CHANNEL ESTIMATION AND EQUALIZATION FOR DOUBLY-SELECTIVE CHANNELS USING BASIS EXPANSION MODELS

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

_____

Liying Song

Certificate of Approval:

_____	_____
Stanley J. Reeves Professor Electrical and Computer Engineering	Jitendra K. Tugnait, Chair James B. Davis Professor Electrical and Computer Engineering
_____	_____
Shiwen Mao Assistant Professor Electrical and Computer Engineering	Joe F. Pittman Interim Dean Graduate School

Channel Estimation And Equalization For Doubly-Selective Channels Using Basis Expansion Models

Liying Song

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
May 10, 2008

CHANNEL ESTIMATION AND EQUALIZATION FOR DOUBLY-SELECTIVE CHANNELS USING BASIS EXPANSION MODELS

Liying Song

Permission is granted to Auburn University to make copies of this dissertation at its discretion, upon the request of individuals or institutions and at their expense. The author reserves all publication rights.

_____
Signature of Author

_____
Date of Graduation

VITA

Liying Song, daughter of Zhoutian Song and Shuqin Qian, was born in Daqing, Heilongjiang Province, China, on May 25, 1980. She received her B. E. and M. S. degree in Electrical Engineering from University of Electronics Science and Technology of China in 2001 and 2004, respectively. Since August 2004 she has been a Ph. D student in Electrical and Computer Engineering Department of Auburn University.

Her research interests are in wireless communications and signal processing, including channel estimation and equalization for time- and frequency-selective channels, OFDM for wireless communications, and MIMO technique.

DISSERTATION ABSTRACT

CHANNEL ESTIMATION AND EQUALIZATION FOR DOUBLY-SELECTIVE CHANNELS USING BASIS EXPANSION MODELS

Liying Song

Doctor of Philosophy, May 10, 2008
(M. S., Univ. of Electronics Sci. and Tech. of China, 2001)
(B. E., Univ. of Electronics Sci. and Tech. of China, 2004)

166 Typed Pages

Directed by Jitendra K. Tugnait

The nature of the wireless channels places fundamental limitations on the performance of wireless communication systems. In addition to the frequency-selectivity characteristics caused by multipath propagation, the high-rate wireless and mobile links often exhibit time-selectivity characteristics caused by the user's mobility, so-called doubly-selective wireless channels. The quality of channel acquisition has a major impact on the overall system performance. Therefore, reliable estimation of doubly-selective channels is well motivated. Equalization is used at the receiver to compensate for intersymbol interference created by multipath propagation and improve received signal quality. Equalizers should be adaptive since the channel is time-varying.

In this dissertation, channel estimation and equalization for doubly-selective channels are considered in Chapter 2 (under single input single output models) and Chapter 3 (under multiple input multiple output models), where the time-varying channel is assumed to be well described by basis expansion models (BEM). Our focus is on time-multiplexed training for channel estimation where the training symbols are periodically inserted and use all transmitted power during their transmission.

The linear equalization and decision feedback equalization (DFE) of doubly-selective channels modeled via BEMs are introduced in Chapter 4. There has been much interest in designing time-variant serial finite impulse response (FIR) linear and DFE equalizers using complex exponential (CE-) BEMs for equalizers in addition to using CE-BEM for modeling the channel itself. In this dissertation we show that the Kalman filter formulation of the linear equalizer and an alternative formulation of the FIR DFE based on a CE-BEM channel model yields the same or an improved BER at a lower computational cost, without incurring the approximation error inherent in CE-BEM modeling of equalizers.

In Chapter 5, an adaptive channel estimation scheme, exploiting the oversampled complex exponential basis expansion model (CE-BEM), is presented for doubly-selective channels where we track the BEM coefficients via a multiple model approach in this dissertation. We propose to use a multiple model framework where several candidate Doppler spread values are used to cover the range from zero to an upper bound, which leads to multiple CE-BEM channel models, each corresponding to an assumed value of the Doppler spread. Subsequently, the well known interacting multiple model (IMM) algorithm is used for symbol detection based on multiple state-space models corresponding to the multiple estimated channels.

Orthogonal Frequency-Division Multiplexing (OFDM), a digital multi-carrier modulation scheme, has developed into a popular scheme for wideband wireless communication due to its high spectral efficiency and simple equalization. We extend the optimum time-multiplexed training based channel estimation introduced in Chapter 2 to OFDM systems under doubly-selective channels in Chapter 6. Compared to the traditional frequency-domain training design, the main advantages of time-domain training for OFDM system is that the information symbols are not contaminated by the training symbols as in the frequency-domain training case.

ACKNOWLEDGMENTS

It has been a great pleasure working with the faculty, staff, and students at the Electrical and Computer Engineering Department, Auburn University, during my tenure as a doctoral student. Completing this work is definitely a high point in my academic career. I could not have come this far without the assistance of many individuals and I want to express my deepest appreciation to them.

My first and most earnest thanks go to my advisor, Dr. Jitendra K. Tugnait for his support, guidance, and kindness. In every sense, none of this work would have been possible without him.

I would like to sincerely thank my committee members Dr. Stanley J. Reeves and Dr. Shiwen Mao who have given their time to read this manuscript and offered valuable advice during my graduate studies. Many thanks to Dr. Douglas A. Leonard who served as my outside reader for providing valuable comments that improved the contents of this dissertation.

I gratefully acknowledge Dr. Xiaoli Ma who gave me the chance to start my PhD journey at Auburn University. I also wish to thank my colleagues and friends whose encouragement over the years has been invaluable.

My final, and most heartfelt, acknowledgment must go to my parents. They have raised me and have been giving their love and support all the time.

I would end this list of acknowledgments not with a ”thank you” but with a warm memory.

Style manual or journal used Journal of Approximation Theory (together with the style known as “aums”). Bibliography follows van Leunen’s *A Handbook for Scholars*.

Computer software used The document preparation package TEX (specifically LATEX) together with the departmental style-file `aums.sty`.

TABLE OF CONTENTS

LIST OF FIGURES xii

LIST OF TABLES xv

1 INTRODUCTION 1
1.1 Characteristics and Representations of Wireless Communication Channels 1
1.1.1 Jakes' Model 3
1.1.2 Complex Exponential Basis Expansion Model (CE-BEM) 5
1.1.3 Discrete Prolate Spheroidal Basis Expansion Model (DPS-BEM) 7
1.1.4 Modeling Error Comparison between CE-BEM and DPS-BEM 9
1.2 Channel Estimation and Equalization Approaches 11
1.2.1 Channel Estimation Approaches 11
1.2.2 Equalization Approaches 13
1.3 Optimal Maximum Likelihood Detector (Viterbi Decoder) 14
1.4 Orthogonal Frequency Division Multiplexing (OFDM) 17
1.5 Outline and Contribution 19

2 SISO DOUBLY-SELECTIVE CHANNEL ESTIMATION USING DISCRETE PROLATE SPHEROIDAL BASIS EXPANSION MODELS AND TIME-MULTIPLEXED TRAINING 22
2.1 Introduction 22
2.2 System Model 24
2.3 Channel Estimation 27
2.3.1 Least-squares Channel Estimation 29
2.3.2 Linear Minimum Mean-Square-Error Channel Estimation 30
2.3.3 Channel Estimation Error 32
2.4 Optimum Training Design 33
2.5 Training Power Allocation 36
2.6 Numerical Examples 40
2.6.1 Example 1: Approximation Errors 41
2.6.2 Example 2: DPS-BEM Channel Estimation Performance 44
2.6.3 Example 3: CE-BEM versus DPS-BEM 46
2.6.4 Example 4. Training Power Allocation 47
2.7 Conclusion 48

3 MIMO DOUBLY-SELECTIVE CHANNEL ESTIMATION USING DISCRETE PROLATE SPHEROIDAL BASIS EXPANSION MODELS AND TIME-MULTIPLEXED TRAINING 50
3.1 Introduction . . . 50
3.2 Multiuser MIMO Channels . . . 51
3.3 Channel Estimation . . . 53
3.3.1 Linear Least-Squares Channel Estimator . . . 55
3.3.2 Linear MMSE Channel Estimator . . . 56
3.3.3 Channel Estimation Error . . . 58
3.4 Optimum Training Design . . . 59
3.5 Training Power Allocation . . . 62
3.6 Numerical Examples . . . 67
3.6.1 Example 1: Approximation Errors . . . 70
3.6.2 Example 2: Channel Estimation Performance . . . 71
3.6.3 Example 3: CE-BEM versus DPS-BEM . . . 74
3.7 Conclusion . . . 75

4 TIME-VARYING EQUALIZATION FOR DOUBLY-SELECTIVE CHANNELS 76
4.1 Introduction . . . 76
4.2 Linear Equalization . . . 78
4.2.1 System Model . . . 78
4.2.2 Existing Linear Time-Varying MMSE Equalizers . . . 80
4.2.3 Kalman Filter with Equalization Delay d . . . 83
4.2.4 Numerical Examples . . . 85
4.3 Decision Feedback Equalization . . . 93
4.3.1 System Model . . . 93
4.3.2 Time-Varying FIR MMSE DFEs . . . 94
4.3.3 Numerical Examples . . . 100
4.4 Conclusion . . . 103

5 A MULTIPLE MODEL APPROACH TO DOUBLY-SELECTIVE CHANNEL ESTIMATION USING EXPONENTIAL BASIS MODELS 104
5.1 Introduction . . . 104
5.2 System Model and Objectives . . . 105
5.2.1 System Model . . . 105
5.2.2 Block-Adaptive Channel Estimation [48] . . . 107
5.2.3 Objectives . . . 108
5.3 Multiple Model Approach . . . 109
5.3.1 Multiple Models . . . 109
5.3.2 Interacting Multiple Model Data Detection . . . 110
5.4 Numerical Examples . . . 112
5.4.1 Example 1: IMM with Three Models . . . 113
5.4.2 Example 2: IMM with Two Models . . . 115

5.5 Conclusion 122

6 DOUBLY-SELECTIVE CHANNEL ESTIMATION FOR OFDM SYSTEMS USING DPS-BEM AND TIME-MULTIPLEXED TRAINING 123
6.1 Introduction 123
6.2 System Model 125
6.3 Doubly-Selective Channel Estimation for OFDM Systems Using Frequency-Domain Training [62] 126
6.4 Numerical Examples 128
6.4.1 Example 1: Channel Estimation Performance 129
6.4.2 Example 2: CE-BEM versus DPS-BEM 130
6.5 Conclusion 130

7 SUMMARY AND FUTURE WORK 133
7.1 Summary of Original Work 133
7.2 Possible Future Directions 135

BIBLIOGRAPHY 138

A ASYMPTOTIC DPS-BEM/ SLEPIAN SEQUENCES 145

B MATHEMATICAL NOTATIONS 147

C ABBREVIATIONS 149

LIST OF FIGURES

1.1 Basis expansion modeling error comparison between CE-BEM and DPS-BEM models 9

1.2 Subchannels are 312 kHz wide in 802.11a and HyperLAN II 18

1.3 An OFDM transceiver diagram 19

2.1 Channel estimation errors (with and without modeling errors):Numerical results derived from $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$, SNR=20dB 41

2.2 MSE lower bound comparison between LS (2.44) and MMSE (2.45) estimators . . 42

2.3 Comparison between the LS channel estimation MSE lower bound in (2.44) and simulation results in (2.65), $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1$ 42

2.4 Comparison between the MMSE channel estimation MSE lower bound in (2.45) and simulation results in (2.65), $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1$ 43

2.5 LS channel estimation MSE with varying maximum normalized Doppler bandwidth, $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1$,$Q_F = 2\lceil \nu_{\mathrm{Dmax}} N \rceil + 1$ 44

2.6 BER with varying maximum normalized Doppler bandwidth using BPSK modulation, $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1$,$Q_F = 2\lceil \nu_{\mathrm{Dmax}} N \rceil + 1$ 45

2.7 BER with varying maximum normalized Doppler bandwidth using QPSK modulation, $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1$,$Q_F = 2\lceil \nu_{\mathrm{Dmax}} N \rceil + 1$ 46

2.8 Simulations-based BER versus β for SNR=15dB 47

2.9 Theoretical β_{opt} (2.64) versus received signal SNR 48

3.1 Channel estimation errors (with and without modeling errors): numerical results derived from $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$, SNR=25dB . . 67

3.2 MSE lower bound comparison between LS (3.37) and MMSE (3.38) estimators . . 68

3.3 Comparison between channel estimation MSEs lower bound in (3.37) and simulation results (3.58) 68

3.4 Comparison between channel estimation MSEs lower bound in (3.38) and simulation results (3.58) 69

3.5 Channel estimation MSE with varying maximum normalized Doppler bandwidth, $Q_F = 2\lceil \nu_{Dmax} N \rceil$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil$ 70

3.6 Channel estimation MSE with varying maximum normalized Doppler bandwidth, $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$ 71

3.7 BER with varying maximum normalized Doppler bandwidth for BPSK modulation, $Q_F = 2\lceil \nu_{Dmax} N \rceil$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil$ 72

3.8 BER with varying maximum normalized Doppler bandwidth for QPSK modulation, $Q_F = 2\lceil \nu_{Dmax} N \rceil$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil$ 73

3.9 BER with varying maximum normalized Doppler bandwidth for BPSK modulation, $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$ 73

3.10 BER with varying maximum normalized Doppler bandwidth for QPSK modulation, $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$ 74

4.1 BER performance versus SNR using BPSK modulation, averaged over 10,000 runs. BLM denotes the method of [5]; FRESH-OPT and FRESH-SUBOPT are methods of [68]. $N = 50$, $P = 100$, $Q = 4$, $L = 3$, $d = 5$, $Q' = 18$, $L' = 6$ 88

4.2 BER performance versus SNR using QPSK modulation, averaged over 10,000 runs 89

4.3 BER performance versus SNR using BPSK modulation, averaged over 1,000 runs. BLM denotes the method of [5]; FRESH-OPT and FRESH-SUBOPT are methods of [68]. Kalman filter is based on $P = 800$ when $Q = 4$ and on $P = 1600$ when $Q = 8$. $P = N$ corresponds to critically sampled CE-BEM and $P = 2N$ corresponds to oversampled CE-BEM. $N = 800$, $P = 800$ or 1600, $Q = 4$ or $Q = 8$, $L = 3$, $d = 10$, $Q' = 12$, $L' = 12$ 90

4.4 BER performance versus SNR using QPSK modulation, averaged over 1,000 runs . 91

4.5 BER performance versus SNR For DPS-BEM and CE-BEM channel modeling using Kalman filtering, BPSK modulation 91

4.6 BER performance versus SNR For DPS-BEM and CE-BEM channel modeling using Kalman filtering, QPSK modulation 92

4.7 BER performance versus SNR, averaged over 1,000 runs, BPSK signal. $N = 800$, $P = 1600$, $Q = 8$, $L = 3$, d=5, Q'=12, l_f=12, $Q'' = 4$, $l_b = 3$. R denotes the number of receive antennas. 100

4.8 BER performance versus SNR, averaged over 1,000 runs, BPSK signal. $N = 800$, $P = 1600$, $Q = 8$, $L = 3$, d=5, Q'=12, l_f=12, $Q'' = 4$, $l_b = 3$. R denotes the number of receive antennas. 101

5.1 IMM with three models, BER vs SNR with BPSK information symbols. 113

5.2 IMM with three models, BER vs SNR with QPSK information symbols. 114

5.3 IMM with two models. BER vs SNR with BPSK information symbols. Step shape time-varying Channel 118

5.4 IMM with two models. BER vs SNR with QPSK information symbols. Step shape time-varying Channel 118

5.5 Average mode probability of IMM. Step shape time-varying channel. SNR=20dB . 119

5.6 IMM with two models. BER vs SNR with BPSK information symbols. Linear Shape Time-Varying Channel 119

5.7 IMM with two models. BER vs SNR with QPSK information symbols. Linear Shape Time-Varying Channel 120

5.8 Average mode probability of IMM. Linear shape time-varying channel. SNR=20dB 120

5.9 Average mode probability of IMM. Time-invariant channel. With training insertion. SNR=20dB 121

5.10 Average mode probability of IMM. Time-invariant channel. Without training insertion. SNR=20dB 121

6.1 OFDM transceiver block diagram in [62] 123

6.2 Proposed OFDM transceiver block diagram 124

6.3 LS Channel estimation MSE comparison between TDE and FDE with varying ν_{Dmax} 129

6.4 LS Channel estimation MSE comparison between DPS-BEM and CE-BEM with varying ν_{Dmax}, TDE 131

6.5 BER comparison between DPS-BEM and CE-BEM with varying ν_{Dmax}, TDE . . 131

LIST OF TABLES

4.1 Operation summary for Kalman filter 85

4.2 Computation complexity; $N = 50$, $P = 2N = 100$ 87

4.3 Computation complexity; $N = 800$, $P = 2N = 1600$ 88

4.4 Computation complexity 102

5.1 Summary of the IMM algorithm (one cycle) 111

CHAPTER 1

INTRODUCTION

The doubly-selective channel estimation and equalization algorithms described in this dissertation involve some background material in the wireless communication area. We will first briefly introduce these topics in this chapter.

Starting with the characteristics and representations of wireless channels, two basis expansion models (BEM) used in this dissertation are introduced and compared. The pros and cons of different channel estimation and equalization approaches are then discussed. The fundamentals of the Viterbi decoder, a commonly used information detector in this dissertation, and the principles of Orthogonal Frequency-Division Multiplexing (OFDM) technology are also introduced.

1.1 Characteristics and Representations of Wireless Communication Channels

In wireless telecommunications, multipath is the propagation phenomenon that results in radio signals' reaching the receiving antenna by two or more paths. Causes of multipath include atmospheric ducting, ionospheric reflection and refraction, and reflection from terrestrial objects, such as mountains and buildings.

The effects of multipath include constructive and destructive interference, and phase shifting of the signal. This causes Rayleigh fading, named after Lord Rayleigh. Rayleigh fading is a statistical model for the effect of a propagation environment on a radio signal, such as that used by wireless devices. It assumes that the power of a signal that has passed through such a transmission medium (also called a communication channel) will vary randomly, or fade, according to a Rayleigh distribution - the radial component of the sum of two uncorrelated Gaussian random variables. Rayleigh

fading is used to refer to the rapid fluctuations of the received signal in both dimensions — time and frequency. If we assume that fading is caused by the superposition of a large number of independent scattered components, then the in-phase and quadrature components of the received signal can be assumed to be independent zero-mean Gaussian processes. The envelope A of the received signal has a Rayleigh probability density function (pdf) given by

$$f_A(a) := \begin{cases} \frac{a}{\sigma^2} \exp\left(-\frac{a^2}{2\sigma^2}\right) & a \geq 0, \\ 0 & a < 0 \end{cases} \tag{1.1}$$

with σ^2 being the time-average power of the received signal before envelope detection. The phase θ of the received signal is uniformly distributed with pdf

$$f_\Theta(\theta) := \frac{1}{2\pi},\ \theta \in [0, 2\pi)\,. \tag{1.2}$$

The autocorrelation function of the received signal for two-dimensional isotropic scattering and an omnidirectional receiving antenna is given by [17]

$$R_A(\tau) = \sigma^2 \cos(\omega_c \tau) J_0(\omega_m \tau) \tag{1.3}$$

where ω_c is the carrier radian frequency, $J_0(\cdot)$ is the zero-order Bessel function of the first kind and ω_m is the maximum Doppler radian frequency spread. The autocorrelation function of the Rayleigh fading channels is periodic in lag and its envelope decays slowly after the initial zero-crossing.

In a multipath propagation environment, several delayed and scaled versions of the transmitted signal arrive at the receiver. The span of path delays is called delay spread. Delay spread causes frequency-selective fading as the channel acts like a tapped delay line filter. Time-selective fading

due to scatter or transmitter/receiver motion results in a Doppler spread, i.e., a pure tone spreads over a finite spectral bandwidth. The frequency- and time- selective Rayleigh fading channel is the channel model we consider in this proposal.

1.1.1 Jakes' Model

As described above, a Rayleigh fading channel itself can be modeled by generating the real and imaginary parts of a complex number according to independent normal Gaussian variables. Any model simulating the Rayleigh fading channel has to exhibit the statistical behaviors given in (1.1) - (1.3).

In [34], Jakes popularized a time-varying model for Rayleigh fading based on summing sinusoids. The model supposes the received signal $g(t)$ at time t is

$$g(t) = E_0 \sum_{l=1}^{L} C_l \cos\left(\omega_c t + \omega_m t \cos A_l + \Phi_l\right) \tag{1.4}$$

where E_0 is the amplitude of the transmitted cosine wave, C_l is a random variable representing the attenuation of the l-th path, A_l is a random variable representing the angle of arrival of the l-th ray with respect to the direction of motion of the receiver, Φ_l is a random variable representing the phase shift undergone by the l-th ray. Note that the stochastic signal $g\left(t\right)$ representing the flat fading signal can be characterized by L sets of triples (C_l, A_l, Φ_l). The random variables C_l, A_l, and Φ_l are assumed statistically independent.

To reduce the complexity, the simplified Jakes' model selects:

$$C_l = \frac{1}{\sqrt{L}}, \tag{1.5a}$$

$$A_l = \frac{2\pi l}{L}, \tag{1.5b}$$

$$\Phi_l = 0, \tag{1.5c}$$

where $l = 1, 2, ..., L$. Furthermore, L is of the form $L = 4M + 2$ where M is a positive integer.

However, the simplifying relationships forced in (1.5) make this simulation model deterministic and wide-sense nonstationary [53]. Various modifications of Jakes' model have been proposed, which we call the family of Jakes' simulators. Among the Jakes' simulator family, [81] is worthy of mention since it generates a wide-sense stationary process and its second-order correlation statistics match desired reference model (1.4) exactly. Following [81], the normalized low-pass fading process of the statistical sum-of-sinusoids simulation model is defined by

$$X(t) = X_c(t) + jX_s(t), \tag{1.6a}$$

$$X_c(t) = \frac{2}{\sqrt{M}} \sum_{l=1}^{M} \cos(\vartheta_l) \cos(\omega_m t \cos\alpha_l + \phi), \tag{1.6b}$$

$$X_s(t) = \frac{2}{\sqrt{M}} \sum_{l=1}^{M} \sin(\vartheta_l) \cos(\omega_m t \cos\alpha_l + \phi) \tag{1.6c}$$

with

$$\alpha_l = \frac{2\pi l - \pi + \theta}{4M},\ l = 1, 2, ..., M$$

where α_l, ϕ_l, and ϑ_l are statistically independent and uniformly distributed over $[-\pi, \pi)$ for all l. As $M \to \infty$, the envelope $|X|$ is Rayleigh distributed and the phase $\Theta_X(t)$ is uniformly distributed

over $[-\pi, \pi)$, for which the pdf's are given by

$$f_{|X|}(x) = x \exp\left(-\frac{x^2}{2}\right),\ x \geq 0,$$
$$f_{\Theta_X}(\theta) = \frac{1}{2\pi},\ \theta \in [-\pi, \pi).$$

A minor defect, however, occurs in model $(1.6a)$ when $\omega_m = 0$ or the Doppler spread is small: A Rayleigh distribution cannot be guaranteed [77]. This problem can be easily resolved by replacing a common phase ϕ by ϕ_l, which is also uniformly distributed over $[-\pi, \pi)$ for all l. The simulation model is revised as [77]:

$$X(t) = X_c(t) + jX_s(t), \tag{1.7a}$$
$$X_c(t) = \frac{2}{\sqrt{M}} \sum_{l=1}^{M} \cos(\vartheta_l) \cos(\omega_m t \cos\alpha_l + \phi_l), \tag{1.7b}$$
$$X_s(t) = \frac{2}{\sqrt{M}} \sum_{l=1}^{M} \sin(\vartheta_l) \cos(\omega_m t \cos\alpha_l + \phi_l). \tag{1.7c}$$

1.1.2 Complex Exponential Basis Expansion Model (CE-BEM)

Statistical modeling of the channel is well motivated when time-varying path delays arise due to a large number of scatterers. Deterministic basis expansion models have gained popularity for wireless applications, especially when the multipath is caused by a few strong reflectors and path delays exhibit variations due to the mobiles [26]. The time-varying taps are expressed as a superposition of time-varying bases (complex exponentials when modeling Doppler effects) with time invariant coefficients. By assigning time variations to the bases, rapidly fading channels with coherence time as small as a few tens of symbols can be captured.

Consider a time-varying channel with impulse response $h(t;\tau)$ (response at time t to an unit impulse at time $t-\tau$) which includes transmit-receive filters as well as doubly-selective propagation effects. Let $s(t)$ denote the complex baseband, continuous-time input signal (with symbol duration T_s), and $x(t)$ denote the complex baseband, continuous-time received signal. The noise-free received signal $x(t)$ is the convolution of $s(t)$ and $h(t;\tau)$ [41]:

$$x(t) = \int_0^\infty h(t;\tau)s(t-\tau)d\tau. \tag{1.8}$$

Let $H(f;\tau) = \int_{-\infty}^{\infty} h(t;\tau)e^{-j2\pi ft}dt$ be the Fourier transform of $h(t;\tau)$; $H(f;\tau)$ is the delay-Doppler spreading function of the channel. If $|H(f;\tau)| \approx 0$ for $|\tau| > \tau_d$, then τ_d is called the (multipath) delay-spread of the channel; if $|H(f;\tau)| \approx 0$ for $|f| > f_d$, the f_d is called the Doppler spread of the channel [13]. If $s(t),\ x(t)$ and $h(t;\tau)$ in (1.8) are sampled at symbol rate, then by [41], for $t = nT_s \in [t_0, t_0 + NT_s)$, the sampled signal $x(n) := x(t)|_{t=nT_s}$ has the representation

$$x(n) = \sum_{l=0}^{L} h(n;l)s(n-l) \tag{1.9}$$

where T_s is the symbol duration. Over the block interval of $[t_0, t_0 + NT_s)$, the channel impulse response $h(n;l)$ can be represented using $Q_F + 1$ coefficients $\{w_q(l)\}_{q=0}^{Q_F}$, which remain invariant during this block but are allowed to change for the next block, and $Q_F + 1$ Fourier basis functions that are used to describe the temporal variation of the channel and are common for each block. Then for the block of $[t_0, t_0 + NT_s)$, the discrete-time baseband equivalent channel model based on complex exponential basis expansion can be described as [13, 41]:

$$h(n;l) = \sum_{q=0}^{Q_F} w_q(l)e^{j\omega_q n}, \tag{1.10}$$

where

$$\omega_q := \frac{2\pi}{T}(q - \frac{Q_F}{2}), \quad q = 0, 1, \ldots, Q_F, \tag{1.11a}$$

$$L := \lfloor \tau_d / T_s \rfloor, \tag{1.11b}$$

$$Q_F \geq 2\lceil f_d N T_s \rceil. \tag{1.11c}$$

There are two slightly different CE-BEMs involved in this dissertation. One is referred to as the critically-sampled CE-BEM because the BEM period equals the length of the observed window. The other uses a longer BEM period and is thus referred to as the oversampled CE-BEM. Section 4.2.1 provides the reader more details about oversampled CE-BEM. Note that the basis functions in critically-sampled CE-BEM are orthogonal, while in over-sampled CE-BEM they are not.

1.1.3 Discrete Prolate Spheroidal Basis Expansion Model (DPS-BEM)

It has been known that the Fourier basis function based CE-BEM model has the major drawback that the rectangular window associated with the discrete Fourier transform (DFT) introduces spectral leakage [44], which results in the floor in the bit error rate (BER) [3]. In [76, 77], the BEM coefficients are expanded by the orthogonal discrete prolate spheroidal (DPS) sequences resulting in a basis expansion model (DPS-BEM). The DPS sequences have a double orthogonality property: They are orthogonal over the finite set $\{0, ..., N-1\}$ and the infinite set $\{-\infty, ..., \infty\} = \mathbb{Z}$ simultaneously. This remarkable property enables parameter estimation without the drawbacks of windowing in the case of the Fourier basis expansion.

Let T_s denote the symbol interval. For a channel with a multipath delay spread of τ_d sec and a Doppler spread of f_d Hz, $\nu_{\mathrm{Dmax}} := f_d T_s$ is the maximum normalized Doppler bandwidth. An ideal basis function should have at least two properties: It is band-limited to the normalized

frequency range $[-f_dT_s, f_dT_s]$; its energy is time-concentrated in a certain time interval $[0, N-1]$. Thus, given the maximum normalized Doppler bandwidth f_dT_s and the window size N, we seek a sequence $\psi(n)$ to maximize

$$\lambda = \frac{\sum_{n=0}^{N-1} |\psi(n)|^2}{\sum_{m=-\infty}^{\infty} |\psi(m)|^2} \tag{1.12}$$

with the band-limited constraint

$$\psi(n) = \int_{-f_dT_s}^{f_dT_s} \Psi(f) e^{j2\pi fn} df \tag{1.13}$$

where

$$\Psi(f) = \sum_{m=-\infty}^{\infty} \psi(m) e^{-j2\pi fm}. \tag{1.14}$$

The DPS sequences $\{\psi_q(n)\}$ give us the solution to the constrained maximization problem [58]. In DPS-BEM, the DPS vectors $\boldsymbol{\psi}_q := [\psi_q(0), ..., \psi_q(N-1)]^T \in \mathbb{R}^N$ (called Slepian sequences in [77], which are time-windowed DPS sequences) with elements $\psi_q(n)$ for $n \in \{0, ..., N-1\}$, are eigenvectors of the matrix $\boldsymbol{C} \in \mathbb{R}^{N \times N}$, fulfilling [77]

$$\boldsymbol{C}\boldsymbol{\psi}_q = \lambda_q \boldsymbol{\psi}_q, \tag{1.15}$$

where $\lambda_1 \geq \lambda_1 \geq \cdots \geq \lambda_N$ are eigenvalues of matrix $\boldsymbol{C}$. The (y, z) entries in matrix $\boldsymbol{C}$ are defined as:

$$[\boldsymbol{C}]_{y,z} = \frac{\sin[2\pi(y-z)\nu_{\mathrm{Dmax}}]}{\pi(y-z)}, \tag{1.16}$$

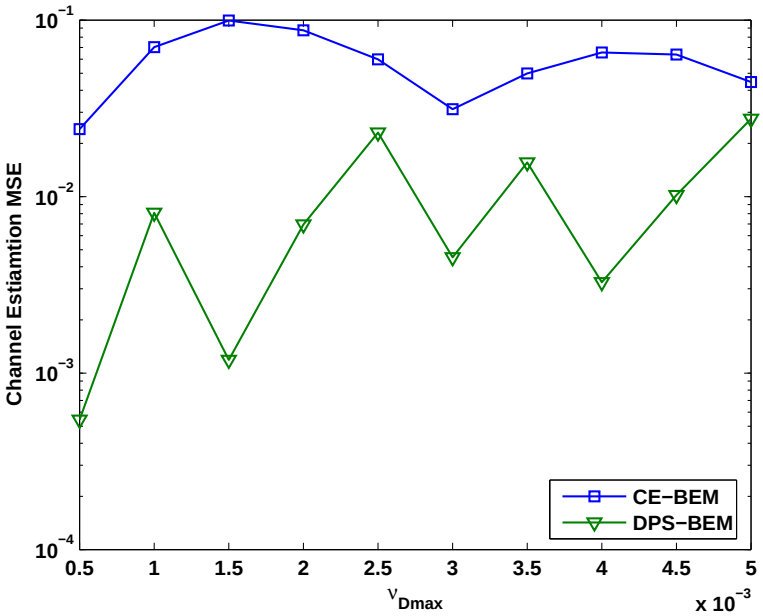

Figure 1.1: Basis expansion modeling error comparison between CE-BEM and DPS-BEM models

where $y, z \in \{0, ..., N-1\}$. In this case one takes $Q_S \geq \lceil 2\nu_{\mathrm{Dmax}} N \rceil$ [77, 58].

1.1.4 Modeling Error Comparison between CE-BEM and DPS-BEM

In the following, we use generic notations Q and $\psi_q(n)$ to indicate that the expressions are applicable to both CE-BEM and DPS-BEM basis functions. In a basis expansion representation over a time-block $n = \{0, 1, ..., N-1\}$, it is assumed that [26]

$$h(n; l) = \sum_{q=0}^{Q} w_q(l)\psi_q(n). \tag{1.17}$$

However, the true channel may not be exactly equal to this basis expansion since modeling error always occurs, so that we have to revise (1.17) as

$$h(n;l) = \sum_{q=0}^{Q} w_q(l)\psi_q(n) + e_{\mathrm{BEM}}(n;l), \tag{1.18}$$

where $e_{\mathrm{BEM}}(n;l)$ is the basis expansion modeling error. By the orthogonality principle, $e_{\mathrm{BEM}}(n;l)$ should be orthogonal to the given basis set $\{\psi_q(n)\}_{q=0}^{Q}$ when the square error $\sum_{n=0}^{N-1}|e(n;l)|^2$ is minimized. Then

$$\sum_{n=0}^{N-1} h(n;l)\psi_{q'}^*(n) = w_{q'}(l)\sum_{n=0}^{N-1}|\psi_{q'}(n)|^2. \tag{1.19}$$

Since the orthogonal basis functions of CE-BEM and DPS-BEM satisfy $\sum_{n=0}^{N-1}|\psi_{q'}(n)|^2 = 1$, the BEM coefficient is derived as

$$w_{q'}(l) = \sum_{n=0}^{N-1} h(n;l)\psi_{q'}^*(n). \tag{1.20}$$

Based on (1.19) and (1.20), the least squares approximation of the channel is given by

$$\hat{h}(n;l) = \sum_{q'=0}^{Q} w_{q'}(l)\psi_{q'}(n). \tag{1.21}$$

Fig. 1.1 shows the modeling error comparison between CE-BEM and DPS-BEM models. In this example, an SISO doubly-selective Rayleigh fading channel defined in (1.7) that is based on the modified Jakes model is considered. Since the Jakes model is more realistic as a practical wireless channel, we are trying to approximate a 3-tap Jakes channel by two BEMs: CE-BEM and DPS-BEM. We choose a data record length of 400 symbols, $T_s = 25\mu$s, average over $M_\alpha = 1000$ realizations of randomly generated channels, and plot the channel estimation mean square error

(MSE), which is defined as:

$$\text{MSE} := \frac{1}{M_\alpha N} \sum_{\alpha=1}^{M_\alpha} \sum_{n=0}^{N-1} \sum_{l=0}^{L} \|h_\alpha(n;l) - \hat{h}_\alpha(n;l)\|^2. \tag{1.22}$$

In Fig 1.1, the basis function dimensions Q_F and Q_S change with the maximum Doppler bandwidth $\nu_{\text{Dmax}} := f_d T_s$, and SNR=20dB. From the results, we notice that the channel estimation MSE of DPS-BEM is usually several orders of magnitude lower than that of the CE-BEM. Clearly DPS-BEM is better than CE-BEM when approximating a band-limited time-varying channel.

1.2 Channel Estimation and Equalization Approaches

1.2.1 Channel Estimation Approaches

For channel state information (CSI) acquisition, three classes of approaches are available: the training-based approach, the blind approach and the semi-blind approach. In the following, we briefly describe these three approaches.

In conventional training-based approaches, training sequences (known to the receiver) are time-multiplexed with the information sequence [21, 32, 63, 79]. Training symbols can be placed either at the beginning of each burst (as a preamble) or regularly throughout the burst. In rapidly fading or quasi-static fading channels, preamble-based training may not work well. This motivates embedding training symbols in each transmitted block, instead of concentrating them at the preamble. More recently, a superimposed training approach has been explored where the training sequences are added (superimposed) at a low power to the information sequence at the transmitter before modulation and transmission [65]. There is no loss in information rate, but the channel estimation using superimposed training will be interfered with by information symbols. The conventional

time-multiplexed training approach is attractive especially when it decouples symbol detection from channel estimation and thus simplifies the receiver implementation and relaxes the required channel identifiability conditions [52].

Blind approaches have been proposed to mitigate the multipath effects in wireless communications. Blind equalization algorithms are usually based on optimization procedures trying to minimize some nonlinear functional(s) of the received samples [66]. Compared with a training-based approach, a blind approach avoids training and thus makes an efficient use of the available bandwidth [66, 25]. But on the other hand, blind algorithms typically require longer data records and entail higher complexity [66]. Other major drawbacks of blind approaches are the slow convergence time of the equalizer and possible misconvergence which takes place when the convergence process reaches a local minimum of the functional to be minimized. For certain applications these disadvantages can be unacceptable.

Semi-blind approaches use a combination of training and blind cost functions. A training sequence is used in a semi-blind approach. At the receiver, the channel estimation depends on both the known training sequence and the unknown information sequence. This way, in addition to the information carried by the training symbols, the unknown information sequence is also exploited to enhance the channel estimation performance. Using a semi-blind approach allows the length of the training sequence to be shortened compared to the traditional training-based approaches [27, 78]. Compared with training-based approaches and blind approaches, semi-blind approaches have a relative better bandwidth efficientcy than training-based approaches and converge faster than blind approaches.

1.2.2 Equalization Approaches

With the emergence of next-generation wireless mobile communications, multimedia services have increasing demands for high data rates and high mobility. The high data rates give rise to frequency selectivity, while the mobility and carrier offset introduce time selectivity. To confront the doubly-selective effects of wireless channels, equalizers are usually employed at the receiver end. The existing equalizers can be divided into two types: block equalizers and serial equalizers.

The interest in block equalization can be motivated as follows. When transmission channels are affected by both frequency and time selectivity, reliable communication can be achieved by dividing the information data stream into short blocks and by adding a header of known data to each block. The known symbols allow us to obtain the reliable channel identification and to prevent interference between two adjacent blocks. Receiver processing can be carried out on a block-by-block basis so that if the transmission channel does not change appreciably during the transmission of each block, the receiver has to cope with the frequency selectivity only. Block equalization strategies are used to compensate for this channel impairment [36, 18]. However, block equalizers are usually complex to design since the inversion of a large matrix is required. Especially, since a doubly-selective channel can not be diagonalized by a channel-independent transformation, the implementation of block time-varying (TV) equalizers, which collect and process in blocks all the available data in the received frame, leads to a very high computational complexity [5].

Serial equalizers process few data at a time and provide a flexible trade-off between complexity and performance [66].

1.3 Optimal Maximum Likelihood Detector (Viterbi Decoder)

In a communication system, the role of channel estimation is to aid in extracting the desired information data from the distorted receive symbols. Next we will briefly review the commonly used symbol detection technique - Viterbi detector.

The Viterbi algorithm, originally introduced as a method for decoding convolutional codes, has become one of the most commonly used detectors in digital communications. Forney [23] has shown in 1972 that the algorithm solves maximum likelihood sequence detection (MLSD) of a pulse amplitude modulated (PAM) sequence of symbols with finite intersymbol interference (ISI) and memoryless noise. The algorithm has earned its place in almost every modern digital communications textbook where it is recognized as the optimal sequence detector for memoryless noise.

Consider a SIMO (single-input multiple-output) FIR (finite impulse response) linear channel with R outputs and discrete-time impulse response $\boldsymbol{h}(n;l)$. Let $\{s(n)\}$ denote the input sequence to the SIMO channel. The noisy channel output is given by

$$\boldsymbol{y}(n) = \sum_{l=0}^{L} \boldsymbol{h}(n;l)s(n-l) + \boldsymbol{\eta}(n) \tag{1.23}$$

where $L+1$ is the multipath channel length and $\boldsymbol{\eta}(n)$ is the white complex Gaussian noise. Assume that the white Gaussian noise $\boldsymbol{\eta}(n)$ is uncorrelated with $\{s(n)\}$, with mean $E\{\boldsymbol{\eta}(n)\} = \mathbf{0}$ and $E\{[\boldsymbol{\eta}(n+\tau)][\boldsymbol{\eta}(n)]^H\} = \sigma_\eta^2 \boldsymbol{I}_R \delta(\tau)$.

Given $s(n)$, $\boldsymbol{y}(n)$ is a R-dim Gaussian random vector with mean $\sum_{l=0}^{L} \boldsymbol{h}(l)s(n-l)$ and variance $\sigma_\eta^2 \boldsymbol{I}_R$. The joint probability density function (pdf) of $\boldsymbol{y}(n)$ given $\{s(n), s(n-1), \ldots, s(n-L)\}$

is

$$p(\boldsymbol{y}(n)|s(n),\ldots,s(n-L)) = \frac{1}{(\pi\sigma_\eta)^R}\exp\left\{-\frac{1}{\sigma_\eta^2}\left\|\boldsymbol{y}(n)-\sum_{l=0}^{L}\boldsymbol{h}(n;l)s(n-l)\right\|^2\right\} \quad (1.24)$$

where $s(n) = 0$ for $n < 0$. The joint pdf of the random vectors $\{\boldsymbol{y}(0),\ \boldsymbol{y}(1),\ \ldots,\ \boldsymbol{y}(N-1)\}$ given the transmitted sequence $\{s(0),\ s(1),\ \ldots,\ s(N-1)\}$ is

$$\begin{aligned}
&p(\boldsymbol{y}(0),\ldots,\boldsymbol{y}(N-1)|s(0),\ldots,s(N-1))\\
&=\prod_{n=0}^{N-1}\frac{1}{(\pi\sigma_\eta)^R}\ \exp\left\{-\frac{1}{\sigma_\eta^2}\left\|\boldsymbol{y}(n)-\sum_{l=0}^{L}\boldsymbol{h}(l)s(n-l)\right\|^2\right\},\\
&=\frac{1}{(\pi\sigma_\eta)^{NR}}\ \exp\left\{-\frac{1}{\sigma_\eta^2}\sum_{n=0}^{N-1}\left\|\boldsymbol{y}(n)-\sum_{l=0}^{L}\boldsymbol{h}(l)s(n-l)\right\|^2\right\}. \qquad (1.25)
\end{aligned}$$

Taking the logarithm on both sides of the equation above, we have

$$\begin{aligned}
&\ln p(\ \boldsymbol{y}(0),\ldots,\boldsymbol{y}(N-1)|s(0),\ldots,s(N-1))\\
&=-NR\ln(\pi\sigma_\eta)-\frac{1}{\sigma_\eta^2}\sum_{n=0}^{N-1}\left\|\boldsymbol{y}(n)-\sum_{l=0}^{L}\boldsymbol{h}(n;l)s(n-l)\right\|^2. \qquad (1.26)
\end{aligned}$$

The ML (maximum likelihood) estimate of the input sequence $\{s(0),\ \ldots,\ s(N-1)\}$ is the one that maximizes

$$p(\boldsymbol{y}(0),\ldots,\boldsymbol{y}(N-1)\mid s(0),\ldots,s(N-1));$$

or equivalently maximizes

$$\ln p(\boldsymbol{y}(0),\ldots,\boldsymbol{y}(N-1)\mid s(0),\ldots,s(N-1));$$

or minimizes the Euclidean distance

$$\sum_{n=0}^{N-1} \left\| \boldsymbol{y}(n) - \sum_{l=0}^{L} \boldsymbol{h}(n;l)s(n-l) \right\|^2 .$$

This MLSE (maximum likelihood sequence estimation) criterion is equivalent to the problem of estimating the state of a discrete-time "finite-state machine". In this case, the finite-state machine is the discrete-time channel with coefficients $\{\boldsymbol{h}(n;l)\}$ and its state at any time n is represented by the L most recent input symbols

$$\text{State}_n = (\, s(n),\ s(n-1), \ldots,\ s(n-L+1)\,)$$

where $s(n) = 0$ for $n < 0$. If the input symbols are M-ary, the finite-state machine has M^L states. Consequently, the channel is described by an M^L-state trellis and the Viterbi algorithm may be used to determine the most probable path through the trellis. In brief, we describe the Viterbi algorithm in the following 3 steps:

Step 1. We begin with $\boldsymbol{y}(L)$, from which we compute the M^{L+1} metrics

$$\sum_{n=0}^{L} \left\| \boldsymbol{y}(n) - \sum_{l=0}^{L} \boldsymbol{h}(n;l)s(n-l) \right\|^2 .$$

The M^{L+1} possible sequences are divided into M^L groups according to the M^L states. From each group, we pick the one with the minimum metric, i.e., the most probable sequence, and assign to the surviving sequence the metric

$$PM_0(s(L),\ \ldots,\ s(1)) = \min_{s(0)} \left\{ \sum_{n=0}^{L} \left\| \boldsymbol{y}(n) - \sum_{l=0}^{L} \boldsymbol{h}(n;l)s(n-l) \right\|^2 \right\} .$$

The $M-1$ remaining sequences from each of the M^L groups are discarded.

Step 2. Upon reception of $\boldsymbol{y}(L+n)$, $n \geq 1$, compute the M^{L+1} metrics

$$\left\| \boldsymbol{y}(L+n) - \sum_{l=0}^{L} \boldsymbol{h}(n;l) s(L+n-l) \right\|^2 + PM_{n-1}(s(L+n-1),\ \ldots,\ s(n)).$$

Again, the M^{L+1} sequences are divided into M^L groups corresponding to the M^L possible state $(s(L+n), \ldots, s(n+1))$ and the most probable sequence from each group is selected while the other $M-1$ sequences are discarded. The surviving metrics are

$$PM_n(s(L+n),\ \ldots,\ s(n+1))$$

$$= \min_{s(n)} \left\{ \left\| \boldsymbol{y}(L+n) - \sum_{l=0}^{L} \boldsymbol{h}(n;l) s(L+n-l) \right\|^2 + PM_{n-1}(s(L+n-1),\ \ldots,\ s(n)) \right\}. \tag{1.27}$$

Step 3. If $\boldsymbol{y}(L+n)$ is the last received sample, from the M^L survivor sequences, pick the one as the ML (maximum likelihood) sequence estimator which has the minimum metric; otherwise, set $n = n+1$ and then go to step 2.

1.4 Orthogonal Frequency Division Multiplexing (OFDM)

Frequency division multiplexing is a technology that transmits multiple signals simultaneously over a single transmission path, such as a cable or wireless system. Each signal travels within its own unique frequency range (carrier), which is modulated by the data (text, voice, video, etc.).

Orthogonal frequency division multiplexing (OFDM) distributes the data over a large number of subchannels that are spaced apart at precise frequencies (see Fig. 1.2). This spacing provides the "orthogonality" in this technique which prevents the demodulators from seeing frequencies other

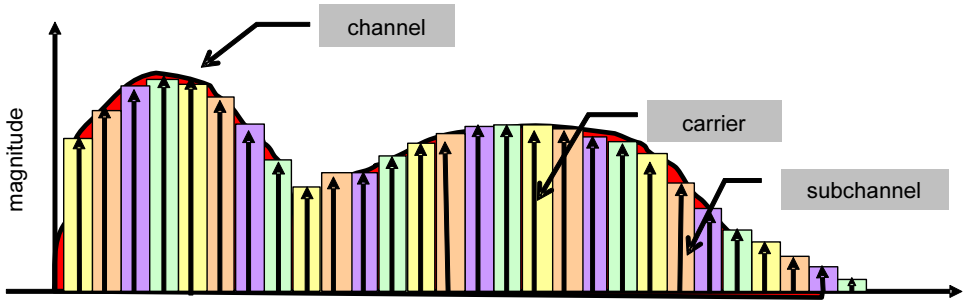

Figure 1.2: Subchannels are 312 kHz wide in 802.11a and HyperLAN II

than their own. The benefits of OFDM are high spectral efficiency, resiliency to radio frequency (RF) interference, and lower multipath distortion. This is useful because in a typical terrestrial broadcasting scenario there are multipath-channels (i.e. the transmitted signal arrives at the receiver using various paths of different length). Since multiple versions of the signal interfere with each other (intersymbol interference) it becomes very hard to extract the original information.

OFDM has already been included in digital audio/video broadcasting (DAB/DVB) standards in Europe, and has been successfully applied to high-speed digital subscriber line (DSL) modems in the United States. Recently, it has also been proposed for digital cable television systems and local area mobile wireless networks, such as those specified in the IEEE802.11a, and the HIPERLAN/2 standards [15]. By implementing an inverse fast Fourier transform (IFFT) at the transmitter and a fast Fourier transform (FFT) at the receiver, OFDM converts an intersymbol interference (ISI) channel into parallel ISI-free subchannels with gains equal to the channel's frequency response values on the FFT grid. To eliminate interblock interference (IBI) between successive IFFT-processed blocks, a cyclic prefix (CP) of length no less than the channel order is inserted per transmitted block. Discarding the CP at the receiver not only suppresses IBI, but also converts the linear channel convolution into circular convolution, which facilitates diagonalization of the associated channel matrix (see, e.g., [71]). An OFDM transceiver diagram is shown in Fig. 1.3.

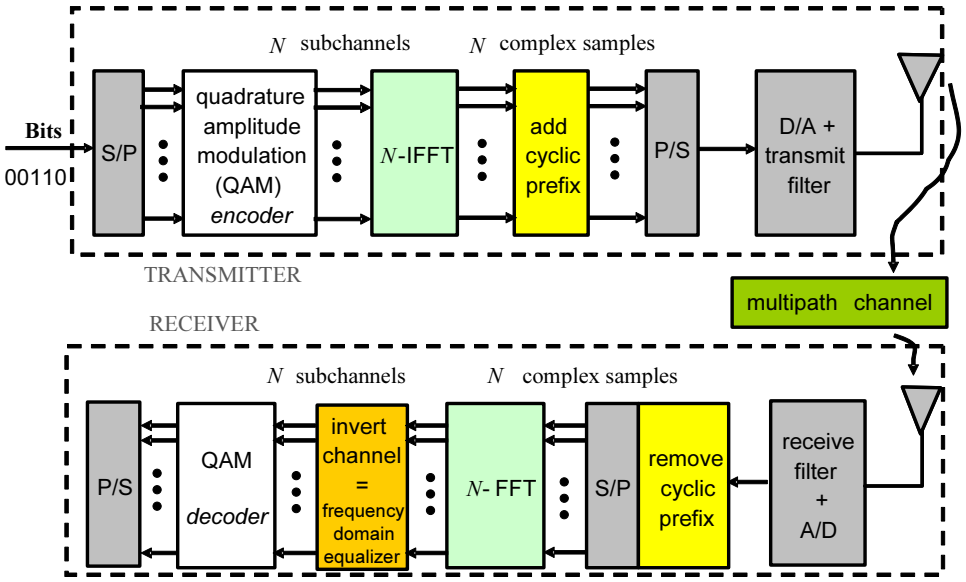

Figure 1.3: An OFDM transceiver diagram

1.5 Outline and Contribution

The dissertation is organized in the following chapters and the author's contributions are as follows:

Chapter 2: The channel estimation for doubly-selective channels is considered using time-multiplexed training. The time-varying channel is assumed to be well-described by a basis expansion model using discrete prolate spheroidal sequences as the bases (DPS-BEM). The popular linear least squares and minimum mean-square-error approaches are exploited to estimate the basis expansion coefficients. Computer simulations based on Monte Carlo runs are provided. With the channel estimation MSE and bit error rate as the performance measures, we find that the channel estimation based on DPS-BEM significantly outperforms the more widely used complex exponential basis expansion model-based channel estimation.

Certain aspects of pilot symbol aided modulation (PSAM) parameter design for DPS-BEM-based doubly-selective channels is investigated, following the CE-BEM results in [41, 75]. The optimum time-multiplexed training structure design based on an asymptotic DPS expression is presented by minimizing the DPS-BEM-based channel estimation MSE. The training power allocation problem is also addressed in this chapter.

Chapter 3: The channel estimation approaches proposed in Chapter 2 are then extended to multiuser MIMO systems in this chapter. We suppose there are multiple users at the transmit side, and each user has one antenna.

Chapter 4: The conventional Kalman filter is used as a time-varying (TV) minimum mean-square-error (MMSE) equalizer for doubly-selective channels. A formulation of FIR decision feedback equalizer is proposed for doubly-selective channels. The proposed equalizers have the main advantages that they do not incur the approximation error inherent in BEM modeling of equalizers. The BER performance and the computational complexity of the proposed design are investigated by means of Monte Carlo computer simulations, and compared with the existing BEM-based TV equalizers.

Chapter 5: An adaptive channel estimation scheme, exploiting the oversampled CE-BEM, is presented for doubly-selective channels where we track the BEM coefficients via a multiple model approach. In the past work the number of BEM coefficients used to model the doubly-selective channels for channel estimation has been based on an upper bound on the channel Doppler spread. The higher the Doppler spread, the more the number of BEM coefficients leading to a higher channel estimation variance. In this chapter we propose to use a multiple model framework where several candidate Doppler spread values are used to cover the range from zero to an upper bound, leading to multiple CE-BEM channel models, each corresponding to an assumed value of the Doppler spread.

Subsequently the well known interacting multiple model (IMM) algorithm is used for symbol detection based on multiple state-space models corresponding to the multiple estimated channels.

Chapter 6: The pilot-aided doubly-selective channel estimation for OFDM systems is considered in this chapter. The time-varying channel is described by CE-BEM or DPS-BEM. The "optimum" training strategies proposed in Chapter 2 are applied to OFDM systems under doubly-selective channels. Compared to the traditional frequency-domain training design, the main advantages of time-domain training for OFDM system is that the information symbols are not contaminated by the training symbols as in the frequency-domain training case. The performance of frequency-domain training-based channel estimation and time-domain training-based channel estimation is presented and compared.

Chapter 7: The dissertation concludes in Chapter 7 with future research topics suggested.

CHAPTER 2

SISO DOUBLY-SELECTIVE CHANNEL ESTIMATION USING DISCRETE PROLATE SPHEROIDAL BASIS EXPANSION MODELS AND TIME-MULTIPLEXED TRAINING

2.1 Introduction

Doubly-selective channel estimation using complex exponential basis expansion model (CE-BEM) and time-multiplexed training is considered in [41, 75], where CE-BEM based on Fourier basis functions is applied to represent the time-variant channel. However, since the Fourier basis expansion has the major drawback that the rectangular window associated with the discrete Fourier transform (DFT) introduces spectral leakage [44], the bit error rate (BER) suffers an error floor [3, 76]. In [76, 77], the linear minimum mean-square-error (MMSE) channel estimation using discrete prolate spheroidal (DPS) sequences is considered. It is shown that DPS-BEM-based approaches significantly outperform CE-BEM-based approaches for the doubly-selective channel estimation and data detection.

To acquire the channel state information at the receiver, training symbols are usually periodically inserted during transmission, which is known as pilot symbol aided modulation (PSAM) [16]. Optimization of the PSAM for CE-BEM based doubly-selective channel models has been considered in [41, 75] where the time-multiplexed training sequence is designed to minimize the channel estimation mean-square-error (MSE). In the case of CE-BEM with independent basis expansion coefficients, minimizing the channel estimation MSE is also shown (in [41, 75]) to be equivalent to maximizing a lower bound on the estimated channel-based average capacity. No such considerations are to be found in [76, 77] where the doubly-selective is represented by DPS-BEM. The linear MMSE channel estimator for DPS-BEM-based MIMO-OFDM (multiple input multiple

output-orthogonal frequency division multiplexing) doubly-selective channels is introduced in [49], but the training design problem is not considered.

In this chapter, we consider the channel estimation for doubly-selective single input single output (SISO) channels described by DPS-BEM. Both linear least squares (LS) and minimum mean-square-error (MMSE) estimators are presented and compared.

Our system model is exactly as in [41] except that instead of CE-BEM as in [41] we use DPS-BEM. In [41, 75], the linear MMSE channel estimator is used, which requires knowledge of the noise variance and of the covariance matrix of the channel basis expansion coefficients. While the former may be known at the receiver, the latter is seldom known. In [41], the latter is assumed to be known and diagonal. For the Jakes' model, the basis expansion coefficients for a given tap are not mutually uncorrelated, hence the diagonal assumption does not always hold true. The linear LS channel estimator does not need to know the covariance matrix or make any assumption regarding its nature. On the other hand, the performance of the MMSE channel estimator is better than that of the LS estimator; however, the difference is negligible at high SNRs. [41] does not consider the LS channel estimation whereas [77] has used the LS channel estimator but not the MMSE estimator.

Certain aspects of PSAM parameter design for DPS-BEM-based doubly-selective channels is also considered, following the CE-BEM results in [41, 75]. Since the Slepian sequences as a solution to (1.15) are hard to work with analytically, we give asymptotic DPS expressions based on some heuristic considerations. Then the optimum time-multiplexed training structure design based on asymptotic DPS expressions is presented by minimizing the DPS-BEM-based channel estimation MSE. The training power allocation problem is finally addressed.

2.2 System Model

Consider a doubly-selective single input single output (SISO) finite impulse response (FIR) linear channel. Let $h(n;l)$ denote the symbol-rate impulse response (the channel response at time n to a unit input at time $n-l$), where $n \in [0,1,...,N-1]$ and $l \in [0,L]$ capture the time- and frequency- selectivity of the channel, respectively. Over a time-block of size N, given N orthonormal functions of $\psi_q(n)$, the following representation is always true

$$h(n;l) = \sum_{q=0}^{N-1} w_q(l)\psi_q(n), \tag{2.1}$$

where $\psi_q(n)$ is the q-th basis function and the basis expansion coefficient $w_q(l)$ is fixed over the data block. As the above representation is not parsimonious, the following BEM is used to approximate model (2.1):

$$h_{\mathrm{BEM}}(n;l) = \sum_{q=0}^{Q} w_q(l)\psi_q(n), \tag{2.2}$$

where only $Q+1 << N$ basis functions are involved. Hence, the channel modeling error can be expressed as:

$$e_{\mathrm{BEM}}(n;l) = h(n;l) - h_{\mathrm{BEM}}(n;l) = \sum_{q=Q+1}^{N-1} w_q(l)\psi_q(n). \tag{2.3}$$

Let T_s denote the symbol interval. For a channel with a multipath delay spread of τ_d sec and a Doppler spread of f_d Hz, in the complex exponential basis expansion model (CE-BEM) [56, 38],

one takes

$$\begin{aligned} \psi_q^{(F)}(n) &:= e^{j\omega_q n}, \\ \omega_q &:= 2\pi(q - \frac{Q_F}{2})/N, \\ L &:= \lfloor \tau_d/T_s \rfloor \text{ and } Q_F := 2\lceil \nu_{\mathrm{Dmax}} N \rceil \end{aligned}$$

where $\nu_{\mathrm{Dmax}} := f_d T_s$ is the maximum normalized Doppler bandwidth. After normalization to unit norm, we have

$$\psi_q^{(F)}(n) = \frac{1}{\sqrt{N}} e^{(i2\pi(q-Q_F/2)n/N)}. \tag{2.4}$$

In the discrete prolate spheroidal sequence-based BEM (DPS-BEM), the DPS vectors $\boldsymbol{\psi}_q^{(S)} \in \mathbb{R}^N$ (called Slepian sequences in [77], which are time-windowed DPS sequences) with elements $\psi_q^{(S)}(n)$ for $n \in \{0, ..., N-1\}$, are eigenvectors of the matrix $\boldsymbol{C} \in \mathbb{R}^{N \times N}$, fulfilling [77]

$$\boldsymbol{C}\boldsymbol{\psi}_q^{(S)} = \lambda_q \boldsymbol{\psi}_q^{(S)}, \tag{2.5}$$

where λ_q are eigenvalues of matrix $\boldsymbol{C}$. The (y, z) entries in matrix $\boldsymbol{C}$ are defined as:

$$[\boldsymbol{C}]_{y,z} = \frac{\sin[2\pi(y-z)\nu_{\mathrm{Dmax}}]}{\pi(y-z)}, \tag{2.6}$$

where $y, z \in \{0, ..., N-1\}$. In this case one takes $Q_S \geq \lceil 2\nu_{\mathrm{Dmax}} N \rceil$ [77, 58].

In the following, as in [77], we will use a general notation for the basis expansion quantities $\psi_q(n)$, Q, w_q, and h to indicate that all expressions are applicable to any set of orthonormal basis functions $\psi_q(n)$.

Let $\{s(n)\}$ denotes transmitter's information sequence. Using the discrete-time baseband equivalent channel model, the received sequence at the receive antenna can be written as

$$x(n) = \sum_{l=0}^{L} h_{\mathrm{BEM}}(n;l)s(n-l) + \eta(n), \tag{2.7}$$

where $\eta(n)$ is the additive complex Gaussian noise at the receive antenna, with zero-mean and variance σ_η^2. Plugging (2.2) into (2.7), we can rewrite $x(n)$ as:

$$x(n) = \sum_{q=0}^{Q} \psi_q(n) \left[\sum_{l=0}^{L} w_q(l)s(n-l) \right] + \eta(n). \tag{2.8}$$

We consider block transmission as in [41], where transmitted symbols are collected into $N \times 1$ blocks with $\boldsymbol{s} = [s(0), s(1), ..., s(N-1)]^T$ as the 0th block and received $x(n)$'s are also collect into blocks with $\boldsymbol{x} = [x(0), x(1), ..., x(N-1)]^T$ as 0th block. To avoid inter-block interference (IBI), as in [41], L guard zeros are inserted in each block at the transmitter. Then the matrix-vector input-output relationship of (2.8) is given by

$$\boldsymbol{x} = \sum_{q=0}^{Q} \boldsymbol{D}_{\psi_q} \boldsymbol{W}_q \boldsymbol{s} + \boldsymbol{\eta}, \tag{2.9}$$

where $\boldsymbol{\eta}$ is defined similarly to $\boldsymbol{x}$, $\boldsymbol{D}_{\psi_q} = \mathrm{diag}[\boldsymbol{\psi}_q]$ with $\boldsymbol{\psi}_q := [\psi_q(0), \psi_q(1), ..., \psi_q(N-1)]^T$, and $\boldsymbol{W}_q$'s are $N \times N$ lower triangular Toeplitz matrices with 1st column $[w_q(0), w_q(1), \ldots, w_q(L), 0, \ldots, 0]^T$.

2.3 Channel Estimation

Since $\boldsymbol{\psi}_q(n)$ are known by the receiver, the objective of the channel estimator is to find basis expansion coefficients in (2.2) from the received samples corresponding to the training symbols. The proposed channel estimation relies on time-multiplexed training symbols at known positions.

As in [41], each transmitted block $\boldsymbol{s}$ consists of J segments (sub-blocks) of training and information symbols $b(n)$ and $c(n)$, respectively, and each segment has the same length. Then the general structure of $\boldsymbol{s}$ is

$$\boldsymbol{s} = [\boldsymbol{b}_1^T, \boldsymbol{c}_1^T, ..., \boldsymbol{b}_J^T, \boldsymbol{c}_J^T]^T, \tag{2.10}$$

where $\boldsymbol{b}_j$ with length N_b and $\boldsymbol{c}_j$ with length N_c, for all $j \in [1, J]$, denote training and information symbol sub-blocks, respectively. Therefore, $N = J(N_b + N_c)$ with $N_b > L$. Let $M = N_b + N_c$ denote the sub-block size. Obviously, the first L symbols in the "training part" of the jth subblock of the received signal are contaminated by information symbols in the preceding $(j-1)$th subblock. In the similar way, the first L symbols in the "information part" of the jth subblock of the received signal are also contaminated by the last L training symbols in the current jth subblock. In order to avoid the inter-subblock interference (ISBI) so that the channel estimation is decoupled from data detection, we will choose the first and the last L symbols in each training subblock to be zeros, as in [41].

Define $\boldsymbol{b}_j := [b((j-1)M), b((j-1)M+1), ..., b((j-1)M+N_b-1)]^T$. Further define $\bar{\boldsymbol{D}}_{\psi_{q,j}} = \text{diag}[\bar{\boldsymbol{\psi}}_{q,j}]$ where $\bar{\boldsymbol{\psi}}_{q,j} = [\psi_q((j-1)M+L), \psi_q((j-1)M+L+1), ..., \psi_q((j-1)M+N_b-1)]^T$. Then the ISBI free received subblock can be written as

$$\bar{\boldsymbol{x}}_{b,j} = \sum_{q=0}^{Q} \bar{\boldsymbol{D}}_{\psi_{q,j}} \bar{\boldsymbol{W}}_q \boldsymbol{b}_j + \bar{\boldsymbol{\eta}}_{b,j}, \tag{2.11}$$

where $\bar{\boldsymbol{x}}_{b,j}^{(r)} := [x_b((j-1)M+L), x_b((j-1)M+L+1), ..., x_b((j-1)M+N_b-1)]^T$, $\bar{\boldsymbol{\eta}}_{b,j}$ is defined similarly, and $(N_b - L) \times N_b$ matrix $\bar{\boldsymbol{W}}_q$ is given by

$$\bar{\boldsymbol{W}}_q = \begin{bmatrix} w_q(L) & \dots & w_q(0) & & \\ & \ddots & \ddots & \ddots & \\ & & w_q(L) & \dots & w_q(0) \end{bmatrix}. \tag{2.12}$$

Gathering training symbols per block, we obtain

$$\bar{\boldsymbol{x}}_b = \sum_{q=0}^{Q} \begin{bmatrix} \bar{\boldsymbol{D}}_{\psi_{q,1}} \bar{\boldsymbol{W}}_q \boldsymbol{b}_1 \\ \vdots \\ \bar{\boldsymbol{D}}_{\psi_{q,J}} \bar{\boldsymbol{W}}_q \boldsymbol{b}_J \end{bmatrix} + \bar{\boldsymbol{\eta}}_b. \tag{2.13}$$

According to the commutativity property of convolution, we have $\bar{\boldsymbol{W}}_q \boldsymbol{b}_j = \boldsymbol{B}_j \boldsymbol{w}_q$ with $\boldsymbol{w}_q := [w_q(0), ..., w_q(L)]^T$ and $\boldsymbol{B}_j$ a $(N_b - L) \times (L+1)$ Toeplitz matrix given by

$$\boldsymbol{B}_j := \begin{bmatrix} b_j(L) & \dots & b_j(0) \\ \vdots & \ddots & \vdots \\ b_j(N_b-1) & \dots & b_j(N_b-L-1) \end{bmatrix}, \tag{2.14}$$

where $b_j(l) := b((j-1)M + l)$. Therefore, (2.13) can be rewritten as

$$\bar{\boldsymbol{x}}_b = \boldsymbol{\Phi} \boldsymbol{w} + \bar{\boldsymbol{\eta}}_b, \tag{2.15}$$

with simple substitutions, where the $[J(N_b - L)] \times [(Q+1)(L+1)]$ matrix

$$\boldsymbol{\Phi} := \begin{bmatrix} \bar{\boldsymbol{D}}_{\psi_{0,1}} \boldsymbol{B}_1 & ... & \bar{\boldsymbol{D}}_{\psi_{Q,1}} \boldsymbol{B}_1 \\ \vdots & \ddots & \vdots \\ \bar{\boldsymbol{D}}_{\psi_{0,J}} \boldsymbol{B}_J & \dots & \bar{\boldsymbol{D}}_{\psi_{Q,J}} \boldsymbol{B}_J \end{bmatrix}, \tag{2.16}$$

and

$$\boldsymbol{w} := [\boldsymbol{w}_0^T, \boldsymbol{w}_1^T, ..., \boldsymbol{w}_Q^T]^T. \tag{2.17}$$

2.3.1 Least-squares Channel Estimation

The linear least-squares (LS) channel estimator based on (2.15) is

$$\hat{\boldsymbol{w}}_{\mathrm{LS}} = \boldsymbol{\Lambda}_{\mathrm{LS}} \bar{\boldsymbol{x}}_b, \tag{2.18}$$

where $\boldsymbol{\Lambda}_{\mathrm{LS}} = (\boldsymbol{\Phi}^H \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^H$. Define the estimation error of BEM parameters as $\tilde{\boldsymbol{w}}_{\mathrm{LS}} := \boldsymbol{w} - \hat{\boldsymbol{w}}_{\mathrm{LS}}$. Then the covariance matrix of $\tilde{\boldsymbol{w}}_{\mathrm{LS}}$ is

$$\boldsymbol{R}_{\tilde{w}_{\mathrm{LS}}} := E[\tilde{\boldsymbol{w}}_{\mathrm{LS}} \tilde{\boldsymbol{w}}_{\mathrm{LS}}^H] = \sigma_\eta^2 \left(\boldsymbol{\Phi}^H \boldsymbol{\Phi}\right)^{-1}. \tag{2.19}$$

As a result, the MSE of $\tilde{\boldsymbol{w}}_{\mathrm{LS}}$ is

$$\sigma_{\tilde{w}_{\mathrm{LS}}}^2 := \mathrm{tr}(\boldsymbol{R}_{\tilde{w}_{\mathrm{LS}}}) = \sigma_\eta^2 \mathrm{tr}\left[(\boldsymbol{\Phi}^H \boldsymbol{\Phi})^{-1}\right]. \tag{2.20}$$

Using [52, Lemma 1], $\sigma^2_{\tilde{w}_{\mathrm{LS}}}$ is lower bounded by ($S := K(Q+1)(L+1)$)

$$\sigma^2_{\tilde{w}_{\mathrm{LS}}} \geq \sigma^2_\eta \sum_{k=1}^{S} \frac{1}{[(\boldsymbol{\Phi}^H\boldsymbol{\Phi})]_{k,k}} = \sigma^2_\eta \sum_{k=1}^{S} \frac{1}{[\boldsymbol{\Phi}^H\boldsymbol{\Phi}]_{k,k}} \tag{2.21}$$

where the equality holds if and only if $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ is a diagonal matrix. By the arithmetic-geometric mean inequality [31, p. 535],

$$\sigma^2_\eta \sum_{k=1}^{S} \frac{1}{[\boldsymbol{\Phi}^H\boldsymbol{\Phi}]_{k,k}} \geq \sigma^2_\eta S \left(\prod_{k=1}^{S} \frac{1}{[\boldsymbol{\Phi}^H\boldsymbol{\Phi}]_{k,k}} \right)^{1/S} \tag{2.22}$$

where the equality holds iff $\left[\boldsymbol{\Phi}^H\boldsymbol{\Phi}\right]_{k,k}$ are all equal. Equivalently, we need $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ to be a diagonal matrix with all its diagonal entries equal.

2.3.2 Linear Minimum Mean-Square-Error Channel Estimation

The linear minimum mean-square-error (MMSE) channel estimator based on (2.15) is

$$\hat{\boldsymbol{w}}_{\mathrm{MMSE}} = \boldsymbol{\Lambda}_{\mathrm{MMSE}} \bar{\boldsymbol{x}}_b, \tag{2.23}$$

where $\boldsymbol{\Lambda}_{\mathrm{MMSE}} = \sigma^2_\eta \boldsymbol{R}_w^{-1} + (\boldsymbol{\Phi}^H\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^H$. This estimator requires $\boldsymbol{R}_w := E\left[\boldsymbol{w}\boldsymbol{w}^H\right]$ to be known at the receiver. Define the estimation error of BEM parameters as $\tilde{\boldsymbol{w}}_{\mathrm{MMSE}} := \boldsymbol{w} - \hat{\boldsymbol{w}}_{\mathrm{MMSE}}$. Then the covariance matrix of $\tilde{\boldsymbol{w}}_{\mathrm{MMSE}}$ is

$$\boldsymbol{R}_{\tilde{w}_{\mathrm{MMSE}}} := E[\tilde{\boldsymbol{w}}_{\mathrm{MMSE}}\tilde{\boldsymbol{w}}^H_{\mathrm{MMSE}}] = \left[\boldsymbol{R}_w^{-1} + \frac{1}{\sigma^2_\eta}(\boldsymbol{\Phi}^H\boldsymbol{\Phi}) \right]^{-1}. \tag{2.24}$$

As a result, the MSE of $\tilde{\boldsymbol{w}}_{\mathrm{MMSE}}$ is

$$\sigma^2_{\tilde{w}_{\mathrm{MMSE}}} := \mathrm{tr}(\boldsymbol{R}_{\tilde{w}_{\mathrm{MMSE}}}) = \mathrm{tr}\left[\left(\boldsymbol{R}_w^{-1} + \frac{1}{\sigma_\eta^2}(\boldsymbol{\Phi}^H\boldsymbol{\Phi})\right)^{-1}\right]. \tag{2.25}$$

Since our analysis allows for correlated channels, the BEM coefficients $w_q(l)$ are not necessarily independent. Eigen-decomposition of $\boldsymbol{R}_w$ yields $\boldsymbol{R}_w = \boldsymbol{U}_w\boldsymbol{\Omega}_w\boldsymbol{U}_w^{-1}$ where $\boldsymbol{U}_w^{-1} = \boldsymbol{U}_w^H$. Since $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$, (2.25) can be rewritten as

$$\sigma^2_{\tilde{w}_{\mathrm{MMSE}}} = \mathrm{tr}\left[\left(\boldsymbol{\Omega}_w^{-1} + \frac{1}{\sigma_\eta^2}\boldsymbol{U}_w^{-1}(\boldsymbol{\Phi}^H\boldsymbol{\Phi})\boldsymbol{U}_w\right)^{-1}\right]. \tag{2.26}$$

Similar to (2.21), by [52, Lemma 1], $\sigma^2_{\tilde{w}_{\mathrm{MMSE}}}$ is lower bounded by

$$\sigma^2_{\tilde{w}_{\mathrm{MMSE}}} \geq \sum_i \frac{1}{\left[\boldsymbol{\Omega}_w^{-1} + \frac{1}{\sigma_\eta^2}\boldsymbol{U}_w^{-1}(\boldsymbol{\Phi}^H\boldsymbol{\Phi})\boldsymbol{U}_w\right]_{i,i}} \tag{2.27}$$

where the equality holds if and only if $\boldsymbol{U}_w^{-1}(\boldsymbol{\Phi}^H\boldsymbol{\Phi})\boldsymbol{U}_w$ is a diagonal matrix. This is true if $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ is a diagonal matrix with all its diagonal elements equal. Similar as in (2.21)-(2.22), $\sigma^2_{\tilde{w}_{\mathrm{MMSE}}}$ is lower bounded by

$$\sigma^2_{\tilde{w}_{\mathrm{MMSE}}} \geq S \times \left[\prod_i \frac{1}{\left[\boldsymbol{\Omega}_w^{-1} + \frac{1}{\sigma_\eta^2}\boldsymbol{U}_w^{-1}(\boldsymbol{\Phi}^H\boldsymbol{\Phi})\boldsymbol{U}_w\right]_{i,i}}\right]^{\frac{1}{S}} \tag{2.28}$$

where the equality holds if and only if $\boldsymbol{\Omega}_w^{-1} + \frac{1}{\sigma_\eta^2}\boldsymbol{U}_w^{-1}\left(\boldsymbol{I}_R \otimes (\boldsymbol{\Phi}^H\boldsymbol{\Phi})\right)\boldsymbol{U}_w$ is a diagonal matrix with all its diagonal entries equal. Equivalently, we need $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ to be a diagonal matrix with all its diagonal entries equal provided that diagonal $\boldsymbol{\Omega}_w$ has all its diagonal entries equal; unfortunately, the latter is not necessarily true (at least for a channel tap with Jakes' spectrum).

2.3.3 Channel Estimation Error

Given estimated BEM parameters $\hat{w}_q(l)$ via LS or MMSE estimators, the channel impulse response is then given by:

$$\hat{h}_{\mathrm{BEM}}(n;l) = \sum_{q=0}^{Q} \hat{w}_q(l)\psi_q(n). \tag{2.29}$$

There are two sources of channel estimation error: one is from the difference between $h_{\mathrm{BEM}}(n;l)$ and $\hat{h}_{\mathrm{BEM}}(n;l)$, the other is from the channel modeling error $e_{\mathrm{BEM}}(n;l)$ in (2.3). Therefore, the mean square value of channel estimation error is expressed as:

$$\sigma_{\tilde{h}}^2 = N^{-1} \sum_{n=0}^{N-1} \sum_{l=0}^{L} E\left\{ \left\| h_{\mathrm{BEM}}(n;l) - \hat{h}_{\mathrm{BEM}}(n;l) + e_{\mathrm{BEM}}(n;l) \right\|^2 \right\}. \tag{2.30}$$

Since the following orthogonality is true for both CE-BEM and DPS-BEM models

$$E\left\{ \left[h_{\mathrm{BEM}}(n;l) - \hat{h}_{\mathrm{BEM}}(n;l) \right]^H e_{\mathrm{BEM}}(n;l) \right\} = 0, \tag{2.31}$$

we have

$$\sigma_{\tilde{h}}^2 = \sigma_{\tilde{w}}^2 + \underbrace{N^{-1} \sum_{n=0}^{N-1} \sum_{l=0}^{L} E\left\{ \left[e_{\mathrm{BEM}}(n;l) \right]^H e_{\mathrm{BEM}}(n;l) \right\}}_{\sigma_{\mathrm{BEM}}^2}. \tag{2.32}$$

The channel modeling error σ_{BEM}^2 has an analytic expression from Niedzwiecki's results in [51]:

$$\sigma_{\mathrm{BEM}}^2 \approx \frac{1}{\pi} \int_0^{\pi} \Psi(n,\omega) \mathrm{tr}\left\{ S_{hh}(\omega) \right\} d\omega, \tag{2.33}$$

where $\Psi(n,\omega)$ is the instantaneous parameter matching characteristics of a basis function estimator:

$$\Psi(n,\omega) = |1 - \mathcal{H}(n,\omega)|^2, \tag{2.34}$$

$\mathcal{H}(n,\omega)$ is the instantaneous frequency response of the basis expansion estimator:

$$\mathcal{H}(n,\omega) = \boldsymbol{f}^T(n) \sum_{n'=0}^{N-1} \boldsymbol{f}(n) e^{-i\omega(n-n')},$$
$$\boldsymbol{f}(n) := [\psi_0(n), \psi_1(n), ..., \psi_Q(n)]^T, \tag{2.35}$$

the power spectral density $S_{hh}(\omega)$ is derived from the autocorrelation of $h(n)$:

$$\mathcal{R}_{hh}(\tau) := E\left[h^*(n+\tau)h(n)\right],$$
$$\mathcal{R}_{hh}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{hh}(\omega)e^{i\omega\tau}d\omega. \tag{2.36}$$

The modeling error for CE-BEM (referred to as σ^2_{CE}) and DPS-BEM (referred to as σ^2_{DPS}) models are analyzed in [77]. We have also shown some simulation results earlier in Chapter 1 (Fig. 1.1). For a 3-tap Jakes' channel, σ^2_{DPS} is several orders of magnitude smaller than σ^2_{CE}.

2.4 Optimum Training Design

Observe from (2.21) that in order to achieve the lower bound of $\sigma^2_{\tilde{w}_{\mathrm{LS}}}$, The LS estimator requires $\mathbf{\Phi}^H\mathbf{\Phi}$ to be a diagonal matrix with all its diagonal entries equal. Observe from (2.27) and (2.28) that in order to achieve the lower bound of $\sigma^2_{\tilde{w}_{\mathrm{MMSE}}}$, The MMSE estimator also requires $\mathbf{\Phi}^H\mathbf{\Phi}$ to be a diagonal matrix to achieve (2.27) and for both it and $\mathbf{\Omega}_w$ to be diagonal with all their respective diagonal entries equal to achieve (2.28). Unfortunately, in general, the diagonal $\mathbf{\Omega}_w$ does not necessarily have all its diagonal entries equal. We will design the training schemes to make $\mathbf{\Phi}^H\mathbf{\Phi}$ to be a diagonal matrix with all its diagonal entries equal which will achieve the lower bound

in (2.22) for the LS channel estimator and in (2.27) (but not in (2.28)) for the linear MMSE channel estimator.

Suppose that we choose

$$\boldsymbol{\Phi}^H\boldsymbol{\Phi} = \alpha \boldsymbol{I}_{(Q+1)(L+1)} \text{ for some } \alpha > 0 \tag{2.37}$$

Then by (2.16) we must have

$$\sum_{j=1}^{J} \boldsymbol{B}_j^H \bar{\boldsymbol{D}}_{\psi_{q1},j}^H \bar{\boldsymbol{D}}_{\psi_{q2},j} \boldsymbol{B}_j = \alpha \boldsymbol{I} \delta(q_1 - q_2). \tag{2.38}$$

For CE-BEM, it turns out that [41]

$$\sum_{j=1}^{J} \bar{\boldsymbol{D}}_{\psi_{q1},j}^H \bar{\boldsymbol{D}}_{\psi_{q2},j} = \frac{1}{M} \boldsymbol{I}_{N_b - L} \delta(q_1 - q_2) \tag{2.39}$$

Note that in [41], (2.4) is not normalized; here it is. By (2.16) and (2.38), for all k's and q's, the training sequence should be designed to satisfy

$$\sum_{j=1}^{J} \boldsymbol{B}_j^H \bar{\boldsymbol{D}}_{\psi_{q1},j}^H \bar{\boldsymbol{D}}_{\psi_{q2},j} \boldsymbol{B}_j = \gamma \sum_{j=1}^{J} \bar{\boldsymbol{D}}_{\psi_{q1},j}^H \bar{\boldsymbol{D}}_{\psi_{q2},j}. \tag{2.40}$$

Under (2.39), following [41] and [75], we pick $N_b = 2L + 1$ and $\boldsymbol{b}_j^T = [\boldsymbol{0}_L^T, b_k, \boldsymbol{0}_L^T]^T$ where $\boldsymbol{0}_L$ is a size L null column, in which case $\boldsymbol{\Phi}^H\boldsymbol{\Phi} = \frac{b^2}{M}\boldsymbol{I}$. Therefore, with $\mathcal{P}_b := Jb^2$ denoting the total training power, we can obtain $\alpha = \frac{\mathcal{P}_b}{N}$ (recall that $N = JM$).

Plugging (2.37) into (2.22), (2.28) and (2.32), the LS and MMSE lower bounds of channel estimation error are derived as

$$\sigma^2_{\tilde{h}_{\mathrm{LS}}} = \frac{N\sigma^2_\eta}{\mathcal{P}_b}(L+1)(Q_F+1) + \sigma^2_{\mathrm{CE}}, \tag{2.41}$$

$$\sigma^2_{\tilde{h}_{\mathrm{MMSE}}} = \sum_{t=1}^{(L+1)(Q_F+1)} \left[\check{\lambda}_t^{-1} + \frac{\mathcal{P}_b}{\sigma^2_\eta N}\right]^{-1} + \sigma^2_{\mathrm{CE}}, \tag{2.42}$$

where $\check{\lambda}_t$ is the t-th diagonal entry of matrix $\boldsymbol{\Omega}_w$, i.e. $\check{\lambda}_t$ is the t-th eigenvalue of $\boldsymbol{R}_w$.

For DPS-BEM, we will use "large" N approximation from the Appendix A for DPS-BEM basis functions to obtain an expression for the modeled part of channel estimation error. Using heuristic asymptotic $(A.6)$, we can easily establish (note that asymptotic $(A.6)$ corresponds to (2.4) and $(A.3)$)

$$\sum_{j=1}^{J} \bar{\boldsymbol{D}}^H_{\psi_{q1},j} \bar{\boldsymbol{D}}_{\psi_{q2},j} \approx \frac{1}{M} \boldsymbol{I}_{N_b-L} \delta(q_1 - q_2). \tag{2.43}$$

Therefore, for DPS-BEM mimicking (2.41) and (2.42), the lower bounds of channel estimation error are

$$\sigma^2_{\tilde{h}_{\mathrm{LS}}} \approx \frac{N\sigma^2_\eta}{\mathcal{P}_b}(L+1)(Q_S+1) + \sigma^2_{\mathrm{DPS}}, \tag{2.44}$$

$$\sigma^2_{\tilde{h}_{\mathrm{MMSE}}} \approx \sum_{t=1}^{(L+1)(Q_S+1)} \left[\check{\lambda}_t^{-1} + \frac{\mathcal{P}_b}{\sigma^2_\eta N}\right]^{-1} + \sigma^2_{\mathrm{DPS}}. \tag{2.45}$$

Remark 2.1 Note that (2.43) is critical for (2.44) and (2.45) to hold true and for the proposed training designs to be valid. The asymptotic Slepian sequences in $(A.6)$ are only "heuristic." But we can "verify" (2.43) by computing the results in (2.20) and (2.22) numerically and compare them with the results in (2.44). This is done in Section 2.6.

2.5 Training Power Allocation

We assume that the time-varying channel $h(n;l)$ is zero-mean, wide sense stationary (in n with fixed l) complex Gaussian with the same variance σ_h^2. We also assume that the channel taps are mutually independent, i.e. $h(n;l)$ is wide sense stationary uncorrelated scattering (WSSUS). To simplify the expressions, in this section we assume that the channel modeling error $e_{\rm BEM}$ in (2.3) is zero.

The received information symbols at the received antenna can be expressed as

$$x_c(n) = \underbrace{\sum_{l=0}^{L} \hat{h}(n;l)c(n-l)}_{:=x_s(n)} + \underbrace{\sum_{l=0}^{L} [h(n;l) - \hat{h}(n;l)]c(n-l) + \eta(n)}_{=x_\eta(n)}, \tag{2.46}$$

where $\hat{h}(n;l) = \sum_{q=0}^{Q} \hat{w}_q(l)\psi_q(n)$ is used for data detection. Therefore, the signal power is given by

$$\begin{aligned}
\sigma_{xs}^2(n) &:= E\left\{|x_s(n)|^2\right\} \\
&= E\left\{\left[\left|\sum_{l=0}^{L} \hat{h}(n;l)c(n-l)\right|^2\right]\right\} \\
&= \bar{\mathcal{P}}_c \sum_{l=0}^{L} \left[E_w\left\{E\left\{\left|h(n;l) - \hat{h}(n;l)\right|^2 |\boldsymbol{w}\right\}\right\} + E\left\{|h(n;l)|^2\right\}\right] \\
&= \bar{\mathcal{P}}_c \left[\sigma_{\tilde{h}}^2(n) + (L+1)\sigma_h^2\right],
\end{aligned} \tag{2.47}$$

where the average power of information symbols is

$$\bar{\mathcal{P}}_c := E\left\{|c(n)|^2\right\}, \tag{2.48}$$

and the effective noise power is:

$$\begin{aligned}\sigma_{x\eta}^2(n) &= E\left\{\left[\left|\sum_{l=0}^{L}[h(n;l)-\hat{h}(n;l)]c(n-l)\right|^2\right]\right\}+E\left\{|\eta(n)|^2\right\}\\ &= \bar{\mathcal{P}}_c\sigma_{\tilde{h}}^2(n)+\sigma_\eta^2,\end{aligned} \tag{2.49}$$

where $\sigma_{\tilde{h}}^2(n) := \sum_{l=0}^{L} E_{\boldsymbol{w}}\left\{E\left\{\left|h(n;l)-\hat{h}(n;l)\right|^2|\boldsymbol{w}\right\}\right\}$. Define

$$\begin{aligned}\mathcal{W}_l &:= [w_0(l), w_1(l), ..., w_Q(l)]^T,\\ \mathcal{W} &:= [\mathcal{W}_0, \mathcal{W}_1, ..., \mathcal{W}_L]^T,\end{aligned} \tag{2.50}$$

By (2.17) and (2.50), we can get the following relationship:

$$\mathcal{W} = \boldsymbol{\Xi}\boldsymbol{w}, \tag{2.51}$$

where

$$\boldsymbol{\Xi} = \begin{bmatrix} 1 & \overbrace{0...0}^{L} & \dots & & \dots & \\ 0 & 0...0 & 1 & \overbrace{0...0}^{L} & \dots & \dots \\ \dots & & \dots & & \dots & \\ 0 & 1 & \overbrace{0...0}^{L} & \dots & \dots & \dots \\ 0 & 0 & \overbrace{0...0}^{L} & 1 & \overbrace{0...0}^{L} & \dots \\ \ddots & & \ddots & & \ddots & \\ \dots & \dots & \dots & \dots & \dots & 1 \end{bmatrix}_{(Q+1)(L+1)\times(Q+1)(L+1)}. \tag{2.52}$$

We also find that $\boldsymbol{\Xi}^H\boldsymbol{\Xi}$ is always an identity matrix. Then $\sigma^2_{\tilde{h}}(n)$ can be rewritten as

$$\begin{aligned}\sigma^2_{\tilde{h}}(n) &= \sum_{l=0}^{L} E_{\mathcal{W}}\left\{E\left\{\left|h(n;l)-\hat{h}(n;l)\right|^2|\mathcal{W}\right\}\right\} \\ &= \mathrm{tr}\left\{\boldsymbol{\Psi}(n)E_{\mathcal{W}}\left\{\mathrm{cov}\{\hat{\mathcal{W}},\hat{\mathcal{W}}|\mathcal{W}\}\right\}\boldsymbol{\Psi}(n)^H\right\}, \end{aligned} \tag{2.54}$$

where

$$\begin{aligned}\boldsymbol{\psi}(n) &:= [\psi_0(n),\psi_1(n),...,\psi_Q(n)], \\ \boldsymbol{\Psi}(n) &:= \boldsymbol{I}_{(L+1)}\otimes\boldsymbol{\psi}(n).\end{aligned}$$

Based on (2.50) and (2.51), we have

$$E_{\mathcal{W}}\left\{\mathrm{cov}\{\hat{\mathcal{W}},\hat{\mathcal{W}}|\mathcal{W}\}\right\} = \boldsymbol{\Xi}\underbrace{E_{\mathcal{W}}\left\{\mathrm{cov}\{\hat{\boldsymbol{w}},\hat{\boldsymbol{w}}|\mathcal{W}\}\right\}}_{\boldsymbol{R}_{\tilde{\boldsymbol{w}}}}\boldsymbol{\Xi}^H. \tag{2.55}$$

Based on the orthonormality of $\psi_q(n)$, we have

$$\sum_{n=1}^{N}\boldsymbol{\Psi}(n)^H\boldsymbol{\Psi}(n) = \boldsymbol{I}_{(L+1)(Q+1)}. \tag{2.56}$$

Therefore, the time-averaged of $\bar{\sigma}^2_{\tilde{h}}$ over information subblocks in the current block is

$$\bar{\sigma}^2_{\tilde{h}} := (N-JN_b)^{-1}\sum_{n}\sigma^2_{\tilde{h}}(n) \approx N^{-1}\sum_{n=0}^{N-1}\sigma^2_{\tilde{h}}(n) = \frac{1}{N}\mathrm{tr}\left\{\boldsymbol{\Xi}\boldsymbol{R}_{\tilde{w}}\boldsymbol{\Xi}^H\right\} = \frac{1}{N}\sigma^2_{\tilde{w}} \tag{2.57}$$

Similarly, the time averaged signal and noise powers turn out to be

$$\bar{\sigma}_{xs}^2 := \frac{1}{N}\sum_{n=1}^{N}\sigma_{xs}^2(n) = \bar{\mathcal{P}}_c[\bar{\sigma}_{\tilde{h}}^2 + (L+1)\sigma_h^2],$$
$$\bar{\sigma}_{x\eta}^2 := \frac{1}{N}\sum_{n=1}^{N}\sigma_{x\eta}^2(n) = \bar{\mathcal{P}}_c\bar{\sigma}_{\tilde{h}}^2 + \sigma_\eta^2. \tag{2.58}$$

Therefore, we obtain an effective average SNR for (2.46) as

$$\mathrm{SNR}_d = \frac{\bar{\sigma}_{xs}^2}{\bar{\sigma}_{x\eta}^2}. \tag{2.59}$$

Define the total information power and received signal power $\mathcal{P}_c := JN_c\bar{\mathcal{P}}_c$ and $\mathcal{P} := \mathcal{P}_b + \mathcal{P}_c$, respectively. Define the training power overhead

$$\beta := \frac{\mathcal{P}_b}{\mathcal{P}_c + \mathcal{P}_b}.$$

Our objective is to maximize SNR with respect to β under the constraint of a fixed $\mathcal{P}$. Thus, incorporating those constraint-carrying variables into (2.59) and using the developed expression for average signal and noise powers in (2.59), we obtain the unconstrained cost

$$\mathrm{SNR}_d(\beta) = \frac{\frac{(1-\beta)\mathcal{P}}{JN_c}[\bar{\sigma}_{\tilde{h}}^2 + (L+1)\sigma_h^2]}{\frac{(1-\beta)\mathcal{P}}{JN_c}\bar{\sigma}_{\tilde{h}}^2 + \sigma_\eta^2}. \tag{2.60}$$

Using the lower bound of the LS estimator in (2.44) (due to the lower bound of the MMSE estimator, the closed form of the optimal β can not be obtained), we can explicitly write (2.59) as

$$\mathrm{SNR}_d(\beta) = \frac{f_1\beta^2 + f_2\beta + f_3}{g_1\beta + g_2}, \tag{2.61}$$

where

$$
\begin{aligned}
f_1 &= -\frac{\mathcal{P}(L+1)\sigma_h^2}{N_c J}, \\
f_2 &= \frac{\mathcal{P}(L+1)\sigma_h^2}{N_c J} - \frac{(L+1)(Q+1)\sigma_\eta^2}{N_c J}, \\
f_3 &= \frac{(L+1)(Q+1)\sigma_\eta^2}{N_c J}, \\
g_1 &= \sigma_\eta^2 - \frac{(L+1)(Q+1)\sigma_\eta^2}{N_c J}, \\
g_2 &= \frac{(L+1)(Q+1)\sigma_\eta^2}{N_c J}.
\end{aligned} \tag{2.62}
$$

Setting the first derivative of $\mathrm{SNR}_d(\beta)$ with respect to β to zero, we obtain a quadratic equation in β

$$
\beta^2 + 2\frac{g_2}{g_1}\beta + \frac{f_2 g_2 - f_3 g_1}{f_1 g_1} = 0 \tag{2.63}
$$

with two roots, one of which is negative ($\beta < 0$), and hence is excluded. The other root is given by

$$
\beta_{\mathrm{opt}} = \frac{g_2}{g_1}\left[-1 + \sqrt{1 + \frac{g_1(f_3 g_1 - f_2 g_2)}{g_2^2 f_1}}\right]. \tag{2.64}
$$

2.6 Numerical Examples

In the following examples, we use binary phase shift keying (BPSK) and quadrature phase shift keying (QPSK) modulation. Each transmitted block has $J = 10$ subblocks, and each subblock has $N_c = 30$ information symbols and $N_b = 2L + 1$ training symbols with optimal structure $[\mathbf{0}_L, b, \mathbf{0}_L]$, $b > 0$. A doubly-selective Rayleigh fading channel $h(n; l)$ is simulated according to [77, 81] with the channel order $L = 2$, carrier frequency of 2GHz, data rate of 40 kbps, and thus, symbol duration $T_s = 25\mu s$. Therefore, each tap of the generated time-variant channel has a Jakes'

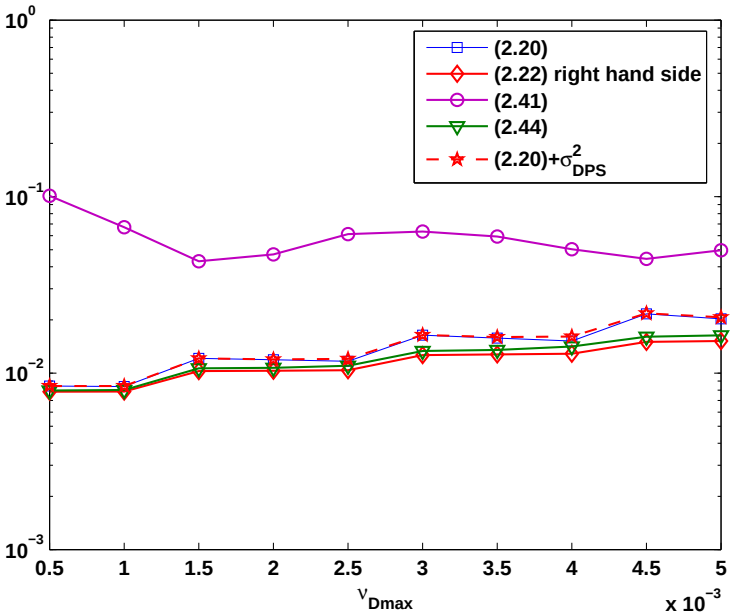

Figure 2.1: Channel estimation errors (with and without modeling errors):Numerical results derived from $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$, SNR=20dB

spectrum; it is not generated using the assumed BEM modeling. Also, 3 taps of the channel are mutually independent. Depending on different maximum Doppler spread f_ds, a varying maximum normalized one-sided Doppler bandwidth $\nu_{\rm Dmax} = f_d T_s$ can be derived. Given $\nu_{\rm Dmax}$, Jakes' spectrum and other information, we can calculate $\boldsymbol{R}_w$ and therefore, $\boldsymbol{\Omega}_w$ and $\check{\lambda}_t$s, needed in (2.28), (2.45) and elsewhere. The SNR refers to the ratio of total signal and training power to the total noise power, each per block.

2.6.1 Example 1: Approximation Errors

As noted earlier in Remark 2.1, here we want to show the influence of approximation errors in $\bar{\boldsymbol{D}}_{\psi_q,j}$ when the true Slepian sequences instead of the approximations in $(A.6)$ are used. We compute (2.20) and (2.22) numerically with true Slepian sequences generated by (2.5), then compare them

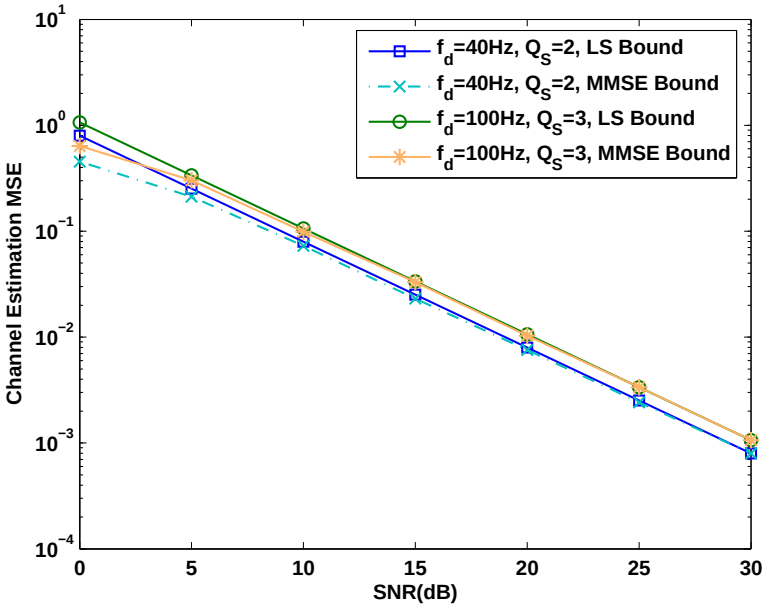

Figure 2.2: MSE lower bound comparison between LS (2.44) and MMSE (2.45) estimators

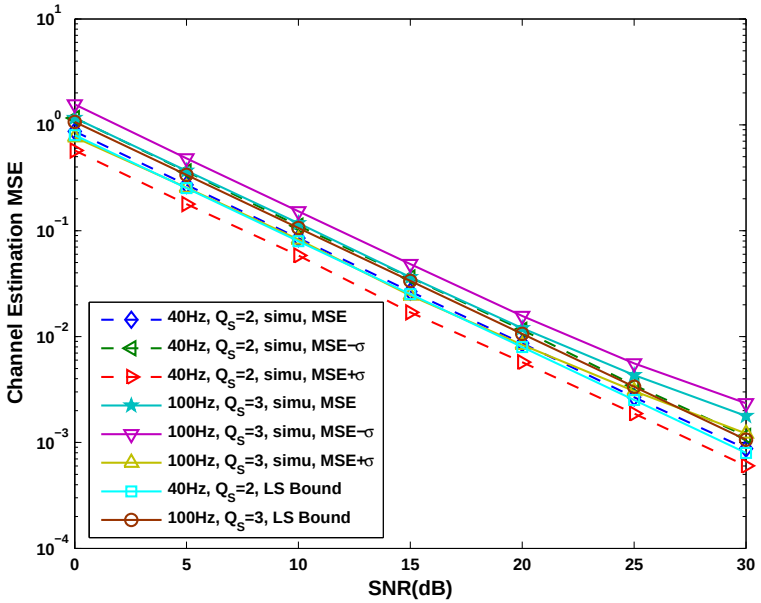

Figure 2.3: Comparison between the LS channel estimation MSE lower bound in (2.44) and simulation results in (2.65), $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1$

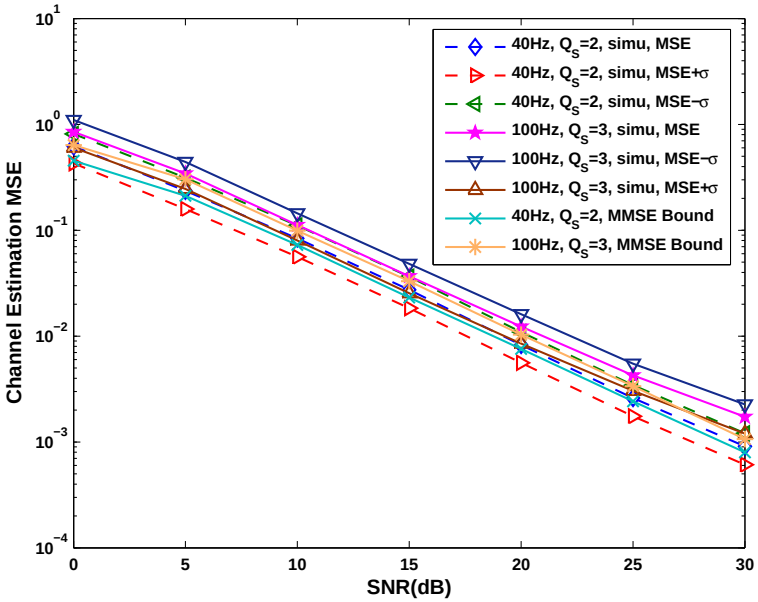

Figure 2.4: Comparison between the MMSE channel estimation MSE lower bound in (2.45) and simulation results in (2.65), $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1$

with (2.41) and (2.44). For the parameters stated earlier in this section (J=10, N_b=8, N_c=30, $M = N_b + N_c$=38, $N = JM$ =380) the results are shown in Fig. 2.1, where SNR=20dB and the dimensions $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$ change with the maximum Doppler bandwidth. [Here the minimum $Q_F = 2\lceil \nu_{Dmax} N \rceil$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil$ are not taken since a slightly higher values of Q's can significantly reduce the modeling error. This is suggested in [77]]. From Fig. 2.1 we see that the CE-BEM lower bound in (2.41) is high due to "large" modeling error σ^2_{CE} even though the error for the modeled part is minimum, whereas the results in (2.20), (2.22) and (2.44) are close to each other due to "small" modeling error σ^2_{DPS} and "close-to- minimum" (compare curves for (2.20) (exact error) and (2.22) (lower bound)) error for the modeled part. [We note that σ^2_{DPS} and σ^2_{CE} were obtained via Monte Carlo averaging, similar to [77].]

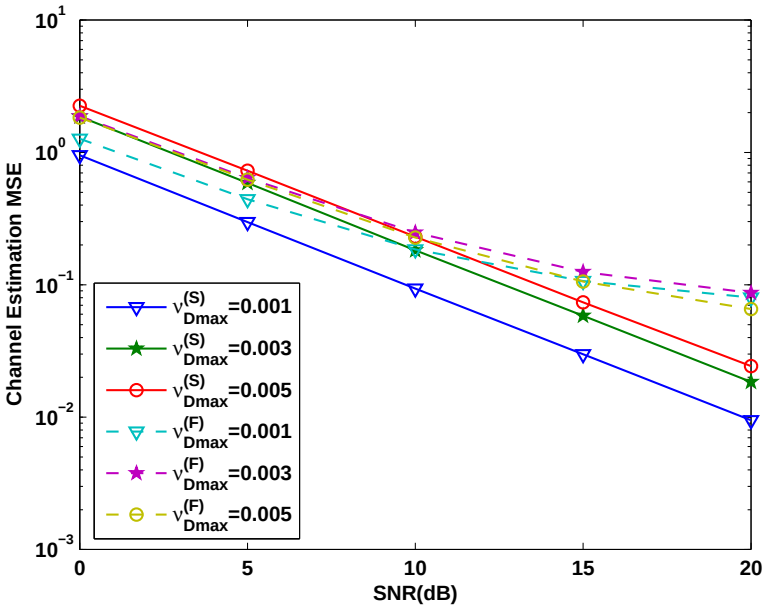

Figure 2.5: LS channel estimation MSE with varying maximum normalized Doppler bandwidth, $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1$,$Q_F = 2\lceil \nu_{\mathrm{Dmax}} N \rceil + 1$

2.6.2 Example 2: DPS-BEM Channel Estimation Performance

In this case we pick $b = 1$ in training. The LS and MMSE estimators are used to estimate $\boldsymbol{w}$, and then the channel is estimated as in (2.29). Based on M_r Monte Carlo runs, the channel estimation MSE is calculated as (h_k corresponds to the kth run)

$$\mathrm{MSE} = (M_r N)^{-1} \sum_{k=1}^{M_r} \sum_{n=0}^{N-1} \sum_{l=0}^{L} |\hat{h}_k(n;l) - h_k(n;l)|^2. \tag{2.65}$$

In Fig. 2.2, the lower bounds in (2.44) and (2.45) are plotted using $Q_S = 2$ for $f_d = 40$Hz and using $Q_S = 3$ for $f_d = 100$Hz. [The MMSE bound needs $\check{\lambda}_t$s which requires knowledge of the Doppler spread. Note also that for both $f_d = 40$Hz and $f_d = 100$Hz, the minimum $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil$ are actually $Q_S = 1$ and $Q_S = 2$, respectively; however, a slightly higher value of

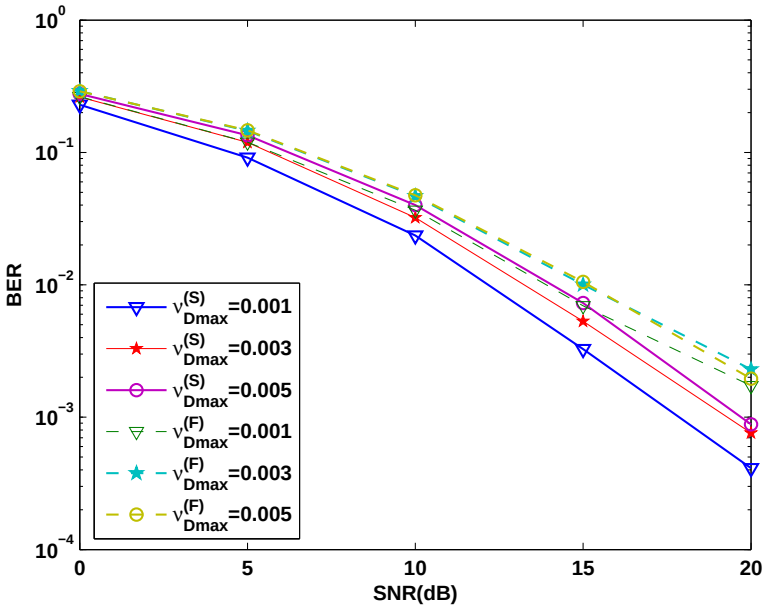

Figure 2.6: BER with varying maximum normalized Doppler bandwidth using BPSK modulation, $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1$,$Q_F = 2\lceil \nu_{\mathrm{Dmax}} N \rceil + 1$

$Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1 = 3$ for $f_d = 100$Hz yields the smaller modeling error, which becomes significant at high SNRs]. Then they are compared with the simulation results (averaged over 200 Monte Carlo runs) in Figs. 2.3 and 2.4 for $f_d = 40$Hz and 100Hz; also shown are $\pm\sigma$ bounds on the simulation averages. It can be seen that the theoretical results are consistent with the simulation results for f_d =40Hz indicating that the optimal pilot design does minimize the channel MSE when using DPS-BEM. For f_d=100 Hz, there is a "small gap" between theory and simulations at high SNRs which is probably attributable to modeling error ("small" but nonzero). Furthermore, the performance of the MMSE estimator outperforms that of the LS estimator at low SNR, and they converge to each other at high SNR.

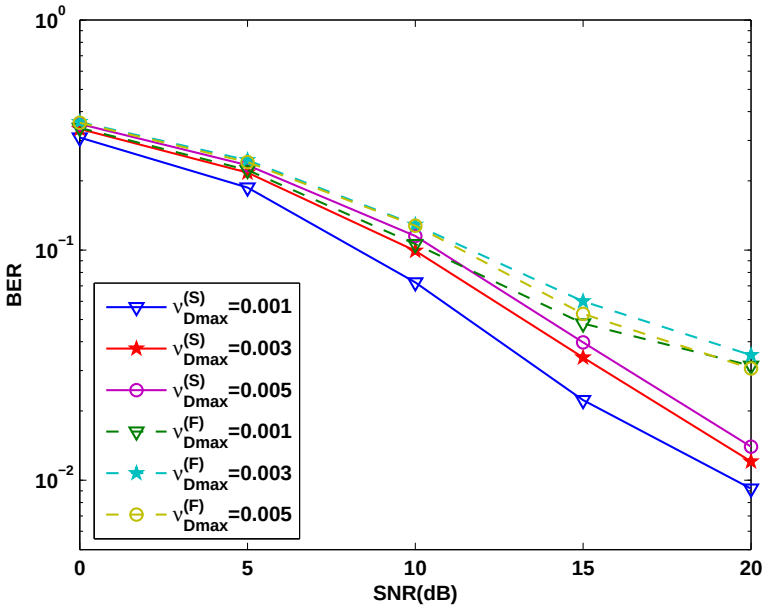

Figure 2.7: BER with varying maximum normalized Doppler bandwidth using QPSK modulation, $Q_S = \lceil 2\nu_{\mathrm{Dmax}} N \rceil + 1$,$Q_F = 2\lceil \nu_{\mathrm{Dmax}} N \rceil + 1$

2.6.3 Example 3: CE-BEM versus DPS-BEM

In Fig. 2.5, the LS channel estimation MSE (2.65) versus SNR under different maximum Doppler bandwidths ($\nu^{(S)}_{\mathrm{Dmax}}$ is for DPS-BEM, $\nu^{(F)}_{\mathrm{Dmax}}$ is for CE-BEM) are plotted. It is clear that the MSE of DPS-BEM is consistently smaller than that of CE-BEM. Fig. 2.6 takes BER (average over 2000 Monte Carlo runs) as a performance measure to compare the performance between DPS-BEM and CE-BEM. A Kalman filter formulation is used for information detection after the channel estimation. Comparing with Fig. 2.5 makes it quite clear that the significantly reduced MSE for the DPS-BEM channel estimation leads to a pronounced reduction in BER compared to the CE-BEM case.

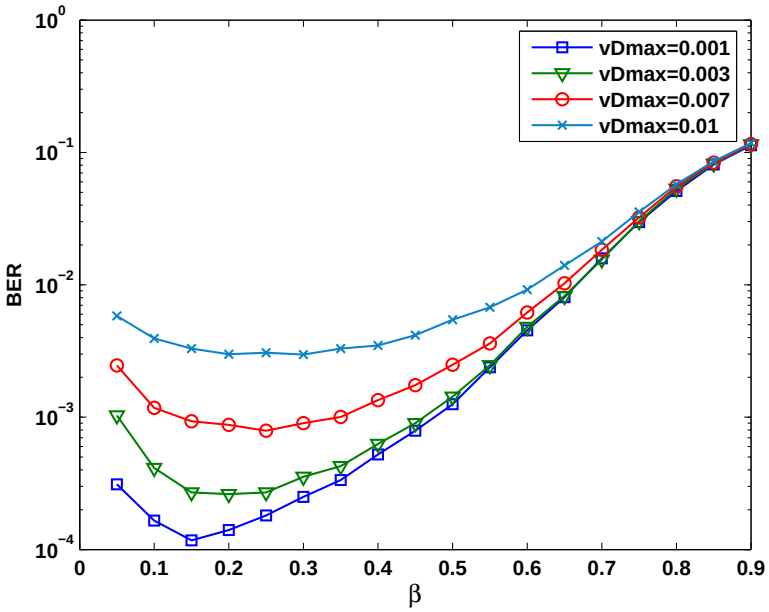

Figure 2.8: Simulations-based BER versus β for SNR=15dB

The plotting in Fig. 2.7 is exactly the same as Fig. 2.6 except that the information sequences are QPSK signals. It is seen from Figs. 2.6 and 2.7 that the BER by using QPSK information symbols are worse than that using BPSK symbols, as expected.

2.6.4 Example 4. Training Power Allocation

Here we vary training power (by varying b) with fixed total transmitted power. The BER versus optimum β based on simulation results (averaged over 1000 Monte Carlo runs) is shown in Fig. 2.8 for SNR=15dB where we used BPSK modulation and a Viterbi detector based on the estimated channel for data detection. We also varied ν_{Dmax} according to different maximum Doppler spread f_d. In Fig. 2.9, we plot the optimum theoretical values of β (derived in (2.64)) versus the received signal SNR. Comparing Figs. 2.8 and 2.9, we see that the two show mutually consistent results

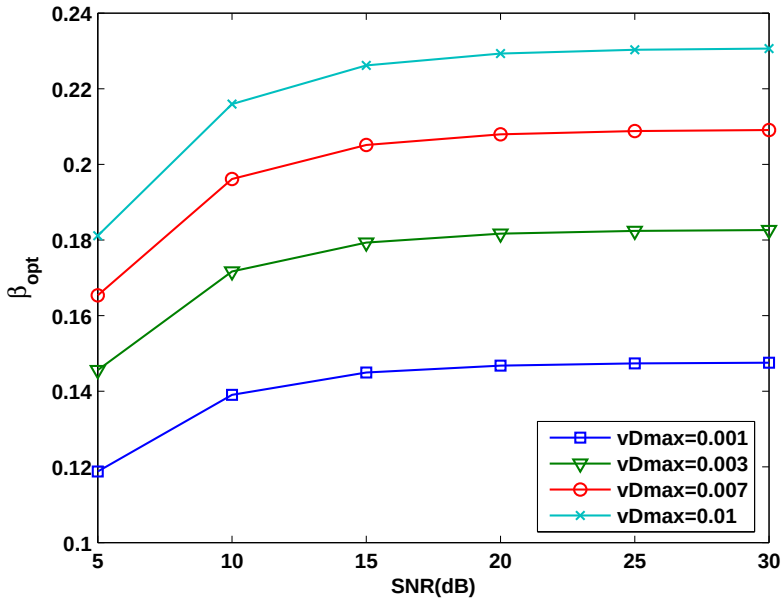

Figure 2.9: Theoretical β_{opt} (2.64) versus received signal SNR

supporting our theoretical results: the optimal (simulations based) β inferred from Fig. 2.8 is in good agreement with the theoretical β_{opt} of Fig. 2.9.

2.7 Conclusion

The channel estimation for doubly-selective channels was considered using time-multiplexed training. The time-varying channel was assumed to be well-described by a basis expansion model using discrete prolate spheroidal sequences as bases. Training designs for time-multiplexed training based on minimization of the channel estimation mean-square-error were investigated. Both least squares and minimum mean-square-error approaches were exploited to estimate the basis expansion coefficients. Computer simulation examples were presented where the channel was generated via

Jakes' models with different Doppler spreads. In these examples the DPS-BEM model significantly outperforms the more widely used complex exponential basis expansion model.

CHAPTER 3

MIMO DOUBLY-SELECTIVE CHANNEL ESTIMATION USING DISCRETE PROLATE SPHEROIDAL BASIS EXPANSION MODELS AND TIME-MULTIPLEXED TRAINING

3.1 Introduction

The prospect of extraordinary improvements in the capacity of wireless networks has drawn considerable attention to multiple-input multiple-output (MIMO) communication techniques. MIMO methods employ multiple transmitter and receiver antennas to increase the data rate and to achieve spatial diversity. Traditionally, multiple antennas have been used on the receiver side to combat the multipath fading. The receive antennas see independently faded versions of the same signal. The receiver combines these signals so that the resultant signal exhibits considerably reduced amplitude variability (fading) in comparison with the signal at any one antenna. This is called diversity gain. Diversity is characterized by the number of independently fading branches, also known as the diversity order and is equal to the number of receive antennas in single-input-multiple output (SIMO) channels. However, recent advances have shown that using multiple antennas at both the transmitter and the receiver can significantly increase the data rate and improve the performance [22, 24]. By employing multiple transmitter antennas, multiple spatial channels are supported in the same frequency band, thus data can be transmitted in parallel which results in an increased data rate.

In this chapter, the approaches proposed in Chapter 2 are extended to the MIMO system. We consider the problem of channel estimation for the doubly-selective MIMO channels described by DPS-BEM. Time-multiplexed training is used to estimate the channel. Illustrative simulation examples are provided to show the performance of channel estimation and data detection under a MIMO system.

3.2 Multiuser MIMO Channels

Consider a multiuser channel with K users and R receive antennas leading to a MIMO system with K inputs and R outputs. Let $h^{(r,k)}(n;l)$ denote the symbol-rate impulse response (the channel response at time n to a unit input at time $n-l$) of the doubly-selective MIMO FIR linear channels between the kth user's transmit antenna and the rth receive antenna, where $n \in [0,1,...,N-1]$ and $l \in [0,L]$ capture the time- and frequency- selectivity of the channel, respectively. In a general basis expansion representation over a time-block, the following is always true:

$$h^{(r,k)}(n;l) = \sum_{q=0}^{N-1} w_q^{(r,k)}(l)\psi_q(n), \tag{3.1}$$

where $\psi_q(n)$ is the q-th basis function and the basis expansion coefficient $w_q^{(r,k)}(l)$ is fixed over the data block. As the above representation is not parsimonious, the following basis expansion model is used to approximate model (3.1):

$$h_{\mathrm{BEM}}^{(r,k)}(n;l) = \sum_{q=0}^{Q} w_q^{(r,k)}(l)\psi_q(n), \tag{3.2}$$

where only $Q+1 \ll N$ basis functions are involved. Hence, the channel modeling error between $h^{(r,k)}(n;l)$ and $h_{\mathrm{BEM}}^{(r,k)}(n;l)$ can be expressed as:

$$e_{\mathrm{BEM}}^{(r,k)}(n;l) := h^{(r,k)}(n;l) - h_{\mathrm{BEM}}^{(r,k)}(n;l) = \sum_{q=Q+1}^{N-1} w_q^{(r,k)}(l)\psi_q(n). \tag{3.3}$$

Let $\{s_k(n)\}$ denote the kth transmitter's information sequence. The received sequence at the rth receive antenna can be written as

$$x^{(r)}(n) = \sum_{k=1}^{K} \sum_{l=0}^{L} h^{(r,k)}(n;l) s_k(n-l) + \eta^{(r)}(n), \tag{3.4}$$

where $\eta^{(r)}(n)$ is the additive complex Gaussian noise at the rth receive antenna, with zero-mean and variance σ_η^2. Plugging (3.2) into (3.4), we can rewrite $x^{(r)}(n)$ as:

$$x^{(r)}(n) = \sum_{k=1}^{K} \sum_{q=0}^{Q} \psi_q(n) \left[\sum_{l=0}^{L} w_q^{(r,k)}(l) s_k(n-l) \right] + \eta^{(r)}(n). \tag{3.5}$$

We consider block transmission as in [41], where transmitted symbols for the kth transmitter are collected into $N \times 1$ blocks with $\boldsymbol{s}_k = [s_k(0), s_k(1), ..., s_k(N-1)]^T$ as the 0th block and received $x^{(r)}(n)$'s are also collect into blocks with $\boldsymbol{x}^{(r)} = [x^{(r)}(0), x^{(r)}(1), ..., x^{(r)}(N-1)]^T$ as 0th block. To avoid inter-block interference (IBI), as in [41], L guard zeros are inserted in each block at the transmitter. Then the matrix-vector input-output relationship of (3.5) is given by

$$\boldsymbol{x}^{(r)} = \sum_{k=1}^{K} \sum_{q=0}^{Q} \boldsymbol{D}_{\psi_q} \boldsymbol{W}_q^{(r,k)} \boldsymbol{s}_k + \boldsymbol{\eta}^{(r)}, \tag{3.6}$$

where $\boldsymbol{\eta}^{(r)}$ is defined similarly to $\boldsymbol{x}^{(r)}$, $\boldsymbol{D}_{\psi_q} = \mathrm{diag}[\boldsymbol{\psi}_q]$ with $\boldsymbol{\psi}_q := [\psi_q(0), \psi_q(1), ..., \psi_q(N-1)]$, and $\boldsymbol{W}_q^{(r,k)}$s are $N \times N$ lower triangular Toeplitz matrices with 1st column $\left[w_q^{(r,k)}(0), w_q^{(r,k)}(1), \ldots, w_q^{(r,k)}(L), 0, \ldots, 0 \right]^T$.

3.3 Channel Estimation

As in [41], each transmitted block $\boldsymbol{s}_k$ consists of J segments (sub-blocks) of training and information symbols $b_k(n)$ and $c_k(n)$, respectively, and each segment has the same length. Then the general structure of $\boldsymbol{s}_k$ is

$$\boldsymbol{s}_k = [\boldsymbol{b}_{k,1}^T, \boldsymbol{c}_{k,1}^T, ..., \boldsymbol{b}_{k,J}^T, \boldsymbol{c}_{k,J}^T]^T, \tag{3.7}$$

where $\boldsymbol{b}_{k,j}$ with length N_b and $\boldsymbol{c}_{k,j}$ with length N_c, for all $j \in [1, J]$, denote training and information symbol sub-blocks, respectively. Therefore, $N = J(N_b + N_c)$ with $N_b > L$. Let $M = N_b + N_c$ denote the sub-block size. Obviously, the first L symbols in the "training part" of the jth subblock of the received signal are contaminated by information symbols in the preceding $(j-1)$th subblock. In the similar way, the first L symbols in the "information part" of the jth subblock of the received signal are also contaminated by the last L training symbols in the current jth subblock. In order to avoid the inter-subblock interference (ISBI) so that the channel estimation is decoupled from data detection, we will choose the first and the last L symbols in each training subblock to be zeros, as in [41].

Define $\boldsymbol{b}_{k,j} := [b_k((j-1)M), b_k((j-1)M+1), ..., b_k((j-1)M+N_b-1)]^T$. Further define $\bar{\boldsymbol{D}}_{\psi_{q,j}} = \mathrm{diag}[\bar{\boldsymbol{\psi}}_{q,j}]$ where $\bar{\boldsymbol{\psi}}_{q,j} = [\psi_q((j-1)M+L), \psi_q((j-1)M+L+1), ..., \psi_q((j-1)M+N_b-1)]^T$. Then the ISBI free received subblock can be written as

$$\bar{\boldsymbol{x}}_{b,j}^{(r)} = \sum_{k=1}^{K} \sum_{q=0}^{Q} \bar{\boldsymbol{D}}_{\psi_{q,j}} \bar{\boldsymbol{W}}_q^{(r,k)} \boldsymbol{b}_{k,j} + \bar{\boldsymbol{\eta}}_{b,j}^{(r)}, \tag{3.8}$$

where $\bar{\boldsymbol{x}}_{b,j}^{(r)} := [x_b((j-1)M+L), x_b((j-1)M+L+1), ..., x_b((j-1)M+N_b-1)]^T$, $\bar{\boldsymbol{\eta}}_{b,j}$ is defined similarly, and $(N_b - L) \times N_b$ matrix $\bar{\boldsymbol{W}}_q^{(r,k)}$ is given by

$$\bar{\boldsymbol{W}}_q^{(r,k)} = \begin{bmatrix} w_q^{(r,k)}(L) & \dots & w_q^{(r,k)}(0) & & \\ & \ddots & \vdots & \ddots & \\ & & w_q^{(r,k)}(L) & \dots & w_q^{(r,k)}(0) \end{bmatrix}.$$

Gathering training symbols per block, we obtain

$$\bar{\boldsymbol{x}}_b^{(r)} = \sum_{k=1}^{K} \sum_{q=0}^{Q} \begin{bmatrix} \bar{\boldsymbol{D}}_{\psi_{q,1}} \bar{\boldsymbol{W}}_q^{(r,k)} \boldsymbol{b}_{k,1} \\ \vdots \\ \bar{\boldsymbol{D}}_{\psi_{q,J}} \bar{\boldsymbol{W}}_q^{(r,k)} \boldsymbol{b}_{k,J} \end{bmatrix} + \bar{\boldsymbol{\eta}}_b^{(r)}. \tag{3.9}$$

According to the commutativity property of convolution, we have $\bar{\boldsymbol{W}}_q^{(r,k)} \boldsymbol{b}_{k,j} = \boldsymbol{B}_{k,j} \boldsymbol{w}_q^{(r,k)}$ with $\boldsymbol{w}_q^{(r,k)} := [w_q^{(r,k)}(0), ..., w_q^{(r,k)}(L)]^T$ and $\boldsymbol{B}_{k,j}$ a $(N_b - L) \times (L+1)$ Toeplitz matrix given by

$$\boldsymbol{B}_{k,j} := \begin{bmatrix} b_{k,j}(L) & \dots & b_{k,j}(0) \\ \vdots & \ddots & \vdots \\ b_{k,j}(N_b - 1) & \dots & b_{k,j}(N_b - L - 1) \end{bmatrix}, \tag{3.10}$$

where $b_{k,j}(l) := b_k((j-1)M + l)$. Therefore, (4.9) can be rewritten as

$$\bar{\boldsymbol{x}}_b^{(r)} = \sum_{k=1}^{K} \boldsymbol{\Phi}_k \boldsymbol{w}^{(r,k)} + \bar{\boldsymbol{\eta}}_b^{(r)} = \boldsymbol{\Phi} \boldsymbol{w}^{(r)} + \bar{\boldsymbol{\eta}}_b^{(r)}, \tag{3.11}$$

with simple substitutions, where the $[J(N_b - L)] \times [(Q+1)(L+1)]$ matrix

$$\boldsymbol{\Phi}_k := \begin{bmatrix} \bar{\boldsymbol{D}}_{\psi_{0,1}} \boldsymbol{B}_{k,1} & ... & \bar{\boldsymbol{D}}_{\psi_{Q,1}} \boldsymbol{B}_{k,1} \\ \vdots & \ddots & \vdots \\ \bar{\boldsymbol{D}}_{\psi_{0,J}} \boldsymbol{B}_{k,J} & \dots & \bar{\boldsymbol{D}}_{\psi_{Q,J}} \boldsymbol{B}_{k,J} \end{bmatrix},$$

$$\boldsymbol{\Phi} := [\boldsymbol{\Phi}_1, ..., \boldsymbol{\Phi}_K] \tag{3.12}$$

and

$$\boldsymbol{w}^{(r,k)} := \begin{bmatrix} \boldsymbol{w}_0^{(r,k)} \\ \vdots \\ \boldsymbol{w}_Q^{(r,k)} \end{bmatrix}, \quad \boldsymbol{w}^{(r)} := \begin{bmatrix} \boldsymbol{w}^{(r,1)} \\ \vdots \\ \boldsymbol{w}^{(r,K)} \end{bmatrix}. \tag{3.13}$$

Since the matrix $\boldsymbol{\Phi}$ is common for all receivers, after collecting all $\bar{\boldsymbol{x}}_b^{(r)}$ and $\boldsymbol{w}^{(r)}$ for different r, we get

$$\boldsymbol{x}_b = (\boldsymbol{I}_R \otimes \boldsymbol{\Phi})\boldsymbol{w} + \boldsymbol{\eta}_b, \tag{3.14}$$

where $\boldsymbol{w}$ is defined as

$$\boldsymbol{w} := \begin{bmatrix} \boldsymbol{w}^{(1)T} & \dots & \boldsymbol{w}^{(R)T} \end{bmatrix}^T, \tag{3.15}$$

$\boldsymbol{x}_b$ and $\boldsymbol{\eta}_b$ are defined similarly.

3.3.1 Linear Least-Squares Channel Estimator

The linear least-squares (LS) channel estimator based on (3.14) is

$$\hat{\boldsymbol{w}}_{\mathrm{LS}} = \boldsymbol{\Lambda}_{\mathrm{LS}} \boldsymbol{x}_b, \tag{3.16}$$

where $\mathbf{\Lambda}_{\mathrm{LS}} = (\boldsymbol{I}_R \otimes (\mathbf{\Phi}^H\mathbf{\Phi}))^{-1}(\boldsymbol{I}_R \otimes \mathbf{\Phi}^H)$. Define the estimation error of BEM parameters as $\tilde{\boldsymbol{w}}_{\mathrm{LS}} := \boldsymbol{w} - \hat{\boldsymbol{w}}_{\mathrm{LS}}$, then the covariance matrix of $\tilde{\boldsymbol{w}}_{\mathrm{LS}}$ is

$$\boldsymbol{R}_{\tilde{w}_{\mathrm{LS}}} := E[\tilde{\boldsymbol{w}}_{\mathrm{LS}}\tilde{\boldsymbol{w}}_{\mathrm{LS}}^H] = \sigma_\eta^2 \left(\boldsymbol{I}_R \otimes (\mathbf{\Phi}^H\mathbf{\Phi})\right)^{-1}. \tag{3.17}$$

As a result, the MSE of $\tilde{\boldsymbol{w}}_{\mathrm{LS}}$ is

$$\sigma^2_{\tilde{w}_{\mathrm{LS}}} := \mathrm{tr}(\boldsymbol{R}_{\tilde{w}_{\mathrm{LS}}}) = \sigma_\eta^2 \mathrm{tr}\left[\left(\boldsymbol{I}_R \otimes (\mathbf{\Phi}^H\mathbf{\Phi})\right)^{-1}\right]. \tag{3.18}$$

Using [52, Lemma 1], $\sigma^2_{\tilde{w}_{\mathrm{LS}}}$ is lower bounded by ($S := K(Q+1)(L+1)$)

$$\sigma^2_{\tilde{w}_{\mathrm{LS}}} \geq \sigma_\eta^2 \sum_{i=1}^{RS} \frac{1}{[\boldsymbol{I}_R \otimes (\mathbf{\Phi}^H\mathbf{\Phi})]_{i,i}} = R\sigma_\eta^2 \sum_{i=1}^{S} \frac{1}{[\mathbf{\Phi}^H\mathbf{\Phi}]_{i,i}} \tag{3.19}$$

where the equality holds if and only if $\mathbf{\Phi}^H\mathbf{\Phi}$ is a diagonal matrix. By the arithmetic-geometric mean inequality [33, p. 535],

$$R\sigma_\eta^2 \sum_{i=1}^{S} \frac{1}{[\mathbf{\Phi}^H\mathbf{\Phi}]_{i,i}} \geq R\sigma_\eta^2 S \left(\prod_{i=1}^{S} \frac{1}{[\mathbf{\Phi}^H\mathbf{\Phi}]_{i,i}}\right)^{1/S} \tag{3.20}$$

where the equality holds iff $\left[\mathbf{\Phi}^H\mathbf{\Phi}\right]_{i,i}$ are all equal. Equivalently, we need $\mathbf{\Phi}^H\mathbf{\Phi}$ to be a diagonal matrix with all its diagonal entries equal.

3.3.2 Linear MMSE Channel Estimator

The linear minimum mean-square-error (MMSE) channel estimator based on (3.14) is

$$\hat{\boldsymbol{w}}_{\mathrm{MMSE}} = \mathbf{\Lambda}_{\mathrm{MMSE}}\boldsymbol{x}_b, \tag{3.21}$$

where $\boldsymbol{\Lambda}_{\mathrm{MMSE}} = \left(\sigma_\eta^2 \boldsymbol{R}_w^{-1} + (\boldsymbol{I}_R \otimes (\boldsymbol{\Phi}^H \boldsymbol{\Phi}))\right)^{-1} (\boldsymbol{I}_R \otimes \boldsymbol{\Phi}^H)$. This estimator requires $\boldsymbol{R}_w := E\left[\boldsymbol{w}\boldsymbol{w}^H\right]$ to be known at the receiver. Define the estimation error of BEM parameters as $\tilde{\boldsymbol{w}}_{\mathrm{MMSE}} := \boldsymbol{w} - \hat{\boldsymbol{w}}_{\mathrm{MMSE}}$. Then the covariance matrix of $\tilde{\boldsymbol{w}}_{\mathrm{MMSE}}$ is

$$\begin{aligned} \boldsymbol{R}_{\tilde{w}_{\mathrm{MMSE}}} &:= E[\tilde{\boldsymbol{w}}_{\mathrm{MMSE}} \tilde{\boldsymbol{w}}_{\mathrm{MMSE}}^H] \\ &= \left[\boldsymbol{R}_w^{-1} + \frac{1}{\sigma_\eta^2}\left(\boldsymbol{I}_R \otimes (\boldsymbol{\Phi}^H \boldsymbol{\Phi})\right)\right]^{-1}. \end{aligned} \tag{3.22}$$

As a result, the MSE of $\tilde{\boldsymbol{w}}_{\mathrm{MMSE}}$ is

$$\begin{aligned} \sigma_{\tilde{w}_{\mathrm{MMSE}}}^2 &:= \mathrm{tr}(\boldsymbol{R}_{\tilde{w}_{\mathrm{MMSE}}}) \\ &= \mathrm{tr}\left[\left(\boldsymbol{R}_w^{-1} + \frac{1}{\sigma_\eta^2}\left(\boldsymbol{I}_R \otimes (\boldsymbol{\Phi}^H \boldsymbol{\Phi})\right)\right)^{-1}\right] \end{aligned} \tag{3.23}$$

Since our analysis allows for correlated channels, the BEM coefficients $w_q^{(r,k)}(l)$ are not necessarily independent. Eigen-decomposition of $\boldsymbol{R}_w$ yields $\boldsymbol{R}_w = \boldsymbol{U}_w \boldsymbol{\Omega}_w \boldsymbol{U}_w^{-1}$ where $\boldsymbol{U}_w^{-1} = \boldsymbol{U}_w^H$. Since $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$, (3.23) can be rewritten as

$$\sigma_{\tilde{w}_{\mathrm{MMSE}}}^2 = \mathrm{tr}\left[\left(\boldsymbol{\Omega}_w^{-1} + \frac{1}{\sigma_\eta^2}\boldsymbol{U}_w^{-1}\left(\boldsymbol{I}_R \otimes (\boldsymbol{\Phi}^H \boldsymbol{\Phi})\right)\boldsymbol{U}_w\right)^{-1}\right]. \tag{3.24}$$

Similar as in (3.19), by [52, Lemma 1], $\sigma_{\tilde{w}_{\mathrm{MMSE}}}^2$ is lower bounded by

$$\sigma_{\tilde{w}_{\mathrm{MMSE}}}^2 \geq \sum_i \frac{1}{\left[\boldsymbol{\Omega}_w^{-1} + \frac{1}{\sigma_\eta^2}\boldsymbol{U}_w^{-1}\left(\boldsymbol{I}_R \otimes (\boldsymbol{\Phi}^H \boldsymbol{\Phi})\right)\boldsymbol{U}_w\right]_{i,i}} \tag{3.25}$$

where the equality holds if and only if $\boldsymbol{U}_w^{-1}\left(\boldsymbol{I}_R \otimes (\boldsymbol{\Phi}^H \boldsymbol{\Phi})\right)\boldsymbol{U}_w$ is a diagonal matrix. This is true if $\boldsymbol{\Phi}^H \boldsymbol{\Phi}$ is a diagonal matrix with all its diagonal elements equal. Similar to (3.19)-(3.20), $\sigma_{\tilde{w}_{\mathrm{MMSE}}}^2$ is

lower bounded by

$$\sigma^2_{\tilde{w}_{\mathrm{MMSE}}} \geq (RS) \times \left[\prod_i \frac{1}{\left[\boldsymbol{\Omega}_w^{-1} + \frac{1}{\sigma_\eta^2} \boldsymbol{U}_w^{-1} \left(\boldsymbol{I}_R \otimes (\boldsymbol{\Phi}^H \boldsymbol{\Phi}) \right) \boldsymbol{U}_w \right]_{i,i}} \right]^{\frac{1}{RS}} \tag{3.26}$$

where the equality holds if and only if $\boldsymbol{\Omega}_w^{-1} + \frac{1}{\sigma_\eta^2} \boldsymbol{U}_w^{-1} \left(\boldsymbol{I}_R \otimes (\boldsymbol{\Phi}^H \boldsymbol{\Phi}) \right) \boldsymbol{U}_w$ is a diagonal matrix with all its diagonal entries equal. Equivalently, we need $\boldsymbol{\Phi}^H \boldsymbol{\Phi}$ to be a diagonal matrix with all its diagonal entries equal provided that diagonal $\boldsymbol{\Omega}_w$ has all its diagonal entries equal; unfortunately, the latter is not necessarily true (at least for a channel tap with Jakes' spectrum).

3.3.3 Channel Estimation Error

After deriving the estimated BEM parameters $\hat{w}_q^{(r,k)}(l)$ by the LS or MMSE estimators, the channel impulse response is then given by:

$$\hat{h}_{\mathrm{BEM}}^{(r,k)}(n;l) = \sum_{q=0}^{Q} \hat{w}_q^{(r,k)}(l) \psi_q(n). \tag{3.27}$$

Here we consider that the channel estimation error is from two sources: one is from the difference between $h_{\mathrm{BEM}}^{(r,k)}(n;l)$ and $\hat{h}_{\mathrm{BEM}}^{(r,k)}(n;l)$, the other is from the channel modeling error $e_{\mathrm{BEM}}^{(r,k)}(n;l)$ in (3.3). Therefore, the mean square error of channel estimation error is expressed as:

$$\sigma_{\tilde{h}}^2 = N^{-1} \sum_{r=1}^{R} \sum_{k=1}^{K} \sum_{n=0}^{N-1} \sum_{l=1}^{L} E \left\{ \left\| h_{\mathrm{BEM}}^{(r,k)}(n;l) - \hat{h}_{\mathrm{BEM}}^{(r,k)}(n;l) + e_{\mathrm{BEM}}^{(r,k)}(n;l) \right\|^2 \right\}. \tag{3.28}$$

Since the following orthogonality is true for both CE-BEM and DPS-BEM models

$$E \left\{ \left[h_{\mathrm{BEM}}^{(r,k)}(n;l) - \hat{h}_{\mathrm{BEM}}^{(r,k)}(n;l) \right]^H e_{\mathrm{BEM}}^{(r,k)}(n;l) \right\} = 0,$$

we have

$$\sigma_{\tilde{h}}^2 = \sigma_{\tilde{w}}^2 + \underbrace{N^{-1} \sum_{r=1}^{R} \sum_{k=1}^{K} \sum_{n=0}^{N-1} \sum_{l=0}^{L} E \left\{ \left[e_{\mathrm{BEM}}^{(r,k)}(n;l) \right]^H e_{\mathrm{BEM}}^{(r,k)}(n;l) \right\}}_{:=\sigma_{\mathrm{BEM}}^2}. \tag{3.29}$$

For simulations presented in Section 3.6 we calculate σ_{BEM}^2 by averaging over Monte Carlo runs.

3.4 Optimum Training Design

Observe from (3.19) that in order to achieve the lower bound of $\sigma_{\tilde{w}_{\mathrm{LS}}}^2$, The LS estimator requires $\mathbf{\Phi}^H \mathbf{\Phi}$ to be a diagonal matrix with all its diagonal entries equal. Observe from (3.25) and (3.26) that in order to achieve the lower bound of $\sigma_{\tilde{w}_{\mathrm{MMSE}}}^2$, The MMSE estimator also requires $\mathbf{\Phi}^H \mathbf{\Phi}$ to be a diagonal matrix to achieve (3.25), and for both it and $\mathbf{\Omega}_w$ to be diagonal with all their respective diagonal entries equal to achieve (3.26). Unfortunately, in general, the diagonal $\mathbf{\Omega}_w$ does not necessarily have all its diagonal entries equal. We will design the training schemes to make $\mathbf{\Phi}^H \mathbf{\Phi}$ to be a diagonal matrix with all its diagonal entries equal which will achieve the lower bound in (3.19) for the LS channel estimator and in (3.25) (but not in (3.26)) for the linear MMSE channel estimator.

Suppose that we choose

$$\mathbf{\Phi}_{k1}^H \mathbf{\Phi}_{k2} = \begin{cases} \alpha \boldsymbol{I}_{(Q+1)(L+1)} & , \quad \text{for } k1 = k2 \\ \mathbf{0} & , \quad \text{for } k1 \neq k2, \end{cases} \tag{3.30}$$

or, equivalently

$$\begin{aligned}
\textstyle\sum_{j=1}^{J} \boldsymbol{B}_{k1,j}^{H} \bar{\boldsymbol{D}}_{\psi_{q1},j}^{H} \bar{\boldsymbol{D}}_{\psi_{q2},j} \boldsymbol{B}_{k2,j} &= \alpha \boldsymbol{I} \quad , \quad \text{for } k1 = k2, q1 = q2 \\
\textstyle\sum_{j=1}^{J} \boldsymbol{B}_{k1,j}^{H} \bar{\boldsymbol{D}}_{\psi_{q1},j}^{H} \bar{\boldsymbol{D}}_{\psi_{q2},j} \boldsymbol{B}_{k2,j} &= \boldsymbol{0} \quad , \quad \text{for } k1 = k2, q1 \neq q2 \\
\textstyle\sum_{j=1}^{J} \boldsymbol{B}_{k1,j}^{H} \bar{\boldsymbol{D}}_{\psi_{q1},j}^{H} \bar{\boldsymbol{D}}_{\psi_{q2},j} \boldsymbol{B}_{k2,j} &= \boldsymbol{0} \quad , \quad \text{for } k1 \neq k2, \text{for all } q
\end{aligned} \tag{3.31}$$

for some $\alpha > 0$.

For CE-BEM, it is shown in [41] that

$$\sum_{j=1}^{J} \bar{\boldsymbol{D}}_{\psi_{q1},j}^{H} \bar{\boldsymbol{D}}_{\psi_{q2},j} = \begin{cases} \frac{1}{M} \boldsymbol{I}_{N_b - L} & , \quad \text{for } \ q1 = q2 \\ \boldsymbol{0} & , \quad \text{for } \ q1 \neq q2 \end{cases} \tag{3.32}$$

By (3.12) and (3.31), for all k's and q's, the training sequence should be designed to satisfy

$$\sum_{j=1}^{J} \boldsymbol{B}_{k1,j}^{H} \bar{\boldsymbol{D}}_{\psi_{q1},j}^{H} \bar{\boldsymbol{D}}_{\psi_{q2},j} \boldsymbol{B}_{k2,j} = \gamma \sum_{j=1}^{J} \bar{\boldsymbol{D}}_{\psi_{q1},j}^{H} \bar{\boldsymbol{D}}_{\psi_{q2},j} \tag{3.33}$$

for some $\gamma > 0$. Under (3.32), following [41] and [75], we pick $N_b = K(L+1) + K$ and $\boldsymbol{b}_{k,j}^{T} = [\boldsymbol{0}_{k(L+1)-1}, b_k, \boldsymbol{0}_{(K-k)(L+1)+L}]$ where $\boldsymbol{0}_{k(L+1)-1}$ is a size $k(L+1)-1$ null column, in which case $\boldsymbol{\Phi}_k^{H} \boldsymbol{\Phi}_k = \frac{b_k^2}{M} \boldsymbol{I}$. Therefore, with $\mathcal{P}_{bk} := J b_k^2$ denoting the total training power from the kth transmitter, we can obtain $\alpha = \frac{\mathcal{P}_{bk}}{N}$ (recall that $N = JM$). Since α should not be a function of k, we take $b_k = b$ leading to $\mathcal{P}_{bk} := \mathcal{P}_b$ for all k.

Plugging (3.30) into (3.20), (3.26) and (3.29), the LS and MMSE lower bounds of channel estimation error are derived as

$$\sigma^2_{\tilde{h}_{\mathrm{LS}}} = \sigma^2_{\tilde{w}_{\mathrm{LS}}} + \sigma^2_{\mathrm{CE}} = \frac{N\sigma^2_\eta}{\mathcal{P}_b} RK(L+1)(Q_F+1) + \sigma^2_{\mathrm{CE}}, \tag{3.34}$$

$$\sigma^2_{\tilde{h}_{\mathrm{MMSE}}} = \sigma^2_{\tilde{w}_{\mathrm{MMSE}}} + \sigma^2_{\mathrm{CE}} = \sum_{t=1}^{RK(L+1)(Q_F+1)} \left[\check{\lambda}_t^{-1} + \frac{\mathcal{P}_b}{\sigma^2_\eta N} \right]^{-1} + \sigma^2_{\mathrm{CE}}, \tag{3.35}$$

where $\check{\lambda}_t$ is the t-th diagonal entry of matrix $\boldsymbol{\Omega}_w$, i.e. $\check{\lambda}_t$ is the t-th eigenvalue of $\boldsymbol{R}_w$.

For DPS-BEM, using the heuristic asymptotic $(A.6)$, we can easily establish

$$\sum_{j=1}^{J} \bar{\boldsymbol{D}}^H_{\psi_{q1},j} \bar{\boldsymbol{D}}_{\psi_{q2},j} = \begin{cases} \frac{1}{M}\boldsymbol{I}_{(N_b-L)} & , \quad \text{for } q1 = q2 \\ \mathbf{0} & , \quad \text{for } q1 \neq q2. \end{cases} \tag{3.36}$$

Therefore, for DPS-BEM mimicking CE-BEM, the lower bounds of the channel estimation error are

$$\sigma^2_{\tilde{h}_{\mathrm{LS}}} = \frac{N\sigma^2_\eta}{\mathcal{P}_b} RK(L+1)(Q_S+1) + \sigma^2_{\mathrm{DPS}}, \tag{3.37}$$

$$\sigma^2_{\tilde{h}_{\mathrm{MMSE}}} = \sum_{t=1}^{RK(L+1)(Q_S+1)} \left[\check{\lambda}_t^{-1} + \frac{\mathcal{P}_b}{\sigma^2_\eta N} \right]^{-1} + \sigma^2_{\mathrm{DPS}}. \tag{3.38}$$

Remark 3.1. Note that (3.36) is critical for (3.37) and (3.38) to hold true and for the proposed training designs to be valid. The asymptotic Slepian sequences in $(A.6)$ are only "heuristic." But we can "verify" (3.36) by computing the results in (3.18) and (3.20) numerically and compare them with the results in (3.37). This is done so in Sec. 3.6.

3.5 Training Power Allocation

We assume that the time-varying channel $h^{(r,k)}(n;l)$ is zero-mean, WSS complex Gaussian with the same variance σ_h^2 for each tap. We also assume that the channel taps are mutually independent, i.e. $h^{(r,k)}(n;l)$ is WSSUS. To further simplify the analysis, we assume the channel correlation matrices are the same across $r \in [1, ..., R]$. The received information symbols at the rth received antenna can be expressed as

$$x_c^{(r)}(n) = \underbrace{\sum_{k=1}^{K}\sum_{l=0}^{L}\hat{h}^{(r,k)}(n;l)c_k(n-l)}_{:=x_s(n)} + \underbrace{\sum_{k=1}^{K}\sum_{l=0}^{L}[h^{(r,k)}(n;l)-\hat{h}^{(r,k)}(n;l)]c_k(n-l)+\eta^{(r)}(n)}_{=x_\eta(n)}, \tag{3.39}$$

where $\hat{h}^{(r,k)}(n;l) = \sum_{q=0}^{Q}\hat{w}_q^{(r,k)}(l)\psi_q(n)$ is used for data detection. Therefore, similar to the single user case in Section 2.5, the signal power of all receivers is given by

$$\begin{aligned}
\sigma_{xs}^2(n) &= \sum_{r=1}^{R} E\left\{|x_s(n)|^2\right\} \\
&= \sum_{r=1}^{R} E\left\{\sum_{k=1}^{K}\left[\left|\sum_{l=0}^{L}\hat{h}^{(r,k)}(n;l)c_k(n-l)\right|^2\right]\right\} \\
&= \bar{\mathcal{P}}_c\sum_{r=1}^{R}\sum_{k=1}^{K}\sum_{l=0}^{L}\left[E_{\boldsymbol{w}}\left\{E\left\{\left|h^{(r,k)}(n;l)-\hat{h}^{(r,k)}(n;l)\right|^2|\boldsymbol{w}\right\}\right\}+E\left\{|h^{(r,k)}(n;l)|^2\right\}\right] \\
&= \bar{\mathcal{P}}_c\left[\sigma_{\hat{h}}^2(n)+RK(L+1)\sigma_h^2\right],
\end{aligned} \tag{3.40}$$

where the average information power per user:

$$\bar{\mathcal{P}}_c := E\left\{|c_k(n)|^2\right\}. \tag{3.41}$$

The effective noise power is:

$$\begin{aligned}\sigma_{x\eta}^2(n) &= \sum_{r=1}^{R} E\left\{\sum_{k=1}^{K}\left[\left|\sum_{l=0}^{L}[h^{(r,k)}(n;l)-\hat{h}^{(r,k)}(n;l)]c_k(n-l)\right|^2\right]\right\}+\sum_{r=1}^{R}E\left\{|\eta^{(r)}(n)|^2\right\}\\ &= \bar{\mathcal{P}}_c\sum_{r=1}^{R}\left[\sum_{k=1}^{K}\sum_{l=0}^{L}E_w\left\{E\left\{\left|h^{(r,k)}(n;l)-\hat{h}^{(r,k)}(n;l)\right|^2|\boldsymbol{w}\right\}\right\}\right]+R\sigma_\eta^2\\ &= \bar{\mathcal{P}}_c\sigma_{\tilde{h}}^2(n)+R\sigma_\eta^2 \end{aligned} \tag{3.42}$$

where $\sigma_{\tilde{h}}^2(n) := \sum_{r=1}^{R}\sum_{k=1}^{K}\sum_{l=0}^{L}E_w\left\{E\left\{\left|h^{(r,k)}(n;l)-\hat{h}^{(r,k)}(n;l)\right|^2|\boldsymbol{w}\right\}\right\}$. Define

$$\begin{aligned}\mathcal{W}_l^{(r,k)} &:= [w_0^{(r,k)}(l), w_1^{(r,k)}(l), ..., w_Q^{(r,k)}(l)]^T,\\ \mathcal{W}^{(r,k)} &:= [\mathcal{W}_0^{(r,k)}, \mathcal{W}_1^{(r,k)}, ..., \mathcal{W}_L^{(r,k)}]^T,\\ \mathcal{W}^{(r)} &:= [\mathcal{W}^{(r,1)}, \mathcal{W}^{(r,2)}, ..., \mathcal{W}^{(r,K)}]^T,\\ \mathcal{W} &:= [\mathcal{W}^{(1)}, \mathcal{W}^{(2)}, ..., \mathcal{W}^{(R)}]^T. \end{aligned} \tag{3.43}$$

By (3.15) and (3.43), we can get the following relationship:

$$\mathcal{W} = \boldsymbol{\Theta}\boldsymbol{w}, \tag{3.44}$$

where

$$\boldsymbol{\Xi} = \begin{bmatrix} 1 & \overbrace{0...0}^{L} & \dots & & \dots & \\ 0 & 0...0 & 1 & \overbrace{0...0}^{L} & \dots & \dots \\ \dots & & \dots & & \dots & \\ 0 & 1 & \overbrace{0...0}^{L} & \dots & \dots & \dots \\ 0 & 0 & \overbrace{0...0}^{L} & 1 & \overbrace{0...0}^{L} & \dots \\ \ddots & & \ddots & & \ddots & \\ \dots & \dots & \dots & \dots & \dots & 1 \end{bmatrix}_{(Q+1)(L+1)\times(Q+1)(L+1)} \tag{3.45}$$

and

$$\boldsymbol{\Theta} = \boldsymbol{I}_{RK} \otimes \boldsymbol{\Xi}. \tag{3.46}$$

We also find that $\boldsymbol{\Theta}^H\boldsymbol{\Theta}$ is always an identity matrix. Then $\sigma^2_{\tilde{h}}(n)$ can be rewritten as

$$\begin{aligned} \sigma^2_{\tilde{h}}(n) &= \sum_{r=1}^{R}\sum_{k=1}^{K}\sum_{l=0}^{L} E_{\mathcal{W}}\left\{E\left\{\left|h^{(r,k)}(n;l) - \hat{h}^{(r,k)}(n;l)\right|^2 |\mathcal{W}\right\}\right\} \\ &= \mathrm{tr}\left\{\boldsymbol{\Psi}(n) E_{\mathcal{W}}\left\{\mathrm{cov}\{\hat{\mathcal{W}},\hat{\mathcal{W}}|\mathcal{W}\}\right\}\boldsymbol{\Psi}(n)^H\right\}, \end{aligned} \tag{3.47}$$

where

$$\begin{aligned} \boldsymbol{\psi}(n) &:= [\psi_0(n), \psi_1(n), ..., \psi_Q(n)], \\ \boldsymbol{\Psi}(n) &:= \boldsymbol{I}_{RK(L+1)} \otimes \boldsymbol{\psi}(n). \end{aligned}$$

Based on (3.43) and (3.44), we have

$$E_{\mathcal{W}}\left\{\mathrm{cov}\{\hat{\mathcal{W}},\hat{\mathcal{W}}|\mathcal{W}\}\right\}=\boldsymbol{\Theta}\underbrace{E_{\mathcal{W}}\left\{\mathrm{cov}\{\hat{\boldsymbol{w}},\hat{\boldsymbol{w}}|\mathcal{W}\}\right\}}_{\boldsymbol{R}_{\tilde{w}}}\boldsymbol{\Theta}^{H}. \tag{3.48}$$

Based on the orthonormality of $\psi_q(n)$, we have

$$\sum_{n=1}^{N}\boldsymbol{\Psi}(n)^{H}\boldsymbol{\Psi}(n)=\boldsymbol{I}_{RK(L+1)(Q+1)}. \tag{3.49}$$

Therefore, the time-averaged of $\bar{\sigma}_{\tilde{h}}^2$ over length N is

$$\bar{\sigma}_{\tilde{h}}^2:=(N-JN_b)^{-1}\sum_{n}\sigma_{\tilde{h}}^2(n)\approx N^{-1}\sum_{n=0}^{N-1}\sigma_{\tilde{h}}^2(n)=\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{\Theta}\boldsymbol{R}_{\tilde{w}}\boldsymbol{\Theta}^{H}\right\}=\frac{1}{N}\sigma_{\tilde{w}}^2. \tag{3.50}$$

In a similar way, the time averaged signal and noise power can be expressed as

$$\begin{aligned}\bar{\sigma}_{xc}^2&=\frac{1}{N}\sum_{n=1}^{N}\sigma_{xc}^2(n)=\bar{\mathcal{P}}_c[\bar{\sigma}_{\tilde{h}}^2+RK(L+1)\sigma_h^2],\\ \bar{\sigma}_{xn}^2&=\frac{1}{N}\sum_{n=1}^{N}\sigma_{xn}^2(n)=\bar{\mathcal{P}}_c\bar{\sigma}_{\tilde{h}}^2+R\sigma_\eta^2.\end{aligned} \tag{3.51}$$

Therefore, we obtain an effective average SNR of (3.39) as

$$\mathrm{SNR}_d=\frac{\bar{\sigma}_{xc}^2}{\bar{\sigma}_{xn}^2}. \tag{3.52}$$

Define the total information power and received signal power $\mathcal{P}_c:=JN_c\bar{\mathcal{P}}_c$ and $\mathcal{P}:=\mathcal{P}_b+\mathcal{P}_c$, respectively. Define the training power overhead $\beta:=\frac{\mathcal{P}_b}{\mathcal{P}_c+\mathcal{P}_b}$. Our objective is to maximize SNR with respect to β under the constraint of a fixed $\mathcal{P}$. Thus, incorporating those constraint-carrying

variables into (3.52) and using the developed expression for average signal and noise powers in (3.52), we obtain the unconstrained cost

$$\mathrm{SNR}_d(\beta) = \frac{\frac{(1-\beta)\mathcal{P}}{JN_c}[\bar{\sigma}_{\tilde{h}}^2 + RK(L+1)\sigma_{\tilde{h}}^2]}{\frac{(1-\beta)\mathcal{P}}{JN_c}\bar{\sigma}_{\tilde{h}}^2 + \sigma_\eta^2}. \tag{3.53}$$

Using the lower bound of the LS estimator in (3.37) (due to the lower bound of the MMSE estimator, the closed form of the optimal β can not be obtained), we can explicitly write (3.52) as

$$\mathrm{SNR}_d(\beta) = \frac{f_1\beta^2 + f_2\beta + f_3}{g_1\beta + g_2}, \tag{3.54}$$

where

$$\begin{aligned}
f_1 &= -\frac{\mathcal{P}RK(L+1)\sigma_{\tilde{h}}^2}{N_cJ}, \\
f_2 &= \frac{\mathcal{P}RK(L+1)\sigma_{\tilde{h}}^2}{N_cJ} - \frac{RK(L+1)(Q+1)\sigma_\eta^2}{N_cJ}, \\
f_3 &= \frac{RK(L+1)(Q+1)\sigma_\eta^2}{N_cJ}, \\
g_1 &= \sigma_\eta^2 - \frac{RK(L+1)(Q+1)\sigma_\eta^2}{N_cJ}, \\
g_2 &= \frac{RK(L+1)(Q+1)\sigma_\eta^2}{N_cJ}.
\end{aligned} \tag{3.55}$$

Setting the first derivative of $\mathrm{SNR}(\beta)$ with respect to β to zero, we obtain a quadratic equation in β

$$\beta^2 + 2\frac{g_2}{g_1}\beta + \frac{f_2g_2 - f_3g_1}{f_1g_1} = 0 \tag{3.56}$$

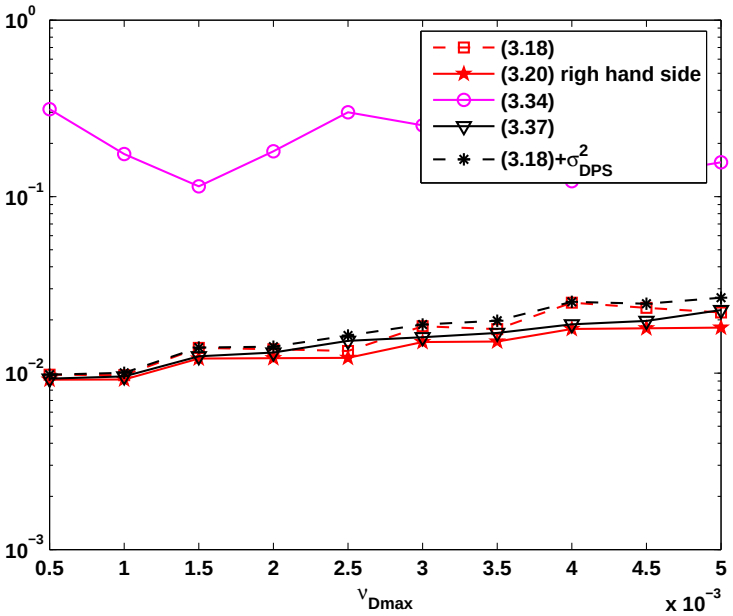

Figure 3.1: Channel estimation errors (with and without modeling errors): numerical results derived from $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$, SNR=25dB

the two roots, one of which is negative $(\beta < 0)$, and hence is excluded. The other root is given by

$$\beta_{\text{opt}} = \frac{g_2}{g_1}\left[-1 + \sqrt{1 + \frac{g_1(f_3 g_1 - f_2 g_2)}{g_2^2 f_1}}\right]. \tag{3.57}$$

3.6 Numerical Examples

In the following examples we consider a multiuser system with $K = 2$ users and $R = 2$ receivers. We use binary phase shift keying (BPSK) and quadrature phase shift keying (QPSK) modulation. Each transmitted block has $J = 10$ subblocks, and each subblock has $N_c = 30$ information symbols and $N_b = 3L + 2$ training symbols with optimal structure $\left[\mathbf{0}_{k(L+1)-1}, b,\right.$

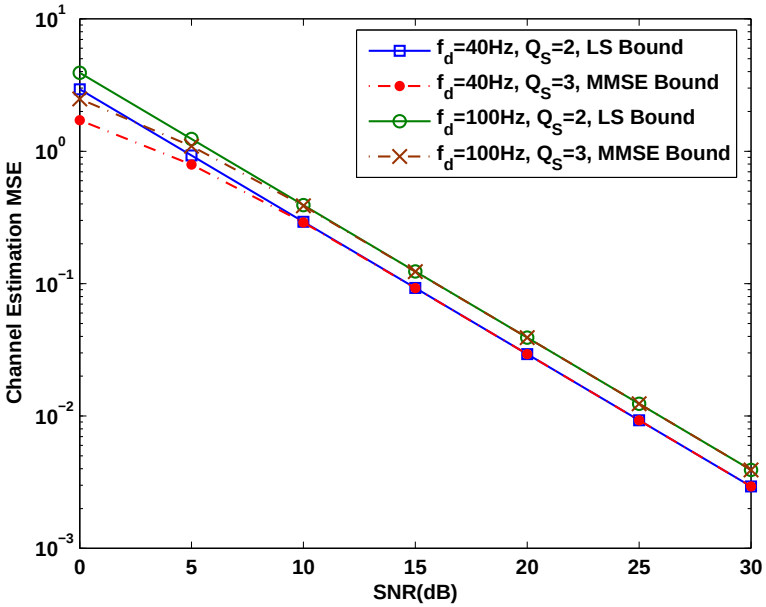

Figure 3.2: MSE lower bound comparison between LS (3.37) and MMSE (3.38) estimators

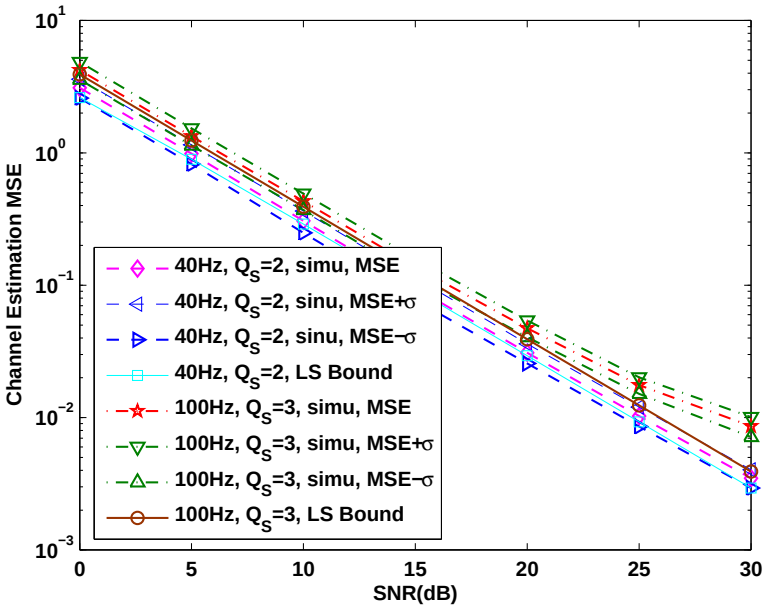

Figure 3.3: Comparison between channel estimation MSEs lower bound in (3.37) and simulation results (3.58)

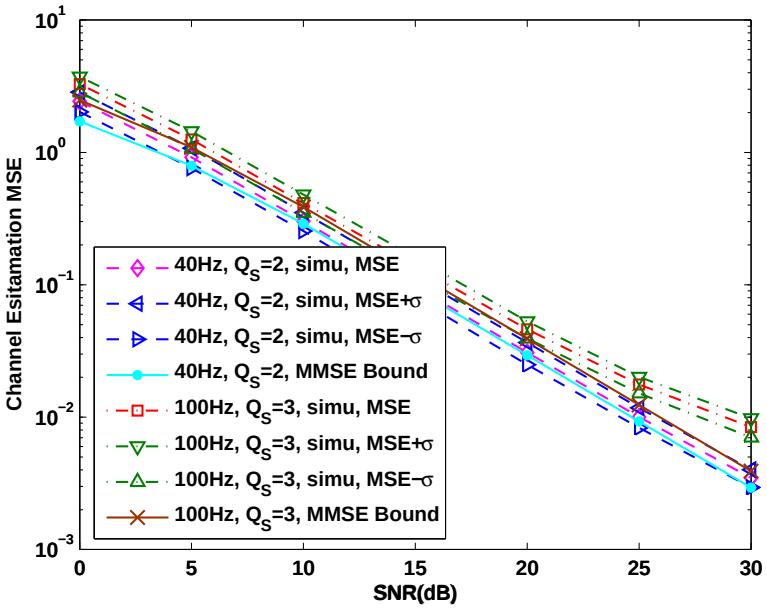

Figure 3.4: Comparison between channel estimation MSEs lower bound in (3.38) and simulation results (3.58)

$\mathbf{0}_{(K-k)(L+1)+L}\big]$, $b > 0$. A doubly-selective Rayleigh fading channel $h^{(r,k)}(n;l)$ is simulated according to [77, 81] with channel order $L = 2$, carrier frequency of 2GHz, data rate of 40 kbps, and thus, symbol duration $T_s = 25\mu s$. Therefore, each tap of the generated time-variant channel has a Jakes' spectrum; it is not generated using the assumed BEM modeling. Also, 3 taps of the channel are mutually independent. Depending on different maximum Doppler spread f_ds, a varying maximum normalized one-sided Doppler bandwidth $\nu_{\mathrm{Dmax}} = f_d T_s$ can be derived. Given ν_{Dmax}, Jakes' spectrum and other information, we can calculate $\boldsymbol{R}_w$ and therefore, $\boldsymbol{\Omega}_w$ and $\check{\lambda}_t$s, needed in (3.25), (3.38) and elsewhere. A Kalman filter formulation is used for information detection after the channel estimation. The SNR refers to $1/\sigma_\eta^2$ where the information sequence power per user is normalized to one and the channel power (per user) is also normalized to unity.

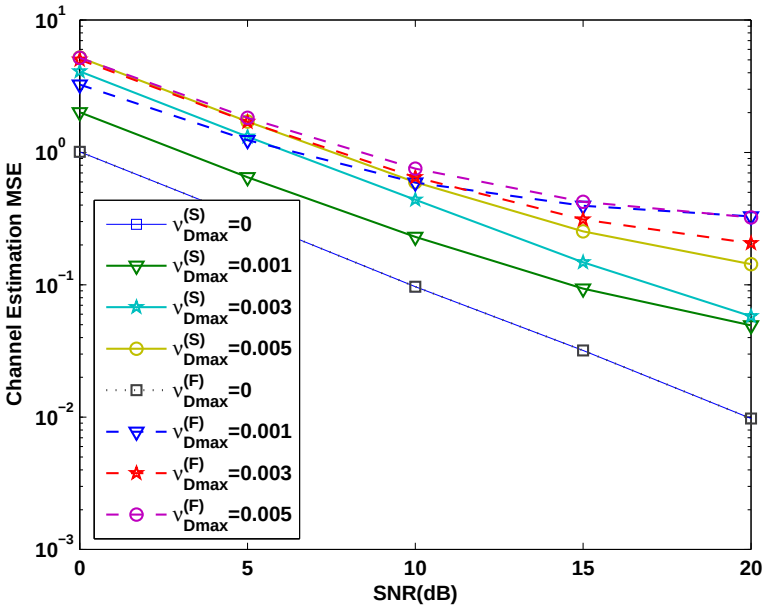

Figure 3.5: Channel estimation MSE with varying maximum normalized Doppler bandwidth, $Q_F = 2\lceil \nu_{Dmax} N \rceil$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil$

3.6.1 Example 1: Approximation Errors

As noted earlier in Remark 3.1, here we want to show the influence of approximation errors in $\bar{\boldsymbol{D}}_{\psi_q,j}$ when the true Slepian sequences instead of the approximations in $(A.6)$ are used. We compute (3.18) and (3.20) numerically with true Slepian sequences generated by (2.5), then compare them with (3.34) and (3.37). For the parameters stated earlier in this section (J=10, N_b=8, N_c=30, $M = N_b + N_c$=38, $N = JM$ =380) the results are shown in Fig. 3.1, where the dimensions $Q_F = 2\lceil \nu_{Dmax} N \rceil$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil$ change with the maximum Doppler bandwidth and SNR=25dB. From Fig. 3.1 we see that the CE-BEM lower bound in (3.34) is high due to "large" modeling error σ^2_{rmCE} even though the error for the modeled part is minimum, whereas the results in (3.18), (3.20) and (3.37) are close to each other due to "small" modeling error σ^2_{DPS} and "close-to-minimum" (compare curves for (3.18) (exact error) and (3.20) (lower bound)) error for

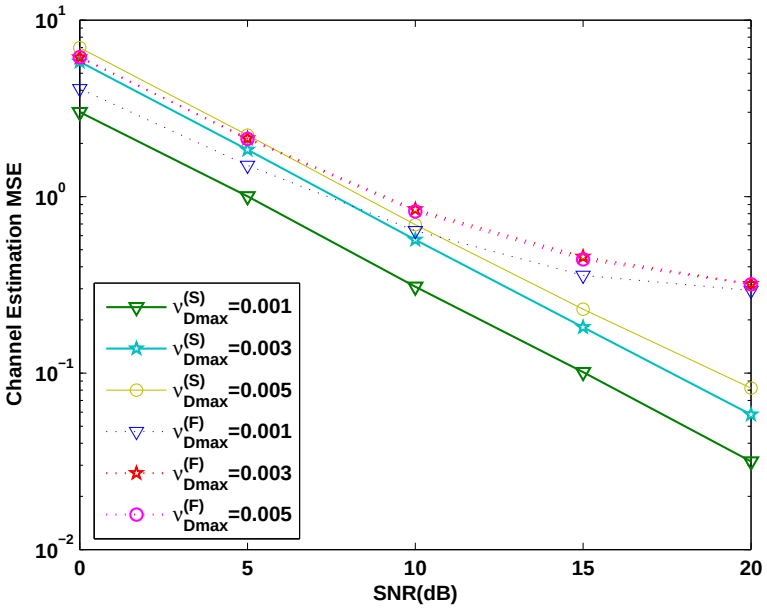

Figure 3.6: Channel estimation MSE with varying maximum normalized Doppler bandwidth, $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$

the modeled part. [We note that σ^2_{DPS} and σ^2_{CE} were obtained via Monte Carlo averaging, similar as in [77].]

3.6.2 Example 2: Channel Estimation Performance

In this case we pick $b = 1$ in training. The LS and MMSE estimators are used to estimate $\boldsymbol{w}$, and then the channel is estimated as $\hat{h}^{(r,k)}(n;l) = \sum_{q=0}^{Q} \hat{w}_q^{(r,k)}(l)\psi_q(n)$. Based on M_r Monte Carlo runs, the channel estimation MSE is calculated as

$$\mathrm{MSE} = (M_r N)^{-1} \sum_{\alpha=1}^{M_r} \sum_{r=1}^{R} \sum_{k=1}^{K} \sum_{n=0}^{N-1} \sum_{l=0}^{L} |\hat{h}_\alpha^{(r,k)}(n;l) - h_\alpha^{(r,k)}(n;l)|^2. \tag{3.58}$$

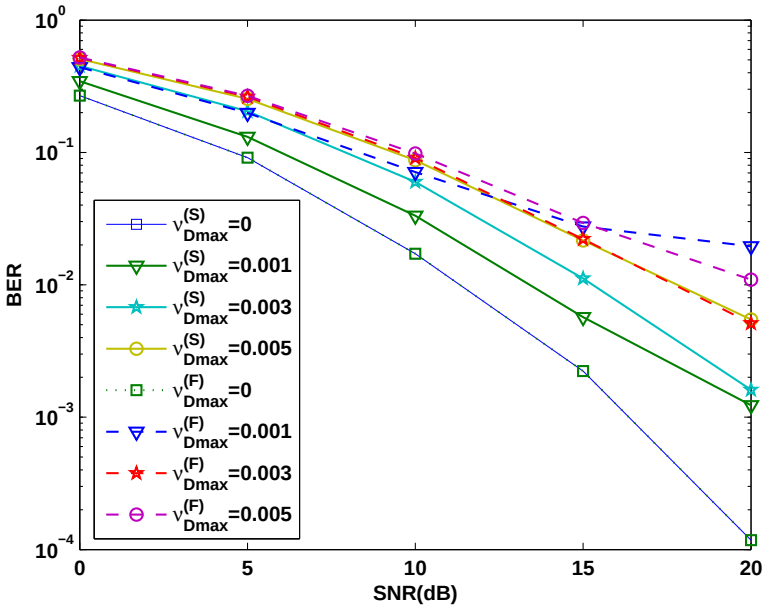

Figure 3.7: BER with varying maximum normalized Doppler bandwidth for BPSK modulation, $Q_F = 2\lceil \nu_{Dmax} N \rceil$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil$

In Fig. 3.2, the lower bounds in (3.37) and (3.38) are plotted using $Q_S = 2$ for $f_d = 40$Hz and using $Q_S = 3$ for $f_d = 100$Hz. [The MMSE bound needs $\check{\lambda}_t$s which requires knowledge of the Doppler spread. Note also that for both $f_d = 40$Hz and $f_d = 100$Hz, the minimum $Q_S = \lceil 2\nu_{\rm Dmax} N \rceil$ are actually $Q_S = 1$ and $Q_S = 2$, respectively; however, a slightly higher value of $Q_S = \lceil 2\nu_{\rm Dmax} N \rceil + 1 = 3$ for $f_d = 100$Hz yields smaller modeling error which has been proved in Example 1]. Then they are compared with the simulation results (averaged over 200 Monte Carlo runs) in Figs. 3.3 and 3.4 for $f_d = 40$Hz and 100Hz; also shown are $\pm\sigma$ bounds on the simulation averages. It can be seen that the theoretical results are consistent with the simulation results for f_d =40Hz indicating that the optimal pilot design does minimize channel MSEs when using DPS-BEM. For f_d=100 Hz, there is a "small gap" between theory and simulations at high SNRs which is probably attributable to modeling error ("small" but nonzero). Furthermore, the performance of

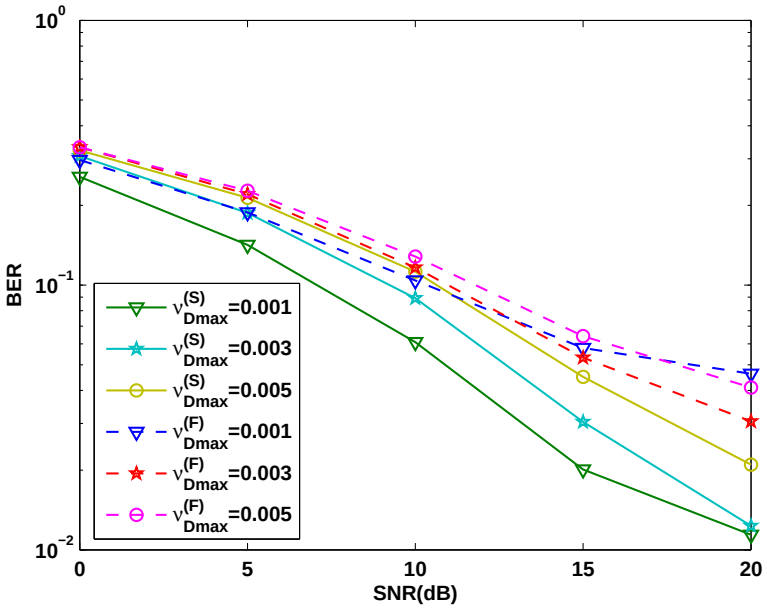

Figure 3.8: BER with varying maximum normalized Doppler bandwidth for QPSK modulation, $Q_F = 2\lceil \nu_{Dmax} N \rceil$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil$

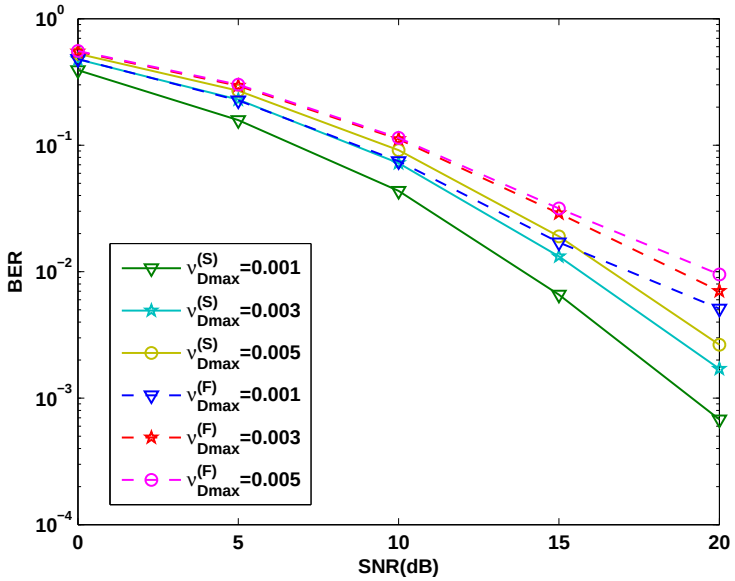

Figure 3.9: BER with varying maximum normalized Doppler bandwidth for BPSK modulation, $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$

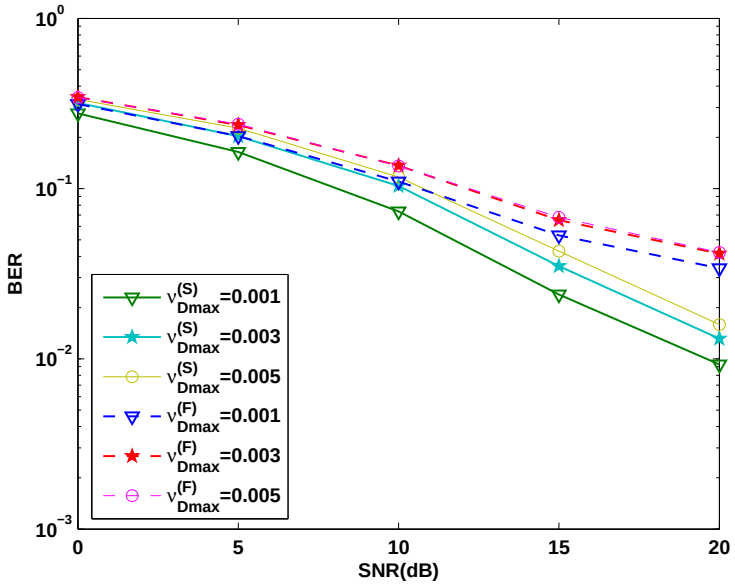

Figure 3.10: BER with varying maximum normalized Doppler bandwidth for QPSK modulation, $Q_F = 2\lceil \nu_{Dmax} N \rceil + 1$ and $Q_S = \lceil 2\nu_{Dmax} N \rceil + 1$

MMSE estimator outperforms that of LS estimator at low SNR, and they converge to each other at high SNR.

3.6.3 Example 3: CE-BEM versus DPS-BEM

In Fig. 3.5, the LS channel estimation MSE (3.58) versus SNR under different maximum Doppler bandwidths ($\nu^{(S)}_{\rm Dmax}$ is for DPS-BEM, $\nu^{(F)}_{\rm Dmax}$ is for CE-BEM) are plotted. It is clear that the MSE of DPS-BEM is consistently smaller than that of CE-BEM. However, we notice that the channel estimation error based on DPS-BEM suffers an error floor, too. This can be alleviated by taking Q_S to be larger than $\lceil 2\nu_{\rm Dmax} N \rceil$ at the cost of more computations, as Fig. 3.6 shows. Fig. 3.7 and Fig. 3.9 take BER (average over 2000 Monte Carlo runs) as a performance measure to compare the performance between DPS-BEM and CE-BEM. Comparing with Fig. 3.5 and Fig. 3.6 make it quite

clear that the drastically reduced MSE for the DPS-BEM channel estimation leads to a pronounced reduction in BER compared to the CE-BEM case. The BER performances under QPSK modulation are plotted in Fig. 3.8 and Fig. 3.10, which are worse than those of BPSK modulation, as expected.

3.7 Conclusion

In this chapter, the channel estimation approaches and optimum training design proposed in Chapter 2 were extended to multiuser MIMO system. Illustrative simulation examples were provided to show the performance of channel estimation and data detection under MIMO system. The system model and performance analysis in Chapter 3 are consistent with those in Chapter 2. The numerical results of MIMO system (Chapter 3) outperforms the SISO system (Chapter 2) due to the diversity gain.

CHAPTER 4

TIME-VARYING EQUALIZATION FOR DOUBLY-SELECTIVE CHANNELS

4.1 Introduction

Well-known techniques of equalization include (a) linear equalization (LE), (b) decision feedback equalization (DFE), and (c) maximum likelihood sequence estimation (MLSE) [66]. Linear equalizers are the simplest computationally whereas MLSE yields (near-)optimum performance but is computationally demanding. DFE typically provides a good compromise between complexity and performance. In this chapter we consider LE and DFE for doubly-selective channels modeled via basis expansion models (BEM).

Design of equalizers for wireless communication systems over doubly-selective channels has been studied in the literature [26, 67, 45, 46, 3, 5, 68]. These equalizers may be divided into two types: block equalizers and serial equalizers. The block equalizers are usually complex to design since inversion of a large matrix is required. Especially, since a doubly-selective channel can not be diagonalized by a channel-independent transformation, the implementation of block time-varying (TV) equalizers, which collect and process in blocks all the available data in the received frame, leads to a very high computational complexity [5]. On the other hand, serial equalizers are more favored since they process few data at a time and provide a flexible trade-off between complexity and performance [66]. The TV serial equalizers in [45, 46, 3, 5, 68] rely on a particular basis expansion model (BEM) of the TV channel impulse response, namely complex exponential (CE) BEM, and the knowledge of the BEM coefficients at the receiver. Furthermore, they also use a CE-BEM representation for the TV equalizer; therefore, when a low-order CE-BEM model is used for the equalizer, it incurs an approximation error inherent in CE-BEM modeling of equalizers.

Thus, their performances are highly related to the choice of the number of BEM coefficients and the equalizer order.

It is well known that the Kalman filter, a linear recursive minimum mean-square-error (MMSE) filter, is the best linear MMSE detector for a given detection delay. It provides a symbol detection method of obtaining the true MMSE performance within an implementable structure having a finite number of filter weights. The Kalman filter has been widely used in channel estimation and equalization over wireless channels [74]. In this chapter we exploit the Kalman filter as a TV linear MMSE equalizer for doubly-selective channels. An alternative formulation of the FIR DFE based on a CE-BEM channel model is also proposed for doubly-selective channels.

The main advantages of the proposed LE and DFE are fourfold: (i) The design process does not rely on a specific basis expansion model for the underlying channel; therefore, it can be applied to any doubly-selective channel model. (ii) The proposed equalizers rely solely on the channel model and therefore, do not incur any approximation error inherent in CE-BEM modeling of equalizers. (iii) Only one parameter, the equalizer delay, will influence the performance of the Kalman filter; only three parameters, the equalizer delay, and feedforward and feedback filter lengths, will influence the performance of proposed DFE. (iv) The computational complexity in terms of the number of flops turns out to be much lower than existing BEM-based TV equalizers. The simulation results will show that the proposed equalizers can achieve the same or an improved bit error rate (BER) performance than the equalizers in [45, 46, 3, 5, 68], without incurring the approximation error inherent in BEM modeling of equalizers.

4.2 Linear Equalization

4.2.1 System Model

Consider a doubly-selective finite impulse response (FIR) linear channel with single input and multiple outputs (SIMO). The discrete time impulse response $h^{(r)}(n;l)$ denotes the time-varying impulse response of channel that includes transmit-receive filters as well as doubly-selective propagation effects between the transmit antenna and the rth receive antenna, where $n \in [0, 1, ..., N-1]$ and $l \in [0, L]$ capture the time- and frequency- selectivity of the channel, respectively. Let $s(n)$ denote the transmitted symbols which is input to the SIMO channel, the received sequence at the rth receive antenna can be written as

$$x^{(r)}(n) = \sum_{l=0}^{L} h^{(r)}(n;l)s(n-l) + \eta^{(r)}(n), \tag{4.1}$$

where $\eta^{(r)}(n)$ is the additive complex Gaussian noise at the rth receive antenna, with zero-mean and variance σ_η^2.

In [45, 46, 3, 5, 68], the designs of TV equalizers depend on a BEM of the time-varying channel, given by [26]:

$$h^{(r)}(n;l) = \sum_{q=0}^{Q} w_q^{(r)}(l)\psi_q(n), \tag{4.2}$$

where $w_q^{(r)}(l)$ is the basis expansion model parameter (coefficient), $\psi_q(n)$ is the basis expansion function, and $Q+1$ is the basis dimension that satisfies $2\lceil \nu_{\mathrm{Dmax}} N \rceil \leq Q \leq N-1$ with ν_{Dmax} the maximum normalized Doppler bandwidth. The above model is valid over a block of N symbols over which the BEM coefficients remain time-invariant. The BEM coefficients may change from block to block. Therefore, we will refer to the equalizers of [45, 46, 3, 5, 68] as BEM-based equalizers.

More specifically, the complex exponential basis functions [26] are used in [45, 46, 3, 5, 68], where

$$\psi_q(n) = e^{j2\pi(q-Q/2)n/P}. \tag{4.3}$$

if $P = N$ where N is the block size in symbols, the CE-BEM is said to be critically sampled whereas when $P > N$, it is said to be oversampled [37]. The oversampling reduces frequency spacing of complex exponentials and gives a better representation of the channel impulse response. [68] uses $P = 2N$ whereas [5] has used $P = N$.

Collecting the symbols $\left\{x^{(r)}(n)\right\}_{r=1}^{r=R}$ received by the R receivers into the vector $\boldsymbol{x}(n) := [x_1(n), x_2(n), ..., x_R(n)]^T \in \mathbb{C}^R$, we can obtain the matrix-vector form of (4.1):

$$\boldsymbol{x}(n) = \sum_{l=0}^{L} \sum_{q=-Q/2}^{Q/2} \boldsymbol{w}_q(l) s(n-l) e^{j2\pi qn/P} + \boldsymbol{\eta}(n), \tag{4.4}$$

where $\boldsymbol{w}_q(l) := [w_q^{(1)}(l), ..., w_q^{(R)}(l)]^T \in \mathbb{C}^T$ and $\boldsymbol{\eta}(n) := [\eta^{(1)}(n), ..., \eta^{(R)}(n)]^R \in \mathbb{C}^R$.

If the input vector of a BEM-based equalizer of order L' is defined as $\boldsymbol{y}(n) := [\boldsymbol{x}^T(n), \boldsymbol{x}^T(n-1), ..., \boldsymbol{x}^T(n-L')]^T \in \mathbb{C}^{R(L'+1)}$, the output of the equalizer can be written as:

$$\hat{s}(n) = \boldsymbol{g}^H(n)\boldsymbol{y}(n), \tag{4.5}$$

where the vector $\boldsymbol{g}(n) \in \mathbb{C}^{R(L'+1)}$ collects all the equalizer parameters and d is the equalization delay.

The following model assumptions are considered throughout this chapter:

(H4.1): The information symbols $s(n)$ consists of zero-mean, finite-alphabet of independent and identically distributed (i.i.d.) random variables, satisfying $E\left\{s(m)s^*(n)\right\} = \sigma_s^2\delta(m-n)$.

(H4.2): The measurement noise $\eta^{(r)}(n)$ in Eq. (3.1) is uncorrelated with $s(n)$, and $E\{\eta(m)\eta^*(n)\} = \sigma_\eta^2\delta(m-n)$.

4.2.2 Existing Linear Time-Varying MMSE Equalizers

In [45, 46, 3, 5, 68] both zeroforcing and MMSE linear equalizers have been considered. Since the Kalman filter yields a linear MMSE equalizer, in this section we will restrict ourselves to existing TV MMSE equalizers. First we briefly review two linear TV MMSE equalizers from [5] and [68], respectively. The design of these equalizers assumes availability of the CE-BEM coefficients $w_q^{(r)}(l)$s of the channel.

BLM Equalizer [5]

In [45, 46, 3, 5] one seeks

$$\hat{s}(n) = \sum_{l'=d}^{L'-d} \sum_{q'=-Q'/2}^{Q'/2} \sum_{r=1}^{R} g_{q'}^{(r)}(l') e^{j2\pi q' n/P} x^{(r)}(n-l'), \tag{4.6}$$

where design parameters are d, L', Q' and equalizer coefficients $g_{q'}^{(r)}(l')$, invariant over $n \in \{0, 1, ..., N-1\}$. Notice that the structure of the equalizer is that of a CE-BEM. Define

$$g^{(r)} := \left[g_{-Q'/2}^{(r)}(-d), ..., g_{-Q'/2}^{(r)}(L'-d), g_{-Q'/2+1}^{(r)}(-d), ..., g_{Q'/2}^{(r)}(-d), ..., g_{Q'/2}^{(r)}(L'-d)\right]^T \tag{4.7}$$

and

$$\boldsymbol{g}_{\text{blm}} = \left[\boldsymbol{g}^{(1)T}, \boldsymbol{g}^{(2)T}, ..., \boldsymbol{g}^{(R)T}\right]^T \in \mathbb{C}^{R(L'+1)(Q'+1)}. \tag{4.8}$$

The TV MMSE equalizer proposed in [45, 46, 3, 5](henceforth referred to as BLM) provides an explicit frequency-domain representation, by turning a TV equalization problem into a simpler time-invariant (TI) filtering design, which only involves the TI basis expansion parameters of the doubly-selective channel. The (TV MMSE) BLM equalizer can be expressed as [5]:

$$\boldsymbol{g}_{\rm blm} = \boldsymbol{H}_{\rm blm} \left[\boldsymbol{H}_{\rm blm}^H \boldsymbol{H}_{\rm blm}(n) + \frac{\sigma_\eta^2}{\sigma_s^2} \boldsymbol{I}_{(Q+Q'+1)(L+L'+1)} \right]^{-1} \boldsymbol{e}_d \tag{4.9}$$

where the $[R(Q'+1)(L'+1)] \times [(Q+Q'+1)(L'+L+1)]$ matrix $\boldsymbol{H}_{\rm blm}$ collects TI BEM parameters and TV complex exponential basis functions as given in Sec. V in [5], $d \in \{0, 1, ..., L+L'\}$ is the delay of the equalizer, and $\boldsymbol{e}_d \in \mathbb{R}^{(Q'+Q+1)(L'+L+1)}$ is an unit vector with the element 1 in the $(d+1)$st position.

Computational Complexity: The inverse in (4.9) requires $\mathcal{O}(K^3)$ flops (one flop is roughly one multiply-and-accumulate operation) where $K = (Q+Q'+1)(L+L'+1)$ and $\mathcal{O}(K^3) = K^3 + 0.5K^2 + 0.5K$ if one uses modified Cholesky decomposition (i.e. UD-decomposition) based approach to matrix inverse ([28] Table 6.13). This has been called design complexity (equalizer design) in [5]. For implementation of (4.6), one needs $NR(Q'+1)(L'+1)$ flops; this has been called implementation complexity in [5]. For numerical comparisons in Sec. 4.2.4, we assume that the inverse of a positive-definite matrix has been computed via the modified Cholesky decomposition method.

FRESH Equalizer [68]

The MMSE solution to the frequency-shift TV equalizer proposed in [68] (referred to as FRESH) gives the canonical frequency-domain representation of the optimal norm TV-MMSE equalizers, which leads the optimal FRESH equalizer design to construct the TI Fourier coefficients of optimal

equalizer weights $\boldsymbol{g}_{\mathrm{opt}}(n)$. The discrete Fourier series (DFS) expansion relationship of $\boldsymbol{g}_{\mathrm{opt}}(n)$ is:

$$\boldsymbol{g}_{\mathrm{opt}}(n) = \sum_{p=0}^{P-1} \boldsymbol{g}_{p,\mathrm{opt}} e^{j2\pi np/P}. \tag{4.10}$$

Estimation of $\boldsymbol{g}_{\mathrm{opt}}(n)$ via estimation of $\boldsymbol{g}_{p,\mathrm{opt}}$ is discussed in [68] for zero-foring equalization with a specified equalization delay. It is straightforward to modify the design to obtain a linear MMSE equalizer; for instance, add the noise variance term to eq. (26) in [68] just as has been done in (4.9).

However, the implementation complexity of the FRESH representation of the optimal TV MMSE equalizer maybe quite large for "large" values of N. Therefore, a low-complexity implementation of the optimal FRESH equalizer (refer to as a sub-optimum FRESH method) is also derived in [68] by evaluate only $Q'+1$ Fourier coefficients in (4.10), which is similar to that considered in BLM equalizer design. So the DFS expansion for sub-optimum FRESH equalizer is given by

$$\boldsymbol{g}_{\mathrm{subopt}}(n) = \sum_{p=-Q'/2}^{Q'/2} \boldsymbol{g}_{p,\mathrm{subopt}} e^{j2\pi np/P} \tag{4.11}$$

where $\boldsymbol{g}_{p,\mathrm{subopt}} = \boldsymbol{g}_{p,\mathrm{opt}}$ for $p = 0, 1, ..., Q'/2$ and $\boldsymbol{g}_{p,\mathrm{subopt}} = \boldsymbol{g}_{p+P,\mathrm{opt}}$ for $p = -1, -2, ..., -Q'/2$.

Computational Complexity: As discussed in [68], design complexity of both optimal and suboptimal FRESH equalizers involves $P \cdot \mathcal{O}[(L + L' + 1)^3] + (P \log_2 P)(L + L' + 1)^2$ flops where $\mathcal{O}(K^3) = K^3 + 0.5K^2 + 0.5K$ when one uses a UD-decomposition approach to positive-definite matrix inversion. For implementation, one needs $NRP(L' + 1)$ flops for optimal FRESH and $NR(Q' + 1)(L' + 1)$ flops for suboptimal FRESH equalizer.

Remark 4.1: The design of both BLM and FRESH equalizers rely on the CE-BEM representation of the channel. They assume that the BEM parameters $w_q^{(r)}(l)$ are known at the receiver. It is still not clear if these design methods apply to other basis expansion models. However, the

simulation results in [5] illustrate that the BEM modeling errors have a significant influence on the performance.

Remark 4.2: Three parameters have to be considered or optimized for BLM and FRESH equalizers design, d, L' and Q'.

4.2.3 Kalman Filter with Equalization Delay d

The Kalman filter has been widely used in channel estimation and equalization over wireless channels; see [74] and references therein. It requires a state-space model of the underlying system. Define the state vector $\boldsymbol{s}(n)$

$$\boldsymbol{s}(n) := [s(n), s(n-1), ..., s(n-d)]^T \in \mathbb{C}^{d+1}, \tag{4.12}$$

the state transition matrix $\boldsymbol{\Phi}$ and the input vector $\boldsymbol{\xi}$

$$\boldsymbol{\Phi} := \begin{bmatrix} \mathbf{0}_{1\times d} & 0 \\ \boldsymbol{I}_d & \mathbf{0}_{d\times 1} \end{bmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}, \;\; \boldsymbol{\xi} := [1, \mathbf{0}_{1\times d}]^T, \in \mathbb{R}^{d+1} \tag{4.13}$$

and the observation matrix $\boldsymbol{H}(n)$

$$\boldsymbol{H}(n) := [\boldsymbol{h}(n;0), \boldsymbol{h}(n;1), ..., \boldsymbol{h}(n;L), \mathbf{0}_{R\times(d-L)}] \in \mathbb{C}^{R\times(d+1)}, \tag{4.14}$$

where $\boldsymbol{h}(n;l) := [h^{(1)}(n,l), ..., h^{(R)}(n,l)]^T \in \mathbb{C}^R$. Based on these definitions and the system model in Section 4.2.1, we have the **time-invariant state equation**:

$$\boldsymbol{s}(n) = \boldsymbol{\Phi}\boldsymbol{s}(n-1) + \boldsymbol{\xi}s(n), \tag{4.15}$$

and the **time-varying observation equation**:

$$\boldsymbol{x}(n) = \boldsymbol{H}(n)\boldsymbol{s}(n) + \boldsymbol{\eta}(n). \tag{4.16}$$

The Kalman filter algorithm is given as follows [74], [59] where $\hat{\boldsymbol{s}}(n|n)$ denotes the linear MMSE estimate of $\boldsymbol{s}(n)$ based on observations $\boldsymbol{x}(k)$, $k = 1, 2, ..., n$.

Initialization: At time $n = 0$, $\hat{\boldsymbol{s}}(1|0) = E\{s(1)\} = 0$ and $\boldsymbol{V}_{\hat{s}}(1|0) = \sigma_s^2 \boldsymbol{I}_{d+1}$.

Filtering: At time $n = 1, 2...$

$$\boldsymbol{V}_z(n) = \boldsymbol{H}(n)\boldsymbol{V}_{\hat{s}}(n|n-1)\boldsymbol{H}^H(n) + \sigma_\eta^2 \boldsymbol{I}_R, \tag{4.17}$$

$$\boldsymbol{K}(n) = \boldsymbol{V}_{\hat{s}}(n|n-1)\boldsymbol{H}^H(n)\boldsymbol{V}_z^{-1}(n), \tag{4.18}$$

$$\boldsymbol{z}(n) = \boldsymbol{x}(n) - \boldsymbol{H}(n)\hat{\boldsymbol{s}}(n|n-1), \tag{4.19}$$

$$\hat{\boldsymbol{s}}(n|n) = \hat{\boldsymbol{s}}(n|n-1) + \boldsymbol{K}(n)z(n), \tag{4.20}$$

$$\boldsymbol{V}_{\hat{s}}(n|n) = [\boldsymbol{I}_{d+1} - \boldsymbol{K}(n)\boldsymbol{H}(n)]\boldsymbol{V}_{\hat{s}}(n|n-1), \tag{4.21}$$

$$\boldsymbol{V}_{\hat{s}}(n+1|n) = \boldsymbol{\Phi}\boldsymbol{V}_{\hat{s}}(n|n)\boldsymbol{\Phi}^H + \sigma_s^2 \boldsymbol{\xi}\boldsymbol{\xi}^H, \tag{4.22}$$

$$\hat{\boldsymbol{s}}(n+1|n) = \boldsymbol{\Phi}\hat{\boldsymbol{s}}(n|n). \tag{4.23}$$

Since $\hat{\boldsymbol{s}}(n|n) = [\hat{s}(n|n), \hat{s}(n-1|n), ..., \hat{s}(n-d|n)]^T$, we extract its last term $\hat{s}(n-d|n)$ as the desired equalized output.

Remark 4.3: For Kalman filter, the design process does not rely on a particular basis expansion model as long as the estimated/fitted channel impulse response is known at the receiver, so it can be applied to any doubly-selective channel model.

Table 4.1: Operation summary for Kalman filter

Operation	**flops**
$\boldsymbol{H}(n)\boldsymbol{V}_{\hat{s}}(n\|n-1)$	$(d+1)^2R$
$\boldsymbol{V}_z(n)$	$\frac{1}{2}(d+1)R^2+\frac{1}{2}(d+1)R$
$\boldsymbol{V}_z(n)^{-1}(n)$	$R^3+\frac{1}{2}R^2+\frac{1}{2}R$
$\boldsymbol{K}(n)$	$(d+1)R^2$
$\boldsymbol{V}_{\hat{s}}(n\|n)$	$\frac{1}{2}(d+1)^2R+\frac{1}{2}(d+1)R$
Total	$C=\frac{3}{2}(d+1)^2R+\frac{3}{2}(d+1)R^2+(d+1)R+R^3+\frac{1}{2}R^2+\frac{1}{2}R$

Remark 4.4: Only one parameter, equalization delay d, is needed for the Kalman filter design and will influence its performance.

Design Complexity: Computations in (4.17),(4.18),(4.21) and (4.22) comprise the design equations for Kalman filtering. Following Table 6.5 in [28], the required number of flops C per time-step are listed in Table 4.1. Therefore, for a transmitted block of size N symbols, one needs NC flops.

Implementation Complexity: Computations in (4.19),(4.20),and (4.23) comprise the implementation equations for Kalman filtering. The number of flops needed per time-step are $2R(d+1)$ where given the nature of $\mathbf{\Phi}$, (4.23) does not require any flops. Therefore, for a transmitted block of size N symbols, one needs $2NR(d+1)$ flops. On the other hand, both BLM and suboptimal FRESH equalizers require $NR(Q'+1)(L'+1)$ flops each.

4.2.4 Numerical Examples

In this section, the BER performance and the computational complexity of the proposed Kalman filter are investigated by means of Monte Carlo computer simulations, and compared with BEM-based BLM [5] and FRESH [68] equalizers. For computational complexity calculations, we assume

that the inverse of a positive-definite matrix has been computed via the modified Cholesky decomposition method.

A random time- and frequency- selective Rayleigh fading channel is simulated according to [81] with channel order $L = 3$ (4 taps). For different l's, $h\,(n;l)$'s are mutually independent, satisfy Jakes' model, and each tap is generated via the method of [81] given the symbol duration T_s and the Doppler spread f_d. It is important to point out that each channel tap follows the Jakes' spectrum, rather then the assumed BEM representation. The data were generated using the double-selective channel described above. However, for equalizer design one needs CE-BEM representation of the true channel; this was obtained by a least-squares fit of the assumed BEM to the true channel in each Monte Carlo run, just as in [5] and [68], to obtain the BEM coefficients (which vary from run-to-run). These BEM coefficients were used in the designs of [5] and [68], as well as in Kalman filtering where $\boldsymbol{h}(n;l)$ in (4.14) is generated via (4.2) with fitted BEM parameters (i.e. $\boldsymbol{h}(n;l)$ in (4.14) is not the true channel, but its BEM approximation). Note that one could have directly used the Jakes' channel $\boldsymbol{h}(n;l)$ in the Kalman filter implementation of the linear MMSE equalizer unlike the approaches of [5] and [68]. However, this would not be a fair comparison with the approaches of [5] and [68]. Most importantly, this would not illustrate the effects of modeling errors since practical channels are rarely Jakes' channels although their use in simulations and analysis is widespread.

In all simulations, the transmitter transmits binary phase shift keying (BPSK) and quadrature phase shift keying (QPSK) modulated symbols. The SNR refers to the energy per bit over one-sided noise spectral density.

Table 4.2: Computation complexity; $N = 50$, $P = 2N = 100$

Equalizer	Design Complexity(flops)	Implementation Complexity(flops)
FRESH-OPT [68]	$P \cdot \mathcal{O}[(L+L'+1)^3] + (P\log_2 P)(L+L'+1)^2 = 171,939$	$NRP(L'+1) = 70,000$
FRESH-SUBOPT [68]	$P \cdot \mathcal{O}[(L+L'+1)^3] + (P\log_2 P)(L+L'+1)^2 = 171,939$	$NR(Q'+1)(L'+1) = 13,300$
BLM [5]	$\mathcal{O}(K^3) = 12,193,565$	$NR(Q'+1)(L'+1) = 13,300$
Kalman filter (proposed)	$NC = 8,350$	$2NR(d+1) = 1,200$

Example 1: BER versus SNR under Simulation Parameters in [68]

We consider the same simulation parameters as in [[68], Experiment 3], i.e., the block size (number of information symbols) is $N = 50$, $P = 2N$ (oversampled CE-BEM with a factor of 2), the number of receive antennas $R = 2$, symbol duration $T_s = 160$us, the maximum Doppler spread $f_d = 100$Hz, the equalization delay $d = 5$ symbols and $Q = 2\lceil f_d P T_s \rceil = 4$. For BEM-based equalizers (both BLM and FRESH), the equalizer order $L' = 6$ and the number of Fourier coefficients (equalizer BEM coefficients) $Q' = 18$, as in [[68], Experiment 3]. The BER averaged over $10,000$ Monte Carlo runs versus SNR is shown in Fig. 4.1 for the four approaches: BLM [5], optimum FRESH [68], suboptimum FRESH [68] and the proposed Kalman filter solution. Note that the Kalman filter does not need L' or Q'. The results for BPSK modulation are plotted in Fig. 4.1. It is seen from Fig. 4.1 that the performances of the four approaches are very close each other. The computational complexity measured in terms of flops for entire block is shown in Table 4.2. Notice that the Kalman filter requires significantly fewer flops than the other two approaches, both for design as well as implementation.

Fig. 4.2 plots the BER versus SNR results for QPSK modulation, where the information sequences are QPSK signals. Although the performance of Kalman filter is slightly worse than the other two equalizers at high SNR, the three approaches have similar performances in this case.

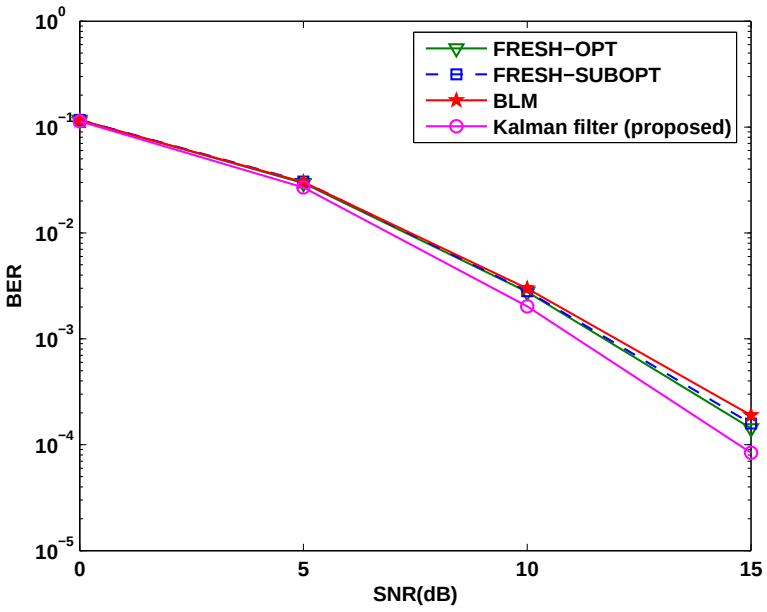

Figure 4.1: BER performance versus SNR using BPSK modulation, averaged over 10,000 runs. BLM denotes the method of [5]; FRESH-OPT and FRESH-SUBOPT are methods of [68]. $N = 50$, $P = 100$, $Q = 4$, $L = 3$, $d = 5$, $Q' = 18$, $L' = 6$

Table 4.3: Computation complexity; $N = 800$, $P = 2N = 1600$

Equalizer	Design Complexity(flops)	Implementation Complexity(flops)
FRESH-OPT [68]	$P \cdot \mathcal{O}[(L+L'+1)^3] + (P\log_2 P)(L+L'+1)^2 = 11,130,924$	$NRP(L'+1) = 33,280,000$
FRESH-SUBOPT [68]	$P \cdot \mathcal{O}[(L+L'+1)^3] + (P\log_2 P)(L+L'+1)^2 = 11,130,924$	$NR(Q'+1)(L'+1) = 274,400$
BLM [5]	$\mathcal{O}(K^3) = 37,989,672$	$NR(Q'+1)(L'+1) = 274,400$
Kalman filter (proposed)	$NC = 22,000$	$2NR(d+1) = 35,200$

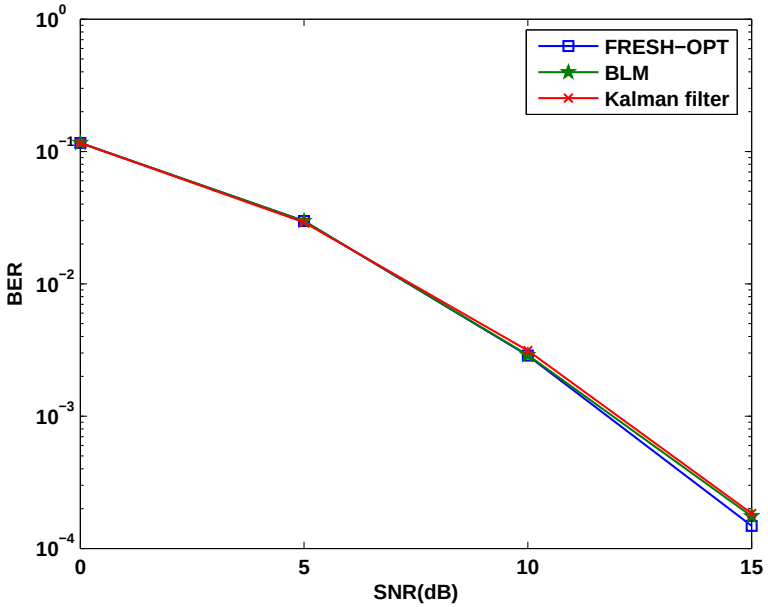

Figure 4.2: BER performance versus SNR using QPSK modulation, averaged over 10,000 runs

Example 2: BER versus SNR under Simulation Parameters in [5]

Now the simulation parameters in [[68], Fig. 11] are used where $N = 800$, $R = 2$, $T_s = 25$us, $f_d = 100$Hz, $d = 10$, $L' = 12$, and $Q' = 12$. Two cases are considered: $P = 2N$ and $P = N$, leading to Q is 8 and 4, respectively. The results are shown in Fig. 4.3 which were obtained by carrying out 1000 Monte Carlo runs. When $P = N$, the performance of the Kalman filter and BLM equalizer [5] are quite close to each other as in Fig. 4.1. The error floors at high SNR are caused by the BEM channel modeling error. However, the Kalman filter significantly outperforms the BLM equalizer [5] when $P = 2N$, the case where the channel modeling errors in approximating a Jakes' channel with CE-BEM are significantly smaller compared the case of $P = N$: recall that the data are generated via the true Jakes' channel. The computational complexity measured in terms of number of flops for entire block is shown in Table 4.3 for the case $P = 2N = 1600$. Notice

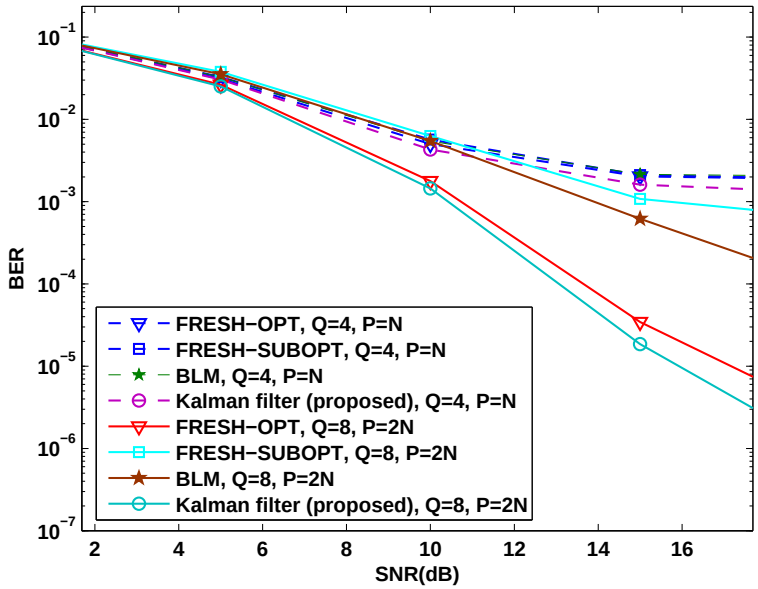

Figure 4.3: BER performance versus SNR using BPSK modulation, averaged over 1,000 runs. BLM denotes the method of [5]; FRESH-OPT and FRESH-SUBOPT are methods of [68]. Kalman filter is based on $P = 800$ when $Q = 4$ and on $P = 1600$ when $Q = 8$. $P = N$ corresponds to critically sampled CE-BEM and $P = 2N$ corresponds to oversampled CE-BEM. $N = 800$, $P = 800$ or 1600, $Q = 4$ or $Q = 8$, $L = 3$, $d = 10$, $Q' = 12$, $L' = 12$

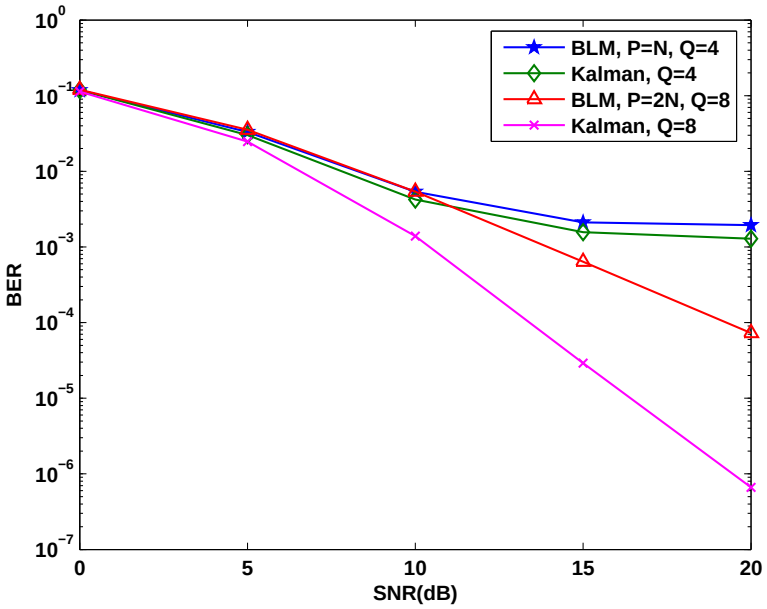

Figure 4.4: BER performance versus SNR using QPSK modulation, averaged over 1,000 runs

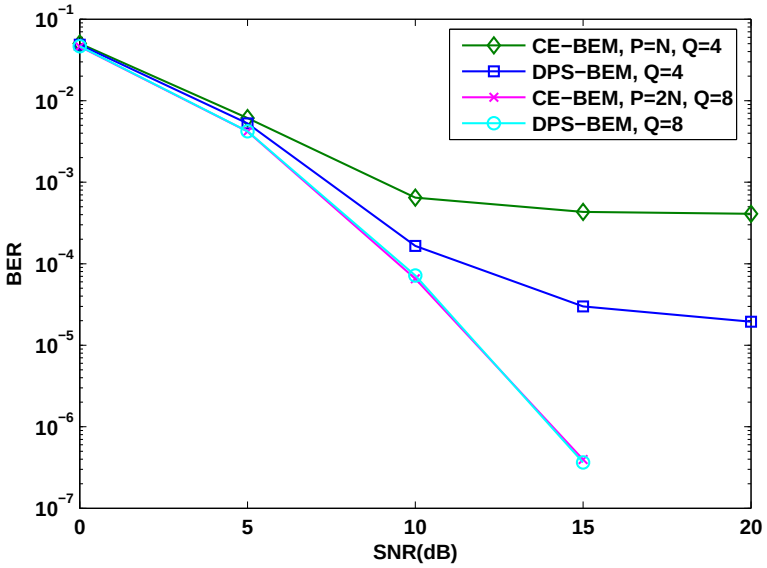

Figure 4.5: BER performance versus SNR For DPS-BEM and CE-BEM channel modeling using Kalman filtering, BPSK modulation

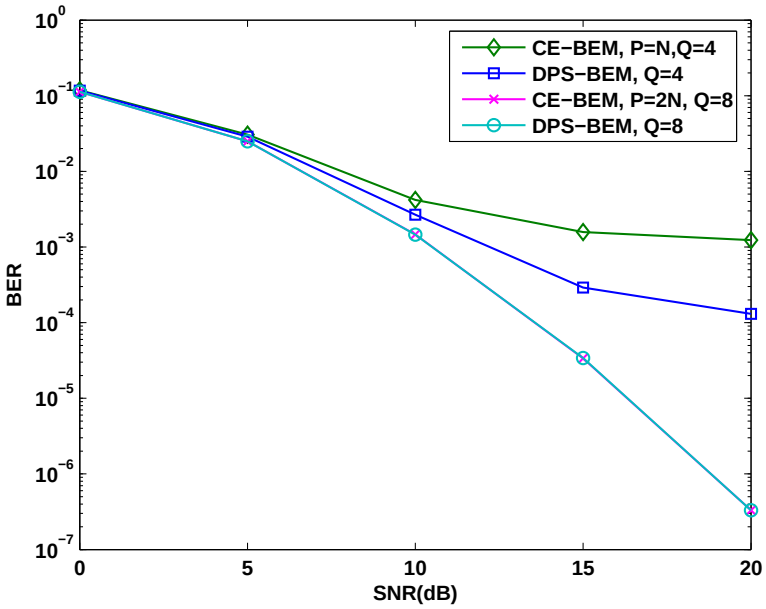

Figure 4.6: BER performance versus SNR For DPS-BEM and CE-BEM channel modeling using Kalman filtering, QPSK modulation

again that the Kalman filter requires significantly fewer flops than the other two approaches, both for design as well as implementation

In Fig. 4.4 the BER performances for QPSK modulation are plotted under two schemes: BLM and Kalman filter. The results in Fig. 4.4 are very similar to Fig. 4.3, except that the BER by using QPSK information symbols are worse than that using BPSK symbols, as expected.

Example 3: Performance of Kalman Filter Using DPS-BEM

Now we test the performance of the Kalman filer by using Discrete Prolate Spheroidal (DPS) BEM , in addition to CE-BEM, in (4.3), for approximating the true Jakes' channel. As we noted earlier, the Kalman filter has the flexibility that it is not restricted to a particular BEM and can be applied to any doubly-selective channel model. In Fig. 4.5, we compare the BER performances of

the Kalman filter for CE-BEM and DPS-BEM cases by using the simulation parameters in Example 2. It is known that the channel modeling error of DPS-BEM is several magnitudes smaller then that of critically sampled ($P = N$) CE-BEM [77], which is the reason why the BER performance of DPS-BEM case is better then CE-BEM for $P = N$, $Q = 4$. For the oversampling case of $P = 2N$ and $Q = 8$, the performances of DPS-BEM and CE-BEM are very close to each other since the modeling error differences are not so pronounced any more. This example shows the flexibility of the Kalman filter: its application is not limited to the CE-BEM channel model unlike FRESH [68] and BLM [5] equalizers.

Without loss of generality, in Fig. 4.6 we again plot the simulation results under QPSK modulation where the information symbols are QPSK signals.

4.3 Decision Feedback Equalization

4.3.1 System Model

Consider a doubly-selective finite impulse response (FIR) linear channel with single input and multiple outputs (SIMO), where the expressions of input and output of the SIMO channel are the same as (4.1)-(4.4) in Section 4.2.1.

Consider an FIR DFE with equalization delay d, feedforward (FF) filter of length l_f and feedback (FB) filter of length l_b. Let $\hat{s}(n)$ be the "soft" estimate of symbol $s(n)$ and let $\tilde{s}(n)$ denote its quantized (hard decision) value. Then the soft output of the DFE is given by

$$\hat{s}(n-d) = \sum_{m=0}^{l_f-1} \boldsymbol{f}_m^T(n)\boldsymbol{x}(n-m) - \sum_{k=1}^{l_b} b_k(n)\tilde{s}(n-d-k) \tag{4.24}$$

where the R-column vectors $\boldsymbol{f}_m(n)$'s and scalars $b_k(n)$'s are the tap-gains of FF and FB time-varying filters at time n. Note that both FF and FB filters are FIR. In the minimum mean-square error (MMSE) design of DFE, one seeks the tap-gains to minimize $E\{|s(n-d)-\hat{s}(n-d)|^2\}$. As discussed in [2] and [57, Sec. 3.3], one first designs the FB filter assuming that $\tilde{s}(n)=s(n)$ (i.e. no decision errors), and then one designs the FF filter given the FB filter.

4.3.2 Time-Varying FIR MMSE DFEs

In this section we first briefly review the TV FIR MMSE DFE from [46],[3]. Then our proposed formulation is presented. The design of these equalizers assumes availability of the CE-BEM coefficients $w_q^{(r)}(l)$s of the channel.

FIR MMSE DFE of [46],[3]

In [46],[3], one approximates time-varying FF and FB tap-gains via another sets of CE-BEMs, leading to

$$\hat{s}(n-d)=\sum_{l'=0}^{l_f-1}\sum_{q'=-\frac{Q'}{2}}^{\frac{Q'}{2}}\sum_{r=1}^{R}f_{q'}^{(r)}(l')e^{j\frac{2\pi qn}{P}}x^{(r)}(n-l')-\sum_{l''=1}^{l_b}\sum_{q''=-\frac{Q''}{2}}^{\frac{Q''}{2}}b_{q''}(l'')e^{j\frac{2\pi qn}{P}}\tilde{s}(n-d-l'') \tag{4.25}$$

where design parameters are d, l_f, l_b, Q', Q'' and equalizer FF coefficients $f_{q'}^{(r)}(l')$ and FB coefficients $b_{q''}(l'')$, invariant over $n\in\{0,1,\ldots,N-1\}$. Notice that the structure of the FF and FB parts of the equalizer is that of a CE-BEM.

The design equations may be found in [46],[3]; in particular, see [46, Sec. 5.2]. The TV FIR MMSE DFE proposed in [46],[3] (henceforth referred to as BLM: Barhumi, Leus, Moonen) provides an explicit "frequency-domain" representation, by turning a TV equalization problem into

a simpler time-invariant (TI) filtering design, which only involves the TI basis expansion parameters of the equalizer tap-gains. For details, we refer the reader to [46, Sec. 5.2].

We now provide details from [46, Sec. 5.2] (modified to conform to the notation in this paper). Define the $[l_b(Q''+1)+1] \times 1$ vector

$$\tilde{\boldsymbol{b}} = [b_{\frac{Q''}{2},l_b}, b_{\frac{Q''}{2},l_b-1}, \ldots, b_{\frac{Q''}{2},1}, \ldots, b_{1,1}, b_{0,l_b}, \ldots, b_{0,1}, 0, b_{-1,l_b}, \ldots, b_{-1,1}, \ldots, b_{\frac{-Q''}{2},1}]^T,$$

the $[l_f(Q'+1)] \times 1$ vector

$$\tilde{\boldsymbol{f}}^{(r)} = [f^{(r)}_{\frac{Q'}{2},l_f-1}, \ldots, f^{(r)}_{\frac{Q'}{2},0}, \ldots, f^{(r)}_{\frac{-Q'}{2},0}]^T,$$

and the $[Rl_f(Q'+1)] \times 1$ vector

$$\tilde{\boldsymbol{f}} = [\tilde{\boldsymbol{f}}^{(1)T}, \ldots, \tilde{\boldsymbol{f}}^{(R)T}]^T.$$

Define the $l_f \times [l_f + L]$ Toeplitz matrix

$$\boldsymbol{W}_q^{(r)} := \begin{bmatrix} w_q^{(r)}(L) & \cdots & w_q^{(r)}(0) & & \\ & \ddots & & \ddots & \\ & & w_q^{(r)}(L) & \cdots & w_q^{(r)}(0) \end{bmatrix},$$

the $[(Q'+1)l_f] \times [(Q+Q'+1)(l_f+L)]$ matrix

$$\boldsymbol{\mathcal{H}}^{(r)} := \begin{bmatrix} \boldsymbol{\Omega}^{\frac{Q}{2}}\boldsymbol{W}_q^{(r)} & \cdots & \boldsymbol{\Omega}^{\frac{-Q}{2}}\boldsymbol{W}_q^{(r)} & & \\ & \ddots & & \ddots & \\ & & \boldsymbol{\Omega}^{\frac{Q}{2}}\boldsymbol{W}_q^{(r)} & \cdots & \boldsymbol{\Omega}^{\frac{-Q}{2}}\boldsymbol{W}_q^{(r)} \end{bmatrix}$$

and the $[R(Q'+1)l_f] \times [(Q+Q'+1)(l_f+L)]$ matrix

$$\boldsymbol{\mathcal{H}} := \left[\boldsymbol{\mathcal{H}}^{(1)T}, \ldots, \boldsymbol{\mathcal{H}}^{(R)T}\right]^T,$$

where $\boldsymbol{\Omega} := \mathrm{diag}\{1, e^{j2\pi/P}, \ldots, e^{j2\pi(l_f-1)/P}\}$. Define a $[(Q''+1)l_f] \times [(Q+Q'+1)(L+l_f)]$ "selection" matrix

$$\boldsymbol{P} := \begin{bmatrix} & \boldsymbol{I}_{\frac{Q''}{2}} \otimes \boldsymbol{P}_1 & & \\ \boldsymbol{0}_{\alpha\times\beta} & & \boldsymbol{P}_2 & & \boldsymbol{0}_{\alpha\times\beta} \\ & & & \boldsymbol{I}_{\frac{Q''}{2}} \otimes \boldsymbol{P}_1 & \end{bmatrix}$$

with $\alpha = (Q''+1)l_b + 1$, $\beta = (Q+Q'-Q'')(L+l_f)/2$,

$$\boldsymbol{P}_1 := [\boldsymbol{0}_{l_b\times[L+l_f-1-l_b-d)]}, \boldsymbol{I}_{l_b}, \boldsymbol{0}_{l_b\times(d+1)}],$$

$$\boldsymbol{P}_2 := [\boldsymbol{0}_{(l_b+1)\times[L+l_f-1-l_b-d)]}, \boldsymbol{I}_{l_b+1}, \boldsymbol{0}_{(l_b+1)\times d}].$$

Let $\boldsymbol{e}$ denote the $[(Q''+1)l_b+1] \times 1$ unit vector with a one in position $1 + l_b + (Q''l_b/2)$. Then the MMSE DFE tap-gains are given by [46, Sec. 5.2]

$$\tilde{\boldsymbol{b}}_{MMSE} = \frac{\boldsymbol{R}_{MMSE}^{-1}\boldsymbol{e}}{\boldsymbol{e}^T\boldsymbol{R}_{MMSE}^{-1}\boldsymbol{e}} - \boldsymbol{e}, \tag{4.26}$$

$$\boldsymbol{R}_{MMSE} = \boldsymbol{P}\left(\sigma_\eta^{-2}\boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{H}} + \sigma_s^{-2}\boldsymbol{I}\right)^{-1}\boldsymbol{P}^H, \tag{4.27}$$

$$\tilde{\boldsymbol{f}}_{MMSE}^T = \tilde{\boldsymbol{b}}_{MMSE}^T\boldsymbol{P}\left(\boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{H}} + \frac{\sigma_\eta^2}{\sigma_s^2}\boldsymbol{I}\right)^{-1}\boldsymbol{\mathcal{H}}^H. \tag{4.28}$$

Computational Complexity: To design the feedforward and feedback filter coefficients, the inverse in (4.28) and (4.26) require $2\cdot\mathcal{O}(K^3)$ flops (one flop is roughly one multiply-and-accumulate

operation) where $K = (Q+Q'+1)(L+l_f)$ and $\mathcal{O}(K^3) = K^3+0.5K^2+0.5K$ if one uses modified Cholesky decomposition (i.e. UD-decomposition) based approach to matrix inverse [2, Table 6.13]. This has been called design complexity (equalizer design) in [3]. For implementation of (4.25), one needs $NR(Q'+1)l_f$ flops for feedforward part and $N(Q''+1)l_b$ flops for feedback part; this has been called implementation complexity in [3]. For numerical comparisons in Section 4.3.3, we assume that the inverse of a positive-definite matrix has been computed via the modified Cholesky decomposition method.

Remark 4.5: The design of BLM DFEs relies on CE-BEM representation of the channel. They assume that the BEM parameters $w_q^{(r)}(l)$ are known at the receiver. It is not clear if these design methods apply to other basis expansion models. However, the simulation results in [5] illustrate that the BEM modeling errors have a significant influence on the equalizer performance.

Remark 4.6: Five parameters have to be considered or optimized for BLM DFE design: d, l_f, l_b, Q' and Q''.

Proposed MMSE DFE with Time-Varying Taps

Here we follow [2] (see also [57, Sec. 3.3]) where a time-invariant channel is considered. Their results extend to time-varying channels in a straight-forward fashion; therefore, we simply state the final results instead of repeating the entire derivation with obvious changes.

Using the estimated channel, the symbol decisions are made by an FIR MMSE-DFE [2]. Given the lengths of the feedforward (FF) and the feedback (FB) filters as l_f and l_b respectively, the estimate of the information symbol at time n with equalization delay d is given by (4.24). Stack the

inputs of the FF filter at time n into a "tall" vector

$$\boldsymbol{x}_f(n) := \begin{bmatrix} \boldsymbol{x}^T(n) & \boldsymbol{x}^T(n-1) & \cdots & \boldsymbol{x}^T(n-l_f+1) \end{bmatrix}^T,$$

and also define $\boldsymbol{\eta}_f(n)$ likewise. By (4.4), we have

$$\boldsymbol{x}_f(n) = \boldsymbol{H}(n)\,\boldsymbol{s}_f(n) + \boldsymbol{\eta}_f(n) \tag{4.29}$$

where

$$\boldsymbol{H}(n) := \begin{bmatrix} \boldsymbol{h}(n;0) & \cdots & \boldsymbol{h}(n;L) & & \\ & \ddots & \ddots & \ddots & \\ & & \boldsymbol{h}(n-l_f+1;0) & \cdots & \boldsymbol{h}(n-l_f+1;L) \end{bmatrix} \tag{4.30}$$

and

$$\boldsymbol{s}_f(n) := \begin{bmatrix} s(n) & s(n-1) & \cdots & s(n-l_f-L+1) \end{bmatrix}^T.$$

We further define

$$\mathbf{s}_b(n) := \begin{bmatrix} s(n-d) & s(n-d-1) & \cdots & s(n-d-l_b) \end{bmatrix}^T.$$

By the assumption that $\{s(n)\}$ are i.i.d. with variance σ_s^2, and based on (4.29), we have

$$\begin{aligned} \boldsymbol{R}_{ss}(n) &:= E\left\{\boldsymbol{s}_b(n)\,\boldsymbol{s}_b^H(n)\right\} = \sigma_s^2 \boldsymbol{I}_{l_b+1}, \\ \boldsymbol{R}_{sx}(n) &:= E\left\{\boldsymbol{s}_b(n)\,\boldsymbol{x}_f^H(n)\right\} = \sigma_s^2 \boldsymbol{\Phi}\boldsymbol{H}^H(n), \\ \boldsymbol{R}_{xx}(n) &:= E\left\{\boldsymbol{x}_f(n)\,\boldsymbol{x}_f^H(n)\right\} = \sigma_s^2 \boldsymbol{H}(n)\,\boldsymbol{H}^H(n) + \sigma_v^2 \boldsymbol{I}_{Rl_f} \end{aligned}$$

where $\boldsymbol{\Phi} := \begin{bmatrix} \mathbf{0}_{(l_b+1)\times d} & \boldsymbol{I}_{l_b+1} & \mathbf{0}_{(l_b+1)\times\left(l_f+L-d-l_b-1\right)} \end{bmatrix}$.

Let $\boldsymbol{f}\left(n\right)$ and $\boldsymbol{b}\left(n\right)$ denote the vectors of time-varying taps of FF and FB filters,

$$\boldsymbol{f}\left(n\right) := \begin{bmatrix} \boldsymbol{f}_0^T\left(n\right) & \boldsymbol{f}_1^T\left(n\right) & \cdots & \boldsymbol{f}_{l_f-1}^T\left(n\right) \end{bmatrix}^T,$$
$$\boldsymbol{b}\left(n\right) := \begin{bmatrix} 1 & b_1\left(n\right) & b_2\left(n\right) & \cdots & b_{l_b}\left(n\right) \end{bmatrix}^T.$$

Assuming the decisions $\{\tilde{s}\left(n\right)\}$ are correct and equal to $\{s\left(n\right)\}$, the FF and the FB time-varying filters of the MMSE-DFE are given by [2]

$$\boldsymbol{b}_{\mathrm{MMSE}}\left(n\right) = \frac{\boldsymbol{R}_\delta^{-1}\boldsymbol{e}_0}{\boldsymbol{e}_0^T\boldsymbol{R}_\delta^{-1}\boldsymbol{e}_0}, \tag{4.31}$$
$$\boldsymbol{f}_{\mathrm{MMSE}}\left(n\right) = \boldsymbol{R}_{xx}^{-1}\left(n\right)\boldsymbol{R}_{sx}^H\left(n\right)\boldsymbol{b}_{\mathrm{MMSE}}\left(n\right) \tag{4.32}$$

where $\boldsymbol{e}_0 := \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}^T$ and

$$\boldsymbol{R}_\delta := \boldsymbol{R}_{ss}\left(n\right) - \boldsymbol{R}_{sx}\left(n\right)\boldsymbol{R}_{xx}^{-1}\left(n\right)\boldsymbol{R}_{sx}^H\left(n\right). \tag{4.33}$$

Computational Complexity: To design the proposed MMSE DFE, we need to compute the inverse of a $Rl_f \times Rl_f$ matrix to obtain the feedforward filter coefficients (as in eq. (4.32)), and we need to compute the inverse of a $(l_b+1)\times(l_b+1)$ matrix to obtain the feedback filter coefficients (as in eq.(4.31)). Therefore, the required number of flops is $\mathcal{O}((Rl_f)^3) + \mathcal{O}((l_b+1)^3)$.

The implementation complexity associated with eq. (4.24) requires NRl_f flops for the feedforward part and Nl_b flops for the feedback part.

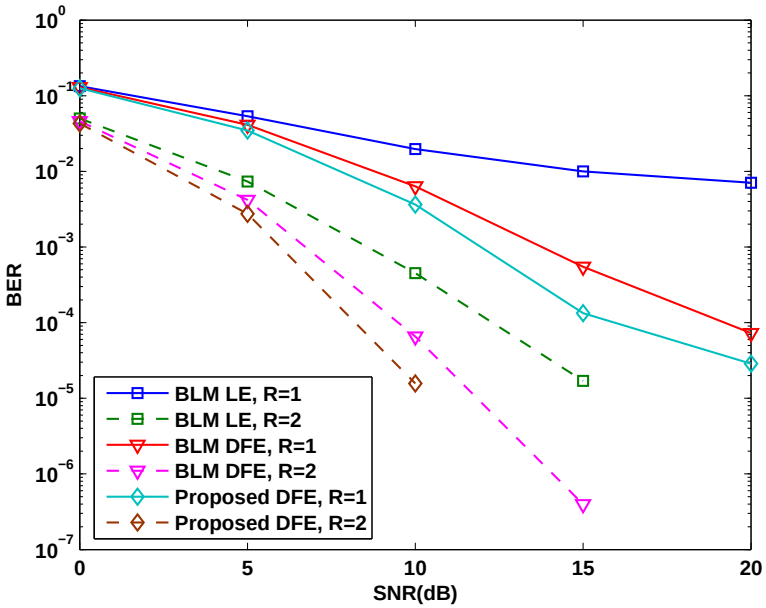

Figure 4.7: BER performance versus SNR, averaged over 1,000 runs, BPSK signal. $N = 800$, $P = 1600$, $Q = 8$, $L = 3$, d=5, Q'=12, l_f=12, $Q'' = 4$, $l_b = 3$. R denotes the number of receive antennas.

Remark 4.7: For the proposed MMSE DFE, the design process does not rely on a particular basis expansion model as long as the estimated/fitted channel impulse response is known at the receiver, so it can be applied to any doubly-selective channel model.

Remark 4.8: Only three parameters, equalization delay d and filter lengths l_f and l_b, are needed for the proposed filter design and will influence its performance.

4.3.3 Numerical Examples

In this section, the BER performance of the proposed DFE solution are investigated by means of Monte Carlo computer simulations, and compared with BEM-based BLM DFE solution of [46],[3].

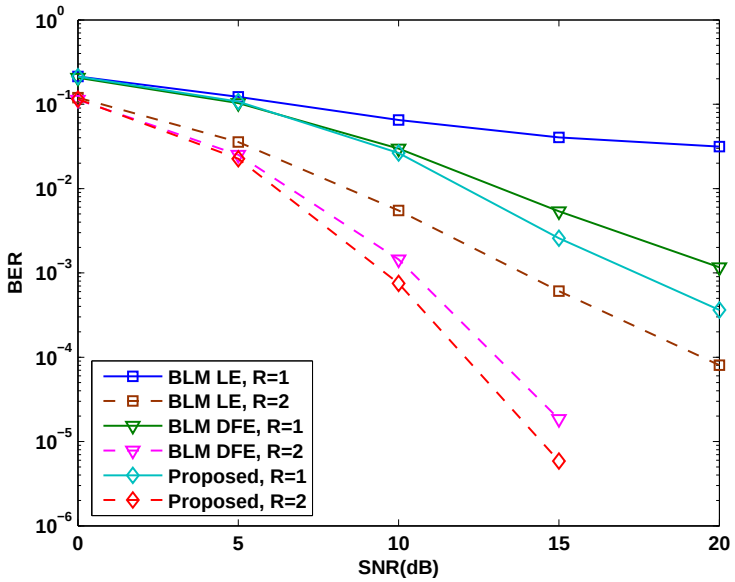

Figure 4.8: BER performance versus SNR, averaged over 1,000 runs, BPSK signal. $N = 800$, $P = 1600$, $Q = 8$, $L = 3$, d=5, Q'=12, l_f=12, $Q'' = 4$, $l_b = 3$. R denotes the number of receive antennas.

A random time- and frequency- selective Rayleigh fading channel is simulated according to [81] with channel order $L = 3$ (4 taps). For different l's, $h\,(n;l)$'s are mutually independent, satisfy Jakes' model, and each tap is generated via the method of [81] given the symbol duration T_s and the Doppler spread f_d. It is important to point out that each channel tap follows the Jakes' spectrum, rather then the assumed BEM representation. The data were generated using the double-selective channel described above. However, for equalizer design one needs CE-BEM representation of the true channel; this was obtained by a least-squares fit of the assumed BEM to the true channel in each Monte Carlo run, just as in [46], [5] and [68], to obtain the BEM coefficients (which vary from run-to-run). These BEM coefficients were used in the designs of [46],[3] and [5], as well as in our proposed DFE solution.

Table 4.4: Computation complexity

Equalizer	Design Flops	Implementation Flops
BLM DFE [46],[3]	62,611,290	261,600
Proposed DFE	14,198	21,600

In our simulations, the transmitter transmits binary or quaternary phase shift keying (BPSK/ QPSK) modulated symbols. The SNR refers to the energy per symbol over one-sided noise spectral density.

We consider the simulation parameters quite similar to those in [46],[3] and [5, Fig. 11], i.e., the block size (number of information symbols) is $N = 800$, $P = 2N$ (oversampled CE-BEM with a factor of 2), the number of receive antennas $R = 1, 2$, symbol duration $T_s = 25$ μs, the maximum Doppler spread $f_d = 100Hz$, the equalization delay $d = 5$ symbols and $Q = 2\lceil f_d P T_s \rceil = 8$. For BEM-based equalizers (both BLM LE and BLM DFE), the equalizer lengths $l_f = 12$ and $l_b = 3$ with corresponding number of Fourier coefficients (equalizer BEM coefficients) $Q' = 12$ and $Q'' = 4$. The BER averaged over 1,000 Monte Carlo runs versus SNR is shown in Fig. 4.7 for the three approaches when using BPSK signal: BLM DFE [46],[3], BLM LE [5], and our proposed DFE solution. Note that our formulation does not need Q' or Q''. It is seen from Fig. 4.7 that the DFE solutions outperform the LE solution, and the performance improves with increasing number of receive antennas R. Furthermore, our proposed solution outperforms the BEM-based DFE of [46],[3]. The computational complexity measured in terms of number of flops for the entire block is shown in Table 4.4. Notice that our DFE formulation requires far fewer flops than the BEM-based DFE of [46],[3].

Finally, Fig. 4.8 shows the BER results for QPSK signals for the same set of parameters as for Fig. 4.7. We see comparative performance similar to Fig. 4.7 in Fig. 4.8.

4.4 Conclusion

In this chapter, we exploited the conventional Kalman filter and DFEs as time-varying minimum mean-square-error equalizer for doubly-selective channels modeled via basis expansion models (BEM). Recently there has been much interest in designing time-variant serial FIR (finite impulse response) linear equalizers and FIR DFEs using complex exponential (CE) BEMs for equalizers in addition to using CE-BEM for modeling the channel itself. We showed that a Kalman filter formulation of the linear equalizer and an alternative formulation of the FIR DFE based on a CE-BEM channel model yields same or improved BER at a much lower computational cost, without incurring the approximation error inherent in CE-BEM modeling of equalizers.

CHAPTER 5

A MULTIPLE MODEL APPROACH TO DOUBLY-SELECTIVE CHANNEL ESTIMATION USING EXPONENTIAL BASIS MODELS

5.1 Introduction

In order to "accurately" model the underlying doubly-selective channel, the number of BEM coefficients used to model the doubly-selective channels for channel estimation has been based on an upper bound on the channel Doppler spread. The higher the Doppler spread, the more the number of BEM coefficients, which leads to a higher channel estimation variance. This, in turn, leads to higher bit error rate (BER) when the estimated channel is used for data detection and the actual Doppler spread is (much) less than the upper bound.

In this chapter we propose to use a multiple model framework where several candidate Doppler spread values are used to cover the range from zero to an upper bound, leading to multiple CE-BEM channel models, each corresponding to an assumed value of the Doppler spread. Subsequently the well known interacting multiple model (IMM) algorithm [11] is used for symbol detection based on multiple state-space models corresponding to the multiple estimated channels.

The multiple model (or hybrid system) approach assumes the system to be in one of a finite number of models (i.e., that is described by one out of a finite number of models). Each model is characterized by its parameters — the models differ in Doppler spread here. For each model a filter "matched" to its parameters is yielding model-conditioned estimates and covariances. A mode probability calculator — a Bayesian model comparator — updates the probability of each mode using the likehood function (innovations) of each filter and the prior probability of each model. Then an estimate combiner computes the overall estimate and the associated covariance as the weighted

sum of the model-conditioned estimates and the corresponding covariance — via the (Gaussian) mixture equations. Among all the multiple model approaches, the interacting multiple model is considered to be the best compromise between complexity and performance. The IMM approach computes the state estimate that accounts for each possible current model using a suitable mixing of the previous model conditioned estimate depending on the current model.

In this chapter an adaptive channel estimation scheme, exploiting the over-sampled complex exponential basis expansion model (CE-BEM), is presented for doubly-selective channels where we track the BEM coefficients via the multiple model approach. The chapter is organized as follows. First, the system model and objectives are provided. Second, the IMM algorithm and configuration are briefly summarized. The performances of the proposed design are finally illustrated by simulation examples.

5.2 System Model and Objectives

5.2.1 System Model

Consider a doubly-selective (time- and frequency-selective) FIR (finite impulse response) linear channel. Let $\{s\,(n)\}$ denotes a scalar sequence which is the input to time-varying channels with the discrete-time response $\{h\,(n;l)\}$ (the channel response at time n to a unit input at time $n-l$). Then the symbol-rate noisy channel output at the rth receive antenna is given by ($n = 0, 1, \ldots$; $r = 1, 2, \ldots, R$)

$$y^{(r)}(n) = \sum_{l=0}^{L} h^{(r)}(n;l)s(n-l) + \eta^{(r)}(n) \tag{5.1}$$

where $v^{(r)}(n)$ is the zero-mean white complex-Gaussian noise with variance σ_v^2. We assume that $\left\{h^{(r)}(n;l)\right\}$ ($r = 1, \ldots, R$) represents a wide-sense stationary uncorrelated scattering (WSSUS) vector channel [54].

In CE-BEM [26, 48, 37], over the k-th block consisting of an observation window of T_B symbols, the channel is represented as ($\bar{n}_k := (k-1)T_B$)

$$h^{(r)}(n;l) = \sum_{q=1}^{Q} w_q^{(r)}(l) e^{j\omega_q n},\ n = \bar{n}_k, \ldots, \bar{n}_k + T_B - 1, \tag{5.2}$$

where one chooses ($l = 0, 1, \ldots, L$, and K is an integer)

$$T := KT_B, \quad K \geq 1, \qquad Q \geq 2\left\lceil f_d T T_s \right\rceil + 1, \tag{5.3}$$

$$\omega_q := \frac{2\pi}{T}\left[q - (Q+1)/2\right], \quad q = 1, 2, \ldots, Q, \tag{5.4}$$

$$L := \left\lfloor \tau_d / T_s \right\rfloor, \tag{5.5}$$

τ_d and f_d are respectively the delay spread and the Doppler spread, and T_s is the symbol duration. The BEM coefficients $w_q^{(r)}(l)$'s remain invariant during this block, but are allowed to change at the next block, and the Fourier basis functions $\left\{e^{j\omega_q n}\right\}$ ($q = 1, 2, \ldots, Q$) are common for each block. If the delay spread τ_d and the Doppler spread f_d of the channel (or at least their upper-bounds) are known, one can infer the basis functions of the CE-BEM [48]. Treating the basis functions as known, estimation of a time-varying process is reduced to estimating the invariant coefficients over a block of length T_B symbols. Note that the BEM period is $T = KT_B$ whereas the block size is T_B symbols. If $K > 1$ (e.g. $K = 2$ or $K = 3$), then the Doppler spectrum is said to be over-sampled [37] compared to the case $K = 1$ where the Doppler spectrum is said to be critically sampled. In

[26, 48] only $K = 1$ (henceforth called CE-BEM) is considered whereas [37] considers $K \geq 2$ (henceforth called over-sampled CE-BEM).

5.2.2 Block-Adaptive Channel Estimation [48]

Here we summarize the time-multiplexed training approach of [48]. In Sec. 5.4 we provide simulation comparisons with results of [48]. In [48] each transmitted block of symbols $\{s(n)\}_{n=0}^{T_B-1}$ is segmented into P subblocks of time-multiplexed training and information symbols. Each sub-block is of equal length l_b symbols with l_d information symbols and l_t training symbols ($l_b = l_d + l_t$). If $\boldsymbol{s}$ denotes a column-vector composed of $\{s(n)\}_{n=0}^{T_B-1}$, then $\boldsymbol{s}$ is arranged as

$$\boldsymbol{s} := \begin{bmatrix} \boldsymbol{b}_0^T & \boldsymbol{c}_0^T & \boldsymbol{b}_1^T & \boldsymbol{c}_1^T & \cdots & \boldsymbol{b}_{P-1}^T & \boldsymbol{c}_{P-1}^T \end{bmatrix}^T \tag{5.6}$$

where $\boldsymbol{b}_p$ ($p = 0, 1, \cdots P-1$) is a column of l_d information symbols and $\boldsymbol{c}_p$ is a column of l_t training symbols. We clearly have $T_B = Pl_b$. Given (5.1) and CE-BEM (5.2), [48] has shown that (5.6) is an optimum structure for $K = 1$ with $l_t = 2L + 1$, $P \geq Q$ and

$$\boldsymbol{c}_p := \begin{bmatrix} \boldsymbol{0}_L^T & \gamma & \boldsymbol{0}_L^T \end{bmatrix}^T, \ \gamma > 0. \tag{5.7}$$

Thus, given a transmission block of size T_B, $(2L+1)P$ symbols have to be devoted to training and the remaining $T_B - (2L+1)P$ are available for information symbols. This design has been used by others for oversampled CE-BEM also [37].

Let $n_p := pl_b + l_d + L$, ($p = 0, 1, \cdots P-1$), denote the location of (nonzero) γ's in the optimum $\boldsymbol{c}_p$'s in the P subblocks. Then by design, received signal (assuming timing synchronization)

$$y^{(r)}(n_p + l) = \gamma h^{(r)}(n_p + l; l) + v^{(r)}(n_p + l) \tag{5.8}$$

for $l = 0, 1, \cdots, L$. Using (5.2) in these $y^{(r)}(n_p + l)$'s, one can uniquely solve for $w_q^{(r)}(l)$'s via a least-squares approach. The channel estimates are given by the CE-BEM (5.2) using the estimated BEM coefficients.

5.2.3 Objectives

Suppose that we collect the received signal over a time interval of $\bar{T}$ symbols. We wish to estimate the time-variant channel using a channel model and time-multiplexed training (such as that discussed in Sec. 5.2.2 and [48]), and subsequently using the estimated channel, estimate the information symbols. For CE-BEM, if we choose $\bar{T}$ as the block size, then in general Q value will be very high requiring estimation of a large number of parameters, thereby degrading the channel estimation performance. If we divide $\bar{T}$ into blocks of size T_B, and then fit CE-BEM block by block, we need smaller Q. This is the solution considered in this chapter (and also [48]). In practical situations, over a large $\bar{T}$, the actual Doppler spread f_d is likely to vary. Absent any prior knowledge, a commonly used solution [48, 37] is to use an upper bound on the anticipated f_d (based on the maximum vehicle speed, e.g.) and pick Q accordingly. In this chapter we investigate a multiple model framework where several candidate Doppler spread values are used to cover the range from zero to an upper bound, leading to multiple CE-BEM channel models, each corresponding to an assumed value of the Doppler spread. Multiple model approach has been extensively used in target tracking applications [11, 7, 64] and more recently, has been used for tracking dispersive DS-CDMA

channels using multiple autoregressive (AR) models in [30]. In this chapter we propose to use such an approach in conjunction with BEM's.

5.3 Multiple Model Approach

5.3.1 Multiple Models

Let $f_{d,u}$ denote an upper bound on the anticipated Doppler spread f_d. Let $f_{d,1}$, $f_{d,2}$,..., $f_{d,M}$ denote our M candidate Doppler spreads and let Q_m, $1 \leq m \leq M$, denote the corresponding values of Q from (5.3). Then we have M candidate channel impulse responses indexed by m over the k-th block consisting of an observation window of T_B symbols,

$$h^{(m,r)}(n;l) = \sum_{q=0}^{Q_m} w_q^{(m,r)}(l) e^{j\omega_q n}, n = \bar{n}_k, \bar{n}_k + 1, ..., \bar{n}_k + T_B - 1. \tag{5.9}$$

We will use a Kalman filter with equalization delay d for data detection using the estimated channel. Define

$$\begin{aligned}
\boldsymbol{y}(n) &:= [y^{(1)}(n),\ y^{(2)}(n),\ ...,\ y^{(R)}(n)]^T, \\
\boldsymbol{s}_d(n) &:= [s(n), s(n-1), ..., s(n-d)]^T, \\
\bar{s}(n) &:= E\{s(n)\},\ \ \tilde{s}(n) := s(n) - \bar{s}(n), \\
\boldsymbol{\Phi} &:= \begin{bmatrix} \mathbf{0}_{1\times d} & 0 \\ \boldsymbol{I}_d & \mathbf{0}_{d\times 1} \end{bmatrix}, \\
\boldsymbol{\Gamma} &:= [1, \mathbf{0}_{1\times d}]^T, \\
\boldsymbol{H}_d(n) &:= [\boldsymbol{h}^{(m)}(n;0), \boldsymbol{h}^{(m)}(n;1), ..., \boldsymbol{h}^{(m)}(n;L), \mathbf{0}_{R\times(d-L)}], \\
\boldsymbol{h}^{(m)}(n;l) &:= [h^{(m,1)}(n,l), ..., h^{(m,R)}(n,l)]^T,
\end{aligned} \tag{5.10}$$

where $\boldsymbol{\eta}(n)$ is defined just as $\boldsymbol{y}(n)$ and integer $d \geq L$. Assume data symbols are zero-mean and white. If $s\,(n)$ is a data symbol, we have $\bar{s}\,(n) := 0$, $\tilde{s}\,(n) := s\,(n)$; if $s\,(n)$ is a training symbol, $\bar{s}\,(n) := s\,(n)$, $\tilde{s}\,(n) := 0$. Then the underlying state-space model corresponding to the mth channel is given by the state and the measurement equations

$$\boldsymbol{s}_d\,(n) = \boldsymbol{\Phi}\boldsymbol{s}_d\,(n-1) + \boldsymbol{\Gamma}\bar{s}\,(n) + \boldsymbol{\Gamma}\tilde{s}\,(n)\,, \tag{5.11}$$

$$\boldsymbol{y}\,(n) = \boldsymbol{H}_d^{(m)}\,(n)\,\boldsymbol{s}_d\,(n) + \boldsymbol{v}\,(n)\,. \tag{5.12}$$

In (5.11) $\bar{\boldsymbol{s}}(n)$ and $\tilde{\boldsymbol{s}}(n)$ are defined just as $\boldsymbol{s}_d(n)$.

Consider a set of T_B received symbols divided up into P subblocks as in Sec. 5.2.2. For model m, we estimate the BEM coefficients $w_q^{(m,r)}(l)$ via the least-squares approach of Sec. 5.2.2 using the training symbols. Then the estimated channel for the mth model is given by $\hat{h}^{(m,r)}(n;l) = \sum_{q=1}^{Q_m} \hat{w}_q^{(m,r)}(l)e^{j\omega_q n}$.

5.3.2 Interacting Multiple Model Data Detection

Using the M estimated channels from each block of received symbols, we obtain the M models with state equation (5.11) and measurement equation

$$\boldsymbol{y}(n) = \hat{\boldsymbol{H}}_d^{(m)T}\,(n)\,\boldsymbol{s}_d\,(n) + \boldsymbol{\eta}\,(n)\,, \tag{5.13}$$

where $\hat{\boldsymbol{H}}_d^{(m)}\,(n)$ is as in Section 5.3.1 with $h^{(m)}\,(n;l)$ replaced with estimated $\hat{h}^{(m)}\,(n;l)$. Now our task is to estimate $\boldsymbol{s}_d(n)$ given $\boldsymbol{y}(k)$, $k \leq n$, and the M models specified by (5.11) and (5.13). In (5.13) we treat $\hat{\boldsymbol{H}}_d^{(m)}(n)$ as true $\boldsymbol{H}_d^{(m)}(n)$.

Table 5.1: Summary of the IMM algorithm (one cycle)

Interaction ($i, j = 1, 2, \cdots M$):

predicted mode probability: $\mu_j^-(k) = \sum_i p_{ij}\mu_i(k-1)$

mixing probability: $\mu_{i|j} = p_{ij}\mu_i(k-1)/\mu_j^-(k)$

$\hat{\boldsymbol{s}}_{0dj}(k-1|k-1) = \sum_i \hat{\boldsymbol{s}}_{di}(k-1|k-1)\mu_{i|j}$

$\boldsymbol{V}_{0dj}(k-1|k-1) = \sum_i \boldsymbol{V}_{di}(k-1|k-1)\mu_{i|j} + \boldsymbol{X}_j$

where the "spread-of-the-means" term in the mixing is

$$\boldsymbol{X}_j = \sum_i [\hat{\boldsymbol{s}}_{di}(k-1|k-1) - \hat{\boldsymbol{s}}_{0dj}(k-1|k-1)] \times [\hat{\boldsymbol{s}}_{di}(k-1|k-1) - \hat{\boldsymbol{s}}_{0dj}(k-1|k-1)]^H \mu_{i|j}$$

Filtering ($i, j = 1, 2, \cdots M$):

$\hat{\boldsymbol{s}}_{dj}(k|k-1) = \boldsymbol{\Phi}\hat{\boldsymbol{s}}_{0dj}(k-1|k-1) + \boldsymbol{\Gamma}\bar{\boldsymbol{s}}(k)$

$\boldsymbol{V}_{dj}(k|k-1) = \boldsymbol{\Phi}\boldsymbol{V}_{0dj}(k-1|k-1)\boldsymbol{\Phi}^H + \sigma_{\tilde{s}}^2\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$

measurement residual: $\boldsymbol{z}_j = \boldsymbol{y}(k) - \boldsymbol{H}_{dj}\hat{\boldsymbol{s}}_{dj}(k|k-1)$

residual cov.: $\boldsymbol{D}_j = \boldsymbol{H}_d^{(j)}\boldsymbol{V}_{dj}(k|k-1)\boldsymbol{H}_d^{(j)H} + \sigma_v^2\boldsymbol{I}_R$

fliter gain: $\boldsymbol{G}_j = \boldsymbol{V}_{dj}(k|k-1)\boldsymbol{H}_d^{(j)H}\boldsymbol{D}_j^{-1}$

$\hat{\boldsymbol{s}}_{dj}(k|k) = \hat{\boldsymbol{s}}_{dj}(k|k-1) + \boldsymbol{G}_j\boldsymbol{z}_j$

$\boldsymbol{V}_{dj}(k|k) = \boldsymbol{V}_{dj}(k|k-1) - \boldsymbol{G}_j\boldsymbol{D}_j\boldsymbol{G}_j^H$

likelihood function: $\boldsymbol{\Lambda}_j = [\det(\pi\boldsymbol{D}_j)]^{-1}e^{-\boldsymbol{z}_j^H\boldsymbol{D}_j^{-1}\boldsymbol{z}_j}$

mode probability: $\mu_j(k) = \frac{\mu_j^-\boldsymbol{\Lambda}_j}{\sum_i \mu_i^-\boldsymbol{\Lambda}_i}$

Combination:

$\hat{\boldsymbol{s}}_d(k|k) = \sum_j \hat{\boldsymbol{s}}_{dj}(k|k)\mu_j$

$\boldsymbol{V}_d(k|k) = \sum_j \boldsymbol{V}_{dj}(k|k)\mu_j + \boldsymbol{X}$

where the "spread-of-the means" term in combination is

$$\boldsymbol{X} = \sum_i [\hat{\boldsymbol{s}}_{di}(k|k) - \hat{\boldsymbol{s}}_d(k|k)][\hat{\boldsymbol{s}}_{di}(k|k) - \hat{\boldsymbol{s}}_d(k|k)]^H \mu_i(k)$$

We propose to use the IMM algorithm [11] to estimate $\boldsymbol{s}_d(n)$. In order to do this, in keeping with [11], we allow transitions among the M models (this also allows consideration of time-varying f_d) where these transitions are governed by a first-order homogeneous Markov chain with transition probabilities p_{ij}, $i, j \in \{1, 2, \ldots, M\}$, $\sum_{j=1}^{M} p_{ij} = 1$. The data symbols input to the channel $\tilde{s}(n)$ are treated as Gaussian random variables. The operation of IMM algorithm in one cycle is summarized in Table 5.1 where $\sigma_{\tilde{s}}^2 = \sigma_s^2 = E\{|s(n)|^2\}$ for information symbol, $= 0$ for training symbol. Table 1 provides one-cycle (one time sample update) of the IMM algorithm. The required initialization for the algorithm is as follows: at time $k = 0$, $\hat{s}(1|0) = E\{s(1)\} = 0$ and its covariance $\boldsymbol{V}_{\hat{s}}(1|0) = \sigma_s^2 \boldsymbol{I}_{d+1}$. Having obtained the IMM estimate $\hat{\boldsymbol{s}}_d(n|n)$ of $\boldsymbol{s}_d(n)$, we estimate $s(n)$ with equalization delay d by quantizing the $(d+1)$st component of $\hat{\boldsymbol{s}}_d(n|n)$.

5.4 Numerical Examples

A random time- and frequency-selective Rayleigh fading channel is considered. We take $L = 2$ (3 taps) in (5.1), number of receive antennas $R = 2$, and $h^{(r)}(n; l)$ are zero-mean, complex Gaussian with variance $\sigma_h^2 = 1/(L+1)$. For different l's and r's, $h^{(r)}(n; l)$'s are mutually independent and satisfy the Jakes' model. To this end, we simulated each single tap following [81] (with a correction in the appendix of [77]).

We consider a communication system with carrier frequency of $2\,\mathrm{GHz}$, data rate of $40\,\mathrm{kBd}$ (kilo-Bauds), therefore $T_s = 25\,\mu\mathrm{s}$, and a varying Doppler spread f_d in the range of $0\,\mathrm{Hz}$ to $200\,\mathrm{Hz}$, or the normalized Doppler spread $f_d T_s$ from 0 to 0.005 (corresponding to a maximum mobile velocity from 0 to $108\,\mathrm{km/h}$). The additive noise was zero-mean complex white Gaussian. The (receiver) SNR refers to the average energy per symbol over one-sided noise spectral density. The time-multiplexed training scheme of [48] described in Sec. 5.2.2 is adopted, where during data

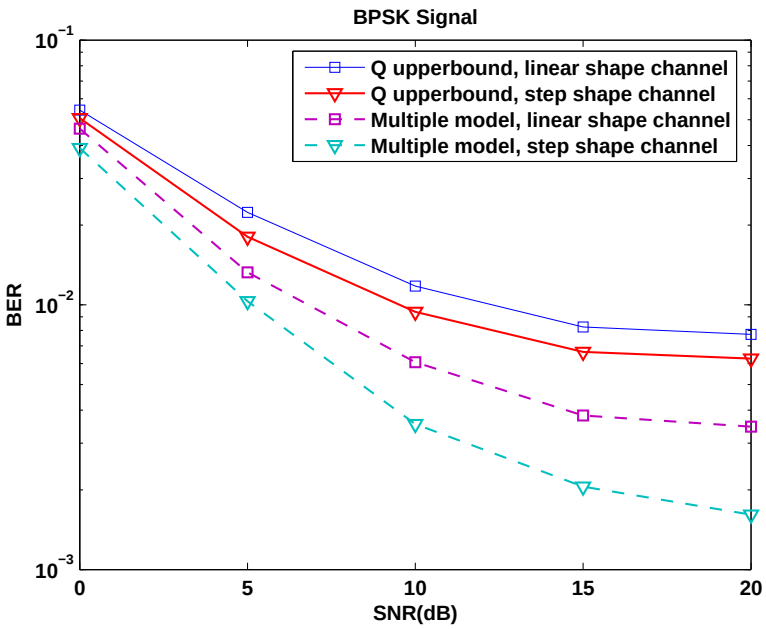

Figure 5.1: IMM with three models, BER vs SNR with BPSK information symbols.

sessions the information sequences is modulated by BPSK or QPSK with unit power. The training session is described by (5.7) with $\gamma = \sqrt{2L+1}$ so that the average symbol power of training sessions is equal to that of data sessions.

5.4.1 Example 1: IMM with Three Models

We generated a random doubly-selective channel as discussed earlier but with two different profiles of varying f_d's as follows:

1. f_d=0 Hz for $1 \leq n \leq 420$, f_d=100 Hz for $421 \leq n \leq 840$, f_d=200 Hz for $841 \leq n \leq 1260$, f_d=100 Hz for $1261 \leq n \leq 1680$, f_d=0 Hz for $1681 \leq n \leq 2100$. We picked $K = 2$, $T_B = 175$ and $P = 5$. Each subblock has 35 symbols with 30 information symbols in the beginning and 5 training symbols at the end (see Sec. 5.2.2). This channel is named as the **Step Shape** time-varying channel.

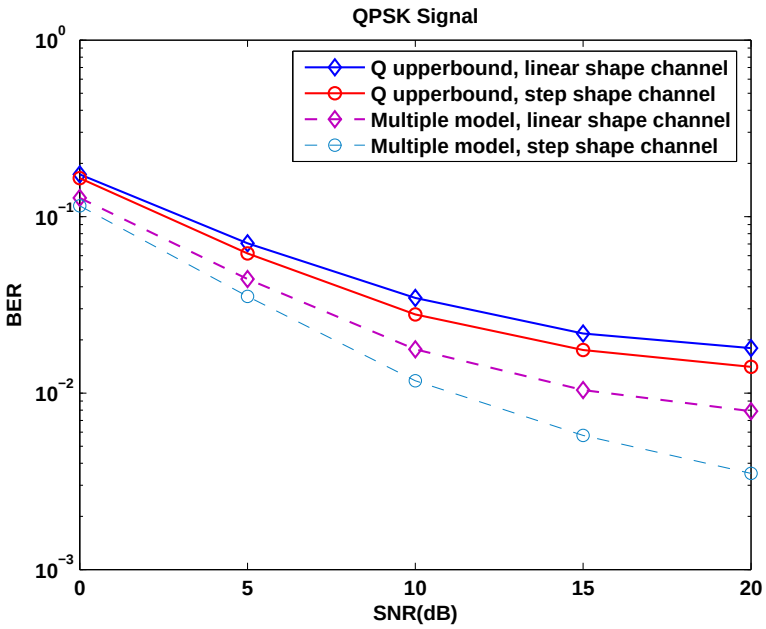

Figure 5.2: IMM with three models, BER vs SNR with QPSK information symbols.

2. Now f_d varies linearly from 0Hz to 200Hz over $1 \leq n \leq 1050$, and f_d varies linearly from 200Hz to 0Hz over $1051 \leq n \leq 2100$. This channel is named as the **Linear Shape** time-varying channel.

Two variations on channel estimation schemes are compared using an equalization delay $d = 5$:

1. **Q Upper bound:** We used a fixed Q for all blocks with Q=5= upper bound (denoted by "Q upper bound" in the figs.). With 5 subblocks per non-overlapping block (total 60 blocks), we estimated the channel for each block via the approach of Sec. 5.2.2. Then we used Kalman filtering with $d = 5$ (no IMM) to detect the information symbols.

2. **Proposed Multiple Model:** Here we used overlapping blocks by shifting blocks by one subblock. We used three models $M = 3$ with Q_1=1, Q_2=3 and Q_3=5. The channels are estimated over one block, then we shifted to the right by one subblock (35 symbols), and

estimated the 3 candidate channels again, and so on. For transition probability matrix we picked

$$\begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.05 & 0.9 & 0.05 \\ 0. & 0.1 & 0.9 \end{bmatrix}$$

which reflects the fact that transitions in f_d do not jump over an intermediate value. The three models had equal initial probabilities of $1/3$.

The bit error rate (BER) of each scheme was studied by averaging over 200 runs where in each run, a symbol sequence of length 2100 is generated and fed into a random doubly-selective channel generated with specified f_d's. The first 70 symbols were discarded in evaluations. In Figs. 5.1 and 5.2, the performances of the two schemes under different SNR's are compared for BPSK and QPSK information sequences, respectively. It is readily seen that overestimating Doppler spread leads to a performance deterioration compared to the proposed IMM approach relying on a multiple model formulation.

5.4.2 Example 2: IMM with Two Models

In this example, the varying Doppler spread f_d is in the range of 0Hz to 100Hz. Again two variations on channel estimation schemes are compared:

1. **Q Upper bound**: We used a fixed Q for all blocks with Q=3= upper bound (denoted by "Q upper bound" in the figures). With 5 subblocks per non-overlapping block (total 50 blocks), we estimated the channel for each block via the approach of Sec.5.2.2. Then we used Kalman filtering with $d = 5$ (no IMM) to detect the information symbols.

2. **Proposed Multiple Model**: Here we used overlapping blocks by shifting blocks by one subblock. We used two models $M = 2$ with Q_1=1 and Q_2=3. The channels are estimated over one block, then we shifted to the right by one subblock (35 symbols), and estimated the 2 candidate channels again, and so on. For transition probability matrix we picked

$$\begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

The two models had equal initial probabilities of $1/2$.

In all numerical results, the bit error rate (BER) of each scheme was studied by averaging over 200 runs where in each run, a symbol sequence of length 1750 (total 50 blocks) is generated and fed into a random doubly-selective channel generated with specified f_d's.

Step Shape Time-Varying Channel

We generate the **Step Shape** random doubly-selective channel with varying f_d as follows: f_d=0 Hz for $1 \leq n \leq 584$, f_d=100 Hz for $585 \leq n \leq 1168$, and f_d=0 Hz for $1169 \leq n \leq 1750$.

In Figs. 5.3 and 5.4, the performances of the two schemes under different SNR's are compared for BPSK and QPSK information sequences, respectively. It is readily seen that overestimating Doppler spread leads to a performance deterioration. The proposed IMM approach relying on a multiple model formulation provides a good performance improvement.

Fig. 5.5 plots how the average mode probabilities μ (in Table 5.1) change with time. It can be seen from the figure that model 2 ($Q_2 = 3$) that is supposed to capture larger Doppler frequency can also capture model 1 ($Q_1 = 1$ for $f_d = 0Hz$), while model 1 can not capture model 2.

Linear Shape Time-Varying Channel

In this example the **Linear Shape** random doubly-selective channel is generated with varying f_d linearly: f_d is from 0Hz to 100Hz for $1 \leq n \leq 875$, and f_d is from 100Hz to 0Hz for $876 \leq n \leq 1750$.

Again the performances of the two schemes under different SNR's are compared in Fig. 5.6 and Fig. 5.7 for BPSK and QPSK information sequences, respectively. Fig. 5.8 plots how the average mode probabilities μ change with time.

The Effect of Training Sequence on Mode Probability

How does the training sequence structure in (5.7) influence the mode probability is tested in this example. The random channel is generated by fixing $f_d = 0$ for $1 \leq n \leq 1750$ (frequency-selective and time-invariant channel). Model 1 with $Q = 1$ and Model 2 with $Q = 3$ are exploited for symbol detection. Suppose the exact channel information regarding Model 1 and Model 2 is available at the receiver. Clearly Model 1 should be the right model being chosen with higher mode probabilities than Model 2.

In one case, the training symbols with structure as (5.7) are used. The resulting average mode probability (averaged over 200 iterations) is plotted in Fig. 5.9. Similar to the plottings in Fig. 5.5 and Fig. 5.8, we see that the mode probabilities between blocks are inconsistent.

In the other case, no training symbols are inserted in the transmitted signal. The average mode probability result is shown in Fig. 5.10. It is seen that the mode probabilities under this case are consistently smooth. The comparison between Fig. 5.9 and Fig. 5.10 illustrates that the inserted zeros in training structure (5.7) result in the inconsistency of mode probability.

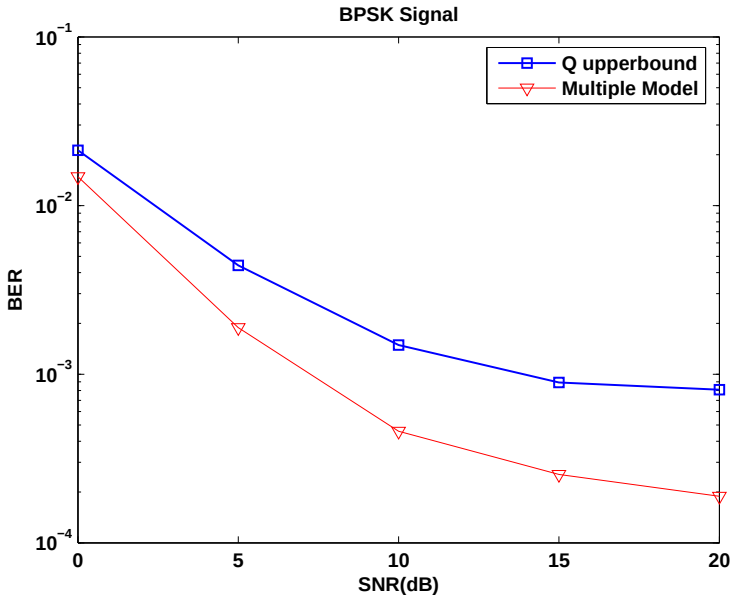

Figure 5.3: IMM with two models. BER vs SNR with BPSK information symbols. Step shape time-varying Channel

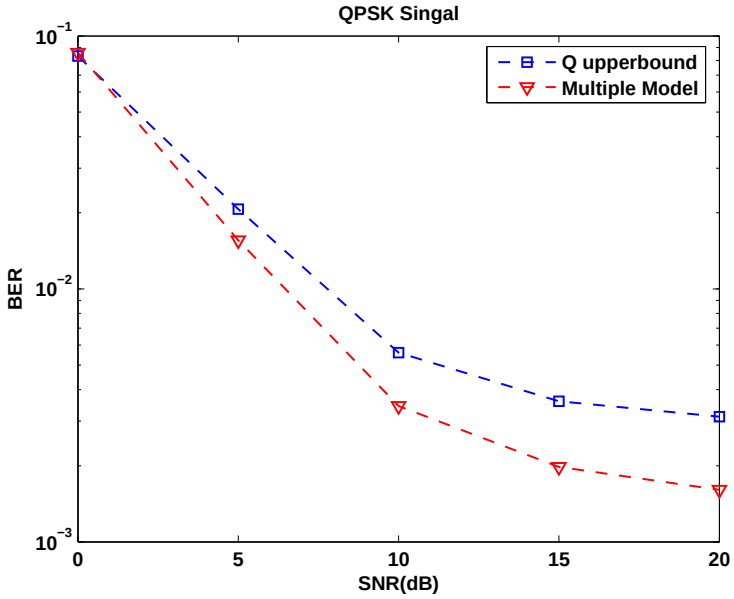

Figure 5.4: IMM with two models. BER vs SNR with QPSK information symbols. Step shape time-varying Channel

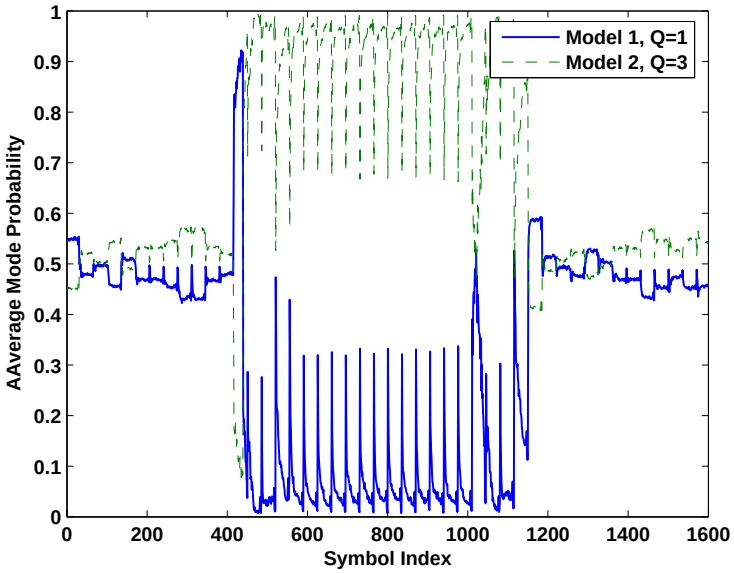

Figure 5.5: Average mode probability of IMM. Step shape time-varying channel. SNR=20dB

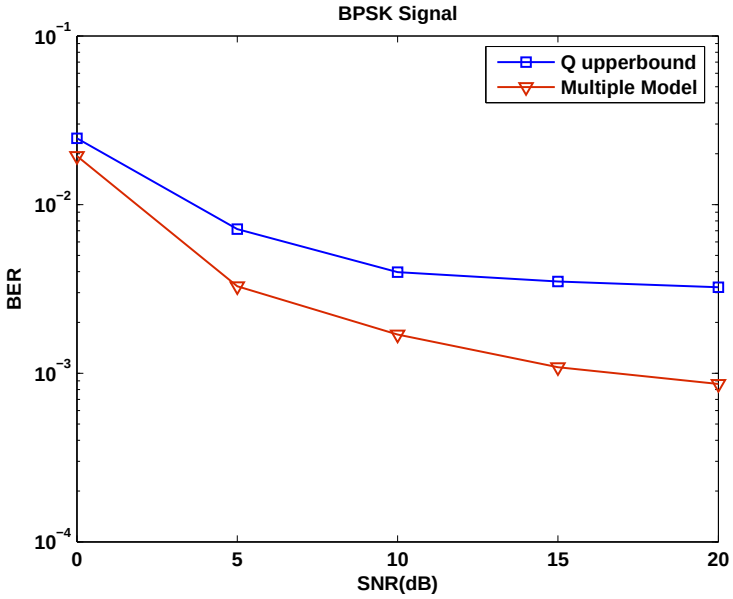

Figure 5.6: IMM with two models. BER vs SNR with BPSK information symbols. Linear Shape Time-Varying Channel

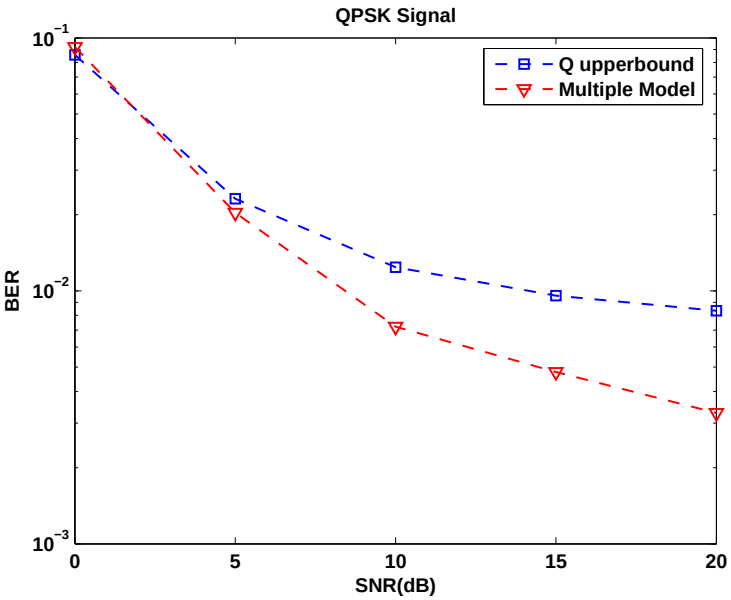

Figure 5.7: IMM with two models. BER vs SNR with QPSK information symbols. Linear Shape Time-Varying Channel

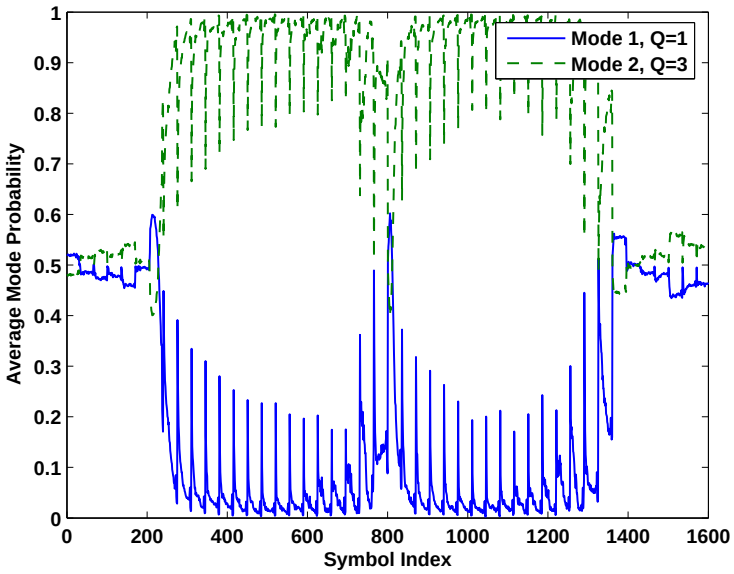

Figure 5.8: Average mode probability of IMM. Linear shape time-varying channel. SNR=20dB

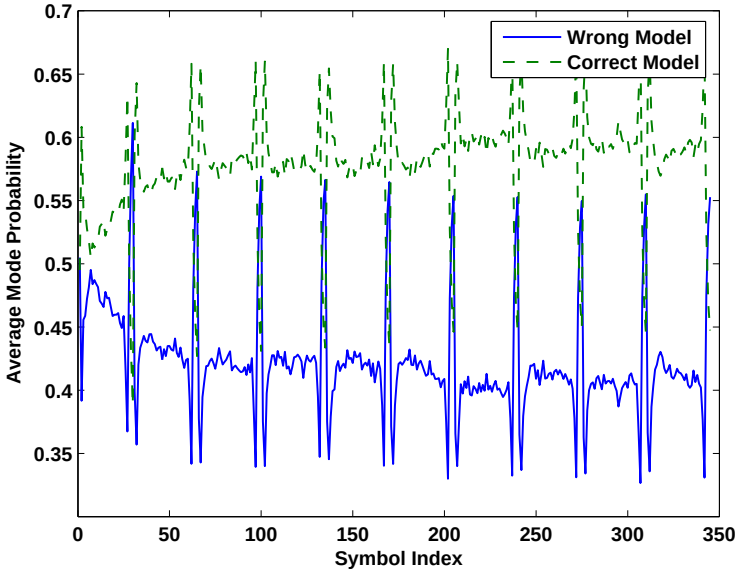

Figure 5.9: Average mode probability of IMM. Time-invariant channel. With training insertion. SNR=20dB

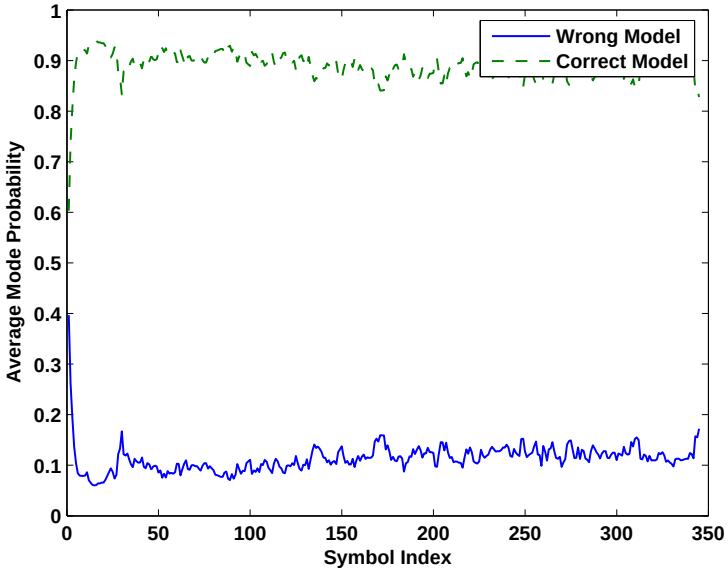

Figure 5.10: Average mode probability of IMM. Time-invariant channel. Without training insertion. SNR=20dB

5.5 Conclusion

An adaptive channel estimation scheme, exploiting the oversampled complex exponential basis expansion model (CE-BEM), was presented for doubly-selective channels where we tracked the BEM coefficients via a multiple model approach. In the past work the number of BEM coefficients used to model the doubly-selective channels for channel estimation has been based on an upper bound on the channel Doppler spread. The higher the Doppler spread, the more the number of BEM coefficients, which leads to a higher channel estimation variance. In this chapter we proposed to use a multiple model framework where several candidate Doppler spread values were used to cover the range from zero to an upper bound, leading to multiple CE-BEM channel models, each corresponding to an assumed value of the Doppler spread. Subsequently the well known interacting multiple model (IMM) algorithm was used for symbol detection based on multiple state-space models corresponding to the multiple estimated channels. Numerical examples were presented to illustrate the proposed approach. The results showed that the IMM approach relying on multiple models had a better performance than traditional overestimating Doppler spread scheme.

CHAPTER 6

DOUBLY-SELECTIVE CHANNEL ESTIMATION FOR OFDM SYSTEMS USING DPS-BEM AND TIME-MULTIPLEXED TRAINING

6.1 Introduction

Wireless multicarrier (MC) communication systems utilize multiple complex exponentials as information-bearing carriers. MC transmissions thus retain their shape and orthogonality when propagating through linear time-dispersive media. They were first conceived and implemented with analog oscillators in the 60s [42, 83], but it was not until their all-digital implementation with the Fast Fourier Transform (FFT), that their attractive features were unravelled and sparked widespread interest for adoption in various single user and multiple access (MA) communication standards [9]. Nowadays, MC systems such as the Orthogonal Frequency Division Multiplexing (OFDM) are included in the Digital Audio/Video Broadcasting (DAB/DVB) standards in Europe while the Discrete Multi-Tone (DMT), its wireline counterpart in the US, has been applied to high-speed Digital Subscriber Line (DSL) modems over twisted pairs [10].

The pilot-aided doubly-selective channel estimation for OFDM systems is considered in [62], where the channels are approximated by BEMs. The transceiver block diagram is shown in Fig. 6.1. Since the channel estimation is based on frequency-domain training, the receiver can find no subcarrier that solely depends on pilots and thus is not contaminated by information symbols. Due to time-variation, the resulting channel matrix (include IFFT and FFT) is no longer diagonal

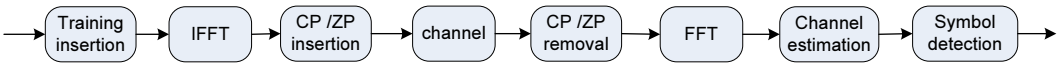

Figure 6.1: OFDM transceiver block diagram in [62]

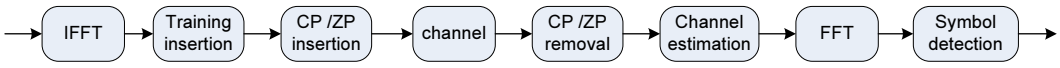

Figure 6.2: Proposed OFDM transceiver block diagram

matrix, which results in the very complicated channel estimation or equalization procedure [6]. In [62], a receiver window is used to suppress the out of band interference, and it is assumed that the channel matrix is approximately banded after windowing. It is also not clear what is the "optimum" strategy to place the frequency-domain training symbols. The linear MMSE channel estimator for DPS-BEM-based MIMO-OFDM (multiple input multiple output-orthogonal frequency division multiplexing) doubly-selective channels is introduced in [49], but their channel estimation suffers from the frequency-domain training problem, too.

In this chapter, we will apply the optimum time-multiplexed training based channel estimation introduced in Chapter 2 to OFDM systems under doubly-selective channels. The time-varying channel is described by CE-BEM or DPS-BEM. The time-domain training clusters are inserted after IFFT at the transmitter and the channel estimation proceeds before FFT at the receiver (see Fig. 6.2). In this way, the "optimum" training strategy and the corresponding channel estimation schemes proposed before can be directly applied to MIMO-OFDM systems. Compared to frequency-domain training design, the main advantage of time-domain training for OFDM system is that the information symbols are not contaminated by the training symbols as the frequency-domain training. The performance of frequency-domain training-based channel estimation and time-domain training-based channel estimation are compared by simulations.

6.2 System Model

Fig. 6.2 depicts the proposed baseband block diagram of a cyclic prefix (CP)-OFDM system, where the jth $N_c \times 1$ information block

$$\tilde{\boldsymbol{c}}_j := [c((j-1)N_c), c((j-1)N_c+1), ..., c((j-1)N_c+N_c-1)]^T \tag{6.1}$$

is first precoded by the IFFT matrix $\boldsymbol{F}_{N_c}^H$ with (m,k)th entry

$$\left[\boldsymbol{F}_{N_c}^H\right]_{m,k} = \frac{1}{\sqrt{N_c}} e^{j2\pi mk/N_c}, \tag{6.2}$$

to yield the so-called "time domain" block vector

$$\boldsymbol{c}_j = \boldsymbol{F}_{N_c}^H \tilde{\boldsymbol{c}}_j. \tag{6.3}$$

Then the training sequence $\boldsymbol{b}_j$ of length N_b and cyclic prefix (CP) $\boldsymbol{c}p_j$ of length N_{cp} is inserted between each $\boldsymbol{c}_j$. CP is inserted at the transmitter and discarded at the receiver to avoid the interblock interference (IBI) [72]. The entries of the resulting redundant block $\boldsymbol{s}_j := [\boldsymbol{c}p_j^T, \boldsymbol{b}_j^T, \boldsymbol{c}_j^T]$ are finally sent sequentially through the channel. Suppose there are totally J blocks for transmission. Then the general structure of transmitted signal $\boldsymbol{s}$ with length $N := N_{cp} + N_b + N_c$ is

$$\boldsymbol{s} := [\boldsymbol{c}p_1^T, \boldsymbol{b}_1^T, \boldsymbol{c}_1^T, ..., \boldsymbol{c}p_J^T, \boldsymbol{b}_J^T, \boldsymbol{c}_J^T]^T. \tag{6.4}$$

Consider the doubly-selective single input single output FIR linear channel, the same as in Section 2.2. Then the matrix-vector input-output relationship is given by

$$\boldsymbol{x} = \sum_{q=0}^{Q} \boldsymbol{D}_{\psi_q} \boldsymbol{W}_q \boldsymbol{s} + \boldsymbol{\eta}, \tag{6.5}$$

where $\boldsymbol{x} := [x(0), x(1), ..., x(N-1)]^T$, $\boldsymbol{\eta}$ is the zero-mean white complex-Gaussian noise defined similarly to $\boldsymbol{x}$, $\boldsymbol{D}_{\psi_q} = \mathrm{diag}[\boldsymbol{\psi}_q]$ with $\boldsymbol{\psi}_q := [\psi_q(0), \psi_q(1), ..., \psi_q(N-1)]^T$, and $\boldsymbol{W}_q$'s are $N \times N$ lower triangular Toeplitz matrices with 1st column $[w_q(0), w_q(1), \ldots, w_q(L), 0, \ldots, 0]^T$.

This system model in (6.5) is exactly the same as in (2.9). Comparing (2.10) with (6.4), and (2.9) with (6.5), we notice that the IBI free received signal is the same as that in Chapter 2 after CP removal at the receiver. Therefore, the LS/MMSE channel estimator and optimum training design presented in Chapter 2 can be directly applied here without any change. We will skip the theoretical expressions which would be the same as in Chapter 2.

6.3 Doubly-Selective Channel Estimation for OFDM Systems Using Frequency-Domain Training [62]

Consider an OFDM system with M subcarriers. The frequency domain transmitted signal is defined as:

$$\tilde{\boldsymbol{x}} := [\boldsymbol{t}_1^T, \boldsymbol{u}_1^T, \boldsymbol{t}_2^T, \boldsymbol{u}_2^T \cdots \boldsymbol{t}_P^T, \boldsymbol{u}_P^T]^T$$

where $\boldsymbol{t}_p$ $(p = 1, 2, \cdots P)$ is a column of M_t training symbols and $\boldsymbol{u}_j$ is a column of M_u information symbols. Let $M = (M_t + M_u)P$. As illustrated in Fig 6.1, the OFDM symbol $\tilde{\boldsymbol{x}}$ is first modulated by the IFFT operation

$$\boldsymbol{x} = \boldsymbol{F}_M^H \tilde{\boldsymbol{x}}. \tag{6.6}$$

We ignore the symbol index in (6.6) since only one OFDM symbol is considered.

The modulated symbol $\boldsymbol{x}$ is then concatenated by a CP and sent over the channel. If the channel is assumed to be time-varying (TV) and approximated by a BEM, as in [62], the received signals at the receiver after CP removal and FFT operation is given by

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{F}_M \boldsymbol{R}_{\rm cp} \boldsymbol{H} \boldsymbol{T}_{\rm cp} \boldsymbol{x} + \boldsymbol{v} \\ &= \sum_{q=0}^{Q} \boldsymbol{D}_q \boldsymbol{\Delta}_q \tilde{\boldsymbol{x}} + \boldsymbol{v}, \end{aligned} \tag{6.7}$$

where $\boldsymbol{T}_{\rm cp} := [\boldsymbol{I}_{N_{\rm cp} \times M}; \boldsymbol{I}_M]$ and $\boldsymbol{R}_{\rm cp} := [\boldsymbol{0}_{M \times N_{\rm cp}}, \boldsymbol{I}_M]$ are CP insertion and CP removal matrices, respectively; $\boldsymbol{H}$ represents the channel matrix in time-domain; $\boldsymbol{v}$ is the zero-mean white complex-Gaussian noise; $\boldsymbol{D}_q$ is a circulant matrix whose first column is the frequency response of the qth basis function

$$\boldsymbol{D}_q := \boldsymbol{F}_M {\rm diag}\{\boldsymbol{\psi}_q\} \boldsymbol{F}_M^H \tag{6.8}$$

and $\boldsymbol{\Delta}_q$ is a diagonal matrix whose diagonal is the frequency response of the BEM coefficients corresponding to the qth basis function [62]

$$\boldsymbol{\Delta}_q = {\rm diag}\{\boldsymbol{F}_L [w_{q,0}, ..., w_{q,L}]^T\}. \tag{6.9}$$

$\boldsymbol{F}_L$ stands for the first $L+1$ columns of the matrix $\boldsymbol{F}_M$. The least squares (LS) channel estimator can be achieved based on (6.7) as in [[62], Section IV. B].

Due to the time-variation, the channel matrix in the frequency domain is no longer diagonal, but approximately banded. Therefore, the channel estimation approach based on (6.7) suffers a large estimation error especially with a high Doppler spread.

6.4 Numerical Examples

In this section, the numerical results of the proposed time-domain (TD) training channel estimation (Sec. 6.2, referred to as TDE) are given and compared with the frequency-domain (FD) channel estimation (Sec. 6.3, referred to as FDE). We use binary phase shift keying (BPSK) modulation in all examples.

To make a relatively fair comparison between TDE and FDE cases, we want to keep the same transmission rates in both schemes. For TDE scheme, the OFDM system has $N_c = 30$ subcarriers. There are $N_b = 2L + 1$ training symbols with optimal structure $[\mathbf{0}_L, b, \mathbf{0}_L]$, $b > 0$ in every OFDM symbol and $J = 10$ OFDM symbols are transmitted sequentially. For FDE scheme, an OFDM system with $M = (M_t + M_u)P$ subcarriers is considered. The subcarriers that are reserved for pilots are grouped in $P = 10$ equidistant clusters, each containing $M_t = 2L + 1$ pilot tones. We pick $M_u = 30$. Inside each cluster, the scheme referred to as "frequency-domain Kronecker delta" (FDKD) in [35] are exploited, where a nonzero pilot is located in the middle of the cluster with zero guard bands on both sides. This equidistant pilot cluster scheme finds its practical advantage in [60] although the channel follows the bathtub-shaped Doppler spectrum in that case. Another important reason of using FDKD training pattern here is that it is convenient to keep the same transmission rate with TDE scheme.

The doubly-selective Rayleigh fading channel is simulated according to [77, 81] with channel order $L = 2$, carrier frequency of 2GHz, data rate of 40 kbps, and thus, symbol duration $T_s = 25\mu s$.

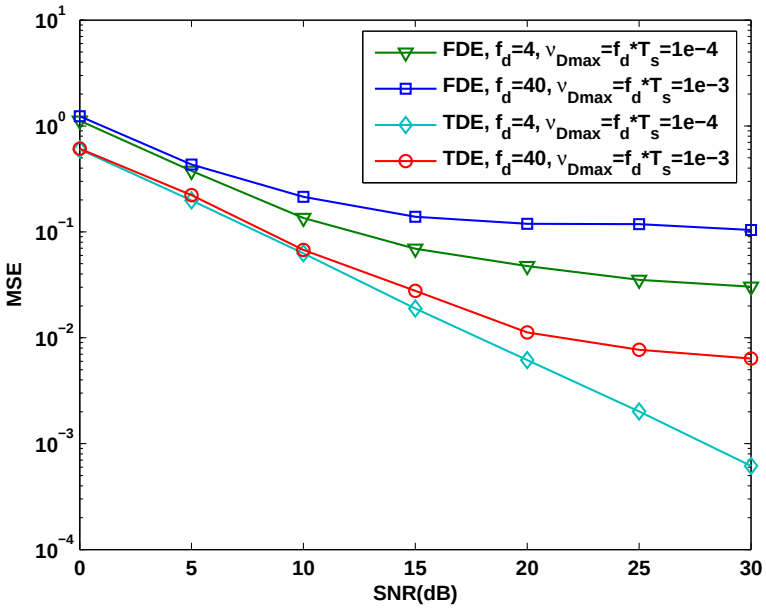

Figure 6.3: LS Channel estimation MSE comparison between TDE and FDE with varying ν_{Dmax}

Therefore, each tap of the generated time-variant channel has a Jakes' spectrum; it is not generated using the assumed BEM modeling. Also, the 3 taps of the channel are mutually independent. Depending on different maximum Doppler spread f_ds, a varying maximum normalized one-sided Doppler bandwidth $\nu_{\mathrm{Dmax}} = f_d T_s$ can be derived, where T_s is the symbol duration. A Kalman filter formulation is used for information detection after the channel estimation. The SNR refers to $1/\sigma_\eta^2$ where the information sequence power is normalized to one and the channel power is also normalized to unity.

6.4.1 Example 1: Channel Estimation Performance

In this example, the LS channel estimator is used in both TDE and FDE schemes to estimate $\boldsymbol{w}$, and then the channel is estimated as in (2.29). The channel estimation MSE is calculated as in (2.65). Fig. 6.3 plots the channel estimation MSE (averaged over 200 Monte Carlo runs) versus

SNR under different normalized maximum Doppler bandwidths, where $f_d = 4$ or 40Hz. It is clear that the MSE of time-domain training channel estimation is consistently smaller than that of frequency-domain training channel estimation.

6.4.2 Example 2: CE-BEM versus DPS-BEM

The LS channel estimator in Example 1 is based on the DPS-BEM channel model. In this example we compare the LS channel estimation MSE of TDE under CE-BEM and DPS-BEM. Fig. 6.4 plots the MSE (2.65) versus SNR under different maximum Doppler bandwidths ($\nu_{Dmax}^{(F)}$ is for CE-BEM, and $\nu_{Dmax}^{(S)}$ is for DPS-BEM). Especially, we take $Q_F = 2\lceil \nu_{Dmax} N \rceil$ for CE-BEM and $Q_S = \lceil 2\nu_{Dmax} N \rceil$ DPS-BEM. It is seen from Fig. 6.4 that the channel estimation performances of DPS-BEM outperform that of CE-BEM for all ν_{Dmax}. Fig. 6.5 takes BER (average over 2000 Monte Carlo runs) as a performance measure to compare the performance between DPS-BEM and CE-BEM for TDE. Comparing with Fig. 6.4 makes it clear that the significantly reduced MSE of the DPS-BEM channel estimation leads to a pronounced reduction in BER compared to the CE-BEM case.

6.5 Conclusion

In this chapter, we applied the optimum time-multiplexed training based channel estimation introduced in Chapter 2 to OFDM systems under doubly-selective channels. The time-varying channel was described by CE-BEM or DPS-BEM. The time-domain training clusters are inserted after IFFT at the transmitter and the channel estimation proceeds before FFT at the receiver. Compared to frequency-domain training design, the main advantage of time-domain training for OFDM system

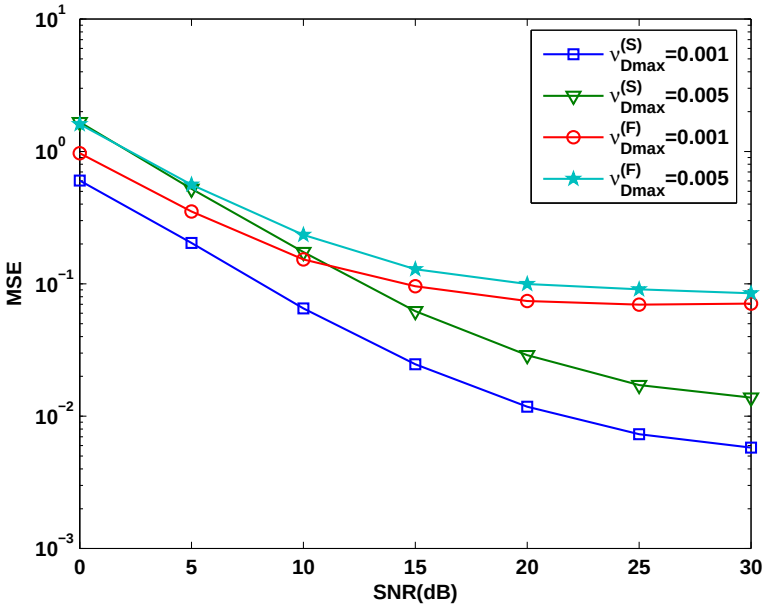

Figure 6.4: LS Channel estimation MSE comparison between DPS-BEM and CE-BEM with varying ν_{Dmax}, TDE

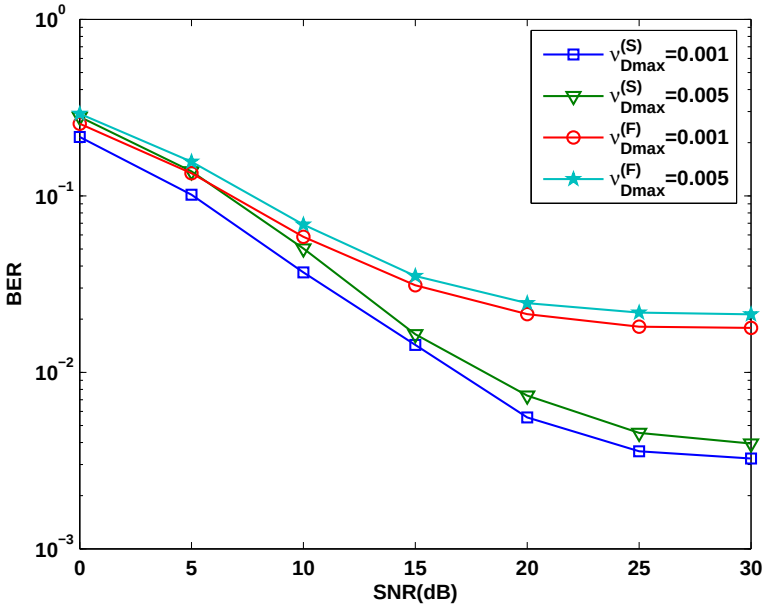

Figure 6.5: BER comparison between DPS-BEM and CE-BEM with varying ν_{Dmax}, TDE

is that the information symbols are not contaminated by the training symbols as in the frequency-domain training case. The simulation results confirmed the claims.

CHAPTER 7

SUMMARY AND FUTURE WORK

With the emergence of next-generation wireless mobile communications, multimedia services have placed increasing demands for high data rates and high mobility. The high data rates give rise to frequency selectivity, while the mobility and carrier offset introduce time selectivity. In this dissertation, the channel estimation and equalization for frequency-selective and time-selective channel were considered.

7.1 Summary of Original Work

In Chapter 2, the channel estimation for single-input single-output frequency- and time- selective channels was considered using time-multiplexed training. The time-varying channel was assumed to be well-described by a basis expansion model using discrete prolate spheroidal sequences as bases (DPS-BEM). Both linear least squares and minimum mean-square-error approaches were exploited to estimate the basis expansion coefficients. Training designs for time-multiplexed training based on minimization of channel estimation mean-square-error were investigated. The issue of training power allocation was addressed. Then the proposed channel estimation approaches in Chapter 2 was extended to multiuser multiple-input multiple-output doubly-selective channels in Chapter 3.

The linear equalization and decision feedback equalization of doubly-selective channels modeled via BEMs were introduced in Chapter 4. There has been much interest in designing time-variant serial finite impulse response linear and DFE equalizers using complex exponential BEMs for equalizers in addition to using CE-BEM for modeling the channel itself. We showed that the

Kalman filter formulation of the linear equalizer and an alternative formulation of the FIR DFE based on a CE-BEM channel model yields the same or improved BER at a lower computational cost, without incurring the approximation error inherent in CE-BEM modeling of equalizers.

An adaptive channel estimation scheme, exploiting the oversampled complex exponential basis expansion model (CE-BEM), was presented for doubly-selective channels where we tracked the BEM coefficients via a multiple model approach in Chapter 5. In the past work the number of BEM coefficients used to model the doubly-selective channels for channel estimation has been based on an upper bound on the channel Doppler spread. The higher the Doppler spread, the more the number of BEM coefficients, which leads to a higher channel estimation variance. We proposed to use a multiple model framework where several candidate Doppler spread values were used to cover the range from zero to an upper bound, leading to multiple CE-BEM channel models, each corresponding to an assumed value of the Doppler spread. Subsequently, the well known interacting multiple model (IMM) algorithm was used for symbol detection based on multiple state-space models corresponding to the multiple estimated channels.

Orthogonal Frequency-Division Multiplexing (OFDM), a digital multi-carrier modulation scheme, has developed into a popular scheme for wideband wireless communication due to its high spectral efficiency and simple equalization. In Chapter 7, we extended the optimum time-multiplexed training-based channel estimation introduced in Chapter 2 to OFDM systems under a doubly-selective channels. Compared to the traditional frequency-domain training design, the main advantage of time-domain training design for OFDM system is that the information symbols are not contaminated by the training symbols as in the frequency-domain training case.

In all chapters, numerical examples based on computer simulations were presented to illustrate the proposed approaches and confirm the conclusions.

7.2 Possible Future Directions

So far we have assumed that the coefficients of basis expansion model (BEM) are fixed over data blocks. In fact, the BEM coefficients may also undergo changes.

Instead of the basis expansion model, another way to model time-varying channels is via autoregressive (AR) process, particularly the first order AR process [19]. Suppose $h(n;l)$ represents a wide-sense stationary uncorrelated scattering (WSSUS) channel. It is common to use the following first-order AR model to describe it:

$$h(n;l) = \alpha h(n-1;l) + \eta(n), \tag{7.1}$$

where α is the AR coefficient, and the driving noise $\eta(n)$ is zero-mean complex Gaussian with variance σ_η^2 and statistically independent of $h(n-1;l)$. Assume that $h(n;l)$ is also zero-mean, complex Gaussian with variance σ_h^2. Then

$$\alpha = \frac{1}{\sigma_h^2} E\left\{h(n;l)h^*(n-1;l)\right\}, \tag{7.2}$$

$$\sigma_\eta^2 = \sigma_h^2(1-|\alpha|^2). \tag{7.3}$$

It is a tractable model, where the channel is assumed to be Markovian; i.e., for the current channel symbol, the effect of channel symbols other than the immediately preceding one is negligible [73]. This Markovian assumption has been verified for Rayleigh fading channels in [73], by considering the mutual information between channel symbols. AR models describe temporal variation on a symbol-by-symbol update basis, while BEMs depict the evolution of the channel over a period (block) of time.

An AR model is not appropriate for a fast fading channels because the channel tracking depends on the inserted training. For information sessions, channel estimates can only be obtained based on the results from the previous training session [40]. However, it could be a good way to track the basis expansion coefficients in BEMs. Up to this point, we suppose that the basis expansion coefficients are fixed over data blocks. In fact, the BEM coefficients may also undergo changes, but not as fast as the channel variations.

In the next step, we will model the doubly-selective channel by exploiting the CE-BEM for the overall time-variant channel and an AR model for the BEM coefficients. Recall the BEM model in (2.2). Stack the BEM coefficients into vectors

$$\boldsymbol{w}_l = [w_1(l), w_2(l), ..., w_Q]^T. \tag{7.4}$$

Suppose the whole data block is divided into J sub-blocks, where the coefficient vector in (7.4) for the jth sub-block $(j = 1, 2, ..., J)$ is denoted as $\boldsymbol{w}_l(j)$ and updated every sub-block. Since a fading channel well follows the Markov model, we further assume that the BEM coefficients of each block are Markovian, too. The first-order AR process for BEM coefficients is then given by:

$$\boldsymbol{w}_l(j) = \alpha \boldsymbol{w}_l(j-1) + \boldsymbol{\eta}_l(j), \tag{7.5}$$

where $\boldsymbol{\eta}_l(j)$ defined similar to $\boldsymbol{w}_l(j)$, is zero-mean complex Gaussian noise and statistically independent of $\boldsymbol{w}_l(j-1)$. Based on the AR model in (7.5), a Kalman filter can be applied to track the BEM coefficient for each block. After deriving the estimated BEM coefficients for jth sub-block $\hat{\boldsymbol{w}}_l(j)$, the estimated channel impulse response for the whole jth sub-block can be given, via the

CE-BEM, as

$$\hat{h}(n;l) = [e^{j\omega_0 n}, e^{j\omega_1 n}, ..., e^{j\omega_Q n}]^T \hat{\boldsymbol{w}}_l(j). \tag{7.6}$$

It would be interesting to explore this approach and compare it with our proposed approach.

Multiple Model Approach. In applying IMM we treated the estimated channel as the true channel. It would be interesting to incorporate estimated channel statistics, e.g., estimation error variance, into the IMM algorithm.

BIBLIOGRAPHY

[1] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for unknown channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2338-2353, Aug. 2002.

[2] N. Al-Dhahir and J.M. Cioffi, "MMSE decision-feedback equalizers: finite-length results," *IEEE Trans. Inform. Theory*, vol. 41, pp. 961-975, July 1995.

[3] I. Barhumi, G. Leus, and M. Moonen, "Time-varying FIR decision feedback equalization of doubly-selective channels," in *Proc. IEEE Global Commun. Conf*, pp. 2263-2268, San Francisco, CA, Dec. 2003.

[4] I. Barhumi, G. Leus, and M. Moonen, "Time-varying FIR equalization for doubly selective channels," in *Proc. IEEE Int. Conf. on Commun.*, vol. 5, no. 1, pp. 3246-3250, May. 2003.

[5] I. Barhumi, G. Leus, and M. Moonen, "Time-varying FIR equalization for doubly selective channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 1, pp. 202-214, Jan. 2005.

[6] I. Barhumi, G. Leus, and M. Moonen, "Equalization for OFDM Over Doubly Selective Channels," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1445-1458, Apr. 2006.

[7] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles, Techniques and Software*. Norwood, MA: Artech House, 1993.

[8] I. Barhumi, G. Leus, and M. Moonen, "Estimation and direct equalization of doubly selective channels," *EURASIP Journal on Applied Signal Process.*, article ID: 62831, pp. 1-15, vol. 2006.

[9] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Communications Magazine*, pp. 5-14, May 1990.

[10] J. A. C. Bingham, *ADSL, VDSL, and Multicarrier Modulation*. Wiley-Interscience, 2000.

[11] H.A.P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Automatic Control*, vol. 33, No. 8, pp. 780- 783, Aug. 1988.

[12] H. B$\ddot{o}$lcskei, R. W. Heath Jr., and A. J. Paulraj, "Blind channel identifi- cation and equalization in OFDM-based multi-antenna systems," *IEEE Trans. Signal Process*., vol. 50, no. 1, pp. 96-109, Jan. 2002.

[13] P.A. Bello, "Characterization of randomly time-variant channels," *IEEE Trans. Comm. Systems*, vol. CS-11, pp. 360-393, Dec. 1963.

[14] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*. New York: Kluwer/Plenum, 1999.

[15] Broadband Radio Access Networks (BRAN); HIPERLAN Type 2 Technical Specification Part 1-Physical Layer, Eur. Telecommun. Standards Inst. (ETSI), DTS/BRAN030 003-1, 1999.

[16] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686-693, Nov. 1991.

[17] R. H. Clarke, "A statistical theory of mobile radio reception", *Bell Systems Technical Journal* pp. 957-1000, July-August 1968.

[18] S. N. Crozier and D. D. Falconer, "Reduced complexity short block data detection techniques for fading dispersive channels," *IEEE Trans. Vehic. Technol.*, vol. 41, pp. 255-265, Aug. 1992.

[19] M. Dong, L. Tong, and B. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, pp. 1403-1418, May. 2004.

[20] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, "Differential space-time transmission for frequency-selective channels," in *Proc. 36th Conf. Information Sciences Systems*. Princeton, NJ, pp. 859-862, Mar. 20-22, 2002.

[21] S. A. Fechtel and H. Meyr, "Optimal parametric feedforward estimation of frequency-selective fading radio channels," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1639-1650, Feb./Mar/Apr. 1994.

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

[23] G. D. Forney Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363-378, March 1972.

[24] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj, "Outdoor MIMO wireless channels: capacity and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926-1934, 2002.

[25] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, *Signal Processing Advances in Wireless and Mobile Communications*, - Volume I, Trends in Channel Estimation and Equalization, Englewood Cliffs, NJ: Prentice-Hall, 2000.

[26] G. B. Giannakis and C. Tepedelenlioğlu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, pp. 1969-1986, Nov. 1998.

[27] A. Gorokhov and P. Loubaton, "Semi-blind second order identification of convolutive channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Munich, Germany, pp. 3905-3908, April 1997.

[28] M. S. Grewal and A. P. Andrews, *Kalman filtering: theory and practice*. Prentice-Hall: Englewood Cliffs, NJ, 1993.

[29] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple- antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951-963, Apr. 2003.

[30] T.-J. Ho and B.-S. Chen, "Tracking of dispersive DS-CDMA channels: An AR-embedded modified interacting multiple-model approach," *IEEE Trans. Wireless Commun*., vol. 6, pp. 166-174, Jan. 2007.

[31] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, UK: Cambridge Univ. Press, 1985.

[32] P. Heher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Mnich, Germany, pp. 1845-1848, 1997.

[33] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge, UK: Cambridge Univ. Press, 1985.

[34] W. C. Jakes, *Microwave Mobile Communications*, New York: John Wiley & Sons Inc., Feb. 1975.

[35] A. P. Kannu and P. Schniter, "MSE-optimal training for linear time-varying channels," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Philadelphia, USA, pp. 789-792, Mar. 2005.

[36] G. K. Khaleh, "Channel equalization for block transmission," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 110-121, Jan. 1995.

[37] G. Leus, "On the estimation of rapidly time-varying channels," in *Proc. European Signal Processing Conf.*, vol. 3, pp. 2227-2230, Vienna, Austria, Sept. 2004.

[38] G. Leus, S. Zhou, and G. B. Giannakis, "Orthogonal multiple access over time- and frequency-selective channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1942-1950, Aug. 2003.

[39] Z. Liu, G. B. Giannakis, S. Barbarossa, and A. Scaglione, "Transmitantennae space-time block coding for generalizedOFDMin the presence of unknown multipath," *IEEE J. Select. Areas Commun*., vol. 19, no. 7, pp. 1352-1364, Jul. 2001.

[40] Z. Liu, X. Ma, and G. B. Giannakis, "Space-time coding and Kalman filtering for time-selective fading channels," *IEEE Trans. Commun.*, vol. 50, no.2, pp. 183-186, 2002.

[41] X. Ma, G. B. Giannakis and S. Ohon, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. on Signal Process.*, vol. 51, pp. 1351-1366, May. 2003.

[42] G. C. Porter, "Error distribution and diversity performance of a frequency differential PSK HF modem," *IEEE Transactions on Communications*, vol. COM-16, pp. 567-575, Aug. 1968.

[43] J. G. Proakis, *Digital Communications*, 3rd edition, Singapore: McGraw-Hill Book Co, 767-768, 1995.

[44] J. G. Proakis and D. G. Manolaks, *Digital Signal Processing*, Third ed., Englewood Cliffs, NJ: Prentice-Hall, 1996.

[45] G. Leus, I. Barhumi, and M. Moonen, "MMSE time-varying FIR equalization of doubly-selective channels," in *Proc. Int. Conf. Acoust, Speech, Signal Process.*, vol. 4, pp. IV 485-488, Apr. 2003.

[46] G. Leus, I. Barhumi, and M. Moonen, "Low-complexity serial equalization of doubly-selective channels," in *Proc. of the Baiona Workshop on Signal Process. in Commun.*, pp. 69-74, Baiona, Spain, Sep. 2003.

[47] X. Li and T. F. Wong, "Turbo equalization with nonlinear Kalman filtering for time-varying frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 691-700, Feb. 2007.

[48] X. Ma, G. B. Giannakis and S. Ohno, "Optimal training for block transmissions over doubly selective channels," *IEEE Trans. Signal Processing*, vol. 51, no. 5, pp. 1352-1366, May 2003.

[49] C. F. Mecklenbräuker, J. Wehinger, T. Zemen, F. Hlawatsch, and H. Artés, *Smart Antennas - State of the Art*. EURASIP book series, Hindawi Publishing Co., Chap. 1.4., 2004.

[50] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933-946, May 2000.

[51] M. Niedzwiecki, *Identification of Time-Varying Processes*. New York: Wiley, 2000.

[52] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inform. Theory.*, vol. 50, pp. 2138-2145, Sept. 2004.

[53] M. F. Pop and N.C. Beaulieu, "Limitations of sum-of-sinusoids fading channel simulators," *IEEE Trans. Commun.*, vol. 49, no. 4, pp. 699-780, Apr. 2001.

[54] T.S. Rappaport, *Wireless Communications: Principles and Practice*. 2nd Ed., *Upper Saddle River, NJ: Prentice-Hall*, 2002.

[55] A. M. Sayeed and B. Aazhang, "Joint multipath-doppler diversity in mobile wireless communications," *IEEE Trans. on Commun.*, vol. 47, pp.123-132, Jan. 1999.

[56] A. M. Sayeed, A. Sendonaris, and B. Aazhang, "Multiuser detection in fast-fading multipath environment," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1691-1701, Dec. 1998.

[57] A.H. Sayed, *Fundamentals of Adaptive Filtering*. Wiley Interscience: Hoboken, NJ, 2003.

[58] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty - V: The discrete case," *Bell Syst. Tech. J.*, vol. 57, pp. 1371-1430, May-Jun. 1978.

[59] M.D. Srinath, P.K Rajasekaran and R. Viswanathen, *Introduction to Statistical Signal Processing with Applications*. Prentice-Hall: Upper Saddle River, NJ, 1996

[60] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2451-2462, Oct. 2002.

[61] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Upper Saddle River, NJ: Prentice-Hall 1997.

[62] Z. Tang, R. C. Cannizzaro, G. Leus, and P. Banelli, "Pilot-Assisted Time-Varying Channel Estimation for OFDM Systems," *IEEE Transactions on Signal Processing*, vol. 55, pp. 2226-2238, May. 2007.

[63] M. K. Tsatsanis and Z. Xu, "Pilot symbol assisted modulation in frequency selective fading wireless channels," *IEEE Trans. Signal Processing*, vol. 48, pp. 2353-2365, Aug. 2000.

[64] J.K. Tugnait, "Tracking of multiple maneuvering targets in clutter using multiple sensors, IMMand JPDA coupled filtering," *IEEE Trans. Aerospace & Electronic Systems*, vol. AES-40, pp. 320-330, Jan. 2004.

[65] J. K. Tugnait and X. Meng, "On Superimposed Training for Channel Estimation: Performance Analysis, Training Power Allocation, and Fram Synchronization," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 752-765, Feb. 2006.

[66] J. K. Tugnait, L. Tong, and Z. Ding, "Single-user channel estimation and equalization," *IEEE Signal Processing Mag.*, vol. 17, no. 3, pp. 17-28, May 2000.

[67] J. K. Tugnait and W. Luo "Linear prediction error method for blind identification of periodically time-varying channels" *IEEE Trans. Signal Processing*, vol. 50, no. 12, pp. 3070-3082, Dec. 2002.

[68] F. Verde, "Frequency-shift zero-forcing time-varying equalization for doubly selective SIMO channels," *EURASIP Journal on Applied Signal Process.*, artile ID: 47261, pp. 1-14, volume 2006.

[69] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "Optimal training for frequency-selective fading channels," in *Proc. Int. Conf. ASSP*, vol. 4, Salt Lake City, UT, pp. 2105-2108, May 7-11, 2001.

[70] X. Wang and H.V. Poor, "Joint channel estimation and symbol detection in Rayleigh flat-fading channels with impulsive noise," *IEEE Commun. Letters*, vol. 1, no. 1, Jan. 1997.

[71] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Processing Mag.*, vol. 47, pp. 29-48, May 2000.

[72] Z. Wang, X. Ma and G. B. Giannakis, "OFDM or single-carrier block transmissions?" *IEEE Trans. Commun.*, vol. 52, pp. 380-394, Mar. 2004.

[73] H. S. Wang and P.-C. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 353-357, May. 1996.

[74] X. Wang and H. V. Poor, "Joint channel estimation and symbol detection in Rayleigh flat-fading channels with impulsive noise," *IEEE Commun. Letters*, vol. 1, no. 1, Jan. 1997.

[75] L. Yang, X. Ma, and G. B. Giannakis, "Optimal training for MIMO fading channels with time- and frequency- selectivity," in *Proc. Int. Conf. Acoust, Speech, Signal Process.*, vol. 3, pp. 17-21, May. 2004.

[76] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel equalization via discrete prolate spheroidal Sequences," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, invited paper, pp. 1288-1292, Nov. 2003.

[77] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, pp. 3597-3607, Sept. 2005.

[78] Q. Zhao and L. Tong, "Semi-Blind equalization by least squares smoothing," in *Proc. IEEE 32nd Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, pp. 645-649, Nov. 1998.

[79] Y. Zhang, M. P. Fitz, and S. B. Gelfand, "A performance analysis and design of equalization with pilot aided channel estimation," in *Proc. 47th Veh. Technol. Conf.*, vol. 2, Phoenix, AZ, May 4-7, pp. 720-724, 1997.

[80] L. Zheng and D. N. C. Tse, "Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359-383, Feb. 2002.

[81] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, pp. 920-928, Jun. 2003.

[82] S. Zhou, B. Muquet, and G. B. Giannakis, "Subspace-based (semi-) blind channel estimation for block precoded space-time OFDM," *IEEE Trans. Signal Process.*, vol. 50, pp. 1215-1228, May 2002.

[83] M. S. Zimmerman and A. L. Kirsch, "The AN/GSC-10 (KATHRYN) variable rate data modem for HF radio," *IEEE Transactions on Communications*, vol. COM-15, pp. 197-205, Apr. 1967.

[84] X. Meng, *Estimation of Wireless Communications Using Superimposed Training: Approaches, Analysis, and Applications*, Ph.D. dissertation, Auburn University, Auburn, AL, May 2005.

[85] S. He, *Doubly-Selective Channel Estimation and Equalization Using Superimposed Training and Basis Expansion Models*, Ph.D. dissertation, Auburn University, Auburn, AL, Aug. 2007.

APPENDIX A

ASYMPTOTIC DPS-BEM/ SLEPIAN SEQUENCES

The Slepian sequences as solutions to (2.5) are hard to work with analytically. Following [61, Sec. 5.3.1] (in a different context), we will use asymptotic expressions (N "large") based on some heuristic considerations (as in [61, Sec. 5.3.1]). The entries in matrix $\boldsymbol{C}$ can also be expressed as

$$[\boldsymbol{C}]_{y,z} = \frac{1}{2\pi} \int_{-2\pi\nu_{\mathrm{Dmax}}}^{2\pi\nu_{\mathrm{Dmax}}} e^{-i(y-z)\omega} d\omega. \tag{A.1}$$

Define $\beta := 2\nu_{\mathrm{Dmax}}$ and $Q_S := N\beta$. Let β be chosen larger than $1/N$ so that we get $Q_S \geq 1$. (To simplify the discussion, Q_S and N are assumed to be even integers in what follows.) For large N, we can approximate the integral in $(A.1)$ as ($\omega = \frac{2\pi}{N}p$):

$$\boldsymbol{C} \simeq \frac{1}{N} \sum_{p=-Q_S/2}^{Q_S/2} \boldsymbol{a}\left(\frac{2\pi}{N}p\right) \boldsymbol{a}^*\left(\frac{2\pi}{N}p\right) := \boldsymbol{C}_0, \tag{A.2}$$

where $\boldsymbol{a}(\omega) := [1, e^{-i\omega}, ..., e^{-i\omega(N-1)}]^T$, and λ_ps are the $Q_S + 1$ largest eigenvalues of the matrix $\boldsymbol{C}$. Based on (2.24), the eigenvectors of the matrix $\boldsymbol{C}_0$ are derived as $\left\{\frac{1}{\sqrt{N}}\boldsymbol{a}(\frac{2\pi}{N}p)\right\}_{p=-Q_S/2}^{Q_S/2}$ with eigenvalues ± 1. Therefore, if $\boldsymbol{C}_0$ is used to approximate $\boldsymbol{C}$, the Slepian sequences in (2.2) and (2.5) can be expressed as

$$\psi_q^{(S)}(n) = \frac{1}{\sqrt{N}} e^{(-i2\pi(q-Q_S/2)n/N)}. \tag{A.3}$$

[As noted in [61, Sec. 5.3.1], $\boldsymbol{C}_0$ does not approximate $\boldsymbol{C}$ in any rigorous sense.]

Obviously, the approximate Slepian sequences in $(A.3)$ are complex-valued. Since the Slepian sequences are real, we want to find a matrix with real eigenvectors that can approximate $\boldsymbol{C}$ based on

$\boldsymbol{C}_0$. Since $\boldsymbol{C}_0$ is a real symmetric positive semi-definite matrix with real eigenvalues ± 1, it readily follows that $\boldsymbol{g}_r := \frac{1}{2\sqrt{N}}(\boldsymbol{a} + \boldsymbol{a}^*)$ and $\boldsymbol{g}_i := \frac{1}{2i\sqrt{N}}(\boldsymbol{a} - \boldsymbol{a}^*)$ are also the eigenvectors of $\boldsymbol{C}_0$ and they are real. From $(A.2)$, we notice that

$$\boldsymbol{a}^* \left(\frac{2\pi}{N}p\right) = \boldsymbol{a}\left(-\frac{2\pi}{N}p\right). \tag{A.4}$$

Therefore, we can rewrite the matrix $\boldsymbol{C}_0$ as

$$\boldsymbol{C}_0 = \sum_{p=-Q_S/2}^{0} \boldsymbol{g}_r\left(\frac{2\pi}{N}p\right)\boldsymbol{g}_r^*\left(\frac{2\pi}{N}p\right) + \sum_{p=-Q_S/2}^{1} \boldsymbol{g}_i\left(\frac{2\pi}{N}p\right)\boldsymbol{g}_i^*\left(\frac{2\pi}{N}p\right). \tag{A.5}$$

With this expression of matrix $\boldsymbol{C}_0$, the normalized eigenvectors of matrix $\boldsymbol{C}_0$ are expressed as $\left\{\boldsymbol{g}_r(\frac{2\pi}{N}p)\right\}_{p=-Q_S/2}^{0}$ and $\left\{\boldsymbol{g}_i(\frac{2\pi}{N}p)\right\}_{p=-Q_S/2}^{1}$. Thus, the corresponding approximate Slepian sequences for $n \in \{0, ..., N-1\}$ are

$$\psi_{q,r}^{(S)}(n) = \sqrt{\frac{2}{N}}\cos(2\pi(q - \frac{Q_S}{2})n/N), \quad \psi_{q,i}^{(S)}(n) = \sqrt{\frac{2}{N}}\sin(2\pi(q - \frac{Q_S}{2})n/N). \tag{A.6}$$

APPENDIX B

MATHEMATICAL NOTATIONS

$\approx$	approximately equal to
$\otimes$	Kronecker product
$\mathbf{0}_{M\times N}$	$M \times N$ all zeros matrix
a	lower case letters for scalars
$\lceil a \rceil$	integer ceiling of a
$\lfloor a \rfloor$	integer floor of a
$\lvert a \rvert$	magnitude of a
$\boldsymbol{a}$	lower case letters in bold face for column vectors
$\lVert \boldsymbol{a} \rVert$	Frobenius norm of $\boldsymbol{a}$
$\boldsymbol{A}$	upper case letters in bold face for matrices
$\boldsymbol{A}^*$	complex conjugate of $\boldsymbol{A}$
$\boldsymbol{A}^\dagger$	Moore-Penrose pseudo-inverse operation
$\boldsymbol{A}^H$	complex conjugate transpose of $\boldsymbol{A}$
$\boldsymbol{A}^T$	transpose of $\boldsymbol{A}$
$[\boldsymbol{A}]_{n,m}$	(n, m)-th entry of $\boldsymbol{A}$
$\boldsymbol{A}$	upper case calligraphic letters for matrices
$\arg\max_x f(x)$	value of x for which $f(x)$ attains its maximum
$\arg\min_x f(x)$	value of x for which $f(x)$ attains its minimum
$\mathrm{cov}\{\cdot\}$	covariance operator

$\delta(\cdot)$	Kronecker delta function, defined as $\delta(n) = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0, t \in \mathbb{Z} \end{cases}$
$J_0(\cdot)$	the zero-order Bessel function of the first kind
$\mathrm{diag}[\boldsymbol{a}]$	diagonal matrix with $\boldsymbol{a}$ on its main diagonal
$E\{\cdot\}$	expectation operator
$E_{\mathcal{H}}\{\cdot\}$	expectation operator with respect to $\mathcal{H}$
$\boldsymbol{I}_N$	$N \times N$ identity matrix
$\max(\cdot)$	maximum value operator
$\min(\cdot)$	minimum value operator
$O(\cdot)$	big O notation: $f(x) = O(g(x))$ as $x \to a$ ($a \in \mathbb{R} \cup \pm\infty$), iff $\lvert f(x)\rvert \leq M \lvert g(x)\rvert$ as $x \to a$ for some constant $M > 0$
$\mathbb{R}$	real field
$\mathrm{tr}\{\boldsymbol{A}\}$	trace of a square matrix $\boldsymbol{A}$
$\mathbb{Z}$	integer field
j	$\sqrt{-1}$

APPENDIX C

ABBREVIATIONS

AM	amplitude modulation
AR	auto-regressive
AWGN	additive white Gaussian noise
BEM	basis expansion model
BER	bit error rate
BPSK	binary phase-shift keying
CE-BEM	complex exponential basis expansion model
CP	cyclic prefex
CSI	channel state information
DAB	digital audio broadcasting
DFS	discrete Fourier series DFS
DFT	discrete Fourier transform
DML	deterministic maximum likelihood
DMT	discrete multi-tone
DPS	discrete prolate spheroidal
DPS-BEM	discrete prolate spheroidal basis expansion model
DS-CDMA	direct sequence - code division multiple access
DSL	digital subscriber line
DVB	digital video broadcasting

FB	feedback
FF	feedforward
FFT	fast Fourier transform
FIR	finite impulse response
FM	frequency modulation
IBI	interblock interference
IMM	interacting multiple model
ISBI	inter-subblock interference
ISI	inter-symbol interference
LS	least squares
LTI	linear time-invariant
MAC	media access control
MC	multicarrier
MIMO	multiple-input multiple-output
ML	maximum likelihood
MLSE	maximum likelihood sequence estimation
MSE	mean square error
MMSE	minimum mean square error
m.s.	mean-square
MUI	multiple-user interference

NCMSE	normalized channel mean square error
OFDM	orthogonal frequency division multiplexing
PAM	pulse amplitude modulation
pdf	probability density function
PN	pseudo-noise
PSAM	pilot symbol aided modulation
QPSK	quadrature phase-shift keying
RF	Radio Frequency
SIMO	single-input multiple-output
SISO	single-input single-output
SNR	signal-to-noise ratio
TI	time invariant
TM	time-multiplexed
TV	time-varying
WSSUS	wide sense stationary uncorrelated scattering