherent metric for every possible bit sequence and then summing over subsets of these metrics [21]. Since the number of possible bit sequences is impractically large, three suboptimal approaches have emerged.

1. If the channel was known, a trellis could be used to efficiently calculate bit posteriors via the forward-backward (i.e., BCJR) algorithm [22]. With an unknown Gauss-Markov channel, forward-backward processing can be applied after expanding the trellis to allow conditional channel estimation at each state,

although not optimally: the performance (and complexity) increase with trellis expansion [23–26]. As an alternative, “fixed-lag” posteriors—which use a limited number of future observations—could be computed using a forward trellis with per-survivor channel estimates [21, 25–27]. However, fixed-lag posteriors are suboptimal for turbo reception. In either case, pilots are critical to ensure good performance with low complexity.

2. In the second approach, the expectation-maximization (EM) approach [28] is applied, resulting in the (iterative) computation of a *single* channel estimate per data block [29–32]. Each EM iteration includes two steps: soft coherent maximum *a posteriori* (MAP) equalization and soft channel estimation. With a sufficiently good initialization (e.g., from pilots), the EM approach will yield the maximum likelihood channel estimate (and the corresponding soft symbol estimates), imparting a sense of optimality.
3. Other approaches generally fall into the third category: a soft coherent equalizer (e.g., [33–36]) is combined with a soft channel estimator (e.g., [37–39]) without attention to *joint* optimality. Recent examples of this “ad hoc” approach include [39, 40].

A potential drawback of near-optimal trellis-based approaches is the need for $2^{q(N_H-1)}$ states, where q is the number of bits/symbol and N_H is the channel spread, or more if the trellis is expanded; for channels with moderate-to-long spreads, the number of states becomes impractical. A potential drawback of near-optimal approaches based on Gauss-Markov channel models is that, while they can efficiently model the time-domain channel trajectories encountered in single-carrier systems,

they do not efficiently model the frequency-domain channel trajectories encountered in multi-carrier systems, where channel variation is best parameterized using a complex exponential basis expansion [41]. Thus, our goals are stated as follows.

1. To provide an efficient and unified modeling of the variation (in time or frequency) of the channel for both single-carrier and multi-carrier systems.
2. To establish optimal (e.g. ML or MAP) noncoherent equalization methods, and then to provide practical algorithms to accomplish a near-optimal noncoherent equalization.

The problems are possibly addressed via adoption of BE and via two new noncoherent equalization schemes. As we will see in Section 2.4.4, generic BE modeling can provide an efficient modeling of the DS channel. Two equalization schemes, i.e., a novel noncoherent metric based sequential equalization and an expectation maximization (EM)-based iterative equalization, will be discussed in Chapter 5.

2.3 Maximum-Diversity Precoding

We are especially interested in the high-SNR regime, where the performance is strongly dependent on the *diversity* order, i.e., the negative slope of the log-error-rate versus log-SNR curve.

For the case where the receiver has channel state information (CSI) and that the channel follows a complex-exponential basis expansion model (CE-BEM), Ma and Giannakis [42] characterized the maximum achievable diversity order and proposed a linear precoding scheme that facilitates maximum-diversity reception. The assumptions of perfect receiver CSI and a CE-BEM channel are quite restrictive, however,

limiting the practical impact of [42]. For example, CSI is not easy to acquire and maintain in the doubly dispersive case, where channel parameters can be multitudinous and quickly varying.

In response, we consider the more difficult but more practical problem of *noncoherent communication* over the DS channel, where neither the transmitter nor the receiver is assumed to have CSI. In this case, the receiver must exploit (a priori known) structure in the transmitted signal in order to decode reliably in the presence of channel uncertainty.

For noncoherent communication over the DS channel, there exists a large body of work on optimal and suboptimal noncoherent reception strategies (e.g., [43–53]). For this case, there also exists several articles on training sequence design (e.g., [54–57]) with the aim of improving explicit channel estimates. But we are not aware of work addressing the general problem of transmitter design (i.e., joint design of data and training sequences) to improve the reliability of communication over the noncoherent DS channel.

In response,

1. we characterize the maximum achievable diversity order for noncoherent (or coherent) communication over the DS channel, and
2. pursue precoding schemes to facilitate the maximum-diversity decoding at the receiver.

2.4 System Model

At the transmitter, we assume that information bits $\{b_m^{(j)}\}$, are rate- R coded, interleaved, and mapped to 2^q -ary QAM symbols, where $j \in \{0, \dots, J-1\}$ is the

transmit block index. Groups of N_s information symbols are then combined with pilot and guard symbols to form symbol blocks of length $N \geq N_s$. We denote the j^{th} symbol block by

$$\mathbf{s}^{(j)} = [s_0^{(j)}, \dots, s_{N-1}^{(j)}]^T, \quad (2.1)$$

where $s_n^{(j)} \in \mathbb{S}$ for symbol alphabet \mathbb{S} , and the corresponding coded bit vector by

$$\mathbf{x}^{(j)} = [x_0^{(j)}, \dots, x_{N_s q-1}^{(j)}]^T, \quad (2.2)$$

where $x_k^{(j)} \in \{0, 1\}$. The symbols are then linearly block-modulated by either a single-carrier scheme or a multi-carrier scheme, represented by $\mathbf{G} \in \mathbb{C}^{N_t \times N}$ with $N_t \geq N$, to form the transmitted signal

$$\mathbf{t}^{(j)} \triangleq [t_0^{(j)}, \dots, t_{N_t-1}^{(j)}]^T \quad (2.3)$$

The we write $\mathbf{t}^{(j)} = \mathbf{G}\mathbf{s}^{(j)}$. The construction of \mathbf{G} will be described later.

At the channel output, the samples in the j^{th} received block

$$\mathbf{r}^{(j)} \triangleq [r_0^{(j)}, \dots, r_{N_r-1}^{(j)}]^T \quad (2.4)$$

are assumed to take the form

$$r_n^{(j)} = \sum_{l=0}^{N_h-1} h_{n,l}^{(j)} t_{n-l}^{(j)} + \nu_n^{(j)}, \quad (2.5)$$

where $h_{n,l}^{(j)}$ is the time- n response of the channel to an impulse applied at time- $(n-l)$, where N_h is the discrete channel delay spread, and where $\{\nu_n^{(j)}\}$ is zero-mean circular white Gaussian noise (CWGN) with covariance σ^2 .

The received vector $\mathbf{r}^{(j)}$ is then linearly (single- or multi-carrier) demodulated via matrix $\mathbf{\Gamma} \in \mathbb{C}^{N \times N_r}$ to yield

$$\mathbf{y}^{(j)} = \mathbf{H}^{(j)} \mathbf{s}^{(j)} + \mathbf{w}^{(j)}, \quad (2.6)$$

where we define $\mathbf{y}^{(j)} = \mathbf{\Gamma}\mathbf{r}^{(j)}$ and

$$\mathbf{H}^{(j)} \triangleq \mathbf{\Gamma}\mathcal{H}^{(j)}\mathbf{G}. \quad (2.7)$$

In (2.6) and (2.7), $\mathbf{w}^{(j)} = \mathbf{\Gamma}\mathbf{v}^{(j)}$ and $\mathcal{H}^{(j)} \in \mathbb{C}^{N_r \times N_t}$ is a convolution matrix constructed from the channel's time-varying impulse response with $[\mathcal{H}^{(j)}]_{n,n-l} = h_{n,l}^{(j)}$. Thus we have $N_r = N_t + N_h - 1$ and $\mathcal{H}^{(j)}$ is banded with bandwidth N_h . Note that $\mathbf{H}^{(j)}$ represents the composite effect of modulation, channel propagation, and demodulation. When either single- or multi-carrier scheme is appropriately designed, $\mathbf{H}^{(j)}$ can be closely approximated by a ‘‘circularly banded’’ matrix [5] with bandwidth N_H , as illustrated in Fig. 5.2(a).

2.4.1 Single-Carrier Modulation/Demodulation

Cyclic-prefixed (CP) single carrier modulation uses the choices of

$$\mathbf{\Gamma} = \mathbf{I}_N \quad (2.8)$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{N_{\text{cp}}} \\ \mathbf{I}_{N-N_{\text{cp}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_{\text{cp}}} \end{bmatrix}. \quad (2.9)$$

The guard length and the CP length is typically set to $N_H - 1$. Thus $\mathbf{H}^{(j)}$, with bandwidth $N_H = N_h$, contains the impulse response coefficients $\{h_{n,l}^{(j)}\}$. Sometimes, the bandwidth N_H , or delay spread, can be very large depending on the channel environment, under which the system operates. For example, delay spread $\tau_{\text{max}} = 20\mu\text{s}$ and sampling frequency $1/T_s = 10$ MHz yield $N_H \triangleq \lceil \tau_{\text{max}}/T_s \rceil = 200$.

2.4.2 Multi-Carrier Modulation/Demodulation

In the multi-carrier case, the modulation is described by

$$\mathbf{G} = \mathcal{D}(\mathbf{g})\mathbf{F}_t^H, \quad (2.10)$$

where $\mathbf{F}_t^H \in \mathbb{C}^{N_t \times N}$ is a period- N unitary inverse discrete Fourier transform (DFT) matrix cyclically extended in the row dimension, and where $\mathcal{D}(\mathbf{g})$ is a diagonal matrix created from a time-domain transmission pulse $\mathbf{g} \in \mathbb{C}^{N_t}$. Then, the demodulation is described by

$$\mathbf{\Gamma} = \mathbf{F}_r \mathcal{D}(\boldsymbol{\gamma} \odot \mathbf{m}), \quad (2.11)$$

where $\mathbf{F}_r \in \mathbb{C}^{N \times N_r}$ is a period- N unitary DFT matrix cyclically extended in the column dimension, $\boldsymbol{\gamma} \in \mathbb{C}^{N_r}$ is a time-domain reception pulse, and $[\mathbf{m}]_n = \exp(j \frac{2\pi}{N} \frac{N_D - 1}{2} n)$. With appropriate design of \mathbf{g} and $\boldsymbol{\gamma}$ [5], the frequency-domain channel matrix $\mathbf{H}^{(j)}$ has bandwidth

$$N_H = N_D \triangleq \lceil 2f_D T_s N \rceil + \alpha, \quad (2.12)$$

where f_D denotes the single-sided Doppler spread (in Hz), T_s denotes the channel-use interval (in sec), and α is a (small) non-negative integer that controls out-of-band coefficient energy. We also define ICI radius as

$$D \triangleq \frac{N_H - 1}{2} \quad (2.13)$$

so that $N_H = 2D + 1$. The off-diagonal elements of $\mathbf{H}^{(j)}$ induce ICI. We assume the last $N_H - 1$ symbols in $\mathbf{s}^{(j)}$ are zero-valued guards, so that $\mathbf{H}^{(j)}$ acts causally on the first $N - N_H + 1$ symbols.

2.4.3 Real-Valued System Model for Tree Search

When we discuss about TSAs in the rest part of the dissertation, particularly in Chapter 3 and Chapter 4, it is sometimes convenient to adopt real-valued system model, since many TSAs are operated only for real-valued systems. One way to

construct real-valued model given complex-valued equation (2.6) is splitting each complex element in vectors $\mathbf{y}^{(j)}$, $\mathbf{s}^{(j)}$ and $\mathbf{w}^{(j)}$ into two real elements and stacking them in an alternating way. In other words, real-valued vector $\bar{\mathbf{y}}^{(j)} \in \mathbb{R}^{2N}$ is constructed so that $[\bar{\mathbf{y}}^{(j)}]_{2n} = \Re\{y_n^{(j)}\}$ and $[\bar{\mathbf{y}}^{(j)}]_{2n+1} = \Im\{y_n^{(j)}\}$ for $0 \leq n < N$. In the similar manner, the real-valued symbol vector $\bar{\mathbf{s}}^{(j)}$ is constructed via $[\bar{\mathbf{s}}^{(j)}]_{2n} = \Re\{s_n^{(j)}\} \triangleq \bar{s}_{2n}^{(j)}$ and $[\bar{\mathbf{s}}^{(j)}]_{2n+1} = \Im\{s_n^{(j)}\} \triangleq \bar{s}_{2n+1}^{(j)}$. With setting $Q^2 = 2^q$, we have Q -ary PAM alphabet $\bar{\mathbb{S}}$ for $\bar{s}_n^{(j)}$ for this real-valued model, instead of the 2^q -ary (i.e., Q^2 -ary) QAM alphabet \mathbb{S} for $s_n^{(j)}$. The channel matrix $\bar{\mathbf{H}}^{(j)} \in \mathbb{R}^{2N \times 2N}$ are built by alternatively placing real and imaginary elements of $\mathbf{H}^{(j)}$, e.g.,

$$\begin{aligned} \bar{\mathbf{H}}^{(j)} &= \begin{bmatrix} \Re\{[\mathbf{H}^{(j)}]_{0,0}\} & -\Im\{[\mathbf{H}^{(j)}]_{0,0}\} & \cdots \\ \Im\{[\mathbf{H}^{(j)}]_{0,0}\} & \Re\{[\mathbf{H}^{(j)}]_{0,0}\} & \\ \vdots & & \ddots \end{bmatrix} \\ &\in \mathbb{R}^{2N \times 2N}. \end{aligned} \quad (2.14)$$

Resulting real-valued system equation

$$\bar{\mathbf{y}}^{(j)} = \bar{\mathbf{H}}^{(j)} \bar{\mathbf{s}}^{(j)} + \bar{\mathbf{w}}^{(j)} \quad (2.15)$$

is used when we discuss preprocessing of TSA and variety of TSAs adopted to equalization in Chapter 3 and Chapter 4.

2.4.4 Basis Expansion Modeling of DS Channel

We assume that the channel is Rayleigh fading and wide-sense stationary (WSS). And the channel's time variation is assumed to follow the "Jakes' channel model," i.e., $\mathbb{E}\{h_{n,l} h_{n+m,l}^*\} = \sigma_l^2 J_0(2\pi f_D T_s m)$, where $J_0(\cdot)$ denotes the zeroth-order Bessel function of the first kind, f_D denotes the single-sided Doppler spread in Hz, T_s denotes the channel-use interval in seconds and σ_l^2 denotes the delay-power profile for the l^{th} channel tap. We also assume the channel retains "uncorrelated scattering" property

across the tap locations, i.e., $\mathbb{E}\{h_{n,l}h_{n+m,l+\ell}^*\} = 0$, for $\ell \neq 0$. Random channel vector defined by the N -sample trajectory of the l^{th} channel tap can be written

$$\mathbf{h}_l^{(j)} \triangleq [h_{0,l}^{(j)}, h_{1,l}^{(j)}, \dots, h_{N-1,l}^{(j)}]^T \quad (2.16)$$

$$= [[\mathbf{H}^{(j)}]_{0,-l}, [\mathbf{H}^{(j)}]_{1,1-l}, \dots, [\mathbf{H}^{(j)}]_{N-1,N-1-l}]^T, \quad (2.17)$$

which corresponds to the l^{th} sub-diagonal of $\mathbf{H}^{(j)}$. Now, it can be expressed (without loss of generality) using a basis matrix $\mathbf{B}^{(j)} \in \mathbb{C}^{N \times N_b}$ as

$$\mathbf{h}_l^{(j)} = \mathbf{B}^{(j)} \boldsymbol{\theta}_l^{(j)}, \quad (2.18)$$

where $\boldsymbol{\theta}_l^{(j)} \in \mathbb{C}^{N_b}$ is a zero-mean circular Gaussian random vector. Examples of BEs which do not require statistical channel knowledge include the polynomial BE model [58], [59]:

$$[\mathbf{B}^{(j)}]_{m,k} = (\sqrt{N_b})^{-1} \left(m - \frac{N_b - 1}{2} \right)^k, \quad (2.19)$$

and oversampled complex exponential BE model with oversampling factor M [60], [61]:

$$[\mathbf{B}^{(j)}]_{m,k} = (\sqrt{N_b})^{-1} e^{j \frac{2\pi}{MN_b} (k - \frac{N-1}{2}) m}. \quad (2.20)$$

BE models which require statistical knowledge include the Slepian BE model [62] and the Karhunen-Lóeve (KL) BE model [63].

The BE channel parameter is collected as a single vector, i.e.,

$$\boldsymbol{\theta}^{(j)} \triangleq [\boldsymbol{\theta}_0^{(j)T}, \dots, \boldsymbol{\theta}_{N_H-1}^{(j)T}]^T \quad (2.21)$$

and thus vector is distributed as $\boldsymbol{\theta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\theta)$, where \mathbf{R}_θ has full rank $N_H N_b$ and its structure depends on the choice of the basis. For example, the covariance matrix

of $\mathbf{h}_l^{(j)}$ can be expressed with KL basis via eigen decomposition and a reduced-rank approximation:

$$\mathbb{E}\{\mathbf{h}_l^{(j)}\mathbf{h}_l^{(j)H}\} = \mathbf{V}^{(j)}\mathbf{\Lambda}_l^{(j)}\mathbf{V}^{(j)H} \quad (2.22)$$

$$\approx \tilde{\mathbf{V}}^{(j)}\tilde{\mathbf{\Lambda}}_l^{(j)}\tilde{\mathbf{V}}^{(j)H}, \quad (2.23)$$

where the rank reduction to N_b eigen-modes is used. This example produces a KL BE with $\mathbf{B}^{(j)} = \tilde{\mathbf{V}}^{(j)}$, $\mathbf{R}_{\theta_l, \theta_l}^{(j)} = \tilde{\mathbf{\Lambda}}_l^{(j)}$ and

$$\mathbf{R}_{\theta}^{(j)} = \begin{bmatrix} \tilde{\mathbf{\Lambda}}_0^{(j)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \tilde{\mathbf{\Lambda}}_{N_H-1}^{(j)} \end{bmatrix}. \quad (2.24)$$

The parameter $N_b \leq N$ quantifies the degrees of variation of the channel tap. In cases of practical interest, the channel varies slowly enough that $N_b \ll N$. For evidence of this claim, Fig. 2.1 plots the effective² degrees of variation for the commonly assumed Rayleigh channel.

For multi-carrier system, N -sample trajectory of the d^{th} frequency-domain (and frequency-varying) channel tap can be defined as the same way as in (2.17) but using frequency domain channel matrix. In this case, we find it convenient to expand $\mathbf{h}_d^{(j)}$ with truncated Fourier basis, e.g.,

$$\mathbf{h}_d^{(j)} = \mathbf{F}^{(j)}\bar{\boldsymbol{\theta}}_d^{(j)} \quad (2.25)$$

$$\approx \mathbf{B}^{(j)}\boldsymbol{\theta}_d^{(j)}, \quad (2.26)$$

where $\mathbf{F}^{(j)} \in \mathbb{C}^{N \times N}$ is the DFT matrix and $\mathbf{B}^{(j)} \in \mathbb{C}^{N \times N_b}$ has N_b columns selected from those of $\mathbf{F}^{(j)}$. Note that when using proper pulse-shaping schemes the approximation can be made very close. A more detailed description of MCM channel

²We define the ‘‘effective degrees of variation’’ as the number of eigenvalues in $\mathbb{E}\{\mathbf{h}_\ell\mathbf{h}_\ell^H\}$ which are larger than 1/1000 of the principle eigenvalue.

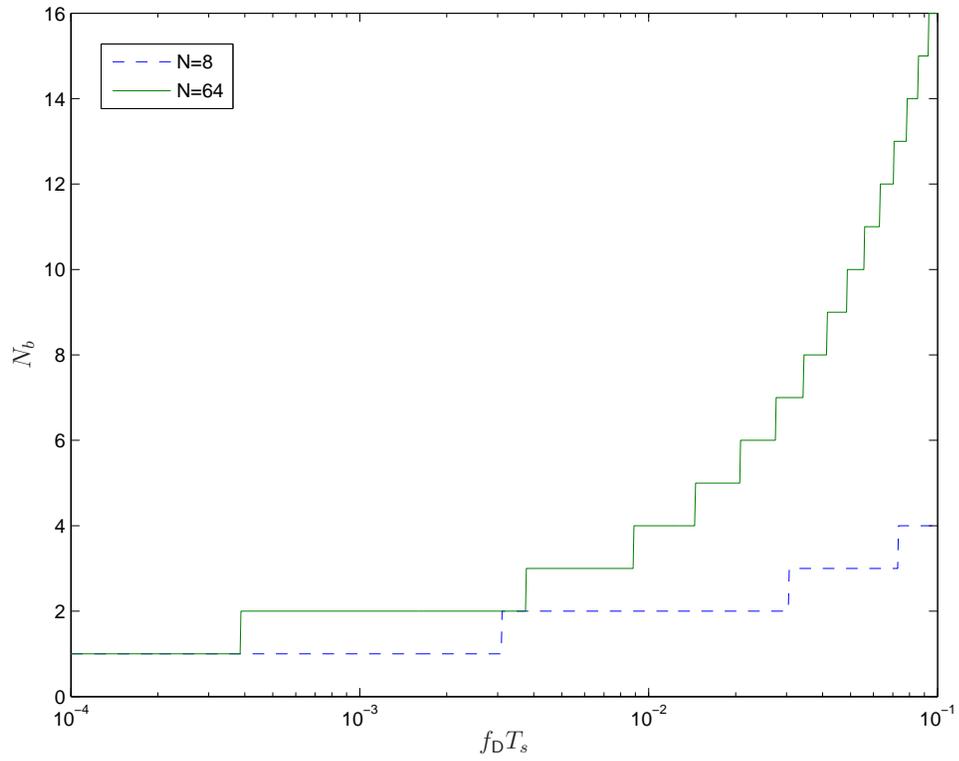


Figure 2.1: Effective degrees of variation N_b versus normalized single-sided Doppler spread $f_D T_s$ for Jakes' channel at different block lengths N .

coefficients and its Fourier basis representation based on sparse channel assumption is provided in Appendix A.

CHAPTER 3

TREE SEARCH ALGORITHMS

TSAs can be employed in many coherent equalization problems as detection or equalization algorithms, i.e., TSAs can perform maximum likelihood sequence detection (MLSD), or TSAs can facilitate a soft maximum *a posteriori* (MAP) equalization by generating a set of best sequences. Also, TSAs can be adopted the same way for the noncoherent equalization problems, as we will see in Chapter 5. In this chapter, we review generic TSAs, by focusing on preprocessing algorithms and several search strategies used for algorithms, especially in the context of MCM application. Then, we discuss the optimality of the minimum mean-squared error (MMSE)-general decision feedback equalizer (GDFE) preprocessing. More specifically, we establish that under some condition the MMSE-GDFE preprocessing does not compromise the optimal search performance.

The MCM described at the end of Section 2.4.2 allows us to focus on a system model (almost) free of ISI and insignificant-ICI. Suppressing the ^(j) notation and adopting the real-valued model in (2.15), the MCM system equation (2.6) becomes

$$\bar{\mathbf{y}} = \bar{\mathbf{H}}\bar{\mathbf{s}} + \bar{\mathbf{w}}, \quad (3.1)$$

where $\bar{\mathbf{H}}$ retains the quasi-banded structure in Fig. 3.1(a). We assume that the symbol is chosen from a Q -ary PAM alphabet written as

$$\bar{\mathcal{S}} = \left\{ -\frac{Q-1}{2}, \dots, \frac{Q-1}{2} \right\}. \quad (3.2)$$

Since (3.1) involves $2N$ -dimensional real-valued vectors, we define $L \triangleq 2N$. The MLSD solution to (3.1) under known \mathbf{H} can be expressed as a minimum distance (MD) search over the signal alphabet:

$$\hat{\mathbf{s}}_{\text{ML}} = \arg \min_{\bar{\mathbf{s}} \in \bar{\mathcal{S}}^L} \left\| \bar{\mathbf{y}} - \bar{\mathbf{H}} \bar{\mathbf{s}} \right\|^2. \quad (3.3)$$

The brute-force approach to finding $\hat{\mathbf{s}}_{\text{ML}}$ requires $\mathcal{O}(Q^L)$ operations, which is impractical for large N . If $\bar{\mathbf{H}}$ was banded with a band radius of D , then the Viterbi algorithm could be used to solve (3.3) with a complexity of $L(2D+1)Q^{2D+1}$ real multiply-accumulate (MAC) operations per block [1]. Since $\bar{\mathbf{H}}$ is only quasi-banded, a different approach is needed. For example, one could instead use a “tail-biting” MLSD which hypothesizes an initial state at an arbitrary location within the block, runs the standard Viterbi algorithm from that state, and forces a termination back to that state. Exhaustively searching among the Q^{2D} possible hypotheses yields an MLSD algorithm with a complexity of $L(2D+1)Q^{4D+1}$ real MACs per frame. However, these Viterbi algorithms, while much cheaper than brute force search, will still be impractical in many applications.

MD search algorithms present an alternative to brute-force and Viterbi MLSD [64]. After converting the linear system (3.1) to upper triangular form, efficient MD search algorithms based on sequential decoding [65, 66] or sphere decoding [67, 68] can be used to implement MLSD with an average complexity far below $\mathcal{O}(Q^L)$. Since the sequential decoding and the sphere decoding are closely related (see, e.g., [69]) in that

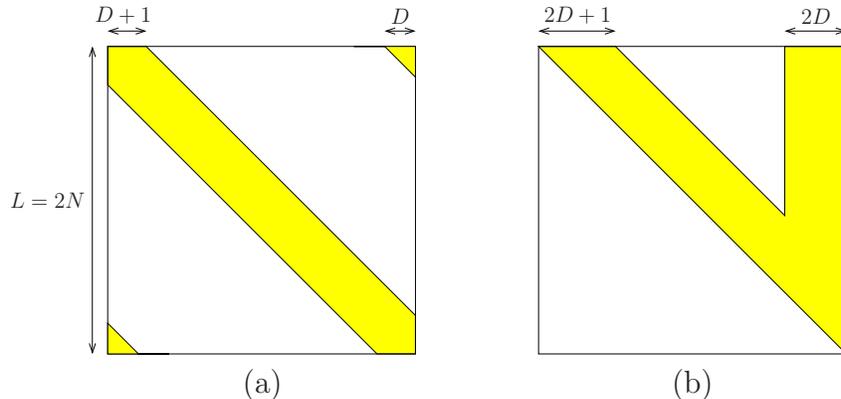


Figure 3.1: Channel matrices associated with MCM: (a) “quasi-banded” channel matrix, (b) “V-shaped” channel matrix.

they explore *tree-shaped* data structure to find the optimal point, we refer to them collectively as TSAs. For the system (3.1) with general (i.e., non-banded) channel matrix $\bar{\mathbf{H}}$, for example, sphere decoding maintains an average complexity of approximately $\mathcal{O}(L^3)$ at high SNR, regardless of constellation size Q [69]. This remarkable fact encourages a more thorough investigation of TSAs capable of leveraging the quasi-banded structure of $\bar{\mathbf{H}}$ for further complexity reduction. In fact, we will show in Section 4.2.2 that quasi-banded $\bar{\mathbf{H}}$ allows near-ML TSA with an average complexity close to $\mathcal{O}(L^2)$. TSA consists of a preprocessing step and a tree search step; both are discussed next.

3.1 Preprocessing

We refer to “TSA preprocessing” as that which converts the linear system (3.1) to upper triangular form. The traditional TSA preprocessing method uses the QR decomposition $\bar{\mathbf{H}} = \mathbf{Q}\mathbf{R}$ to transform (3.1) into the equivalent system $\mathbf{y}' = \mathbf{Q}^T \bar{\mathbf{y}} = \mathbf{R}\bar{\mathbf{s}} + \mathbf{w}'$, where \mathbf{R} is upper triangular and \mathbf{w}' is statistically equivalent to $\bar{\mathbf{w}}$. This is

called zero-forcing (ZF)-DFE preprocessing and with ZF-DFE preprocessing optimal detection criterion still holds, i.e., the detection problem (3.3) is equivalently restated as

$$\hat{\mathbf{s}}_{\text{ML}} = \arg \min_{\bar{\mathbf{s}} \in \bar{\mathbb{S}}^L} \|\mathbf{y}' - \mathbf{R}\bar{\mathbf{s}}\|^2. \quad (3.4)$$

It is not unusual for the preprocessed channel matrix \mathbf{R} to be ill-conditioned. When this is the case, the complexity of near-ML TSA is known to grow significantly [70].

MMSE-GDFE preprocessing [69,71] was recently proposed as an alternative to the traditional ZF-DFE preprocessing. It is motivated by the well known fact that, under perfect decision feedback, the MMSE-GDFE [72] exhibits higher signal to interference-plus-noise ratio (SINR) than the ZF-DFE at the decision point. We now outline the main ideas behind the MMSE-GDFE preprocessing algorithm in [71]. Under the assumptions that \mathbf{s} and \mathbf{w} are zero-mean uncorrelated random vectors with covariance matrices $\sigma_s^2 \mathbf{I}_L$ and $\sigma^2 \mathbf{I}_L$, respectively, we define $\gamma \triangleq \sigma_s^2 / \sigma^2$ and the augmented channel matrix $\tilde{\mathbf{H}}$ in (3.5):

$$\tilde{\mathbf{H}} \triangleq \begin{bmatrix} \bar{\mathbf{H}} \\ \frac{1}{\sqrt{\gamma}} \mathbf{I}_L \end{bmatrix} \quad (3.5)$$

$$= \tilde{\mathbf{Q}} \tilde{\mathbf{R}} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} \tilde{\mathbf{R}}. \quad (3.6)$$

Equation (3.6) gives the QR decomposition of $\tilde{\mathbf{H}}$, where $\tilde{\mathbf{Q}}$ has orthonormal columns and $\tilde{\mathbf{R}}$ is upper triangular with positive diagonal entries. MMSE-GDFE preprocessing produces the transformed observation $\boldsymbol{\rho} \triangleq \mathbf{Q}_1^T \bar{\mathbf{y}}$ which is used in the detection problem

$$\hat{\mathbf{s}}_{\text{PP}} = \arg \min_{\bar{\mathbf{s}} \in \bar{\mathbb{S}}^L} \|\boldsymbol{\rho} - \tilde{\mathbf{R}}\bar{\mathbf{s}}\|^2. \quad (3.7)$$

Because $\mathbf{Q}_1 \in \mathbb{R}^{L \times L}$ is not guaranteed to be orthogonal, we cannot claim (for general³ constellations $\bar{\mathbb{S}}$) that $\hat{\mathbf{s}}_{\text{PP}} = \hat{\mathbf{s}}_{\text{ML}}$. When $\bar{\mathbf{H}}$ is fully populated (i.e., not quasi-banded) as in flat-fading multi-antenna communication, Damen [71] demonstrated that, at moderate-to-high SNR, $\hat{\mathbf{s}}_{\text{PP}}$ is near-ML and can be found, via TSA, at an average search complexity of $O(L^3)$, regardless of constellation size Q . We note, for later use, that the error $\mathbf{n} \triangleq \boldsymbol{\rho} - \tilde{\mathbf{R}}\bar{\mathbf{s}}$, while signal dependent and non-Gaussian, is white with covariance $\sigma^2 \mathbf{I}_L$ [73].

It is important to realize that, when $\bar{\mathbf{H}}$ has the quasi-banded structure in Fig. 3.1(a), $\tilde{\mathbf{R}}$ will have the ‘‘V-shaped’’ structure in Fig. 3.1(b). Since, as we shall see, the V-shaped structure can have a profound affect on TSA behavior, it is worthwhile to consider the conditions under which this V-shaping arises. As suggested by Fig. 3.1, we measure the degree of V-shaping by the ratio $\frac{4D+1}{2N}$; as $\frac{4D+1}{2N}$ decreases below 1, the V-shaping becomes more prominent. Recalling $D = \lceil 2f_{\text{D}}T_cN \rceil + \alpha$ and assuming the typical choice $N = 4N_h$, where $N_h \triangleq T_h/T_c$ denotes the normalized delay spread, we find

$$\frac{4D+1}{2N} = \frac{4\lceil 8f_{\text{D}}T_cN_h \rceil + 4\alpha + 1}{8N_h} = \frac{1.125 + 0.5\alpha}{N_h}, \quad (3.8)$$

where the second equality in (3.8) holds for all reasonable spreading factors, i.e., for $0 < 2f_{\text{d}}T_h \leq 0.5$. When $\alpha = 4$ (as used in Section 6.4), $\frac{4D+1}{2N} = \frac{3.125}{N_h}$, and so $\tilde{\mathbf{R}}$ will be V-shaped for $N_h > 3$. In most applications of interest, though, we have $N_h \gg 3$, in which case $\tilde{\mathbf{R}}$ is prominently V-shaped.

Additional TSA preprocessing might also be considered. For example, relaxing the constraint $\bar{\mathbf{s}} \in \bar{\mathbb{S}}^L$ in (3.7) to $\bar{\mathbf{s}} \in \left(\mathbb{Z} + \frac{1}{2}\right)^L$ allows more freedom in the choice

³It will be established in Section 3.3 that $\hat{\mathbf{s}}_{\text{ML}} = \hat{\mathbf{s}}_{\text{PP}}$ when the data is uncoded and constant modulus.

of lattice basis [70]. In our application, however, we are interested in preserving the quasi-banded structure of $\bar{\mathbf{H}}$, which limits the types of preprocessing that can be performed. These issues will be discussed further in Section 4.1.1.

3.2 Tree Search Algorithms

The preprocessed sequence detection problems (3.4) and (3.7) both correspond to tree search over a tree with depth L , where every tree node has Q children. A brute-force approach to tree search would entail the examination of the Euclidean metrics (3.4) and (3.7) at each of the Q^L leaf nodes. We are interested in search algorithms which prune branches that are unlikely to contain the ML path, thus drastically reducing the search complexity. Unlike their ML counterparts, near-ML tree search algorithms can, in some cases, discard the ML path, and hence return a suboptimal sequence estimate. Thus, each near-ML algorithm achieves a particular tradeoff between performance and complexity.

Tree search algorithms can be categorized as breadth-first, depth-first, or best-first search algorithms [66, 70]. Breadth-first search algorithms include, e.g., the M-algorithm [66], T-algorithm [74], statistical pruning algorithms [75], Wozencraft's TSA [76], and Pohst's sphere decoder [77]. Depth-first search algorithms include, e.g., the Schnor-Euchner's sphere decoder and its variants [67–69]. Best-first search algorithms include, e.g., the stack and Fano algorithms [65, 70, 78].

3.2.1 TSAs for MCM Application

We focus on a few representative TSAs and discuss their strengths and weaknesses in the context of solving (3.7) *for the DS-channel MCM application*, i.e., when $\tilde{\mathbf{R}}$ has the V-shaped structure in Fig. 3.1(b), as opposed to the general case of (3.7)

that results from, e.g., flat-fading multi-antenna channels and time-dispersive single-antenna channels—neither⁴ of which yield V-shaped $\tilde{\mathbf{R}}$. In fact, we find that the structure of $\tilde{\mathbf{R}}$ has a profound effect on TSA behavior.

We now briefly discuss depth-first, breadth-first, and best-first TSAs to gain insight into their behavior in the DS-channel MCM application. But first, some notation: We associate every node on the “ i^{th} level” of the tree ($i \geq 0$) with a realization of the partial path

$$\bar{\mathbf{s}}^i \triangleq [\bar{s}_i, \bar{s}_{i+1}, \dots, \bar{s}_{L-1}]^T \in \bar{\mathbb{S}}^{L-i}. \quad (3.9)$$

The root node corresponds to the L^{th} level and the leaf nodes to the 0^{th} level. The Euclidean partial-path metric associated with $\bar{\mathbf{s}}^i$ is defined in (3.10) using $\tilde{r}_{k,l} \triangleq [\tilde{\mathbf{R}}]_{k,l}$:

$$\mathcal{M}(\bar{\mathbf{s}}^i) \triangleq \sum_{k=i}^{L-1} \left| \rho_k - \sum_{l=k}^{L-1} \tilde{r}_{k,l} \bar{s}_l \right|^2. \quad (3.10)$$

- Depth-First Search: Depth-first search (DFS) algorithms proceed down the tree by following the minimum-cost branch at each level. The first full path obtained in this manner, corresponding to the classical DFE sequence estimate, is kept as a reference. The DFS algorithm then backs up one level at a time, re-examining the discarded branches at each level and pursuing any that have a chance at beating the reference. If a new best-sequence is found, it is used as the new reference and the process is repeated. DFS yields very low search complexity when the initial (i.e., DFE) sequence estimate is ML, since no other branches will be re-examined. For this reason, DFS complexity approaches DFE complexity

⁴The ICI span of properly designed MCM (i.e., N_H) will be much shorter than the ISI span of an equivalent single-carrier system (i.e., N_h). Thus, while a time-domain channel matrix would be banded, it would have a much wider band than our quasi-banded \mathbf{H} . Unless \mathbf{H} has a narrow band, $\tilde{\mathbf{R}}$ will not be V-shaped.

at high SNR. At low SNR, however, DFS can waste a lot of effort on non-ML paths, leading to very costly searches.

When $\tilde{\mathbf{R}}$ is V-shaped, as in MCM-shaped DS channels, and the SNR is moderate-to-low, DFS will not be efficient in solving (3.7). To see why, consider Fig. 3.2, which shows that \bar{s}_{L-2D-1} does not affect $\{\rho_0, \dots, \rho_{L-4D-1}\}$. Consequently, an error in \hat{s}_{L-2D-1} will be invisible to the branch metrics at levels $i \in \{0, \dots, L - 4D - 2\}$. When such an error occurs, all DFS branch re-examinations at levels $i \in \{0, \dots, L - 4D - 2\}$ will be performed in vain. Similar situations occur with errors in \hat{s}_k for $k \in \{2D + 1, \dots, L - 2D - 2\}$. Note that this behavior does not manifest for general upper-triangular $\tilde{\mathbf{R}}$. Thus, while DFS algorithms like the Schnorr-Euchner's sphere decoder may be attractive in multi-antenna or time-dispersive channels, they are not well suited to MCM-shaped DS channels. These notions will be confirmed numerically in Section 4.2.

- Best-First Search: Best-first search (BeFS) algorithms maintain a sorted list of the best partial paths (of possibly different lengths). At each iteration, BeFS extends the best partial path, replaces its list entry with that of its children, and re-sorts the list. BeFS terminates as soon as the best partial path reaches a leaf node, since, at that point, all other partial paths are destined to yield inferior full-path metrics. The Fano algorithm is a near-ML BeFS algorithm that uses the biased partial-path metric

$$\mathcal{M}_{\text{Fano}}(\bar{\mathbf{s}}^i) \triangleq \sum_{k=i}^{L-1} \left| \rho_k - \sum_{l=k}^{L-1} \tilde{r}_{k,l} \bar{s}_l \right|^2 - (L-i)b \quad \text{for } b > 0. \quad (3.11)$$

Larger b biases Fano in favor of longer paths, yielding quicker searches; for very large b , Fano behaves like DFS, greedily extending the best path at every level

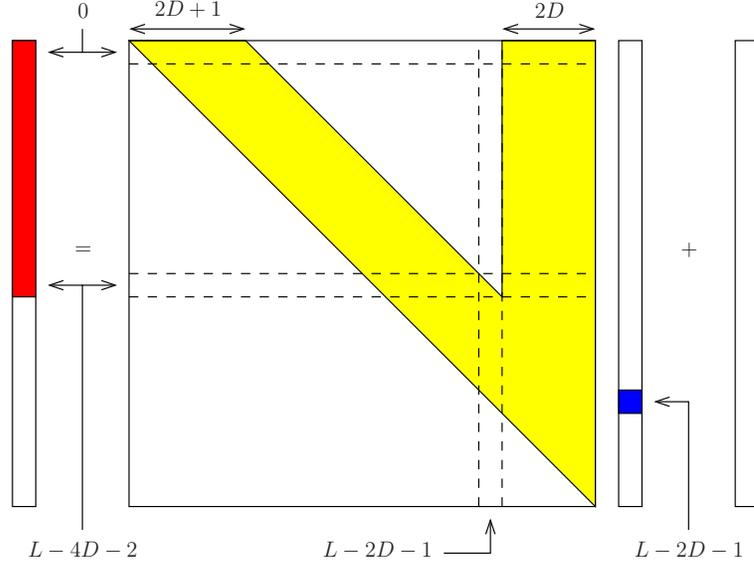


Figure 3.2: Illustration of $\boldsymbol{\rho} = \tilde{\mathbf{R}}\bar{\mathbf{s}} + \mathbf{n}$ for V-shaped $\tilde{\mathbf{R}}$. The PAM symbol \bar{s}_{L-2D-1} does not affect $\{\rho_0, \dots, \rho_{L-4D-2}\}$.

and returning the DFE sequence estimate. In practice, b is chosen to achieve a particular complexity/performance tradeoff.

A recent comprehensive comparison [70] suggested that a properly-designed Fano algorithm achieves a better complexity/performance tradeoff than all other known TSAs when $\tilde{\mathbf{R}}$ has a fully populated upper triangle. For V-shaped $\tilde{\mathbf{R}}$, however, BeFS algorithms (like Fano) can face difficulties. Recalling Fig. 3.2, when the best partial path includes an error in \hat{s}_{L-2D-1} , the branch metrics at levels $i \in \{0, \dots, L-4D-2\}$ will be non-informative about this error, and thus BeFS algorithms can waste lots of time pursuing extensions of this “best” path in vain. Similar situations occur with errors in \hat{s}_k for $k \in \{2D+1, \dots, L-2D-2\}$. Furthermore, best-partial-path errors in any of these \hat{s}_k ’s will be gradually de-emphasized by the Fano bias term in (3.11) as these “best”

partial paths are extended, making the Fano algorithm less likely to revisit the shorter stack elements without the error in \hat{s}_k . Consequently, Fano exhibits an exploding complexity at low SNR and an inferior complexity/performance tradeoff at high SNR when used with the $\tilde{\mathbf{R}}$ that results from MCM-shaped DS channels. These notions will be confirmed numerically in Section 6.4.

- Breadth-First Search: As we saw earlier, the complexity of DFS and BeFS explodes at low SNR because a huge amount of searching is needed to eliminate suboptimal paths, and the problem is exacerbated by V-shaped $\tilde{\mathbf{R}}$. Breadth-first search (BrFS) complexity, in contrast, is much less sensitive to SNR and the structure of $\tilde{\mathbf{R}}$, suggesting that it might be advantageous in our application. The M-algorithm, for example, has complexity that is *invariant* to both SNR and $\tilde{\mathbf{R}}$. The M-algorithm starts at the root node (i.e., level L) and chooses the M best child nodes at level $L - 1$. The children of these level- $(L - 1)$ nodes are then evaluated, and the M best are chosen. This process repeats at every level, extending M nodes per level, until finally the best leaf node is chosen as the sequence estimate.

At high SNR, however, the M-algorithm is much more expensive than DFS and BeFS because it is not aggressive enough in branch pruning. Hence, a better complexity/performance tradeoff might be achieved by a BrFS algorithm that varies the number of nodes considered at each level. For example the T-algorithm only extends paths from nodes whose Euclidean metrics lie in the interval $[\mathcal{M}(\bar{\mathbf{s}}_*^i), \mathcal{M}(\bar{\mathbf{s}}_*^i) + T)$, where $\mathcal{M}(\bar{\mathbf{s}}_*^i)$ denotes the minimum Euclidean metric among all considered nodes, and where T is a threshold parameter that

is chosen to achieve a particular complexity/performance tradeoff. Several approaches to the design of T have been proposed. For example, [74] took an experimental approach, while [79] and [80] used SNR and code structure. In Section 4.1.2 we propose an *adaptive* T-algorithm which uses the elements in $\tilde{\mathbf{R}}$, as well as SNR, to optimize T at each level. We shall see that this adaptive T-algorithm results in a superior complexity/performance tradeoff for MCM-shaped DS channels.

3.3 Optimality of MMSE-GDFE Preprocessed Estimates

In this section, we establish that MMSE-GDFE preprocessing does not compromise the ML-optimality of minimum distance (MD) search for $N \times M$ MIMO systems under constant modulus signaling. This property holds for systems of arbitrary size (i.e., over- or under-determined linear channels), though not for non-constant modulus constellations. The result is attractive because MMSE-GDFE preprocessing is known to yield significant reductions in the average search complexity of TSAs, especially in moderate-to-low SNR ranges and/or with ill-conditioned/under-determined linear channels. Therefore, our finding implies that the complexity savings of MMSE-GDFE preprocessing can be leveraged for *true* ML (rather than only *near*-ML) detection.

Consider the generic MIMO system model

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w}, \quad (3.12)$$

where $\mathbf{H} \in \mathbb{C}^{N \times M}$. Let us define $\mathcal{Y}(\mathbf{s})$ as the set of (non-preprocessed) observations for which the ML estimate of the constant modulus sequence $\mathbf{s} \in \mathbb{S}^M$ is the optimal MD solution. In other words,

$$\mathcal{Y}(\mathbf{s}) \triangleq \{\mathbf{y} : \|\mathbf{y} - \mathbf{H}\mathbf{s}\| \leq \|\mathbf{y} - \mathbf{H}\mathbf{s}'\|, \forall \mathbf{s}' \in \mathbb{S}^M\}, \quad (3.13)$$

where \mathbf{s}' is an arbitrary constant modulus sequence. Since $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w}$, we find that

$$\begin{aligned} \|\mathbf{y} - \mathbf{H}\mathbf{s}\| &\leq \|\mathbf{y} - \mathbf{H}\mathbf{s}'\| \\ \Leftrightarrow \|\mathbf{w}\| &\leq \|\mathbf{w} - \mathbf{H}(\mathbf{s}' - \mathbf{s})\| \\ \Leftrightarrow 2\Re\{\mathbf{w}^H \mathbf{H}(\mathbf{s}' - \mathbf{s})\} &\leq \|\mathbf{H}(\mathbf{s}' - \mathbf{s})\|^2. \end{aligned} \quad (3.14)$$

Now we define an error sequence $\boldsymbol{\alpha}(\mathbf{s}) \triangleq \mathbf{s}' - \mathbf{s}$ and a set of the error sequences $\mathcal{A}(\mathbf{s}) \triangleq \{\mathbf{s}' - \mathbf{s} : \mathbf{s}' \in \mathbb{S}^M\}$ provided that \mathbf{s} is the minimum distance solution. Then, putting (3.13) and (3.14) together, we see that the following equivalence holds for arbitrary $\mathbf{s} \in \mathbb{S}^M$.

$$\mathbf{y} \in \mathcal{Y}(\mathbf{s}) \Leftrightarrow 2\Re\{\mathbf{w}^H \mathbf{H}\boldsymbol{\alpha}(\mathbf{s})\} \leq \|\mathbf{H}\boldsymbol{\alpha}(\mathbf{s})\|^2, \forall \boldsymbol{\alpha}(\mathbf{s}) \in \mathcal{A}(\mathbf{s}). \quad (3.15)$$

Next let us define $\tilde{\mathcal{Y}}(\mathbf{s})$ as the set of MMSE-GDFE preprocessed observations for which the MD estimate of the sequence \mathbf{s} will be the minimum distance solution. In other words,

$$\tilde{\mathcal{Y}}(\mathbf{s}) \triangleq \{\tilde{\mathbf{y}} : \|\tilde{\mathbf{y}} - \tilde{\mathbf{R}}\mathbf{s}\| \leq \|\tilde{\mathbf{y}} - \tilde{\mathbf{R}}\mathbf{s}'\|, \forall \mathbf{s}' \in \mathbb{S}^M\} \quad (3.16)$$

where

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{Q}_1^H \mathbf{y} \\ &= \mathbf{Q}_1^H (\mathbf{H}\mathbf{s} + \mathbf{w}) \\ &= \tilde{\mathbf{R}}\mathbf{s} + \underbrace{(\mathbf{Q}_1^H \mathbf{H} - \tilde{\mathbf{R}})\mathbf{s} + \mathbf{Q}_1^H \mathbf{w}}_{\triangleq \mathbf{n}}. \end{aligned} \quad (3.17)$$

Recall that, since \mathbf{Q}_1 is typically non-orthogonal, we cannot claim that $\mathbf{Q}_1^H \mathbf{H} = \tilde{\mathbf{R}}$. Repeating the argument in (3.14), we obtain the following equivalence, which holds for arbitrary $\mathbf{s} \in \mathbb{S}^M$

$$\tilde{\mathbf{y}} \in \tilde{\mathcal{Y}}(\mathbf{s}) \Leftrightarrow 2\Re\{\mathbf{n}^H \tilde{\mathbf{R}}\boldsymbol{\alpha}(\mathbf{s})\} \leq \|\tilde{\mathbf{R}}\boldsymbol{\alpha}(\mathbf{s})\|^2, \forall \boldsymbol{\alpha}(\mathbf{s}) \in \mathcal{A}(\mathbf{s}). \quad (3.18)$$

Finally, let us define the \mathbf{Q}_1^H -transformation of the region $\mathcal{Y}(\mathbf{s})$:

$$\mathcal{Y}_{\mathbf{Q}_1^H}(\mathbf{s}) \triangleq \{\mathbf{Q}_1^H \mathbf{y} : \mathbf{y} \in \mathcal{Y}(\mathbf{s})\}. \quad (3.19)$$

Note that $\mathbf{y} \in \mathcal{Y}(\mathbf{s}) \Leftrightarrow \tilde{\mathbf{y}} \in \mathcal{Y}_{\mathbf{Q}_1^H}(\mathbf{s})$ since $\tilde{\mathbf{y}} \triangleq \mathbf{Q}_1^H \mathbf{y}$. Thus (3.15) implies

$$\tilde{\mathbf{y}} \in \mathcal{Y}_{\mathbf{Q}_1^H}(\mathbf{s}) \Leftrightarrow 2\mathbf{w}^H \mathbf{H} \boldsymbol{\alpha}(\mathbf{s}) \leq \|\mathbf{H} \boldsymbol{\alpha}(\mathbf{s})\|^2, \forall \boldsymbol{\alpha}(\mathbf{s}) \in \mathcal{A}(\mathbf{s}) \quad (3.20)$$

for arbitrary $\mathbf{s} \in \mathbb{S}^M$.

Lemma 1 $\mathcal{Y}_{\mathbf{Q}_1^H}(\mathbf{s}) = \tilde{\mathcal{Y}}(\mathbf{s})$ for arbitrary constant modulus signaling vector $\mathbf{s} \in \mathbb{S}^M$ and SNR γ .

proof : Examining the left side of the inequality in (3.18), we see that

$$\begin{aligned} \mathbf{n}^H \tilde{\mathbf{R}} \boldsymbol{\alpha}(\mathbf{s}) &= \left(\mathbf{w}^H \mathbf{Q}_1 + \mathbf{s}^H \left(\mathbf{H}^H \mathbf{Q}_1 - \tilde{\mathbf{R}}^H \right) \right) \tilde{\mathbf{R}} \boldsymbol{\alpha}(\mathbf{s}) \\ &= \mathbf{w}^H \mathbf{H} \boldsymbol{\alpha}(\mathbf{s}) + \mathbf{s}^H \left(\mathbf{H}^H \mathbf{H} - \tilde{\mathbf{R}}^H \tilde{\mathbf{R}} \right) \boldsymbol{\alpha}(\mathbf{s}) \\ &= \mathbf{w}^H \mathbf{H} \boldsymbol{\alpha}(\mathbf{s}) - \gamma^{-1} \mathbf{s}^H \boldsymbol{\alpha}(\mathbf{s}), \end{aligned} \quad (3.21)$$

where we have used the facts that $\mathbf{Q}_1 \tilde{\mathbf{R}} = \mathbf{H}$ and $\tilde{\mathbf{R}}^H \tilde{\mathbf{R}} = \mathbf{H}^H \mathbf{H} + \gamma^{-1} \mathbf{I}_M$. The latter fact also implies

$$\|\tilde{\mathbf{R}} \boldsymbol{\alpha}(\mathbf{s})\|^2 = \|\mathbf{H} \boldsymbol{\alpha}(\mathbf{s})\|^2 + \gamma^{-1} \|\boldsymbol{\alpha}(\mathbf{s})\|^2. \quad (3.22)$$

Equations (3.21)-(3.22) can be used to rewrite (3.18) as

$$\begin{aligned} \tilde{\mathbf{y}} \in \tilde{\mathcal{Y}}(\mathbf{s}) \Leftrightarrow 2\Re\{\mathbf{w}^H \mathbf{H} \boldsymbol{\alpha}(\mathbf{s})\} \leq \underbrace{\|\mathbf{H} \boldsymbol{\alpha}(\mathbf{s})\|^2 + \gamma^{-1} \left(\|\boldsymbol{\alpha}(\mathbf{s})\|^2 + 2\Re\{\mathbf{s}^H \boldsymbol{\alpha}(\mathbf{s})\} \right)}_{\triangleq E}, \\ \forall \boldsymbol{\alpha}(\mathbf{s}) \in \mathcal{A}(\mathbf{s}) \end{aligned} \quad (3.23)$$

When \mathbf{s} is a constant-modulus vector, it can be seen that $E = 0$, i.e.,

$$E = \gamma^{-1} \left(\|\mathbf{s}' - \mathbf{s}\|^2 + 2\Re\{\mathbf{s}^H(\mathbf{s}' - \mathbf{s})\} \right) \quad (3.24)$$

$$= \gamma^{-1} \left(\|\mathbf{s}'\|^2 - \|\mathbf{s}\|^2 \right) \quad (3.25)$$

$$= 0 \quad (3.26)$$

Thus, comparing (3.23) to (3.20), we see that $\mathcal{Y}_{\mathbf{Q}_1^H}(\mathbf{s}) = \tilde{\mathcal{Y}}(\mathbf{s})$ for arbitrary $\mathbf{s} \in \{-\frac{1}{2}, \frac{1}{2}\}^M$. ■

From Lemma 1, we see that the decision regions with MMSE-GDFE and ZF-DFE preprocessing are the same, which leads to the following corollary.

Corollary 2 *Consider model (3.12) with arbitrary \mathbf{H} , \mathbf{w} , and constant modulus \mathbf{s} . Then we have $\hat{\mathbf{s}}_{\text{ML}} = \hat{\mathbf{s}}_{\text{PP}}$.*

Corollary 2 establishes the fact that MMSE-GDFE preprocessing does not affect the optimality of ML detection when constant modulus signaling is used. From the proof of Lemma 1, it is clear that the property $\mathcal{Y}_{\mathbf{Q}_1^H}(\mathbf{s}) = \tilde{\mathcal{Y}}(\mathbf{s})$ will not hold for non-constant modulus, e.g., PAM/QAM, constellations, though, implying that MMSE-GDFE preprocessing can render MD decision making sub-optimal when non-constant modulus constellations are used. This sub-optimality will be investigated further in Section 3.3.1.

It is interesting to note that the MMSE-GDFE property holds for arbitrary positive γ . (This fact will be confirmed numerically in Section 3.3.1.) Thus, ML estimates can be obtained via MMSE-GDFE preprocessing without knowledge of SNR. However, the search complexity remains a function of γ , and numerical experiments suggest choosing γ as specified in [71].

3.3.1 Numerical Experiments

In this section we present the results of three numerical experiments. In all experiments, the MIMO channel matrix $\mathbf{H} \in \mathbb{R}^{N \times M}$ was generated from i.i.d. zero-mean Gaussian elements with variance M^{-1} , and the noise vector from i.i.d. zero-mean Gaussian elements with variance $(4\text{SNR})^{-1}$ for BPSK and $(4\text{SNR}/5)^{-1}$ for 4-PAM, recalling that $\mathbb{S} = \{-\frac{1}{2}, \frac{1}{2}\}$ for BPSK and $\mathbb{S} = \{-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}\}$ for 4-PAM. In other words, SNR is the signal-energy to noise-energy ratio at each receive antenna. Unless noted otherwise, we set $\gamma = \text{SNR}$, as specified in [71].

The first experiment shows the typical reduction of search complexity that comes from the use of MMSE-GDFE preprocessing and ordering in place of the traditional ZF-DFE preprocessing and ordering. We employed the greedy ordering scheme suggested by [69] which was originally proposed for V-BLAST in [81]. Figure 3.3 shows average Schnorr-Euchner (SE) sphere decoder (SpD) search complexity for a system with $M = N = 32$ under BPSK signaling (or, equivalently, $m = n = 16$ under QPSK signaling). By “search complexity” we mean the number of real multiplications plus additions per frame consumed by the SpD search stage⁵ of the SE-SpD algorithm from [69]. Note that the complexity is reported on a \log_N scale. Here the SE-SpD sphere radius was initialized at 1.5 times the average distance between the observation and the closest lattice point. Figure 3.3 demonstrates that MMSE-GDFE preprocessing can lead to significant complexity savings over a moderate-to-low SNR range. For example, a factor of about 10 in complexity savings can be observed for

⁵This definition assumes slow-fading, where the matrix computations associated with preprocessing could be amortized over many frames, thereby making the SpD search complexity dominant.

SNR at 8 dB. More detailed investigations of MMSE-GDFE complexity savings can be found in [69–71].

The second experiment compares the frame error rate (FER) achieved by the ML detector (3.3) to that achieved by the MMSE-GDFE preprocessed MD detector (3.7) under BPSK signaling for several combinations of M and N and for several choices of MMSE-GDFE parameter γ . Specifically, Fig. 3.4 examines $(M, N) \in \{(6, 8), (8, 8), (8, 6)\}$, which includes over- and under-determined linear channels, and $\gamma \in \{\frac{1}{10}\text{SNR}, \text{SNR}, 10\text{SNR}\}$. Consistent with Corollary 2, the FER of MMSE-GDFE sphere decoding is identical to that of ML detection in all cases. In this experiment, the ML detector was implemented using the ZF-DFE preprocessed SE-SpD algorithm from [69].

The third experiment verifies the sub-optimality of MMSE-GDFE preprocessed MD estimates when non-BPSK/QPSK constellations are used. In Fig. 3.5, the FER of the ML detector is compared to that of the MMSE-GDFE preprocessed SpD for a 4-PAM system with $M = N = 8$. Observe that the FER degradation caused by MMSE-GDFE preprocessing is small but measurable. As before, the ML detector was implemented using the ZF-DFE preprocessed SE-SpD algorithm from [69].

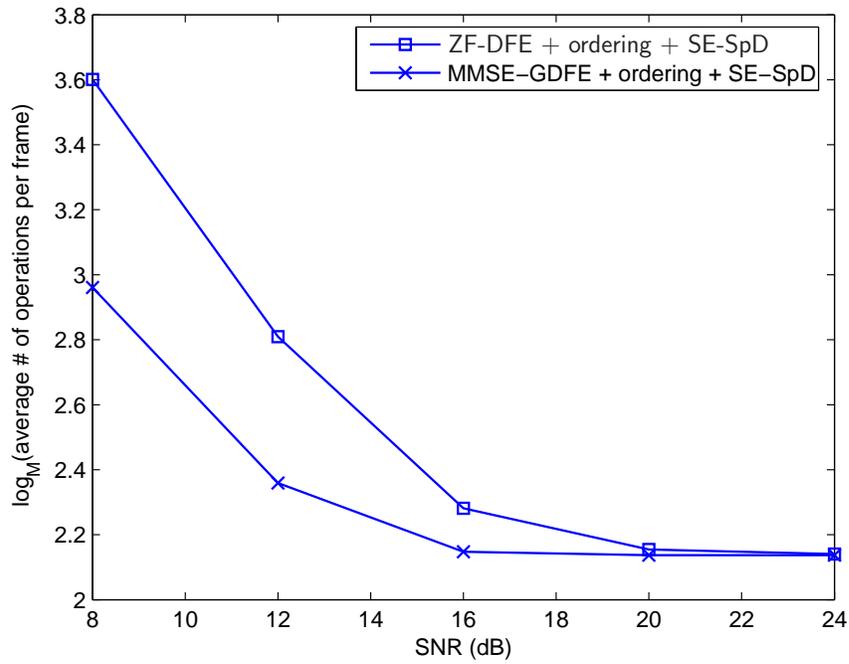


Figure 3.3: A comparison of the per-frame average search complexity of Schnorr-Euchner (SE) SpD with two forms of preprocessing: ZF-DFE versus MMSE-GDFE, both with greedy ordering. A BPSK system of dimension $M = N = 32$ was employed and $\gamma = \text{SNR}$ was used in MMSE-GDFE.

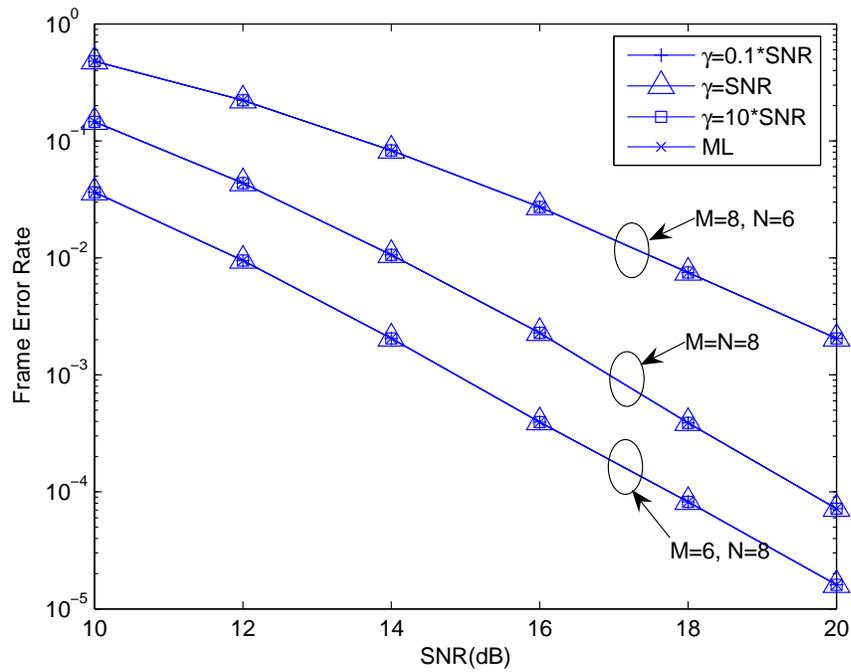


Figure 3.4: Frame error rate of the ML detector (via ZF-DFE preprocessed SpD) versus the MMSE-GDFE preprocessed SpD for a BPSK system under several channel dimensions (M, N) and several values of MMSE-GDFE parameter γ .

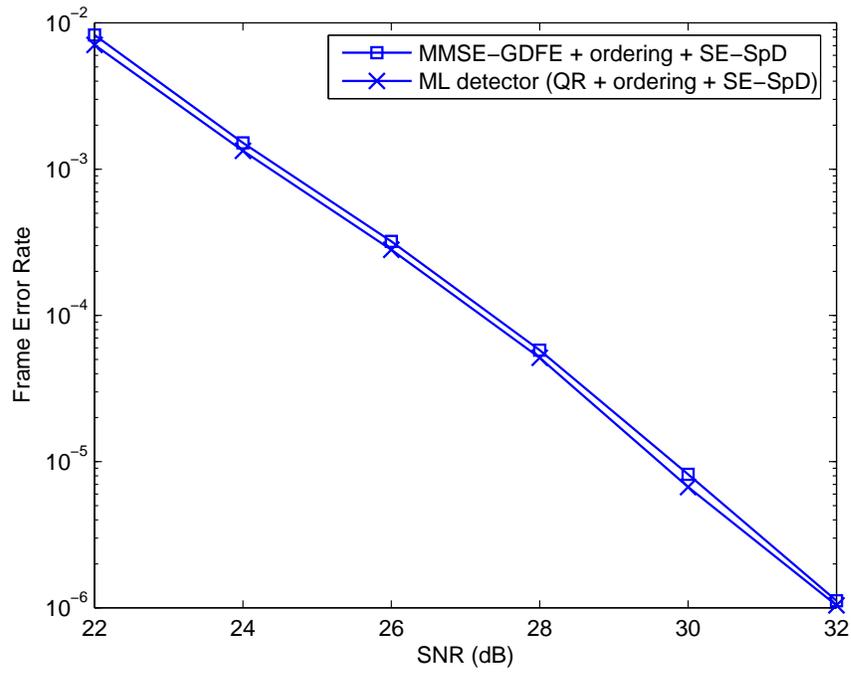


Figure 3.5: Frame error rate of the ML detector (via ZF-DFE preprocessed SpD) versus the MMSE-GDFE preprocessed SpD for a 4-PAM system with $M = N = 8$.

CHAPTER 4

NEAR-OPTIMAL COHERENT EQUALIZATION FOR PS-MCM

In this chapter, we combine the PS-MCM previously proposed by Schniter and Das in [5, 17] with near-ML TSAs [65, 66, 70] with rank-reduced pilot-aided and BE-based Wiener channel estimation for high-spectral-efficiency, high-performance, and low-complexity multicarrier communication over the DS channel. By “near-ML,” we mean FER performance equivalent to that attained by maximum-likelihood (ML) sequence detection (SD) at a fraction-of-a-dB lower SNR. We tolerate this small loss because, as we shall see, it enables huge complexity savings relative to true MLSD. We choose the PS-MCM scheme from [5, 17] because of its high spectral efficiency and excellent ISI/ICI suppression. We propose TSA based on a novel fast MMSE-GDFE preprocessor [71] and on a novel channel-adaptive T-algorithm [74], both of which are specifically tailored to the ISI/ICI structure induced by PS-MCM over the DS channel. We’ve discussed, in Section 3.2.1, the shortcomings of traditional TSAs on these channels. Numerical experiments are conducted to evaluate the efficacy of the PS-MCM scheme, the proposed TSA, the channel estimator, and their combination, relative to other designs.

4.1 Proposed PS-MCM Coherent Equalization

In the proposed PS-MCM receiver, a fast TSA preprocessing is applied to the subchannel outputs $\bar{\mathbf{y}}$ prior to TSA via the adaptive T-algorithm. The channel coefficients used in TSA are estimated via pilot symbols. Below, we describe each receiver component in detail.

4.1.1 TSA Preprocessing

In this section we describe low-complexity TSA preprocessing which leverages the quasi-banded structure in $\bar{\mathbf{H}}^{(j)}$. For ease of TSA discussion, we assume real-valued system model in (2.15) and suppress the $^{(j)}$ notation in this section. Also the signal is assumed to be chosen from the Q -ary PAM alphabet described in (3.2). We start with a description of a low-complexity implementation of MMSE-GDFE preprocessing. Then a simple ordering scheme which preserves the quasi-banded structure in $\bar{\mathbf{H}}$ is presented.

Fast MMSE-GDFE Preprocessing

The MMSE-GDFE preprocessing originally proposed in [71] involves QR decomposition with complexity $\mathcal{O}(L^3)$. In this section, we propose an $\mathcal{O}(D^2L)$ implementation of MMSE-GDFE preprocessing that leverages the quasi-banded structure of $\bar{\mathbf{H}}$ found in our application. We note connections to the fast MMSE-DFE in [15], which was formulated for a banded (as opposed to quasi-banded) matrix $\bar{\mathbf{H}}$ that occurs when the edge subcarriers are inactive.

Recall the augmented channel matrix $\tilde{\mathbf{H}}$ in (3.5) and its QR decomposition (3.6). Note that, while $\bar{\mathbf{H}}$ is quasi-banded with $2D + 1$ active diagonals [as illustrated

in Fig. 3.1(a)], $\tilde{\mathbf{H}}$ is not quasi-banded. However, the matrix $\tilde{\mathbf{H}}^T \tilde{\mathbf{H}}$, which can be computed in $(4D^2 + 4D + 2)L$ MACs, is quasi-banded with $4D + 1$ active diagonals. Now, since $\tilde{\mathbf{Q}}$ is an orthogonal matrix, we know $\tilde{\mathbf{H}}^T \tilde{\mathbf{H}} = \tilde{\mathbf{R}}^T \tilde{\mathbf{R}}$. Hence, $\tilde{\mathbf{R}}$ can be obtained via Cholesky factorization [82] of $\tilde{\mathbf{H}}^T \tilde{\mathbf{H}}$ in $O(D^2L)$ operations. Table 4.1 details the fast Cholesky factorization $\mathbf{A} = \mathbf{G}\mathbf{G}^T$, where $\mathbf{A} \triangleq \tilde{\mathbf{H}}^T \tilde{\mathbf{H}}$ and where $\mathbf{G} \triangleq \tilde{\mathbf{R}}^T$ is the lower triangular Cholesky factor. This fast computation of $\tilde{\mathbf{R}}$ can be shown to consume $(10D^2 + 11D + 2)L - \frac{1}{3}(74D^3 + 133D^2 + 44D + 3)$ MAC operations.

Next, we consider the implementation of the preprocessing operation $\boldsymbol{\rho} = \mathbf{Q}_1^T \mathbf{y}$. Multiplication of this equality by $\tilde{\mathbf{R}}^T$ yields

$$\tilde{\mathbf{R}}^T \boldsymbol{\rho} = \tilde{\mathbf{R}}^T \mathbf{Q}_1^T \mathbf{y} = \bar{\mathbf{H}}^T \mathbf{y} \triangleq \mathbf{b}. \quad (4.1)$$

Due to quasi-banded $\bar{\mathbf{H}}$, the vector \mathbf{b} can be computed in $(2D + 1)L$ MAC operations. From \mathbf{b} we can solve (4.1) for $\boldsymbol{\rho}$ using forward substitution in $O(DL)$ additional operations, because $\tilde{\mathbf{R}}^T$ has the sparse “V-shaped” structure in Fig. 3.1(b). In total, this consumes $(6D + 2)L - 6D^2 - 3D$ MAC operations. Combining forward substitution with fast Cholesky decomposition, our fast MMSE-GDFE preprocessing requires $(14D^2 + 21D + 6)L - \frac{76}{3}D^3 - 53D^2 - \frac{53}{3}D - 1$ real MAC operations.

Circular Ordering

In [69], Damen et al. outline three stages of TSA preprocessing: lattice reduction, column ordering, and MMSE-GDFE preprocessing. In our application, the lattice reduction and column ordering would destroy the quasi-banded structure of $\bar{\mathbf{H}}$, in which case the subsequent MMSE-GDFE preprocessing would require a complexity of $O(L^3)$. Since, in practice, $L = 2N$ can be quite large (e.g., in the hundreds or thousands), such a complexity would be impractical. For these reasons, we restrict

Table 4.1: Fast Cholesky factorization of quasi-banded \mathbf{A} .

<p>Say $\mathbf{A} = \mathbf{G}\mathbf{G}^T$, where \mathbf{G} is lower triangular and $\mathbf{A} \in \mathbb{R}^{L \times L}$ is quasi-banded with $\pm 2D$ diagonals.</p> <p>for $j = 0 : L - 4D - 1$ $v_{j:L-1} = [\mathbf{A}]_{j:L-1,j}$ $m_1 = \max\{0, j - 2D - 1\}$ $m_2 = j + 2D - 1$ for $i = m_1 : j - 1$ $v_{j:m_2} = v_{j:m_2} - [\mathbf{G}]_{j,i}[\mathbf{G}]_{j:m_2,j}$ $v_{L-2D-1:L-1} = v_{L-2D-1:L-1}$ $\quad - [\mathbf{G}]_{j,i}[\mathbf{G}]_{L-2D-1:L-1,j}$ end $[\mathbf{G}]_{j:m_2,j} = v_{j:m_2} / \sqrt{v_j}$ $[\mathbf{G}]_{L-2D-1:L-1,j} = v_{L-2D-1:L-1} / \sqrt{v_j}$ end for $j = L - 4D : L - 2D - 1$ $v_{j:L-1} = [\mathbf{A}]_{j:L-1,j}$ $m_1 = \max\{0, j - 2D - 1\}$ for $i = m_1 : j - 1$ $v_{j:L-1} = v_{j:L-1} - [\mathbf{G}]_{j,i}[\mathbf{G}]_{j:L-1,j}$ end $[\mathbf{G}]_{j:L-1,j} = v_{j:L-1} / \sqrt{v_j}$ end for $j = L - 2D : L - 1$ $v_{j:L-1} = [\mathbf{A}]_{j:L-1,j}$ for $i = 0 : j - 1$ $v_{j:L-1} = v_{j:L-1} - [\mathbf{G}]_{j,i}[\mathbf{G}]_{j:L-1,j}$ end $[\mathbf{G}]_{j:L-1,j} = v_{j:L-1} / \sqrt{v_j}$ end</p>

ourselves to preprocessing operations which preserve the quasi-banded structure of $\bar{\mathbf{H}}$.

One admissible preprocessing operation is an n -place circular shift in column order of $\bar{\mathbf{H}}$. Using the left circular shift matrix \mathbf{J} , the shifting operation transforms (3.1) into the equivalent system (4.2) with channel matrix $\bar{\mathbf{H}}\mathbf{J}^{-n}$:

$$\mathbf{y} = (\bar{\mathbf{H}}\mathbf{J}^{-n})\mathbf{J}^n\bar{\mathbf{s}} + \mathbf{w} \quad (4.2)$$

$$\mathbf{J} \triangleq \begin{bmatrix} \mathbf{0}_{L-1} & \mathbf{I}_{L-1} \\ 1 & \mathbf{0}_{L-1}^T \end{bmatrix}. \quad (4.3)$$

Though $\mathbf{H}\mathbf{J}^{-n}$ is not quasi-banded in the sense of Fig. 3.1a, the matrix $\tilde{\mathbf{H}}^T\tilde{\mathbf{H}} = \tilde{\mathbf{R}}^T\tilde{\mathbf{R}}$ is, allowing the fast MMSE-GDFE processing from Section 4.1.1. Among the unique shifts $n \in \{0, \dots, L-1\}$, we choose the one which maximizes the norm of the rightmost column of $\bar{\mathbf{H}}\mathbf{J}^{-n}$, i.e., the norm of the rightmost column of $\tilde{\mathbf{R}}$. Thus, the PAM symbol contributing the most energy to \mathbf{y} is placed at the root of the tree. The complexity of this circular ordering stage is dominated by the evaluation of column norms, requiring $\mathcal{O}(DN)$ operations. We have observed, numerically, that this ‘‘circular ordering’’ scheme yields a modest improvement in terms of the performance/complexity tradeoff.

4.1.2 Channel-Adaptive T-algorithm

In this section we propose a channel-adaptive version of the T-algorithm in which the threshold parameter T_i is adjusted at the i^{th} level in the tree according to the channel realization and noise variance. Recall that the T-algorithm is a breadth-first search algorithm which, at the i^{th} level, discards all partial paths $\bar{\mathbf{s}}^i$ whose metric $\mathcal{M}(\bar{\mathbf{s}}^i)$ exceeds that of the best partial path $\bar{\mathbf{s}}_*^i \triangleq \arg \min_{\bar{\mathbf{s}}^i} \mathcal{M}(\bar{\mathbf{s}}^i)$ by an amount $\geq T_i$. (See Fig. 4.1.) Thus, the T-algorithm will make a block error if the *true* partial path $\bar{\mathbf{s}}_{\dagger}^i$ is discarded at any level $i \in \{L-1, L-2, \dots, 0\}$.

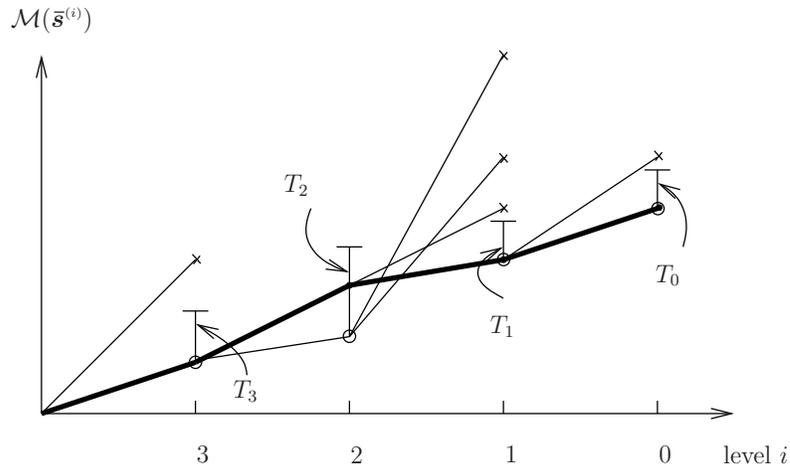


Figure 4.1: Illustration of path evolution in the T-algorithm when $Q = 2$ and $L = 4$. The circled points denote the minimum path metrics, the crossed points denote the discarded path metrics, and the bold line denotes the true path. Note that, in this example, $\mathcal{M}(\bar{\mathbf{s}}_*^2) < \mathcal{M}(\bar{\mathbf{s}}_{\top}^2)$.

In our adaptive T-algorithm, we set the threshold T_i so that the true path is discarded with probability ϵ_o when the true path is not the best partial path.

$$\Pr\{\mathcal{M}(\bar{\mathbf{s}}_{\top}^i) > \mathcal{M}(\bar{\mathbf{s}}_*^i) + T_i \mid \mathcal{M}(\bar{\mathbf{s}}_{\top}^i) > \mathcal{M}(\bar{\mathbf{s}}_*^i)\} < \epsilon_o. \quad (4.4)$$

Note that this is different from simply setting T_i so that the true path is discarded with probability ϵ_o . In the latter case, T_i will increase—thereby increasing search complexity—at low SNR. Intuition, however, tells us that it is not worthwhile to search extensively at low SNR because, even if found, the ML path is more likely to be in error.

With

$$\mu^i \triangleq \mathcal{M}(\bar{\mathbf{s}}_{\top}^i) - \mathcal{M}(\bar{\mathbf{s}}_*^i) \quad (4.5)$$

we can rewrite (4.4) as

$$\Pr\{\mu^i > T_i \mid \mu^i > 0\} < \epsilon_o. \quad (4.6)$$

We now analyze the random variable μ^i . To do this, we define $\boldsymbol{\rho}^i \triangleq [\rho_i, \rho_{i+1}, \dots, \rho_{L-1}]^T$ and construct $\tilde{\mathbf{R}}^i \in \mathbb{R}^{(L-i) \times (L-i)}$ from the last $L-i$ rows and columns of $\tilde{\mathbf{R}}$, i.e., $[\tilde{\mathbf{R}}^i]_{j,k} = [\tilde{\mathbf{R}}]_{j+i,k+i}$. This way, the definition of $\mathcal{M}(\bar{\mathbf{s}}^i)$ in (3.10) can be rewritten as $\mathcal{M}(\bar{\mathbf{s}}^i) = \|\boldsymbol{\rho}^i - \tilde{\mathbf{R}}^i \bar{\mathbf{s}}^i\|^2$. Using the error vector $\mathbf{e}^i \triangleq \bar{\mathbf{s}}_\star^i - \bar{\mathbf{s}}_\top^i$ and the interference vector $\mathbf{n}^i \triangleq \boldsymbol{\rho}^i - \tilde{\mathbf{R}}^i \bar{\mathbf{s}}_\top^i$, we find

$$\begin{aligned} \mu^i &= \|\boldsymbol{\rho}^i - \tilde{\mathbf{R}}^i \bar{\mathbf{s}}_\top^i\|^2 - \|\boldsymbol{\rho}^i - \tilde{\mathbf{R}}^i \bar{\mathbf{s}}_\star^i\|^2 \\ &= \|\mathbf{n}^i\|^2 - \|\mathbf{n}^i - \tilde{\mathbf{R}}^i \mathbf{e}^i\|^2 \\ &= 2\mathbf{n}^{iT} \tilde{\mathbf{R}}^i \mathbf{e}^i - \|\tilde{\mathbf{R}}^i \mathbf{e}^i\|^2. \end{aligned} \quad (4.7)$$

Since the statistics of \mathbf{e}^i are difficult to characterize, we approximate \mathbf{e}^i by the simple error event most likely to occur at the i^{th} level, i.e., an error vector of the form $\mathbf{e}^i = [0, \dots, 0, \pm 1, 0, \dots, 0]^T$. The partial metric $\mathcal{M}(\bar{\mathbf{s}}^i) = \|\boldsymbol{\rho}^i - \tilde{\mathbf{R}}^i \bar{\mathbf{s}}^i\|^2$ suggests that this error will occur at the index of the “weakest” column of $\tilde{\mathbf{R}}^i$. Thus we assume $[\mathbf{e}^i]_l = \pm \delta_{l-l_i}$ for

$$l_i \triangleq \arg \min_l \|\tilde{\mathbf{r}}_l^i\|, \quad (4.8)$$

where $\tilde{\mathbf{r}}_l^i \in \mathbb{R}^{L-i}$ denotes the l^{th} column of $\tilde{\mathbf{R}}^i$. In this case,

$$\mu^i = \pm 2\mathbf{n}^{iT} \tilde{\mathbf{r}}_{l_i}^i - \|\tilde{\mathbf{r}}_{l_i}^i\|^2. \quad (4.9)$$

Recall from our discussion in Section 4.1.1 that the interference vector \mathbf{n} is zero-mean, white, and Gaussian in the case of ZF-GDFE preprocessing; and zero-mean, white, and non-Gaussian in the case of MMSE-GDFE preprocessing. In the latter

case, the non-Gaussianity of \mathbf{n} is due to a contribution from not-yet-detected PAM symbols, which we treat as random since their values are unknown when designing T_i . To proceed further, we approximate \mathbf{n} as Gaussian with covariance $\sigma^2 \mathbf{I}_L$. With these assumptions,

$$\mu^i \sim \mathcal{N}\left(-\|\tilde{\mathbf{r}}_{l_i}^i\|^2, 4\|\tilde{\mathbf{r}}_{l_i}^i\|^2\sigma^2\right). \quad (4.10)$$

Using the statistical description (4.10), we can solve for T_i in (4.6) given a particular ϵ_o . From Bayes rule we find

$$\Pr\{\mu^i > T_i | \mu^i > 0\} = \begin{cases} \frac{\Pr\{\mu^i > T_i\}}{\Pr\{\mu^i > 0\}} & T_i \geq 0, \\ 1 & \text{else,} \end{cases} \quad (4.11)$$

from which it is straightforward to show that

$$T_i = 2\sigma\|\tilde{\mathbf{r}}_{l_i}^i\| \mathcal{Q}^{-1}\left(\epsilon_o \mathcal{Q}\left(\frac{\|\tilde{\mathbf{r}}_{l_i}^i\|}{2\sigma}\right)\right) - \|\tilde{\mathbf{r}}_{l_i}^i\|^2 \quad (4.12)$$

using the tabulated function

$$\mathcal{Q}(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{x^2}{2}} dx. \quad (4.13)$$

From (4.12) we can see that the desired error probability ϵ_o is “weighted” by an SNR-dependent quantity; as SNR increases, so does the $\mathcal{Q}^{-1}(\cdot)$ term.

4.1.3 Channel Estimation

Here we propose a rank-reduced pilot-aided Wiener channel estimation scheme. We discuss the pilot pattern first and the estimation scheme later.

We choose a pilot pattern where one out of every $P \geq 2$ multicarrier symbols is used as a pilot. These pilot symbols are then used to estimate the channel coefficients of the $P - 1$ multicarrier data symbols in-between. Pilot patterns of this form are

relatively common, having been used in several other works (e.g., [13,62]). We choose this pattern over one where each multicarrier symbol contains a mixture of pilot and data sub-carriers for the following reason. Assuming a significant ICI radius equal to D , the pilot and data sub-carriers would interfere unless a frequency-domain guard with radius $2D$ was placed around each pilot tone. Since Nyquist sampling considerations imply the need for at least N_h pilot tones, prevention of pilot/data interference would require that at least $(4D + 1)N_h$ sub-carriers are spared from data transmission. For many applications of interest (e.g., the setup in Section 4.2), however, $(4D + 1)N_h > N$, making this scheme impractical. Since design of optimal pilot symbols appears to be a challenging problem, we used values obtained from a semi-exhaustive search.

We now define some quantities that follow from our pilot pattern. Say that, for all indices m corresponding to pilot symbols, we have $\bar{\mathbf{s}}^{(m)} = \mathbf{p}$. For these m , (2.15) implies that

$$\mathbf{y}^{(m)} = \mathbf{P}\mathbf{h}^{(m)} + \mathbf{w}^{(m)} \quad (4.14)$$

$$\mathbf{h}^{(m)} \triangleq \left[\text{diag}_{-D}(\bar{\mathbf{H}}^{(m)T}), \dots, \text{diag}_D(\bar{\mathbf{H}}^{(m)T}) \right]^T \in \mathbb{R}^{(2D+1)L} \quad (4.15)$$

$$\mathbf{P} \triangleq \left[\mathbf{J}^D \mathcal{D}(\mathbf{p}) \quad \dots \quad \mathbf{J}^{-D} \mathcal{D}(\mathbf{p}) \right], \quad (4.16)$$

where $\text{diag}_k(\cdot)$ extracts the k^{th} sub-diagonal of its matrix argument, i.e.,

$$\text{diag}_k(\bar{\mathbf{H}}) \triangleq \left[[\bar{\mathbf{H}}]_{k,0}, [\bar{\mathbf{H}}]_{k+1,1}, \dots, [\bar{\mathbf{H}}]_{k+L-1,L-1} \right]^T \quad (4.17)$$

with modulo- L indexing assumed. Recall that \mathbf{J} was defined in (4.3). Our goal is to estimate the local-ICI coefficients

$$\underline{\mathbf{h}}^{(m)} \triangleq [\mathbf{h}^{(m+1)T}, \dots, \mathbf{h}^{(m+P-1)T}]^T \quad (4.18)$$

from the pilot observations

$$\underline{\mathbf{y}}^{(m)} \triangleq [\mathbf{y}^{(m)T}, \mathbf{y}^{(m+P)T}]^T. \quad (4.19)$$

Say that $\mathbf{h}^{(m)} = \mathbf{C}\mathbf{g}^{(m)}$, where $\mathbf{g}^{(m)} \in \mathbb{C}^{N_\beta N_h}$ contains all complex-baseband time-domain impulse response coefficients that affect the m^{th} observation with N_β being the length of receive pulse b_n , and where \mathbf{C} is a function of the MCM pulse shapes $\{a_n\}$ and $\{b_n\}$.

The linear MMSE estimate of $\underline{\mathbf{h}}^{(m)}$ from $\underline{\mathbf{y}}^{(m)}$ is [83]

$$\hat{\underline{\mathbf{h}}}^{(m)} = \mathbf{R}_{hy} \mathbf{R}_{yy}^{-1} \underline{\mathbf{y}}^{(m)}, \quad (4.20)$$

where $\mathbf{R}_{hy} \triangleq \text{E}\{\underline{\mathbf{h}}^{(m)} \underline{\mathbf{y}}^{(m)T}\}$ and $\mathbf{R}_{yy} \triangleq \text{E}\{\underline{\mathbf{y}}^{(m)} \underline{\mathbf{y}}^{(m)T}\}$. We can write

$$\mathbf{R}_{hy} = \begin{bmatrix} \mathbf{R}_{hy}^{(1)} & \mathbf{R}_{hy}^{(1-P)} \\ \mathbf{R}_{hy}^{(2)} & \mathbf{R}_{hy}^{(2-P)} \\ \vdots & \vdots \\ \mathbf{R}_{hy}^{(P-1)} & \mathbf{R}_{hy}^{(-1)} \end{bmatrix} \quad (4.21)$$

$$\mathbf{R}_{yy} = \begin{bmatrix} \mathbf{R}_{yy}^{(0)} & \mathbf{R}_{yy}^{(-P)} \\ \mathbf{R}_{yy}^{(P)} & \mathbf{R}_{yy}^{(0)} \end{bmatrix} \quad (4.22)$$

with

$$\mathbf{R}_{hy}^{(q)} \triangleq \mathbf{C} \text{E}\{\mathbf{g}^{(m)} \mathbf{g}^{(m-q)H}\} \mathbf{C}^H \mathbf{P}^T \quad (4.23)$$

$$\mathbf{R}_{yy}^{(q)} \triangleq \mathbf{P} \mathbf{C} \text{E}\{\mathbf{g}^{(m)} \mathbf{g}^{(m-q)H}\} \mathbf{C}^H \mathbf{P}^T + \delta_q \sigma_z^2 \mathbf{I}_{2L} \quad (4.24)$$

Note that $\text{E}\{\mathbf{g}^{(m)} \mathbf{g}^{(m-q)H}\}$ is easily calculated from the time-domain channel autocorrelation function.

Because each of the $2N_h$ real-valued channel taps changes slowly over the pilot/data/pilot interval (i.e., $N_\beta + PN$ channel uses), it contributes only $K = 1 + \lceil 2f_D T_c (N_\beta + PN) \rceil$ non-negligible singular values to $\mathbf{R}_{hy} \mathbf{R}_{yy}^{-1}$. Thus, as in [13], optimal rank reduction [83] can be used to significantly reduce the complexity of channel

estimation with little performance degradation. The optimal rank- $2N_hK$ estimate of $\underline{\mathbf{h}}^{(m)}$ is constructed as follows [83]. From the singular value decomposition (SVD)

$$\mathbf{R}_{hy}\mathbf{R}_{yy}^{-1} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \quad (4.25)$$

we build \mathbf{U}_K and \mathbf{V}_K from the first $2N_hK$ columns of \mathbf{U} and \mathbf{V} , respectively, and we build $\mathbf{\Sigma}_K$ from the first $2N_hK$ rows and columns of $\mathbf{\Sigma}$. We find that

$$\mathbf{R}_{hy}\mathbf{R}_{yy}^{-1} \approx \mathbf{U}_K\mathbf{F}_K^H \quad (4.26)$$

for $\mathbf{U}_K \in \mathbb{R}^{(P-1)(2D+1)L \times 2N_hK}$ and $\mathbf{F}_K \triangleq \mathbf{V}_K\mathbf{\Sigma}_K \in \mathbb{C}^{2L \times 2N_hK}$. Note that \mathbf{U}_K can be interpreted as the MMSE-optimal order- $2N_hK$ basis expansion for $\underline{\mathbf{h}}^{(m)}$ and \mathbf{F}_K^H can be interpreted as the linear MMSE estimator of the corresponding basis coefficients $\underline{\boldsymbol{\lambda}}^{(m)}$. The resulting rank-reduced estimation procedure

$$\hat{\underline{\boldsymbol{\lambda}}}^{(m)} = \mathbf{F}_K^H \underline{\mathbf{y}}^{(m)} \quad (4.27)$$

$$\hat{\underline{\mathbf{g}}}^{(m)} = \mathbf{U}_K \hat{\underline{\boldsymbol{\lambda}}}^{(m)} \quad (4.28)$$

requires only $2N_hK[2L + (P-1)(2D+1)L]$ complex MACs per $P-1$ frames. In Section 4.2 we demonstrate that, with $K=2$, the complexity of this channel estimation method is on par with that of preprocessed TSA. Experiments have confirmed that the rank-reduced performance is nearly indistinguishable from the full-rank performance [84].

4.2 Numerical Results

Our experiments employed the ICI/ISI-corrupted MCM system specified in the real-valued form by (2.15). Uncoded QPSK symbols $\{\bar{s}_k^{(m)}\}_{k=0}^{N-1}$ (i.e., $Q=2$) were communicated over $N=64$ MCM subcarriers (i.e., $L=128$), and the demodulator

outputs $\bar{\mathbf{y}}^{(m)}$ were used to detect the QPSK sequence $\bar{\mathbf{s}}^{(m)}$. For sequence detection (SD), we focused on the case where only the “significant” ICI coefficients $\bar{\mathbf{H}}^{(m)}$ were known, in which case ISI and residual ICI were treated as unknown interference.

Several methods of SD were examined: MLSD, near-ML TSA, and MMSE-DFE. In each case, we first apply circular ordering and fast MMSE-GDFE preprocessing to arrive at the detection problem (3.7), since, in the case of uncoded QPSK, solutions to (3.7) are known to be ML as shown in Section 3.3. For MLSD, we solve (3.7) via SE-SpD, while for near-ML TSA, we obtain an approximate solution to (3.7) via suboptimal tree search. For MMSE-DFE, we decode the bits $\{\bar{s}_k^{(m)}\}_{k=0}^{L-1}$ in the order $\bar{s}_{L-1}^{(m)}, \bar{s}_{L-2}^{(m)}, \dots, \bar{s}_0^{(m)}$ by first making a hard decision on each bit and then subtracting its (estimated) contribution from $\bar{\mathbf{y}}^{(m)}$ [72].

We assumed a wide sense stationary uncorrelated scattering (WSSUS) Rayleigh fading channel [85] whose realizations were generated using Jakes method. The channel had a uniform delay-profile with normalized⁶ delay spread $N_h = T_h/T_c = 16$ and a normalized single-sided Doppler spread $f_D T_c \in \{0.001, 0.003\}$. These parameters correspond to, e.g., a system with subcarrier spacing $F_s = 20$ kHz, carrier frequency $f_c = 10$ GHz, delay spread $T_h = 12.25 \mu\text{s}$, and effective⁷ velocities of 138 km/hr and 414 km/hr, respectively. We defined SNR as the ratio of signal energy to noise energy in (pulse-shaped and sampled) receiver inputs.

Four FFT-based MCM schemes were considered: CP-OFDM [7], ZP-OFDM [8], Strohmer and Beaver’s “optimal” OFDM (S-OFDM) [9], and PS-MCM [5, 17]. Each

⁶These quantities are normalized to the “channel-use interval” or “chip interval” $T_c = \frac{1}{N F_s}$.

⁷Effective velocity v can result from mobile velocity $v/3$ in, e.g., a “triple-Doppler” situation, when the downlink signal bounces off of a reflector traveling directly towards (away from) the base station at velocity $v/3$ and is received by a mobile traveling away from (towards) the base station at velocity $v/3$.

of these schemes was allowed the same transmitted energy per information bit. With the exception of ZP-OFDM, all guarantee white subchannel noise samples (i.e., $E\{\bar{\mathbf{w}}^{(m)}\bar{\mathbf{w}}^{(m)T}\} = \sigma^2/2\mathbf{I}$). For CP-OFDM and ZP-OFDM, we employed a length- $N_g = 16$ guard to avoid ISI, yielding a spectral efficiency of 0.8 QPSK-symbols/sec/Hz. For S-OFDM, $N = 64$ QPSK symbols were transmitted every 80 channel uses, also yielding a spectral efficiency of 0.8 QPSK-symbols/sec/Hz. For PS-MCM, N QPSK symbols were transmitted every N channel uses, yielding a spectral efficiency of 1 QPSK-symbol/sec/Hz. The dilation factor σ of the Gaussian pulse $g^\sigma(t)$ orthogonalized for S-OFDM (see [9, p. 1114]) was numerically optimized to minimize the total power of ICI plus ISI. The PS-MCM transmitter pulse $\{a_n\}$ was length $N_a = \frac{3}{2}N$ and the PS-MCM receiver pulse $\{b_n\}$ was length- N rectangular and preceded by $\frac{3}{2}N_h$ zeros.

4.2.1 FER Performance

Figure 4.2 examines the frame⁸ error rate (FER) performance of the four MCM schemes with MLSD. When MLSD was too costly, the matched filter bound (MFB) was used as an approximation. When the MLSD has perfect *global*-ICI knowledge [i.e., knowledge of full $\bar{\mathbf{H}}^{(m)}$], PS-MCM and ZP-OFDM performed similarly, and significantly outperformed S-OFDM and CP-OFDM. S-OFDM performed poorly due to a high level of ISI. Better S-OFDM performance was observed when the dilation factor was chosen to decrease ISI and increase ICI (which is incorporated in MLSD), but, since that was inconsistent with the S-OFDM design methodology in [9], we do not present those results here. CP-OFDM suffers from high FER because it wastes energy on a CP that is discarded by the receiver, and because CP-OFDM does not

⁸We use the terms “frame” and “multicarrier symbol” interchangeably in this paper.

make for easy extraction of delay diversity with uncoded transmissions. When using MLSD with global ICI knowledge, all MCM schemes other than S-OFDM benefit from additional Doppler diversity at higher $f_D T_c$. S-OFDM, in contrast, reacts to the $f_D T_c$ increase by dilating the pulse in such a way that both ICI and ISI increase, but the increase in ISI hurts more than the increase in ICI helps. In comparing the schemes, it is important to remember that PS-MCM operates at 25% higher spectral efficiency than the other schemes.

When the receiver has only *local*-ICI knowledge up to ± 3 subcarriers [i.e., knowledge of $2D + 1 = 13$ diagonals of $\bar{\mathbf{H}}^{(m)}$], Fig. 4.2 shows that the MLSD performance of PS-MCM is indistinguishable from that with global-ICI knowledge. This confirms that PS-MCM suppresses non-local ICI well below the noise floor over the SNR range of interest. In contrast, the MLSD performance of ZP-OFDM and CP-OFDM collapse when only the local ICI is known; while S-OFDM avoids this collapse, it does so at the expense of high ISI power. Note that [1], which applies Viterbi SD to CP-OFDM under local-ICI knowledge, is lower-bounded by the “CP-OFDM MFB D=6” trace, and hence performs far worse than the proposed PS-MCM scheme.

Figure 4.3 examines FER performance of the four MCM schemes under MMSE-DFE detection. It is interesting to note that, when the MMSE-DFE detector is given perfect global-ICI knowledge, the PS-MCM FER floors at high SNR. This is consistent with [15], which showed similar MMSE-DFE performance for max-SINR *reception*-pulse (MSRP) MCM. We conjecture that max-SINR pulse shaped schemes are more prone to DFE error because the channel energy is not as well concentrated in the main diagonal of $\tilde{\mathbf{R}}$ as it is for CP-OFDM, ZP-OFDM, and S-OFDM. When the MMSE-DFE has only *local*-ICI knowledge up to ± 3 subcarriers, the FER performances of

ZP-OFDM and CP-OFDM collapse, while the performance of PS-OFDM remains the same as that with global-ICI knowledge. As before, S-OFDM avoids this collapse, though at the cost of high ISI power. Once again, this confirms that PS-MCM suppresses non-local ICI well below the noise floor over the SNR range of interest. Note, however, that in all cases, the MMSE-DFE performances are significantly worse than their MLSD counterparts in Fig. 4.2. Fortunately, there is no need to use MMSE-DFE detection on PS-MCM because (as we shall see) TSA can be used to achieve near-ML performance with MMSE-DFE-like complexity.

Figure 4.4 shows the FER performance of various TSAs operating on DS-channel PS-MCM with perfect knowledge of ± 3 subcarriers of local ICI (i.e., $D = 6$). For the M-algorithm, we set $M = 8$, and, for the T-algorithm, we set $T = 0.4$ when $f_D T_c = 0.001$ and $T = 0.5$ when $f_D T_c = 0.003$. For both algorithms, these were the computationally cheapest parameter settings which guaranteed near-ML performance over the SNR range of interest. For the adaptive T-algorithm, we set $\epsilon_o = 10^{-5}$ and limited the maximum list size to 8. For the Fano algorithm of [70], we set⁹ the bias $b = \sigma^2/2$ and the “step size” = σ^2 . For the Schnorr-Euchner’s sphere decoder (SE-SpD), we initialized the squared sphere radius at 1.2 times the average squared Euclidean distance between the observation and the ML point. From Fig. 4.4 we see that, with the exception of MMSE-DFE, all TSAs give near-ML performance. In fact, the ML and SE-SpD traces are identical since MMSE-GDFE preprocessed SE-SpD yields ML performance with uncoded QPSK (i.e. see Section 3.3). The MMSE-DFE error floor is consistent with that observed in [15] for MSRP-MCM.

⁹These settings were personally suggested by the authors of [70].

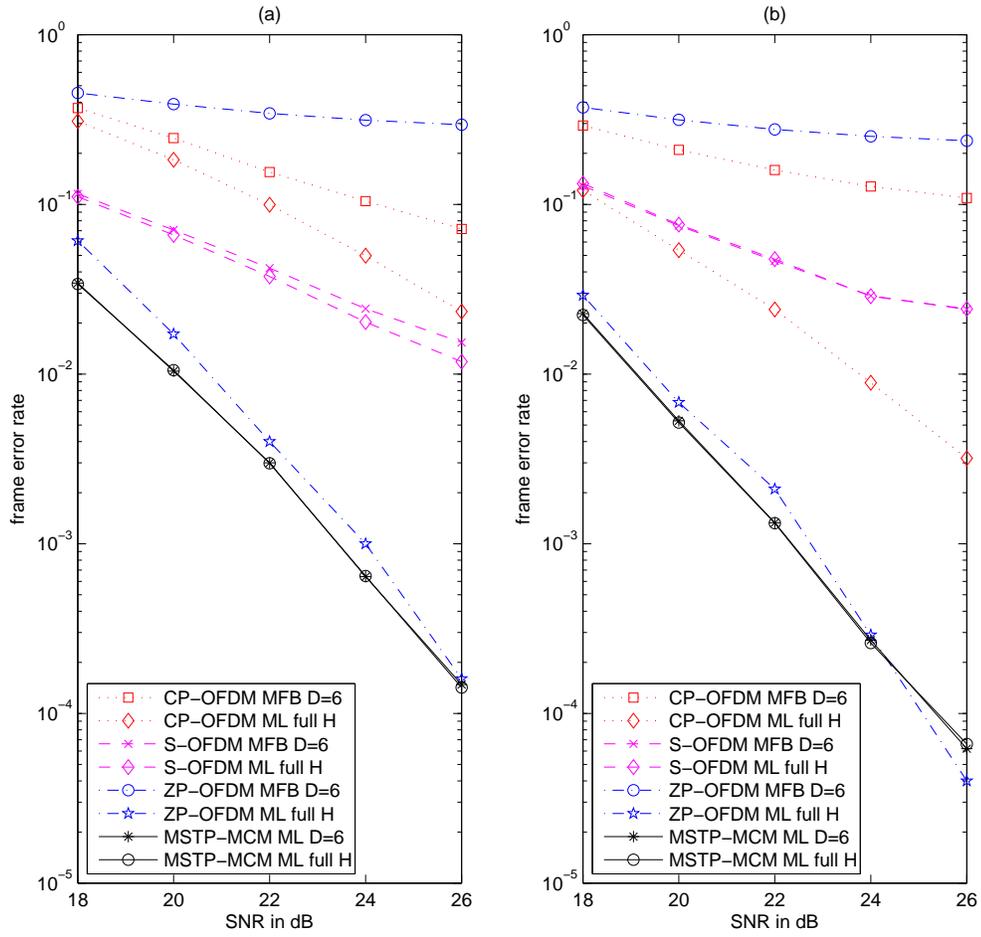


Figure 4.2: ML and MFB performance of several MCM schemes using global ICI (“full H”) or local ICI (“D=6”) at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$.

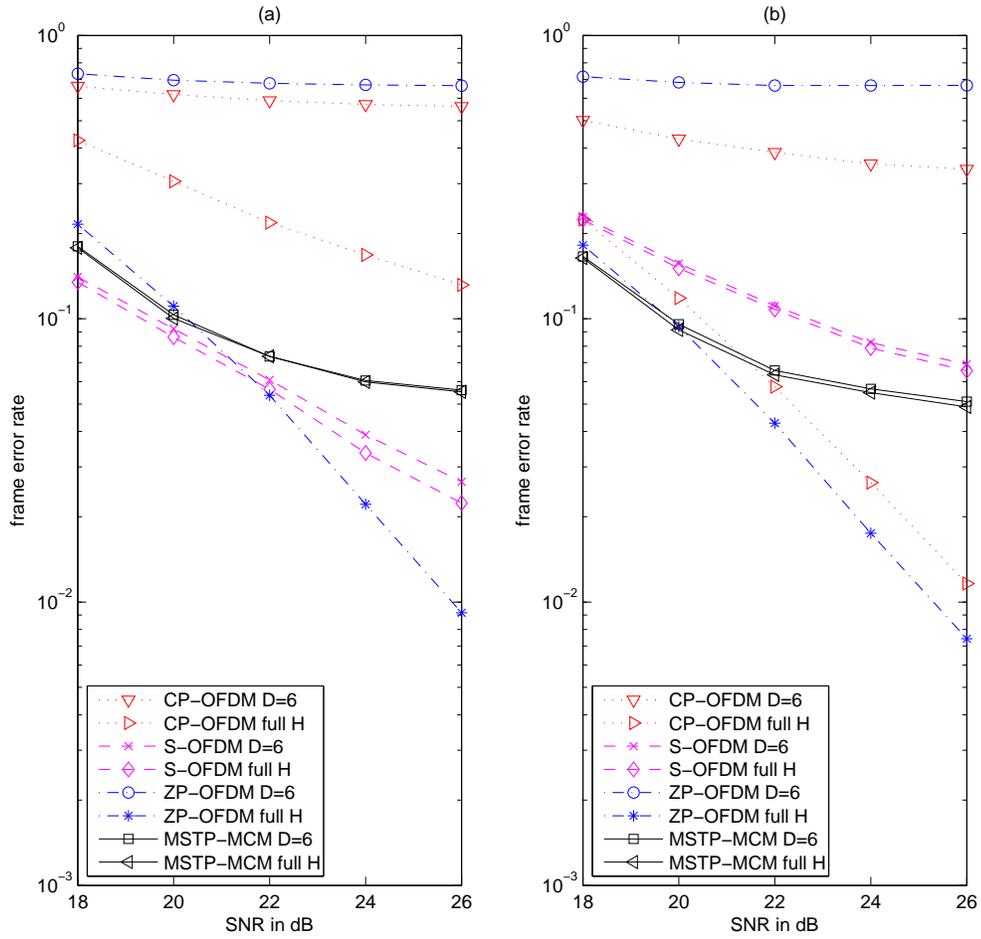


Figure 4.3: MMSE-DFE performance of several MCM schemes using global ICI (“full H”) or local ICI (“D=6”) at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$.

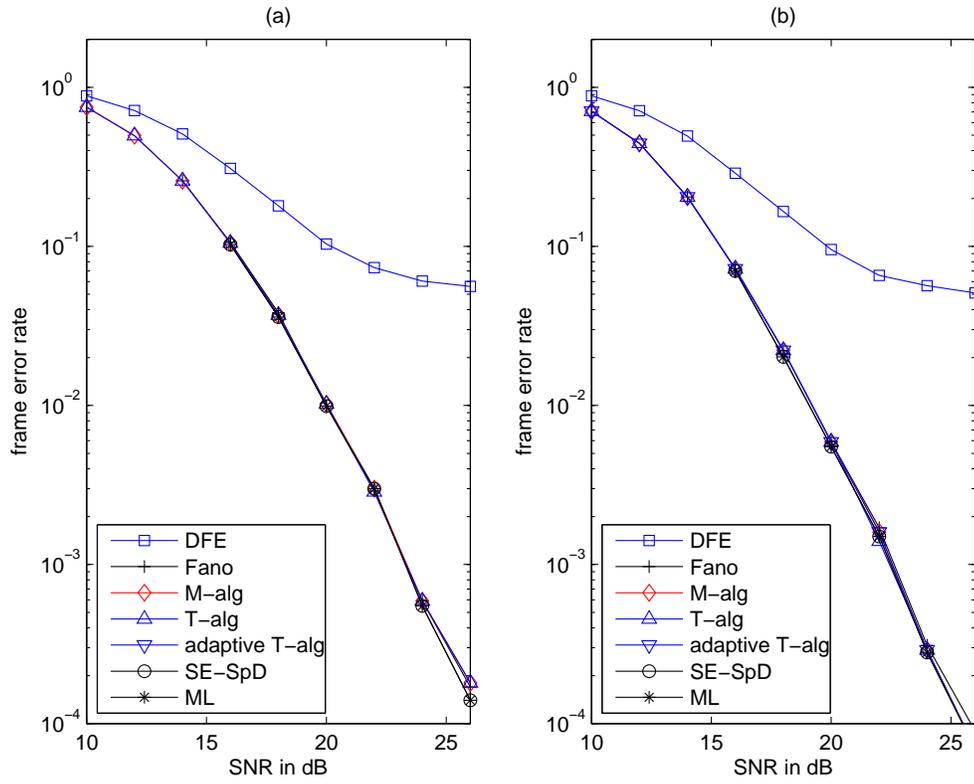


Figure 4.4: Performance of several TSAs on doubly selective PS-MCM with perfect knowledge of local ICI (i.e., $D = 6$) at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$.

4.2.2 Complexity

Figure 4.5 compares the average complexity of the TSAs and the Viterbi algorithm used in [1] when operating on PS-MCM¹⁰ with perfect knowledge of ± 3 subcarriers of local ICI (i.e., $D = 6$). Here “complexity” is measured in real MAC operations per decoded frame and is plotted on a log (base- L) scale, as in other near-ML TSA studies (e.g., [69, 70]). For the TSAs, we plot the average number of operations required to achieve the FER performance of Fig. 4.4, including that required for PS-MCM demodulation, circular ordering, and fast MMSE-GDFE preprocessing.

For the reasons discussed in Section 3.2.1, both the SE-SpD and Fano algorithms exhibit reasonable complexity at high SNR but explosive complexity at low SNR. As expected, the M-algorithm has the same complexity at all SNRs. Interestingly, the T-algorithm has almost the same complexity as the M-algorithm. Remarkably, the adaptive T-algorithm yields DFE-like complexity at high SNR and complexity that is no higher than that of the M- and T-algorithms at low SNR. As discussed in Section 4.1.2, the adaptive T-algorithm’s excellent low-SNR complexity is a consequence of its BrFS nature, while its excellent high-SNR complexity is a consequence of the fact that it uses channel knowledge to intelligently guide the search process.

The $D = 6$ Viterbi complexity is much larger than that of the near-ML BrFS algorithms and the MMSE-DFE. Furthermore, the Viterbi complexity plotted in Fig. 4.5 (i.e., $L(2D + 1)Q^{(4D+1)} = L^{3.39}$) is valid only when where D edge subcarriers are inactive. For our case, where all subcarriers are active, the “tail-biting” Viterbi approach proposed in Chapter 3 would be an admissible MLSD with complexity

¹⁰Since Viterbi complexity is invariant to the channel realization, the values reported in Fig. 4.5 apply equally well to CP-OFDM transmission, as was used in [1].

$L(2D + 1)Q^{(4D+1)} = L^{5.10}$ (which is well outside the range of Fig. 4.5). In conclusion, Fig. 4.5 shows that, by sacrificing a fraction-of-a-dB in performance relative to MLSD, TSA can be implemented with near-MMSE-DFE complexity, even when all subcarriers are active.

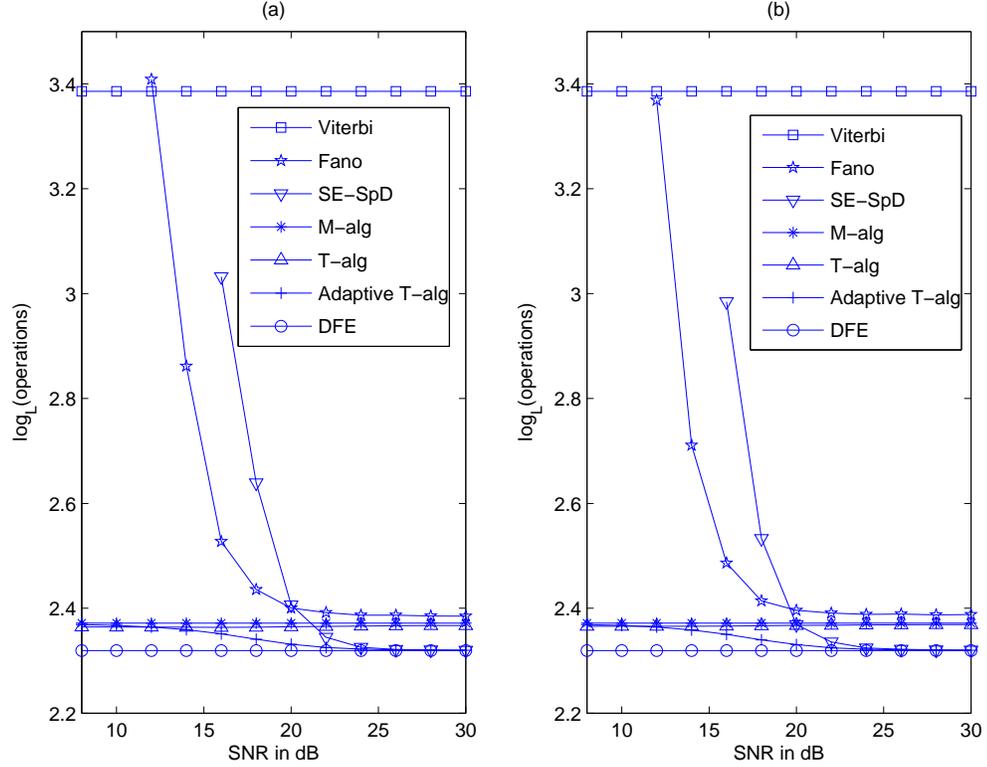


Figure 4.5: Number of real MAC operations per frame for doubly selective PS-MCM reception at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$. Viterbi complexity above assumes inactive edge subcarriers (as in [1]); for active subcarriers, the tail-biting version would require $L^{5.10}$ MACs/frame.

4.2.3 Effect of Imperfect Channel Estimates

In Figures 4.6-4.7 we investigate the effect of channel estimation error on TSA performance in DS-channel PS-MCM. For this, we use the rank- $2N_hK$ pilot-aided Wiener estimator proposed in Section 4.1.3 with $K = 2$. With pilot-spacing $P = 2$, Fig. 4.6 shows that imperfect channel estimates yield an SNR loss of about 1dB at $f_D T_c = 0.001$ and about 2dB at $f_D T_c = 0.003$ for all the TSAs examined, i.e., MMSE-DFE, the M-algorithm, the adaptive T-algorithm, and MLSD. For our parameter choices, the channel estimation complexity reported in Section 4.1.3 translates into about $L^{2.4}$ real MACs per frame, which is comparable to the complexity of pre-processed TSA. With pilot-spacing $P = 3$, Fig. 4.7 shows that imperfect channel estimates yield a significantly higher SNR loss, especially at $f_D T_c = 0.003$, making this choice impractical. A more thorough investigation of reduced-rank pilot-aided Wiener channel estimation appears in [84].

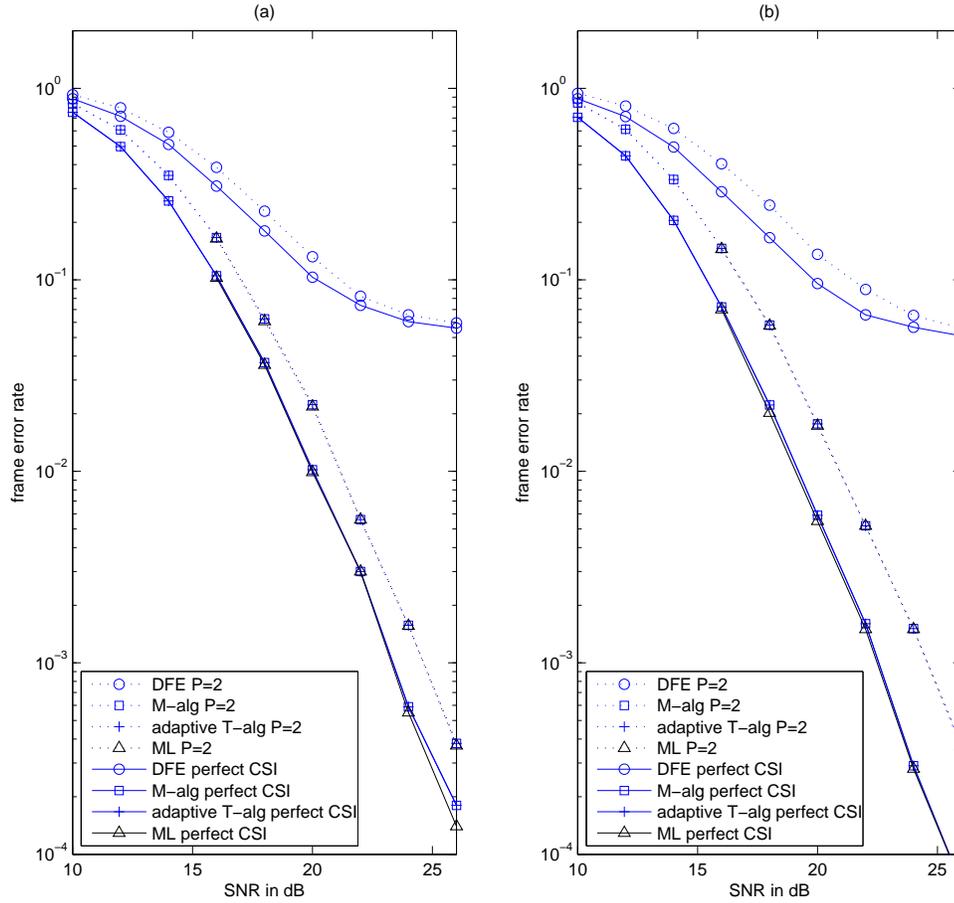


Figure 4.6: Performance of several TSAs operating on doubly selective PS-MCM with rank-reduced pilot-aided estimates of local-ICI (i.e., $D = 6$) at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$, using pilot spacing $P = 2$

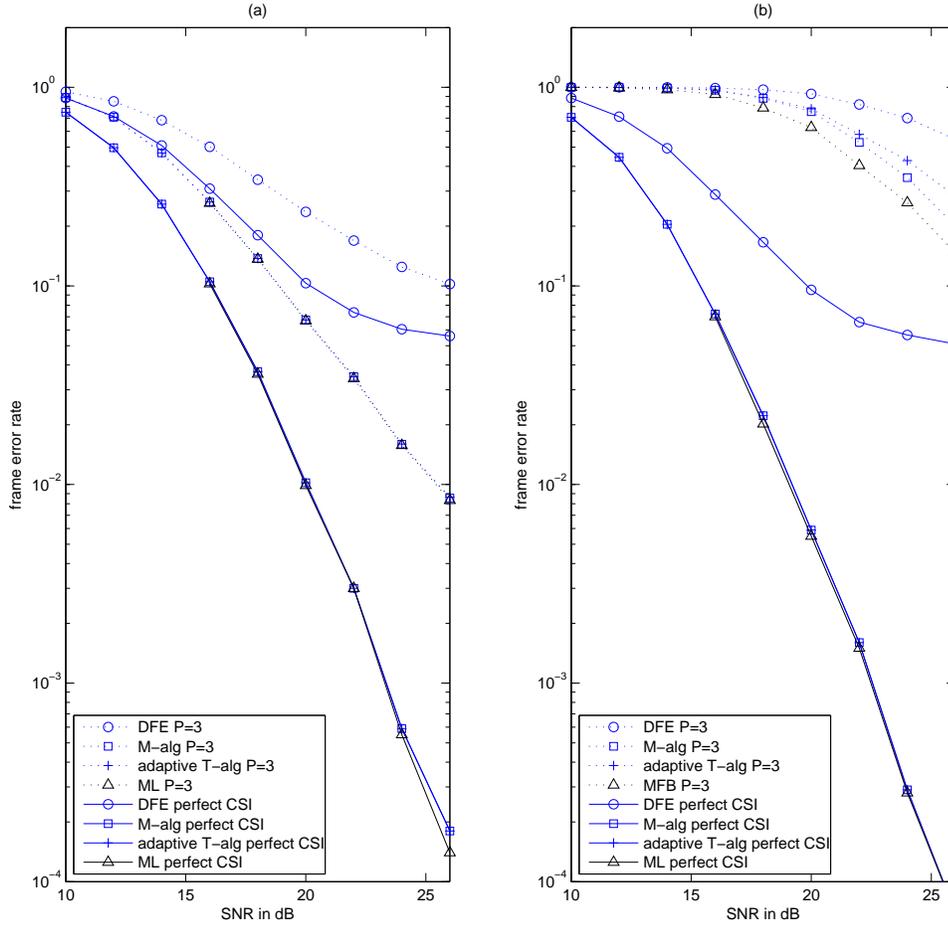


Figure 4.7: Performance of several TSAs operating on doubly selective PS-MCM with rank-reduced pilot-aided estimates of local-ICI (i.e., $D = 6$) at (a) $f_D T_c = 0.001$; (b) $f_D T_c = 0.003$, using pilot spacing $P = 3$

CHAPTER 5

NEAR-OPTIMAL NONCOHERENT EQUALIZATION

In this chapter, we approach the problem of near-optimal soft noncoherent equalization via basis expansion (BE) channel modeling and soft TSA. Our use of generic BE modeling [41, 58, 86] described in Chapter 2 allows an efficient and unified treatment of different channels (e.g., time-variant ISI channels, frequency-variant ICI channels, and sparse versions of those channels), and our use of soft TSA leverages recent ideas from the flat-fading multiple-input multiple-output (MIMO) literature (e.g., [87, 88]) that facilitate an efficient tradeoff between performance and complexity. Furthermore, when BE channel modeling is used, we reason that TSA is more appropriate than trellis-search, since channel parameters have a global rather than local influence.

The contributions of this chapter are as follows. Within the framework of generic BE channel modeling, we first derive the optimal soft noncoherent equalizer, which requires the computation of a noncoherent metric for every possible bit sequence. Next, we show that each these metrics can be recursively computed, meanwhile establishing links to per-sequence BE-coefficient estimation. Due to the high complexity of optimal equalization, we then propose a practical near-optimal equalizer based on the M-algorithm, with $\mathcal{O}(NN_b^2N_H^2)$ complexity, where N denotes the block length and N_bN_H

the number of BE coefficients. Taking a different approach, we then show that the Bayesian EM (EMB) algorithm [28] can also be used to design practical near-optimal BE-based soft noncoherent equalizers. For this, we first derive the EMB iteration, which alternates between soft coherent MAP equalization and a particular form of soft BE-coefficient estimation. For the former, we propose $\mathcal{O}(NN_bN_H)$ -complexity TSA as an alternative to conventional trellis-based processing, and for the latter, we propose a low-complexity ($\mathcal{O}(NN_b^2N_H^2)$) sequential-Bayes scheme [83].

Within the existing literature, we are not aware of other sequential or EM-based near-optimal soft noncoherent equalizers for BE-modeled DS channels that are suitable for turbo reception (apart from our preliminary work [89, 90]). However, BE-based channel estimators have been used in non-turbo receivers, and they have been combined in “ad hoc” ways with soft coherent equalizers for turbo reception, as described below. In [30], Antón-Haro, Fonollosa, and Fonollosa proposed a scheme for EM-based blind channel tracking in uncoded GMSK systems using a polynomial BE channel model. Their scheme does not easily generalize to generic BE models, generic codes, and pilot-aided transmission, though. In [52], El-Mahdy proposed a noncoherent sequence detector which used the Viterbi algorithm with per-path BE-coefficient tracking. It is not applicable to turbo reception, though, and BER performance shows an error floor. In [91], Cui and Tellambura proposed a polynomial BE-based sequence detection scheme for ICI-corrupted OFDM. Their technique uses an ad-hoc alternation of (hard) equalization and (hard) channel estimation and is incompatible with turbo reception. More recently [40], Fang, Rugini, and Leus combined a soft coherent equalizer (based on linear MMSE [33]) with a soft BE-modelled channel estimator for turbo equalization of ICI-corrupted OFDM. Unlike our (EM-derived) soft

BE-coefficient estimator, which applies BE in the frequency domain and costs only $\mathcal{O}(NN_b^2N_H^2)$, they apply BE in the time domain, with a cost of $\mathcal{O}(N^3)$. A related turbo equalization/channel-estimation scheme for ICI-corrupted OFDM is the one recently proposed by Liu and Fitz [36, 39]. It iterates between soft coherent equalization, which uses a reduced-state approximation of optimal trellis-based MAP, and soft channel estimation, which uses periodic insertion of OFDM symbols with 100% pilot subcarriers, and SVDs, and requires $\approx KN^2$ operations, for very large K (e.g., $K = 31360$ in the simulations).

In Section 5.1 we describe the system model, in Section 5.2 the optimal soft noncoherent equalizer and its sequential approximation, and in Section 5.3 the EM-based equalizer. In Section 5.4 we discuss implementational details, in Section 5.5 we present numerical results.

5.1 Basis Expansion Modeling

The equalizer employs an N_b -term BE model for the variation of the composite channel over the block. Recalling (2.6), we rewrite the BE model with the d^{th} “cyclic” diagonal of $\mathbf{H}^{(j)}$, i.e., $\mathbf{h}_d^{(j)} \triangleq [[\mathbf{H}^{(j)}]_{0,-d}, [\mathbf{H}^{(j)}]_{1,1-d}, \dots, [\mathbf{H}^{(j)}]_{N-1,N-1-d}]^T$, as

$$\mathbf{h}_d^{(j)} = \mathbf{B}\boldsymbol{\eta}_d^{(j)}, \quad d = 0, \dots, N_H - 1, \quad (5.1)$$

where $\mathbf{B} \in \mathbb{C}^{N \times N_b}$ is a matrix of basis vectors and $\boldsymbol{\eta}_d^{(j)} \in \mathbb{C}^{N_b}$ is a vector of BE coefficients. Note that the approximation in (5.1) can be made arbitrarily accurate via large enough N_b . With single-carrier modulation, the BE models channel variation in the time domain, so that $N_b = N_D$ suffices (with appropriate choice of \mathbf{B} and α). With multi-carrier modulation, the BE models channel variation in the frequency domain, so that $N_b = N_h$ suffices, with \mathbf{B} being a truncated DFT matrix [41]. In

either case, $N_b N_H = N_h N_D$. Under the BE model (5.1), the received vector $\mathbf{y}^{(j)}$ from (2.6) becomes

$$\mathbf{y}^{(j)} = \mathbf{A}^{(j)} \boldsymbol{\theta}^{(j)} + \mathbf{z}^{(j)}, \quad (5.2)$$

where $\boldsymbol{\theta}^{(j)} \triangleq [\boldsymbol{\eta}_0^{(j)T}, \dots, \boldsymbol{\eta}_{N_H-1}^{(j)T}]^T \in \mathbb{C}^{N_b N_H}$ and

$$\mathbf{A}^{(j)} \triangleq [\mathcal{D}_0(\mathbf{s}^{(j)})\mathbf{B}, \dots, \mathcal{D}_{N_H-1}(\mathbf{s}^{(j)})\mathbf{B}]. \quad (5.3)$$

The receiver infers the information bits $\{b_m^{(j)}\}$ using the “turbo” principle: “soft” information on the coded bits $\mathbf{x}^{(j)}$, in the form of log-likelihood ratios (LLRs), is iteratively refined through alternating soft-equalization and soft-decoding steps, as shown in Fig. 5.1. The equalizer’s task is to produce extrinsic LLRs given the observation $\mathbf{y}^{(j)}$ and the prior LLRs provided by the decoder (or, in the first turbo iteration, from pilots).

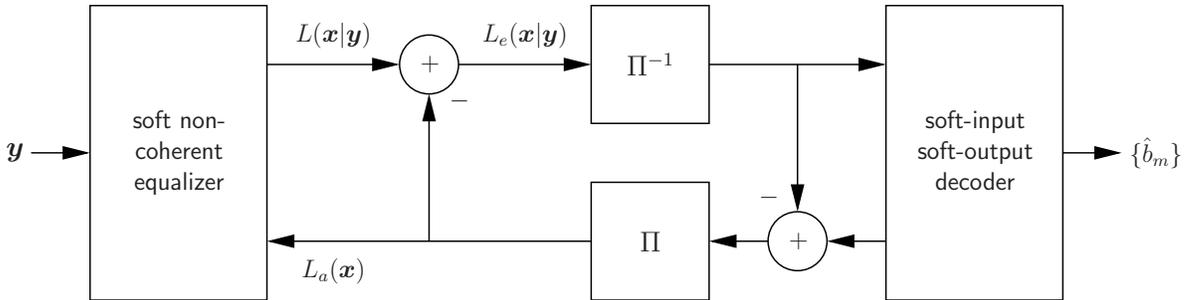


Figure 5.1: Turbo receiver with soft noncoherent equalizer.

The equalizers we propose are “noncoherent” in that they treat the channel realization $\boldsymbol{\theta}^{(j)}$ as unknown. They treat the statistics as known, however, and assume that $\mathbf{w}^{(j)} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$ and $\boldsymbol{\theta}^{(j)} \sim \mathcal{CN}(\bar{\boldsymbol{\theta}}^{(j)}, \mathbf{R}_\theta)$ for full rank \mathbf{R}_θ . The selection of $\bar{\boldsymbol{\theta}}^{(j)}$ and \mathbf{R}_θ is discussed in Section 5.4.1.

In Section 5.2, we describe the optimal noncoherent equalizer and a practical implementation based on TSA, and in Section 5.3 we describe equalization based on the Bayesian EM algorithm. Because the equalization procedure is invariant to block index j , we suppress the “ (j) ” notation in the sequel.

5.2 Sequential Noncoherent Equalization

5.2.1 Optimum Soft Noncoherent Equalization

The log-likelihood ratio (LLR) of coded bit x_k given \mathbf{y} , i.e.,

$$L(x_k|\mathbf{y}) \triangleq \ln \frac{\Pr[x_k = 1|\mathbf{y}]}{\Pr[x_k = 0|\mathbf{y}]}, \quad k \in \{0, \dots, N_s q - 1\}, \quad (5.4)$$

can be written in the form [87]

$$L(x_k|\mathbf{y}) = \ln \frac{\sum_{\mathbf{x}:x_k=1} p(\mathbf{y}|\mathbf{x}) \exp \mathbf{l}^T \mathbf{x}}{\sum_{\mathbf{x}:x_k=0} p(\mathbf{y}|\mathbf{x}) \exp \mathbf{l}^T \mathbf{x}}, \quad (5.5)$$

where $\mathbf{l} \triangleq [L_a(x_0), \dots, L_a(x_{N_s q})]^T$ such that $L_a(x_k) \triangleq \ln(P[x_k = 1]/P[x_k = 0])$ is the *a priori* LLR of x_k . The “extrinsic” LLR $L_e(x_k|\mathbf{y}) \triangleq L(x_k|\mathbf{y}) - L_a(x_k)$ then becomes

$$L_e(x_k|\mathbf{y}) = \ln \frac{\sum_{\mathbf{x}:x_k=1} \exp \mu(\mathbf{x})}{\sum_{\mathbf{x}:x_k=0} \exp \mu(\mathbf{x})} - L_a(x_k) \quad (5.6)$$

using the noncoherent MAP sequence metric

$$\mu(\mathbf{x}) \triangleq \ln p(\mathbf{y}|\mathbf{x}) + \mathbf{l}^T \mathbf{x}. \quad (5.7)$$

Since $\boldsymbol{\theta}$ and \mathbf{z} in (5.2) are both Gaussian distributed, we have

$$\mathbf{y}|\mathbf{x} \sim \mathcal{CN}(\mathbf{A}\bar{\boldsymbol{\theta}}, \mathbf{A}\mathbf{R}_\theta\mathbf{A}^H + \sigma^2\mathbf{I}_N), \quad (5.8)$$

where \mathbf{A} depends on the coded bits \mathbf{x} through the corresponding symbols \mathbf{s} . Thus, with $\boldsymbol{\Phi} \triangleq \mathbf{A}\mathbf{R}_\theta\mathbf{A}^H + \sigma^2\mathbf{I}_N$, we get

$$\mu(\mathbf{x}) = -\|\mathbf{y} - \mathbf{A}\bar{\boldsymbol{\theta}}\|_{\boldsymbol{\Phi}^{-1}}^2 - \ln(\pi^N \det \boldsymbol{\Phi}) + \mathbf{l}^T \mathbf{x}. \quad (5.9)$$

The sequence metrics $\mu(\mathbf{x})$ can be evaluated using an N_s -stage Q -ary tree, where, the partial metrics

$$\mu(\mathbf{x}_n) \triangleq \ln p(\mathbf{y}_n | \mathbf{x}_n) + \mathbf{l}_n^T \mathbf{x}_n \quad (5.10)$$

are evaluated recursively. In (5.10), $\mathbf{x}_n \triangleq [\underline{\mathbf{x}}_0^T, \dots, \underline{\mathbf{x}}_n^T]^T$ with $\underline{\mathbf{x}}_i \triangleq [x_{iq}, \dots, x_{iq+q-1}]^T$, $\mathbf{l}_n \triangleq [\underline{l}_0^T, \dots, \underline{l}_n^T]^T$ with $\underline{l}_i \triangleq [L_a(x_{iq}), \dots, L_a(x_{iq+q-1})]^T$, and $\mathbf{y}_n \triangleq [y_0, \dots, y_n]^T$. Note that $\underline{\mathbf{x}}_i$ and \underline{l}_i correspond to the i^{th} symbol. The recursion is derived in Appendix B.1 and summarized in Table 5.1, where \mathbf{b}_n^H denotes the n^{th} row of \mathbf{B} . It is straightforward to show that each recursion consumes $N_b^2 N_H^2 + 3N_b N_H + 7$ multiplications.

Table 5.1: Fast recursion for $\mu(\mathbf{x}_n)$

<p>from the old quantities: $\mu(\mathbf{x}_{n-1}), \hat{\boldsymbol{\theta}}_{n-1}, \boldsymbol{\Sigma}_{n-1}^{-1}, [s_{n-1}, \dots, s_{n-N_H+1}]$, and the inputs: $y_n, s_n, \underline{l}_n, \underline{\mathbf{x}}_n$, calculate the new quantities: $\mu(\mathbf{x}_n), \hat{\boldsymbol{\theta}}_n, \boldsymbol{\Sigma}_n^{-1}, [s_n, \dots, s_{n-N_H+2}]$, using the recursion: $\mathbf{a}_n = [s_n \mathbf{b}_n^H, \dots, s_{n-N_H+1} \mathbf{b}_n^H]^H$ $\mathbf{d}_n = \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{a}_n$ $\alpha_n = (1 + \mathbf{a}_n^H \mathbf{d}_n)^{-1}$ $e_n = y_n - \mathbf{a}_n^H \hat{\boldsymbol{\theta}}_{n-1}$ $\boldsymbol{\Sigma}_n^{-1} = \boldsymbol{\Sigma}_{n-1}^{-1} - \alpha_n \mathbf{d}_n \mathbf{d}_n^H$ $\mu(\mathbf{x}_n) = \mu(\mathbf{x}_{n-1}) - \frac{\alpha_n}{\sigma^2} e_n ^2 + \ln\left(\frac{\alpha_n}{\pi \sigma^2}\right) + \underline{l}_n^T \underline{\mathbf{x}}_n$ $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_{n-1} + \alpha_n e_n \mathbf{d}_n$, initializing (iff $n = 0$) with: $\mu(\mathbf{x}_{-1}) = 0, \hat{\boldsymbol{\theta}}_{-1} = \bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{-1}^{-1} = \sigma^{-2} \mathbf{R}_\theta$.</p>
--

In the Table 5.1, the quantity $\hat{\boldsymbol{\theta}}_n$ can be written (see Appendix B.1):

$$\hat{\boldsymbol{\theta}}_n = \bar{\boldsymbol{\theta}} + \mathbf{R}_\theta \mathbf{A}_n^H \boldsymbol{\Phi}_n^{-1} (\mathbf{y}_n - \mathbf{A}_n \bar{\boldsymbol{\theta}}), \quad (5.11)$$

which can be recognized as the \mathbf{x}_n -conditional MMSE estimate of $\boldsymbol{\theta}_n$ from \mathbf{y}_n . Using this fact, Appendix B.2 shows that

$$\begin{aligned} \mu(\mathbf{x}_n) &= -\frac{1}{\sigma^2} \|\mathbf{y}_n - \mathbf{A}_n \hat{\boldsymbol{\theta}}_n\|^2 + \mathbf{l}_n^T \mathbf{x}_n - \ln(\pi^N \det \boldsymbol{\Phi}_n) \\ &\quad - \|\hat{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}\|_{\mathbf{R}_\theta}^2. \end{aligned} \quad (5.12)$$

From (5.12), we see that the noncoherent MAP metric $\mu(\mathbf{x})$ is the sum of a “coherent MAP metric” $-\frac{1}{\sigma^2} \|\mathbf{y}_n - \mathbf{A}_n \hat{\boldsymbol{\theta}}_n\|^2 + \mathbf{l}_n^T \mathbf{x}_n$, a “bias term” $-\ln(\pi^N \det \boldsymbol{\Phi}_n)$, and a term $-\|\hat{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}\|_{\mathbf{R}_\theta}^2$ which penalizes the deviation of the conditional estimate $\hat{\boldsymbol{\theta}}$ from the prior statistics $\boldsymbol{\theta}_n \sim \mathcal{CN}(\bar{\boldsymbol{\theta}}, \mathbf{R}_\theta)$. Thus, the recursive MAP sequence metric evaluation implicitly uses per-survivor processing (PSP) [47].

5.2.2 Practical Sequential Soft Noncoherent Equalization

From (5.6), computation of exact soft outputs $L_e(x_k|\mathbf{y})$ is impractical because it requires evaluating and summing $\mu(\mathbf{x})$ for all $2^{N_s q}$ hypotheses of \mathbf{x} . However, we expect the set $\{\exp \mu(\mathbf{x})\}$ to be dominated by a few “significant” bit vectors \mathbf{x} , which we collect into the set \mathcal{S} . Thus, we reason that near-optimal soft outputs will result from restricting the summations in (5.6) to $\mathbf{x} \in \mathcal{S}$, i.e.,

$$L_e(x_k|\mathbf{y}) \approx \ln \frac{\sum_{\mathbf{x} \in \mathcal{S} \cap \{\mathbf{x}: x_k=1\}} \exp \mu(\mathbf{x})}{\sum_{\mathbf{x} \in \mathcal{S} \cap \{\mathbf{x}: x_k=0\}} \exp \mu(\mathbf{x})} - L_a(x_k). \quad (5.13)$$

If desired, the “max-log” approximation $\sum_{\mathbf{x}: x_k=x} \exp \mu(\mathbf{x}) \approx \max_{\mathbf{x}: x_k=x} \mu(\mathbf{x})$ could be applied for further simplification:

$$L_e(x_k|\mathbf{y}) \approx \max_{\mathbf{x} \in \mathcal{S} \cap \{\mathbf{x}: x_k=1\}} \mu(\mathbf{x}) - \max_{\mathbf{x} \in \mathcal{S} \cap \{\mathbf{x}: x_k=0\}} \mu(\mathbf{x}) - L_a(x_k). \quad (5.14)$$

To find the significant bit vectors \mathcal{S} and their metrics $\{\mu(\mathbf{x})\}_{\mathbf{x} \in \mathcal{S}}$, we suggest a suboptimal breadth-first TSA such as the M-algorithm or the T-algorithm [66].

The M-algorithm is particularly convenient because it yields a complexity that is invariant to channel realization and SNR. With search breadth M and the recursion in Table 5.1, soft noncoherent equalization consumes only $\mathcal{O}(M2^q N N_h^2 N_D^2)$ operations (since $N_b N_H = N_h N_D$ and $|\mathbb{S}| = 2^q$). Furthermore, when the symbol constellation \mathbb{S} satisfies a multi-level bit mapping, the complexity can be made nearly independent of q , as discussed in [88], which is useful when q is large.

Note that $\mathcal{S} \cap \{\mathbf{x} : x_k = 1\}$ or $\mathcal{S} \cap \{\mathbf{x} : x_k = 0\}$ may be empty for some k , which would make $L_e(x_k|\mathbf{y})$ infinite. For this, a simple solution is to clip $L_e(x_k|\mathbf{y})$ [88]. Note also that (arbitrarily placed) pilot symbols are easily incorporated by setting their a priori bit LLRs \underline{l}_i to very large values.

5.3 Equalization via the Bayesian EM Algorithm

5.3.1 Exact Bayesian-EM Soft Equalization

In this section, we propose a soft noncoherent equalizer based on the Bayesian EM (EMB) algorithm [28]. Given the iteration- i parameter estimate $\hat{\boldsymbol{\theta}}[i]$, the “incomplete data” \mathbf{y} , and the “missing data” \mathbf{s} , the EMB algorithm specifies the following iteration- $(i+1)$ parameter estimate (for $i \geq 1$):

$$\hat{\boldsymbol{\theta}}[i+1] \triangleq \arg \max_{\boldsymbol{\theta}} \mathbb{E} \left\{ \ln p(\mathbf{y}, \mathbf{s}|\boldsymbol{\theta}) \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i] \right\} + \ln p(\boldsymbol{\theta}). \quad (5.15)$$

As before, we treat $\boldsymbol{\theta} \sim \mathcal{CN}(\bar{\boldsymbol{\theta}}, \mathbf{R}_{\boldsymbol{\theta}})$. It is well known that the EMB estimates will converge to $\hat{\boldsymbol{\theta}}_{\text{MAP}} \triangleq \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{y})$ iff¹¹ the initialization $\hat{\boldsymbol{\theta}}[0]$ is within the region-of-convergence of the global maximum of $p(\boldsymbol{\theta}|\mathbf{y})$. The choice of $\hat{\boldsymbol{\theta}}[0]$ is discussed in the sequel.

¹¹Since $p(\boldsymbol{\theta}|\mathbf{y})$ is generally not unimodal in $\boldsymbol{\theta}$, the EM algorithm may converge to a local maximum other than the global maximum.

Due to the independence between data and channel, $p(\mathbf{y}, \mathbf{s}|\boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta})p(\mathbf{s})$, where $\mathbf{y}|\mathbf{s}, \boldsymbol{\theta} \sim \mathcal{CN}(\mathbf{A}\boldsymbol{\theta}, \sigma^2\mathbf{I})$. Plugging the resulting pdfs into (5.15) and simplifying, we get

$$\begin{aligned}\hat{\boldsymbol{\theta}}[i+1] &= \arg \min_{\boldsymbol{\theta}} \mathbb{E}\left\{\|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|^2 \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]\right\} \\ &\quad + \sigma^2\|\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\|_{\mathbf{R}_{\theta}^{-1}}^2.\end{aligned}\quad (5.16)$$

Zeroing the partial derivative w.r.t $\hat{\boldsymbol{\theta}}$ gives

$$\hat{\boldsymbol{\theta}}[i+1] = (\mathbf{C} + \sigma^2\mathbf{R}_{\theta}^{-1})^{-1}(\bar{\mathbf{A}}^H\mathbf{y} + \sigma^2\mathbf{R}_{\theta}^{-1}\bar{\boldsymbol{\theta}}) \quad (5.17)$$

$$= \bar{\boldsymbol{\theta}} + (\mathbf{C} + \sigma^2\mathbf{R}_{\theta}^{-1})^{-1}(\bar{\mathbf{A}}^H\mathbf{y} - \mathbf{C}\bar{\boldsymbol{\theta}}) \quad (5.18)$$

for $\bar{\mathbf{A}} \triangleq \mathbb{E}\{\mathbf{A} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]\}$ and $\mathbf{C} \triangleq \mathbb{E}\{\mathbf{A}^H\mathbf{A} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]\}$. Recalling (5.3), we can write

$$\bar{\mathbf{A}} = [\mathcal{D}_0(\mathbf{m})\mathbf{B}, \dots, \mathcal{D}_{N_H-1}(\mathbf{m})\mathbf{B}] \quad (5.19)$$

$$\mathbf{C} = \bar{\mathbf{A}}^H\bar{\mathbf{A}} + \begin{bmatrix} \mathbf{B}^H\mathcal{D}_0(\mathbf{v})\mathbf{B} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{B}^H\mathcal{D}_{N_H-1}(\mathbf{v})\mathbf{B} \end{bmatrix}, \quad (5.20)$$

using the vector $\mathbf{m} \triangleq [m_0, \dots, m_{N-1}]^T$ of posterior means $m_n \triangleq \mathbb{E}\{s_n \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]\}$, and the vector $\mathbf{v} \triangleq [v_0, \dots, v_{N-1}]^T$ of posterior variances $v_n \triangleq \mathbb{E}\{|s_n - m_n|^2 \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]\}$, computed as

$$m_n = \sum_{s \in \mathbb{S}} s \Pr[s_n = s \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]] \quad (5.21)$$

$$v_n = \sum_{s \in \mathbb{S}} |s - m_n|^2 \Pr[s_n = s \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]]. \quad (5.22)$$

In (5.21)-(5.22), the symbol probabilities $\{\Pr[s_n = s \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]]\}_{n=0}^{N-1}$ can be calculated from the posterior bit LLRs

$$L(x_k \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]) \triangleq \ln \frac{\Pr[x_k = 1 \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]]}{\Pr[x_k = 0 \mid \mathbf{y}, \hat{\boldsymbol{\theta}}[i]]}. \quad (5.23)$$

With QPSK (i.e., $\mathbb{S} = \{\pm 1 \pm j\}$) for example, we find that $m_n = \tanh\{\frac{1}{2}L(x_{2n}|\mathbf{y}, \hat{\boldsymbol{\theta}}[i])\} + j \tanh\{\frac{1}{2}L(x_{2n+1}|\mathbf{y}, \hat{\boldsymbol{\theta}}[i])\}$ and $v_n = 2 - |m_n|^2$. The posterior LLRs can be written as

$$L(x_k|\mathbf{y}, \hat{\boldsymbol{\theta}}[i]) = \ln \frac{\sum_{\mathbf{x}:x_k=1} \exp \mu(\mathbf{x}|\hat{\boldsymbol{\theta}}[i])}{\sum_{\mathbf{x}:x_k=0} \exp \mu(\mathbf{x}|\hat{\boldsymbol{\theta}}[i])}, \quad (5.24)$$

using the coherent MAP sequence metric (recall (5.12)):

$$\mu(\mathbf{x}|\hat{\boldsymbol{\theta}}[i]) \triangleq \ln p(\mathbf{y}|\mathbf{x}, \hat{\boldsymbol{\theta}}[i])p(\mathbf{x}) \quad (5.25)$$

$$= -\frac{1}{\sigma^2}\|\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\theta}}[i]\|^2 + \mathbf{l}^T \mathbf{x}. \quad (5.26)$$

We identify (5.18)–(5.22) as the “M-step” and (5.24)–(5.26) as the “E-step” of the EM recursion. The M-step updates the channel estimate $\hat{\boldsymbol{\theta}}[i+1]$ using the latest posterior bit LLRs $\{L(x_k|\mathbf{y}, \hat{\boldsymbol{\theta}}[i])\}_{k=0}^{N_s Q-1}$, and the E-step updates the posterior LLRs using the latest channel estimate. In this way, the EM-based noncoherent equalizer iterates between “soft” channel estimation and soft coherent equalization. The LLR inputs $L_a(x_k)$ for the first EM iteration ($i=0$) come from the decoder, and the LLRs outputs $L(x_k|\mathbf{y}, \hat{\boldsymbol{\theta}}[K-1])$ from the last EM iteration ($i=K-1$) are passed to the decoder.

Regarding the choice of $\hat{\boldsymbol{\theta}}[0]$, observe that the simple choice $\hat{\boldsymbol{\theta}}[0] = \mathbf{0}$ leads to $\mu(\mathbf{x}|\hat{\boldsymbol{\theta}}[0]) = -\frac{1}{\sigma^2}\|\mathbf{y}\|^2 + \mathbf{l}^T \mathbf{x}$, of which the constant term $-\frac{1}{\sigma^2}\|\mathbf{y}\|^2$ cancels in (5.24), leading to $L(x_k, \mathbf{y}, \hat{\boldsymbol{\theta}}[0]) = L_a(x_k)$, which is a reasonable choice for 0^{th} iteration posteriors. The 0^{th} iteration posteriors can be improved, however, through a more informed choice of $\hat{\boldsymbol{\theta}}[0]$. For example, $\hat{\boldsymbol{\theta}}[0]$ could be chosen as the MMSE estimate of $\boldsymbol{\theta}$ based on pilots in the current and neighboring blocks. An even more informative choice for $\hat{\boldsymbol{\theta}}[0]$ would be the LMMSE estimate of $\boldsymbol{\theta}$ based on previously demodulated symbols, in addition to pilots, in the current and/or neighboring blocks. The details are essentially the same as those discussed for the selection of $\bar{\boldsymbol{\theta}}$ in Section 5.4.1.

5.3.2 Practical EM-Based Soft Noncoherent Equalization

One can observe two potential challenges for practical implementation of the EM-based equalizer:

1. calculation of the posterior LLRs in (5.24), and
2. the $(N_b N_H \times N_b N_H)$ matrix inversion in (5.18).

Recall that (5.24) is essentially *coherent* MAP equalization, a well-understood problem. One option is to generate an estimate of \mathbf{H} from $\hat{\boldsymbol{\theta}}[i]$ and then to leverage the N_H -banded structure of \mathbf{H} via the forward-backward algorithm [22], consuming $\mathcal{O}(N_H N 2^{q 2^{N_H}})$ operations. Another option is to compute approximate LLRs using only the “significant” bit vectors found during a breadth-first TSA, as proposed for the noncoherent case in Section 5.2.2. The coherent tree-search is simpler, though: comparing the coherent metric (5.26) to the noncoherent metric (5.12), we see that the coherent metric update can skip the calculation of \mathbf{d}_n , α_n , $\boldsymbol{\Sigma}_n^{-1}$ and $\hat{\boldsymbol{\theta}}_n$, and set $\mu(\mathbf{x}_n | \hat{\boldsymbol{\theta}}[i]) = \mu(\mathbf{x}_{n-1} | \hat{\boldsymbol{\theta}}[i]) - \sigma^{-2} |e_n|^2 + \underline{\mathbf{l}}_n^T \underline{\mathbf{x}}_n$, thus consuming only $\mathcal{O}(M 2^q N N_b N_H)$ operations per EM iteration. This will be cheaper than the forward-backward for large enough N_H . (The max-log approximation, discussed after (5.13), could also be employed here.)

We now propose an approach to circumvent the matrix inversion in (5.18): a sequential method and an iterative method. If we assume that $\mathbf{v} \approx \mathbf{0}$, then $\mathbf{C} \approx \bar{\mathbf{A}}^H \bar{\mathbf{A}}$ and (5.18) becomes

$$\hat{\boldsymbol{\theta}}[i+1] \approx \bar{\boldsymbol{\theta}} + \left(\bar{\mathbf{A}}^H \bar{\mathbf{A}} + \sigma^2 \mathbf{R}_\theta^{-1} \right)^{-1} \bar{\mathbf{A}}^H (\mathbf{y} - \bar{\mathbf{A}} \bar{\boldsymbol{\theta}}), \quad (5.27)$$

facilitating the application of the sequential-Bayes recursion summarized in Table 5.2, with complexity $\mathcal{O}(N N_h^2 N_D^2)$. We justify $\mathbf{v} \approx \mathbf{0}$ by noting that, as the turbo iterations

proceed, the symbol estimates should become more reliable, thereby reducing the variances v_n . (In Section 6.4, we demonstrate numerically that this approximation is benign.)

Table 5.2: N -step computation of $\hat{\boldsymbol{\theta}}[i+1]$

$\boldsymbol{\Sigma}_{-1}^{-1} = \sigma^{-2} \mathbf{R}_\theta$ and $\hat{\boldsymbol{\theta}}_{-1} = \bar{\boldsymbol{\theta}}$ for $n = 0, 1, 2, \dots, N-1$, $\mathbf{a}_n = [m_n \mathbf{b}_n^H, \dots, m_{n-N_H+1} \mathbf{b}_n^H]^H$ $\mathbf{d}_n = \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{a}_n$ $\alpha_n = (1 + \mathbf{a}_n^H \mathbf{d}_n)^{-1}$ $e_n = y_n - \mathbf{a}_n^H \hat{\boldsymbol{\theta}}_{n-1}$ $\boldsymbol{\Sigma}_n^{-1} = \boldsymbol{\Sigma}_{n-1}^{-1} - \alpha_n \mathbf{d}_n \mathbf{d}_n^H$ $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_{n-1} + \alpha_n e_n \mathbf{d}_n$ end $\hat{\boldsymbol{\theta}}[i+1] = \hat{\boldsymbol{\theta}}_{N-1}$
--

5.3.3 Refinement of the Prior

Though (5.26) specifies the use of the priors $\{L_a(x_k)\}_{k=0}^{N_s q-1}$ at every EM iteration, we have observed that performance improves significantly if the most recently calculated posteriors $\{L(x_k|\mathbf{y}, \hat{\boldsymbol{\theta}}[i-1])\}_{k=0}^{N_s q-1}$ are used in place of the priors in (5.26) for EM iterations $i \geq 2$.

5.4 Implementation Details

5.4.1 Choice of $\bar{\boldsymbol{\theta}}^{(j)}$ and \mathbf{R}_θ

Recall that both the sequential noncoherent (SNC) equalizer of Section 5.2 and the EMB equalizer of Section 5.3 employ the channel prior $\boldsymbol{\theta}^{(j)} \sim \mathcal{CN}(\bar{\boldsymbol{\theta}}^{(j)}, \mathbf{R}_\theta)$. There

are several methods to choose $\bar{\boldsymbol{\theta}}^{(j)}$. In the simplest case, one could assume Rayleigh fading and set $\bar{\boldsymbol{\theta}}^{(j)} = \mathbf{0}$. However, improved performance generally results when $\bar{\boldsymbol{\theta}}^{(j)}$ is more informative. Thus, one might instead set $\bar{\boldsymbol{\theta}}^{(j)}$ equal to the MMSE estimate of $\boldsymbol{\theta}^{(j)}$ based on pilots in the current and neighboring blocks. This estimator would be linear and a function of the assumed correlation¹² between $\boldsymbol{\theta}^{(j)}$ and $\{\boldsymbol{\theta}^{(i)}\}_{i \neq j}$ and, for a given pilot pattern, could be computed offline and implemented using only $\mathcal{O}(N_b N_H N_P)$ operations, where N_P denotes the total number of pilot observations used for the estimate. Reduced-rank techniques can be used to further reduce the complexity. Details can be found in [84].

A yet more informative choice for $\bar{\boldsymbol{\theta}}^{(j)}$ would be when, in addition to pilots, the previously estimated symbols $\{\mathbf{s}^{(i)}\}_{i \neq j}$ in current¹³ and/or neighboring blocks are used to compute the LMMSE estimate of $\boldsymbol{\theta}^{(j)}$. In this case, the estimator cannot be computed in advance and so estimation would be potentially more complex than in the pilot-only case. However, iterative approximation techniques shown in Table 5.2 could be used to avoid matrix inversion.

The simplest choice of the channel covariance \mathbf{R}_θ would be based on assumed worst-case spreading (e.g., maximal Doppler spread). While, conceivably, pilots and/or previously demodulated symbols could be used to estimate \mathbf{R}_θ , our numerical experiments have not suggested that performance improvements result from this approach.

¹²The estimation would also be a function of the assumed mean of $\boldsymbol{\theta}^{(j)}$, which for this purpose we could set to zero.

¹³Soft symbol estimates for the current block are available only after the first turbo iteration.

5.4.2 Pilot and Guard Patterns

In Section 7.1, we assumed that the last $N_H - 1$ symbols in \mathbf{s} are zero-valued guards, so that \mathbf{H} acts causally on the first $N - N_H + 1$ symbols. With this guard, the last $N_H - 1$ columns of \mathbf{H} become inconsequential; zeroing them (w.l.o.g) yields a lower-triangular N_H -banded matrix. The banded property was exploited for both noncoherent and coherent TSA.

In choosing the pilot pattern, one goal is that the pilots are maximally informative about the BE parameters $\boldsymbol{\theta}$, e.g., that they minimize the MMSE of the MMSE-optimal $\boldsymbol{\theta}$ estimate. For DS channels whose time-variation obeys a complex-exponential BE model [86], and for estimators which use only pilots within the current block, the MMSE-optimal pilot pattern has been derived for single-carrier zero-padded schemes in [54], and for general affine precoded schemes (including multi-carrier schemes) in [92], showing similarity to earlier heuristic proposals in, e.g., [93]. Another goal is that the pilot pattern does not introduce too much redundancy, i.e., that the overall pilot/data scheme attains the maximum achievable rate. For DS channels whose time-variation obeys a complex-exponential BE model, the maximum achievable rate was derived in [94], in addition to a pilot-based system which achieves this rate. The optimal pilot pattern for a practical suboptimal transmitter and receiver, however, is difficult to derive in closed form.

We now discuss the interaction of pilots and the noncoherent TSA performed by the SNC equalizer. For the optimal MAP equalizer discussed in Section 5.2.1, a single pilot symbol is sufficient to resolve the inherent channel/data phase ambiguity of the noncoherent metric $\mu(\mathbf{x})$. For the practical SNC equalizer based on suboptimal TSA, however, we want that the per-survivor channel estimates $\hat{\boldsymbol{\theta}}_n$ are well adjusted *before*

the suboptimal TSA prunes paths. For this reason, in-block pilots placed near the beginning of the block are especially effective. As discussed earlier, pilots outside the demodulation block can be leveraged simply by initializing $\hat{\boldsymbol{\theta}}_n$ (i.e., by setting $\bar{\boldsymbol{\theta}}$) as an MMSE estimate of $\boldsymbol{\theta}$ based on those pilots. Pilots are especially important during the first turbo iteration, when the in-block a priori data LLRs are zero; for later turbo iterations, the a priori data LLRs returned by the decoder help to “guide” the suboptimal TSA.

For the EMB noncoherent equalizer of Section 5.3, pilots can be leveraged both through the choice of the prior mean $\bar{\boldsymbol{\theta}}$ as well as the initialization $\hat{\boldsymbol{\theta}}[0]$, as previously discussed.

Based on all of the aforementioned considerations, we use the pilot patterns shown in Fig. 5.2(b)-(c) for the numerical experiments in Section 5.5. In both single- and multi-carrier cases, we use a total of N_p pilot/guards per N -length block. In the single-carrier case, each block contains $N_p - N_h + 1$ non-zero leading pilots and $N_h - 1$ zero-valued guards. In the multi-carrier case, each block contains $L \geq 1$ pilot/guard clusters, where each cluster contains $N_D - 1$ leading¹⁴ zero-valued guards and $N_p/L - N_D + 1$ trailing non-zero pilots. The cluster pattern repeats every $P = N/N_p$ blocks, and the cluster locations are staggered¹⁵ so that each subcarrier appears in a cluster exactly once every P blocks.

¹⁴Here, we use only leading guards because the frequency-domain channel is causal. When $N_p = N_h N_D$ and $L = N_h$, this multi-carrier pattern is equivalent to the “Kronecker delta” pattern discussed in [92, 93] for (non-causal) frequency-domain channels.

¹⁵Note that, by cyclically the elements of both \mathbf{y} and \mathbf{s} , it is possible to place $N_H - 1$ guards at the end of the block while maintaining the “circularly banded” structure of \mathbf{H} illustrated in Fig. 5.2(a).

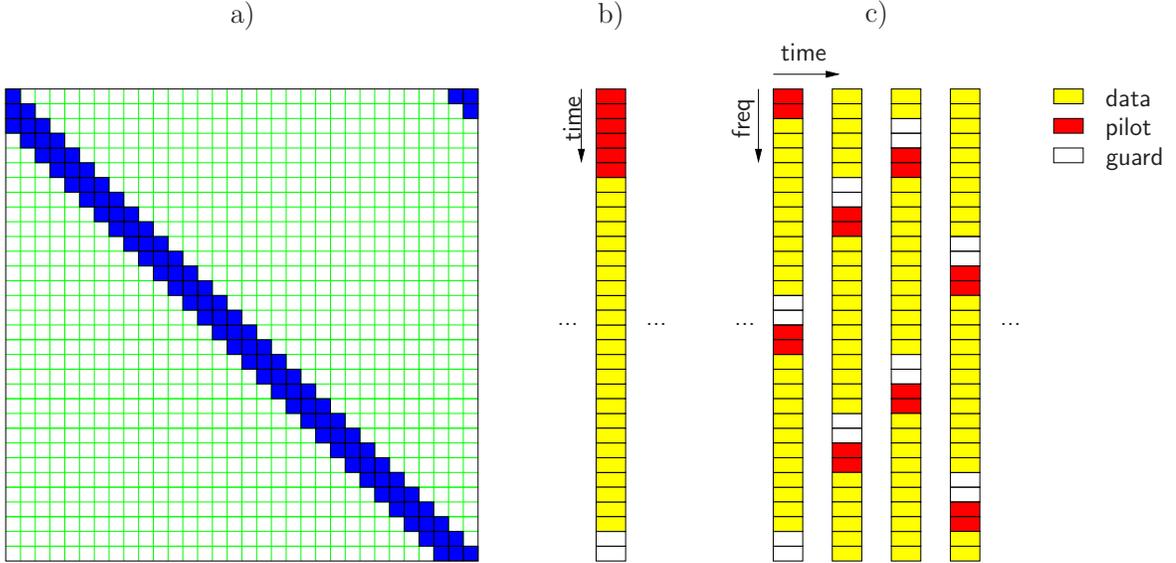


Figure 5.2: For $N = 32$, $N_H = 3$, and $N_p = 8$, illustration of a) $\mathbf{H}^{(j)}$ support, b) single-carrier pilot pattern, and c) multi-carrier pilot pattern with $P = 4$ and $L = 2$.

5.4.3 Complexity

The computation complexity of the proposed equalizers (per turbo iteration) is summarized in Table 5.3, where K denotes the number of EM iterations. When interpreting these complexities, it is important to keep in mind that the block size N is typically chosen to be a small multiple of the channel length N_h (e.g., $N \sim 4N_h$). In this case, the complexities of the Table 5.1-based SNC-equalizer as well as the Table 5.2-based EM-equalizer, grow cubically with the channel length N_h . While we should not forget that the complexity EM-based equalizers scales with number of EM iterations, K , we show in Section 5.5 that K is usually much less than 10.

Table 5.3: Summary of computational complexity

Sequential algorithm using Table 5.1	$\mathcal{O}(M2^q NN_h^2 N_D^2)$
EM-based algorithm using Table 5.2	$\mathcal{O}(KNN_h^2 N_D^2)$

5.5 Numerical Results

For the numerical experiments, Jakes method [95] was employed to generate realizations of a wide-sense stationary uncorrelated (WSSUS) Rayleigh fading channel with temporal autocorrelation $\rho_m = J_0(2\pi f_D T_s m)$ and uniform delay-power profile $\sigma_l^2 = 1/N_h$. Here, f_D denotes the normalized single-sided Doppler spread and $J_0(\cdot)$ the 0th-order Bessel function of the first kind. The values $f_D T_s = 0.002$ and $N_h = 3$ were assumed unless otherwise noted.

The transmitter employed rate- $R = \frac{1}{2}$ irregular low density parity check (LDPC) codes with average column-weight 3, generated by the publicly available software [96]. The coded bits were mapped to QPSK symbols (i.e., $q=2$) and partitioned into data blocks of length N_s , each of which was merged with N_p pilot/guards, as discussed in Section 5.4.2, to form a transmission block of length $N = N_s + N_p$. So that each codeword spanned $J = 32$ data blocks, $(JqN_s, RJqN_s)$ -LDPC codes were employed. Throughout, we used block length $N = 64$ with $N_p = 8$ pilot/guards per block.

Though our soft noncoherent equalizers can be applied to either single- or multi-carrier communication, we mainly present single-carrier experiment here and a brief multi-carrier experiment follows in the later part of the section, and more realistic multi-carrier experiments for ncT-BE will be presented in Chapter 6. Our single-carrier soft noncoherent equalizer used the Karhunen Lóeve (KL) BE [58] with $N_b = 3$ to

model channel variation. In other words, \mathbf{B} was constructed column-wise from the N_b principal eigenvectors of $\mathbf{R}_h \triangleq \text{E}\{\mathbf{h}_d \mathbf{h}_d^H\}$ and diagonal \mathbf{R}_θ was constructed from the N_b principal eigenvalues of \mathbf{R}_h . Since the channel was Rayleigh, we set $\bar{\boldsymbol{\theta}} = \mathbf{0}$ as the channel mean for the first turbo iteration. The M-algorithm used the search parameter $M = 64$, where the LLR magnitudes were clipped to 2.3 for noncoherent TSA (ncT) based LLR and 8 for coherent TSA (cT) based LLR. The LDPC decoder by MacKay and Neal [97] was used with a maximum of 60 LDPC iterations, and equalization/decoding were iterated using a maximum of 8 turbo iterations. We specify the *maximum* number of iterations because the receiver breaks out of both the LDPC and turbo loops as soon as the LDPC syndrome check indicates error-free decoding. The initialization of channel parameter $\hat{\boldsymbol{\theta}}[0]$ for the EM algorithm was done by MMSE estimation as in (5.27) but with only pilots. The iteration number of EM was set to $K = 3$, unless otherwise noted. Through experiments, the simplified EM-recursion in Table 5.2 was used.

In this section, to precisely refer to the algorithms, we use a naming convention as follows. The proposed sequential algorithm is denoted as

- noncoherent TSA coupled with BE-channel estimation: “ncT-BE,”

and the proposed EM-based algorithm as

- coherent TSA coupled with soft BE-channel estimation after K iterations: “(cT+sBE)^K.”

5.5.1 Effect of System Parameters

Figure 5.3 shows coded BER of sequential algorithm (ncT-BE) under different choices of the search width for M-algorithm: $M \in \{16, 32, 64, 128\}$. The figure shows

that performance increases with M , although gains from the use of $M > 64$ are quite small (e.g., ≈ 0.2 dB). Since EM-based algorithm $((cT+sBE)^3)$ shows almost the same traces, we omit them for brevity.

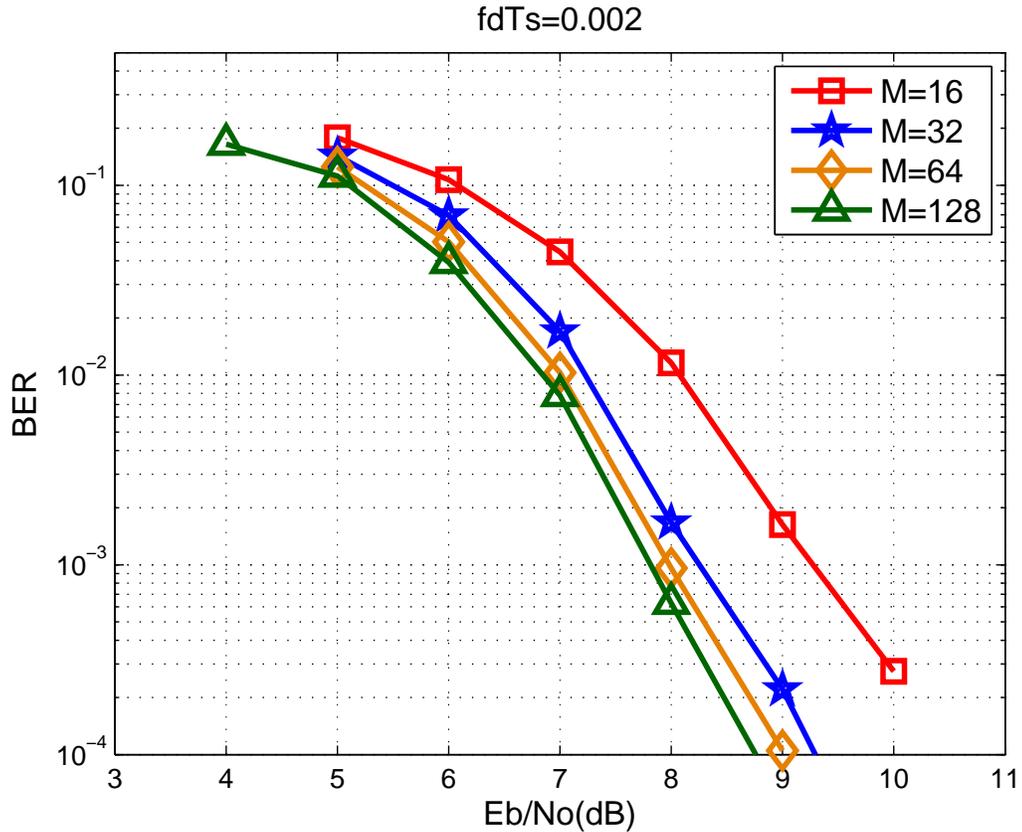


Figure 5.3: Coded BER vs. E_b/N_o for ncT-BE with M-algorithm parameter $M \in \{16, 32, 64, 128\}$.

In Figure 5.4, traces for coded BER versus the maximum number of turbo iterations for ncT-BE and $((cT+sBE)^3)$ algorithms are shown. There, ncT-BE shows fast convergence, i.e., even at 4th iteration the performance almost saturates, showing small gain toward 8th iteration. The performance of $((cT+sBE)^3)$, however, improves

very slowly along the iterations. With full 16 iterations, it outperforms that of ncT-BE. The slow convergence of $(\text{cT+sBE})^K$ along the iterations is more prominent on higher Doppler, which will be seen in Section 5.5.2.

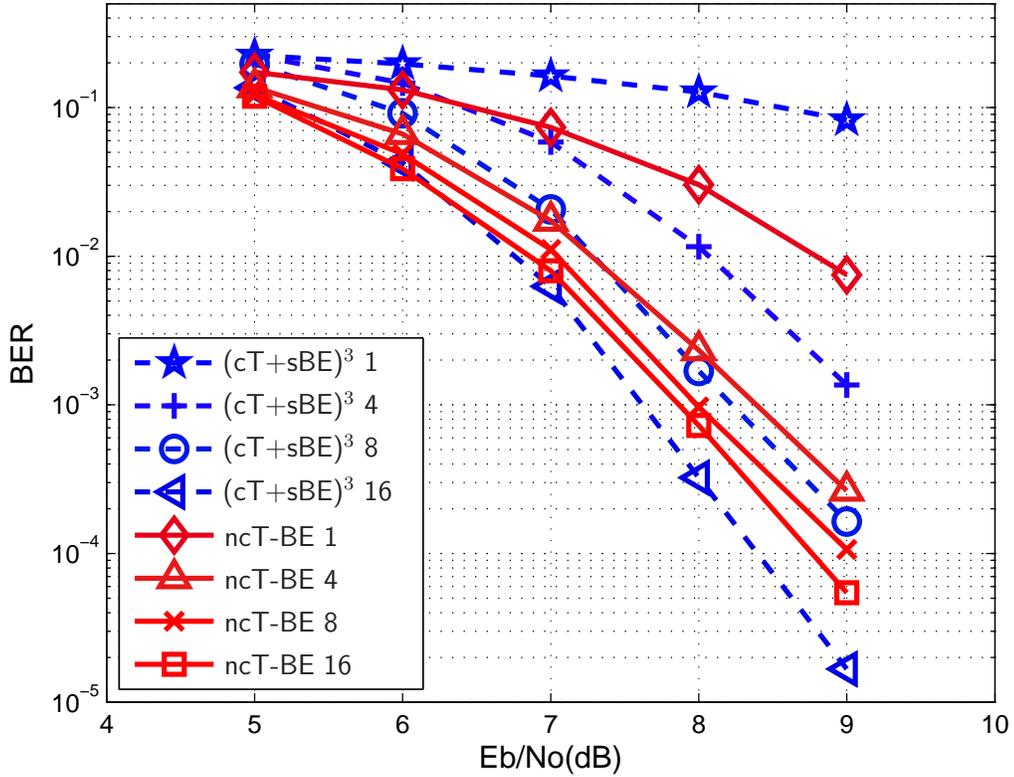


Figure 5.4: BER vs. E_b/N_o for turbo iterations $\in \{1, 4, 8\}$. ncT-BE and $(\text{cT+sBE})^3$ schemes were tested.

Figure 5.5 shows the effect of N_p , the number of pilots per block, on coded BER of ncT-BE. As predicted in Section 5.4.2, performance increases with N_p until $N_p = 6$, after which it saturates and with $N_p = 12$ the performance starts to degrade. This is because beyond the saturation point the improvement in channel estimation error

was canceled out by the penalty on E_b/N_o , as too much energy on pilot was invested for small improvement in (implicit) channel estimate. Theoretically, noncoherent equalizers need no pilot or at least one (for symmetric constellations) to work. Thus, we demonstrate the ability of ncT-BE in that it can run with only one pilot symbol. It doesn't show severe performance degradation, e.g., just 1.5 dB away from that of $N_p = 6$ with almost the same slope of error curve (i.e., diversity order).

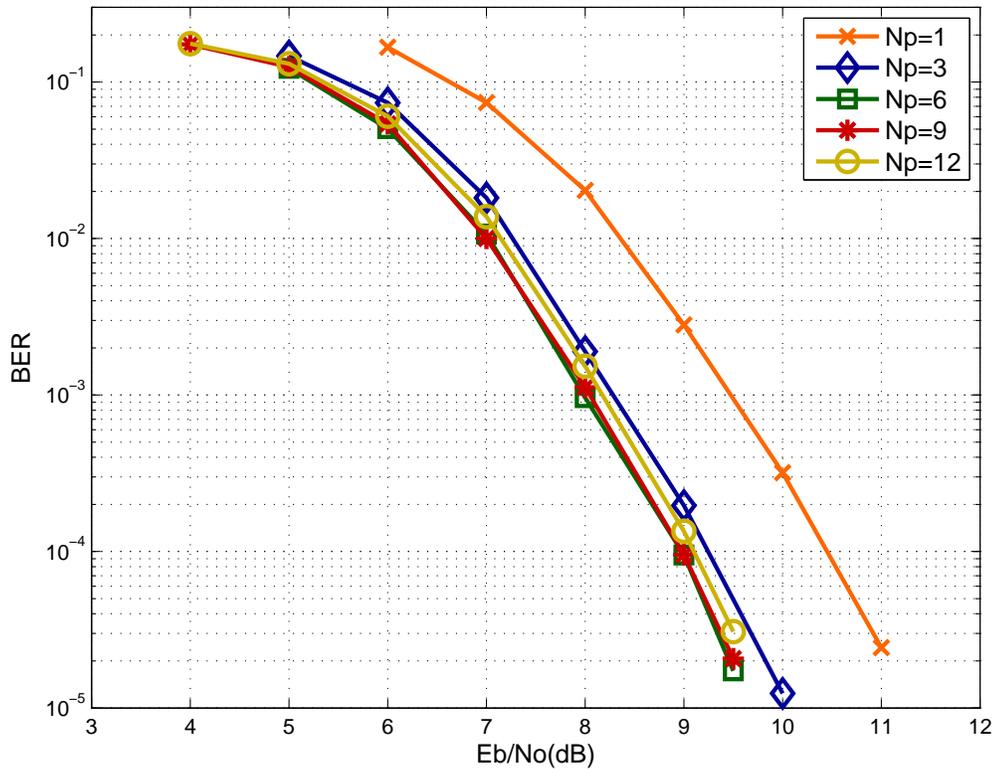


Figure 5.5: ncT-BE algorithm's BER vs. number of pilots $\in \{1, 3, 6, 9, 12\}$.

5.5.2 Performance Comparison

In Fig. 5.6, the soft noncoherent equalizers (ncT-BE and (cT+sBE)^K) proposed in Section 5.2 and Section 5.3 were compared to several other soft noncoherent equalizers:

- coherent TSA coupled with *exact* soft BE-channel estimation after K iterations: “(cT+esBE)^K,”
- coherent TSA using a soft BE-channel estimate *non-iteratively*: “cT+sBE,”
- coherent TSA using a soft Gauss-Markov estimate *non-iteratively*: “cT+sGM,”

and two genie-aided performance bounds:

- coherent TSA based on a *perfect* estimate of the channel \mathbf{H} : “cT+pH,”
- coherent TSA based on a BE-channel estimate constructed using *perfect LLR feedback*: “cT+pllrBE.”

We now elaborate on the procedures cT, sBE, esBE, sGM, pH, and pllrBE mentioned above. As discussed in Section 5.3.2, *coherent TSA* (cT) uses the M-algorithm to sequentially maximize the metric $\ln p(\mathbf{x}|\mathbf{y}, \hat{\mathbf{H}})$ for externally supplied $\hat{\mathbf{H}}$ —a direct application of the MIMO technique [88]. *Soft BE-channel estimation* (sBE) uses the simplified EM-recursion (5.27) and Table 5.2, while *exact soft BE-channel estimation* (esBE) uses the “exact” recursion (5.18). Finally, *soft Gauss-Markov channel estimation* (sGM) refers to the Kalman technique proposed in [37], for which we employed a second-order Gauss-Markov (GM) model.

Figure 5.6 shows that the proposed (cT+sBE)³ approach, a simplification of the expensive (cT+esBE)³ approach, gives almost identical performance. Furthermore,

the proposed ncT-BE and (cT+sBE)³ performs only 2 dB from the perfect-CSI bound cT+pH and only 1.7 dB from the perfect-LLR-feedback bound cT+pHrBE. The traces cT+sBE and cT+sGM can be used to compare between the use of BE versus GM channel models; it can be seen that they are almost the same except high SNR, where the BE approach shows slightly better performance. To see the gain from multiple EM iterations, one can compare (cT+sBE)³ (where $K = 3$) to cT+sBE (where $K = 1$); about 0.5 dB improvement can be observed. No additional gains were observed for $K > 3$.

Figure 5.7 shows the performance under very quickly varying channel with $f_D T_s = 0.005$. ncT-BE still performs well with 3 dB gap from the cT+pHrBE bound, also outperforming all other schemes considered in this simulation. However, (cT+sBE)³ performs bad indicating worse error slope. As we indicated in Fig. 5.4, (cT+sBE)³ converge slowly so that 3 LDPC- and 8 turbo-iterations were not sufficient, underperforming to cT+sGM. Thus, for comparison, we show additional traces for (cT+sBE)⁶ and cT+sBE with doubled (16) turbo iterations, which display much improved performance to match cT+sGM with 16 turbo iterations.

5.5.3 Robustness to Statistical Mismatch

Though the proposed noncoherent equalization schemes operated without knowledge of the channel state, it did assume knowledge of channel distribution in the form of the BE coefficient covariance matrix \mathbf{R}_θ . We now examine the robustness of the schemes to knowledge of Doppler spread $f_D T_s$, the most important parameter in the construction of \mathbf{R}_θ , by comparing the required E_b/N_o to achieve 10^{-2} BER versus $f_D T_s$ performance of the equalizers with perfect knowledge of $f_D T_s$ to ones which assume

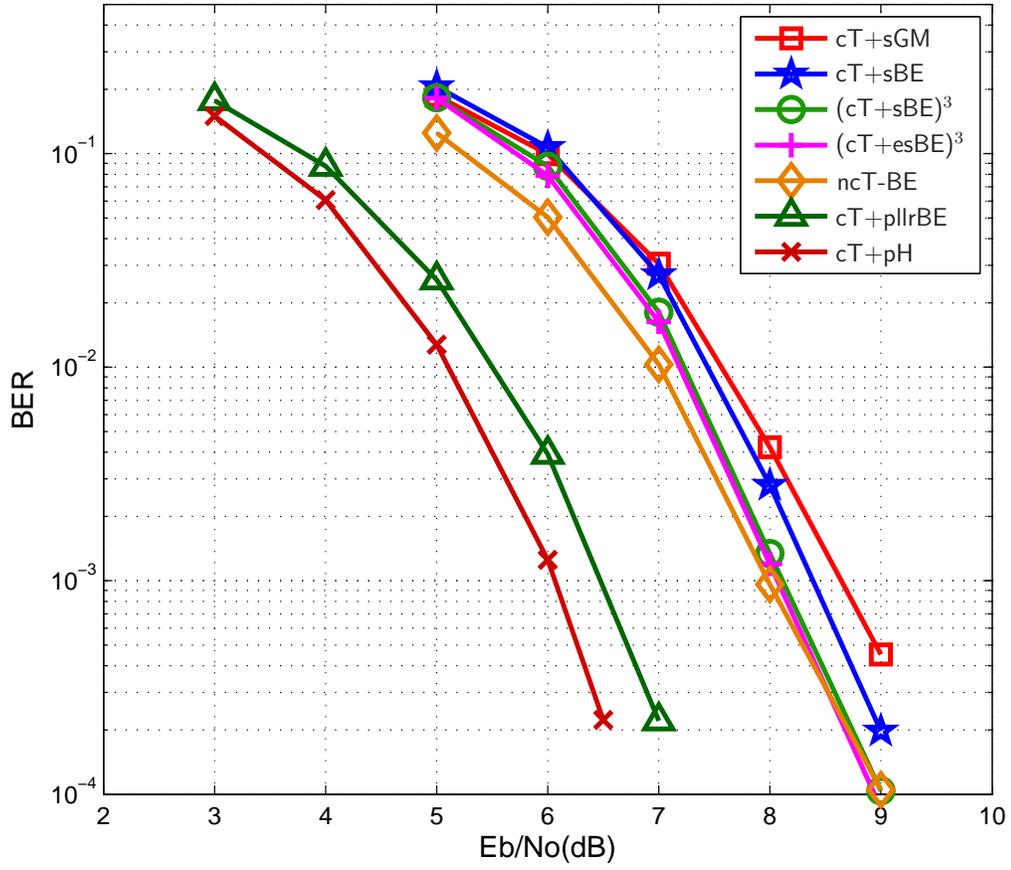


Figure 5.6: BER vs. E_b/N_o for various equalization schemes under $f_D T_s = 0.002$.

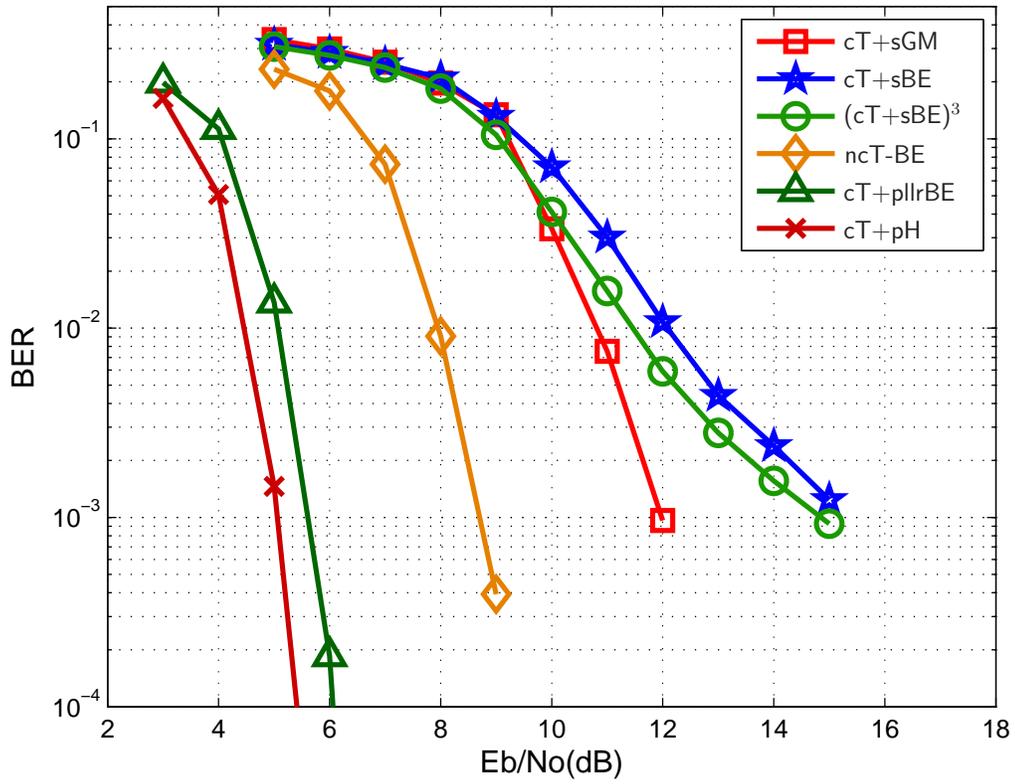


Figure 5.7: BER vs. E_b/N_o for various equalization schemes under very high Doppler with $f_D T_s = 0.005$.

the fixed value $f_D T_s = 0.002$. Figure 5.8 demonstrates that the proposed equalization scheme is robust to mismatch in Doppler-spread: the “mismatched” scheme stays close to the “matched” scheme over the entire range of tested Doppler spreads. Note that, as $f_D T_s$ decreases, the BER for the matched scheme increases due to a lack of diversity; Likewise, the BER for the matched scheme increases sharply with $f_D T_s$ due to the limitations of the $N_b = 3$ BE model. Similar behavior was observed for soft noncoherent equalizers in multi-carrier transmission in [98].

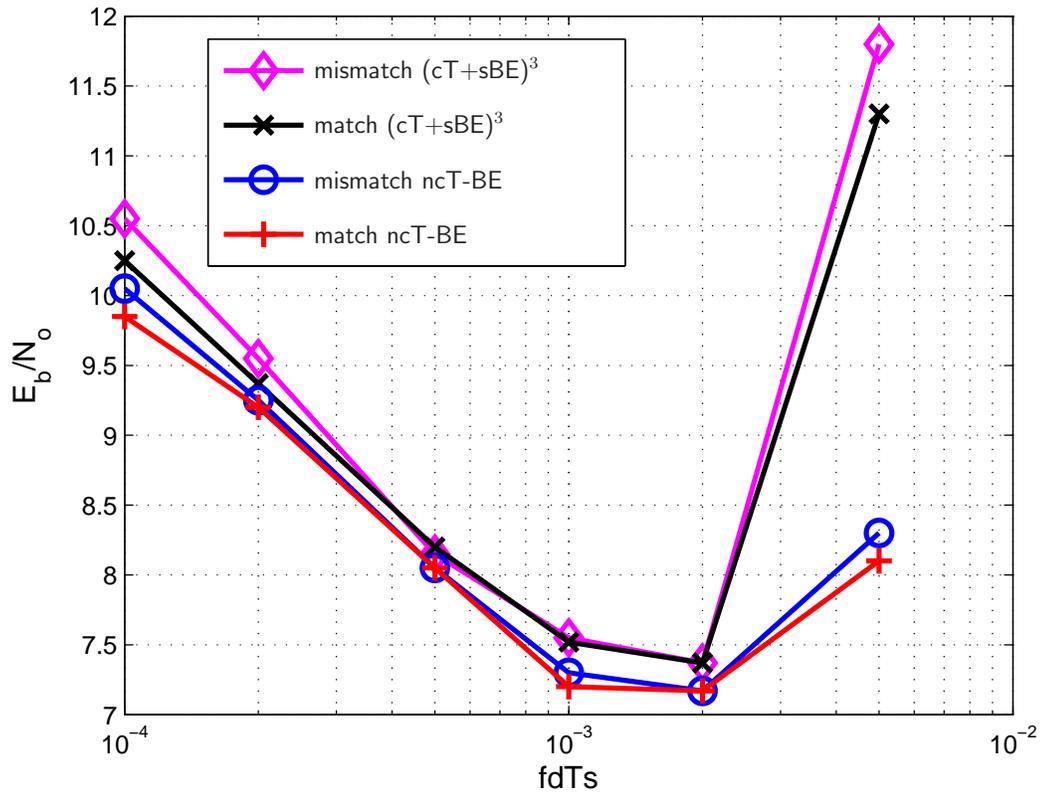


Figure 5.8: Required E_b/N_0 to achieve 10^{-2} BER vs. true $f_D T_s$ for $f_D T_s$ -matched and fixed- $f_D T_s$ (i.e. mismatched) reception.

5.5.4 Multi-carrier Transmission

We tested the noncoherent equalization algorithms considered in Section 5.5.2 under the multi-carrier transmission scenarios. As described in Section 7.1, the multi-carrier transmission scheme employed transmission pulse shaping as in [5] to confine the frequency-domain channel matrix to be banded, so that $N_H = N_D$ and $N_b = N_h$ with \mathbf{B} being the truncated DFT matrix. The receivers for the multi-carrier transmission were equipped with pilot pattern as described in [98]. In this experiment, the length of the channel impulse response was set to unchanged ($N_h = 3$) and the frequency spread (N_D) was set to 3. The BE covariance matrix \mathbf{R}_θ was constructed as in [98]. The EM iteration number was set to $K = 6$. Experiments were performed on two different scenarios on the pilot configuration: In the first scenario, number of pilots was set to $N_p = 16$ and a single ($L = 1$) received symbol was taken for the noncoherent equalization. In the second scenario, $N_p = 9$ pilots were set and at the receiver not only the current received symbol, but also three adjacent received symbols (e.g., $\mathbf{y}^{(j-1)}, \dots, \mathbf{y}^{(j+2)}$), thus total of $L = 4$ received symbols, were taken for $\mathbf{s}^{(j)}$'s equalization. In the later scenario, the receiver effectively used as many as $LN_p = 36$ pilots for a single equalization. For the comparison to the GM-based channel model, cT+sGM algorithm was implemented for the frequency domain to track the channel variation along the frequency axis (not along the time axis).

The BER performance traces of the noncoherent equalizers under the first scenario ($L = 1$) are shown in Fig. 5.9. The turbo iteration was set to 16 for all equalizers. ncT-BE outperforms others, displaying neither error floor (cT+sGM) nor insufficient error slopes ((cT+sBE)⁶ and cT+sBE). The bad performance of EM-based algorithms and GM-based algorithm is caused by noisy initial channel estimate. To see the

impact of this on the performance, in the next experiment we increased the effective pilot numbers, thereby improving the quality of the initial channel estimate.

In Fig. 5.10, we can see that with multiple ($L = 4$) symbols, the equalizers perform much better. Here, the turbo iteration was set as half (8) as that of the previous scenario. Note that (cT+sBE)⁶ outperformed ncT-BE slightly and cT+sGM by 2 dB. Three BE-based noncoherent equalizer algorithms performed almost the same in this experiment.

Throughout the experiments in this section, the ncT-BE consistently performed well regardless of the Doppler, transmission scheme and iteration number. (cT+sBE)^K has lower complexity, and, in the easy channels (e.g., with $f_D T_s = 0.002$ with sufficient pilots) it performed as well as ncT-BE, whereas in the difficult channels (e.g., with $f_D T_s = 0.005$ and/or with insufficient pilots) it underperformed.

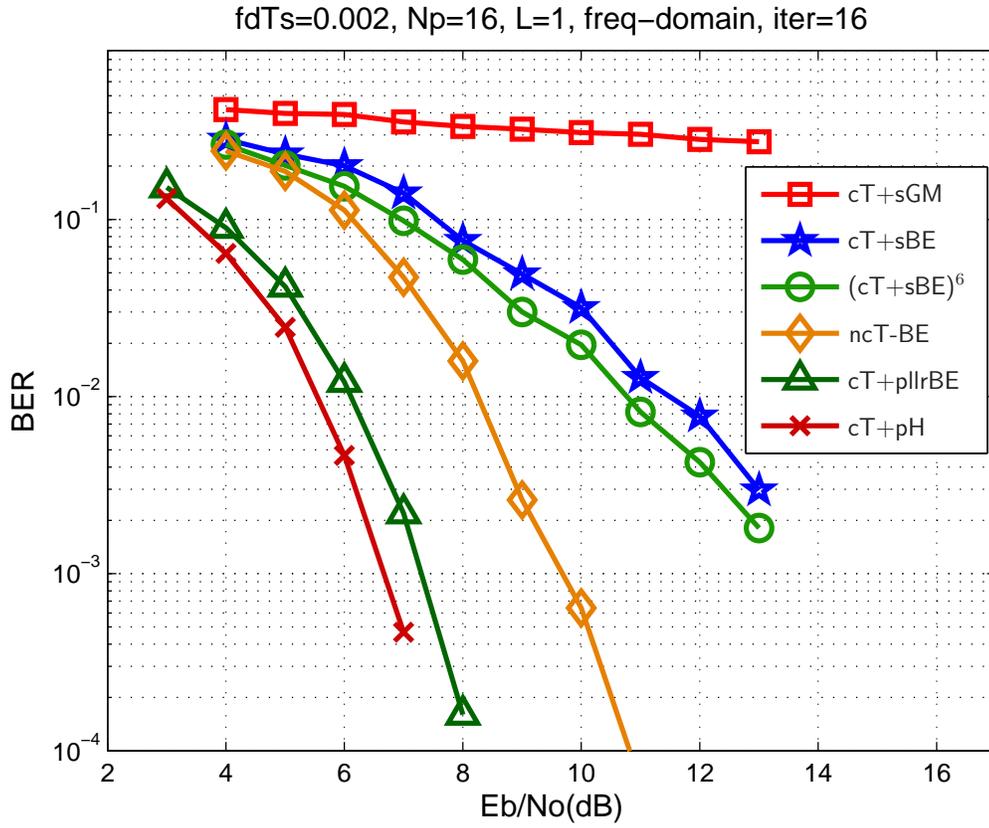


Figure 5.9: BER vs. E_b/N_o for various frequency-domain equalization schemes under $f_D T_s = 0.002$ using $L = 1$.

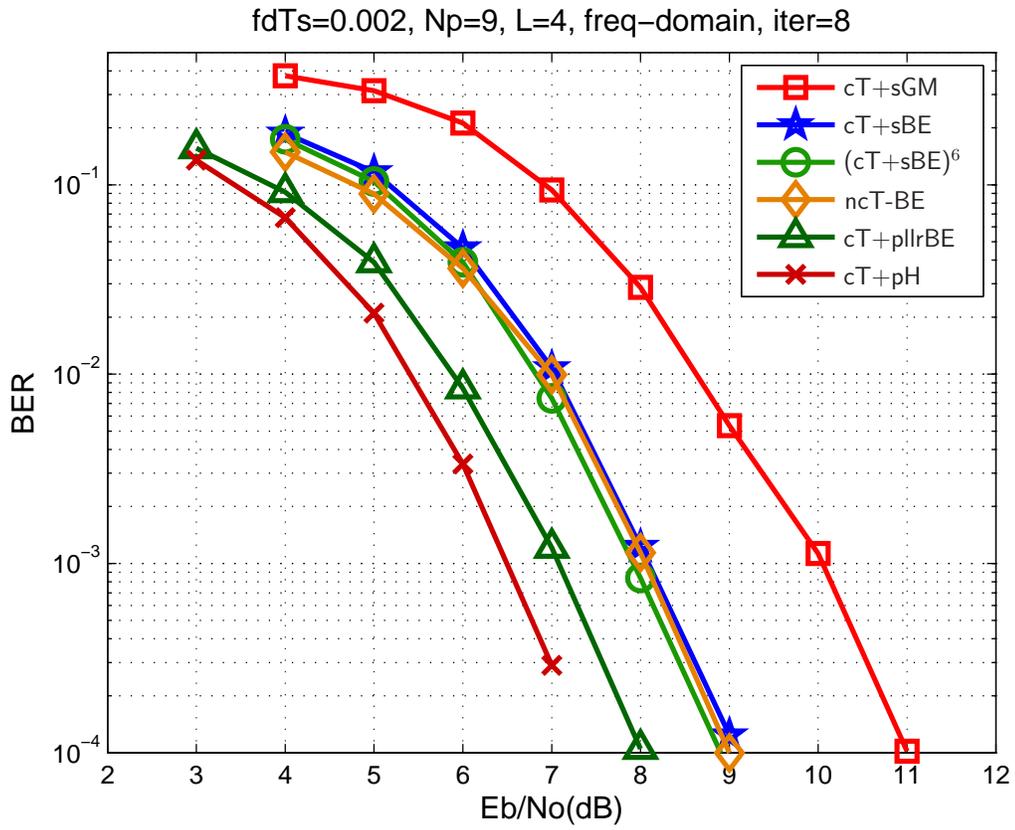


Figure 5.10: BER vs. E_b/N_o for various frequency-domain equalization schemes under $f_D T_s = 0.002$ using $L = 4$.

CHAPTER 6

APPLICATION TO UAC COMMUNICATION

The underwater acoustic channel (UAC) has been referred to as “quite possibly nature’s more unforgiving wireless medium” [99]. The physical characteristics of the UAC are highly dependent on the distance and relative movement between the transmitter and receiver; the proximity, roughness, and motion of the scattering surfaces; and the presence of ambient interference. However, the factors that pose the primary challenges for data communication over the UAC can be summarized as *simultaneously large delay- and Doppler-spreads, limited bandwidth, and limited receiver complexity*. These challenges can be understood as follows. Large delay-spread implies that single-carrier communication will be plagued by inter-symbol interference (ISI) that, for practical signal bandwidths, spans hundreds of symbols. Large Doppler-spread then implies that this ISI response will change quickly in time. Since optimal mitigation of this long and quickly-varying ISI response becomes computationally infeasible, practitioners have resorted to simple sub-optimal strategies such as the adaptive decision-feedback equalization (DFE) [100, 101]. However, these implementable single-carrier techniques perform far short of optimal and fail altogether in very highly spread environments such as the surf zone [102].

As an alternative, multi-carrier modulation (MCM) has been proposed to increase the symbol interval and thereby decrease the ISI span. While a number of MCM proposals for the UAC have been made over the years (see, e.g., the recent work [103–105]), none seem to have been successful enough to displace single-carrier/DFE as the practical method of choice. The primary difficulty in applying MCM to the doubly dispersive UAC is that, as the symbol interval is increased (to reduce ISI span), the subcarrier spacing must be decreased (to preserve data rate and transmission bandwidth), making the system more susceptible to Doppler-spread-induced inter-carrier interference (ICI). Thus, barring a decrease in spectral efficiency, ISI reduction comes at the expense of ICI escalation.

A close look at recent MCM proposals helps to illuminate the challenges in applying MCM to the UAC. The works [103] and [104], for example, proposed classical ZP-OFDM signaling schemes assuming that the Doppler-spread was small enough to induce *negligible* ICI. For these schemes, the universal ICI bound¹⁶ in [106] implies that an ICI power of -25 dB (which we consider to be “negligible”) occurs when $f_D T_s = 0.03$, where f_D denotes the single-sided Doppler-spread and T_s the MCM-symbol duration. Since T_s was chosen as $7T_h$ (as $3.4T_h$) in [103] (in [104]), where T_h denotes delay-spread, we deduce that these schemes can handle UACs with a delay/Doppler-spread product of at most $f_D T_h = 0.004$ ($f_D T_h = 0.009$). The surf-zone channels¹⁷ described in [102], however, yield $f_D T_h \approx 0.1$, which is $25\times$ ($10\times$) as severe. As another example, in the non-traditional MCM approach [105], the symbol length was chosen shorter than that needed for perfect ISI-suppression, in order to

¹⁶The analysis in [106] confirms that this bound is tight for $f_D T_s = 0.03$.

¹⁷For these channels it was found that $f_D \approx 15$ Hz and $T_h \approx 7$ ms.

tolerate high Doppler-spread while keeping ICI negligible. The resulting ISI-span was short enough to enable the use of sophisticated joint estimation/detection techniques (i.e., LMS/Viterbi per-survivor processing [47]), which were shown to significantly outperform the traditional adaptive DFE. But [105] only demonstrated the ability to handle $f_D T_h = 0.0035$ via simulation and $f_D T_h \approx 0.002$ experimentally, which are over 30 times milder than the surf-zone channels discussed in [102]. Thus, the literature seems to fall short of an MCM scheme that is suitable for communication over highly spread UACs.

The general problem of communicating over doubly dispersive channels, i.e., those with simultaneously large delay and Doppler spreads, has received significant attention from theoreticians over the last two decades. Most of the theoreticians have approached this problem through the design of MCM pulse-shapes which minimize total ISI/ICI power (e.g., [9, 107–109]), with the goal of making ISI/ICI negligible. But, even with optimized pulses, ISI/ICI remains¹⁸ non-negligible for channels whose delay/Doppler product $f_D T_h$ is commensurate with that of the surf-zone—a fundamental consequence of the Balian-Low theorem from Gabor theory [9]. This fact led the authors to propose a non-traditional approach in which a small ICI span (e.g., 1 or 2 subcarriers) is tolerated and MCM pulse shapes are designed to minimize *residual* ISI/ICI power [5]. In this case, near-perfect residual-ISI/ICI suppression can be accomplished *without* loss of spectral efficiency, and high-performance dominant-ICI mitigation can be accomplished with low complexity. In this paper, we discuss how such an approach can be applied for communication over the UAC.

¹⁸The ISI/ICI power is significant unless the MCM subcarrier/symbol spacing is increased to about twice the Nyquist spacing, thereby incurring a significant decrease in spectral efficiency which would be very undesirable for the bandwidth-limited UAC.

The aforementioned ISI/ICI mitigation schemes require (implicitly or explicitly) accurate channel state information (CSI). Maintaining this CSI is especially difficult when the channel is doubly dispersive, due to the typically large number of channel coefficients and their fast rate-of-change. Pilot-aided transmission (PAT) [110] is a practical means of aiding data reception in the presence of channel uncertainty. PAT is often used with *decoupled* channel-estimation/data-decoding, where a channel estimate is first obtained via pilots and later used for coherent data decoding. To minimize the MMSE of pilot channel estimates, it is necessary to keep the channel estimates free of interference from unknown data [92] (especially at high SNR), though doing so with a doubly dispersive channel requires time/frequency guards have been shown to sacrifice achievable spectral efficiency [111]. When PAT is used with *joint* estimation/decoding (JED), however, there is no need to separate pilots and data, allowing spectrally efficiency communication over the doubly dispersive channel [111]. These facts motivate the consideration of PAT with JED for the UAC.

As one would expect, the complexity of optimal JED is prohibitive. Practical JED requires the use of simplified channel models and approximations of the optimal maximum *a posteriori* (MAP) decoding metric. For example, it is common to model the channel as first-order Gauss-Markov and to apply trellis-based decoding methods with either forward-backward or fixed-lag MAP processing [31]. Still, this framework does not appear practical for highly dispersive channels like the UAC, whose impulse response spans hundreds of symbols and changes quickly. With this in mind, the authors proposed a novel noncoherent equalizer [89] which uses soft tree-search and leverages a basis-expansion model (BEM) [86] for the time-varying channel. The result is near-MAP performance with a per-symbol complexity that

scales as only $\mathcal{O}(N_h^2 D^2)$, where $N_h = T_h/T_c$ denotes the delay-spread (in chips) and $D = \lceil f_D T_c N \rceil$ the single-sided Doppler spread (in subcarriers). Here, T_c denotes the “chip” (or channel use) interval, i.e., inverse transmission bandwidth. With surf-zone UAC parameters, however, the complexity remains prohibitive. For example, if 512 subcarriers were used to transmit a 14 kHz bandwidth signal over the surf-zone UAC from [102], one can expect $N_h \approx 100$ and $f_D T_c N \approx 0.55$, for which the N_h^2 dependence may be problematic.

The key to solving the complexity riddle may lie in the sparse nature of realistic UAC responses [112, 113]. For example, if only 1/3 of the channel’s N_h delay taps are significant, then a reception algorithm whose complexity is quadratic in the *active* delay taps (versus total delay taps) will save by a factor of 9. But designing a receiver capable of leveraging sparsity in this manner is non-trivial, especially for MCM schemes like [103–105]. In fact, most sparsity-leveraging algorithms are based on adaptive DFE (e.g., [113]), whose performance is known to fall far short of optimal. Another challenge to leveraging sparsity is accurate tracking of the locations of sparse taps, which can change quite rapidly (see, e.g., [102]). While clever order-recursive matching-pursuit algorithms have been proposed for this purpose (e.g., [113]), their complexity remains quadratic in the non-sparse channel length N_h , which (as we have seen) can be very large; we would like something that is much simpler.

In this paper, we propose a turbo JED receiver, building on our earlier work [89, 114], that operates in a pulse-shaped MCM framework and which takes full advantage of sparsity in the channel delay profile. Our approach uses a (sparse) Fourier-BEM to model frequency-domain channel variation and a fast tree-search to compute the soft noncoherent equalizer outputs. In addition, we propose a simple pilot-aided means

of tracking the channel's quickly-varying delay-power profile. The performance of our algorithm is evaluated numerically using simulated channels whose sparsity and delay/Doppler spreads mimic those of the surf-zone channels from [102]. The proposed scheme is found to exhibit excellent BER performance relative to genie-aided bounds, while maintaining high spectral efficiency and relatively low complexity.

6.1 System Model

First we describe a discrete-time complex-baseband model of our system, which includes coded multicarrier modulation (MCM) and a sparse doubly dispersive channel.

At the transmitter, information bits are rate- R coded and mapped to 2^q -ary scalar data symbols. Groups of N_s scalar data symbols are then combined with scalar pilot/guard symbols to form transmission blocks of length $N \geq N_s$. (Pilot and guard details will be given later.) Using N subcarriers, the j^{th} MCM-symbol is composed of the scalar symbols $\{c_k^{(j)}\}_{k=0}^{N-1}$ corresponding to the coded bits $\{x_n^{(j)}\}_{n=0}^{N_s q-1}$. In particular, the n^{th} scalar symbol is mapped from the coded bits $\underline{\mathbf{x}}_n^{(j)} \triangleq [x_{nq}^{(j)}, \dots, x_{nq+q-1}^{(j)}]^T$.

6.1.1 Modifications for Noncoherent Soft Equalization

In Section 5.2.2, we describe a noncoherent soft equalization scheme based on tree search. A frequency-domain guard pattern that facilitates this tree search will now be described.

From (A.15), it can be seen that every element in $\mathbf{y}^{(j)}$ sees contributions from $2D+1$ subcarriers. For tree search, we would like that the first observation contains a contribution from only one unknown scalar symbol, the second contains contributions from only two unknown scalar symbols, the third from only three unknown scalar

symbols, and so on. One way to ensure this is to set $\{c_{\langle k \rangle_N}^{(j)}\}_{k=-D}^{D-1} = 0$, i.e., to “turn off” the first and last D subcarriers—a technique commonly used to prevent adjacent-channel interference in channelized systems. Note that the resulting loss in spectral efficiency will be small when $2D \ll N$.

To proceed further, it is convenient to define the D -shifted quantities $\check{c}_k^{(j)} = c_{\langle k+D \rangle_N}^{(j)}$ and $\check{\mathbf{c}}^{(j)} \triangleq [\check{c}_0^{(j)}, \dots, \check{c}_{N-1}^{(j)}]^T$, noticing that the last $2D$ elements in $\check{\mathbf{c}}^{(j)}$ constitute a zero-valued guard interval. Since $\mathcal{D}_k(\check{\mathbf{c}}^{(j)}) = \mathcal{D}_{k-D}(\mathbf{c}^{(j)})$ for any k , we can rewrite (A.15) as

$$\mathbf{y}^{(j)} = \mathbf{A}^{(j)} \boldsymbol{\theta}^{(j)} + \mathbf{w}^{(j)} \quad (6.1)$$

$$\mathbf{A}^{(j)} \triangleq [\mathcal{D}_{2D}(\check{\mathbf{c}}^{(j)}) \mathbf{B}^{(j)}, \dots, \mathcal{D}_0(\check{\mathbf{c}}^{(j)}) \mathbf{B}^{(j)}] \quad (6.2)$$

and see that, for each $k \in \{0, \dots, N-1\}$, the observations $\{y_d^{(j)}\}_{d=0}^k$ depend only on $\{\check{c}_d^{(j)}\}_{d=0}^k$. The now “causal” ICI channel allows us to write the partial observation $\mathbf{y}_k^{(j)} \triangleq [y_0^{(j)}, \dots, y_k^{(j)}]^T$ as

$$\mathbf{y}_k^{(j)} = \mathbf{A}_k^{(j)} \boldsymbol{\theta}^{(j)} + \mathbf{w}_k^{(j)}, \quad (6.3)$$

where $\mathbf{w}_k^{(j)} \triangleq [w_0^{(j)}, \dots, w_k^{(j)}]^T$ and where $\mathbf{A}_k^{(j)}$ appends a new row $\mathbf{a}_k^{(j)H} \in \mathbb{C}^{(2D+1)N_a}$ with each k :

$$\mathbf{A}_k^{(i)} = \begin{bmatrix} \mathbf{a}_0^{(j)H} \\ \vdots \\ \mathbf{a}_k^{(j)H} \end{bmatrix} \quad (6.4)$$

$$\mathbf{a}_k^{(j)H} = [\check{c}_{k-2D}^{(j)} \mathbf{b}_k^{(j)H}, \dots, \check{c}_k^{(j)} \mathbf{b}_k^{(j)H}]. \quad (6.5)$$

In (6.5), $\mathbf{b}_k^{(i)H}$ denotes the k^{th} row of $\mathbf{B}^{(i)}$. Note that the full-block quantities $\mathbf{y}_{N-1}^{(j)}$, $\mathbf{A}_{N-1}^{(j)}$, $\check{\mathbf{c}}_{N-1}^{(j)}$, and $\mathbf{w}_{N-1}^{(j)}$ are identical to the previously defined $\mathbf{y}^{(j)}$, $\mathbf{A}^{(j)}$, $\check{\mathbf{c}}^{(j)}$, and $\mathbf{w}^{(j)}$, respectively.

6.2 Noncoherent Soft Equalization

As illustrated in Fig. 6.1, the receiver consists of a soft noncoherent equalizer and a soft decoder, connected in a “turbo” configuration, as well as a delay-power profile (DPP) estimator. The DPP estimator uses the observations and pilots to estimate the active-tap indices $\mathcal{L}^{(j)}$, as will be described in Section 6.3. The equalizer uses the observations $\mathbf{y}^{(j)}$, as well as any *a priori* information provided by the decoder, to generate soft information on the coded bits $\mathbf{x}^{(j)} \triangleq [\mathbf{x}_0^{(j)T}, \dots, \mathbf{x}_{N_s-1}^{(j)T}]^T$, leveraging its knowledge of the pilot symbols and statistical channel structure, including $\mathcal{L}^{(j)}$. The decoder then uses the soft equalizer outputs (from possibly many MCM-symbols) to refine the soft information on the coded bits, leveraging its knowledge of the code structure. After a sufficient number of turbo iterations, the decoder outputs a hard estimate of the information bits.

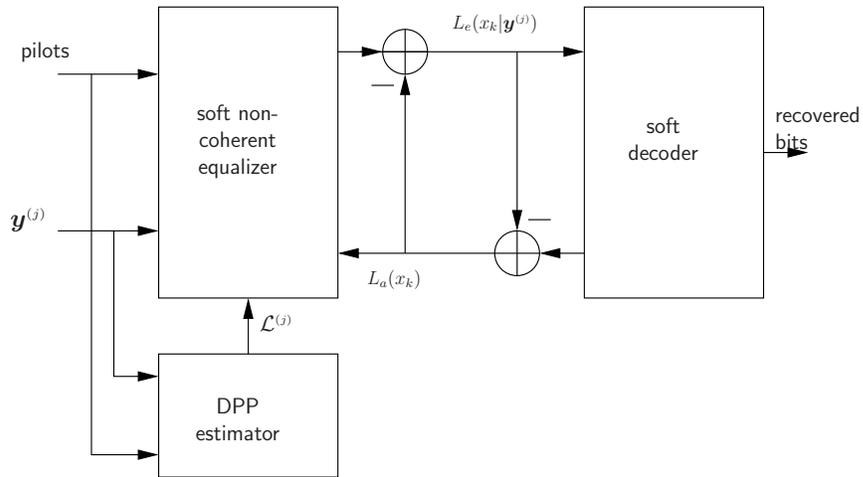


Figure 6.1: Receiver structure.

6.3 Tracking Sparsity

In this section, we present an efficient means of learning the active-tap locations $\mathcal{L}^{(j)}$ and compensating for the fact that, in practice, “inactive” taps are non-zero (but small).

6.3.1 Active-Tap Identification

To estimate the locations of active taps, we compute a pilot-based minimum mean-squared error (MMSE) estimate of the non-sparse BEM coefficient vector $\check{\boldsymbol{\theta}}^{(j)} \triangleq [\check{\boldsymbol{\theta}}_{-D}^{(j)T}, \dots, \check{\boldsymbol{\theta}}_D^{(j)T}]^T$ and from that estimate the DPP $\{\sigma_l^{(j)2}\}_{l=0}^{N_h-1}$ as follows.

$$\widehat{\sigma_l^{(j)2}} = \sum_{d=-D}^D \left| [\check{\boldsymbol{\theta}}_d^{(j)}]_l \right|^2. \quad (6.6)$$

Note that this approach permits accurate DPP estimation even when the channel gain $h_{n,l}^{(j)}$ varies significantly over the MCM-symbol interval. Once the DPP has been estimated, we set¹⁹ $\mathcal{L}^{(j)}$ as the largest N_a indices of the DPP. Note that the ability to assign $\mathcal{L}^{(j)}$ directly from the DPP is a consequence of our BEM’s orthogonality (i.e., \mathbf{F}). Without BEM orthogonality, estimating $\mathcal{L}^{(j)}$ becomes much more complicated (e.g., [113]).

As illustrated in Fig. 6.2, our pilot pattern employs $N_p = N/P$ pilot/guard subcarriers per MCM-symbol and repeats after every P MCM-symbols. Each MCM-symbol contains $K \geq 1$ pilot clusters, where each cluster is comprised of $N_p - 2KD$ non-zero pilots with D zero-valued guards on either edge. The cluster locations are staggered so that each subcarrier is used in a cluster exactly once every P MCM-symbols. Note

¹⁹Note that this approach yields a fixed complexity; if a fixed performance was more important, one could set $\mathcal{L}^{(j)}$ to be the indices of DPP values that lie above a threshold.

that, on average, $\frac{P-1}{P}$ subcarriers are used for data. We experiment with different choices of K in Section 6.4.

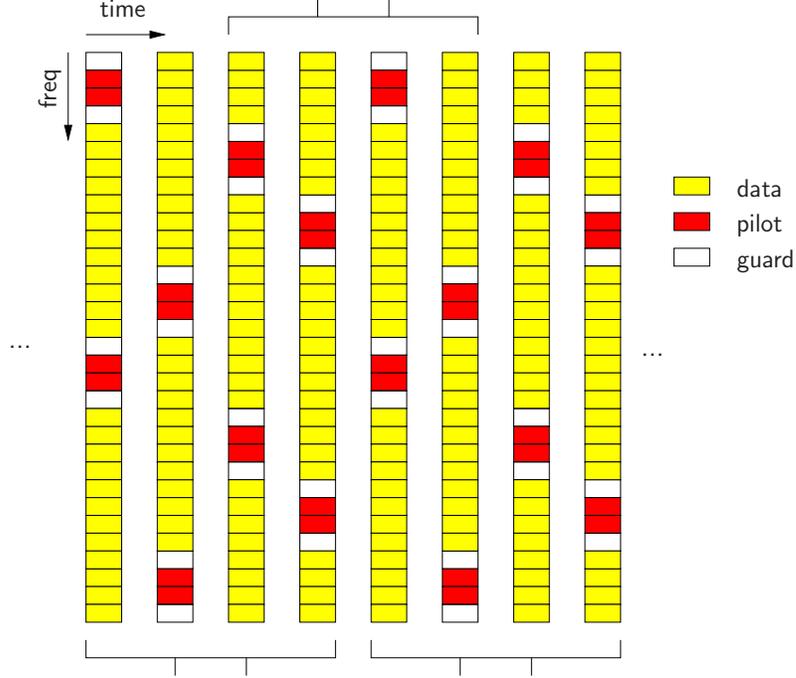


Figure 6.2: Illustration of pilot pattern with $N = 32$, $P = 4$, $K = 2$, $N_p = 8$, and $D = 1$. The columns represent $\check{\mathbf{c}}^{(j-P+2)}, \dots, \check{\mathbf{c}}^{(j+2P-3)}$, respectively.

The pilots play a twofold²⁰ role. First, as discussed in Section 5.4.2, they have the potential to significantly improve the complex/performance tradeoff of suboptimal tree-search. Second, they facilitate the tracking of active taps $\mathcal{L}^{(j)}$. Recall that DPP sparsity can lead to significant reductions in equalization complexity since the latter is quadratic in the number of modelled taps (i.e., $N_a = |\mathcal{L}^{(j)}|$). The guards also play a

²⁰Note that pilots could also be used for adjustment of *large* timing clock offsets and carrier frequency offsets. For example, the receiver's sampling and carrier frequencies could be adjusted so that the support of the measured delay/Doppler profile stays close to the origin. There is no need to compensate for small timing/carrier offsets, though, since they are subsumed by the doubly dispersive channel.

twofold role. First, as described in Section 6.1.1, they facilitate tree-search by making the ICI channel (6.1) appear causal. Second, they ensure that $N_p - 2KD$ subcarriers in $\mathbf{y}^{(j)}$ will be free of interference from unknown data, thereby improving the quality of pilot-based estimates of $\mathcal{L}^{(j)}$.

We now detail our pilot-based DPP estimation procedure, assuming power-of-two P . At even symbol indices j , we use the pilot-only subcarriers in $\{\mathbf{y}^{(j)}, \dots, \mathbf{y}^{(j+P-1)}\}$ to jointly estimate the channel vectors $\{\check{\boldsymbol{\theta}}^{(j+P/4)}, \dots, \check{\boldsymbol{\theta}}^{(j+3P/4-1)}\}$. Here, estimates for the $P/4$ left and right “edge” vectors are not attempted because we anticipate that they would be unreliable. However, since estimation is performed at every even symbol index i , all channel vectors will eventually be estimated. (See Fig. 6.2 for a $P = 4$ example.) Notice that a total of $P(N_p - 2KD) = N - 2PKD$ scalar observations are used to estimate $\frac{P}{2}(2D + 1)N_h$ scalar BEM coefficients during every other MCM-symbol index. Since MMSE estimation is implemented by multiplication with a (fixed and known) $\frac{P}{2}(2D + 1)N_h \times (N - 2PKD)$ matrix, the cost of DPP estimation is only $\frac{P}{4}(2D + 1)N_h(N - 2PKD)$ multiplications per MCM-symbol. Recalling Section 5.2.2, this cost is small relative to that of iterative noncoherent soft equalization.

The MMSE estimator matrix can be designed as follows. Say that we collect the pilot-only observations from $\{\mathbf{y}^{(j)}, \dots, \mathbf{y}^{(j+P-1)}\}$ into the vector $\underline{\mathbf{y}}^{(j)} \in \mathbb{C}^{N-2PKD}$ and the corresponding noise samples $\{w_k\}$ into the vector $\underline{\mathbf{w}}^{(j)}$. Then, considering (6.1) with non-sparse coefficients $\check{\boldsymbol{\theta}}^{(j)}$, we can write

$$\underline{\mathbf{y}}^{(j)} = \underline{\mathbf{A}}\check{\boldsymbol{\theta}}^{(j)} + \underline{\mathbf{w}}^{(j)}, \quad (6.7)$$

where $\underline{\mathbf{A}}$ is a block-diagonal pilot matrix (with P blocks) and $\check{\underline{\boldsymbol{\theta}}}^{(j)} \triangleq [\check{\boldsymbol{\theta}}^{(j)T}, \dots, \check{\boldsymbol{\theta}}^{(j+P-1)T}]^T$. The MMSE estimate of $\check{\underline{\boldsymbol{\theta}}}^{(j)}$ from $\underline{\mathbf{y}}^{(j)}$ is then

$$\hat{\underline{\boldsymbol{\theta}}}^{(j)} = \mathbf{R}_{\check{\underline{\boldsymbol{\theta}}}} \underline{\mathbf{A}}^H (\underline{\mathbf{A}} \mathbf{R}_{\check{\underline{\boldsymbol{\theta}}}} \underline{\mathbf{A}}^H + \sigma^2 \mathbf{I})^{-1} \underline{\mathbf{y}}^{(j)}, \quad (6.8)$$

where $\mathbf{R}_{\check{\underline{\boldsymbol{\theta}}}}$ denotes the autocovariance matrix for $\check{\underline{\boldsymbol{\theta}}}^{(j)}$, which can be constructed from the WSSUS model (A.18)-(A.20) under an *a priori* uniform DPP, i.e., $\sigma_{l_\nu^{(j)}}^{(j)2} = N_h^{-1}$ for $\nu \in \{0, \dots, N_h - 1\}$ and $l_\nu^{(j)} = \nu$. Finally, DPP estimates of $\{\sigma_l^{(j+P/4)2}, \dots, \sigma_l^{(j+3P/4-1)2}\}_{l=0}^{N_h-1}$ are computed via (6.6).

6.3.2 Residual Tap Compensation

Because non-active channel taps are ignored by the non-coherent equalizer, they have the effect of contributing additional noise. Because the size and number of non-active channel taps can vary, so can the power of the additional noise. Thus, it should be tracked for use by the equalizer. Fortunately, doing so is relatively easy. For example, after the noncoherent sequence detection algorithm has estimated $\mathbf{A}^{(j)}$ and $\boldsymbol{\theta}^{(j)}$ in (6.1), the residual interference $\hat{\mathbf{w}}^{(j)} \triangleq \mathbf{y}^{(j)} - \hat{\mathbf{A}}^{(j)} \hat{\boldsymbol{\theta}}^{(j)}$ can be computed and its energy calculated to retrieve an estimate of the effective noise power $\sigma^{(j)2}$. This estimate could then be employed for noncoherent equalization of the $(j+1)^{th}$ MCM-symbol, since the sparsity is not expected to change significantly from one MCM-symbol to the next.

6.4 Numerical Results

Numerical tests of the proposed multicarrier system were conducted using MATLAB, with Rayleigh channel coefficients generated according to Jakes method [95].

6.4.1 Setup

Two types of channel were considered, a “perfectly sparse” channel and a “sparse” channel; the latter aims to be realistic while the former is useful as a reference. The impulse response of the “perfectly sparse” channel had $N_c = 5$ nonzero fading coefficients spread over $N_h = 100$ chip intervals. These coefficients were samples of zero-mean circular²¹ Gaussian random processes, each with identical power $\sigma_l^2 = 1/N_c$ and autocorrelation $\rho_n = J_0(2\pi f_D T_c n)$. Here, $J_0(\cdot)$ denotes the 0th-order Bessel function of the first kind, and $f_D T_c = 0.0005$ was chosen for the single-sided normalized Doppler spread. Our choice of $\{f_D T_c, N_h\}$ corresponds to, for example, Doppler spread 10 Hz and delay spread 10 ms if the transmission bandwidth was $\frac{1}{T_c} = 10$ kHz, or Doppler spread 5 Hz and delay spread 20 ms if the transmission bandwidth was $\frac{1}{T_c} = 5$ kHz. Furthermore, it corresponds to a delay/Doppler-spread product of $2f_D T_h = 2f_D T_c N_h = 0.1$, matching that of the surf-zone channel from [102]. To generate the more realistic “sparse” channel model, 2% of the active-tap energy was leaked into the inactive taps. This was accomplished by convolving the “perfectly sparse” impulse response $\{h_{n,l}\}_{l=0}^{N_h-1}$, at each time n , with the truncated sinc sequence $[-0.0721, 0.0739, 0.9893, 0.0739, -0.0721]$.

To model a time-varying DPP, the discrete delays of the active taps were varied in time as follows. While the delays of the first and second taps were fixed at 2 and 15 chips, the delay of the third changed from 30 to 34 chips with a period of 225 MCM-symbol intervals. Furthermore, the delay of the fourth changed from 56 to 66 chips, and then back, with a period of 60 MCM-symbol intervals, and the delay of the

²¹Here, and throughout the paper, we refer to the complex-baseband equivalent model of the channel.

fifth changed from 87 to 97 chips, and then back, with a period of 60 MCM-symbol intervals. Thus, it can be seen that the delay spread of the “sparse” channel was $N_h = 100$. As a consequence of the time-varying tap delays, the effective (normalized single-sided) Doppler spread is actually larger than $f_D T_c = 0.0005$.

At the transmitter, information bits were coded via rate- $R = \frac{1}{2}$ irregular low density parity check (LDPC) codes with average column-weight 3, generated via the publicly available software [96]. The coded bits were then mapped to BPSK symbols (i.e., $Q = 1$) and partitioned into data blocks of length $N_s = N - N_p$, each of which was merged with $N_p = 128$ pilot/guard symbols to form an MCM-symbol of length $N = 512$. We used random BPSK pilots arranged as in Fig. 6.2 with $P = 4$, $D = 1$, and $K = 2$ (unless otherwise noted). So that each codeword spanned exactly $J = 32$ MCM-symbol intervals, $(JQN_s, RJQN_s)$ -LDPC codes were employed. For the MCM pulses $\{\alpha_n\}$ and $\{\beta_n\}$, we used the “transmitter optimized max-SINR” design from [5], which specifies a smooth modulation pulse of length $N_\alpha = 1.5N$ and a rectangular demodulation pulse of length $N_\beta = N$. Since we employed no MCM guard interval, our modulation efficiency (taking the pilots/guards into account) was 0.75 symbols/sec/Hz.

The receiver employed an ICI radius of $D = \lceil f_D T_c N \rceil = 1$ and used $N_a = 10$ sparse taps in its BEM. For noncoherent soft equalization, sequential algorithm described in Table 5.1 is employed with the LLR clipping threshold set at 3.0. The M-algorithm used a search breadth of $M = 32$. A maximum of 60 sum-product decoding iterations were allowed, while a maximum of 8 turbo (i.e., equalization/decoding) iterations were allowed. Note, however, that decoder and turbo iterations are terminated as

soon as the LDPC decoder senses that the bits have been decoded without error (which usually happens very quickly).

6.4.2 BER versus Number of Pilot Clusters K

We first investigated the bit error rate (BER) versus the number of pilot clusters K . As will be seen, the proper choice of K is quite important. However, the theory of choosing K (e.g., [92, 115]) is based on simplified (e.g., non-sparse) channel models and MSE-minimization (rather than coded-BER minimization) and thus falls short of being useful. In Fig. 6.3, we plot the BER versus $K \in \{1, 2, 4, 8, 16, 32\}$. Since we fix $N_p = 128$, these values of K correspond to cluster sizes of $\{128, 64, 32, 16, 8, 4\}$, respectively. Figure 6.3 shows that $K = 2$ yields the minimum BER, with $K = 1$ also performing quite well. Thus, for our setup, a few large clusters seems to perform better than many smaller clusters.

6.4.3 BER versus SNR

Next we investigated BER versus E_b/N_o , i.e., the ratio of energy-per-information-bit to noise power-spectral-density. (Note that pilot/guards do not contribute to the information-bit energy.) In addition to simulating the BER of the proposed soft noncoherent receiver, we also simulated the BER of several genie-aided reference receivers. In one reference receiver, we replace our soft noncoherent equalizer with the soft *coherent* equalizer from [88] that uses an MMSE estimate of $\boldsymbol{\theta}^{(j)}$ computed under the assumption that, in addition to the pilot/guards, *all data subcarriers are known for the purpose of estimating $\boldsymbol{\theta}^{(j)}$* . Notice that this “genie-aided- $\hat{\boldsymbol{\theta}}$ coherent reference” upper bounds the performance of any noncoherent equalizer with a bound tighter than that of coherent equalization under *perfect* channel knowledge. Since,

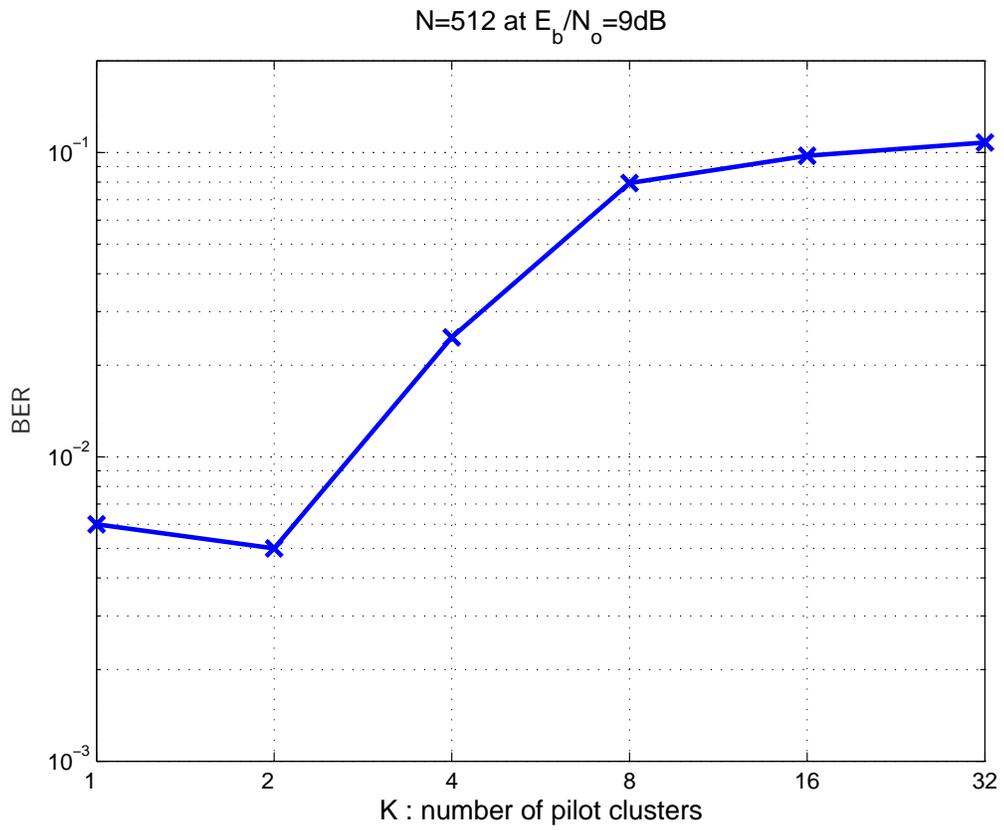


Figure 6.3: BER versus number of pilot clusters for the “sparse” channel at $E_b/N_o=9$ dB.

however, both the noncoherent and genie-aided- $\hat{\theta}$ equalizers still need to track the active coefficients $\mathcal{L}^{(j)}$, we consider these two schemes with perfectly known $\mathcal{L}^{(j)}$ as additional references.

Fig. 6.4 shows BER versus E_b/N_o for the “perfectly sparse” channel. The performance of the proposed soft noncoherent receiver is very close to the genie-aided- $\hat{\theta}$ coherent receiver; less than 1.5 dB SNR loss (at 10^{-2} BER) can be observed. It can also be seen that reception under estimated $\mathcal{L}^{(j)}$ performs nearly as well as reception under perfectly known $\mathcal{L}^{(j)}$.

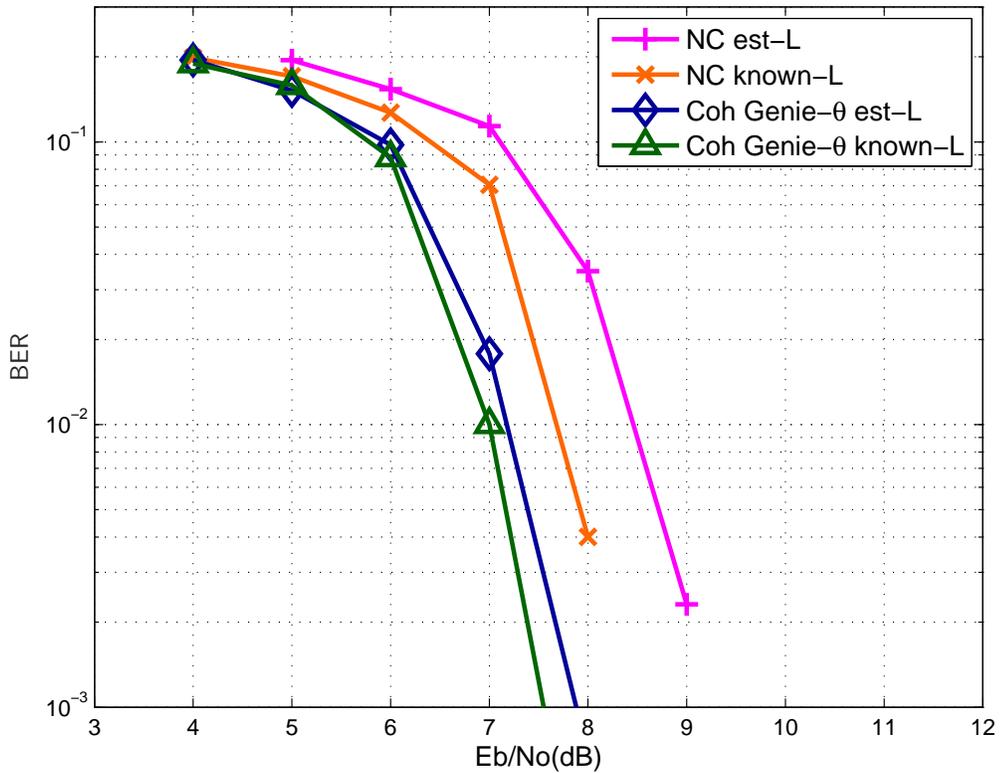


Figure 6.4: BER versus SNR for the “perfectly sparse” channel. The noncoherent soft equalizer is compared to coherent soft equalizer with genie-estimated $\theta^{(j)}$ for both true and estimated $\mathcal{L}^{(j)}$.

Fig. 6.5 shows BER versus E_b/N_o for the more realistic “sparse” channel. Here again, the proposed noncoherent receiver performs less than 1.5 dB worse (at 10^{-2} BER) than the genie-aided- θ reference, and both schemes suffer by less than 1 dB when an imperfect estimate of $\mathcal{L}^{(j)}$ is used. Since $N_a \ll N_h$, many residual channel taps remain unmodeled. Thus even the best selection of $\mathcal{L}^{(j)}$ leaves about 1.6% of the received signal energy as unmodeled and thus acting as noise. The resulting noise floor could be suppressed, at the expense of receiver complexity, by increasing N_a .

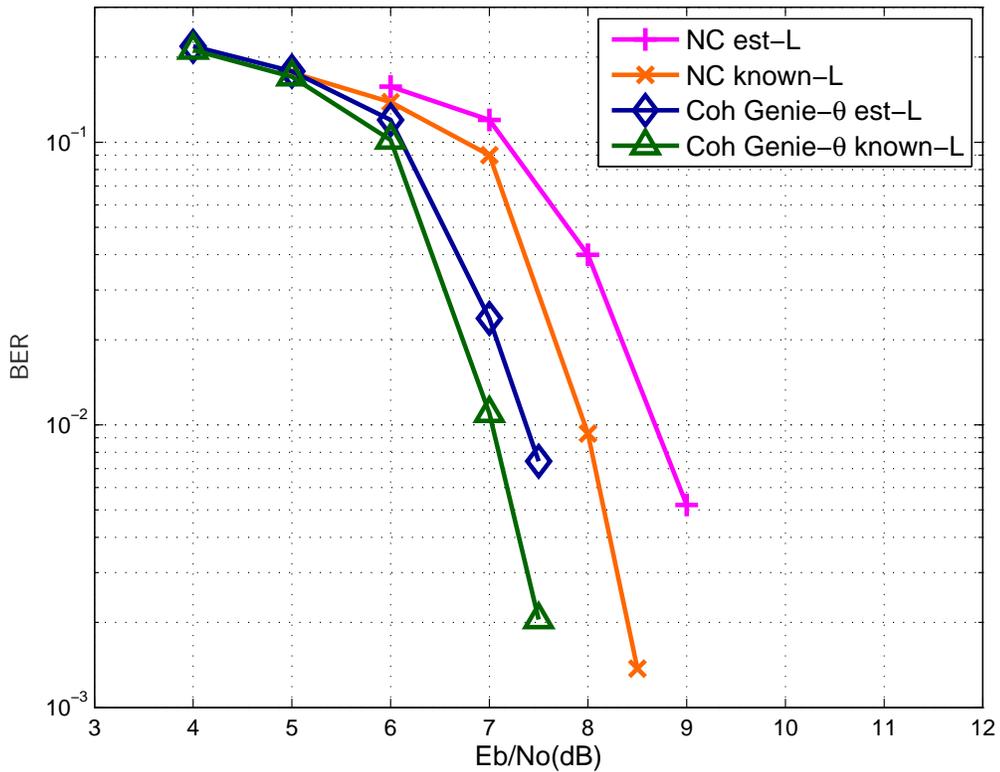


Figure 6.5: BER versus SNR for the “sparse” channel. The noncoherent soft equalizer is compared to coherent soft equalizer with genie-estimated $\theta^{(j)}$ for both true and estimated $\mathcal{L}^{(j)}$.

CHAPTER 7

MAXIMUM-DIVERSITY PRECODING

We first characterize the maximum achievable diversity order for noncoherent (coherent) communication over the DS channel, and find (for wide-sense stationary uncorrelated scattering (WSSUS) channels with limited time-frequency spread) that the diversity order equals the product of temporal and spectral diversity orders, thereby coinciding with the maximum diversity order for coherent communication over the DS channel [42,116]. For our analysis, we leverage certain asymptotic results from the noncoherent pairwise error probability (PWEPP) analysis in [117,118]. Next, we show that (under mild channel conditions) *almost any* affine precoder facilitates maximum diversity reception. We also show that linear precoding [119,120] does not facilitate maximum diversity reception for commonly used symbol alphabets (e.g., uncoded QAM or PSK) for noncoherent channel. But for coherent channel, we establish the fact that *almost any* linear precoding facilitates the maximum diversity reception (under mild channel condition), which is shown as a special case of affine precoding results. Recall that linear (affine) precoding [121] refers to the general class of schemes which combine linear processing of the information symbols (with additive training). It is interesting to note that, while the maximum-diversity precoder proposed for the coherent case in [42] led to a high degree of transmit-signal redundancy, both the

affine noncoherent precoders and linear coherent precoders considered here are not rate-constrained in any way.

7.1 System Model

For convenience of the discussion, we consider a reverse indexing of the codeword and channel, i.e., consider block transmission of a codeword $\mathbf{c} = [c_{N-1}, c_{N-2}, \dots, c_0] \in \mathcal{C}$, where $\mathcal{C} \subset \mathbb{C}^N$ is a finite set of candidate codewords, $\mathbf{h}_\ell \triangleq [h_{N-1,\ell}, h_{N-2,\ell}, \dots, h_{0,\ell}]^T$. The random vector \mathbf{h}_ℓ , can be expressed (without loss of generality) using its Karhunen-Lòeve (KL) expansion as $\mathbf{h}_\ell = \mathbf{B}_\ell \boldsymbol{\theta}_\ell$, where $\mathbf{B}_\ell \in \mathbb{C}^{N \times N_b}$ is a fixed basis matrix such that $\mathbf{B}_\ell^H \mathbf{B}_\ell = \mathbf{I}_{N_b}$, and where $\boldsymbol{\theta}_\ell \in \mathbb{C}^{N_b}$ is a zero-mean circular Gaussian random vector. The parameter $N_b \leq N$ quantifies the degrees of variation in the channel tap. In cases of practical interest, the channel varies slowly enough that $N_b \ll N$. We furthermore assume that our channel exhibits WSS uncorrelated scattering (WS-SUS), so that $\boldsymbol{\theta} \triangleq [\boldsymbol{\theta}_0^T, \dots, \boldsymbol{\theta}_{N_h-1}^T]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\theta)$, where \mathbf{R}_θ has full rank $N_h N_b$. In addition, we assume that each tap has the same Doppler profile, so that $\mathbf{B}_\ell = \mathbf{B} \forall \ell$.

Using \mathbf{b}_n^H to denote the row of \mathbf{B} such that $h_{n,\ell} = \mathbf{b}_n^H \boldsymbol{\theta}_\ell$, the model (A.2) can be rewritten, for $n \in \{0, \dots, N-1\}$, as

$$\mathbf{y} = \mathbf{C}\boldsymbol{\theta} + \mathbf{w}, \quad (7.1)$$

where

$$\mathbf{C} = \begin{bmatrix} c_{N-1} \mathbf{b}_{N-1}^H & \cdots & c_{N-N_h} \mathbf{b}_{N-1}^H \\ \vdots & & \vdots \\ c_1 \mathbf{b}_1^H & \cdots & c_{-N_h+2} \mathbf{b}_1^H \\ c_0 \mathbf{b}_0^H & \cdots & c_{-N_h+1} \mathbf{b}_0^H \end{bmatrix}. \quad (7.2)$$

For simplicity, we assume that $c_n = 0$ for $n < 0$, as occurs when block transmissions are separated by zero-valued guards with duration $\geq N_h - 1$. However, we note that

such guards may not be needed in the high-SNR regime, where good estimates of $\{c_n\}_{n<0}$ are available from previously detected blocks and thus do not pose a problem when detecting the unknown codeword \mathbf{c} .

We assume that the receiver knows the channel statistics, i.e., \mathbf{B} and \mathbf{R}_θ , but not the channel realization. In this case, the noncoherent ML estimate of $\mathbf{c} \in \mathcal{C}$ has the well known form [53, 117]

$$\hat{\mathbf{c}}_{\text{ML,nc}} = \arg \min_{\mathbf{c} \in \mathcal{C}} \mathbf{y}^H \Phi^{-1} \mathbf{y} + \log \det \Phi \quad (7.3)$$

$$\Phi \triangleq \mathbf{C} \mathbf{R}_\theta \mathbf{C}^H + \sigma^2 \mathbf{I}_N. \quad (7.4)$$

On the other hand, the coherent ML estimate of the codeword is expressed as

$$\hat{\mathbf{c}}_{\text{ML,c}} = \arg \min_{\mathbf{c} \in \mathcal{C}} \|\mathbf{y} - \mathbf{C}\theta\|^2. \quad (7.5)$$

7.2 Diversity-Order Analysis

In this section, we quantify the diversity order attained by the noncoherent ML detector over DS channel via pairwise error probability (PWE) analysis, leveraging the work of Brehler and Varanasi [117] and Siwamogsatham, Fitz, and Grimm [118].

7.2.1 Noncoherent Receiver

Let \mathbf{c}_k denote the k^{th} codeword in \mathcal{C} , and let the corresponding versions of \mathbf{C} and Φ be denoted by \mathbf{C}_k and Φ_k , respectively. Then E_{kl} , the event that \mathbf{c}_k is transmitted and $\mathbf{c}_{l \neq k}$ is chosen by the ML detector, becomes

$$E_{kl} = \{\mathbf{y}^H \Phi_k^{-1} \mathbf{y} + \log \det \Phi_k > \mathbf{y}^H \Phi_l^{-1} \mathbf{y} + \log \det \Phi_l\}. \quad (7.6)$$

A closed-form expression for the PWEF $\Pr\{E_{kl}\}$ has been derived [117, 118] for the high-SNR asymptotic case, i.e., $\sigma^2 \rightarrow 0$. Adapted to the specifics of our model, the result can be summarized as follows:

Lemma 3 (High-SNR PWEF for Noncoherent Detector [117, 118]) *If the matrix*

$$\mathbf{M}_{kl} \triangleq \mathbf{C}_k^H (\mathbf{I}_N - \mathbf{C}_l (\mathbf{C}_l^H \mathbf{C}_l)^{-1} \mathbf{C}_l^H) \mathbf{C}_k \quad (7.7)$$

has full rank $N_h N_b$, then, as $\sigma^2 \rightarrow 0$,

$$\Pr\{E_{kl}\} \rightarrow \left(\frac{1}{\sigma^2}\right)^{-N_h N_b} \det(\mathbf{R}_\theta \mathbf{M}_{kl})^{-1} \sum_{m=0}^{N_h N_b - 1} \binom{2N_h N_b - m}{N_h N_b} \eta_m, \quad (7.8)$$

where

$$\eta_m \triangleq \begin{cases} \frac{(-\zeta_{lk})^m e^{\zeta_{lk}}}{m!}, & \text{for } \zeta_{lk} \leq 0 \\ \frac{\zeta_{lk}^m}{m!}, & \text{for } \zeta_{lk} > 0 \end{cases} \quad (7.9)$$

$$\zeta_{kl} \triangleq N \log \frac{\det(\mathbf{C}_k^H \mathbf{C}_k)}{\det(\mathbf{C}_l^H \mathbf{C}_l)} \quad (7.10)$$

Lemma 3 establishes that the maximum achievable diversity order equals $N_h N_b$, and that achieving this maximum diversity order requires that \mathbf{M}_{kl} be full rank for all k and all $l \neq k$.

We now translate the full-rank condition on \mathbf{M}_{kl} to a more convenient form.

Lemma 4 *\mathbf{M}_{kl} has full rank $N_h N_b$ if and only if $[\mathbf{C}_k, \mathbf{D}_{lk}]$ has full rank $2N_h N_b$, where $\mathbf{D}_{lk} \triangleq \mathbf{C}_l - \mathbf{C}_k$.*

proof : From (7.7), we see that \mathbf{M}_{kl} shares the rank of $\mathbf{\Pi}_l^\perp \mathbf{C}_k \mathbf{C}_k^H$, where $\mathbf{\Pi}_l^\perp \triangleq \mathbf{I}_N - \mathbf{C}_l (\mathbf{C}_l^H \mathbf{C}_l)^{-1} \mathbf{C}_l^H$ accomplishes projection onto the null space of \mathbf{C}_l . Since $\mathbf{C}_k \in \mathbb{C}^{N \times N_b N_h}$, full rank \mathbf{M}_{kl} occurs iff the following two conditions are satisfied: \mathbf{C}_k has full rank $N_b N_h$, and the column space of \mathbf{C}_k is contained in the null space of \mathbf{C}_l , i.e., the column spaces of \mathbf{C}_k and \mathbf{C}_l share no common subspace. In other words, \mathbf{M}_{kl} has full rank iff $[\mathbf{C}_k, \mathbf{C}_l]$ has full rank $2N_h N_b$. Furthermore, since rank is not affected

by subtracting the first $N_h N_b$ columns from the last, the rank of $[\mathbf{C}_k, \mathbf{C}_l]$ equals the rank of $[\mathbf{C}_k, \mathbf{D}_{lk}]$. ■

Lemma 4 states that, for full diversity noncoherent detection, the following must hold for all k and $l \neq k$: both the codeword matrix \mathbf{C}_k and the codeword-difference matrix \mathbf{D}_{lk} must be full rank, and their column spaces must not intersect. Notice that the full-rank condition requires that $N \geq 2N_h N_b$. This latter condition specifies the maximum degree of time-frequency spreading for which maximum-diversity reception is possible. Notice that the condition $N \geq 2N_h N_b$ is stronger than $N > N_h N_b$, the condition for an “underspread” channel.

Linear Precoding

We refer to the class of schemes in which the codewords are generated according to

$$\mathbf{c} = \mathbf{P}\mathbf{s}, \quad (7.11)$$

for general $\mathbf{P} \in \mathbb{C}^{N \times N_s}$, as *linear precoders* [119,120]. In this case, we associate the k^{th} codeword \mathbf{c}_k with the k^{th} symbol vector $\mathbf{s}_k \in \mathcal{S}$, where $\mathcal{S} \subset \mathbb{C}^{N_s}$ is a finite set.

Lemma 5 *Linear precoding does not facilitate maximum-diversity detection when $\exists \mathbf{s}_k, \mathbf{s}_l \in \mathcal{S}$ and $a \in \mathbb{C}$ such that $\mathbf{s}_k = a\mathbf{s}_l$, i.e., when \mathcal{S} contains symbol vectors which differ only by a scale factor.*

proof: With linear precoding, $\mathbf{s}_k = a\mathbf{s}_l$ implies $\mathbf{C}_k = a\mathbf{C}_l$, and hence $[\mathbf{C}_k, \mathbf{D}_{lk}] = [\mathbf{C}_k, (1-a)\mathbf{C}_k]$. Since this $[\mathbf{C}_k, \mathbf{D}_{lk}]$ has rank of at most $N_h N_b$, Lemmas 3 and 4 establish that this rank is insufficient for maximum-diversity detection. ■

The situation described in Lemma 5 is common and arises, e.g., when \mathbf{s} is composed of uncoded QAM or PSK symbols.

Affine Precoding

We refer to the class of schemes in which the codewords are generated according to

$$\mathbf{c} = \mathbf{P}\mathbf{s} + \mathbf{t}, \quad (7.12)$$

for general $\mathbf{P} \in \mathbb{C}^{N \times N_s}$ and $\mathbf{t} \in \mathbb{C}^N$, as *affine precoders* [121]. Here again, we associate the k^{th} codeword \mathbf{c}_k with the k^{th} symbol vector $\mathbf{s}_k \in \mathcal{S}$, where $\mathcal{S} \subset \mathbb{C}^{N_s}$ is a finite set. The affine precoder described in (7.12) is parameterized by a precoding matrix \mathbf{P} and a (superimposed) training vector \mathbf{t} . In this section, we demonstrate that *almost any* choice of $\{\mathbf{P}, \mathbf{t}\}$ is sufficient to facilitate maximum-diversity detection under some mild channel conditions. Before stating our result, we define $\tilde{\mathbf{B}}$ as the matrix created from the top $N - N_h + 1$ rows of \mathbf{B} , i.e.,

$$\tilde{\mathbf{B}} \triangleq \begin{bmatrix} \mathbf{b}_{N-1}^H \\ \mathbf{b}_{N-2}^H \\ \vdots \\ \mathbf{b}_{N_h-1}^H \end{bmatrix}. \quad (7.13)$$

Lemma 6 *If $N \geq 2N_hN_b$, if $\tilde{\mathbf{B}}$ is full rank, and if $[\mathbf{P}, \mathbf{t}]$ is chosen randomly from a distribution whose support contains an open ball in $\mathbb{C}^{N \times (N_s+1)}$, then $[\mathbf{C}_k, \mathbf{D}_{lk}]$ is full rank w.p.1. $\forall k$ and $\forall l \neq k$.*

proof: See Appendix C.1. ■

We now make some observations. First, Lemma 6 holds for general N_s , i.e., for precoders of arbitrary rate. Second, the rank condition on $\tilde{\mathbf{B}}$ is quite mild, and states that the first $N_h - 1$ samples (out of $N \geq 2N_hN_b$) of each tap trajectory are not essential to experiencing the N_b degrees of variation in channel tap. This is expected behavior for WSS channels. (Recall that \mathbf{B} satisfied $\mathbf{B}^H \mathbf{B} = \mathbf{I}_{N_b}$.)

Systematic Precodings

We refer such precodings having the form

$$\mathbf{c} = \begin{bmatrix} \mathbf{P}' \\ \mathbf{I}_{N_s} \end{bmatrix} \mathbf{s} + \mathbf{t}$$

as systematic precodings, where $\mathbf{P}' \in \mathbb{C}^{N_p \times N_s}$ with $N = N_p + N_s$.

This form of precodings can facilitate near-ML sequential detection at very low complexity (e.g., $\mathcal{O}(N)$ in [89]). With this precoding, it can be seen from (7.1) that the zeroth observation y_0 contains contribution from only one unknown symbol s_0 , the first observation y_1 contains contributions from only two unknown symbols, s_0 and s_1 , and so on. Otherwise, i.e., if the matrix \mathbf{P} were full matrix (without \mathbf{I}_{N_s}), then all the data symbols would contribute to the observation for every time instance, in which case the symbols could not be decoded by the sequential decoders.

Lemma 7 *For the maximum diversity, we need $N_p \geq N_h N_b - 1$.*

proof : For the maximum diversity, we need \mathbf{D}_{lk} to be full rank as a necessary condition. Considering the worst case, i.e., the lower N_s elements of $\mathbf{c}_l - \mathbf{c}_k$ are all zero except one position, the minimum N_p required for \mathbf{D}_{lk} to have $N_h N_b$ nonzero rows is $N_h N_b - 1$ over all k and $l \neq k$. ■

Again, we establish in this section that *almost any* choice of $\{\mathbf{P}', \mathbf{t}\}$ is sufficient to facilitate the maximum diversity under some conditions. To do this, we define $\check{\mathbf{B}}$ as the matrix created from the top $N_p - N_h + 1$ rows of \mathbf{B} , i.e.,

$$\check{\mathbf{B}} \triangleq \begin{bmatrix} \mathbf{b}_{N-1}^H \\ \mathbf{b}_{N-2}^H \\ \vdots \\ \mathbf{b}_{N_s+N_h-1}^H \end{bmatrix}. \quad (7.14)$$

Lemma 8 *If $N \geq 2N_h N_b$ and $N_p \geq N_h N_b - 1$, if $\check{\mathbf{B}}$ is full rank, and if entries of $\{\mathbf{P}', \mathbf{t}\}$ are chosen randomly from a distribution whose support contains an open ball in $\mathbb{C}^{N+N_p N_s}$, then $[\mathbf{C}_k, \mathbf{D}_{lk}]$ is full rank w.p.1. $\forall k$ and $\forall l \neq k$.*

proof : See Appendix C.2. ■

7.2.2 Coherent Receiver

We will follow the similar approach for the noncoherent receiver in Section 7.2.1. Then E_{kl} , the event that \mathbf{c}_k is transmitted and $\mathbf{c}_{l \neq k}$ is chosen by the coherent ML detector, becomes

$$E_{kl} = \{\|\mathbf{y} - \mathbf{C}_k \boldsymbol{\theta}\|^2 > \|\mathbf{y} - \mathbf{C}_l \boldsymbol{\theta}\|^2\}. \quad (7.15)$$

A closed-form expression for the PWEF $\Pr\{E_{kl}\}$ has been derived [117, 118] for the high-SNR asymptotic case, i.e., $\sigma^2 \rightarrow 0$. Adapted to the specifics of our model, the result can be summarized as follows:

Lemma 9 (High-SNR PWEF for Coherent Detector [117, 118]) *If the matrix*

$$\mathbf{M}_{kl} \triangleq (\mathbf{C}_k - \mathbf{C}_l)^H (\mathbf{C}_k - \mathbf{C}_l) \quad (7.16)$$

has full rank $N_h N_b$, then, as $\sigma^2 \rightarrow 0$,

$$\Pr\{E_{kl}\} \rightarrow \left(\frac{1}{\sigma^2}\right)^{-N_h N_b} \det(\mathbf{R}_\theta \mathbf{M}_{kl})^{-1} \binom{2N_h N_b - 1}{N_h N_b}. \quad (7.17)$$

We now state the full-rank condition on \mathbf{M}_{kl} for coherent detector in a more convenient form than Lemma 9.

Lemma 10 *$\mathbf{M}_{kl} = (\mathbf{C}_k - \mathbf{C}_l)^H (\mathbf{C}_k - \mathbf{C}_l)$ has full rank $N_h N_b$ if and only if \mathbf{D}_{lk} has full rank $N_h N_b$,*

Proof: This is obvious from the definition of \mathbf{D}_{lk} . ■

The following Lemma states that almost all linear precoding facilitates the maximum-diversity coherent ML detection.

Lemma 11 *If $N \geq N_h N_b$, if $\tilde{\mathbf{B}}$ is full rank, and if \mathbf{P} is chosen randomly from a distribution whose support contains an open ball in $\mathbb{C}^{N \times (N_s+1)}$, then \mathbf{D}_{lk} is full rank w.p.1. $\forall k$ and $\forall l \neq k$.*

Proof: We follow the same approach as the proof of Lemma 6: characterizing the \mathbf{P} which causes \mathbf{D}_{lk} to be rank deficient, and showing that these problematic \mathbf{P} are avoided w.p.1. We set $\boldsymbol{\alpha} = \mathbf{0}$ and $\mathbf{t} = \mathbf{0}$ in (C.1) and examine the rank deficiency condition to conclude that $\mathbf{p} \notin \mathcal{N}(\mathbf{H})$ w.p.1. ■

7.3 Numerical Examples

Figure 7.1 plots average PWEF versus SNR (σ^{-2}) for a randomly chosen affine precoder assuming an energy-preserving two-tap (i.e., $N_h = 2$) channel whose time evolution is governed by Jakes' model with $f_D T_s = 0.003$. By “average” PWEF, we mean that the PWEF is averaged across symbol pairs. Our experiments assumed $N = 8$, for which the channel model yields $N_b = 2$ (see Fig. 2.1). To demonstrate that the results hold for general N_s , Fig. 7.1 investigates $N_s \in \{6, 8, 10\}$, which covers the cases that $N_s > N$, $N_s = N$, and $N_s < N$. In all cases, it can be seen that the asymptotic slope of the average PWEF equals $-N_b N_h = -4$, which confirms full-diversity reception.

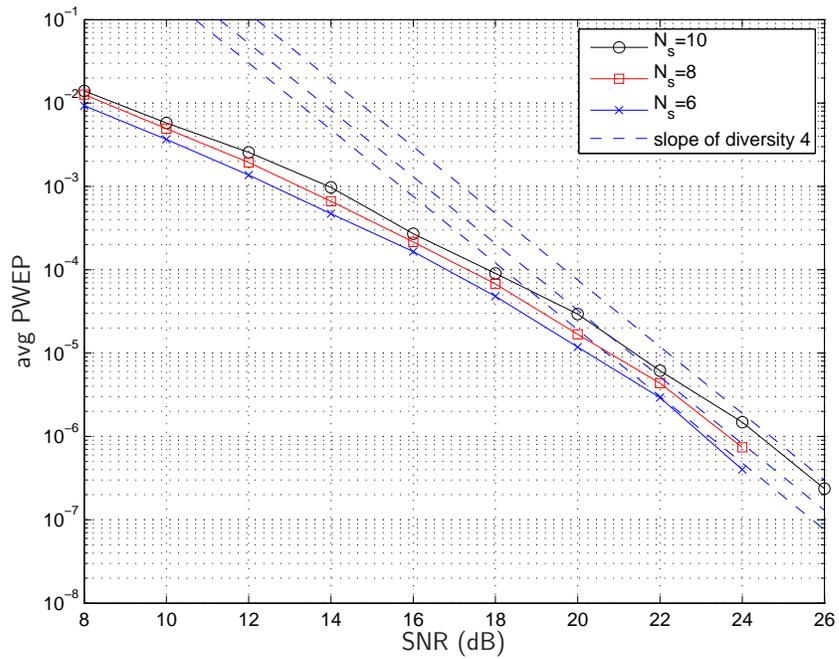


Figure 7.1: Average PWE versus SNR for $N = 8$, $N_h = N_b = 2$, and various N_s . The dashed line confirms the asymptotic slope of -4 .

CHAPTER 8

CONCLUSIONS

8.1 Summary of the Work

The problems in DS channel equalization have been an active research topics since channel estimation and optimal equalization are not straightforward compared to those in singly-selective channels. In response, we proposed practical near-optimal equalization schemes for both coherent receivers and noncoherent receivers. To achieve these tasks, efficient and near-optimal procedures for channel and symbol estimation methods were inevitable.

For efficient modeling of DS channel, we adopted basis expansion modeling and provided a unified description of the channel regardless whether the channel is modeled in time domain or in the frequency domain. The unified channel model could be made possible when we use cyclic-prefixed single-carrier modulation for the time-domain modeling and pulse-shaped multi-carrier modulation for the frequency-domain modeling.

For efficient near-optimal symbol estimation methods, we proposed to use tree-search algorithms (TSAs), which are suitable for both coherent equalization and noncoherent equalizations. Then, we discussed the performance of TSA preprocessing, and established that MMSE-GDFE preprocessing does not compromise the

ML-optimality of minimum-distance decisions for systems (including generic MIMO systems) with uncoded constant modulus signaling. This property holds for systems of arbitrary size (i.e., over- or under-determined linear channels), though not for non-constant modulus constellations (e.g., PAM and QAM). The result is attractive because MMSE-GDFE pre-processing is known to yield significant reductions in the average search complexity of optimal TSAs (e.g., SpD), especially in moderate-to-low SNR ranges and/or with ill-conditioned/under-determined linear channels.

As a coherent receiver structure for optimal (or near-optimal) DS channel equalizer, we proposed a means of high-spectral-efficiency MCM with practically realizable near-ML TSA suitable for communication over the DS channel. Our solution consisted of three components: 1) MCM that guarantees a small number of significant ICI/ISI coefficients while maintaining high spectral efficiency and white subchannel noise samples; 2) near-ML SD which leverages the pulse-shaped DS-channel ICI/ISI structure for low complexity over a wide SNR range; and 3) rank-reduced pilot-aided BE parameter estimation of significant ICI/ISI coefficients. Because traditional MCM schemes (i.e., CP-OFDM, ZP-OFDM, and S-OFDM) were shown (numerically) to yield too many non-negligible ISI/ICI coefficients, we utilized the PS-MCM scheme previously proposed by the authors in [5,17]. PS-MCM used in this research combined a SINR-maximizing transmission pulse with a rectangular reception pulse, permitting ICI/ISI truncation with negligible loss in performance. Because traditional ML and near-ML TSAs (e.g., Viterbi, Fano, and sphere decoders) were shown (numerically) to incur high search complexity when used with the PS-MCM DS channel, a novel TSA was proposed. The new TSA combines a fast MMSE-GDFE preprocessor with a channel-adaptive T-algorithm, of which both components were specifically tuned

to the PS-MCM DS-channel's ICI/ISI structure. The new TSA demonstrated FER performance indistinguishable from that of MLSD with a complexity that is approximately quadratic in the frame length, i.e., similar to that of fast MMSE-DFE. Finally, the use of rank-reduced pilot-aided channel estimates was observed to cause only a small loss in FER performance (relative to perfect channel knowledge) without significantly increasing the overall receiver complexity.

For near optimal noncoherent equalization, we proposed two schemes for use in a turbo receiver: sequential algorithm and Bayesian EM algorithm. For sequential algorithm, we started from the optimal noncoherent metric and we derived a novel fast algorithm to efficiently evaluate the metric. The other soft noncoherent equalizer was based on the use of the Bayesian EM algorithm to estimate the channel parameters, and it manifested as iterations between a soft coherent equalizer and a soft channel estimator. The proposed receivers modeled the channel via basis expansion (BE), and performed soft TSA. Efficient operations were accomplished using fast algorithms whose overall complexities grow linearly in the block size and quadratically in the number of BE parameters. Numerical studies show that two soft noncoherent equalizers perform equally well in mid-Doppler and remain robust to mismatch assumed Doppler spread. In particular, the sequential algorithm based noncoherent equalizer performs well in very high Doppler.

We also presented a novel multicarrier strategy for communication over UACs with simultaneously large delay and Doppler spreads. A multicarrier scheme employing a smooth transmission pulse was chosen to transform a time-varying ISI span of one hundred taps to an ICI span of three taps. Careful design of the pulse eliminated the need for a bandwidth-wasting cyclic (or zero) prefix. A turbo receiver, which passes

soft bit information between a noncoherent equalizer and an off-the-shelf decoder, was then described. The complexity of the noncoherent equalizer, which requires knowledge of the channel's statistics but not its realizations, is relatively low due to the use of suboptimal TSA and the leveraging of sparsity in the channel's delay-power profile. Although the noncoherent equalizer can function with only a single pilot subcarrier, a more extensive pilot pattern is proposed to track the (time-varying) sparsity profile as well as to reduce the complexity of near-optimal TSA. Simulations with highly spread channels showed that the performance of the proposed noncoherent algorithm was about 1.5 dB away from coherent detection using a genie-estimated channel.

Finally, we have characterized the maximum diversity-order that can be attained for both noncoherent detection and coherent detection of block transmissions under DS channels, and we have provided a set of sufficient conditions under which this maximum diversity-order can be attained. Specifically, we have shown that *almost any* affine (linear) precoder will facilitate maximum-diversity noncoherent (coherent) ML detection. In addition, we have shown that linear precoding does not facilitate maximum-diversity detection for certain commonly used symbol alphabets.

8.2 Possible Future Research

To further improve the fast coherent/noncoherent equalization algorithms, other tree searching algorithms (TSAs) could be considered. Although the biggest advantage of the M-algorithm is the complexity that is invariant to SNRs and channel

realizations, its performance is known to be inferior to other TSAs such as the adaptive T-algorithm in Chapter 4, Schnorr-Euchner (SE) sphere decoder [69] and Fano sequential decoder [70].

As shown in [70], TSAs can successfully be used as the joint coherent detection and decoding algorithms, where the channel structure and the code structure are both taken care of. In response, one could pursue a practical joint noncoherent receiver structure that jointly performs equalization, channel estimation and decoding together without performing three tasks in three different modules.

The max-diversity condition could be refined by using the fact that N_b is a function of N , i.e., $N_b = \phi(N)$, where $\phi(\cdot)$ is a monotonically increasing function. With proper identification and/or approximation of the function $\phi(\cdot)$, the noncoherent max-diversity condition $N \geq 2N_bN_h$ could provide more insightful relationship between the block length and the number of channel parameters.

We can see from Lemma 6 that the focus was taken on the diversity gain of the precoders, while no attention is paid on the coding gain. Thus, in an effort to get practical precoding schemes for the DS channel, structured affine precoding schemes should be pursued so that not only the maximum diversity, but also an excellent coding gain could be achieved. Toward this end, packing in complex Grassmannian space [122] or a few Grassmannian noncoherent code design methods in [123, 124] for singly selective channel could provide useful tools for the design of the precoder with good coding gain.

APPENDIX A

FOURIER BASIS REPRESENTATION OF MCM CHANNEL COEFFICIENTS

The transmitted sequence $\{t_n\}$ is generated by transforming the i^{th} MCM-symbol sequence $\{c_k^{(i)}\}_{k=0}^{N-1}$ with an N -point inverse discrete Fourier transform (DFT), applying an N_α -point modulation pulse $\{\alpha_n\}_{n=0}^{N_\alpha-1}$ to its cyclic extension, and superimposing the result N samples behind the contribution from the $(i-1)^{\text{th}}$ MCM-symbol:

$$t_n = \sum_{i=-\infty}^{\infty} \alpha_{n-iN} \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k^{(i)} e^{j\frac{2\pi}{N}kn}. \quad (\text{A.1})$$

(See [5] for a slightly more general scheme.) A noisy linear time-varying channel then produces the received samples

$$r_n = \sum_{l=0}^{N_h-1} h_{n,l} t_{n-l} + v_n, \quad (\text{A.2})$$

where $\{h_{n,l}\}_{l=0}^{N_h-1}$ denotes the length- N_h discrete impulse response at time n , and where $\{v_n\}$ is zero-mean circular white Gaussian noise (CWGN) with covariance σ^2 . In relation to the i^{th} MCM-symbol, we define $r_n^{(i)} \triangleq r_{iN+n}$, $v_n^{(i)} \triangleq v_{iN+n}$, and $h_{n,l}^{(i)} \triangleq h_{iN+n,l}$ and rewrite (A.2) as

$$r_n^{(i)} = \sum_{l=0}^{N_h-1} h_{n,l}^{(i)} \sum_{\ell=-\infty}^{\infty} \alpha_{\ell N+n-l} \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k^{(i-\ell)} e^{j\frac{2\pi}{N}k(n-l)} + v_n^{(i)}. \quad (\text{A.3})$$

For demodulation, the receiver applies the length- N_β pulse $\{\beta_n\}_{n=0}^{N_\beta-1}$ prior to an N -point DFT, yielding the frequency-domain observations $\{y_d^{(i)}\}_{d=0}^{N-1}$:

$$y_d^{(i)} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N_\beta-1} r_n^{(i)} \beta_n e^{-j\frac{2\pi}{N}dn}. \quad (\text{A.4})$$

Putting (A.1)-(A.4) together, it is straightforward to show that

$$y_d^{(i)} = \sum_{\ell=-\infty}^{\infty} \sum_{k=0}^{N-1} H_{d-k,k}^{(i,\ell)} c_k^{(i-\ell)} + w_d^{(i)}, \quad (\text{A.5})$$

where

$$H_{d,k}^{(i,\ell)} \triangleq \frac{1}{N} \sum_{n=0}^{N_\beta-1} \sum_{l=0}^{N_h-1} h_{n,l}^{(i)} \beta_n \alpha_{\ell N+n-l} e^{-j\frac{2\pi}{N}(dn+kl)} \quad (\text{A.6})$$

$$w_d^{(i)} \triangleq \frac{1}{\sqrt{N}} \sum_{n=0}^{N_\beta-1} \beta_n v_n^{(i)} e^{-j\frac{2\pi}{N}dn}. \quad (\text{A.7})$$

In writing (A.5), we used the fact that $H_{d,k}^{(i,\ell)}$ is N -cyclic in the indices d and k . Note that $H_{d,k}^{(i,\ell)}$ can be interpreted as the response, at MCM-symbol i and subcarrier $k+d$, to a frequency-domain impulse applied at MCM-symbol $i-\ell$ and subcarrier k . Using $\mathbf{y}^{(i)} \triangleq [y_0^{(i)}, \dots, y_{N-1}^{(i)}]^T$, $\mathbf{c}^{(i)} \triangleq [c_0^{(i)}, \dots, c_{N-1}^{(i)}]^T$, and $\mathbf{w}^{(i)} \triangleq [w_0^{(i)}, \dots, w_{N-1}^{(i)}]^T$, (A.5)

can be written in vector form as

$$\mathbf{y}^{(i)} = \sum_{\ell=-\infty}^{\infty} \sum_{k=0}^{N-1} \mathcal{D}_k(\mathbf{c}^{(i-\ell)}) \mathbf{H}_{-k}^{(i,\ell)} + \mathbf{w}^{(i)} \quad (\text{A.8})$$

$$\mathbf{H}_d^{(i,\ell)} \triangleq [H_{d,-d}^{(i,\ell)}, H_{d,-d+1}^{(i,\ell)}, \dots, H_{d,N-1-d}^{(i,\ell)}]^T \in \mathbb{C}^N, \quad (\text{A.9})$$

where $\mathcal{D}_k(\cdot)$ denotes the diagonal matrix created from the k -place cyclic downward shift of its vector argument, i.e., $[\mathcal{D}_k(\mathbf{c})]_{d,d} = c_{(d-k)_N}$.

Although from (A.1) it can be seen that no time-domain guard is employed by the transmitter, it is possible (see, e.g., [5]) to design pulses $\{\alpha_n\}$ and $\{\beta_n\}$ that yield both negligible ISI (i.e., $H_{d,k}^{(i,\ell)} \approx 0$ for $\ell \neq 0$) and negligible ICI beyond a radius of

$D \triangleq \lceil f_D T_c N \rceil$ subcarriers (i.e., $H_{d,k}^{(i,\ell)} \approx 0$ for $D < d < N - D$). Here, f_D denotes the single-sided Doppler spread in Hz and T_c denotes the sampling (or ‘‘chip’’) interval in seconds. With zero ISI and an ICI radius of D , (A.8) becomes

$$\mathbf{y}^{(i)} = \sum_{k=-D}^D \mathcal{D}_k(\mathbf{c}^{(i)}) \mathbf{H}_{-k}^{(i,0)} + \mathbf{w}^{(i)}. \quad (\text{A.10})$$

A.1 A Sparse Frequency-Domain BEM

From (A.6) and (A.9) it can be seen that

$$\mathbf{H}_d^{(i,0)} = \mathbf{F} \check{\boldsymbol{\theta}}_d^{(i)} \quad (\text{A.11})$$

$$\check{\boldsymbol{\theta}}_d^{(i)} \triangleq \sqrt{N} \mathcal{D}(\mathbf{f}_d^*) (\mathcal{H}^{(i)} \odot \mathcal{P})^T \underline{\mathbf{f}}_d \in \mathbb{C}^N, \quad (\text{A.12})$$

where $\mathcal{H}^{(i)}, \mathcal{P} \in \mathbb{C}^{N_\beta \times N}$ are defined element-wise as $[\mathcal{H}^{(i)}]_{n,l} \triangleq h_{n,l}^{(i)}$ and $[\mathcal{P}]_{n,l} \triangleq \beta_n \alpha_{n-l}$, where \mathbf{F} denotes the unitary N -DFT matrix (i.e., $[\mathbf{F}]_{n,m} = \frac{1}{\sqrt{N}} e^{-j \frac{2\pi}{N} nm}$) with d^{th} column $\mathbf{f}_d \in \mathbb{C}^N$, and where $\underline{\mathbf{f}}_d \triangleq [e^{-j \frac{2\pi}{N} d \cdot 0}, e^{-j \frac{2\pi}{N} d \cdot 1}, \dots, e^{-j \frac{2\pi}{N} d (N_\beta - 1)}]^T$ denotes the latter’s N_β -length (cyclic) extension. Equation (A.11) can be recognized as an N^{th} -order BEM for the frequency-domain channel vector $\mathbf{H}_d^{(i,0)}$; the columns of \mathbf{F} are the basis vectors and elements of $\check{\boldsymbol{\theta}}_d^{(i)}$ are the BEM coefficients.

The BEM order can be reduced if the impulse response is known to be sparse. In particular, if only $N_a < N_h$ taps of the impulse response $\{h_{n,l}^{(i)}\}_{l=0}^{N_h-1}$ are non-zero over the time duration $n \in \{0, \dots, N_\beta - 1\}$, then only N_a columns of $\mathcal{H}^{(i)}$ will be non-zero, implying that only N_a BEM coefficients in $\check{\boldsymbol{\theta}}_d^{(i)}$ will be non-zero. More precisely, let us denote the set of channel taps active during the i^{th} MCM-symbol interval by

$$\mathcal{L}^{(i)} = \left\{ l : h_{n,l}^{(i)} \neq 0 \text{ for some } n \in \{0, \dots, N_\beta - 1\} \right\}, \quad (\text{A.13})$$

where $|\mathcal{L}^{(i)}| = N_a$. Constructing $\mathbf{B}^{(i)} \in \mathbb{C}^{N \times N_a}$ from the columns of \mathbf{F} with indices in $\mathcal{L}^{(i)}$, and constructing $\boldsymbol{\theta}_d^{(i)} \in \mathbb{C}^{N_a}$ from the corresponding elements of $\check{\boldsymbol{\theta}}_d^{(i)}$, (A.11)

can be restated as

$$\mathbf{H}_d^{(i,0)} = \mathbf{B}^{(i)} \boldsymbol{\theta}_d^{(i)}. \quad (\text{A.14})$$

Using this sparse BEM, (A.10) can be rewritten as

$$\mathbf{y}^{(i)} = \left[\mathcal{D}_D(\mathbf{c}^{(i)}) \mathbf{B}^{(i)}, \dots, \mathcal{D}_{-D}(\mathbf{c}^{(i)}) \mathbf{B}^{(i)} \right] \boldsymbol{\theta}^{(i)} + \mathbf{w}^{(i)} \quad (\text{A.15})$$

$$\boldsymbol{\theta}^{(i)} \triangleq [\boldsymbol{\theta}_{-D}^{(i)t}, \dots, \boldsymbol{\theta}_D^{(i)t}]^T. \quad (\text{A.16})$$

Notice that the BEM can change with the MCM-symbol index.

A.2 Covariance of the BE Parameter

Finally, we investigate the crosscovariance matrix $\mathbf{R}_{\boldsymbol{\theta}_d, \boldsymbol{\theta}_k}^{(i)} \triangleq \text{E}\{\boldsymbol{\theta}_d^{(i)} \boldsymbol{\theta}_k^{(i)H}\}$ for later use in tree-search. We assume that the channel obeys the wide-sense stationary uncorrelated scattering (WSSUS) assumption over the duration of one MCM-symbol, i.e.,

$$\text{E}\{h_{n,l}^{(i)} h_{n-m,l'}^{(i)*}\} = \rho_m^{(i)} \sigma_l^{(i)2} \delta_{l-l'} \text{ for } n \in \{0, \dots, N_\beta - 1\}, \quad (\text{A.17})$$

where $\{\sigma_l^{(i)2}\}_{l=0}^{N_h-1}$ is the delay-power profile (DPP) and $\{\rho_m^{(i)}\}$ is the normalized (i.e., $\rho_0^{(i)} = 1$) tap autocorrelation sequence during the i^{th} MCM-symbol interval. Furthermore, we assume that $\{\rho_m^{(i)}\}$ is invariant to i and thus suppress the superscript notation. It can be shown straightforwardly that $\mathbf{R}_{\boldsymbol{\theta}_d, \boldsymbol{\theta}_k}^{(i)}$ is diagonal with elements

$$\left[\mathbf{R}_{\boldsymbol{\theta}_d, \boldsymbol{\theta}_k}^{(i)} \right]_{\nu, \nu} = \sigma_{l_\nu^{(i)}}^{(i)2} \sum_{m=-N_\beta+1}^{N_\beta-1} \rho_m \gamma_{d,k,m,\nu}^{(i)}, \quad (\text{A.18})$$

where $l_\nu^{(i)}$ denotes the index of the ν^{th} sparse tap (i.e., $\mathcal{L}^{(i)} = \{l_0^{(i)}, l_1^{(i)}, \dots, l_{N_a-1}^{(i)}\}$) and

$$\gamma_{d,k,m,\nu}^{(i)} = \frac{1}{N} \sum_{p,q:p-q=m} \beta_p \check{\alpha}_{d,p-l_\nu^{(i)}} \beta_q^* \check{\alpha}_{k,q-l_\nu^{(i)}}^* \quad (\text{A.19})$$

$$\check{\alpha}_{d,p} \triangleq \alpha_p e^{-j \frac{2\pi}{N} dp}. \quad (\text{A.20})$$

The values $\{[\mathbf{R}_{\boldsymbol{\theta}_d, \boldsymbol{\theta}_d}^{(i)}]_{\nu, \nu}\}_{d=0}^{N-1}$ can be interpreted as a sampling of the (pulse-shaped) Doppler spectrum of the ν^{th} sparse tap.

APPENDIX B

DERIVATIONS FOR NONCOHERENT EQUALIZATION ALGORITHMS

B.1 A Fast Recursive Update for $\mu(\mathbf{x}_n)$ (Table 5.1)

First we write (5.10) as

$$\mu(\mathbf{x}_n) = -\|\mathbf{y}_n - \mathbf{A}_n \bar{\boldsymbol{\theta}}\|_{\Phi_n^{-1}}^2 - \ln(\pi^{n+1} \det \Phi_n) + \mathbf{l}_n^T \mathbf{x}_n \quad (\text{B.1})$$

$$\Phi_n \triangleq \mathbf{A}_n \mathbf{R}_\theta \mathbf{A}_n^H + \sigma^2 \mathbf{I}_{n+1}, \quad (\text{B.2})$$

In the sequel, we use $\tilde{\mathbf{y}}_n \triangleq \mathbf{y}_n - \mathbf{A}_n \bar{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}_n \triangleq \hat{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}$. In the two sections below, we derive fast recursions for the first two terms in (B.1): $\mu_1(\mathbf{x}_n) \triangleq \tilde{\mathbf{y}}_n^H \Phi_n^{-1} \tilde{\mathbf{y}}_n$ and $\mu_2(\mathbf{x}_n) \triangleq \ln(\pi^{n+1} \det \Phi_n)$. Together, these recursions yield Table 5.1.

B.1.1 Recursion for $\mu_1(\mathbf{x}_n)$

Rewriting Φ_n with the aid of $\mathbf{A}_n = \begin{bmatrix} \mathbf{A}_{n-1} \\ \mathbf{a}_n^H \end{bmatrix}$, where \mathbf{a}_n^H denotes the n^{th} row of \mathbf{A} , we have

$$\Phi_n^{-1} = \begin{bmatrix} \Phi_{n-1} & \mathbf{A}_{n-1} \mathbf{R}_\theta \mathbf{a}_n \\ \mathbf{a}_n^H \mathbf{R}_\theta \mathbf{A}_{n-1}^H & \mathbf{a}_n^H \mathbf{R}_\theta \mathbf{a}_n + \sigma^2 \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{P}_n & \mathbf{p}_n \\ \mathbf{p}_n^H & p_n \end{bmatrix}, \quad (\text{B.3})$$

for the block-inverse quantities

$$\mathbf{P}_n \triangleq \Phi_{n-1}^{-1} + p_n^{-1} \mathbf{p}_n \mathbf{p}_n^H \quad (\text{B.4})$$

$$\mathbf{p}_n \triangleq -\Phi_{n-1}^{-1} \mathbf{A}_{n-1} \mathbf{R}_\theta \mathbf{a}_n p_n \quad (\text{B.5})$$

$$p_n^{-1} \triangleq \sigma^2 + \mathbf{a}_n^H \left(\mathbf{R}_\theta - \mathbf{R}_\theta \mathbf{A}_{n-1}^H \Phi_{n-1}^{-1} \mathbf{A}_{n-1} \mathbf{R}_\theta \right) \mathbf{a}_n. \quad (\text{B.6})$$

Writing $\mu_1(\mathbf{x}_n)$ using (B.3) and $\tilde{\mathbf{y}}_n = \begin{bmatrix} \tilde{y}_{n-1} \\ \tilde{y}_n \end{bmatrix}$, we get

$$\mu_1(\mathbf{x}_n) = \tilde{\mathbf{y}}_{n-1}^H \mathbf{P}_n \tilde{\mathbf{y}}_{n-1} + 2\Re\{\tilde{\mathbf{y}}_{n-1}^H \mathbf{p}_n \tilde{y}_n\} + p_n |\tilde{y}_n|^2. \quad (\text{B.7})$$

Using the \mathbf{s}_{n-1} -conditional MMSE estimate of $\tilde{\boldsymbol{\theta}}$ from $\tilde{\mathbf{y}}_{n-1}$:

$$\tilde{\boldsymbol{\theta}}_{n-1} = \mathbf{R}_\theta \mathbf{A}_{n-1}^H \Phi_{n-1}^{-1} \tilde{\mathbf{y}}_{n-1}, \quad (\text{B.8})$$

we see that $\mathbf{r}_{n-1}^H \mathbf{p}_n = -\tilde{\boldsymbol{\theta}}_{n-1}^H \mathbf{a}_n p_n$. Applying this relationship to (B.4)-(B.6), we can rewrite (B.7) as

$$\begin{aligned} \mu_1(\mathbf{x}_n) &= \tilde{\mathbf{y}}_{n-1}^H \Phi_{n-1}^{-1} \tilde{\mathbf{y}}_{n-1} + p_n \tilde{\boldsymbol{\theta}}_{n-1}^H \mathbf{a}_n \mathbf{a}_n^H \tilde{\boldsymbol{\theta}}_{n-1} \\ &\quad - 2p_n \Re\{\tilde{\boldsymbol{\theta}}_{n-1}^H \mathbf{a}_n \tilde{y}_n\} + p_n |\tilde{y}_n|^2 \end{aligned} \quad (\text{B.9})$$

$$= \mu_1(\mathbf{x}_{n-1}) + p_n |\tilde{y}_n - \mathbf{a}_n^H \tilde{\boldsymbol{\theta}}_{n-1}|^2. \quad (\text{B.10})$$

Now we concentrate on p_n . Defining $\boldsymbol{\Sigma}_{n-1}$ and applying the matrix inversion lemma (MIL):

$$\boldsymbol{\Sigma}_{n-1} \triangleq \mathbf{A}_{n-1}^H \mathbf{A}_{n-1} + \sigma^2 \mathbf{R}_\theta^{-1} \quad (\text{B.11})$$

$$\sigma^2 \boldsymbol{\Sigma}_{n-1}^{-1} = \mathbf{R}_\theta - \mathbf{R}_\theta \mathbf{A}_{n-1}^H \Phi_{n-1}^{-1} \mathbf{A}_{n-1} \mathbf{R}_\theta, \quad (\text{B.12})$$

we see from (B.6) that $p_n^{-1} = \sigma^2 (1 + \mathbf{a}_n^H \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{a}_n)$. Using the fact that $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_{n-1} + \mathbf{a}_n \mathbf{a}_n^H$, a second application of the MIL yields $\boldsymbol{\Sigma}_n^{-1} = \boldsymbol{\Sigma}_{n-1}^{-1} - \alpha_n \mathbf{d}_n \mathbf{d}_n^H$ for

$$\mathbf{d}_n \triangleq \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{a}_n \quad (\text{B.13})$$

$$\alpha_n \triangleq (1 + \mathbf{a}_n^H \mathbf{d}_n)^{-1} = p_n \sigma^2. \quad (\text{B.14})$$

Together, this gives a fast update for $p_n = \alpha_n/\sigma^2$.

Finally, we tackle $\tilde{\boldsymbol{\theta}}_n$. Using the MIL again,

$$\boldsymbol{\Phi}_n^{-1} = \sigma^{-2}(\mathbf{I}_{n+1} - \mathbf{A}_n \boldsymbol{\Sigma}_n^{-1} \mathbf{A}_n^H), \quad (\text{B.15})$$

which applied to (B.8) yields

$$\tilde{\boldsymbol{\theta}}_n = \frac{1}{\sigma^2} \mathbf{R}_\theta (\boldsymbol{\Sigma}_n - \mathbf{A}_n^H \mathbf{A}_n) \boldsymbol{\Sigma}_n^{-1} \mathbf{A}_n^H \tilde{\mathbf{y}}_n \quad (\text{B.16})$$

$$= \boldsymbol{\Sigma}_n^{-1} \mathbf{A}_n^H \tilde{\mathbf{y}}_n \quad (\text{B.17})$$

$$= (\boldsymbol{\Sigma}_{n-1}^{-1} - \alpha_n \mathbf{d}_n \mathbf{d}_n^H) (\mathbf{A}_{n-1}^H \tilde{\mathbf{y}}_{n-1} + \mathbf{a}_n \tilde{y}_n). \quad (\text{B.18})$$

Expanding (B.18) and applying $\mathbf{a}_n^H \boldsymbol{\Sigma}_n^{-1} \mathbf{a}_n = \alpha_n^{-1} - 1$, we get

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_n &= \tilde{\boldsymbol{\theta}}_{n-1} + \mathbf{d}_n \tilde{y}_n - \alpha_n \mathbf{d}_n \mathbf{a}_n^H \tilde{\boldsymbol{\theta}}_{n-1} - \alpha_n \mathbf{d}_n (\alpha_n^{-1} - 1) \tilde{y}_n \\ &= \tilde{\boldsymbol{\theta}}_{n-1} + \alpha_n (\tilde{y}_n - \mathbf{a}_n^H \tilde{\boldsymbol{\theta}}_{n-1}) \mathbf{d}_n. \end{aligned} \quad (\text{B.19})$$

Notice that, in (B.10) and (B.19), $\tilde{y}_n - \mathbf{a}_n^H \tilde{\boldsymbol{\theta}}_{n-1} = y_n - \mathbf{a}_n^H \hat{\boldsymbol{\theta}}_{n-1}$.

B.1.2 Recursion for $\mu_2(\mathbf{x}_n)$

From (B.3), we can write

$$\boldsymbol{\Phi}_n = \begin{bmatrix} \boldsymbol{\Phi}_{n-1} & \boldsymbol{\phi}_n \\ \boldsymbol{\phi}_n^H & \phi_n \end{bmatrix}, \quad (\text{B.20})$$

The Schur complement $\gamma_n \triangleq \phi_n - \boldsymbol{\phi}_n^H \boldsymbol{\Phi}_{n-1}^{-1} \boldsymbol{\phi}_n$ obeys [125]

$$\det(\boldsymbol{\Phi}_n) = \gamma_n \det(\boldsymbol{\Phi}_{n-1}). \quad (\text{B.21})$$

Identifying $\boldsymbol{\phi}_n$ and ϕ_n from (B.3),

$$\begin{aligned} \gamma_n &= \sigma^2 + \mathbf{a}_n^H \mathbf{R}_\theta \mathbf{a}_n - \mathbf{a}_n^H \mathbf{R}_\theta \mathbf{A}_{n-1}^H \boldsymbol{\Phi}_{n-1}^{-1} \mathbf{A}_{n-1} \mathbf{R}_\theta \mathbf{a}_n \\ &= \frac{\sigma^2}{\alpha_n} \end{aligned} \quad (\text{B.22})$$

using (B.12) and (B.14) for (B.22). Finally, taking the log of (B.21),

$$\mu_2(\mathbf{x}_n) = \mu_2(\mathbf{x}_{n-1}) + \ln \frac{\pi\sigma^2}{\alpha_n}. \quad (\text{B.23})$$

B.2 Derivation of (5.12)

The derivation is performed for full-block vectors rather than partial ones (e.g., \mathbf{x} rather than \mathbf{x}_n), but applies to both. Applying the MIL to Φ^{-1} , the first term of (5.9) becomes

$$\tilde{\mathbf{y}}^H \Phi^{-1} \tilde{\mathbf{y}} = \frac{1}{\sigma^2} (\tilde{\mathbf{y}}^H \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^H \mathbf{A} \Sigma^{-1} \mathbf{A}^H \tilde{\mathbf{y}}), \quad (\text{B.24})$$

where $\Sigma \triangleq \mathbf{A}^H \mathbf{A} + \sigma^2 \mathbf{R}_\theta^{-1} = \Sigma_{N-1}$ (via (B.11)). Writing

$$\tilde{\mathbf{y}}^H \mathbf{A} \Sigma^{-1} \mathbf{A}^H \tilde{\mathbf{y}} = 2\Re\{\tilde{\mathbf{y}}^H \mathbf{A} \Sigma^{-1} \mathbf{A}^H \tilde{\mathbf{y}}\} - \tilde{\mathbf{y}}^H \mathbf{A} \Sigma^{-1} \mathbf{A}^H \tilde{\mathbf{y}}$$

and plugging in $\tilde{\boldsymbol{\theta}} \triangleq \Sigma^{-1} \mathbf{A} \tilde{\mathbf{y}} = \tilde{\boldsymbol{\theta}}_{N-1}$ (via (B.17)), we find

$$\tilde{\mathbf{y}}^H \Phi^{-1} \tilde{\mathbf{y}} = \frac{1}{\sigma^2} (\tilde{\mathbf{y}}^H \tilde{\mathbf{y}} - 2\Re\{\tilde{\mathbf{y}}^H \mathbf{A} \tilde{\boldsymbol{\theta}}\} + \tilde{\boldsymbol{\theta}}^H \Sigma \tilde{\boldsymbol{\theta}}) \quad (\text{B.25})$$

$$= \frac{1}{\sigma^2} \|\tilde{\mathbf{y}} - \mathbf{A} \tilde{\boldsymbol{\theta}}\|^2 + \|\tilde{\boldsymbol{\theta}}\|_{\mathbf{R}_\theta^{-1}}^2. \quad (\text{B.26})$$

The definitions of $\tilde{\mathbf{y}}$ and $\tilde{\boldsymbol{\theta}}$ from Appendix B.1 then yield (5.12).

APPENDIX C

PROOFS FOR CHAPTER 7

C.1 Proof of Lemma 6

Our strategy is to characterize the $[\mathbf{P}, \mathbf{t}]$ which cause $[\mathbf{C}_k, \mathbf{D}_{lk}]$ to be rank deficient, and show that these problematic $[\mathbf{P}, \mathbf{t}]$ are avoided w.p.1. In the sequel, we consider arbitrary k and arbitrary $l \neq k$, and we use the abbreviations $\mathbf{s} = \mathbf{s}_k$, $\boldsymbol{\delta} = \mathbf{s}_l - \mathbf{s}_k$, and $[\mathbf{C}, \mathbf{D}] = [\mathbf{C}_k, \mathbf{D}_{lk}]$.

Rank deficiency occurs when $\exists [\boldsymbol{\alpha}, \boldsymbol{\beta}] \neq \mathbf{0}$ such that $[\mathbf{C}, \mathbf{D}][\boldsymbol{\alpha}, \boldsymbol{\beta}] = \mathbf{0}_N$. We would like to rewrite $[\mathbf{C}, \mathbf{D}][\boldsymbol{\alpha}, \boldsymbol{\beta}]$ so that the role of $[\mathbf{P}, \mathbf{t}]$ is explicit. From the construction of \mathbf{C} , and from the partitions $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_0^T, \boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_{N_h-1}^T]^T$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_{N_h-1}^T]^T$

where $\alpha_\ell, \beta_\ell \in \mathbb{C}^{N_b}$, we rewrite $[\mathbf{C}, \mathbf{D}][\begin{smallmatrix} \alpha \\ \beta \end{smallmatrix}] = [\mathbf{F}, \mathbf{G}][\mathbf{d}]$ with

$$\mathbf{F} = \begin{bmatrix} \mathbf{b}_{N-1}^H \alpha_0 & \cdots & \mathbf{b}_{N-1}^H \alpha_{N_h-1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \mathbf{b}_{N_h-1}^H \alpha_0 & \cdots & \mathbf{b}_{N_h-1}^H \alpha_{N_h-1} \\ \vdots & \ddots & \ddots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \mathbf{b}_0^H \alpha_0 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{b}_{N-1}^H \beta_0 & \cdots & \mathbf{b}_{N-1}^H \beta_{N_h-1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \mathbf{b}_{N_h-1}^H \beta_0 & \cdots & \mathbf{b}_{N_h-1}^H \beta_{N_h-1} \\ \vdots & \ddots & \ddots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \mathbf{b}_0^H \beta_0 \end{bmatrix}.$$

for \mathbf{c} defined in (7.12) and $\mathbf{d} \triangleq \mathbf{P}\delta$. Here we used the fact that $\{d_n = 0\}_{n < 0}$ and $\{c_n = 0\}_{n < 0}$. Using \mathbf{p}_n^H to denote the row of \mathbf{P} such that $c_n = \mathbf{p}_n^H \mathbf{s}$, we can then write

$$\begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix} = \left[\begin{array}{c|ccc} \mathbf{s}^T & & 1 & \\ \dots & & \dots & \\ & \mathbf{s}^T & & 1 \\ \hline \delta^T & & 0 & \\ \dots & & \dots & \\ & \delta^T & & 0 \end{array} \right] \begin{bmatrix} \mathbf{p} \\ \mathbf{t} \end{bmatrix}$$

$$\mathbf{p} = [\mathbf{p}_{N-1}^H, \mathbf{p}_{N-2}^H, \dots, \mathbf{p}_0^H]^T$$

$$\mathbf{t} = [t_{N-1}, t_{N-2}, \dots, t_0]^T.$$

Putting these together, we have

$$[\mathbf{C}, \mathbf{D}] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = [\mathbf{H}, \mathbf{F}] \begin{bmatrix} \mathbf{p} \\ \mathbf{t} \end{bmatrix} \quad (\text{C.1})$$

with

$$\mathbf{H} = \begin{bmatrix} \mathbf{b}_{N-1}^H (\alpha_0 \mathbf{s}^T + \beta_0 \delta^T) & \cdots & \mathbf{b}_{N-1}^H (\alpha_{N_h-1} \mathbf{s}^T + \beta_{N_h-1} \delta^T) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \mathbf{b}_{N_h-1}^H (\alpha_0 \mathbf{s}^T + \beta_0 \delta^T) & \cdots & \mathbf{b}_{N_h-1}^H (\alpha_{N_h-1} \mathbf{s}^T + \beta_{N_h-1} \delta^T) \\ \vdots & \ddots & \ddots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \mathbf{b}_0^H (\alpha_0 \mathbf{s}^T + \beta_0 \delta^T) \end{bmatrix}$$

and with \mathbf{F} as defined earlier. Thus, $[\mathbf{C}, \mathbf{D}][\frac{\alpha}{\beta}] = \mathbf{0}_N$ becomes equivalent to $[\frac{\mathbf{p}}{\mathbf{t}}] \in \mathcal{N}([\mathbf{H}, \mathbf{F}])$.

Notice that, if $[\mathbf{H}, \mathbf{F}] \neq \mathbf{0}_{N \times N(N_s+1)}$, then $\mathcal{N}([\mathbf{H}, \mathbf{F}])$ is a strict subspace of $\mathbb{C}^{N(N_s+1)}$. In this case, our assumptions on the distribution of $[\frac{\mathbf{p}}{\mathbf{t}}]$ imply that the set $\mathcal{N}([\mathbf{H}, \mathbf{F}])$ has measure zero, so that $[\frac{\mathbf{p}}{\mathbf{t}}] \notin \mathcal{N}([\mathbf{H}, \mathbf{F}])$ w.p.1. Thus, we need to show that $[\mathbf{H}, \mathbf{F}] \neq \mathbf{0}$ for all \mathbf{s} , for all nonzero $\boldsymbol{\delta}$, and for all nonzero $[\frac{\alpha}{\beta}]$. To do this, we consider two cases.

Case 1) $\alpha \neq \mathbf{0}$: Here we show that $[\mathbf{H}, \mathbf{F}] \neq \mathbf{0}$ by showing that $\mathbf{F} \neq \mathbf{0}$. Since $\alpha \neq \mathbf{0}$, we know that $\alpha_\ell \neq 0$ for some ℓ . The assumption of full rank $\tilde{\mathbf{B}}$ then implies that $\tilde{\mathbf{B}}\alpha_\ell \neq \mathbf{0}$ for some ℓ , which ensures that $\mathbf{b}_n^H \alpha_\ell \neq 0$ for some $n \in \{N_h - 1, \dots, N - 1\}$. The latter condition implies $\mathbf{F} \neq \mathbf{0}$. Clearly, this occurs for any $\{\mathbf{s}, \boldsymbol{\delta}\}$.

Case 2) $\alpha = \mathbf{0}$: Here it is evident that $\beta \neq \mathbf{0}$, $\mathbf{F} = \mathbf{0}$, and

$$\mathbf{H} = \begin{bmatrix} \mathbf{b}_{N-1}^H \boldsymbol{\beta}_0 \boldsymbol{\delta}^T & \cdots & \mathbf{b}_{N-1}^H \boldsymbol{\beta}_{N_h-1} \boldsymbol{\delta}^T & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \mathbf{b}_{N_h-1}^H \boldsymbol{\beta}_0 \boldsymbol{\delta}^T & \cdots & \mathbf{b}_{N_h-1}^H \boldsymbol{\beta}_{N_h-1} \boldsymbol{\delta}^T \\ \vdots & \ddots & \ddots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \mathbf{b}_0^H \boldsymbol{\beta}_0 \boldsymbol{\delta}^T \end{bmatrix}$$

Thus, we need to show that there is no combination of \mathbf{s} , nonzero $\boldsymbol{\delta}$, and nonzero $\boldsymbol{\beta}$ that yields $\mathbf{H} = \mathbf{0}$. But, since $\boldsymbol{\delta} \neq \mathbf{0}$, the condition $\mathbf{H} = \mathbf{0}$ is equivalent to $\mathbf{G} = \mathbf{0}$. Now, since $\tilde{\mathbf{B}}$ is full rank and $\beta_\ell \neq \mathbf{0}$ for some ℓ , we know that $\mathbf{b}_n^H \beta_\ell \neq 0$ for some $n \in \{N_h - 1, \dots, N - 1\}$, which ensures that $\mathbf{G} \neq \mathbf{0}$. Clearly, this occurs for any \mathbf{s} and any nonzero $\boldsymbol{\delta}$.

C.2 Proof of Lemma 8

We follow the same procedure as the proof of Lemma 6 to arrive at

$$\begin{aligned} [\mathbf{C}, \mathbf{D}] \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} &= [\mathbf{H}_1, \mathbf{H}_2, \mathbf{F}] \begin{bmatrix} \mathbf{p}' \\ \mathbf{e} \\ \mathbf{t} \end{bmatrix} \\ \mathbf{p}' &= [\mathbf{p}_{N-1}^H, \dots, \mathbf{p}_{N_s}^H]^T \\ \mathbf{e} &= [\mathbf{e}_0^T, \dots, \mathbf{e}_{N_s-1}^T]^T, \end{aligned}$$

where we define \mathbf{e}_n^T as the n th row vector of \mathbf{I}_{N_s} . Also $\mathbf{H}_1 \in \mathbb{C}^{N \times N_p N_s}$ and $\mathbf{H}_2 \in \mathbb{C}^{N \times N_s^2}$ are defined such that $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2]$. The matrices \mathbf{C} , \mathbf{D} , \mathbf{H} and \mathbf{F} and the vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and \mathbf{p}_n^H are defined as in the proof of Lemma 6. It can be easily seen that the equation $[\mathbf{C}, \mathbf{D}][\boldsymbol{\beta}] = \mathbf{0}$ becomes

$$[\mathbf{H}_1, \mathbf{F}] \begin{bmatrix} \mathbf{p}' \\ \mathbf{t} \end{bmatrix} = -\mathbf{H}_2 \mathbf{e}. \quad (\text{C.2})$$

We can see that if $[\mathbf{H}_1, \mathbf{F}] \neq \mathbf{0}$, then the solution of $[\mathbf{p}'_{\mathbf{t}}]$ is a set of measure zero with respect to $\mathbb{C}^{N+N_p N_s}$. More specifically, under the condition of $[\mathbf{H}_1, \mathbf{F}] \neq \mathbf{0}$, when $\mathbf{H}_2 \mathbf{e} \in \mathcal{R}([\mathbf{H}_1, \mathbf{F}])$, the solution is uniquely determined, and when $\mathbf{H}_2 \mathbf{e} \notin \mathcal{R}([\mathbf{H}_1, \mathbf{F}])$, there is no solution. The condition “ $[\mathbf{H}_1, \mathbf{F}] = \mathbf{0}$ and $\mathbf{H}_2 \mathbf{e} = \mathbf{0}$ ” only yields the solution $[\mathbf{p}'_{\mathbf{t}}]$ being $\mathbb{C}^{N+N_p N_s}$. Thus, as long as $[\mathbf{H}_1, \mathbf{F}] \neq \mathbf{0}$, the solution of $[\mathbf{p}'_{\mathbf{t}}]$ is at most a point (i.e., a set of measure zero). This implies that with $[\mathbf{H}_1, \mathbf{F}] \neq \mathbf{0}$ and with the distribution of $[\mathbf{p}'_{\mathbf{t}}]$, the solution $[\mathbf{p}'_{\mathbf{t}}]$ occurs with probability zero.

Then, with full rank $\check{\mathbf{B}}$, when we consider the cases $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\alpha} \neq \mathbf{0}$ (as in the proof for Lemma 6), it can be seen that we cannot have $[\mathbf{H}_1, \mathbf{F}] = \mathbf{0}$ for any nonzero $[\check{\boldsymbol{\beta}}]$, for any \mathbf{s} and for any nonzero $\boldsymbol{\delta}$, implying that $[\mathbf{C}, \mathbf{D}]$ is full rank w.p.1 for all k and $l \neq k$.

BIBLIOGRAPHY

- [1] S. Ohno, “Maximum likelihood inter-carrier interference suppression for wireless OFDM with null subcarriers,” in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, 2005, pp. 849–852.
- [2] “Al Gore: wireless a bright spot in global economy,” FierceWireless, Apr. 2009, <http://www.fiercewireless.com/ctialive/story/al-gore-wireless-bright-spot-global-economy/2009-04-03>.
- [3] “3G: Data rate,” <http://en.wikipedia.org/wiki/3G>.
- [4] L. Zheng and D. N. C. Tse, “Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel,” *IEEE Trans. on Information Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
- [5] S. Das and P. Schniter, “Max-SINR ISI/ICI-shaped multi-carrier communication over the doubly dispersive channel,” *IEEE Trans. on Signal Processing*, vol. 55, no. 12, pp. 5782–5795, Dec. 2007.
- [6] J. A. C. Bingham, “Multicarrier modulation for data transmission: An idea whose time has come,” *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, May 1990.
- [7] L. J. Cimini, Jr., “Analysis and simulation of a digital mobile radio channel using orthogonal frequency division multiplexing,” *IEEE Trans. on Communications*, vol. 33, pp. 665–765, July 1985.
- [8] B. Muquet, Z. Wang, G. B. Giannakis, M. De Courville, and P. Duhamel, “Cyclic-prefixed or zero-padded multicarrier transmissions?” *IEEE Trans. on Communications*, vol. 50, no. 12, Dec. 2002.
- [9] T. Strohmer and S. Beaver, “Optimal OFDM design for time-frequency dispersive channels,” *IEEE Trans. on Communications*, vol. 51, no. 7, pp. 1111–1122, July 2003.

- [10] W. G. Jeon, K. H. Chang, and Y. S. Cho, "An equalization technique for orthogonal frequency-division multiplexing systems in time-variant multipath channels," *IEEE Trans. on Communications*, vol. 47, no. 1, pp. 27–32, Jan. 1999.
- [11] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Trans. on Signal Processing*, vol. 50, no. 10, pp. 2451–2464, Oct. 2002.
- [12] I. Barhumi, G. Leus, and M. Moonen, "Time-domain and frequency-domain per-tone equalization for OFDM in doubly-selective channels," *Signal Processing*, vol. 84, pp. 2055–2066, Nov. 2004.
- [13] Y.-S. Choi, P. J. Voltz, and F. A. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Trans. on Communications*, vol. 49, no. 8, pp. 1375–1387, Aug. 2001.
- [14] X. Cai and G. B. Giannakis, "Bounding performance and suppressing inter-carrier interference in wireless mobile OFDM," *IEEE Trans. on Communications*, vol. 51, no. 12, pp. 2047–2056, Dec. 2003.
- [15] L. Rugini, P. Banelli, and G. Leus, "Block DFE and windowing for Doppler-affected OFDM systems," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communication*, 2005.
- [16] P. Schniter, "Low-complexity equalization of OFDM in doubly-selective channels," *IEEE Trans. on Signal Processing*, vol. 52, no. 4, pp. 1002–1011, Apr. 2004.
- [17] S. Das and P. Schniter, "A new pulse shaped frequency division multiplexing technique for doubly dispersive channels," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, 2004.
- [18] K. Matheus and K.-D. Kammeyer, "Optimal design of a multicarrier systems with soft impulse shaping including equalization in time or frequency direction," in *IEEE Globecom*, vol. 1, 1997, pp. 310–314.
- [19] K. Matheus, K. Knoche, M. Feuersänger, and K.-D. Kammeyer, "Two-dimensional (recursive) channel equalization for multicarrier systems with soft impulse shaping (MCSIS)," in *Proc. IEEE Global Telecommunications Conf.*, vol. 2, 1998, pp. 956–961.
- [20] K. Sigloch, M. R. Andrews, P. P. Mitra, and D. J. Thomson, "Communicating over nonstationary nonflat wireless channels," *IEEE Trans. on Signal Processing*, vol. 53, no. 6, pp. 2216–2227, 2005.

- [21] R. A. Iltis, J. J. Shynk, and K. Giridhar, "Bayesian algorithms for blind equalization using parallel adaptive filtering," *IEEE Trans. on Communications*, vol. 42, pp. 1017–1032, Feb./Mar./Apr. 1994.
- [22] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Information Theory*, vol. 20, pp. 284–287, Mar. 1974.
- [23] M. J. Gertsman and J. H. Lodge, "Symbol-by-symbol MAP demodulation of CPM and PSK signals on Rayleigh flat-fading channels," *IEEE Trans. on Communications*, vol. 45, pp. 788–789, Jul. 1997.
- [24] L. M. Davis, I. B. Collings, and P. Hoeher, "Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels," *IEEE Trans. on Communications*, vol. 49, no. 12, pp. 2106–2114, Dec. 2001.
- [25] B. D. Hart and S. Pasupathy, "Innovations-based map detection for time-varying frequency-selective channels," *IEEE Trans. on Communications*, vol. 48, pp. 1507–1519, Sept 2000.
- [26] A. Anostasopoulos and K. M. Chugg, "Iterative detection for channels with memory," *IEEE Trans. on Communications*, vol. 48, pp. 1638–1649, Oct. 2000.
- [27] Y. Zhang, M. P. Fitz, and S. B. Gelfand, "Soft output demodulation on frequency-selective Rayleigh fading channels using AR channel models," in *Proc. IEEE Global Telecommunications Conf.*, Oct. 1997.
- [28] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.
- [29] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. on Communications*, vol. 42, no. 7, pp. 2406–2413, Jul. 1994.
- [30] C. Antón-Haro, J. A. R. Fonollosa, and J. R. Fonollosa, "Blind channel estimation and data detection using hidden Markov models," *IEEE Trans. on Signal Processing*, vol. 45, no. 1, pp. 241–247, Jan. 1997.
- [31] M. Nissilä and S. Pasupathy, "Adaptive Bayesian and EM-based detectors for frequency-selective fading channels," *IEEE Trans. on Communications*, vol. 51, no. 8, pp. 1325–1336, Aug. 2003.
- [32] H. Nguyen and B. C. Levy, "Blind equalization of dispersive fast fading Ricean channels via the EMV algorithm," *IEEE Trans. on Vehicular Technology*, vol. 54, pp. 1793–1801, Sep. 2005.

- [33] M. Tüchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. on Communications*, vol. 50, no. 5, pp. 754–767, May 2002.
- [34] P. Schniter, "A low-complexity receiver for OFDM in doubly-selective channels," in *Proc. IEEE Global Telecommunications Conf.*, 2003, (to appear).
- [35] F. Peng and W. E. Ryan, "A low-complexity soft demapper for OFDM fading channels with ICI," in *Proc. IEEE Wireless Communication and Networking Conf.*, 2006, pp. 1549–1554.
- [36] D. Liu and M. P. Fitz, "Reduced state iterative MAP equalization and decoding in wireless mobile coded OFDM," in *Proc. Allerton Conf. on Communication, Control, and Computing*, Sep. 2007.
- [37] S. Song, A. C. Singer, and K.-M. Sung, "Soft input channel estimation for turbo equalization," *IEEE Trans. on Signal Processing*, vol. 52, no. 10, pp. 2885–2894, Oct. 2004.
- [38] R. Otnes and M. Tüchler, "Iterative channel estimation for turbo equalization of time-varying frequency-selective channels," *IEEE Trans. on Wireless Communications*, vol. 3, no. 6, pp. 1918–1923, Nov. 2004.
- [39] D. Liu and M. P. Fitz, "Joint turbo channel estimation and data recovery in fast fading mobile coded OFDM," in *Proc. IEEE Internat. Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2008.
- [40] K. Fang, L. Rugini, and G. Leus, "Iterative channel estimation and turbo equalization for time-varying OFDM systems," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, 2008.
- [41] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. on Communications*, vol. 46, pp. 931–939, July 1996.
- [42] X. Ma and G. B. Giannakis, "Maximum-diversity transmissions over doubly selective wireless channels," *IEEE Trans. on Information Theory*, vol. 49, no. 7, pp. 1832–1840, July 2003.
- [43] R. A. Iltis, "A Bayesian maximum-likelihood sequence estimation algorithm for a priori unknown channels and symbol timing," *IEEE Journal on Selected Areas In Communications*, vol. 10, pp. 579–588, Apr. 1992.
- [44] Q. Dai and E. Shwedyk, "Detection of bandlimited signals over frequency selective Raleigh fading channels," *IEEE Trans. on Communications*, vol. 42, no. 2-4, pp. 941–950, Feb.-Apr. 1994.

- [45] H. Kubo, M. Murakami, and T. Fujino, "An adaptive maximum-likelihood sequence estimator for fast time-varying intersymbol interference channels," *IEEE Trans. on Communications*, vol. 42, pp. 1872–1880, Feb./Mar./Apr. 1994.
- [46] 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.
- [47] R. Raheli, A. Polydoros, and C. K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. on Communications*, vol. 43, pp. 354–364, Feb.-Apr. 1995.
- [48] D. K. Borah and B. D. Hart, "Receiver structures for time-varying frequency-selective fading channels," *IEEE Journal on Selected Areas In Communications*, vol. 17, pp. 1863–1875, Nov. 1999.
- [49] B. D. Hart, "Maximum likelihood sequence detection using a pilot tone," *IEEE Trans. on Vehicular Technology*, vol. 49, no. 2, pp. 550–560, March 2000.
- [50] H. Chen, K. Buckley, and R. Perry, "Time-recursive maximum likelihood based sequence estimation for unknown ISI channels," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, 2000, pp. 1005–1009.
- [51] H. Chen, R. Perry, and K. Buckley, "On MLSE algorithms for unknown fast time-varying channels," *IEEE Trans. on Communications*, vol. 51, no. 5, pp. 730–734, May 2003.
- [52] A. E.-S. El-Mahdy, "Adaptive channel estimation and equalization for rapidly mobile communication channels," *IEEE Trans. on Communications*, vol. 52, no. 7, pp. 1126–1135, July 2004.
- [53] S.-J. Hwang and P. Schniter, "Near-optimal noncoherent sequence detection for doubly dispersive channels," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Oct. 2006.
- [54] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective wireless fading channels," *IEEE Trans. on Signal Processing*, vol. 51, no. 5, pp. 1351–1366, May 2003.
- [55] A. P. Kannu and P. Schniter, "MSE-optimal training for linear time-varying channels," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, 2005.
- [56] —, "Capacity analysis of MMSE pilot patterns for doubly selective channels," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communication*, 2005.

- [57] J. K. Tugnait, X. Meng, and S. He, “Doubly selective channel estimation using superimposed training and exponential bases models,” pp. Article ID 85 303, 11 pages, 2006.
- [58] D. K. Borah and B. D. Hart, “Frequency-selective fading channel estimation with a polynomial time-varying channel model,” *IEEE Trans. on Communications*, vol. 47, no. 6, pp. 862–873, June 1999.
- [59] A. Gorokhov and J.-P. Linnartz, “Robust OFDM receivers for dispersive time-varying channels: Equalization and channel acquisition,” *IEEE Trans. on Communications*, vol. 52, no. 4, pp. 572–583, Apr. 2004.
- [60] T. A. Thomas and F. W. Vook, “Multi-user frequency-domain channel identification, interference suppression, and equalization for timevarying broadband wireless communications,” in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Boston, MA, Mar. 2000, pp. 444–448.
- [61] G. Leus, “On the estimation of rapidly time-varying channels,” in *Proc. European Signal Processing Conf.*, Sep. 2004.
- [62] T. Zemen and C. F. Mecklenbräuker, “Time-variant channel estimation using discrete prolate spheroidal sequences,” *IEEE Trans. on Signal Processing*, vol. 53, pp. 3597–3607, Sep. 2005.
- [63] K. D. Teo and S. Ohno, “Optimal mmse finite parameter model for doubly-selective channels,” in *Proc. IEEE Global Telecommunications Conf.*, St. Louis, MO, Nov. 2005.
- [64] W. H. Mow, “Maximum likelihood sequence estimation from the lattice viewpoint,” *IEEE Trans. on Information Theory*, vol. 40, no. 5, pp. 1591–1600, 1994.
- [65] A. J. Viterbi and J. Omura, *Principles of Digital Communication and Decoding*. New York: McGraw-Hill, 1979.
- [66] J. B. Anderson and S. Mohan, “Sequential decoding algorithms: A survey and cost analysis,” *IEEE Trans. on Communications*, vol. 32, pp. 169–172, 1984.
- [67] E. Viterbo and J. Boutros, “A universal lattice code decoder for fading channels,” *IEEE Trans. on Information Theory*, vol. 45, no. 5, pp. 1639–1642, July 1999.
- [68] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, “Closest point search in lattices,” *IEEE Trans. on Information Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.

- [69] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.
- [70] A. Murugan, H. El Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding: Rediscovering the sequential decoder," *IEEE Trans. on Information Theory*, vol. 52, pp. 933–953, Mar. 2006.
- [71] M. O. Damen, H. El Gamal, and G. Caire, "MMSE-GDFE lattice decoding for under-determined linear channels," in *Proc. Conf. on Information Science and Systems*, 2004.
- [72] N. Al-Dhahir and A. H. Sayed, "The finite-length MIMO MMSE-DFE," *IEEE Trans. on Signal Processing*, vol. 48, no. 10, pp. 2921–2936, Oct. 2000.
- [73] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of mimo channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 428–436, June 2004.
- [74] S. J. Simmons, "Breadth-first trellis decoding with adaptive effort," *IEEE Trans. on Communications*, vol. 38, no. 1, pp. 3–12, Jan. 1990.
- [75] R. Gowaikar and B. Hassibi, "Efficient statistical pruning algorithms for maximum likelihood decoding," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, 2003, pp. 49–52.
- [76] J. M. Wozencraft and B. Reiffen, *Sequential Decoding*. New York: MIT Press and Wiley, 1961.
- [77] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice," *Math. of Comput.*, vol. 44, pp. 463–471, Apr. 1985.
- [78] R. M. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Trans. on Information Theory*, vol. 9, no. 2, pp. 64–74, Apr. 1963.
- [79] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive t-algorithm in mlsd/mlsde receivers for fading channels," in *Proc. IEEE Intern. Conf. on Communication*, vol. 1, June 1999, pp. 539–543.
- [80] M. Kokkonen and K. Kalliojarvi, "Soft-decision decoding of binary linear block codes using reduced breadth-first search algorithms," *IEEE Trans. on Communications*, vol. 48, no. 6, pp. 905–905, June 2000.
- [81] G. J. Foschini, G. Golden, R. Valanzuela, and P. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE Journal on Selected Areas In Communications*, vol. 17, pp. 1841–1852, Nov. 1999.

- [82] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: John Hopkins University Press, 1996.
- [83] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.
- [84] P. Schniter, “On doubly dispersive channel estimation for pilot-aided pulse-shaped multi-carrier modulation,” in *Proc. Conf. on Information Science and Systems*, Princeton, NJ, Mar. 2006.
- [85] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.
- [86] M. K. Tsatsanis and G. B. Giannakis, “Modeling and equalization of rapidly fading channels,” *Internat. Journal of Adaptive Control & Signal Processing*, vol. 10, pp. 159–176, Mar. 1996.
- [87] B. M. Hochwald and S. ten Brink, “Achieving near-capacity on a multiple-antenna channel,” *IEEE Trans. on Communications*, vol. 51, no. 3, pp. 389–399, March 2003.
- [88] Y. L. C. de Jong and T. J. Willink, “Iterative tree search detection for MIMO wireless systems,” *IEEE Trans. on Communications*, vol. 53, pp. 930–935, June 2005.
- [89] S.-J. Hwang and P. Schniter, “Fast noncoherent decoding of block transmissions over doubly dispersive channels,” in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov. 2007.
- [90] ———, “EM-based soft noncoherent equalization of doubly selective channels using tree search and basis expansion,” in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communication*, Perugia, Italy, June 2009.
- [91] T. Cui and C. Tellambura, “Blind receiver design for OFDM systems over doubly selective channels,” *IEEE Trans. on Communications*, vol. 55, pp. 906–917, May 2007.
- [92] A. P. Kannu and P. Schniter, “Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels,” *TSP*, vol. 56, pp. 1148–1160, 2008.
- [93] B. D. Hart and D. P. Taylor, “Extended MLSD diversity receiver for the time and frequency selective channel,” *IEEE Trans. on Communications*, vol. 45, pp. 322–333, Mar. 1997.

- [94] A. P. Kannu and P. Schniter, "On the spectral efficiency of noncoherent doubly selective channels," in *Proc. Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Sept. 2006.
- [95] W. C. Jakes, *Microwave Mobile Communications*. Wiley, 1974.
- [96] I. Kozintsev, "Matlab programmes for encoding and decoding of LDPC codes over $GF(2^m)$," downloadable at <http://www.kozintsev.net/soft.html>.
- [97] J. D. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electronics Letters*, vol. 33, no. 6, pp. 457–458, Mar. 1997.
- [98] S.-J. Hwang and P. Schniter, "Efficient multicarrier communication for highly spread underwater acoustic channels," *IEEE Journal on Selected Areas In Communications*, vol. 26, no. 9, pp. 1674–1683, Dec. 2008.
- [99] D. Brady and J. C. Preisig, "Underwater acoustic communications," in *Wireless Communications: Signal Processing Perspectives*, H. V. Poor and G. W. Wornell, Eds. Prentice-Hall, 1998, ch. 8, pp. 330–379.
- [100] M. Stojanovic, J. A. Catipovic, and J. G. Proakis, "Phase-coherent digital communications for underwater acoustic channels," *IEEE Journal of Oceanic Engineering*, vol. 19, pp. 100–111, Jan. 1994.
- [101] J. C. Preisig, "Performance analysis of adaptive equalization for coherent acoustic communications in the time-varying ocean environment," *Journal of the Acoustical Society of America*, vol. 118, pp. 263–278, Jul. 2005.
- [102] J. C. Preisig and G. Deane, "Surface wave focusing and acoustic communications in the surf zone," *Journal of the Acoustical Society of America*, vol. 116, pp. 2067–2080, Oct. 2005.
- [103] B. Li, S. Zhou, M. Stojanovic, and L. Freitag, "Pilot-tone based ZP-OFDM demodulation for an underwater acoustic channel," Sep. 2006.
- [104] M. Stojanovic, "Low complexity OFDM detector for underwater acoustic channels," Sep. 2006.
- [105] A. K. Morozov and J. C. Preisig, "Underwater acoustic communications with multi-carrier modulation," Sep. 2006.
- [106] Y. Li and L. J. Cimini, Jr., "Bounds on the interchannel interference of OFDM in time-varying impairments," *IEEE Trans. on Communications*, vol. 49, no. 3, pp. 401–404, Mar. 2001.

- [107] B. Le Floch, M. Alard, and C. Berrou, "Coded orthogonal frequency division multiplex," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 982–996, June 1995.
- [108] W. Kozek and A. F. Molisch, "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE Journal on Selected Areas In Communications*, vol. 16, no. 8, pp. 1579–1589, Oct. 1998.
- [109] K. Liu, T. Kadous, and A. M. Sayeed, "Orthogonal time-frequency signaling over doubly dispersive channels," *IEEE Trans. on Information Theory*, vol. 50, no. 11, pp. 2583–2603, Nov. 2004.
- [110] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 12–25, Nov. 2004.
- [111] A. P. Kannu and P. Schniter, "On the spectral efficiency of noncoherent doubly selective block-fading channels," *TIT*, submitted.
- [112] T. H. Eggen, A. B. Baggeroer, and J. C. Preisig, "Communication over Doppler spread channels—Part i: Channel and receiver presentation," *IEEE Journal of Oceanic Engineering*, vol. 25, pp. 62–71, Jan. 2000.
- [113] W. Li and J. C. Preisig, "Estimation and equalization of rapidly varying sparse acoustic communication channels," Sep. 2006.
- [114] S.-J. Hwang and P. Schniter, "Efficient communication over highly spread underwater acoustic channels," in *The Second ACM International Workshop on Underwater Networks (WUWNet)*, Montreal, Quebec, Sept. 2007.
- [115] Z. Tang, R. C. Cannizzaro, G. Leus, and P. Banelli, "Pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Trans. on Signal Processing*, vol. 55, pp. 2226–2238, May 2007.
- [116] A. M. Sayeed and B. Aazhang, "Joint multipath-doppler diversity in mobile wireless communications," *IEEE Trans. on Communications*, vol. 47, no. 1, pp. 123–132, Jan. 1999.
- [117] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. on Information Theory*, vol. 47, no. 6, pp. 2383–2399, Sep 2001.
- [118] S. Siwamogsatham, M. Fitz, and J. 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.

- [119] A. Scaglione, G. B. Giannakis, and S. Barbarossa, “Redundant filterbank precoders and equalizers Part I: Unification and optimal designs,” *IEEE Trans. on Signal Processing*, vol. 47, no. 7, pp. 1988–2006, July 1999.
- [120] —, “Redundant filterbank precoders and equalizers Part II: Blind channel estimation, synchronization, and direct equalization,” *IEEE Trans. on Signal Processing*, vol. 47, no. 7, pp. 2007–2022, July 1999.
- [121] J. H. Manton, I. Y. Mareels, and Y. Hua, “Affine precoders for reliable communications,” in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, June 2000, pp. 2749–2752.
- [122] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, “Packing lines, planes, etc.: packing in Grassmannian spaces,” *Experiment. Math.*, vol. 5, pp. 139–159, Sept. 1996.
- [123] 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.
- [124] I. Kammoun and J.-C. Belfiore, “A new family of Grassmann space-time codes for non-coherent MIMO systems,” *IEEE Communications Letters*, vol. 7, no. 11, pp. 528–530, Nov. 2003.
- [125] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins University Press, 1983.