

DOUBLY-SELECTIVE CHANNEL ESTIMATION AND EQUALIZATION USING SUPERIMPOSED
TRAINING AND 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.

Shuangchi He

Certificate of Approval:

Stanley J. Reeves
Professor
Electrical and Computer Engineering

Jitendra K. Tugnait, Chair
Professor
Electrical and Computer Engineering

Soo-Young Lee
Professor
Electrical and Computer Engineering

Joe F. Pittman
Interim Dean
Graduate School

DOUBLY-SELECTIVE CHANNEL ESTIMATION AND EQUALIZATION USING SUPERIMPOSED
TRAINING AND BASIS EXPANSION MODELS

Shuangchi He

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
August 4, 2007

DOUBLY-SELECTIVE CHANNEL ESTIMATION AND EQUALIZATION USING SUPERIMPOSED
TRAINING AND BASIS EXPANSION MODELS

Shuangchi He

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

DISSERTATION ABSTRACT

DOUBLY-SELECTIVE CHANNEL ESTIMATION AND EQUALIZATION USING SUPERIMPOSED
TRAINING AND BASIS EXPANSION MODELS

Shuangchi He

Doctor of Philosophy, August 4, 2007
(M.S., Tsinghua University, 2003)
(B.E., Tsinghua University, 2000)

261 Typed Pages

Directed by Jitendra K. Tugnait

Owing to multipath propagation and Doppler spread, typical wireless channels are both frequency- and time-selective (doubly-selective). In this dissertation, we concentrate on channel estimation and equalization over doubly-selective channels, by exploiting both superimposed training and basis expansion models (BEM).

In contrast to the conventional time-multiplexed (TM) training schemes, at the transmitter, a periodic training sequence is arithmetically added at low power to the information sequence in superimposed training schemes. There is no loss in data transmission rate, but some useful power has to be allocated to superimposed training. We also employ various BEM's to describe the temporal variations of the doubly-selective channel so that the estimation of a time-varying process can be reduced to estimating fewer invariant BEM coefficients.

Firstly, a channel estimator is presented using superimposed training and the first-order statistics of the observations, based on various BEM's, where information sequences act as

interference in channel estimation. By using user-specific training sequences, the estimator can be extended to multiple-user systems.

We next analyze the information-induced self-interference of this estimator. The performance analysis and the parameter optimizations are investigated.

We propose two schemes to alleviate the self-interference in channel estimation. Using the channel estimates by the first-order statistics-based estimator as an initial guess, a deterministic maximum likelihood (DML) approach is used to jointly estimate the channel and the information sequence. Exploiting the channel estimates and the detected information data from the previous iteration, the self-interference can be significantly reduced at the present iteration. We also propose a data-dependent superimposed training scheme. The training sequence is designed based on the current information sequence so that the self-interference can be entirely eliminated at the receiver. However, total elimination of the interference may lead to information loss. We then modify the scheme to the partially-data-dependent (PDD) training, striking a compromise between interference cancelation and information integrity.

Using superimposed training and a BEM, direct equalization of doubly-selective channels is also considered, without estimating the channel first. The direct equalizer is also extended to a multiple-user scenario, which can be used in a wireless *ad hoc* network.

The proposed approaches are illustrated by computer simulation examples, and compared with conventional TM training-based approaches. When self-interference is sufficiently suppressed by our proposed schemes, the performance of superimposed training-based approaches are competitive with the ones using the conventional TM training, without incurring any data-rate loss.

ACKNOWLEDGMENTS

I have had a phenomenal time during my four years in Auburn. I attribute my good fortune to the wonderful people I have come to know through the past four years.

First and foremost, I would like to thank my advisor, Prof. Jitendra K. Tugnait, for his generosity and support over the years. As the guide of my research career, he gave me careful and rigorous instructions, from which I will surely benefit for life.

Many thanks also go to my committee members, Profs. Stanley J. Reeves and Soo-Young Lee; they have provided me with invaluable guidance and friendliness during my studies and contributed much to my dissertation work. Thanks to Prof. Douglas A. Leonard who served as my outside reader, for his suggestions and instruction. These acknowledgements would be far from complete without thanks to Profs. Tin-Yau Tam, Xiaoli Ma, and Shiwen Mao, for their sage advice during my study.

I want to thank Weilin Luo and Xiaohong Meng for their foundational work relating to my dissertation—and especially Xiaohong, for her elegant simulation programs that offered me a good model to follow.

Finally, I would like to express my gratitude to my parents and my wife, Fan, whose love and faith in me are the source of my strength.

My studies were funded by National Science Foundation under Grant ECS 0424145 and Vodafone Fellowship.

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 T_EX (specifically L^AT_EX) together with the departmental style-file `aums.sty`.

TABLE OF CONTENTS

LIST OF FIGURES		xi
1	INTRODUCTION	1
1.1	Previous Work	3
1.2	Contributions	8
1.3	Organization	10
2	REPRESENTATIONS OF WIRELESS CHANNELS	11
2.1	Introduction	11
2.2	Jakes' Model	13
2.3	Complex Exponential Basis Expansion Model (CE-BEM)	16
2.4	Orthogonal Polynomial Basis Expansion Model (OP-BEM)	18
2.5	Discrete Prolate Spheroidal Basis Expansion Model (DPS-BEM)	20
2.6	Modeling Error of BEM's	23
2.6.1	LS Approximation by Basis Expansion Models	23
2.6.2	Simulation Model	24
2.6.3	Simulation Example: Modeling Error of CE-, OP-, and DPS-BEM's in Approximating a Doubly-Selective Channel	25
2.7	Conclusions	27
3	FIRST-ORDER STATISTICS-BASED ESTIMATION OF DOUBLY-SELECTIVE CHANNELS	28
3.1	Introduction	28
3.2	First-Order Statistics-Based Channel Estimation Using CE-BEM [81]	28
3.3	First-Order Statistics-Based Channel Estimation Using DPS-BEM	36
3.4	First-Order Statistics-Based Channel Estimation Using OP-BEM	40
3.5	First-Order Statistics-Based Channel Estimation: Multiple-User (MIMO) Channels	47
3.6	Simulation Examples	59
3.6.1	First-Order Statistics-Based Estimator: Single User	59
3.6.2	First-Order Statistics-Based Estimator: Multiple Users	64
3.7	Conclusions	69
4	PERFORMANCE ANALYSIS AND PARAMETER DESIGN FOR FIRST-ORDER STATISTICS-BASED ESTIMATOR	71
4.1	Introduction	71
4.2	Performance Analysis for the First-Order Statistics-Based Estimator Using BEM	72

4.2.1	Performance Analysis for CE-BEM-Based Estimator	74
4.2.2	Performance Analysis for DPS-BEM-Based Estimator	77
4.2.3	Performance Analysis for OP-BEM-Based Estimator	81
4.2.4	Performance Analysis for Multiple-User (MIMO) Channels	84
4.3	Performance Analysis for the First-Order Statistics-Based Estimator: with Modeling Error	89
4.4	Training Power Allocation	93
4.5	Bias-Variance Trade-Off	99
4.6	Simulation Examples	106
4.6.1	Performance Analysis for the First-Order Statistics-Based Estimator	106
4.6.2	Training Power Allocation	111
4.6.3	Bias-Variance Trade-Off	114
4.7	Conclusions	116
5	DETERMINISTIC MAXIMUM LIKELIHOOD (DML) APPROACH	118
5.1	Introduction	118
5.2	DML Approach Using BEM	118
5.3	DML Approach: Multiple-User (MIMO) Channels	125
5.4	Simulation Examples	132
5.4.1	DML Approach: Single User	132
5.4.2	DML Approach: Multiple Users	139
5.5	Conclusions	142
6	DOUBLY-SELECTIVE CHANNEL ESTIMATION USING DATA-DEPENDENT SUPERIMPOSED TRAINING	144
6.1	Introduction	144
6.2	Data-Dependent Superimposed Training Using CE-BEM	145
6.2.1	Data-Dependent Processing at the Transmitter	146
6.2.2	Data Detection	148
6.2.3	Performance Analysis	149
6.3	Data-Dependent Superimposed Training Using DPS-BEM	151
6.3.1	Partially-Data-Dependent (PDD) Superimposed Training	153
6.3.2	Performance Analysis	154
6.3.3	Power Allocation and Self-Interference Suppression	156
6.3.4	Recovery of Suppressed Frequencies via DML Approach	163
6.4	Simulation Examples	167
6.4.1	Data-Dependent Superimposed Training Using CE-BEM	167
6.4.2	(Partially) Data-Dependent Superimposed Training Using DPS-BEM	174
6.5	Conclusions	183

7	DIRECT FIR LINEAR EQUALIZATION OF DOUBLY-SELECTIVE CHANNELS BASED ON SUPERIMPOSED TRAINING	185
7.1	Introduction	185
7.2	Direct FIR Linear Equalization Using CE-BEM	186
7.2.1	Time-Varying FIR Equalizers	187
7.2.2	Linear LS Equalizers Based on CE-BEM	188
7.3	Direct FIR Linear Equalization: Multiple Users	196
7.3.1	User-Specific Training Sequences	199
7.3.2	Linear LS Equalizers for the Desired User	199
7.4	Simulation Examples	204
7.4.1	Direct FIR Equalization: Single User	204
7.4.2	Direct FIR Equalization: Multiple Users	207
7.5	Conclusions	210
8	CONCLUDING REMARKS AND FUTURE WORK	211
8.1	Summary of Original Work	212
8.2	Possible Future Directions	214
	BIBLIOGRAPHY	216
	APPENDICES	224
A	OPTIMAL TIME-MULTIPLEXED TRAINING FOR BLOCK TRANSMISSIONS OVER DOUBLY-SELECTIVE CHANNELS [40]	225
B	SYMBOL DETECTION	230
B.1	Maximum Likelihood Sequence Detector (Viterbi Detector) [64]	230
B.2	Kalman Filtering	234
C	MATHEMATICAL NOTATIONS	239
D	ABBREVIATIONS	241

LIST OF FIGURES

2.1	Modeling error of CE-, OP-, and DPS-BEM's in approximating a three-tap ($L = 2$) Rayleigh fading channel following Jakes' model.	25
3.1	First-order statistics-based estimator (SISO): BER vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 1$. The curves for CE-, OP- and DPS-BEM's completely overlap, since the three basis functions are all constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.) . . .	55
3.2	First-order statistics-based estimator (SISO): BER vs SNR under $f_d = 50$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)	56
3.3	First-order statistics-based estimator (SISO): BER vs SNR under $f_d = 100$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)	57
3.4	First-order statistics-based estimator (SISO): BER vs SNR under $f_d = 200$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)	58
3.5	First-order statistics-based estimator (SISO): NCMSE vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 1$. The curves for CE-, OP- and DPS-BEM's completely overlap, since the three basis functions are all constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.) . . .	59
3.6	First-order statistics-based estimator (SISO): NCMSE vs SNR under $f_d = 50$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)	60
3.7	First-order statistics-based estimator (SISO): NCMSE vs SNR under $f_d = 100$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)	61
3.8	First-order statistics-based estimator (SISO): NCMSE vs SNR under $f_d = 200$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)	62

3.9	First-order statistics-based estimator (MIMO): BER vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 2$. The curves for CE- and DPS-BEM's completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; Kalman: a Kalman filter as the symbol detector; Viterbi: a Viterbi detector as the symbol detector.)	63
3.10	First-order statistics-based estimator (MIMO): BER vs SNR under $f_d = 50$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; Kalman: a Kalman filter as the symbol detector; Viterbi: a Viterbi detector as the symbol detector.)	64
3.11	First-order statistics-based estimator (MIMO): BER vs SNR under $f_d = 100$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; Kalman: a Kalman filter as the symbol detector; Viterbi: a Viterbi detector as the symbol detector.)	65
3.12	First-order statistics-based estimator (MIMO): BER vs SNR under $f_d = 200$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; Kalman: a Kalman filter as the symbol detector; Viterbi: a Viterbi detector as the symbol detector.)	66
3.13	First-order statistics-based estimator (MIMO): NCMSE vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 2$. The curves for CE- and DPS-BEM's completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM.)	67
3.14	First-order statistics-based estimator (MIMO): NCMSE vs SNR under $f_d = 50$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM.)	68
3.15	First-order statistics-based estimator (MIMO): NCMSE vs SNR under $f_d = 100$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM.)	69
3.16	First-order statistics-based estimator (MIMO): NCMSE vs SNR under $f_d = 200$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM.)	70

4.1	Estimation variance: NCMSE vs SNR under $f_d = 0$ Hz (time-invariant). The curves for CE-, OP- and DPS-BEM's completely overlap, since the three basis functions are all constant for time-invariant channels ($Q = 1$). (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM; MSE_1 : defined in (4.2); MSE: defined in (4.33).)	103
4.2	Estimation variance: NCMSE vs SNR under $f_d = 50$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM; MSE_1 : defined in (4.2); MSE: defined in (4.33).)	104
4.3	Estimation variance: NCMSE vs SNR under $f_d = 100$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM; MSE_1 : defined in (4.2); MSE: defined in (4.33).)	105
4.4	Estimation variance: NCMSE vs SNR under $f_d = 200$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM; MSE_1 : defined in (4.2); MSE: defined in (4.33).)	106
4.5	Training power allocation: BER vs β under $f_d = 0$ Hz (time-invariant). The curves for CE-, OP- and DPS-BEM's completely overlap, since the three basis functions are all constant for time-invariant channels ($Q = 1$). (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.) . .	107
4.6	Training power allocation: BER vs β under $f_d = 50$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)	108
4.7	Training power allocation: BER vs β under $f_d = 100$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)	109
4.8	Training power allocation: BER vs β under $f_d = 200$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)	110
4.9	Training power allocation: optimum β vs SNR for CE-BEM. (“sim.”: simulation results; “analy.”: β_{opt} in (4.46); “analy. app.”: β'_{opt} in (4.48).)	111
4.10	Training power allocation: optimum β vs SNR for OP-BEM. (“sim.”: simulation results; “analy.”: β_{opt} in (4.46); “analy. app.”: β'_{opt} in (4.48).)	112
4.11	Training power allocation: optimum β vs SNR for DPS-BEM. (“sim.”: simulation results; “analy.”: β_{opt} in (4.46); “analy. app.”: β'_{opt} in (4.48).) . . .	113
4.12	Bias-variance trade-off: BER vs Q under $TIR = 0.3$ for different f_d 's.	114

4.13	Bias-variance trade-off: $\text{SNR}_d(Q)$ (defined in (4.50)) vs Q under TIR = 0.3 for different f_d 's.	115
4.14	Bias-variance trade-off: BER vs Q under TIR = 1.0 for different f_d 's.	116
4.15	Bias-variance trade-off: $\text{SNR}_d(Q)$ (defined in (4.50)) vs Q under TIR = 1.0 for different f_d 's.	117
5.1	DML approach (SISO): BER vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 1$. The curves for CE- and DPS-BEM's completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; "step 1": the first-order statistics-based estimator; "1st iter.": the first DML iteration; "2nd iter.": the second DML iteration; "3rd iter.": the third DML iteration.)	128
5.2	DML approach (SISO): BER vs SNR under $f_d = 50$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; "step 1": the first-order statistics-based estimator; "1st iter.": the first DML iteration; "2nd iter.": the second DML iteration; "3rd iter.": the third DML iteration.)	129
5.3	DML approach (SISO): BER vs SNR under $f_d = 100$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; "step 1": the first-order statistics-based estimator; "1st iter.": the first DML iteration; "2nd iter.": the second DML iteration; "3rd iter.": the third DML iteration.)	130
5.4	DML approach (SISO): BER vs SNR under $f_d = 200$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; "step 1": the first-order statistics-based estimator; "1st iter.": the first DML iteration; "2nd iter.": the second DML iteration; "3rd iter.": the third DML iteration.)	131
5.5	DML approach (SISO): NCMSE vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 1$. The curves for CE- and DPS-BEM's completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; "step 1": the first-order statistics-based estimator; "1st iter.": the first DML iteration; "2nd iter.": the second DML iteration; "3rd iter.": the third DML iteration.)	132

5.6	DML approach (SISO): NCMSE vs SNR under $f_d = 50$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)	133
5.7	DML approach (SISO): NCMSE vs SNR under $f_d = 100$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)	134
5.8	DML approach (SISO): NCMSE vs SNR under $f_d = 200$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)	135
5.9	DML approach (MIMO): BER vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 2$. The curves for CE- and DPS-BEM’s completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)	136
5.10	DML approach (MIMO): BER vs SNR under $f_d = 50$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)	137
5.11	DML approach (MIMO): BER vs SNR under $f_d = 100$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)	138
5.12	DML approach (MIMO): BER vs SNR under $f_d = 200$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)	139

5.13	DML approach (MIMO): NCMSE vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 2$. The curves for CE- and DPS-BEM's completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; "step 1": the first-order statistics-based estimator; "1st iter.": the first DML iteration; "2nd iter.": the second DML iteration; "3rd iter.": the third DML iteration.)	140
5.14	DML approach (MIMO): NCMSE vs SNR under $f_d = 50$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; "step 1": the first-order statistics-based estimator; "1st iter.": the first DML iteration; "2nd iter.": the second DML iteration; "3rd iter.": the third DML iteration.)	141
5.15	DML approach (MIMO): NCMSE vs SNR under $f_d = 100$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; "step 1": the first-order statistics-based estimator; "1st iter.": the first DML iteration; "2nd iter.": the second DML iteration; "3rd iter.": the third DML iteration.)	142
5.16	DML approach (MIMO): NCMSE vs SNR under $f_d = 200$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; "step 1": the first-order statistics-based estimator; "1st iter.": the first DML iteration; "2nd iter.": the second DML iteration; "3rd iter.": the third DML iteration.)	143
6.1	Data-dependent superimposed training (CE-BEM): BER vs SNR for non-data-dependent, data-dependent, and time-multiplexed training, under $f_d = 0$ and 50 Hz. (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)	167
6.2	Data-dependent superimposed training (CE-BEM): BER vs SNR for non-data-dependent, data-dependent, and time-multiplexed training, under $f_d = 100$ and 200 Hz. (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)	168
6.3	Data-dependent superimposed training (CE-BEM): NCMSE vs SNR for non-data-dependent, data-dependent, and time-multiplexed training, under $f_d = 0$ and 50 Hz. (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)	169

6.4	Data-dependent superimposed training (CE-BEM): NCMSE vs SNR for non-data-dependent, data-dependent, and time-multiplexed training, under $f_d = 100$ and 200 Hz. (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)	170
6.5	Data-dependent superimposed training (CE-BEM): BER vs SNR for non-data-dependent, data-dependent, and time-multiplexed training, under $f_d = 100$ Hz and $N = 1, 2,$ and 3 . (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)	171
6.6	Data-dependent superimposed training (fast fading): BER vs SNR under $f_d = 100$ and 250 Hz. (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)	172
6.7	Estimation variance: NCMSE vs f_d under SNR = 25 dB for comparison between analytical and simulation-based results of non-data-dependent and data-dependent superimposed training. ($\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; σ : standard deviation.)	173
6.8	PDD superimposed training: NCMSE vs SNR for CE- and DPS-BEM-based estimators, under $f_d = 100$ Hz. ($\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)	174
6.9	PDD superimposed training: BER vs SNR for CE- and DPS-BEM-based estimators, under $f_d = 100$ Hz. ($\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)	175
6.10	PDD superimposed training: BER vs (β, γ) under SNR = 15 dB and $f_d = 100$ Hz.	176
6.11	PDD superimposed training: optimum (β, γ) vs SNR under $f_d = 100$ Hz.	177
6.12	PDD superimposed training: BER vs SNR under $f_d = 0$ Hz (time-invariant). (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)	178

6.13	PDD superimposed training: BER vs SNR under $f_d = 100$ Hz. (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)	179
6.14	PDD superimposed training: BER vs SNR under $f_d = 200$ Hz. (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)	180
6.15	PDD superimposed training: NCMSE vs SNR under $f_d = 0$ Hz (time-invariant). (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)	181
6.16	PDD superimposed training: NCMSE vs SNR under $f_d = 100$ Hz. (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)	182
6.17	PDD superimposed training: NCMSE vs SNR under $f_d = 200$ Hz. (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)	183
7.1	Single-user direct FIR equalization: BER vs SNR under $f_d = 0$ Hz and length of equalizer $L_e = 6$ with different TIR and number of receivers.	205
7.2	Single-user direct FIR equalization: BER vs SNR under $f_d = 50$ Hz and length of equalizer $L_e = 6$ with different TIR and number of receivers.	206
7.3	Single-user direct FIR equalization: BER vs SNR under $f_d = 100$ Hz and length of equalizer $L_e = 6$ with different TIR and number of receivers.	207
7.4	Multiple-user direct FIR equalization (<i>ad hoc</i>): BER vs SNR under $f_d = 100$ Hz and length of equalizer $L_e = 4$ with different number of receivers.	208

7.5	Multiple-user direct FIR equalization (<i>ad hoc</i>): BER vs Doppler spread f_d under SNR = 25 dB and length of equalizer $L_e = 4$ with different number of receivers.	209
7.6	Multiple-user direct FIR equalization (<i>ad hoc</i>): BER vs length of equalizer L_e under $f_d = 50$ Hz and SNR = 25 dB with different number of receivers.	210

CHAPTER 1

INTRODUCTION

With the emergence of next-generation wireless mobile communications, multimedia services have increasing demands for higher data rates, better quality of service, and higher network capacity. In efforts to support such demands, researchers have paid special attention to wireless channels. Phenomena occurring in wireless channels, such as fading, delay spread, Doppler spread, co-channel interference, and multi-user interference, may impair signal transmission and data reception. A wireless channel is a challenging communications medium with limited bandwidth, relatively low capacity per unit bandwidth, random amplitude and phase fluctuations, and inter-symbol interference (ISI). To design a physical link with data rates approaching the fundamental information capacity limits of the wireless channel, accurate knowledge of the channel state information (CSI) becomes a prerequisite for many physical layer approaches. Channel estimation thus plays a key role. At the receive ends, equalizers are usually used to compensate for the signal distortion. One may design an equalizer based on a channel estimate, or by directly using the received signals.

Traditionally, receivers rely on a transmitter-assisted training session to extract the desired reference signal for channel estimation or equalization [64]. In a fast-varying environment, training sessions have to be transmitted frequently and periodically to keep up with the temporal variation of the channel. For a band-limited wireless application, frequent use of training sessions decreases the effective information rate. To save valuable spectrum resources, blind (self-recovering) channel estimation and equalization, based solely on the noisy received data, exploiting the statistical or other properties, has attracted researchers'

interest, where no training sessions are available nor are used [23]. Semi-blind channel estimation, combining explicit (time-multiplexed) training and blind cost functions, has also attracted considerable attention due to the need for fast and robust channel estimation and the fact that, for many packet transmission systems, embedded known symbols can be exploited for channel estimation. In semi-blind approaches, there are training sessions but one uses information data also to improve the training-based results [23,87]. More recently, superimposed training-based approaches have been explored where the training sequence is “on” all the time, and is transmitted (at low power) concurrently with (superimposed on) the information data. In contrast to explicit training, there is no loss in data transmission rate. On the other hand, some useful power is wasted in superimposed training sequences which could have otherwise been allocated to the information data.

In this dissertation, we will discuss doubly-selective channel estimation and equalization using superimposed training. Common wireless channels are frequency-selective (due to delay spread and multipath propagation) and time-selective (due to mobility). An accurate model of realistic wireless channels can be complicated and involve too many parameters for estimation purposes. Therefore, a parsimonious representation is preferred. We employ basis expansion models (BEM) to represent the doubly-selective channel with many fewer parameters [24]. In a BEM, the channel is represented as a finite impulse response (FIR) filter where each tap is a superposition of distinct basis functions that describe the temporal variations of the channel. Three BEM’s are considered: the complex exponential basis expansion model (CE-BEM), the orthogonal polynomial basis expansion model (OP-BEM), and the discrete prolate spheroidal basis expansion model (DPS-BEM).

1.1 Previous Work

In this section, we summarize the previous research work on superimposed training-based channel estimation, equalization, and related areas.

To the best of our knowledge, the idea of superimposed training (simultaneous transmission of information-bearing signal and channel sounding) was first proposed in [36] in 1965 for analog communications, where a pseudo-random channel sounding signal was superimposed upon a frequency-modulated (FM) information-bearing signal by amplitude modulation (AM). This idea was extended to digital systems in [17] in 1995, where both least squares (LS) and least mean squares (LMS) methods were considered to build an adaptive filter, treating the known superimposed training sequence as the input and the received signal as the desired output. Periodic superimposed training sequences allowed for the use of first-order statistics (time-varying mean) of the received signal, which were also exploited for time-invariant channel estimation in [59, 82, 98], among others. Using CE-BEM, such periodic superimposed training schemes were extended to doubly-selective channel environments in [81, 97]. Direct design of FIR equalizers using periodic superimposed training was investigated in [58] for time-invariant channels. The Cramér-Rao lower bound (CRLB) on channel estimation variance was given in [98], under the assumption of Gaussian source symbols for a special class of training sequences. Such bounds were extended to a general class of training sequences in [59]. Non-periodic random or pseudo-noise (PN) superimposed training sequences (known at the receiver) were used in [33, 44]. A linear predictor was designed in [33] to estimate the time-varying flat fading channel, and based on the minimum mean square error (MMSE) criterion channel estimation and equalization were discussed in [44] for M -quadrature amplitude modulated (QAM) symbols.

The formulations of [59, 82] allowed for the presence of an unknown “direct current” (DC) offset at the receiver, whereas [17, 98] did not. The two schemes in [59, 82] were compared by [45], where their structural equivalence was verified and therefore identical estimates would be got for zero (or known) DC offset. In the presence of an unknown DC offset, the basic approach of [59] yielded biased channel estimates, so that estimation of the DC offset was required by using the biased estimates and received data by finding the roots of a fifth-degree polynomial [59]. In contrast, the method of [82] yielded unbiased channel estimates directly. Performance analysis (a closed-form solution for the channel estimation variance) was also performed in [59] for zero (or known) DC offset, which was then used for an optimal training sequence synthesis to yield a channel-independent performance. Unfortunately, the synthesized training sequences in [59] do not necessarily have a small peak-to-average power ratio, whereas that of the training sequence in [82] has the optimal value of one. The performance analysis that is valid for any DC offset of the approach proposed by [82] was conducted in [84, 85], where power allocation for superimposed training in Rayleigh fading channels was also addressed by maximizing the equivalent signal-to-noise ratio (SNR) for equalizer design under a fixed power constraint. As in [17], the period of the superimposed training sequence of [59] was equal to the number of channel taps, whereas this condition was relaxed in [82] to be greater or equal to the number of channel taps. Synchronization of the training sequence (frame synchronization), based on the correlation and the fourth-order cumulant functions of the observations, was also discussed in [59]. Under mis-synchronization, however, the estimated channel will consequently yield a circularly shifted estimate whose “shift” cannot be resolved via the first-order statistics of the data [59]. Synchronization of the approach in [82] was discussed in [83, 85], where

the problem of shifted channel estimate was avoided. A synchronization technique based on subspace projections was discussed in [1]. The estimator proposed in [82] offers the fundamental approach to channel estimation in this dissertation.

To exploit the enormous capacity potential of multiple-input multiple-output (MIMO) communications [18], superimposed training-based channel estimation was considered in [4, 5, 21, 47, 48] for MIMO systems. Superimposed training in multi-carrier systems was considered in [9, 10, 13, 93]. We also note that a more general framework of superimposed training, engaged in affine precoding, has now attracted much interest and was investigated in [42, 56, 89], among others.

Since superimposed training-based methods usually use statistical properties of the information data, they can be treated as semi-blind approaches [17]. In contrast to slow convergence and possible convergence toward incorrect solutions occurring in blind methods [92], identifiability conditions for superimposed training-based methods are much less stringent [17]. Furthermore, blind approaches cannot resolve complex scaling factor ambiguity, so that differential coding and decoding resulting in 3 dB SNR loss is required [85], whereas power allocated to superimposed training is typically much less than 3 dB (1 dB or less in [17]).

In superimposed training-based approaches, the unknown information data are typically incorporated in the noise term, in essence yielding a lower SNR [73]. In other words, the information data may act as interference at the receiver to the superimposed training and adversely affect channel estimation and data detection performance—no loss in data transmission rate might be at the price of degradation in data reception.

Several methods were investigated to reduce the interference from information data (we call it *self-interference* in this dissertation since it comes from the transmitted signal itself). A selective superimposed training scheme was proposed in [76], where the selection of the training sequence depends on each frame of information data. From a candidate set of orthogonal sequences maintained at the transmitter, a training sequence is chosen and superimposed on the incoming information frame so as to minimize the correlation between training and the information frame. In [20], a data-dependent superimposed training scheme was proposed for time-invariant channel estimation, where the training sequence is distorted before transmission in order that the self-interference is eliminated at the receiver. This scheme was extended to MIMO systems in [21], and was modified to allow for unknown DC offset in [19]. However, in Chapter 6 we will show that the cost of using data-dependent superimposed training for interference cancelation is information loss to some extent. Channel estimation and data detection can also be enhanced in an iterative way, i.e., the detected data can be utilized to cancel the self-interference in the next iteration. Such applications are available in [13, 49–51, 97].

Now, one may wonder, in comparison with conventional time-multiplexed (TM) training, what are the advantages and disadvantages of superimposed training? Since the ultimate goal of communications is to improve the capacity of communication systems to the Shannon bound, how can superimposed training help?

A superimposed training-based scheme for space-time coded transmission over flat block fading (quasi-static) channels was considered in [8]. The analysis revealed the weakness of superimposed training in block-stationary (and thus time-invariant) environments, showing that superimposed training has higher CRLB than that of TM training due to

the presence of self-interference. On the other hand, if training must be included in every block and the channel estimation is accurate, the superimposed scheme gives higher mutual information (capacity) [8]. Performance bounds for TM and superimposed training-based semi-blind estimation of time-varying flat fading channels were considered in [16]. Under the same overall power allocation, it was shown that the superimposed training performs better for fast fading channels, which confirms the intuition that the constant presence of training has considerable benefit. For slow fading and high SNR, such an advantage disappears and there is a penalty for using superimposed training, since data transmission interferes with channel estimation [74]. This viewpoint was also confirmed by [73], which showed that when the coherence time is relatively short (compared with the time devoted to training in each block), superimposed training achieves higher capacity than that of TM training. It is because superimposed training allows for data transmission over the entire block, whereas TM training sessions occupy a large portion of time in this situation and hence not much time remains for information transmission. Capacity of superimposed training-based MIMO systems was considered in [5], where similar results have been derived, i.e., in the scenarios of high SNR, many receiver antennas, and short coherence time, it is beneficial to employ superimposed training; otherwise, TM training will be better.

An important conclusion appeared in [4], where the author answered the following question: How much will the capacity increase by allowing re-estimation of the channel when the detected symbols are available? It was shown that the capacity after re-estimation can be very close to the fundamental capacity of the non-coherent channel, especially when the channel coherence time is short. After iterations, significant improvement is achieved—of course, at the expense of increased complexity.

1.2 Contributions

In this dissertation, we investigate superimposed training-based approaches to the estimation and equalization of doubly-selective channels. In order to model channel variation by a parsimonious representation, we explore various BEM's, including CE-, OP-, and DPS-BEM's, to approximate the multipath channel with Doppler spread.

Our starting point is the first-order statistics-based estimator proposed by [81], using CE-BEM. In approximating band-limited time-varying channels, the modeling error of CE-BEM is noticeable. Therefore, we extend this estimator using DPS- and OP-BEM's to reduce the modeling error. A more general estimator that applies to arbitrary BEM's is also provided. We then further apply this estimator to a multiple-user scenario. Channel estimation across different users is decoupled by means of user-specific superimposed training sequences.

Performance analysis is then conducted for the first-order statistics-based estimator for doubly-selective channels, in which we demonstrate that the interference in estimation mainly comes from the unknown information sequence (self-interference). Based on the results of the performance analysis, we cast the issues of power allocation and bias-variance trade-off as ones of optimizing an SNR for equalizer design, following the method proposed for time-invariant channels in [85].

The major drawback of superimposed training is that the self-interference from information data may adversely affect channel estimation and data reception performance. To alleviate the effect of self-interference, we propose two methods: A deterministic maximum likelihood (DML) approach is employed at the receiver, to enhance the channel estimation

iteratively, exploiting the detected symbols from the previous iteration to reduce the self-interference. We can also achieve this by transmitter-end processing—the superimposed training sequence is modified based on the information sequence, or equivalently, the information sequence is distorted before transmission so that training and information data occupy distinct frequencies and hence can be separated at the receiver. However, distortion of the information sequence may cause “information” loss before transmission, which cannot be fully recovered by receiver-end processing. A partially-data-dependent (PDD) superimposed training scheme is proposed in order to strike a trade-off between interference cancelation and information integrity.

We also design a direct equalizer, without first estimating the channel, using superimposed training and CE-BEM. With the aid of periodic white training sequences, we show that the optimal linear equalizer for the training sequence is also a scaled version of the optimal equalizer for the information sequence. By employing user-specific training sequences, this direct equalizer can be extended to a multiple-user scenario, which can be used in a wireless *ad hoc* network.

Computer simulation examples illustrate our proposed approaches. Analytical results are also compared with simulation results to show their validity. Comparisons with conventional TM training-based approaches are also presented—when self-interference is sufficiently suppressed by our proposed schemes, the performance of superimposed training-based approaches are competitive with the ones using TM training, without incurring any data-rate loss.

1.3 Organization

The rest of this dissertation is organized as follows.

In Chapter 2, representations of time-varying wireless channels are briefly reviewed, including Jakes' model, CE-, OP-, and DPS-BEM's.

Chapter 3 introduces the first-order statistics-based channel estimator using superimposed training. We explore this estimator under different channel representations, and extend it from a single-user scenario to a multiple-user situation by exploiting user-specific superimposed training sequences.

We consider performance analysis for the first-order statistics-based estimators in Chapter 4. Training power allocation and bias-variance trade-off are also optimized from the viewpoint of equalization.

The DML approach is considered in Chapter 5. Exploiting detected symbols from the previous iteration, the self-interference is reduced at the current iteration and therefore, channel estimation and data reception performance are improved. A multiple-user scenario is also considered.

In Chapter 6, we investigate the data-dependent superimposed training scheme. The data-dependent processing at the transmitter results in information loss. We propose a PDD superimposed training scheme to mend this problem. Performance analysis and parameter design are also provided.

Superimposed training-based direct equalization is considered in Chapter 7, by the aid of periodic white training sequences. This algorithm is also extended to a multiple-user wireless *ad hoc* network.

The dissertation concludes in Chapter 8. Future directions are also suggested.

CHAPTER 2

REPRESENTATIONS OF WIRELESS CHANNELS

2.1 Introduction

Due to multipath propagation and Doppler spread, wireless channels are characterized by frequency- and time-selectivity [66]. A radio signal, experiencing distortions through transmission by fading, background noise, and interference of every sort, becomes stochastic to an observer at the receiver. Small-scale fading (or simply fading) is the term to describe the rapid fluctuations of the amplitudes, phases, or multipath delays of a signal over a short period of time or travel distance, so that large-scale path loss may be ignored [66]. The goal of channel estimation and equalization is mainly to combat small-scale fading.

Fading can be attributed to physical factors including multipath propagation, relative motion between the transmitter and the receiver or surrounding objects, and the transmission bandwidth of the signal, etc [35]. The presence of reflecting objects and scatterers makes the wireless channel constantly changing, which dissipates the signal energy and distorts the signal in amplitude, phase, and time. Multiple versions of the transmitted signal arrive at the receiver through different paths. The random amplitudes and phases of the different multipath components induce fading. The relative motion between the transmitter and the receiver as well as the motion of the objects within the wireless channel, induces Doppler spreads, which are typically time-varying and become a source of fading also.

For channel estimation or tracking purposes, accurate modeling of the temporal evolution of the channel plays an important role. A parsimonious and accurate channel representation is always preferred. Among various models for channel time variations, the autoregressive (AR) process, particularly the first-order AR model, is regarded as a tractable model to describe a time-varying channel, 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 [90]. This Markovian assumption has been verified for Rayleigh fading channels in [90], by considering the mutual information between channel symbols. The AR model has been used for time-varying channel estimation in [7, 11, 16, 37, 38].

The AR model, based on symbol-by-symbol update, is suitable for sequential time-domain processing. When we deal with block processing schemes, it is more convenient to use block-based channel models such as BEM's.

The BEM that is optimal in MSE is the discrete Karhuen-Loève BEM (DKL-BEM), which is a reduced-rank decomposition of a certain type of Doppler spectrum [72]. The CE-BEM can be viewed as a special DKL-BEM based on a white Doppler spectrum, and the DPS-BEM corresponds to the DKL-BEM with a rectangular Doppler spectrum [72].

In this chapter, we briefly review representations of time-varying channels. In Section 2.2, Jakes' model is introduced, which will be used as the model of the "real" channel in the simulation examples of this dissertation. In Sections 2.3–2.5, CE-, OP-, and DPS-BEM representations are discussed. The modeling error of these BEM's is compared one another in Section 2.6 via a simulation example. Section 2.7 summarizes this chapter.

2.2 Jakes' Model

If we assume that many statistically independent scattering waves with random amplitudes and phases reach the receiver with the phases uniformly lying in $[0, 2\pi)$, and there is no dominant non-fading signal component present (no line-of-sight), by the central limit theorem, the real and imaginary parts of the sum of the scattering waves are both Gaussian. The signal envelope A as a function of time t obeys a Rayleigh distribution, which has a probability density function (pdf) given by

$$f_A(a) := \begin{cases} \frac{a}{\sigma^2} \exp\left(-\frac{a^2}{2\sigma^2}\right) & a \geq 0, \\ 0 & a < 0 \end{cases} \quad (2.1)$$

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

$$f_\Theta(\theta) := \frac{1}{2\pi}, \quad \theta \in [0, 2\pi). \quad (2.2)$$

The autocorrelation function of the received signal for two-dimensional isotropic scattering and an omnidirectional receiving antenna is given by [12, 71]

$$R_A(\tau) = \sigma^2 \cos(\omega_c \tau) J_0(\omega_m \tau) \quad (2.3)$$

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. Any model that attempts to model the Rayleigh flat fading narrow-band wireless channel has to exhibit the statistical behaviors given by (2.1)–(2.3).

Clarke summarized the important characteristics of fading channels and provided a useful mathematical model [12]. According to this model, Jakes proposed a sum-of-sinusoids-based simulator [35] that has been widely used and studied over the past decades. The simulator supposes the received signal $S(t)$ to be a superposition of waves

$$S(t) = E_0 \sum_{n=1}^N C_n \cos(\omega_c t + \omega_m t \cos A_n + \Phi_n)$$

where E_0 is the amplitude of the transmitted cosine wave, C_n is a random variable representing the attenuation of the n -th path, A_n is a random variable representing the angle of arrival of the n -th ray with respect to the direction of motion of the receiver, Φ_n is a random variable representing the phase shift undergone by the n -th ray. Note that the stochastic signal $S(t)$ representing the flat fading signal can be characterized by N sets of triples (C_n, A_n, Φ_n) . The random variables C_n , A_n , and Φ_n are assumed statistically independent.

To reduce the complexity, Jakes' model selects

$$C_n = \frac{1}{\sqrt{N}}, \quad n = 1, 2, \dots, N, \quad (2.4a)$$

$$A_n = \frac{2\pi n}{N}, \quad n = 1, 2, \dots, N, \quad (2.4b)$$

$$\Phi_n = 0, \quad n = 1, 2, \dots, N. \quad (2.4c)$$

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

However, the simplification in (2.4) makes this simulation model deterministic and wide-sense nonstationary [63, 96]. In [96], a modified Jakes' simulator was proposed. It is wide-sense stationary and its autocorrelation and cross correlation functions match the

desired reference model exactly. Following [96], the normalized low-pass fading process of the statistical sum-of-sinusoids simulation model is defined by

$$X(t) = X_c(t) + jX_s(t), \quad (2.5a)$$

$$X_c(t) = \frac{2}{\sqrt{M}} \sum_{n=1}^M \cos(\psi_n) \cos(\omega_m t \cos \alpha_n + \phi), \quad (2.5b)$$

$$X_s(t) = \frac{2}{\sqrt{M}} \sum_{n=1}^M \sin(\psi_n) \cos(\omega_m t \cos \alpha_n + \phi) \quad (2.5c)$$

with

$$\alpha_n = \frac{2\pi n - \pi + \theta}{4M}, \quad n = 1, 2, \dots, M$$

where θ , ϕ , and ψ_n are statistically independent and uniformly distributed over $[-\pi, \pi)$ for all n . As $M \rightarrow \infty$, 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

$$f_{|X|}(x) = x \exp\left(-\frac{x^2}{2}\right), \quad x \geq 0,$$

$$f_{\Theta_X}(\theta) = \frac{1}{2\pi}, \quad \theta \in [-\pi, \pi).$$

A minor defect, however, occurs in model (2.5) when $\omega_m = 0$ or the Doppler spread is small: A Rayleigh distribution cannot be guaranteed [94]. This problem can be easily resolved by replacing a common phase ϕ by ϕ_n , which is also uniformly distributed over

$[-\pi, \pi)$ for all n . The simulation model is revised as [94]:

$$X(t) = X_c(t) + jX_s(t), \quad (2.6a)$$

$$X_c(t) = \frac{2}{\sqrt{M}} \sum_{n=1}^M \cos(\psi_n) \cos(\omega_m t \cos \alpha_n + \phi_n), \quad (2.6b)$$

$$X_s(t) = \frac{2}{\sqrt{M}} \sum_{n=1}^M \sin(\psi_n) \cos(\omega_m t \cos \alpha_n + \phi_n). \quad (2.6c)$$

2.3 Complex Exponential Basis Expansion Model (CE-BEM)

Recently, deterministic complex exponential basis expansion models (CE-BEM) have been widely investigated in wireless applications, especially when the multipath is caused by a few strong reflectors, and path delays exhibit variations due to the kinematics of the mobiles [24]. In these models, the time-varying taps are expressed as a superposition of time-varying basis functions in modeling Doppler effects, with time-invariant coefficients. By assigning temporal variations to basis functions, rapidly fading channels with coherence time as small as a few tens of symbols can be captured. If the delay spread and the Doppler spread of the channel (or at least the upper bounds of them) are known, one can infer the basis functions of the CE-BEM [40]. Treating the basis functions as known parameters, estimation of a time-varying process is reduced to estimate time-invariant coefficients.

Consider a time-varying channel with impulse response $h(t; \tau)$ (response at time t to a 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 $h(t; \tau)$ [64]:

$$x(t) = \int_0^{\infty} h(t; \tau) s(t - \tau) d\tau. \quad (2.7)$$

Let $H(f; \tau) = \int_{-\infty}^{\infty} h(t; \tau) e^{-j2\pi ft} dt$ be the Fourier transform of $h(t; \tau)$. If $|H(f; \tau)| \approx 0$ for $|\tau| > \tau_d$, then τ_d is defined as the delay-spread of the channel; if $|H(f; \tau)| \approx 0$ for $|f| > f_d$, then f_d is defined as the Doppler spread of the channel [40]. Sampling $s(t)$, $x(t)$ and $h(t; \tau)$ in (2.7) at the symbol rate, then for $t = nT_s \in [t_0, t_0 + TT_s)$, the sampled signal $x(n) := x(t)|_{t=nT_s}$ has the representation

$$x(n) = \sum_{l=0}^L h(n; l) s(n - l). \quad (2.8)$$

Over the block interval of $[t_0, t_0 + TT_s)$, the channel impulse response $\{h(n; l)\}_{n=0}^{T-1}$ can be represented by Q coefficients $\{h_q(l)\}_{q=1}^Q$ (which remain invariant throughout this block but are allowed to change at the next block) and the corresponding Q Fourier basis functions that are common for each block. Then over the interval $[t_0, t_0 + TT_s)$, the discrete-time baseband equivalent channel model for the block can be described as [39, 40]:

$$h(n; l) = \sum_{q=1}^Q h_q(l) e^{j\omega_q n} \quad (2.9a)$$

$$Q := 2\lceil f_d TT_s \rceil + 1, \quad (2.9b)$$

$$L := \lfloor \tau_d / T_s \rfloor, \quad (2.9c)$$

$$\omega_q := \frac{2\pi}{T} \left(q - \frac{Q+1}{2} \right), \quad q = 1, \dots, Q. \quad (2.9d)$$

2.4 Orthogonal Polynomial Basis Expansion Model (OP-BEM)

A time-varying channel over a fixed time interval can also be expressed as a superposition of polynomials with invariant coefficients. Following [6], by a Taylor series expansion, the continuous-time channel impulse response $h(t; \tau)$ within a window of time interval $[t_0, t_0 + TT_s)$ with respect to a midpoint $\bar{n}T_s + t_0$ is given by

$$h(t; \tau) = \sum_{i=0}^K \alpha_{\bar{n}}^{(i)}(\tau) \left(\frac{t - \bar{n}T_s - t_0}{T_s} \right)^i + R_K(t; \tau), \quad (2.10)$$

where the coefficients $\alpha_{\bar{n}}^{(i)}(\tau)$ are defined as

$$\alpha_{\bar{n}}^{(i)}(\tau) := \frac{T_s^i}{i!} \left[\frac{d^i h(t; \tau)}{dt^i} \right]_{t=\bar{n}T_s+t_0} \quad (2.11)$$

and $R_K(t; \tau)$ is the remainder of the Taylor series, given by

$$R_K(t; \tau) := \frac{(t - \bar{n}T_s - t_0)^{K+1}}{(K+1)!} \left[\frac{d^{K+1} h(t; \tau)}{dt^{K+1}} \right]_{t=s'} \quad (2.12)$$

for some $s' \in [t, \bar{n}T_s + t_0]$. The polynomials $[(t - \bar{n}T_s - t_0)/T_s]^i$ ($i = 0, 1, \dots, K$) serve as the basis functions in (2.10). In mobile wireless channels, the bandwidth of $h(t; \tau)$ in t (the Doppler spread) is strictly bounded above by v/λ where v is the velocity of the mobile and λ is the carrier wavelength. Therefore, $h(t; \tau)$ can be differentiated to any order with respect to t in the mean square sense, and so (2.11) and (2.12) are well defined [6].

Since $h(t; \tau)$ is band-limited in t , for a given window size T , $\lim_{K \rightarrow \infty} |R_K(t; \tau)|^2 = 0$. Thus with increasing K , the polynomial approximation becomes more and more accurate. As pointed out in [6], increasing the polynomial order K allows the window size to be

increased significantly, without the remainder term (2.12) becoming large. Sampling $h(t; \tau)$ every T_s seconds ($t = nT_s + t_0 \in [t_0, t_0 + TT_s]$) and ignoring the error remainder term, we have the discrete time-varying channel impulse response as follows

$$h(n; l) = \sum_{i=0}^K \alpha^{(i)}(l)(n - \bar{n})^i, \quad (2.13)$$

which is valid over a duration of TT_s seconds (T samples).

The polynomials $\{1, t, t^2, \dots, t^K\}$ are linearly independent over $[-1, 1]$, but not orthogonal. A QR-decomposition was suggested in [6] and generated a unitary matrix offering an orthonormal set of basis vectors. Or equivalently, via the Gram-Schmidt procedure over the interval $[-1, 1]$, we get the Legendre polynomials [53]. By appropriate scaling and translation of the (original) Legendre polynomials, we can obtain modified Legendre polynomials which are orthonormal over the interval $[t_0, t_0 + TT_s]$. Sampling these polynomials at the symbol interval T_s , we get the orthogonal polynomial basis expansion model (OP-BEM). Let $p^{(i)}(\tilde{t})$ denote the orthonormal Legendre polynomial of degree (order) i over the interval $[-1, 1]$. To extend $[-1, 1]$ to $[t_0, t_0 + TT_s]$, we set $t = (TT_s/2)\tilde{t} + t_0 + (TT_s/2)$, leading to $\tilde{t} = (2/(TT_s))[t - t_0] - 1$ and modified Legendre polynomials $p^{(i)'}(t) = p^{(i)}((2/(TT_s))[t - t_0] - 1)$ orthonormal over the interval $[t_0, t_0 + TT_s]$. Sample $p^{(i)'}(t)$'s at $t = nT_s + t_0$, ($n = 0, 1, \dots$) to obtain the discretized modified Legendre polynomials

$$\varphi_i(n) := p^{(i)}\left(\frac{2n}{T} - 1\right). \quad (2.14)$$

The discretized modified Legendre polynomials up to degree (order) five are as follows:

$$\begin{aligned}
\varphi_0(n) &= 1, \\
\varphi_1(n) &= c_1 \left[\frac{2}{T}n - 1 \right], \\
\varphi_2(n) &= c_2 \left[\left(\frac{2}{T}n - 1 \right)^2 - \frac{1}{3} \right], \\
\varphi_3(n) &= c_3 \left[\left(\frac{2}{T}n - 1 \right)^3 - \frac{3}{5} \left(\frac{2}{T}n - 1 \right) \right], \\
\varphi_4(n) &= c_4 \left[\left(\frac{2}{T}n - 1 \right)^4 - \frac{6}{7} \left(\frac{2}{T}n - 1 \right)^2 + \frac{3}{35} \right], \\
\varphi_5(n) &= c_5 \left[\left(\frac{2}{T}n - 1 \right)^5 - \frac{10}{9} \left(\frac{2}{T}n - 1 \right)^3 + \frac{5}{21} \left(\frac{2}{T}n - 1 \right) \right]
\end{aligned}$$

where $0 \leq n \leq T - 1$ and $c_i = 1/\sqrt{\sum_{n=0}^{T-1} \varphi_i(n)^2}$ for $i = 0, 1, \dots, 5$.

Using the basis functions of (2.14), the polynomial representation of equation (2.13) is written as

$$h(n; l) = \sum_{i=0}^K h_i(l) \varphi_i(n), \quad 0 \leq l \leq L. \quad (2.15)$$

Note that the number of basis functions is $Q = K + 1$.

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

Using CE-BEM to describe a band-limited channel, it has been observed that this truncated discrete Fourier transform (DFT)-based model has the following drawback: The rectangular window associated with the DFT introduces spectral leakage—the energy at each individual frequency leaks to the full frequency range [65]. An effect similar to Gibbs phenomenon results in significant amplitude and phase distortion at the beginning and the end of the block [94]. The modeling error of CE-BEM may cause a noticeable floor in bit

error rate (BER) curves, as shown in [2]. Taking advantage of OP-BEM will reduce the spectral leakage to some extent [51], whereas the polynomial functions are neither time-limited nor band-limited. Its square bias varies significantly over the range of the Doppler spread [94].

An ideal basis function should have at least two properties: It is band-limited to the normalized frequency range $[-f_d T_s, f_d T_s]$; and its energy is time-concentrated in a certain time interval $0 \leq n \leq T - 1$. Given the maximum normalized Doppler bandwidth $f_d T_s$ and the window size T , we seek a sequence to maximize

$$\lambda = \frac{\sum_{n=0}^{T-1} |u(n)|^2}{\sum_{m=-\infty}^{\infty} |u(m)|^2} \quad (2.16)$$

with the band-limited constraint

$$u(n) = \int_{-f_d T_s}^{f_d T_s} U(f) e^{j2\pi f n} df$$

where $U(f) = \sum_{m=-\infty}^{\infty} u(m) e^{-j2\pi f m}$.

The discrete prolate spheroidal (DPS) sequences $\{u_i(n)\}$ give us the solution of the above constrained maximization problem [69], which is defined as the real-valued solution of

$$\sum_{n=0}^{T-1} \frac{\sin[2\pi(n-m)f_d T_s]}{\pi(n-m)} u_i(n) = \lambda_i u_i(m)$$

for $i = 1, \dots, T$ and $-\infty < m < \infty$. For the discrete time index $0 \leq n \leq T - 1$, the i -th time-limited DPS vector $\mathbf{u}_i := \begin{bmatrix} u_i(0) & u_i(1) & \dots & u_i(T-1) \end{bmatrix}^T$ is the i -th eigenvector of a matrix \mathbf{C} :

$$\mathbf{C} \mathbf{u}_i = \lambda_i \mathbf{u}_i, \quad (2.17)$$

where the (n, m) -th entry of the $T \times T$ matrix \mathbf{C} is

$$[\mathbf{C}]_{n,m} = \frac{\sin [2\pi (n - m) f_d T_s]}{\pi (n - m)}$$

and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_T$ are the eigenvalues of \mathbf{C} .

The DPS sequences are orthonormal on the finite time interval $0 \leq n \leq T - 1$, and orthogonal on the doubly infinite interval, i.e.,

$$\sum_{n=0}^{T-1} u_i(n) u_k(n) = \lambda_i \sum_{m=-\infty}^{\infty} u_i(m) u_k(m) = \delta(i - k).$$

The band-limited (infinite) sequence $\{u_1(m)\}$ has the maximum energy concentration in $0 \leq m \leq T - 1$, $\{u_2(m)\}$ is the next band-limited sequence that has the most energy concentration among the DPS sequences orthogonal to $\{u_1(m)\}$, and so on. By (2.16), the eigenvalues λ_i 's are a measure for energy concentration, which are clustered near 1 for $i \leq \lceil 2f_d T_s T \rceil + 1$ and drop rapidly toward zero when $i > \lceil 2f_d T_s T \rceil + 1$ [94]. Therefore, the number of dimensions of time-limited snapshots of a band-limited channel is approximately given by [69]

$$Q = \lceil 2f_d T_s T \rceil + 1. \quad (2.18)$$

All the properties described so far make it possible to greatly reduce spectral leakage induced by the CE-BEM, by using several DPS sequences to form the basis set to approximate a band-limited time-varying channel.

In a DPS-BEM representation, we assume that

$$h(n; l) = \sum_{q=1}^Q h_q(l) u_q(n). \quad (2.19)$$

The square bias of the DPS-BEM to approximate a time-varying channel is several magnitudes lower than that of the CE-BEM over the range of Doppler spreads [94].

2.6 Modeling Error of BEM's

We illustrate the modeling error of CE-, OP-, and DPS-BEM's by a simulation example.

2.6.1 LS Approximation by Basis Expansion Models

In a BEM, we assume that the time-varying channel satisfies

$$h(n; l) = \sum_{q=1}^Q h_q(l) \psi_q(n), \quad (2.20)$$

where $\psi_q(n)$ is the q -th basis function (corresponding to $e^{j\omega_q n}$ in (2.9), $\varphi_{q-1}(n)$ in (2.15), and $u_q(n)$ in (2.19)), and Q is the number of basis functions ($Q = K + 1$ for (2.15)). However, the true channel may not exactly follow this expression, for modeling error always occurs. We revise (2.20) as

$$h(n; l) = \sum_{q=1}^Q h_q(l) \psi_q(n) + e(n; l)$$

where $e(n; l)$ denotes the modeling error. By the orthogonality principle, $e(n; l)$ is orthogonal to the basis set $\{\psi_q(n)\}_{q=1}^Q$ when the square error $\sum_{n=0}^{T-1} |e(n; l)|^2$ is minimized.

Then

$$\sum_{n=0}^{T-1} h(n; l) \psi_{q'}^*(n) = \sum_{q=1}^Q h_q(l) \sum_{n=0}^{T-1} \psi_q(n) \psi_{q'}^*(n) = h_{q'}(l) \sum_{n=0}^{T-1} |\psi_{q'}(n)|^2.$$

Therefore

$$h_{q'}(l) = \frac{\sum_{n=0}^{T-1} h(n; l) \psi_{q'}^*(n)}{\sum_{n=0}^{T-1} |\psi_{q'}(n)|^2}.$$

The LS approximation by a BEM is given by

$$\hat{h}(n; l) = \sum_{q=1}^Q \frac{\sum_{n=0}^{T-1} h(n; l) |\psi_q(n)|^2}{\sum_{n=0}^{T-1} |\psi_q(n)|^2}. \quad (2.21)$$

2.6.2 Simulation Model

In the simulations of this dissertation, we use the modified Jakes' model (2.6) to represent the “real” channel. We emphasize that BEM representations are only used for processing at the receiver. A discrete-time baseband Rayleigh fading channel (which can be SISO, SIMO, or MIMO in the subsequent chapters) of order L is generated (see (2.8)). For different taps (i.e., different l 's), $h(n; l)$'s are mutually independent, and for a given tap, we follow (2.6) to generate $h(n; l)$ by sampling $X(t)$ with symbol period T_s :

$$h(n; l) = X(t) |_{t=nT_s}.$$

In simulations, we take $M = 25$ in (2.6).

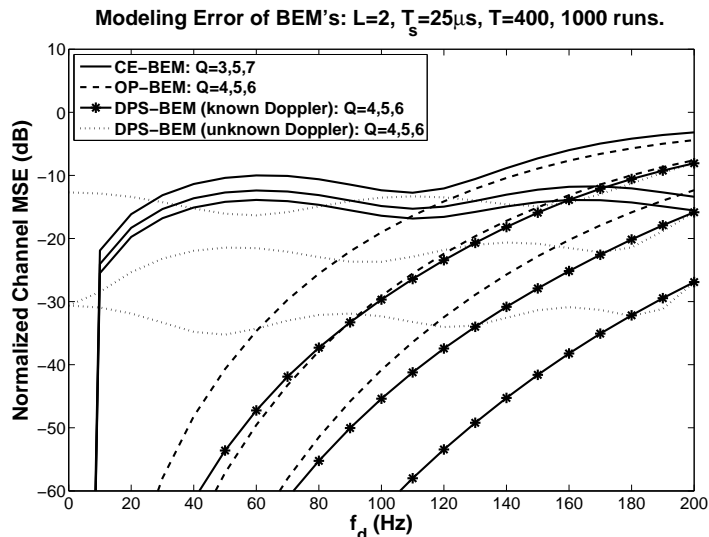


Figure 2.1: Modeling error of CE-, OP-, and DPS-BEM's in approximating a three-tap ($L = 2$) Rayleigh fading channel following Jakes' model.

2.6.3 Simulation Example: Modeling Error of CE-, OP-, and DPS-BEM's in Approximating a Doubly-Selective Channel

We consider a system with carrier frequency of 2 GHz, data rate of 40 kBd (kilo-Bauds), therefore, $T_s = 25 \mu s$. The maximum Doppler spread (in Hz) $f_d = \omega_m/2\pi$ ranges from 0 Hz to 200 Hz (or the normalized Doppler shift ranges from 0 to 0.005), corresponding to a maximum mobile velocity in the range 0 to 108km/h.

A SISO three-tap ($L = 2$) Rayleigh fading channel is generated using (2.6). We try to approximate this Jakes' model using different BEM's: CE-BEM, OP-BEM, DPS-BEM with known Doppler spread, and DPS-BEM with Doppler spread unknown. If the Doppler spread is known, we can follow (2.17) to obtain the DPS sequences using the exact Doppler

spread; whereas we can only use predetermined DPS sequences if the Doppler spread is unknown. In this example, we assume $f_d = 200$ Hz in (2.17) to get the DPS sequences.

We pick up a data record length of 400 symbols, average over 1000 realizations of randomly generated channels, and plot the normalized channel mean square error (NCMSE) defined as

$$\text{NCMSE} := \frac{\sum_{i=1}^{1000} \sum_{n=0}^{T-1} \sum_{l=0}^2 \left| h^{(i)}(n;l) - \hat{h}^{(i)}(n;l) \right|^2}{\sum_{i=1}^{1000} \sum_{n=0}^{T-1} \sum_{l=0}^2 \left| h^{(i)}(n;l) \right|^2},$$

where $h^{(i)}(n;l)$ denotes the realization of the channel in the i -th run, and $\hat{h}^{(i)}(n;l)$ denotes the BEM-based approximation by (2.21). For CE-BEM, we plot the NCMSE curves with $Q = 3, 5,$ and 7 (note that in (2.9), only odd Q is allowed). For OP- and DPS-BEM representations, we take $Q = 4, 5,$ and 6 .

Using more basis functions apparently reduces modeling error, which is confirmed by Figure 2.1. The OP- and DPS-BEM with known Doppler spreads outperform CE-BEM significantly for small f_d 's. As f_d grows, OP-BEM deteriorates quickly, whereas DPS-BEM with known Doppler spread is still much better than the other two. If the Doppler spread is unknown, the performance of DPS-BEM is a little worse, but it still outperforms CE-BEM. In Figure 2.1, the NCMSE of DPS-BEM is at least two orders of magnitude lower than that of the CE-BEM, whether with known Doppler spread or not—DPS-BEM is undoubtedly the best among the three to describe a band-limited channel.

For a fixed Q , the NCMSE's of the CE-BEM and DPS-BEM with unknown Doppler spread fluctuate mildly over the range of Doppler spreads. In these two scenarios, the BEM's are both band-limited. If the "true" channel (that is also band-limited within the Doppler shifts) lies within the frequency band of the BEM, the resulting error should not fluctuate significantly for different Doppler spreads.

2.7 Conclusions

In this chapter, we reviewed characteristics and representations of wireless channels. We first discussed Jakes' model, which will be used as the “real” channel in simulation examples in the following chapters. For channel estimation and data processing at the receiver, a BEM, a more parsimonious representation that is independent of the “real” channel, will be used to describe the temporal variation of the channel. We discussed CE-, OP-, and DPS-BEM's. Although the CE-BEM is more convenient in the theoretical analysis, its modeling error is noticeable due to the spectral leakage. We may employ OP- and DPS-BEM's to reduce this phenomenon. We also compared the modeling error of the three BEM's—the DPS-BEM has the minimum modeling error among the three, since spectral leakage is greatly reduced due to the energy concentration of the DPS sequences.

CHAPTER 3

FIRST-ORDER STATISTICS-BASED ESTIMATION OF DOUBLY-SELECTIVE CHANNELS

3.1 Introduction

An estimator for time-invariant frequency-selective channels, using periodic superimposed training and the first-order statistics of the observations, was proposed in [82]. This estimator was soon extended to doubly-selective channels by exploiting CE-BEM in [81]. We start with this CE-BEM-based channel estimator in [81], which offers us the basic framework of our superimposed training-based channel estimation schemes.

In this chapter, we first review the first-order statistics-based doubly-selective channel estimator of [81] in Section 3.2. By exploiting the band-limitedness of DPS sequences, in Section 3.3 this estimator is extended to DPS-BEM. Since polynomial models are not band-limited, we propose a more general estimator for OP-BEM in Section 3.4, which can apply to arbitrary BEM representations. In Section 3.5, we extend our CE- and DPS-BEM-based channel estimators to multiuser systems. Exploiting the band-limitedness of CE- and DPS-BEM's, the channel estimation across different users is decoupled by assigning user-specific training sequences to different frequencies for distinct users. Our approaches are illustrated by simulation examples in Section 3.6. Section 3.7 concludes this chapter.

3.2 First-Order Statistics-Based Channel Estimation Using CE-BEM [81]

Consider a single-input multiple-output (SIMO) discrete-time baseband communication system. Let $\{s(n)\}$ denote the input symbol sequence that is transmitted over an FIR linear channel with N outputs and discrete-time impulse response $\{\mathbf{h}(n;l)\}$ (N -column vector

channel response at time n to a unit input at time $n - l$). The vector channel may be the result of multiple receiver antennas or over-sampling at the receiver. The channel output is given by

$$\mathbf{x}(n) = \sum_{l=0}^L \mathbf{h}(n; l) s(n - l). \quad (3.1)$$

The noisy measurement is given by

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n) \quad (3.2)$$

where $\mathbf{v}(n)$ is an N -column white complex-Gaussian noise vector. To allow for mean-value ambiguity, we take $E\{\mathbf{v}(n)\} = \mathbf{m}$, with \mathbf{m} unknown. In practice, linear systems arise because of linearization about some operating (set) point—“bias” in BJT/FET amplifiers. These set points are typically unknown (at least not known precisely) a priori, and one does not normally worry about them since unknown means are estimated and removed before processing (blocked by capacitor-coupling etc.) and they are not needed in any processing. However, we will initially use the first-order statistics, i.e., $E\{\mathbf{y}(n)\}$, of the noisy data. Then we must include a term such as nonzero \mathbf{m} .

Channel taps are superpositions of complex exponentials weighted by time-invariant coefficients in CE-BEM. The time-varying SIMO channel response is given by (2.9)

$$\mathbf{h}(n; l) = \sum_{q=1}^Q \mathbf{h}_q(l) e^{j\omega_q n}, \quad (3.3)$$

for

$$\begin{aligned}
Q &:= 2\lceil f_d T T_s \rceil + 1, \\
L &:= \lfloor \tau_d / T_s \rfloor, \\
\omega_q &:= \frac{2\pi}{T} \left(q - \frac{1}{2} - \frac{Q}{2} \right), \quad q = 1, 2, \dots, Q,
\end{aligned}$$

where τ_d is the (multipath) delay-spread and f_d is the Doppler spread. The above representation is valid over a duration of $T T_s$ sec. with symbol interval T_s sec. If τ_d and f_d (or their upper bounds) are known (typically true), then $\mathbf{h}(n; l)$ is unknown up to only time-invariant quantities $\mathbf{h}_q(l)$'s.

In superimposed training, the transmitted signal $\{s(n)\}$ is the superposition of the information sequence $\{b(n)\}$ and a training sequence $\{c(n)\}$, i.e.,

$$s(n) = b(n) + c(n). \quad (3.4)$$

Assume the following:

- (H3.2.1) The time-varying channel $\{\mathbf{h}(n; l)\}$ satisfies (3.3) where the frequencies ω_q 's ($q = 1, 2, \dots, Q$) are distinct and known with $\omega_q \in [0, 2\pi)$. Also $N \geq 1$.
- (H3.2.2) The information sequence $\{b(n)\}$ is zero-mean, white with $E\{|b(n)|^2\} = \sigma_b^2$.
- (H3.2.3) The measurement noise $\{\mathbf{v}(n)\}$ may be nonzero-mean ($E\{\mathbf{v}(n)\} = \mathbf{m}$), white, uncorrelated with $\{b(n)\}$, with $E\{[\mathbf{v}(n + \tau) - \mathbf{m}][\mathbf{v}(n) - \mathbf{m}]^H\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$. The mean vector \mathbf{m} may be unknown.

(H3.2.4) The superimposed training sequence $c(n) = c(n+P)$ for all n is a non-random periodic sequence with period P .

By (H3.2.4), we have

$$c_m := \frac{1}{P} \sum_{n=0}^{P-1} c(n) e^{-j\alpha_m n}, \quad (3.5a)$$

$$c(n) = \sum_{m=0}^{P-1} c_m e^{j\alpha_m n}, \text{ for all } n, \quad (3.5b)$$

$$\alpha_m = 2\pi m/P, \quad m = 0, 1, \dots, P-1. \quad (3.5c)$$

The coefficients c_m 's are known at the receiver since the training sequence $\{c(n)\}$ is known.

We then have

$$\mathbf{y}(n) = \sum_{l=0}^L \sum_{q=1}^Q \mathbf{h}_q(l) e^{j\omega_q n} \left[b(n-l) + \sum_{m=0}^{P-1} c_m e^{j\alpha_m(n-l)} \right] + \mathbf{v}(n).$$

By (H3.2.2) and (H3.2.3), the expectation of the observation at time n is

$$E\{\mathbf{y}(n)\} = \sum_{q=1}^Q \sum_{m=0}^{P-1} \left[\sum_{l=0}^L c_m \mathbf{h}_q(l) e^{-j\alpha_m l} \right] e^{j(\omega_q + \alpha_m)n} + \mathbf{m}.$$

Defining

$$\mathbf{d}_{mq} := \sum_{l=0}^L c_m \mathbf{h}_q(l) e^{-j\alpha_m l}, \quad (3.6)$$

we then have

$$E\{\mathbf{y}(n)\} = \sum_{q=1}^Q \sum_{m=0}^{P-1} \mathbf{d}_{mq} e^{j(\omega_q + \alpha_m)n} + \mathbf{m}. \quad (3.7)$$

Suppose that we pick P to be such that $(\omega_q + \alpha_m)$'s are all distinct for any choice of m and q ; for instance, take $TP^{-1} \geq Q$. In fact, we will take $TP^{-1} = K \geq Q$ where K is an integer, so that ω_q 's and α_m 's are on the same frequency grid of resolution T^{-1} . Then the sequence $E\{\mathbf{y}(n)\}$ is periodic with cycle frequencies $(\omega_q + \alpha_m)$, for $1 \leq q \leq Q$ and $0 \leq m \leq P-1$.

By (3.7),

$$\mathbf{y}(n) = E\{\mathbf{y}(n)\} + \mathbf{e}(n) = \sum_{q=1}^Q \sum_{m=0}^{P-1} \mathbf{d}_{mq} e^{j(\omega_q + \alpha_m)n} + \mathbf{m} + \mathbf{e}(n)$$

where $\{\mathbf{e}(n)\}$ is a zero-mean random "error" sequence.

Define the cost function

$$J = \sum_{n=0}^{T-1} \|\mathbf{e}(n)\|^2 = \sum_{n=0}^{T-1} \left\| \mathbf{y}(n) - \sum_{q=1}^Q \sum_{m=0}^{P-1} \mathbf{d}_{mq} e^{j(\omega_q + \alpha_m)n} + \mathbf{m} \right\|^2. \quad (3.8)$$

Choose \mathbf{d}_{mq} 's to minimize J . We must have

$$\frac{\partial J}{\partial \mathbf{d}_{mq}^*} \Big|_{\mathbf{d}_{mq} = \hat{\mathbf{d}}_{mq}} = \sum_{n=0}^{T-1} e^{-j(\omega_q + \alpha_m)n} \left[\mathbf{y}(n) - \sum_{q'=1}^Q \sum_{m'=0}^{P-1} \mathbf{d}_{mq'} e^{j(\omega_{q'} + \alpha_{m'})n} + \mathbf{m} \right] = 0$$

for each \mathbf{d}_{mq} . A consistent mean-square (m.s.) estimate of \mathbf{d}_{mq} , for $\omega_q + \alpha_m \neq 0$, follows as

$$\hat{\mathbf{d}}_{mq} = \frac{1}{T} \sum_{n=0}^{T-1} \mathbf{y}(n) e^{-j(\omega_q + \alpha_m)n}. \quad (3.9)$$

It follows from (3.7) and (3.9) that

$$E\{\hat{\mathbf{d}}_{mq}\} = \mathbf{d}_{mq} \quad (3.10)$$

for $\omega_q + \alpha_m \neq 0$. As $T \rightarrow \infty$, $\hat{\mathbf{d}}_{mq} \rightarrow \mathbf{d}_{mq}$ m.s. if $\omega_q + \alpha_m \neq 0$, and $\hat{\mathbf{d}}_{mq} \rightarrow \mathbf{d}_{mq} + \mathbf{m}$ m.s. if $\omega_q + \alpha_m = 0$.

It is established in [82] that given \mathbf{d}_{mq} for $1 \leq q \leq Q$ and $1 \leq m \leq P - 1$, we can (uniquely) estimate $\mathbf{h}_q(l)$'s if $P \geq L + 2$, $\alpha_m \neq 0$, and $c_m \neq 0$ for all $m \neq 0$. Since \mathbf{m} is unknown and $\omega_q + \alpha_m = 0$ only when $m = 0$, we will omit the term $m = 0$ for further discussion. For $1 \leq m \leq P - 1$, define the NQ -column vector

$$\mathbf{D}_m := [\mathbf{d}_{m1}^T, \mathbf{d}_{m2}^T, \dots, \mathbf{d}_{mQ}^T]^T, \quad (3.11)$$

and for $0 \leq l \leq L$, define the NQ -column vector

$$\mathbf{H}_l := [\mathbf{h}_1^T(l), \mathbf{h}_2^T(l), \dots, \mathbf{h}_Q^T(l)]^T. \quad (3.12)$$

Then by (3.6), we have

$$\mathbf{D}_m = \sum_{l=0}^L c_m e^{-j\alpha_m l} \mathbf{H}_l$$

for $1 \leq m \leq P - 1$. Define the $NQ(P - 1) \times NQ(L + 1)$ matrix

$$\mathcal{C} := (\text{diag}\{c_1, \dots, c_{P-1}\} \mathbf{V}) \otimes \mathbf{I}_{NQ} \quad (3.13)$$

where

$$\mathbf{V} := \begin{bmatrix} 1 & e^{-j\alpha_1} & \dots & e^{-j\alpha_1 L} \\ 1 & e^{-j\alpha_2} & \dots & e^{-j\alpha_2 L} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\alpha_{P-1}} & \dots & e^{-j\alpha_{P-1} L} \end{bmatrix}, \quad (3.14)$$

the $NQ(L+1)$ -column vector

$$\mathcal{H} := \left[\mathbf{H}_0^T, \mathbf{H}_1^T, \dots, \mathbf{H}_L^T \right]^T, \quad (3.15)$$

and the $NQ(P-1)$ -column vector

$$\mathcal{D} := \left[\mathbf{D}_1^T, \mathbf{D}_2^T, \dots, \mathbf{D}_{P-1}^T \right]^T. \quad (3.16)$$

Then (3.6) leads to

$$\mathcal{C}\mathcal{H} = \mathcal{D}. \quad (3.17)$$

Since α_m 's are distinct and $c_m \neq 0$ for all m , $\text{rank}(\mathcal{C}) = NQ(L+1)$ if $P \geq L+2$; hence, we can determine $\mathbf{h}_q(l)$'s uniquely. Define $\hat{\mathbf{D}}_m$ as in (3.11) with \mathbf{d}_{mq} replaced with $\hat{\mathbf{d}}_{mq}$. Define $\hat{\mathcal{D}}$ as in (3.16) with \mathbf{D}_m replaced with $\hat{\mathbf{D}}_m$. Then the estimate of \mathcal{H} is given by

$$\hat{\mathcal{H}} = (\mathcal{C}^H \mathcal{C})^{-1} \mathcal{C}^H \hat{\mathcal{D}}. \quad (3.18)$$

By (3.10) and (3.17), it follows that

$$E \left\{ \hat{\mathcal{H}} \right\} = \mathcal{H}. \quad (3.19)$$

Denote the corresponding estimate of $\mathbf{h}_q(l)$ by $\hat{\mathbf{h}}_q(l)$ for $q = 1, 2, \dots, Q$ and $l = 0, 1, \dots, L$.

Following (3.3), the time-varying channel coefficients are given by

$$\hat{\mathbf{h}}(n; l) = \sum_{q=1}^Q \hat{\mathbf{h}}_q(l) e^{j\omega_q n}, \quad l = 0, 1, \dots, L, \quad 0 \leq n \leq T-1. \quad (3.20)$$

We summarize this estimator in the following lemmas:

Lemma 3.2.1: Under assumptions **(H3.2.1)**–**(H3.2.4)**, the channel estimator (3.18) is unbiased by (3.19) if the periodic training sequence is such that $c_m \neq 0$ for all $m \neq 0$, $P \geq L + 2$, and P and T are such that $T = KP$ for integer $K \geq Q$.

Lemma 3.2.2: Under assumptions **(H3.2.1)**–**(H3.2.4)**, the channel estimator (3.18) is consistent in probability if the periodic training sequence is such that $c_m \neq 0$ for all $m \neq 0$, $P \geq L + 2$ and P is such that $\omega_q + \alpha_m \neq 0$ for $q = 1, 2, \dots, Q$ and $m \neq 0$, and Q is fixed as T becomes large.

Remark 3.2.1: If the channel length L is unknown, an upper bound L_u will suffice. Then we are estimating $\hat{\mathbf{h}}(n; l)$ for $l = 0, 1, \dots, L_u$ and $\lim_{T \rightarrow \infty} \hat{\mathbf{h}}(n; l) = 0$ for $l > L$.

Remark 3.2.2: We do not need $c_m \neq 0$ for every m . We need at least $L + 2$ nonzero c_m 's.

Remark 3.2.3: If the noise $\mathbf{v}(n)$ is zero-mean, i.e., $\mathbf{m} = \mathbf{0}$, we do not have to discard \mathbf{d}_{0q} . Thus, by setting

$$\tilde{\mathbf{V}} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-j\alpha_1} & \cdots & e^{-j\alpha_1 L} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\alpha_{P-1}} & \cdots & e^{-j\alpha_{P-1} L} \end{bmatrix}, \quad (3.21)$$

$$\tilde{\mathbf{C}} := (\text{diag}\{c_0, c_1, \dots, c_{P-1}\} \mathbf{V}) \otimes \mathbf{I}_{NQ}, \quad (3.22)$$

$$\tilde{\mathbf{D}} := \begin{bmatrix} \mathbf{D}_0^T & \mathbf{D}_1^T & \cdots & \mathbf{D}_{P-1}^T \end{bmatrix}^T, \quad (3.23)$$

then we have $\tilde{\mathbf{C}}\mathcal{H} = \tilde{\mathbf{D}}$ and

$$\hat{\mathcal{H}} = (\tilde{\mathbf{C}}^H \tilde{\mathbf{C}})^{-1} \tilde{\mathbf{C}}^H \hat{\tilde{\mathbf{D}}}. \quad (3.24)$$

To identify the BEM coefficients, we now need $P \geq L + 1$. All our results hold true if we use appropriate substitutions.

3.3 First-Order Statistics-Based Channel Estimation Using DPS-BEM

The first-order statistics-based channel estimator described in Section 3.2, using CE-BEM representation, can be easily extended to DPS-BEM.

As we discussed in Section 2.5, band-limitedness and energy-concentration of DPS sequences greatly reduce the spectral leakage intrinsic to CE-BEM. Therefore, better performance can be expected if we use DPS-BEM instead in the first-order statistics-based estimator.

In the DPS-BEM representation, we assume that

$$\mathbf{h}(n; l) = \sum_{q=1}^Q \mathbf{h}_q(l) u_q(n), \quad (3.25)$$

where $\{u_q(n)\}$ is the q -th DPS sequence.

Similar to **(H3.2.1)**–**(H3.2.4)**, we assume:

(H3.3.1) The time-varying channel $\{\mathbf{h}(n; l)\}$ satisfies (3.25) with the DPS sequences $\{u_q(n)\}$ known at the receiver. Also $N \geq 1$.

(H3.3.2) The information sequence $\{b(n)\}$ is zero-mean, white with $E\{|b(n)|^2\} = \sigma_b^2$.

(H3.3.3) The measurement noise $\{\mathbf{v}(n)\}$ is zero-mean, white, uncorrelated with $\{b(n)\}$, with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$.

(H3.3.4) The superimposed training sequence $c(n) = c(n+P)$ for all n is a non-random periodic sequence with period P .

Note that for the time being we assume that the measurement noise is zero-mean, i.e., $\mathbf{m} = \mathbf{0}$.

By the SIMO channel model (3.1), (3.2), and the DPS-BEM representation (3.25), we have

$$\begin{aligned} \mathbf{y}(n) &= \sum_{l=0}^L \sum_{q=1}^Q \mathbf{h}_q(l) u_q(n) [b(n-l) + c(n-l)] + \mathbf{v}(n) \\ &= \sum_{l=0}^L \sum_{q=1}^Q \mathbf{h}_q(l) u_q(n) \left[b(n-l) + \sum_{m=0}^{P-1} c_m e^{j\alpha_m(n-l)} \right] + \mathbf{v}(n). \end{aligned}$$

By (H3.3.2) and (H3.3.3), the expectation of the observation at time n is

$$E\{\mathbf{y}(n)\} = \sum_{q=1}^Q \sum_{m=0}^{P-1} \left[\sum_{l=0}^L c_m \mathbf{h}_q(l) e^{-j\alpha_m l} \right] u_q(n) e^{j\alpha_m n}. \quad (3.26)$$

Using (3.6), we have

$$E\{\mathbf{y}(n)\} = \sum_{q=1}^Q \sum_{m=0}^{P-1} \mathbf{d}_{mq} u_q(n) e^{j\alpha_m n}. \quad (3.27)$$

It follows that

$$\mathbf{y}(n) = \sum_{q=1}^Q \sum_{m=0}^{P-1} \mathbf{d}_{mq} u_q(n) e^{j\alpha_m n} + \mathbf{e}(n)$$

where $\{\mathbf{e}(n)\}$ is a zero-mean random sequence.

Define the cost function as in (3.8)

$$J = \sum_{n=0}^{T-1} \|\mathbf{e}(n)\|^2 = \sum_{n=0}^{T-1} \left\| \mathbf{y}(n) - \sum_{q=1}^Q \sum_{m=0}^{P-1} \mathbf{d}_{mq} u_q(n) e^{j\alpha_m n} \right\|^2.$$

Choose \mathbf{d}_{mq} 's to minimize J . We must have

$$\left. \frac{\partial J}{\partial \mathbf{d}_{mq}^*} \right|_{\mathbf{d}_{mq} = \hat{\mathbf{d}}_{mq}} = \sum_{n=0}^{T-1} \left[\mathbf{y}(n) - \sum_{q'=1}^Q \sum_{m'=0}^{P-1} \mathbf{d}_{m'q'} u_{q'}(n) e^{j\alpha_{m'}n} \right] u_q(n) e^{-j\alpha_m n} = 0,$$

which leads to

$$\sum_{q'=1}^Q \sum_{m'=0}^{P-1} \hat{\mathbf{d}}_{m'q'} \left[\sum_{n=0}^{T-1} u_{q'}(n) u_q(n) e^{j(\alpha_{m'} - \alpha_m)n} \right] = \sum_{n=0}^{T-1} \mathbf{y}(n) u_q(n) e^{-j\alpha_m n}. \quad (3.28)$$

We then define

$$\mathbf{g}_{mq} := \sum_{n=0}^{T-1} \mathbf{y}(n) u_q(n) e^{-j\alpha_m n}$$

and substitute it into (3.28)

$$\mathbf{g}_{mq} = \sum_{q'=1}^Q \sum_{m'=0}^{P-1} \hat{\mathbf{d}}_{m'q'} \left[\sum_{n=0}^{T-1} u_{q'}(n) u_q(n) e^{j(\alpha_{m'} - \alpha_m)n} \right]. \quad (3.29)$$

By the definitions (3.11), (3.12), (3.15), and (3.21)–(3.23), we also have

$$\tilde{\mathcal{C}}\mathcal{H} = \tilde{\mathcal{D}}. \quad (3.30)$$

If $P \geq L + 1$, then $\text{rank}(\tilde{\mathcal{C}}) = NQ(L + 1)$ [82]. We can determine the $\mathbf{h}_q(l)$'s uniquely.

Define $\hat{\tilde{\mathcal{D}}}$ and \mathcal{G} in the similar way as (3.23) with \mathbf{d}_{mq} replaced with $\hat{\mathbf{d}}_{mq}$ or \mathbf{g}_{mq} , then (3.29) turns out to be

$$\mathcal{G} = (\Psi \otimes \mathbf{I}_N) \hat{\tilde{\mathcal{D}}}$$

where the entries of the $PQ \times PQ$ matrix Ψ are

$$[\Psi]_{mQ+q, m'Q+q'} = \sum_{n=0}^{T-1} u_{q'}(n) u_q(n) e^{j(\alpha_{m'} - \alpha_m)n}.$$

The estimate of $\tilde{\mathcal{D}}$ is given by

$$\hat{\tilde{\mathcal{D}}} = (\Psi^{-1} \otimes \mathbf{I}_N) \mathcal{G}. \quad (3.31)$$

By (3.30) and (3.31) we have the estimate of channel coefficient

$$\hat{\mathcal{H}} = \mathcal{C}^\dagger \hat{\tilde{\mathcal{D}}} = (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \tilde{\mathcal{C}}^H (\Psi^{-1} \otimes \mathbf{I}_N) \mathcal{G}. \quad (3.32)$$

The channel estimate is then given by

$$\hat{\mathbf{h}}(n; l) = \sum_{q=1}^Q \hat{\mathbf{h}}_q(l) u_q(n). \quad (3.33)$$

Remark 3.3.1: Since DPS sequences are approximately band-limited to the normalized frequency range $[-f_d T_s, f_d T_s]$, it follows that

$$\sum_{n=0}^{T-1} u_{q'}(n) u_q(n) e^{j(\alpha_{m'} - \alpha_m)n} \approx \delta(m' - m) \delta(q' - q) \quad (3.34)$$

when $f_d T_s \ll 1/P$ and T is a multiple of P or T is “large”. This is usually true, and a short period P helps to achieve it. Under the assumption (3.34),

$$\Psi \approx \mathbf{I}_{PQ}.$$

By (3.28) and (3.34),

$$\hat{\mathbf{d}}_{mq} = \sum_{n=0}^{T-1} \mathbf{y}(n) u_q(n) e^{-j\alpha_m n}. \quad (3.35)$$

The estimate (3.32) is then given by

$$\hat{\mathcal{H}} = \left(\tilde{\mathcal{C}}^H \tilde{\mathcal{C}} \right)^{-1} \tilde{\mathcal{C}}^H \hat{\mathcal{D}}. \quad (3.36)$$

Remark 3.3.2: If the mean of the noise $\mathbf{v}(n)$ is unknown, suppose $E\{\mathbf{v}(n)\} = \mathbf{m}$. Under the approximation (3.34), we should omit the first row (corresponding to α_0) of $\tilde{\mathbf{V}}$ in (3.21) (denote the resulting $(P-1) \times (L+1)$ matrix by \mathbf{V} , as in (3.14)), and also omit the block \mathbf{D}_0 from $\tilde{\mathcal{D}}$ in (3.23) (denote the resulting matrix by \mathcal{D} , as in (3.16)). We have $\mathcal{C}\mathcal{H} = \mathcal{D}$ and

$$\hat{\mathcal{H}} = (\mathcal{C}^H \mathcal{C})^{-1} \mathcal{C}^H \hat{\mathcal{D}}, \quad (3.37)$$

where \mathcal{C} is defined in (3.13) and $\hat{\mathcal{D}}$ is acquired by (3.35). To identify the BEM coefficients, we now need $P \geq L + 2$. All our results hold true if appropriate substitutions are used.

3.4 First-Order Statistics-Based Channel Estimation Using OP-BEM

The first-order statistics-based channel estimators, using CE-and DPS-BEM's, exploit band-limitedness of the basis functions. OP-BEM, however, does not have this property.

We assume the SIMO channel with N outputs satisfies the OP-BEM representation (2.15), i.e.,

$$\mathbf{h}(n; l) = \sum_{q=0}^K \mathbf{h}_q(l) \varphi_q(n), \quad 0 \leq l \leq L \quad (3.38)$$

where $\varphi_q(n)$ is the discretized modified Legendre polynomial of degree q .

We make the following assumptions:

- (H3.4.1) The time-varying channel $\{\mathbf{h}(n;l)\}$ satisfies (3.38) where $\{\varphi_q(n)\}_{q=0}^K$ are known at the receiver. $N \geq 1$.
- (H3.4.2) The information sequence $\{b(n)\}$ is zero-mean, white with $E\{|b(n)|^2\} = \sigma_b^2$.
- (H3.4.3) The measurement noise $\{\mathbf{v}(n)\}$ may be nonzero-mean ($E\{\mathbf{v}(n)\} = \mathbf{m}$), white, uncorrelated with $\{b(n)\}$, with $E\{[\mathbf{v}(n+\tau) - \mathbf{m}][\mathbf{v}(n) - \mathbf{m}]^H\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$. The mean vector \mathbf{m} may be unknown.
- (H3.4.4) The superimposed training sequence $c(n) = c(n+P)$ for all n is a non-random periodic sequence with period P .

By the SIMO channel model (3.1)–(3.2), and the OP-BEM representation (3.38), we have

$$\mathbf{y}(n) = \sum_{l=0}^L \sum_{q=0}^K \mathbf{h}_q(l) \varphi_q(n) [b(n-l) + c(n-l)] + \mathbf{v}(n)$$

with mean

$$E\{\mathbf{y}(n)\} = \sum_{l=0}^L \sum_{q=0}^K \mathbf{h}_q(l) \varphi_q(n) c(n-l) + \mathbf{m}.$$

It follows that

$$\mathbf{y}(n) = E\{\mathbf{y}(n)\} + \mathbf{e}(n)$$

where $\{\mathbf{e}(n)\}$ is a zero-mean random “error” sequence and

$$\mathbf{e}(n) = \mathbf{y}(n) - \sum_{l=0}^L \sum_{q=0}^K \mathbf{h}_q(l) \varphi_q(n) c(n-l) - \mathbf{m}.$$

We define the cost function as

$$J = \sum_{n=0}^{T-1} \|\mathbf{e}(n)\|^2 = \sum_{n=0}^{T-1} \left\| \mathbf{y}(n) - \sum_{l=0}^L \sum_{q=0}^K \mathbf{h}_q(l) \varphi_q(n) c(n-l) - \mathbf{m} \right\|^2.$$

Choose \mathbf{m} and $\mathbf{h}_q(l)$'s ($q = 0, 1, \dots, K; l = 0, 1, \dots, L$) to minimize the cost function J . We must have

$$\frac{\partial J}{\partial \mathbf{m}^*} \Big|_{\substack{\mathbf{m}=\hat{\mathbf{m}} \\ \mathbf{h}_q(l)=\hat{\mathbf{h}}_q(l)}} = \mathbf{0}, \text{ and } \frac{\partial J}{\partial \mathbf{h}_q^*(l)} \Big|_{\substack{\mathbf{m}=\hat{\mathbf{m}} \\ \mathbf{h}_q(l)=\hat{\mathbf{h}}_q(l)}} = \mathbf{0},$$

leading to

$$\hat{\mathbf{m}} = \frac{1}{T} \sum_{n=0}^{T-1} \left[\mathbf{y}(n) - \sum_{l=0}^L \sum_{q=0}^K \mathbf{h}_q(l) \varphi_q(n) c(n-l) \right] \quad (3.39)$$

and

$$\begin{aligned} & \sum_{l=0}^L \sum_{q=0}^K \hat{\mathbf{h}}_q(l) \left[\frac{1}{T} \sum_{n=0}^{T-1} \varphi_q(n) \varphi_{q_1}^*(n) c(n-l) c^*(n-l_1) \right] \\ &= \frac{1}{T} \sum_{n=0}^{T-1} [\mathbf{y}(n) - \hat{\mathbf{m}}] \varphi_{q_1}^*(n) c^*(n-l_1). \end{aligned} \quad (3.40)$$

Substitute (3.39) in (3.40), we then have

$$\sum_{l=0}^L \sum_{q=0}^K \hat{\mathbf{h}}_q(l) \phi[(q, l), (q_1, l_1)] = \frac{1}{T} \sum_{n=0}^{T-1} \mathbf{y}(n) \left[\varphi_{q_1}^*(n) c^*(n-l_1) - \frac{1}{T} \sum_{n=0}^{T-1} \varphi_{q_1}^*(n) c^*(n-l_1) \right], \quad (3.41)$$

where we define

$$\begin{aligned} \phi[(q, l), (q_1, l_1)] &:= \frac{1}{T} \sum_{n=0}^{T-1} \varphi_q(n) \varphi_{q_1}^*(n) c(n-l) c^*(n-l_1) \\ &\quad - \left[\frac{1}{T} \sum_{n=0}^{T-1} \varphi_q(n) c(n-l) \right] \left[\frac{1}{T} \sum_{n=0}^{T-1} \varphi_{q_1}^*(n) c^*(n-l_1) \right]. \end{aligned} \quad (3.42)$$

We further define a $(K+1)(L+1) \times (K+1)(L+1)$ matrix

$$\mathbf{\Phi}_{\text{OP}} := \begin{bmatrix} \phi[(0,0),(0,0)] & \cdots & \phi[(K,0),(0,0)] & \cdots & \phi[(K,1),(0,0)] & \cdots & \phi[(K,L),(0,0)] \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi[(0,0),(K,0)] & \cdots & \phi[(K,0),(K,0)] & \cdots & \phi[(K,1),(K,0)] & \cdots & \phi[(K,L),(K,0)] \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi[(0,0),(K,1)] & \cdots & \phi[(K,0),(K,1)] & \cdots & \phi[(K,1),(K,1)] & \cdots & \phi[(K,L),(K,1)] \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi[(0,0),(K,L)] & \cdots & \phi[(K,0),(K,L)] & \cdots & \phi[(K,1),(K,L)] & \cdots & \phi[(K,L),(K,L)] \end{bmatrix} \quad (3.43)$$

whose $[(K+1)l_1 + q_1 + 1, (K+1)l + q + 1]$ -th entry is $\phi[(q,l),(q_1,l_1)]$, a $(K+1)(L+1)$ -column vector

$$\mathbf{\Delta}(n) := \begin{bmatrix} \varphi_0^*(n) c^*(n) - \frac{1}{T} \sum_{n=0}^{T-1} \varphi_0^*(n) c^*(n) \\ \vdots \\ \varphi_K^*(n) c^*(n) - \frac{1}{T} \sum_{n=0}^{T-1} \varphi_K^*(n) c^*(n) \\ \varphi_0^*(n) c^*(n-1) - \frac{1}{T} \sum_{n=0}^{T-1} \varphi_0^*(n) c^*(n-1) \\ \vdots \\ \varphi_K^*(n) c^*(n-1) - \frac{1}{T} \sum_{n=0}^{T-1} \varphi_K^*(n) c^*(n-1) \\ \vdots \\ \varphi_K^*(n) c^*(n-L) - \frac{1}{T} \sum_{n=0}^{T-1} \varphi_K^*(n) c^*(n-L) \end{bmatrix},$$

and the channel coefficient vectors

$$\hat{\mathcal{H}}_l := \left[\hat{\mathbf{h}}_0^T(l) \quad \hat{\mathbf{h}}_1^T(l) \quad \cdots \quad \hat{\mathbf{h}}_K^T(l) \right]^T, \quad (3.44)$$

$$\hat{\mathcal{H}} := \left[\hat{\mathcal{H}}_0^T \quad \hat{\mathcal{H}}_1^T \quad \cdots \quad \hat{\mathcal{H}}_L^T \right]^T. \quad (3.45)$$

We also define \mathcal{H}_l and \mathcal{H} as the vectors of true values of channel coefficients, corresponding to (3.44) and (3.45), respectively. Then (3.41) can be written into

$$(\Phi_{\text{OP}} \otimes \mathbf{I}_N) \hat{\mathcal{H}} = \frac{1}{T} \sum_{n=0}^{T-1} \Delta(n) \otimes \mathbf{y}(n),$$

which yields

$$\hat{\mathcal{H}} = \frac{1}{T} \sum_{n=0}^{T-1} \Phi_{\text{OP}}^\dagger \Delta(n) \otimes \mathbf{y}(n). \quad (3.46)$$

The estimate of the time-varying channel $\mathbf{h}(n; l)$ is then given by

$$\hat{\mathbf{h}}(n; l) = \sum_{q=0}^K \hat{\mathbf{h}}_q(l) \varphi_q(n). \quad (3.47)$$

Remark 3.4.1: We did not use assumption **(H3.4.4)** throughout the manipulations, so that aperiodic training can be used also.

Remark 3.4.2: Neither orthogonality nor properties of polynomials was used to obtain (3.46). We can use any basis set $\{\psi_q(n)\}_{n=0}^{T-1}$ ($q = 1, 2, \dots, Q$), as in (2.20), instead of polynomials in (3.46). The basis functions are only required to be linearly independent, not necessarily orthogonal. This offers a general channel estimator using BEM's and superimposed training.

If we consider CE-BEM in (3.42) and replace $\varphi_q(n)$ with $\psi_q(n) = e^{j\omega_q n}$ for $\omega_q := \frac{2\pi}{T}(q - \frac{1}{2} - \frac{Q}{2})$, we have

$$\begin{aligned} \phi[(q, l), (q_1, l_1)] &= \frac{1}{T} \sum_{n=0}^{T-1} e^{j\omega_q n} e^{-j\omega_{q_1} n} c(n-l) c^*(n-l_1) \\ &\quad - \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{j\omega_q n} c(n-l) \right] \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{-j\omega_{q_1} n} c^*(n-l_1) \right]. \end{aligned} \quad (3.48)$$

Taking (3.5) into account, we have

$$\frac{1}{T} \sum_{n=0}^{T-1} e^{j\omega_q n} e^{-j\omega_{q_1} n} c(n-l) c^*(n-l_1) = \sum_{m=0}^{P-1} |c_m|^2 e^{-j\alpha_m(l-l')} \delta(q - q_1) \quad (3.49)$$

and

$$\frac{1}{T} \sum_{n=0}^{T-1} e^{j\omega_q n} c(n-l) = c_0 \delta\left(q - \frac{Q+1}{2}\right), \quad (3.50)$$

so that (3.48) becomes

$$\phi[(q, l), (q_1, l_1)] = \sum_{m=0}^{P-1} |c_m|^2 e^{-j\alpha_m(l-l_1)} \delta(q - q_1) - |c_0|^2 \delta\left(q - \frac{Q+1}{2}\right) \delta\left(q_1 - \frac{Q+1}{2}\right). \quad (3.51)$$

Substituting (3.50) and (3.51) into (3.41), it follows that

$$\begin{aligned} &\sum_{m=0}^{P-1} c_m^* e^{j\alpha_m l_1} \left[\sum_{l=0}^L c_m \hat{\mathbf{h}}_q(l) e^{-j\alpha_m l} \right] - \sum_{l=0}^L \hat{\mathbf{h}}_{\frac{Q+1}{2}}(l) |c_0|^2 \delta\left(q - \frac{Q+1}{2}\right) \\ &= \sum_{m=0}^{P-1} c_m^* e^{j\alpha_m l_1} \left[\frac{1}{T} \sum_{n=0}^{T-1} \mathbf{y}(n) e^{-j(\omega_q + \alpha_m)n} \right] - \frac{1}{T} \sum_{n=0}^{T-1} \mathbf{y}(n) c_0^* \delta\left(q - \frac{Q+1}{2}\right). \end{aligned} \quad (3.52)$$

By (3.6) and (3.9) we have

$$\sum_{l=0}^L c_m \hat{\mathbf{h}}_q(l) e^{-j\alpha_m l} = \frac{1}{T} \sum_{n=0}^{T-1} \mathbf{y}(n) e^{-j(\omega_q + \alpha_m)n} \quad (3.53)$$

for $q = 1, 2, \dots, Q$ and $m = 1, 2, \dots, P-1$. If we discard the terms corresponding to $m = 0$, then (3.52) reduces to

$$\sum_{m=1}^{P-1} c_m^* e^{j\alpha_m l_1} \left[\sum_{l=0}^L c_m \hat{\mathbf{h}}_q(l) e^{-j\alpha_m l} \right] = \sum_{m=1}^{P-1} c_m^* e^{j\alpha_m l_1} \left[\frac{1}{T} \sum_{n=0}^{T-1} \mathbf{y}(n) e^{-j(\omega_q + \alpha_m)n} \right] \quad (3.54)$$

for $q = 1, 2, \dots, Q$. The solution to (3.54) coincides with that of (3.53) if the matrix \mathcal{C} in (3.13) has full column rank. This result demonstrates that the estimator given by (3.46) is indeed the same as the one proposed in Section 3.2. The only difference appears that since \mathbf{m} is unknown, we simply discard all the terms corresponding to $m = 0$ in Section 3.2. However, omitting the terms corresponding to $\omega_q + \alpha_m = 0$ is enough, and the terms corresponding to $m = 0$ but $\omega_q + \alpha_m \neq 0$ are still useful—this is the estimator proposed in this section. Using the band-limitedness of the DPS-BEM (3.34), we have

$$\sum_{l=0}^L c_m \hat{\mathbf{h}}_q(l) e^{-j\alpha_m l} = \sum_{n=0}^{T-1} \mathbf{y}(n) u_q(n) e^{-j\alpha_m n},$$

which is similar to (3.53). Thus the estimator proposed in Section 3.3 is also a special case of the estimator defined by (3.41).

Remark 3.4.3: If the mean value of the noise \mathbf{m} is zero or known, then instead of (3.41), the solution of the estimator is given by (3.40) with $\hat{\mathbf{m}} = \mathbf{m}$. We define

$$\tilde{\phi}[(q, l), (q_1, l_1)] := \frac{1}{T} \sum_{n=0}^{T-1} \varphi_q(n) \varphi_{q_1}^*(n) c(n-l) c^*(n-l_1) \quad (3.55)$$

and $\tilde{\Phi}_{\text{OP}}$ in the same way as in (3.43) with $\phi[(q, l), (q_1, l_1)]$ replaced with $\tilde{\phi}[(q, l), (q_1, l_1)]$. Also define

$$\tilde{\Delta}(n) := \begin{bmatrix} c^*(n) & \cdots & c^*(n-L) \end{bmatrix}^T \otimes \begin{bmatrix} \varphi_0^*(n) & \cdots & \varphi_K^*(n) \end{bmatrix}^T.$$

The estimator is given by

$$\hat{\mathcal{H}} = \frac{1}{T} \sum_{n=0}^{T-1} \tilde{\Phi}_{\text{OP}}^\dagger \tilde{\Delta}(n) \otimes \mathbf{y}(n). \quad (3.56)$$

Note that the estimators given by (3.24) and (3.36), with zero or known \mathbf{m} , are special cases of the estimator of (3.56), using CE- and DPS-BEM respectively.

3.5 First-Order Statistics-Based Channel Estimation: Multiple-User (MIMO) Channels

The first-order statistics-based channel estimator using CE- or DPS-BEM can be easily extended to a multiple-user (MIMO) system, by exploiting the band-limitedness of the basis functions.

Consider an MIMO FIR linear channel with K inputs and N outputs. Let $\{s_k(n)\}$ denote k -th user's information sequence which is input to the MIMO channel with the k -th

user's time-varying discrete-time impulse response $\{\mathbf{h}_k(n; l)\}$ (channel response for the k -th user at time instance n to a unit input at time instance $n - l$). The symbol-rate, N -column channel output vector is given by

$$\mathbf{x}(n) = \sum_{k=1}^K \sum_{l=0}^L \mathbf{h}_k(n; l) s_k(n - l). \quad (3.57)$$

The noisy measurement is given by

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n), \quad (3.58)$$

where $\mathbf{v}(n)$ is the additive N -column vector white complex Gaussian noise.

In superimposed training-based approaches, for the k -th user, one takes

$$s_k(n) = b_k(n) + c_k(n) \quad (3.59)$$

where $\{b_k(n)\}$ and $\{c_k(n)\}$ are information sequence and non-random periodic training sequence of the k -th user. Then the noisy channel output becomes

$$\mathbf{y}(n) = \sum_{k=1}^K \sum_{l=0}^L \mathbf{h}_k(n; l) [b_k(n - l) + c_k(n - l)] + \mathbf{v}(n).$$

Assume the following:

(H3.5.1) The time-varying channel $\{\mathbf{h}_k(n; l)\}$ satisfies CE- or DPS-BEM, i.e.,

$$\mathbf{h}_k(n; l) = \sum_{q=1}^Q \mathbf{h}_{qk}(l) e^{j\omega_q n} \quad (3.60)$$

where $\mathbf{h}_{qk}(l)$ is the N -column time-invariant coefficient vector for the k -th user, or

$$\mathbf{h}_k(n; l) = \sum_{q=1}^Q \mathbf{h}_{qk}(l) u_q(n). \quad (3.61)$$

Also $N \geq 1$.

(H3.5.2) The information sequence $\{b_k(n)\}$ is zero-mean, white with $E\{|b_k(n)|^2\} = \sigma_{bk}^2$ and mutually independent for $k = 1, 2, \dots, K$.

(H3.5.3) The measurement noise $\{\mathbf{v}(n)\}$ is zero-mean, white, uncorrelated with $\{b_k(n)\}$, with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$.

(H3.5.4) The superimposed training sequence $c_k(n) = c_k(n + P)$ for all n is a non-random periodic sequence with period P such that $c_{mk} \neq 0$ for all m, k , and \tilde{P} is integer with $P = \tilde{P}K$.

The expected value of the noisy channel output is given by

$$E\{\mathbf{y}(n)\} = \sum_{k=1}^K \sum_{l=0}^L \mathbf{h}_k(n; l) c_k(n - l). \quad (3.62)$$

We pick user-specific training sequences so that channel estimation is decoupled across various users—this allows us to use the single user superimposed training based approach outlined in Section 3.2. We assign distinct cycle frequencies of the periodic training sequences to distinct users. Suppose that for each user k , $\{c_k(n)\}$ is periodic with period $P = \tilde{P}K$ where \tilde{P} is a positive integer. Then

$$c_k(n) = \sum_{m'=0}^{P-1} c_{m'k} e^{j(2\pi m'/P)n} \quad \text{for all } n.$$

Define

$$c_{m'k} := \frac{1}{P} \sum_{n=0}^{P-1} c_k(n) e^{-j(2\pi m'/P)n}.$$

Pick $\{c_k(n)\}$ so that only \tilde{P} coefficients (out of total P) $c_{m'k}$'s, associated with \tilde{P} distinct frequencies, are nonzero. For instance, we may choose

$$c_k(n) = \sum_{m=0}^{\tilde{P}-1} c_{mk} e^{j(2\pi/P)(Km+k-1)n}, \text{ for all } n \quad (3.63)$$

such that $c_{mk} \neq 0$ for all m, k . Define the frequencies

$$\alpha_{mk} := \frac{2\pi}{P}(Km + k - 1) \quad (3.64)$$

for $m = 0, 1, \dots, \tilde{P} - 1$ and $k = 1, 2, \dots, K$. For example, we show how to use m -sequences (maximal length pseudo-random binary sequences) for training. Pick $\tilde{P} = 2^n - 1$ for some integer n such that $\tilde{P} \geq L + 2$. Let $\{\bar{c}_0(n)\}$ be an m -sequence of length \tilde{P} . Pick the superimposed training sequence $\{c_1(n)\}_{n=0}^{P-1}$ for user 1 as K repetitions of $\{\bar{c}_0(n)\}$ multiplied by a factor σ_{c_1} so that $P^{-1} \sum_{n=0}^{P-1} |c_1(n)|^2 = \sigma_{c_1}^2$. This choice satisfies (3.63) and (3.64) for $k = 1$. Pick $\bar{c}_k(n) = \bar{c}_1(n) e^{j(2\pi/P)(k-1)n}$ for $k = 2, 3, \dots, K$ and $c_k(n) = \sigma_{c_k} \bar{c}_k(n)$. Then $\{c_k(n)\}_{n=0}^{P-1}$ satisfies (3.63) and (3.64) for $k \geq 2$. The above procedures can be used to generate a user-specific training sequence of period $P = \tilde{P}K$ from a sequence of period \tilde{P} .

It follows that

$$E\{\mathbf{y}(n)\} = \sum_{k=1}^K \sum_{l=0}^L \mathbf{h}_k(n; l) \left[\sum_{m=0}^{\tilde{P}-1} c_{mk} e^{j\alpha_{mk}} \right], \text{ for all } n. \quad (3.65)$$

We can write (3.60) and (3.61) into a unified form as (2.20)

$$\mathbf{h}_k(n; l) = \sum_{q=1}^Q \mathbf{h}_{qk}(l) \psi_q(n), \quad (3.66)$$

where $\psi_q(n) = e^{j\omega_q n}$ in CE-BEM, and $\psi_q(n) = u_q(n)$ in DPS-BEM. The expected value of the observations (3.65) can be rewritten as

$$E\{\mathbf{y}(n)\} = \sum_{k=1}^K \sum_{m=0}^{\tilde{P}-1} \sum_{q=1}^Q \left[\sum_{l=0}^L \mathbf{h}_{qk}(l) c_{mk} e^{-j\alpha_{mk} l} \right] \psi_q(n) e^{j\alpha_{mk} n}, \quad \text{for all } n.$$

Define

$$\mathbf{d}_{mqk} := \sum_{l=0}^L \mathbf{h}_{qk}(l) c_{mk} e^{-j\alpha_{mk} l}. \quad (3.67)$$

We have

$$E\{\mathbf{y}(n)\} = \sum_{k=1}^K \sum_{m=0}^{\tilde{P}-1} \sum_{q=1}^Q \mathbf{d}_{mqk} \psi_q(n) e^{j\alpha_{mk} n}.$$

It follows that

$$\mathbf{y}(n) = \sum_{k=1}^K \sum_{q=1}^Q \sum_{m=0}^{P-1} \mathbf{d}_{mqk} \psi_q(n) e^{j\alpha_{mk} n} + \mathbf{e}(n) \quad (3.68)$$

where $\{\mathbf{e}(n)\}$ is a zero-mean random sequence.

Define the cost function by (3.68)

$$J = \sum_{n=0}^{T-1} \|\mathbf{e}(n)\|^2 = \sum_{n=0}^{T-1} \left\| \mathbf{y}(n) - \sum_{k=1}^K \sum_{q=1}^Q \sum_{m=0}^{P-1} \mathbf{d}_{mqk} \psi_q(n) e^{j\alpha_{mk} n} \right\|^2.$$

Choose \mathbf{d}_{mqk} 's to minimize J . We must have

$$\begin{aligned} & \left. \frac{\partial J}{\partial \mathbf{d}_{mqk}^*} \right|_{\mathbf{d}_{mqk} = \hat{\mathbf{d}}_{mqk}} \\ &= \sum_{n=0}^{T-1} \left[\mathbf{y}(n) - \sum_{k'=1}^K \sum_{q'=1}^Q \sum_{m'=0}^{P-1} \mathbf{d}_{m'q'k'} \psi_{q'}(n) e^{j\alpha_{m'k'}n} \right] \psi_q^*(n) e^{-j\alpha_{mk}n} \\ &= 0, \end{aligned}$$

leading to

$$\sum_{k'=1}^K \sum_{q'=1}^Q \sum_{m'=0}^{P-1} \hat{\mathbf{d}}_{m'q'k'} \left[\sum_{n=0}^{T-1} \psi_{q'}(n) \psi_q^*(n) e^{j(\alpha_{m'k'} - \alpha_{mk})n} \right] = \sum_{n=0}^{T-1} \mathbf{y}(n) \psi_q^*(n) e^{-j\alpha_{mk}n}. \quad (3.69)$$

For CE-BEM, suppose that we pick P to be such that $(\omega_q + \alpha_{mk})$'s are all distinct for any choice of m , k and q , e.g., take $T/P \geq Q$. Then the sequence $E\{\mathbf{y}(n)\}$ is periodic [14] with cycle frequencies $(\omega_q + \alpha_{mk})$, where $1 \leq q \leq Q$, $0 \leq m \leq P-1$ and $1 \leq k \leq K$, so that we have

$$\sum_{n=0}^{T-1} e^{j(\omega_{q'} - \omega_q + \alpha_{m'k'} - \alpha_{mk})n} = T \delta(m' - m) \delta(k' - k) \delta(q' - q). \quad (3.70)$$

It follows that

$$\hat{\mathbf{d}}_{mqk} = \frac{1}{T} \sum_{n=1}^T \mathbf{y}(n) e^{-j(\omega_q + \alpha_{mk})n}. \quad (3.71)$$

If DPS-BEM applies, we use the fact that DPS sequences are approximately band-limited to the normalized frequency range $[-f_d T_s, f_d T_s]$. Then if $2f_d T_s \ll 1/P$, we use the approximation

$$\sum_{n=0}^{T-1} u_{q'}(n) u_q(n) e^{j(\alpha_{m'k'} - \alpha_{mk})n} \approx \delta(k' - k) \delta(m' - m) \delta(q' - q). \quad (3.72)$$

By using the approximation (3.72),

$$\hat{\mathbf{d}}_{mqk} = \sum_{n=1}^T \mathbf{y}(n) u_q(n) e^{-j\alpha_{mk}n}. \quad (3.73)$$

For $0 \leq m \leq \tilde{P} - 1$ and $1 \leq k \leq K$, define an NQ -column vector

$$\mathbf{D}_{mk} := \begin{bmatrix} \mathbf{d}_{m1k}^T & \mathbf{d}_{m2k}^T & \cdots & \mathbf{d}_{mQk}^T \end{bmatrix}^T, \quad (3.74)$$

and for $0 \leq l \leq L$ and $1 \leq k \leq K$, define an NQ -column vector

$$\mathbf{H}_{kl} := \begin{bmatrix} \mathbf{h}_{1k}^H(l) & \mathbf{h}_{2k}^H(l) & \cdots & \mathbf{h}_{Qk}^H(l) \end{bmatrix}^H.$$

Then by equation (3.67), we have

$$\mathbf{D}_{mk} = \sum_{l=0}^L c_{mk} e^{-j\alpha_{mk}l} \mathbf{H}_{kl} \quad (3.75)$$

for $0 \leq m \leq \tilde{P} - 1$ and $1 \leq k \leq K$. Define a $\tilde{P} \times (L + 1)$ matrix

$$\tilde{\mathbf{V}}_k := \begin{bmatrix} 1 & e^{-j\alpha_{0k}} & \dots & e^{-j\alpha_{0k}L} \\ 1 & e^{-j\alpha_{1k}} & \dots & e^{-j\alpha_{1k}L} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\alpha_{(\tilde{P}-1)k}} & \dots & e^{-j\alpha_{(\tilde{P}-1)k}L} \end{bmatrix}, \quad (3.76)$$

an $NQ\tilde{P} \times NQ(L+1)$ matrix

$$\tilde{\mathcal{C}}_k := (\text{diag} \{c_{0k}, c_{1k}, \dots, c_{(\tilde{P}-1)k}\} \mathbf{V}_k) \otimes \mathbf{I}_{NQ},$$

an $NQ(L+1)$ -column vector

$$\mathcal{H}_k = \begin{bmatrix} \mathbf{H}_{k0}^H & \mathbf{H}_{k1}^H & \dots & \mathbf{H}_{kL}^H \end{bmatrix}^H \quad (3.77)$$

and an NQP -column vector

$$\tilde{\mathcal{D}}_k = \begin{bmatrix} \mathbf{D}_{0k}^H & \mathbf{D}_{1k}^H & \dots & \mathbf{D}_{(\tilde{P}-1)k}^H \end{bmatrix}^H. \quad (3.78)$$

Then (3.75) leads to

$$\tilde{\mathcal{C}}_k \mathcal{H}_k = \tilde{\mathcal{D}}_k. \quad (3.79)$$

Since α_{mk} 's are distinct and $c_{mk} \neq 0$ for all m, k , $\text{rank}(\tilde{\mathcal{C}}_k) = NQ(L+1)$ if $\tilde{P} \geq L+1$; hence, we can determine $\mathbf{h}_{qk}(l)$'s uniquely. Define $\hat{\mathbf{D}}_{mk}$ as in (3.74) with \mathbf{d}_{mqk} replaced with $\hat{\mathbf{d}}_{mqk}$ and define $\hat{\tilde{\mathcal{D}}}_k$ as in (3.78) with \mathbf{D}_{mk} replaced with $\hat{\mathbf{D}}_{mk}$. The estimate of \mathcal{H}_k is given by

$$\hat{\mathcal{H}}_k = (\tilde{\mathcal{C}}_k^H \tilde{\mathcal{C}}_k)^{-1} \tilde{\mathcal{C}}_k^H \hat{\tilde{\mathcal{D}}}_k. \quad (3.80)$$

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=0\text{Hz}$, 500 runs.

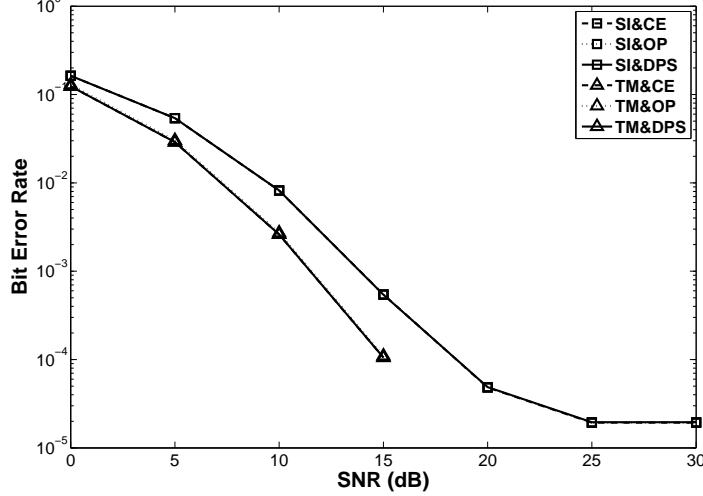


Figure 3.1: First-order statistics-based estimator (SISO): BER vs SNR under $f_d = 0\text{Hz}$ (time-invariant) and $K = N = 1$. The curves for CE-, OP- and DPS-BEM's completely overlap, since the three basis functions are all constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

Denote the corresponding estimate of $\mathbf{h}_{qk}(l)$ as $\hat{\mathbf{h}}_{qk}(l)$. Following (3.66), for $k = 1, 2, \dots, K$, $l = 0, 1, \dots, L$ and for all n , the estimate of time-varying channel is given by

$$\hat{\mathbf{h}}_k(n; l) = \sum_{q=1}^Q \hat{\mathbf{h}}_{qk}(l) \psi_q(n). \quad (3.81)$$

Remark 3.5.1: If the additive noise $\mathbf{v}(n)$ is nonzero-mean ($E\{\mathbf{v}(n)\} = \mathbf{m}$) with the mean unknown. Then (3.69) can be modified as

$$\sum_{k'=1}^K \sum_{q'=1}^Q \sum_{m'=0}^{P-1} \hat{\mathbf{d}}_{m'q'k'} \left[\sum_{n=0}^{T-1} \psi_{q'}(n) \psi_q^*(n) e^{j(\alpha_{m'k'} - \alpha_{mk})n} \right] = \sum_{n=0}^{T-1} [\mathbf{y}(n) - \mathbf{m}] \psi_q^*(n) e^{-j\alpha_{mk}n}.$$

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=50\text{Hz}$, 500 runs.

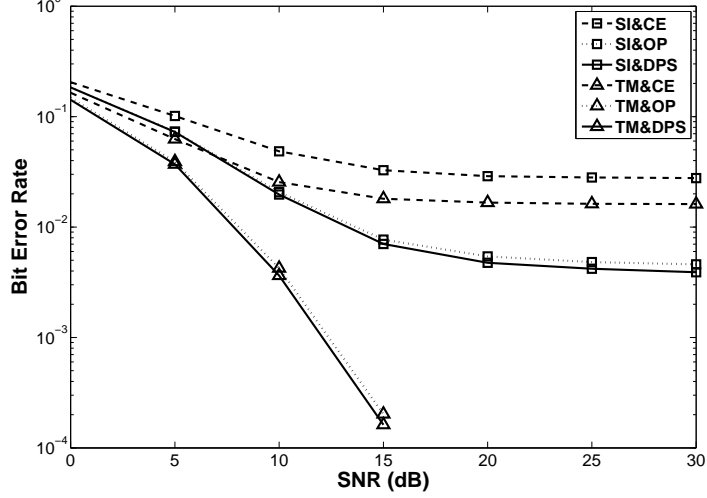


Figure 3.2: First-order statistics-based estimator (SISO): BER vs SNR under $f_d = 50$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

Note that

$$\sum_{n=0}^{T-1} e^{-j(\omega_q + \alpha_{mk})n} = T\delta(\omega_q + \alpha_{mk}),$$

and if $2f_d T_s \ll 1/P$, approximately we have

$$\sum_{n=0}^{T-1} u_q(n) e^{-j\alpha_{mk}n} \approx 0, \quad \text{for all } \alpha_{mk} \neq 0.$$

Since $T/P \geq Q$ and $\alpha_{mk} = 0$ only happens when $m = 0$ and $k = 1$, if $m = 0$ we have

$$\sum_{n=0}^{T-1} \mathbf{m}\psi_q^*(n) e^{-j\alpha_{mk}n} = 0.$$

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $TIR=0.3$, $P=7$, $f_d=100\text{Hz}$, 500 runs.

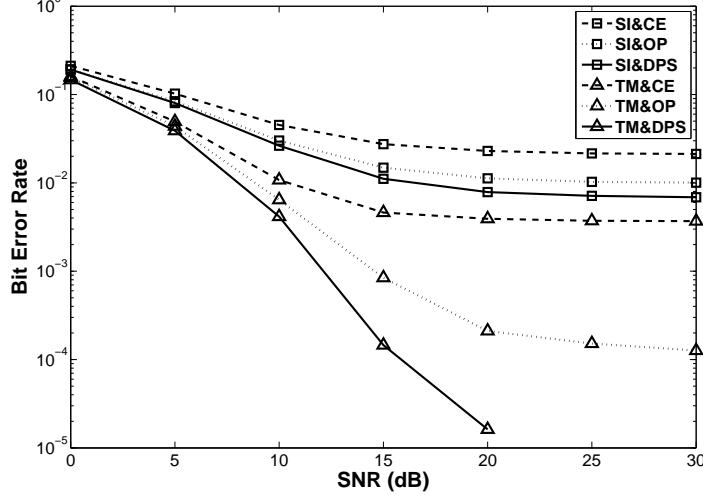


Figure 3.3: First-order statistics-based estimator (SISO): BER vs SNR under $f_d = 100$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

We hence omit the terms corresponding to $m = 0$, and then \mathbf{m} has no effect on estimation.

We should only omit the first row of $\tilde{\mathbf{V}}_k$ in (3.76), and the resulting $(\tilde{P} - 1) \times (L + 1)$ matrix is denoted by

$$\mathbf{V}_k := \begin{bmatrix} 1 & e^{-j\alpha_{1k}} & \dots & e^{-j\alpha_{1k}L} \\ 1 & e^{-j\alpha_{2k}} & \dots & e^{-j\alpha_{2k}L} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\alpha_{(\tilde{P}-1)k}} & \dots & e^{-j\alpha_{(\tilde{P}-1)k}L} \end{bmatrix}.$$

We also define an $NQ(\tilde{P} - 1) \times NQ(L + 1)$ matrix

$$\mathcal{C}_k := \left(\text{diag} \left\{ c_{1k}, c_{2k}, \dots, c_{(\tilde{P}-1)k} \right\} \mathbf{V}_k \right) \otimes \mathbf{I}_{NQ}$$

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $TIR=0.3$, $P=7$, $f_d=200\text{Hz}$, 500 runs.

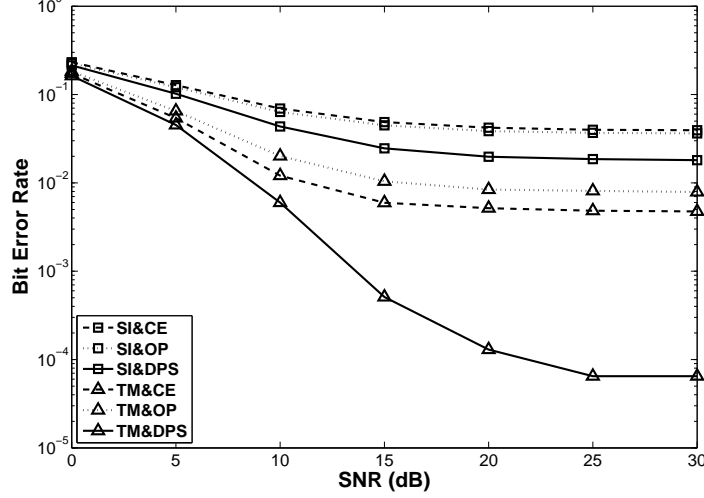


Figure 3.4: First-order statistics-based estimator (SISO): BER vs SNR under $f_d = 200$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

and an $NQ(P-1)$ -column vector

$$\mathcal{D}_k = \left[\mathbf{D}_{1k}^H \quad \mathbf{D}_{2k}^H \quad \dots \quad \mathbf{D}_{(\tilde{P}-1)k}^H \right]^H.$$

Then as in (3.79), we have

$$\mathcal{C}_k \mathcal{H}_k = \mathcal{D}_k,$$

so that the channel estimation is given by

$$\hat{\mathcal{H}}_k = (\mathcal{C}_k^H \mathcal{C}_k)^{-1} \mathcal{C}_k^H \hat{\mathcal{D}}_k,$$

where $\hat{\mathcal{D}}_k$ is also acquired by (3.73). For identifiability, we now need $\tilde{P} \geq L + 2$.

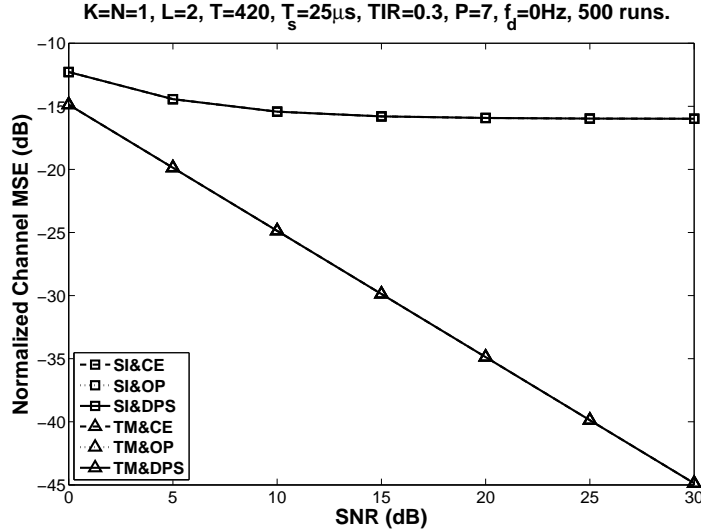


Figure 3.5: First-order statistics-based estimator (SISO): NCMSE vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 1$. The curves for CE-, OP- and DPS-BEM's completely overlap, since the three basis functions are all constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

3.6 Simulation Examples

3.6.1 First-Order Statistics-Based Estimator: Single User

In this example, we generate a doubly-selective Rayleigh fading channel as we mentioned in Section 2.6.2, with $N = 1$ and $L = 2$, satisfying modified Jakes' model. We also employ the communication system described in Section 2.6.3. We emphasize again that BEM's are only used for processing at the receiver; the "true" channel follows Jakes' model, not BEM's.

In simulations, we pick a data record length of 420 symbols (time duration of approximately 10 ms). We consider the system operating under different Doppler spreads. For the

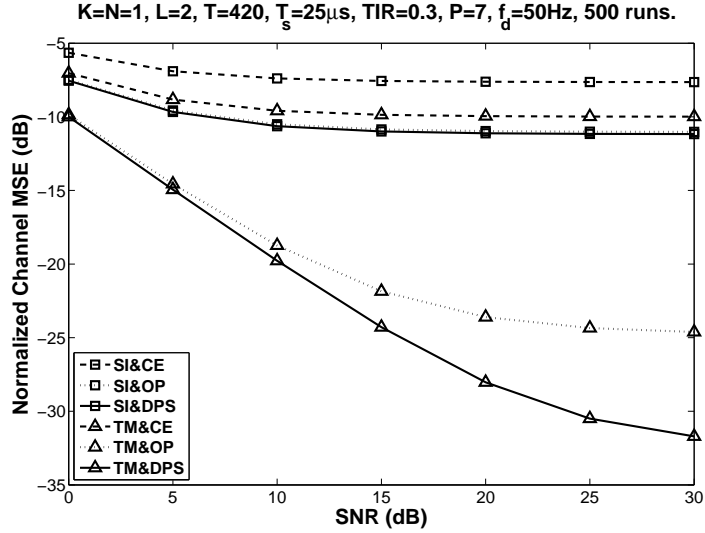


Figure 3.6: First-order statistics-based estimator (SISO): NCMSE vs SNR under $f_d = 50$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

Doppler spreads $f_d = 0, 50, 100,$ and 200 Hz, we take $Q = 1, 3, 5,$ and 7 in CE-BEM by (2.9b), and $Q = 1, 3, 4,$ and 6 for OP- and DPS-BEM representations by (2.18). The average transmitted power in $\{c(n)\}$ is 0.3 of that in $\{b(n)\}$, leading to a training-to-information power ratio (TIR) of 0.3.

We consider a single-user scenario. The information sequence $\{b(n)\}$ and the training sequence $\{c(n)\}$ are all modulated by binary phase-shift keying (BPSK). The periodic training sequence $\{c(n)\}$ is generated from the m -sequence of period $P = 7$, one period of which is given by

$$\{c_1(n)\}_{n=0}^6 = \{1, -1, -1, 1, 1, 1, -1\}. \quad (3.82)$$

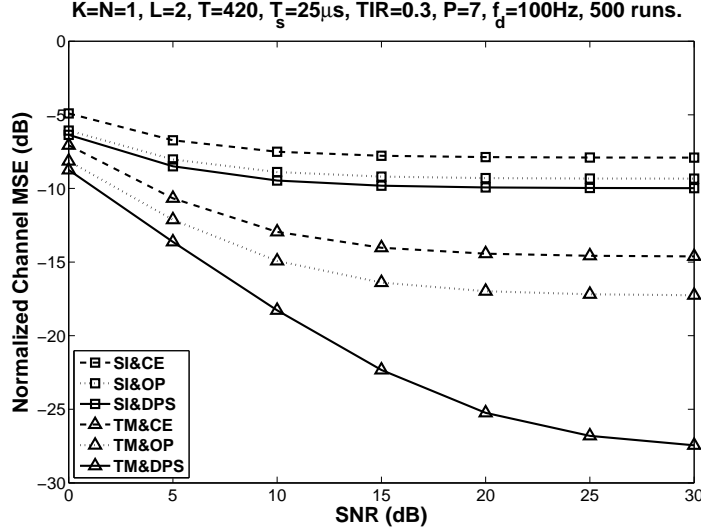


Figure 3.7: First-order statistics-based estimator (SISO): NCMSE vs SNR under $f_d = 100\text{Hz}$ and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

To explore different estimators under equal conditions, we assume the additive noise $\{\mathbf{v}(n)\}$ is zero-mean (i.e., $\mathbf{m} = 0$), white complex-Gaussian, uncorrelated with $\{b(n)\}$ with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$, so that no terms are discarded. The (receiver) SNR refers to the energy per bit over one-sided noise spectral density with both information and superimposed training sequence counting toward the bit energy.

The results are shown in Figures 3.1–3.8 for various Doppler spreads and SNR's. The results are based on 500 Monte Carlo runs for Viterbi detectors (see Appendix B.1). For comparison, CE-, OP- and DPS-BEM-based periodically placed TM training with zero-padding (see Appendix A) is also considered for doubly-selective channel estimation. We take a training session of length $2L + 1 = 5$ symbols with the training sequence $\{0, 0, \sqrt{2L + 1}, 0, 0\}$, and at the receiver an LS estimation is performed. A data session of 17 symbols is inserted

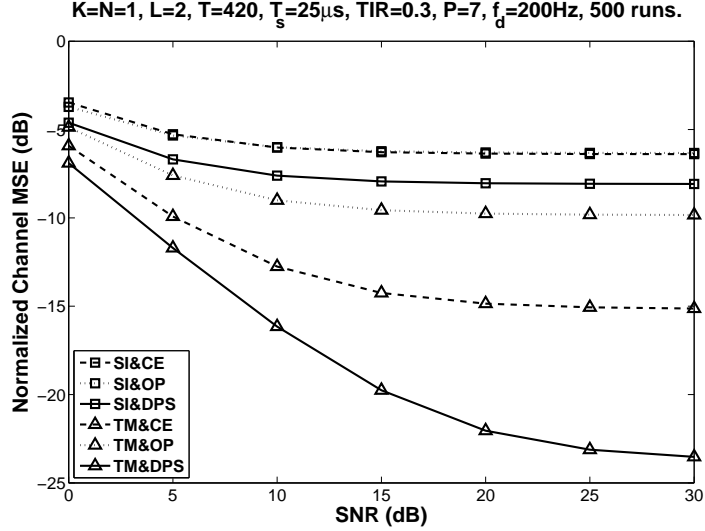


Figure 3.8: First-order statistics-based estimator (SISO): NCMSE vs SNR under $f_d = 200$ Hz and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

between two such training sessions to form a frame of length 22 symbols. Such a block is repeated over a record length of 418 symbols. Thus, we have training-to-information bit ratio as well as training-to-information power ratio of about 0.3.

For comparison, we plot the results of CE-, OP-, and DPS-based superimposed and TM training approaches in each figure. Figures 3.1–3.4 show the BER's with a Viterbi detector at the receiver. Figures 3.5–3.8 show the normalized channel mean square error (NCMSE) correspondingly, which is defined as

$$\text{NCMSE} := \frac{\sum_{i=1}^{M_c} \sum_{n=0}^{T-1} \sum_{l=0}^2 \left\| \mathbf{h}^{(i)}(n;l) - \hat{\mathbf{h}}^{(i)}(n;l) \right\|^2}{\sum_{i=1}^{M_c} \sum_{n=0}^{T-1} \sum_{l=0}^2 \left\| \mathbf{h}^{(i)}(n;l) \right\|^2} \quad (3.83)$$

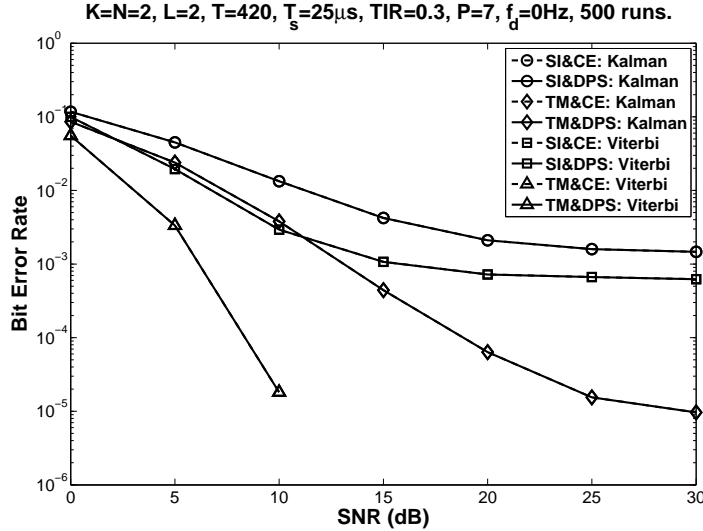


Figure 3.9: First-order statistics-based estimator (MIMO): BER vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 2$. The curves for CE- and DPS-BEM's completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; Kalman: a Kalman filter as the symbol detector; Viterbi: a Viterbi detector as the symbol detector.)

where M_c denotes the number of Monte Carlo runs, $\mathbf{h}^{(i)}(n;l)$ denotes the i -th realization of the time-varying channel, and $\hat{\mathbf{h}}^{(i)}(n;l)$ denotes the acquired channel estimate.

It is seen from the figures that the DPS-BEM-based estimators, no matter whether superimposed or TM training is used, outperform the CE- and OP-BEM-based solutions. (For $f_d = 0$ and $Q = 1$, all the three models gives the same results, since the three models all use constants as the basis functions. The performances of superimposed and TM training, however, are different.) It is consistent with the fact that DPS-BEM can efficiently remove spectral leakage, and it is a much better model in describing a band-limited channel. Due to severe spectral leakage, the CE-BEM-based results are often the worst among the three.

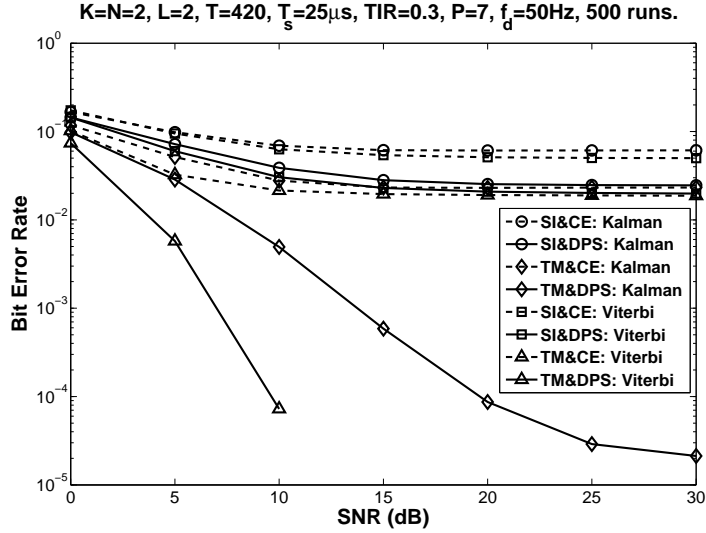


Figure 3.10: First-order statistics-based estimator (MIMO): BER vs SNR under $f_d = 50$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; Kalman: a Kalman filter as the symbol detector; Viterbi: a Viterbi detector as the symbol detector.)

Comparing superimposed training with TM training, we see that the superimposed training-based estimators performs worse than their TM counterparts, although superimposed training can offer higher data transmission rate. As we have discussed in Chapter 1, the major issue of superimposed training is the information-induced interference (self-interference), which results in a notable error floor at BER and NCMSE curves. We will discuss the issue of self-interference in Chapters 4–6.

3.6.2 First-Order Statistics-Based Estimator: Multiple Users

In this example, we follow the conditions addressed in Section 3.6.1 except that a multiple-user scenario is considered.

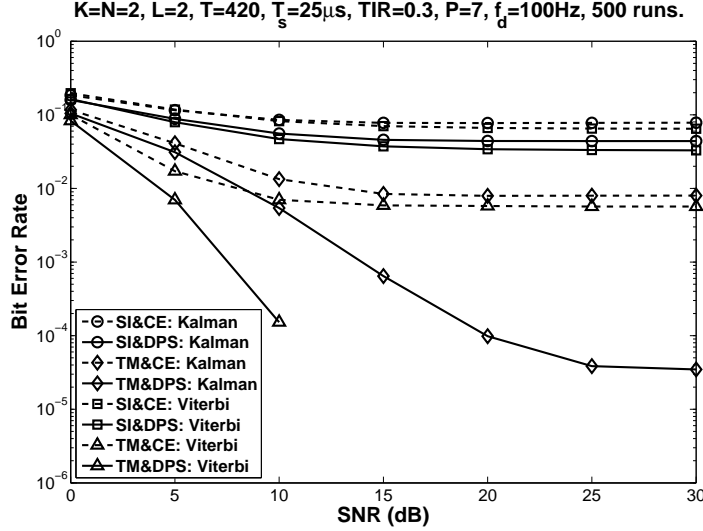


Figure 3.11: First-order statistics-based estimator (MIMO): BER vs SNR under $f_d = 100$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; Kalman: a Kalman filter as the symbol detector; Viterbi: a Viterbi detector as the symbol detector.)

In simulations, all the users have the same transmitted power in training and information data. The average transmitted power in $\{c_k(n)\}$ is 0.3 of that in $\{b_k(n)\}$ ($k = 1, 2, \dots, K$), leading to the same TIR as in Section 3.6.1. We consider a simple two-user scenario, i.e., $K = 2$, with two receive antennas, i.e., $N = 2$. The information sequences $\{b_k(n)\}$ and the training sequences $\{c_k(n)\}$ are all BPSK modulated. The training sequence is generated from the m -sequence of period $\tilde{P} = 7$ by the procedure we introduced in Section 3.5. The training sequences are of length $P = 14$, and the training sequence for the first user is

$$\{c_1(n)\}_{n=0}^{13} = \{1, -1, -1, 1, 1, 1, -1, 1, -1, -1, 1, 1, 1, -1\}, \quad (3.84)$$

the repetitions of the m -sequence of period $\tilde{P} = 7$.

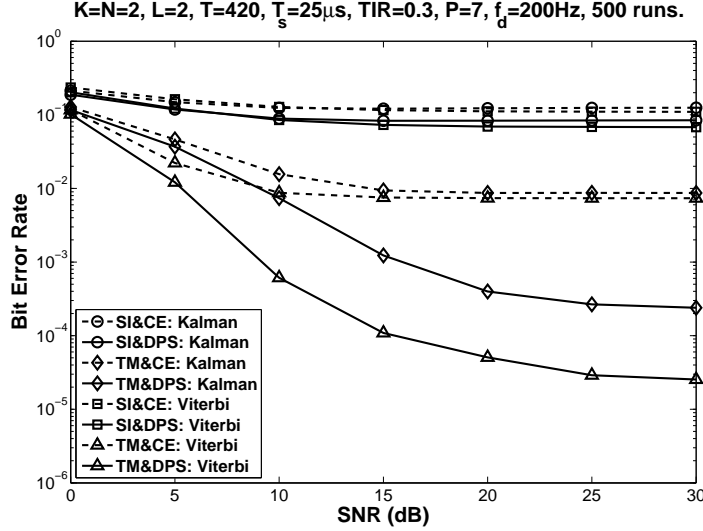


Figure 3.12: First-order statistics-based estimator (MIMO): BER vs SNR under $f_d = 200$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; Kalman: a Kalman filter as the symbol detector; Viterbi: a Viterbi detector as the symbol detector.)

The additive noise $\{\mathbf{v}(n)\}$ is also zero-mean, white complex-Gaussian, uncorrelated with $\{b_k(n)\}$ with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_2 \delta(\tau)$. The (receiver) SNR refers to the energy per bit per user over one-sided noise spectral density with both information and superimposed training sequence counting toward the bit energy.

At the receiver end, a Viterbi detector or a Kalman filter (see Appendix B.2) acts as the symbol detector. We consider different Doppler spreads of $f_d = 0, 50, 100,$ and 200 Hz for this communications system. We also pick Q for CE-BEM as $1, 3, 5, 7$ by (2.9b) and DPS-BEM as $1, 3, 4, 6$ by (2.18).

The results for a record length of $T = 420$ symbols are shown in Figures 3.9–3.16 for various Doppler spreads and SNR's. The results are based on 500 Monte Carlo runs

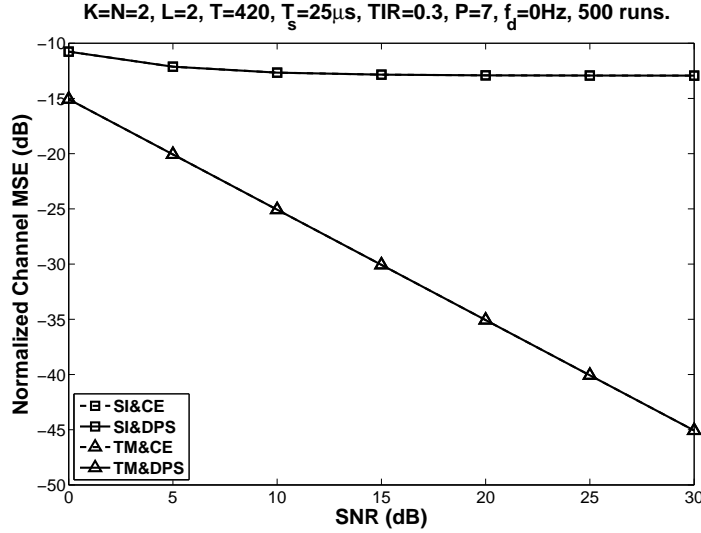


Figure 3.13: First-order statistics-based estimator (MIMO): NCMSE vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 2$. The curves for CE- and DPS-BEM's completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM.)

for Viterbi detectors, and 1000 runs for Kalman filters. For comparison, CE-BEM and DPS-BEM-based periodically placed TM training (see Appendix A) is also considered for doubly-selective channel estimation. We take a training session of length of $(K + 1)L + K = 8$ symbols with the first user's training sequence $\{0, 0, \sqrt{(K + 1)L + K}, 0, 0, 0, 0, 0\}$ and the second user's $\{0, 0, 0, 0, 0, \sqrt{(K + 1)L + K}, 0, 0\}$. An information data session of 27 symbols is inserted between two such training sessions to form a frame of length 35 symbols. Such a frame is repeated over a record length of 420 symbols. Thus, we have a training-to-information bit and power ratio of about 0.3. For multiple-user communications, the

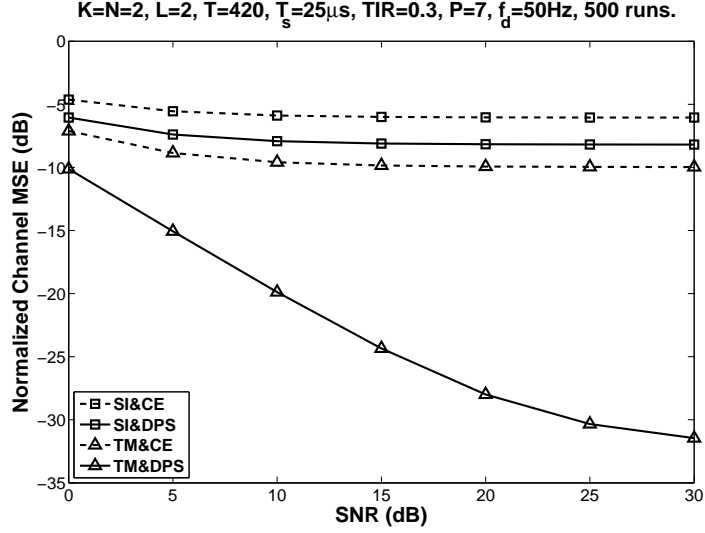


Figure 3.14: First-order statistics-based estimator (MIMO): NCMSE vs SNR under $f_d = 50$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM.)

NCMSE is defined as

$$\text{NCMSE} := \frac{\sum_{k=1}^K \sum_{i=1}^{M_c} \sum_{n=0}^{T-1} \sum_{l=0}^2 \left\| \mathbf{h}_k^{(i)}(n; l) - \hat{\mathbf{h}}_k^{(i)}(n; l) \right\|^2}{\sum_{k=1}^K \sum_{i=1}^{M_c} \sum_{n=0}^{T-1} \sum_{l=0}^2 \left\| \mathbf{h}_k^{(i)}(n; l) \right\|^2} \quad (3.85)$$

We plot the curves for Viterbi detectors and Kalman filters in each figure. The discussion in Section 3.6.1 for the SISO channel applies to the MIMO channel also: DPS-BEM performs best and CE-BEM is the worst; TM training outperforms its superimposed rival. The optimal Viterbi detector shows its advantage in error probability over the Kalman filter, at the expense of increased computational complexity.

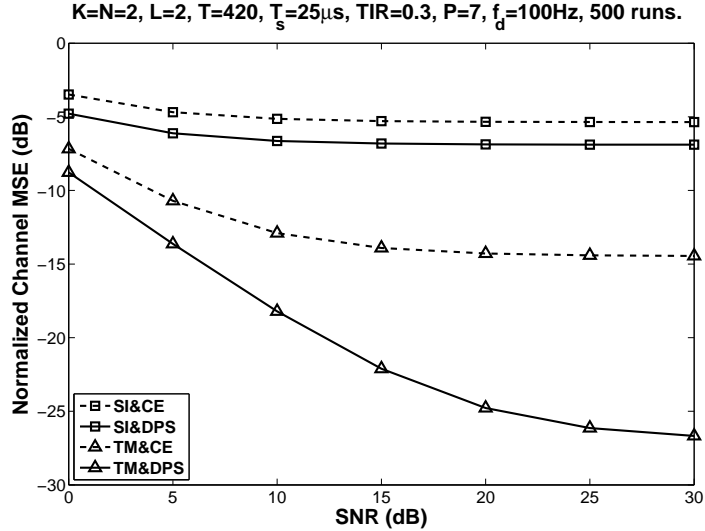


Figure 3.15: First-order statistics-based estimator (MIMO): NCMSE vs SNR under $f_d = 100$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM.)

3.7 Conclusions

In this chapter, we discussed a first-order statistics-based estimator of doubly-selective channels using superimposed training and BEM's. Our starting point was the CE-BEM-based estimator proposed by [81]. Due to the spectral leakage of CE-BEM, this estimator does not perform well in estimating a band-limited channel. We thus extended the estimator to using DPS- and OP-BEM's to reduce the modeling error. We further considered this estimator in a multiple-user scenario. By assigning distinct cycle frequencies of the periodic training sequences to distinct users, channel estimation across various users is decoupled so that the single-user estimator can be used. Our schemes are illustrated by simulation examples, and compared with the conventional TM training: Although higher transmission rate has been achieved by superimposed training, the performance of the proposed estimator

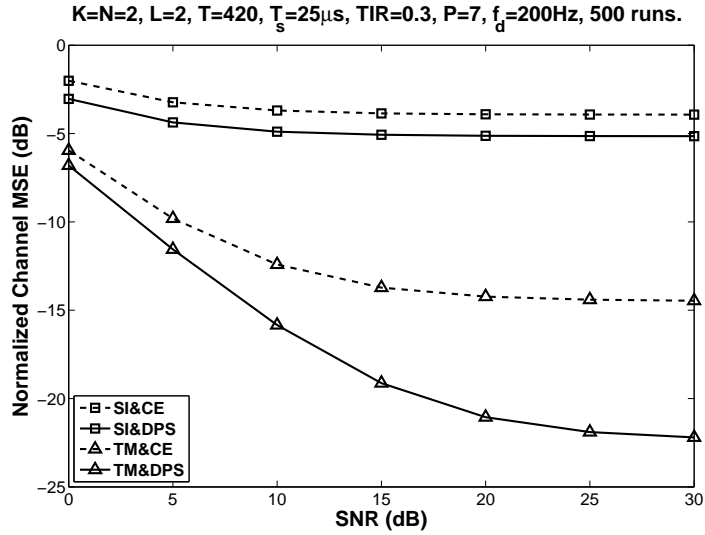


Figure 3.16: First-order statistics-based estimator (MIMO): NCMSE vs SNR under $f_d = 200\text{Hz}$ and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM.)

using superimposed training is inferior to that of TM training, due to the existence of information-induced self-interference. Analyzing and reducing self-interference will be the topic of the following three chapters.

CHAPTER 4

PERFORMANCE ANALYSIS AND PARAMETER DESIGN FOR FIRST-ORDER

STATISTICS-BASED ESTIMATOR

4.1 Introduction

A first-order statistics-based channel estimator, using superimposed training and various BEM's, was discussed in Chapter 3. We present performance analysis of this estimator in this chapter.

Several parameters may affect the performance of the estimator. For example, a portion of transmitted power is allocated to the superimposed training, and this portion will apparently affect the estimator's behavior—more training power leads to higher estimation accuracy, but suppression of information power worsens the effective information SNR in the meantime. Therefore, a trade-off must be made to achieve a balance for the power allocation between training and information.

A similar consideration lies in selecting the number of basis functions for BEM's. More basis functions yield more accurate approximation (or less bias, see Section 2.6), and higher estimation variance. A trade-off between bias and variance should also be studied to achieve better estimation.

In Section 4.2, assuming that the “true” channel follows a BEM, performance analysis for the estimators using CE-, DPS-, and OP-BEM's is explored; performance analysis for the estimator of multiple-user channels is also discussed in this section. In Section 4.3, modeling error of a BEM is counted in performance analysis. Based on the results of performance analysis, we cast the issues of power allocation (Section 4.4) and bias-variance

trade-off (Section 4.5) as ones of optimizing an SNR for equalizer design. Simulation results are provided in Section 4.6. Section 4.7 concludes this chapter.

4.2 Performance Analysis for the First-Order Statistics-Based Estimator Using BEM

We assume the following.

(H4.2.1) The SIMO channel satisfies a BEM representation as in (2.20), i.e.,

$$\mathbf{h}(n; l) = \mathbf{h}_{\text{BEM}}(n; l) = \sum_{q=1}^Q \mathbf{h}_q(l) \psi_q(n), \quad (4.1)$$

where $\psi_q(n)$ is the q -th basis function (corresponding to $e^{j\omega_q n}$ in CE-BEM (2.9), $\varphi_{q-1}(n)$ in OP-BEM (2.15), and $u_q(n)$ in DPS-BEM (2.19)), and Q is the number of basis functions (note that $Q = K + 1$ for K in (2.15)). Also $N \geq 1$.

(H4.2.2) The information sequence $\{b(n)\}$ is zero-mean, white with $E\{|b(n)|^2\} = \sigma_b^2$.

(H4.2.3) The measurement noise $\{\mathbf{v}(n)\}$ is zero-mean, white, uncorrelated with $\{b(n)\}$, with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$.

(H4.2.4) The superimposed training sequence $c(n) = c(n + P)$ for all n is a non-random periodic sequence with period P , with average power $\sigma_c^2 := \sum_{n=0}^{P-1} |c(n)|^2 / P$.

(H4.2.5) The time-varying channel $\{\mathbf{h}(n; l)\}$ is zero-mean, complex Gaussian with variance σ_h^2 , and mutually independent for distinct l 's: $E\{\mathbf{h}(n; l) \mathbf{h}^H(n; l)\} = \sigma_h^2 \mathbf{I}_N$ and $E\{\mathbf{h}(n_1; l_1) \mathbf{h}^H(n_2; l_2)\} = \mathbf{0}$, for $l_1 \neq l_2$, for all n_1, n_2 , i.e., different channel taps are

independent of each other and are identically distributed zero-mean complex Gaussian.

We wish to evaluate the MSE of channel estimation when the true channel follows (4.1). The MSE of estimation is defined as

$$\text{MSE}_1 = \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \left\| \mathbf{h}_{\text{BEM}}(n; l) - \hat{\mathbf{h}}_{\text{BEM}}(n; l) \right\|^2 \right\} \quad (4.2)$$

where $\hat{\mathbf{h}}_{\text{BEM}}(n; l)$ is also given by the BEM

$$\hat{\mathbf{h}}_{\text{BEM}}(n; l) = \sum_{q=1}^Q \hat{\mathbf{h}}_q(l) \psi_q(n). \quad (4.3)$$

By **(H4.2.4)**, we define the normalized training sequence as

$$\bar{c}(n) = c(n) / \sigma_c, \quad \text{for all } n$$

and

$$\bar{c}_m := c_m / \sigma_c, \quad m = 0, 1, \dots, P-1$$

where

$$\begin{aligned} c_m &:= \frac{1}{P} \sum_{n=0}^{P-1} c(n) e^{-j\alpha_m n}, \\ c(n) &= \sum_{m=0}^{P-1} c_m e^{j\alpha_m n}, \quad \text{for all } n, \\ \alpha_m &= 2\pi m / P, \quad m = 0, 1, \dots, P-1. \end{aligned}$$

4.2.1 Performance Analysis for CE-BEM-Based Estimator

Since we have assumed the measurement noise $\{\mathbf{v}(n)\}$ is zero-mean, the CE-BEM based channel estimate is acquired through (3.24), and

$$\hat{\mathbf{h}}_{\text{BEM}}(n; l) = \sum_{q=1}^Q \hat{\mathbf{h}}_q(l) e^{j\omega_q n}, \quad l = 0, 1, \dots, L, \quad 0 \leq n \leq T-1.$$

Let

$$\mathcal{E}_m(n) = \begin{bmatrix} e^{-j(\omega_1 + \alpha_m)n} & e^{-j(\omega_2 + \alpha_m)n} & \dots & e^{-j(\omega_Q + \alpha_m)n} \end{bmatrix}^T,$$

and

$$\mathcal{E}(n) = \begin{bmatrix} \mathcal{E}_0^H(n) & \mathcal{E}_1^H(n) & \dots & \mathcal{E}_{P-1}^H(n) \end{bmatrix}^H. \quad (4.4)$$

By (3.6),

$$\hat{\mathcal{D}} = \frac{1}{T} \sum_{n=0}^{T-1} \mathcal{E}(n) \otimes \mathbf{y}(n),$$

and by (3.24), we have

$$\hat{\mathcal{H}} = \frac{1}{T} (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \tilde{\mathcal{C}}^H \sum_{n=0}^{T-1} \mathcal{E}(n) \otimes \mathbf{y}(n). \quad (4.5)$$

We define

$$\tilde{\mathbf{x}}(n) := \mathbf{y}(n) - E\{\mathbf{y}(n) | \mathcal{H}\} = \sum_{l=0}^L \mathbf{h}(n; l) b(n-l) + \mathbf{v}(n), \quad (4.6)$$

then

$$\begin{aligned} & E \{ \tilde{\mathbf{x}}(n_1) \tilde{\mathbf{x}}^H(n_2) | \mathcal{H} \} \\ &= \sum_{l_1=0}^L \sum_{l_2=0}^L \mathbf{h}(n_1; l_1) \mathbf{h}^H(n_2; l_2) \sigma_b^2 \delta(n_1 - n_2 - l_1 + l_2) + \sigma_v^2 \mathbf{I}_N \delta(n_1 - n_2). \end{aligned}$$

Using **(H4.2.5)**,

$$E_{\mathcal{H}} \{ E \{ \tilde{\mathbf{x}}(n_1) \tilde{\mathbf{x}}^H(n_2) | \mathcal{H} \} \} = [(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2] \mathbf{I}_N \delta(n_1 - n_2).$$

Since $\sum_{n=0}^{T-1} \mathcal{E}(n) \mathcal{E}^H(n) = T \mathbf{I}_{PQ}$, by defining $\text{cov}\{\hat{\mathcal{H}}, \hat{\mathcal{H}} | \mathcal{H}\} := E\{[\hat{\mathcal{H}} - \mathcal{H}][\hat{\mathcal{H}} - \mathcal{H}]^H\}$ and using (4.5) we have (see also (3.21)–(3.23))

$$\begin{aligned} & E_{\mathcal{H}} \left\{ \text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} | \mathcal{H} \right\} \right\} \\ &= \frac{1}{T^2} (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \tilde{\mathcal{C}}^H \text{cov} \left\{ \sum_{n=0}^{T-1} \mathcal{E}(n) \otimes \mathbf{y}(n), \sum_{n=0}^{T-1} \mathcal{E}(n) \otimes \mathbf{y}(n) \right\} \tilde{\mathcal{C}}^H (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \\ &= \frac{1}{T^2} (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \tilde{\mathcal{C}}^H \left[\sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} \mathcal{E}(n_1) \mathcal{E}^H(n_2) \right] \otimes \text{cov} \{ \mathbf{y}(n), \mathbf{y}(n) \} \tilde{\mathcal{C}}^H (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \\ &= \frac{(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2}{T^2} (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \tilde{\mathcal{C}}^H \left[\sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} \mathcal{E}(n_1) \mathcal{E}^H(n_2) \right] \otimes \mathbf{I}_N \delta(n_1 - n_2) \tilde{\mathcal{C}}^H (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \\ &= \frac{(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2}{T} (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \tilde{\mathcal{C}}^H \mathbf{I}_{PQ} \otimes \mathbf{I}_N \tilde{\mathcal{C}}^H (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \\ &= \frac{(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2}{T} (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \\ &= \frac{(L+1) \sigma^2 \sigma_b^2 + \sigma_v^2}{T} \left(\tilde{\mathbf{V}}^H \text{diag} \{ |c_0|^2, |c_1|^2, \dots, |c_{P-1}|^2 \} \tilde{\mathbf{V}} \right)^{-1} \otimes \mathbf{I}_{NQ}. \end{aligned}$$

Since $\hat{\mathcal{D}} = \tilde{\mathcal{C}}\hat{\mathcal{H}}$, then $E\{\hat{\mathcal{D}}\} = \tilde{\mathcal{C}}E\{\hat{\mathcal{H}}\}$. By (3.9), it follows that

$$\begin{aligned}
E\{\hat{\mathbf{d}}_{mq}\} &= \frac{1}{T} \sum_{n=0}^{T-1} E\{\mathbf{y}(n)\} e^{-j(\omega_q + \alpha_m)n} \\
&= \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L \mathbf{h}(n; l) c(n-l) e^{-j(\omega_q + \alpha_m)n} \\
&= \sum_{l=0}^L \sum_{q'=1}^Q \mathbf{h}_{q'}(l) \sum_{m'=0}^{P-1} c_{m'} e^{-j\alpha_{m'}l} \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{-j(\omega_q - \omega_{q'} + \alpha_m - \alpha_{m'})n} \right] \\
&= \sum_{l=0}^L \sum_{q'=1}^Q \mathbf{h}_{q'}(l) \sum_{m'=0}^{P-1} c_{m'} e^{-j\alpha_{m'}l} \delta(m - m') \delta(q - q') \\
&= \sum_{l=0}^L \mathbf{h}_q(l) c_m e^{-j\alpha_m l}.
\end{aligned}$$

Therefore, $E\{\hat{\mathcal{D}}\} = \tilde{\mathcal{C}}\mathcal{H}$, or $\tilde{\mathcal{C}}E\{\hat{\mathcal{H}}\} = \tilde{\mathcal{C}}\mathcal{H}$. Since $\tilde{\mathcal{C}}$ is full column-rank and $P \geq L + 1$, we have $E\{\hat{\mathcal{H}}\} = \mathcal{H}$. Now we evaluate the MSE of the channel estimate given by (4.3):

$$\begin{aligned}
&\text{MSE}_1 \\
&= \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \left(\sum_{q_1=1}^Q [\mathbf{h}_{q_1}(l) - \hat{\mathbf{h}}_{q_1}(l)]^H e^{-j\omega_{q_1}n} \right) \left(\sum_{q_2=1}^Q [\mathbf{h}_{q_2}(l) - \hat{\mathbf{h}}_{q_2}(l)] e^{j\omega_{q_2}n} \right) \right\} \\
&= \sum_{l=0}^L E \left\{ \sum_{q_1=1}^Q \sum_{q_2=1}^Q [\mathbf{h}_{q_1}(l) - \hat{\mathbf{h}}_{q_1}(l)]^H [\mathbf{h}_{q_2}(l) - \hat{\mathbf{h}}_{q_2}(l)] \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{-j(\omega_{q_1} - \omega_{q_2})n} \right] \right\} \\
&= \sum_{l=0}^L E \left\{ \sum_{q_1=1}^Q \sum_{q_2=1}^Q [\mathbf{h}_{q_1}(l) - \hat{\mathbf{h}}_{q_1}(l)]^H [\mathbf{h}_{q_2}(l) - \hat{\mathbf{h}}_{q_2}(l)] \delta(q_1 - q_2) \right\} \\
&= \text{tr} \left\{ E_{\mathcal{H}} \left\{ \text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} \middle| \mathcal{H} \right\} \right\} \right\} \\
&= \frac{[(L+1)\sigma_h^2\sigma_b^2 + \sigma_v^2]NQ}{T\sigma_c^2} \text{tr} \left\{ \left(\tilde{\mathbf{V}}^H \text{diag} \left\{ |\bar{c}_0|^2, |\bar{c}_1|^2, \dots, |\bar{c}_{P-1}|^2 \right\} \tilde{\mathbf{V}} \right)^{-1} \right\}. \quad (4.7)
\end{aligned}$$

Remark 4.2.1.1: If the zero-mean assumption of **(H4.2.1.3)** is relaxed, i.e., we follow the assumption **(H3.2.3)** instead, then the estimator follows (3.18) and the terms corresponding to $m = 0$ are discarded. We omit the entry $\mathcal{E}_0(n)$ in (4.4), and use \mathcal{C} , \mathbf{V} , and $\hat{\mathcal{D}}$ in (3.13) (3.14), and (3.16) instead of $\tilde{\mathcal{C}}$, $\tilde{\mathbf{V}}$, and $\hat{\tilde{\mathcal{D}}}$. We have the MSE of the channel estimation when the mean of the noise is unknown:

$$\text{MSE}_1 = \frac{[(L+1)\sigma_h^2\sigma_b^2 + \sigma_v^2]NQ}{T\sigma_c^2} \text{tr} \left\{ \left(\mathbf{V}^H \text{diag} \left\{ |\bar{c}_1|^2, |\bar{c}_2|^2, \dots, |\bar{c}_{P-1}|^2 \right\} \mathbf{V} \right)^{-1} \right\}. \quad (4.8)$$

Remark 4.2.1.2: If we define an interference factor I_f as

$$I_f = \frac{NQ}{T\sigma_c^2} \text{tr} \left\{ \left(\tilde{\mathbf{V}}^H \text{diag} \left\{ |\bar{c}_0|^2, |\bar{c}_1|^2, \dots, |\bar{c}_{P-1}|^2 \right\} \tilde{\mathbf{V}} \right)^{-1} \right\}$$

or

$$I_f = \frac{NQ}{T\sigma_c^2} \text{tr} \left\{ \left(\mathbf{V}^H \text{diag} \left\{ |\bar{c}_1|^2, |\bar{c}_2|^2, \dots, |\bar{c}_{P-1}|^2 \right\} \mathbf{V} \right)^{-1} \right\},$$

we can clearly see that the MSE of the channel estimation consists of two parts: one is given by $(L+1)\sigma_h^2\sigma_b^2 I_f$ coming from the self-interference, and the other one $\sigma_v^2 I_f$ is the noise-induced part. Normally $(L+1)\sigma_h^2\sigma_b^2 \gg \sigma_v^2$, so that the estimation error mainly comes from the interference from information data.

4.2.2 Performance Analysis for DPS-BEM-Based Estimator

Consider (3.35). From observation $\mathbf{y}(n)$, the estimate $\hat{\mathbf{d}}_{mq}$ has contributions from the information sequence $\{b(n)\}$, which is unknown at the receiver, the superimposed training $\{c(n)\}$, which is known at the receiver, and the measurement noise $\mathbf{v}(n)$. It follows (3.1),

(3.2), (3.4), (3.25), and (3.35) that

$$\begin{aligned}
\hat{\mathbf{d}}_{mq} &= \sum_{n=0}^{T-1} \mathbf{y}(n) u_q(n) e^{-j\alpha_m n} \\
&= \sum_{n=0}^{T-1} \left[\sum_{l=0}^L \mathbf{h}(n;l) c(n-l) + \sum_{l=0}^L \mathbf{h}(n;l) b(n-l) + \mathbf{v}(n) \right] u_q(n) e^{-j\alpha_m n} \\
&= \sum_{n=0}^{T-1} \left[E\{\mathbf{y}(n)\} + \sum_{l=0}^L \mathbf{h}(n;l) b(n-l) + \mathbf{v}(n) \right] u_q(n) e^{-j\alpha_m n}.
\end{aligned}$$

Then by **(H4.2.1)**–**(H4.2.4)**, (3.27), (3.34), and (3.35)

$$\begin{aligned}
E\{\hat{\mathbf{d}}_{mq}\} &= \sum_{n=0}^{T-1} E\{\mathbf{y}(n)\} u_q(n) e^{-j\alpha_m n} \\
&= \sum_{n=0}^{T-1} \sum_{q'=1}^Q \sum_{m'=0}^{P-1} \mathbf{d}_{m'q'} u_{q'}(n) e^{j\alpha_{m'} n} u_q(n) e^{-j\alpha_m n} \\
&= \sum_{q'=1}^Q \sum_{m'=0}^{P-1} \mathbf{d}_{m'q'} \delta(m' - m) \delta(q' - q) \\
&= \mathbf{d}_{mq}.
\end{aligned} \tag{4.9}$$

Define

$$\mathbf{w}_{mq} := \sum_{n=0}^{T-1} \mathbf{v}(n) u_q(n) e^{-j\alpha_m n},$$

which is zero-mean and by **(H4.2.3)** and (3.34)

$$\begin{aligned}
E\{\mathbf{w}_{m'q'} \mathbf{w}_{mq}^H\} &= \sum_{n=0}^{T-1} \sum_{n'=0}^{T-1} E\{\mathbf{v}(n') \mathbf{v}^H(n)\} u_{q'}(n') u_q(n) e^{j(\alpha_m n - \alpha_{m'} n')} \\
&= \sum_{n=0}^{T-1} \sum_{n'=0}^{T-1} \sigma_v^2 \mathbf{I}_N \delta(n' - n) u_{q'}(n') u_q(n) e^{j(\alpha_m n - \alpha_{m'} n')} \\
&= \sigma_v^2 \mathbf{I}_N \delta(m' - m) \delta(q' - q).
\end{aligned} \tag{4.10}$$

Thus,

$$\hat{\mathbf{d}}_{mq} = \mathbf{d}_{mq} + \mathbf{s}_{mq} + \mathbf{w}_{mq}, \quad (4.11)$$

where

$$\mathbf{s}_{mq} := \sum_{n=0}^{T-1} \left[\sum_{l=0}^L \mathbf{h}(n; l) b(n-l) \right] u_q(n) e^{-j\alpha_m n}. \quad (4.12)$$

Clearly, the information sequence's contribution, which is given by \mathbf{s}_{mq} above, interferes with the estimation of \mathbf{d}_{mq} , hence with channel estimation from the observations (see (3.36)).

Since $\tilde{\mathcal{C}}$ is full column-rank when $P \geq L + 1$, by (3.30) and (4.9), we have $E\{\hat{\mathcal{D}}\} = \tilde{\mathcal{D}}$ and $E\{\hat{\mathcal{H}}\} = \mathcal{H}$. Then from (3.36)

$$\text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} \middle| \mathcal{H} \right\} = \left(\tilde{\mathcal{C}}^H \tilde{\mathcal{C}} \right)^{-1} \tilde{\mathcal{C}}^H \text{cov} \{ \hat{\mathcal{D}}, \hat{\mathcal{D}} \} \tilde{\mathcal{C}} \left(\tilde{\mathcal{C}}^H \tilde{\mathcal{C}} \right)^{-1}. \quad (4.13)$$

Consider the zero-mean interference \mathbf{s}_{mq} in (4.12), by **(H4.2.2)**, **(H4.2.5)**, and (3.34)

$$\begin{aligned} E \{ \mathbf{s}_{m'q'} \mathbf{s}_{mq}^H \} &= \sum_{n'=0}^{T-1} \sum_{l'=0}^L \sum_{n=0}^{T-1} \sum_{l=0}^L E \{ \mathbf{h}(n'; l') \mathbf{h}^H(n; l) \} E \{ b(n' - l') b^*(n - l) \} \\ &\quad \times u_q(n) u_{q'}(n') e^{j(\alpha_m n - \alpha_{m'} n')} \\ &= (L + 1) \sigma_h^2 \sigma_b^2 \mathbf{I}_N \sum_{n=0}^{T-1} u_q(n) u_{q'}(n) e^{j(\alpha_m - \alpha_{m'})n} \\ &= (L + 1) \sigma_h^2 \sigma_b^2 \mathbf{I}_N \delta(m' - m) \delta(q' - q). \end{aligned} \quad (4.14)$$

By **(H4.2.3)**, since the noise $\mathbf{v}(n)$ uncorrelated with the zero-mean information sequence $\{b(n)\}$

$$E \{ \mathbf{s}_{m'q'} \mathbf{w}_{mq}^H \} = \mathbf{0}.$$

Then by (4.10), (4.11) and (4.14),

$$\begin{aligned}
E \left\{ [\hat{\mathbf{d}}_{mq} - \mathbf{d}_{mq}] [\hat{\mathbf{d}}_{m'q'} - \mathbf{d}_{m'q'}]^H \right\} &= E \left\{ [\mathbf{s}_{mq} + \mathbf{w}_{mq}] [\mathbf{s}_{m'q'} + \mathbf{w}_{m'q'}]^H \right\} \\
&= E \left\{ \mathbf{s}_{m'q'} \mathbf{s}_{mq}^H \right\} + E \left\{ \mathbf{w}_{m'q'} \mathbf{w}_{mq}^H \right\} \\
&= [(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2] \mathbf{I}_N \delta(m' - m) \delta(q' - q).
\end{aligned}$$

We further have

$$\text{cov}\{\hat{\mathcal{D}}, \hat{\mathcal{D}}\} = [(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2] \mathbf{I}_{NPQ}. \quad (4.15)$$

Substitute (4.15) for (4.13)

$$\text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} \middle| \mathcal{H} \right\} = [(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2] \left(\tilde{\mathcal{C}}^H \tilde{\mathcal{C}} \right)^{-1}.$$

Using the orthonormality of the DPS sequence, the MSE in channel estimation (4.2) is then given by

$$\begin{aligned}
&\text{MSE}_1 \\
&= \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \left(\sum_{q_1=1}^Q [\mathbf{h}_{q_1}(l) - \hat{\mathbf{h}}_{q_1}(l)]^H u_{q_1}(n) \right) \left(\sum_{q_2=1}^Q [\mathbf{h}_{q_2}(l) - \hat{\mathbf{h}}_{q_2}(l)] u_{q_2}(n) \right) \right\} \\
&= \frac{1}{T} \sum_{l=0}^L E \left\{ \sum_{q_1=1}^Q \sum_{q_2=1}^Q [\mathbf{h}_{q_1}(l) - \hat{\mathbf{h}}_{q_1}(l)]^H [\mathbf{h}_{q_2}(l) - \hat{\mathbf{h}}_{q_2}(l)] \delta(q_1 - q_2) \right\} \\
&= \frac{1}{T} \text{tr} \left\{ \text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} \middle| \mathcal{H} \right\} \right\} \\
&= \frac{[(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2] NQ}{T \sigma_c^2} \text{tr} \left\{ \left(\tilde{\mathbf{V}}^H \text{diag} \left\{ |\bar{c}_0|^2, |\bar{c}_1|^2, \dots, |\bar{c}_{P-1}|^2 \right\} \tilde{\mathbf{V}} \right)^{-1} \right\}. \quad (4.16)
\end{aligned}$$

Remark 4.2.2.1: If the mean of noise $\mathbf{v}(n)$ is unknown, we replace $\tilde{\mathcal{C}}$ in (4.16) with \mathcal{C} as in (3.13). Then the MSE in channel estimation is

$$\text{MSE}_1 = \frac{[(L+1)\sigma_h^2\sigma_b^2 + \sigma_v^2]NQ}{T\sigma_c^2} \text{tr} \left\{ \left(\mathbf{V}^H \text{diag} \left\{ |\bar{c}_1|^2, |\bar{c}_2|^2, \dots, |\bar{c}_{P-1}|^2 \right\} \mathbf{V} \right)^{-1} \right\}. \quad (4.17)$$

Whether the noise is zero-mean or not, the interference from the information sequence contributes a majority of the estimation error.

Remark 4.2.2.2: We assume (3.34) which holds precisely for CE-BEM if replacing $u_q(n)$ with $e^{j\omega_q n}$, so that the CE-BEM-based estimator and the DPS-BEM-based estimator give us the same MSE results (compare (4.7) and (4.8) with (4.16) and (4.17)), given the assumption that the “true” channel follows CE- or DPS-BEM respectively. In this section, the modeling error of a BEM in describing a real channel has been omitted.

4.2.3 Performance Analysis for OP-BEM-Based Estimator

Now we turn to the OP-BEM-based channel estimator (3.46). The noise $\mathbf{v}(n)$ is of unknown mean \mathbf{m} . Similar to (4.6) we define

$$\tilde{\mathbf{x}}(n) := \mathbf{y}(n) - E\{\mathbf{y}(n) | \mathcal{H}\} = \sum_{l=0}^L \mathbf{h}(n; l) b(n-l) + \mathbf{v}(n) - \mathbf{m},$$

then

$$\begin{aligned} & E\{\tilde{\mathbf{x}}(n_1) \tilde{\mathbf{x}}^H(n_2) | \mathcal{H}\} \\ &= \sum_{l_1=0}^L \sum_{l_2=0}^L \mathbf{h}(n_1; l_1) \mathbf{h}^H(n_2; l_2) \sigma_b^2 \delta(n_1 - n_2 - l_1 + l_2) + \sigma_v^2 \mathbf{I}_N \delta(n_1 - n_2). \end{aligned}$$

Since $\mathbf{h}(n; l)$'s are independent for different l 's,

$$\begin{aligned}
& E_{\mathcal{H}} \left\{ E \left\{ \tilde{\mathbf{x}}(n_1) \tilde{\mathbf{x}}^H(n_2) \middle| \mathcal{H} \right\} \right\} \\
&= E \left\{ \sum_{l=0}^L \mathbf{h}(n_1; l) \mathbf{h}^H(n_2; l) \right\} \sigma_b^2 \delta(n_1 - n_2) + \sigma_v^2 \mathbf{I}_N \delta(n_1 - n_2) \\
&= [(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2] \mathbf{I}_N \delta(n_1 - n_2). \tag{4.18}
\end{aligned}$$

By (3.46) and (4.18), we have

$$\begin{aligned}
& E_{\mathcal{H}} \left\{ \text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} \middle| \mathcal{H} \right\} \right\} \\
&= \frac{1}{T^2} \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} \mathbf{\Phi}_{\text{OP}}^\dagger \mathbf{\Delta}(n_1) \mathbf{\Delta}^H(n_2) \mathbf{\Phi}_{\text{OP}}^{\dagger H} \otimes E_{\mathcal{H}} \left\{ E \left\{ \tilde{\mathbf{x}}(n_1) \tilde{\mathbf{x}}^H(n_2) \middle| \mathcal{H} \right\} \right\} \\
&= \frac{[(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2]}{T^2} \mathbf{\Phi}_{\text{OP}}^\dagger \left[\sum_{n=0}^{T-1} \mathbf{\Delta}(n) \mathbf{\Delta}^H(n) \right] \mathbf{\Phi}_{\text{OP}}^{\dagger H} \otimes \mathbf{I}_N. \tag{4.19}
\end{aligned}$$

Note that $\frac{1}{T} \sum_{n=0}^{T-1} \mathbf{\Delta}(n) \mathbf{\Delta}^H(n) = \mathbf{\Phi}_{\text{OP}}$ and $\mathbf{\Phi}_{\text{OP}} = \mathbf{\Phi}_{\text{OP}}^H$. We define the normalized $\bar{\mathbf{\Phi}}_{\text{OP}} := \mathbf{\Phi}_{\text{OP}} / \sigma_c^2$, then (4.19) becomes

$$E_{\mathcal{H}} \left\{ \text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} \middle| \mathcal{H} \right\} \right\} = \frac{[(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2]}{T \sigma_c^2} \bar{\mathbf{\Phi}}_{\text{OP}}^\dagger \otimes \mathbf{I}_N.$$

Let

$$\mathbf{\Phi}(n) = \begin{bmatrix} \varphi_0(n) \mathbf{I}_N & \varphi_1(n) \mathbf{I}_N & \cdots & \varphi_K(n) \mathbf{I}_N \end{bmatrix},$$

$$\mathbf{\Psi}(n) = \mathbf{I}_{L+1} \otimes \mathbf{\Phi}(n).$$

By the orthonormality of $\{\varphi_q(n)\}_{n=0}^{T-1}$, $\sum_{n=1}^T \Psi^H(n) \Psi(n) = \mathbf{I}_{N(L+1)(K+1)}$. It follows the OP-BEM (3.38) and (3.47) that

$$\mathbf{h}_{\text{BEM}}(n; l) = \Phi(n) \mathcal{H}_l \text{ and } \hat{\mathbf{h}}_{\text{BEM}}(n; l) = \Phi(n) \hat{\mathcal{H}}_l.$$

The estimation MSE is given by

$$\begin{aligned} \text{MSE}_1 &= \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E_{\mathcal{H}} E \left\{ \left\{ \left| \hat{\mathbf{h}}_{\text{BEM}}(n; l) - \mathbf{h}_{\text{BEM}}(n; l) \right|^2 \middle| \mathcal{H} \right\} \right\} \\ &= \frac{1}{T} \sum_{n=0}^{T-1} \text{tr} \left\{ \Psi(n) E_{\mathcal{H}} \left\{ \text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} \middle| \mathcal{H} \right\} \right\} \Psi^H(n) \right\} \\ &= \frac{(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2}{T^2 \sigma_c^2} \sum_{n=0}^{T-1} \text{tr} \left\{ \Psi(n) \left(\bar{\Phi}_{\text{OP}}^\dagger \otimes \mathbf{I}_N \right) \Psi^H(n) \right\} \\ &= \frac{(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2}{T^2 \sigma_c^2} \sum_{n=0}^{T-1} \text{tr} \left\{ \mathbf{I}_{N(L+1)(K+1)} \left(\bar{\Phi}_{\text{OP}}^\dagger \otimes \mathbf{I}_N \right) \right\} \\ &= \frac{[(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2] N}{T^2 \sigma_c^2} \text{tr} \bar{\Phi}_{\text{OP}}^\dagger. \end{aligned}$$

Remark 4.2.3.1: If the measurement noise is zero-mean, i.e., $\mathbf{m} = \mathbf{0}$, the channel MSE is given by

$$\text{MSE}_1 = \frac{[(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2] N}{T^2 \sigma_c^2} \text{tr} \tilde{\Phi}_{\text{OP}}^\dagger, \quad (4.20)$$

where $\tilde{\Phi}_{\text{OP}} = \bar{\Phi}_{\text{OP}} / \sigma_c^2$ (see **Remark 3.4.3** for the definition of $\bar{\Phi}_{\text{OP}}$ and related discussions).

Remark 4.2.3.2: In **Remark 3.4.2**, we have shown that this estimator can apply to any BEM representation. To reconfirm this, we now consider the channel MSE given by (4.7), (4.16), and (4.20), where we assume $\mathbf{m} = \mathbf{0}$ in all the three cases. The entries of $\tilde{\Phi}_{\text{OP}}$

are

$$\tilde{\phi}(q, q_1, l, l_1) := \frac{1}{T} \sum_{n=0}^{T-1} \varphi_q(n) \varphi_{q_1}^*(n) \bar{c}(n-l) \bar{c}^*(n-l_1).$$

We denote $\tilde{\Phi}_{\text{CE}}$ and $\tilde{\Phi}_{\text{DPS}}$ if we replace $\varphi_q(n)$ with $e^{j\omega_q n}/\sqrt{T}$ and $u_q(n)$ in $\tilde{\Phi}_{\text{OP}}$. Since we have

$$\frac{1}{T} \sum_{n=0}^{T-1} \frac{e^{j\omega_q n}}{\sqrt{T}} \frac{e^{-j\omega_{q_1} n}}{\sqrt{T}} \bar{c}(n-l) \bar{c}^*(n-l_1) = \frac{1}{T} \sum_{m=0}^{P-1} |\bar{c}_m|^2 e^{-j\alpha_m(l-l')} \delta(q - q_1)$$

by (3.49), and

$$\frac{1}{T} \sum_{n=0}^{T-1} u_q(n) u_{q_1}^*(n) \bar{c}(n-l) \bar{c}^*(n-l_1) = \frac{1}{T} \sum_{m=0}^{P-1} |\bar{c}_m|^2 e^{-j\alpha_m(l-l')} \delta(q - q_1)$$

by (3.34), then it follows that

$$\tilde{\Phi}_{\text{DPS}} \approx \tilde{\Phi}_{\text{CE}} = \tilde{\mathbf{V}}^H \text{diag} \left\{ |c_0|^2, |c_1|^2, \dots, |c_{P-1}|^2 \right\} \tilde{\mathbf{V}} \otimes \mathbf{I}_Q.$$

Thus, if we replace $\varphi_q(n)$ with $e^{j\omega_q n}/\sqrt{T}$ or $u_q(n)$ in $\tilde{\Phi}_{\text{OP}}$, (4.20) gives us the same result as (4.7) or (4.16).

4.2.4 Performance Analysis for Multiple-User (MIMO) Channels

We now analyze the estimation performance of the MIMO estimator proposed in Section 3.5.

Due to band-limitedness, the analysis approach used in Section 4.2.2 can also apply to CE-BEM. We hence consider the performance of CE- and DPS-BEM-based MIMO channel estimator in this way.

We assume the following:

(H4.2.4.1) The time-varying channel satisfies a BEM representation, as in (2.20), i.e.,

$$\mathbf{h}_k(n; l) = \mathbf{h}_{\text{BEM}k}(n; l) = \sum_{q=1}^Q \mathbf{h}_{qk}(l) \psi_q(n),$$

where $\psi_q(n)$ is the q -th basis function (corresponding to $e^{j\omega_q n}$ in CE-BEM (2.9) and $u_q(n)$ in DPS-BEM (2.19)), and Q is the number of basis functions. Also $N \geq 1$.

(H4.2.4.2) The information sequence $\{b_k(n)\}$ is zero-mean, white with $E\{|b_k(n)|^2\} = \sigma_{b_k}^2$ and mutually independent for $k = 1, 2, \dots, K$.

(H4.2.4.3) The measurement noise $\{\mathbf{v}(n)\}$ is zero-mean ($\mathbf{m} = \mathbf{0}$), white, uncorrelated with $\{b_k(n)\}$, with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$.

(H4.2.4.4) The superimposed training sequence $c_k(n) = c_k(n + P)$ for all n is a non-random periodic sequence with period P and average power $\sigma_{c_k}^2 := \sum_{n=0}^{P-1} |c_k(n)|^2 / P$ such that $c_{mk} \neq 0$ for all m, k , and \tilde{P} is integer with $P = \tilde{P}K$.

(H4.2.4.5) The time-varying channel $\{\mathbf{h}_k(n; l)\}$ is zero-mean, complex Gaussian with variance σ_{hk}^2 , and mutually independent for distinct l 's: $E\{\mathbf{h}_k(n; l) \mathbf{h}_k^H(n; l)\} = \sigma_{hk}^2 \mathbf{I}_N$ and $E\{\mathbf{h}_k(n_1; l_1) \mathbf{h}_k^H(n_2; l_2)\} = \mathbf{0}$, for $l_1 \neq l_2$, for all n_1, n_2 , i.e., different channel taps are independent of each other and are identically distributed zero-mean complex Gaussian. In addition, $E\{\mathbf{h}_{k'}(n_1; l_1) \mathbf{h}_k^H(n_2; l_2)\} = \mathbf{0}$ for for all n_1, n_2, l_1 , and l_2 if $k' \neq k$, i.e., the channels of different users are mutually independent.

Considering (3.71) and (3.73), we have

$$\hat{\mathbf{d}}_{mqk} = \sum_{n=1}^T \mathbf{y}(n) \psi_q^*(n) e^{-j\alpha_{mk}n}, \quad (4.21)$$

where $\psi_q(n) := e^{j\omega_q n} / \sqrt{T}$ for the CE-BEM and $\psi_q(n) := u_q(n)$ for the DPS-BEM. By (3.70) and (3.72),

$$\sum_{n=0}^{T-1} \psi_{q'}(n) \psi_q^*(n) e^{j(\alpha_{m'k'} - \alpha_{mk})n} \approx \delta(k' - k) \delta(m' - m) \delta(q' - q). \quad (4.22)$$

From $\mathbf{y}(n)$, the estimate $\hat{\mathbf{d}}_{mqk}$ has contributions from the information sequence $\{b_k(n)\}$ ($k = 1, 2, \dots, K$) unknown at the receiver, the superimposed training $\{c_k(n)\}$ known at the receiver, and the measurement noise $\mathbf{v}(n)$. It follows (3.57)–(3.59) that

$$\begin{aligned} \hat{\mathbf{d}}_{mqk} &= \sum_{n=0}^{T-1} \mathbf{y}(n) \psi_q^*(n) e^{-j\alpha_{mk}n} \\ &= \sum_{n=0}^{T-1} \left[E\{\mathbf{y}(n)\} + \sum_{k=1}^K \sum_{l=0}^L \mathbf{h}_k(n; l) b(n-l) + \mathbf{v}(n) \right] \psi_q^*(n) e^{-j\alpha_{mk}n}. \end{aligned}$$

By (H4.2.4.1)–(H4.2.4.5), (3.62) (note that $\mathbf{m} = \mathbf{0}$), (4.21), and (4.22)

$$\begin{aligned} E\{\hat{\mathbf{d}}_{mqk}\} &= \sum_{n=0}^{T-1} E\{\mathbf{y}(n)\} \psi_q^*(n) e^{-j\alpha_{mk}n} \\ &= \sum_{n=0}^{T-1} \sum_{k=1}^K \sum_{m=0}^{\tilde{P}-1} \sum_{q=1}^Q \mathbf{d}_{m'q'k'} \psi_{q'}(n) e^{j\alpha_{m'k'}n} \psi_q^*(n) e^{-j\alpha_{mk}n} \\ &= \mathbf{d}_{mqk}. \end{aligned} \quad (4.23)$$

Define

$$\mathbf{w}_{mqk} := \sum_{n=0}^{T-1} \mathbf{v}(n) \psi_q(n) e^{-j\alpha_{mk}n},$$

which is zero-mean and by **(H4.2.4.3)** and (4.22)

$$\begin{aligned} E \{ \mathbf{w}_{m'q'k'} \mathbf{w}_{mqk}^H \} &= \sum_{n=0}^{T-1} \sum_{n'=0}^{T-1} E \{ \mathbf{v}(n') \mathbf{v}^H(n) \} \psi_{q'}(n') \psi_q^*(n) e^{j(\alpha_{mk}n - \alpha_{m'k'}n')} \\ &= \sigma_v^2 \mathbf{I}_N \delta(m' - m) \delta(q' - q) \delta(k' - k). \end{aligned} \quad (4.24)$$

Thus $\hat{\mathbf{d}}_{mqk} = \mathbf{d}_{mqk} + \mathbf{s}_{mqk} + \mathbf{w}_{mqk}$ where

$$\mathbf{s}_{mqk} := \sum_{n=0}^{T-1} \left[\sum_{k'=1}^K \sum_{l=0}^L \mathbf{h}_{k'}(n; l) b_{k'}(n-l) \right] \psi_q^*(n) e^{-j\alpha_{mk}n}. \quad (4.25)$$

Also, the information sequence's contribution, given by \mathbf{s}_{mqk} above, interferes with the estimation of \mathbf{d}_{mqk} , hence with channel estimation from the observations.

Since $\tilde{\mathcal{C}}_k$ is full column-rank when $\tilde{P} \geq L+1$, by (3.79) and (4.23), we have $E\{\hat{\mathcal{D}}_k\} = \tilde{\mathcal{D}}_k$ and $E\{\hat{\mathcal{H}}_k\} = \mathcal{H}_k$. Then by (3.79)

$$\text{cov} \left\{ \hat{\mathcal{H}}_k, \hat{\mathcal{H}}_k \middle| \mathcal{H}_k \right\} = \left(\tilde{\mathcal{C}}_k^H \tilde{\mathcal{C}}_k \right)^{-1} \tilde{\mathcal{C}}_k^H \text{cov} \{ \hat{\mathcal{D}}_k, \hat{\mathcal{D}}_k \} \tilde{\mathcal{C}}_k \left(\tilde{\mathcal{C}}_k^H \tilde{\mathcal{C}}_k \right)^{-1}. \quad (4.26)$$

Consider the zero-mean interference \mathbf{s}_{mqk} in (4.25), by **(H4.2.4.2)**, **(H4.2.4.5)**, and (4.22)

$$\begin{aligned} &E \{ \mathbf{s}_{m'q'k'} \mathbf{s}_{mqk}^H \} \\ &= \sum_{n'=0}^{T-1} \sum_{l'=0}^L \sum_{n=0}^{T-1} \sum_{l=0}^L \sum_{k'=1}^K \sum_{k=1}^K E \{ \mathbf{h}_{k'}(n'; l') \mathbf{h}_k^H(n; l) \} E \{ b_{k'}(n' - l') b_k^*(n - l) \} \end{aligned}$$

$$\begin{aligned}
& \times \psi_q(n) \psi_{q'}^*(n') e^{j(\alpha_{mk}n - \alpha_{m'k'}n')} \\
& = \sum_{n=0}^{T-1} \sum_{l=0}^L \sum_{k'=1}^K \sum_{k=1}^K E \{ \mathbf{h}_k(n; l) \mathbf{h}_k^H(n; l) \} \sigma_{bk}^2 \delta(k' - k) \psi_q(n) \psi_{q'}^*(n) e^{j(\alpha_{mk}n - \alpha_{m'k'}n')} \\
& = \sum_{k=1}^K (L+1) \sigma_{hk}^2 \sigma_{bk}^2 \mathbf{I}_N \sum_{n=0}^{T-1} \psi_q(n) \psi_{q'}^*(n) e^{j(\alpha_{mk}n - \alpha_{m'k'}n')} \\
& = (L+1) \mathbf{I}_N \left(\sum_{k=1}^K \sigma_{hk}^2 \sigma_{bk}^2 \right) \delta(m' - m) \delta(q' - q) \delta(k' - k). \tag{4.27}
\end{aligned}$$

By **(H4.2.4.3)**, since the noise $\mathbf{v}(n)$ are uncorrelated with $\{b_k(n)\}$, $E \{ \mathbf{s}_{m'q'k'} \mathbf{w}_{mqk}^H \} = \mathbf{0}$.

Then by (4.24) and (4.27),

$$\begin{aligned}
& E \left\{ [\hat{\mathbf{d}}_{mqk} - \mathbf{d}_{mqk}] [\hat{\mathbf{d}}_{m'q'k'} - \mathbf{d}_{m'q'k'}]^H \right\} \\
& = E \{ \mathbf{s}_{m'q'k'} \mathbf{s}_{mqk}^H \} + E \{ \mathbf{w}_{m'q'k'} \mathbf{w}_{mqk}^H \} \\
& = \left[(L+1) \sum_{k=1}^K \sigma_{hk}^2 \sigma_{bk}^2 + \sigma_v^2 \right] \mathbf{I}_N \delta(m' - m) \delta(q' - q) \delta(k' - k).
\end{aligned}$$

We have

$$\text{cov} \{ \hat{\mathcal{D}}_k, \hat{\mathcal{D}}_k \} = \left[(L+1) \sum_{k=1}^K \sigma_{hk}^2 \sigma_{bk}^2 + \sigma_v^2 \right] \mathbf{I}_{N\hat{P}Q}. \tag{4.28}$$

Substitute (4.28) for (4.26)

$$\text{cov} \left\{ \hat{\mathcal{H}}_k, \hat{\mathcal{H}}_k \middle| \mathcal{H}_k \right\} = \left[(L+1) \sum_{k=1}^K \sigma_{hk}^2 \sigma_{bk}^2 + \sigma_v^2 \right] \left(\tilde{\mathcal{C}}_k^H \tilde{\mathcal{C}}_k \right)^{-1}.$$

Using orthonormality of the basis functions, the channel MSE for the k -th user is then given

by

$$\text{MSE}_{1k} = \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \left\| \sum_{q=1}^Q \left[\mathbf{h}_{qk}(l) - \hat{\mathbf{h}}_{qk}(l) \right]^H \psi_{qk}^*(n) \right\|^2 \right\}$$

$$\begin{aligned}
&= \frac{1}{T} \text{tr} \{ \text{cov} \{ \hat{\mathcal{H}}_k, \hat{\mathcal{H}}_k \mid \mathcal{H}_k \} \} \\
&= \left[(L+1) \sum_{k=1}^K \sigma_{hk}^2 \sigma_{bk}^2 + \sigma_v^2 \right] \text{tr} \left\{ \left(\tilde{\mathcal{C}}_k^H \tilde{\mathcal{C}}_k \right)^{-1} \right\} \\
&= \frac{\left[(L+1) \sum_{k=1}^K \sigma_{hk}^2 \sigma_{bk}^2 + \sigma_v^2 \right] NQ}{T\sigma_c^2} \text{tr} \left\{ \left(\tilde{\mathbf{V}}_k^H \text{diag} \left\{ |\bar{c}_0|^2, |\bar{c}_1|^2, \dots, |\bar{c}_{\tilde{P}-1}|^2 \right\} \tilde{\mathbf{V}}_k \right)^{-1} \right\}.
\end{aligned} \tag{4.29}$$

Remark 4.2.4.1: If the mean of $\mathbf{v}(n)$ is unknown, we have

$$\begin{aligned}
&\text{MSE}_{1k} \\
&= \frac{\left[(L+1) \sum_{k=1}^K \sigma_{hk}^2 \sigma_{bk}^2 + \sigma_v^2 \right] NQ}{T\sigma_c^2} \text{tr} \left\{ \left(\mathbf{V}_k^H \text{diag} \left\{ |\bar{c}_1|^2, |\bar{c}_2|^2, \dots, |\bar{c}_{\tilde{P}-1}|^2 \right\} \mathbf{V}_k \right)^{-1} \right\}.
\end{aligned}$$

Remark 4.2.4.2: Compare (4.29) with (4.7) and (4.16), we can see that for the estimation of a multiple-user channel, multiple-user interference (MUI) exists and information data from all users act as interference. The MUI linearly increases as more users join the system.

4.3 Performance Analysis for the First-Order Statistics-Based Estimator: with Modeling Error

In Section 4.2, we assumed that the channel (SIMO or MIMO) follows a BEM representation, i.e., the modeling error was omitted. In practice, modeling error always exists (see Section 2.6). In approximating a band-limited channel, CE-, OP-, and DPS-BEM have

distinct performances. If modeling error is considered, (4.1) is now revised as

$$\mathbf{h}(n; l) = \mathbf{h}_{\text{BEM}}(n; l) + \mathbf{e}_{\text{BEM}}(n; l) = \sum_{q=1}^Q \mathbf{h}_q(l) \psi_q(n) + \mathbf{e}_{\text{BEM}}(n; l)$$

where $\mathbf{e}_{\text{BEM}}(n; l)$ is the modeling error that is intrinsic to the BEM representation and has nothing to do with BEM-based channel estimation. Therefore, the MSE of channel estimation consists of two parts: One comes from the estimation, which has been discussed in Section 4.2; the other part arises from the modeling error.

Consider a “complete” basis matrix

$$\mathbf{\Psi}_T := \begin{bmatrix} \psi_1(0) & \psi_2(0) & \cdots & \psi_T(0) \\ \psi_1(1) & \psi_2(1) & \cdots & \psi_T(1) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(T-1) & \psi_2(T-1) & \cdots & \psi_T(T-1) \end{bmatrix},$$

where $\{\psi_q(n)\}_{q=1}^T$ represents $\left\{e^{j2\pi[q-(T+1)/2]/T}/\sqrt{T}\right\}_{q=1}^T$ in CE-BEM, or modified Legendre polynomials of degree 0 to $T-1$ in OP-BEM, or the total T eigenvectors of the matrix \mathbf{C} defined in (2.17). Note that $\mathbf{\Psi}_T$ is unitary, and for an arbitrary channel $\{\mathbf{h}(n; l)\}_{n=0}^{T-1}$, the following formula is always true:

$$\mathbf{h}(n; l) = \sum_{q=1}^T \mathbf{h}_q(l) \psi_q(n),$$

or equivalently

$$\left[\mathbf{h}^T(0;l) \quad \mathbf{h}^T(1;l) \quad \cdots \quad \mathbf{h}^T(T-1;l) \right]^T = (\mathbf{\Psi}_T \otimes \mathbf{I}_N) \left[\mathbf{h}_1^T(l) \quad \mathbf{h}_2^T(l) \quad \cdots \quad \mathbf{h}_T^T(l) \right]^T, \quad (4.30)$$

where

$$\mathbf{h}_q(l) = \sum_{n=0}^{T-1} \mathbf{h}(n;l) \psi_q^*(n). \quad (4.31)$$

Given the above representations, one may view the BEM representation (2.20) as an approximation where we use only Q (out of total T) basis functions to describe the channel.

In a BEM, (4.30) is approximated by using

$$\begin{aligned} & \left[\mathbf{h}_{\text{BEM}}^T(0;l) \quad \mathbf{h}_{\text{BEM}}^T(1;l) \quad \cdots \quad \mathbf{h}_{\text{BEM}}^T(T-1;l) \right]^T \\ &= (\mathbf{\Psi}_Q \otimes \mathbf{I}_N) \left[\mathbf{h}_1^T(l) \quad \mathbf{h}_2^T(l) \quad \cdots \quad \mathbf{h}_Q^T(l) \right]^T \end{aligned}$$

where $\mathbf{\Psi}_Q$ consists of the first Q columns of $\mathbf{\Psi}_T$. The modeling error is given by

$$\begin{aligned} & \left[\mathbf{e}_{\text{BEM}}^T(0;l) \quad \mathbf{e}_{\text{BEM}}^T(1;l) \quad \cdots \quad \mathbf{e}_{\text{BEM}}^T(T-1;l) \right]^T \\ &= (\mathbf{\Psi}_{T-Q} \otimes \mathbf{I}_N) \left[\mathbf{h}_{Q+1}^T(l) \quad \mathbf{h}_{Q+2}^T(l) \quad \cdots \quad \mathbf{h}_T^T(l) \right]^T, \end{aligned} \quad (4.32)$$

where $\mathbf{\Psi}_{T-Q}$ consists of the last $T - Q$ columns of $\mathbf{\Psi}_T$.

The MSE in channel estimation is now given by

$$\text{MSE}_c = \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \left\| \mathbf{h}(n;l) - \hat{\mathbf{h}}_{\text{BEM}}(n;l) \right\|^2 \right\} \quad (4.33)$$

where $\hat{\mathbf{h}}_{\text{BEM}}(n; l)$ follows (4.3). Since $\mathbf{\Psi}_Q^H \mathbf{\Psi}_{T-Q} = \mathbf{0}$, we have

$$\frac{1}{T} \sum_{n=0}^{T-1} \left[\mathbf{h}_{\text{BEM}}(n; l) - \hat{\mathbf{h}}_{\text{BEM}}(n; l) \right]^H \mathbf{e}_{\text{BEM}}(n; l) = 0,$$

so that

$$\begin{aligned} \text{MSE}_c &= \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \left\| \mathbf{h}_{\text{BEM}}(n; l) - \hat{\mathbf{h}}_{\text{BEM}}(n; l) \right\|^2 \right\} + \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \left\| \mathbf{e}_{\text{BEM}}(n; l) \right\|^2 \right\} \\ &= \text{MSE}_1 + \text{MSE}_2 \end{aligned} \quad (4.34)$$

where we define

$$\text{MSE}_2 := \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \left\| \mathbf{e}_{\text{BEM}}(n; l) \right\|^2 \right\}$$

as the mean square modeling error. It follows by (4.31) and (4.32) that

$$\begin{aligned} \text{MSE}_2 &= \frac{1}{T} \sum_{q=Q+1}^T \sum_{l=0}^L E \left\{ \left\| \mathbf{h}_q(l) \right\|^2 \right\} \\ &= \frac{1}{T} \sum_{q=Q+1}^T \sum_{l=0}^L \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} R_h(n_1; n_2) \psi_q(n_1) \psi_q^*(n_2) \end{aligned}$$

where

$$R_h(n_1, n_2; l) := E \left\{ \mathbf{h}^H(n_1; l) \mathbf{h}(n_2; l) \right\}.$$

For example, we consider the CE-BEM representation for a modified Jakes' channel.

We follow all the other assumptions in Section 4.2 except **(H4.2.1)**. Then

$$R_h(n_1, n_2; l) = N \sigma_h^2 J_0(2\pi f_d T_s (n_1 - n_2))$$

where $J_0(\cdot)$ denotes the zero-th order Bessel function of the first kind. We have

$$\begin{aligned} \text{MSE}_2 &= \frac{L+1}{T^2} \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} N\sigma_h^2 J_0(2\pi f_d T_s (n_1 - n_2)) \left[\sum_{q=\frac{Q+1}{2}}^{T-\frac{Q+1}{2}} e^{j\frac{2\pi q(n_1-n_2)}{T}} \right] \\ &= \frac{L+1}{T} N\sigma_h^2 \left[(T-Q) - 2 \sum_{\tau=1}^{T-1} \left(1 - \frac{\tau}{T}\right) J_0(2\pi f_d T_s \tau) \frac{\sin \frac{\pi \tau Q}{T}}{\sin \frac{\pi \tau}{T}} \right]. \end{aligned} \quad (4.35)$$

4.4 Training Power Allocation

We address the issue of superimposed training power allocation in this section, i.e., we seek the optimal power assignment to training and information under a fixed transmitted power budget. To this end, the more general OP-BEM-based estimator in Section 3.4 is considered that can easily apply to other BEM representations. For our convenience, we assume **(H4.2.1)** holds, i.e., the channel satisfies a BEM so that modeling error is omitted.

We define the training power overhead β as

$$\beta := \frac{\frac{1}{P} \sum_{n=1}^P |c(n)|^2}{\frac{1}{P} \sum_{n=1}^P E \left\{ |s(n)|^2 \right\}} = \frac{\sigma_c^2}{\sigma_b^2 + \sigma_c^2}. \quad (4.36)$$

For a fixed SNR or transmitted power budget, higher β implies smaller effective SNR at the receiver, due to decreased power in the information sequence, but higher channel estimation accuracy. Removing the estimated time-varying mean from the received data, define

$$\tilde{\mathbf{y}}(n) := \mathbf{y}(n) - \sum_{l=0}^L \hat{\mathbf{h}}_{\text{BEM}}(n; l) c(n-l) - \hat{\mathbf{m}}.$$

Assuming that $\hat{\mathbf{m}} = \mathbf{m}$, we have

$$\begin{aligned} \tilde{\mathbf{y}}(n) &\approx \sum_{l=0}^L \hat{\mathbf{h}}_{\text{BEM}}(n; l) b(n-l) \\ &+ \sum_{l=0}^L \left[\mathbf{h}_{\text{BEM}}(n; l) - \hat{\mathbf{h}}_{\text{BEM}}(n; l) \right] [b(n-l) + c(n-l)] + \tilde{\mathbf{v}}(n). \end{aligned} \quad (4.37)$$

In (4.37), define the effective signal as

$$\mathbf{x}_s(n) := \sum_{l=0}^L \hat{\mathbf{h}}_{\text{BEM}}(n; l) b(n-l), \quad (4.38)$$

and the effective noise as

$$\mathbf{w}(n) := \sum_{l=0}^L \left[\mathbf{h}_{\text{BEM}}(n; l) - \hat{\mathbf{h}}_{\text{BEM}}(n; l) \right] [b(n-l) + c(n-l)] + \tilde{\mathbf{v}}(n). \quad (4.39)$$

When using $\hat{\mathbf{h}}_{\text{BEM}}(n; l)$ for equalization or detection, the variance of the effective noise $\mathbf{w}(n)$ contains channel estimation error variance as a component, which in turn depends on β . An ‘‘optimum’’ value of β for the superimposed training method may be obtained by maximizing the SNR in (4.37) with respect to β , which is defined as

$$\text{SNR}_e(\beta, n) = \frac{\sigma_{x_s}^2(n)}{\sigma_w^2(n)} \quad (4.40)$$

under the constraint of a fixed transmitted power, i.e., $\sigma_b^2 + \sigma_c^2 = P_T$. In (4.38), the signal power at time n is given by

$$\begin{aligned}
\sigma_{xs}^2(n) &= E \left\{ \|\mathbf{x}_s(n)\|^2 \right\} = \sigma_b^2 \sum_{l=0}^L E_{\mathcal{H}} \left\{ E \left\{ \left\| \hat{\mathbf{h}}_{\text{BEM}}(n;l) \right\|^2 \middle| \mathcal{H} \right\} \right\} + O(1/T^2) \\
&= \sigma_b^2 \sum_{l=0}^L \left[E_{\mathcal{H}} \left\{ E \left\{ \left\| \mathbf{h}_{\text{BEM}}(n;l) - \hat{\mathbf{h}}_{\text{BEM}}(n;l) \right\|^2 \middle| \mathcal{H} \right\} \right\} + E \left\{ \|\mathbf{h}_{\text{BEM}}(n;l)\|^2 \right\} \right] \\
&\quad + O(1/T^2), \tag{4.41}
\end{aligned}$$

where the $O(1/T^2)$ term accounts for the dependence between $\{\hat{\mathbf{h}}_{\text{BEM}}(n;l)\}$ and $\{b(n)\}$ (see Appendix of [32] for the corresponding details in the time-invariant case). Therefore, the time average of signal power is given by (omitting $O(1/T^2)$ terms)

$$\bar{\sigma}_{xs}^2 := \frac{1}{T} \sum_{n=0}^{T-1} \sigma_{xs}^2(n) = \sigma_b^2 [\text{MSE}_1 + (L+1)\sigma_h^2]. \tag{4.42}$$

The noise power at time n is given by

$$\begin{aligned}
\sigma_w^2(n) &= E \left\{ \|\mathbf{w}(n)\|^2 \right\} = \sigma_b^2 \sum_{l=0}^L E_{\mathcal{H}} \left\{ E \left\{ \left\| \mathbf{h}_{\text{BEM}}(n;l) - \hat{\mathbf{h}}_{\text{BEM}}(n;l) \right\|^2 \middle| \mathcal{H} \right\} \right\} + N\sigma_v^2 \\
&\quad + \sigma_c^2 \sum_{l_1=0}^L \sum_{l_2=0}^L E \left\{ \left[\hat{\mathbf{h}}_{\text{BEM}}(n;l_1) - \mathbf{h}_{\text{BEM}}(n;l_1) \right]^H \left[\hat{\mathbf{h}}_{\text{BEM}}(n;l_2) - \mathbf{h}_{\text{BEM}}(n;l_2) \right] \right\} \\
&\quad \times \bar{c}(n-l_1)^* \bar{c}(n-l_2) + O(1/T^2), \tag{4.43}
\end{aligned}$$

where in a manner similar to (3.41), the $O(1/T^2)$ term accounts for the dependence between $\{\hat{\mathbf{h}}_{\text{BEM}}(n; l)\}$ and $\{b(n)\}$. We define

$$\mathcal{E}(n) := \begin{bmatrix} \bar{c}(n) \mathbf{I}_{N(K+1)} & \bar{c}(n-1) \mathbf{I}_{N(K+1)} & \cdots & \bar{c}(n-L) \mathbf{I}_{N(K+1)} \end{bmatrix}$$

and consider

$$\begin{aligned} & \sum_{l_1=0}^L \sum_{l_2=0}^L E \left\{ \left[\hat{\mathbf{h}}_{\text{BEM}}(n; l_1) - \mathbf{h}_{\text{BEM}}(n; l_1) \right]^H \left[\hat{\mathbf{h}}_{\text{BEM}}(n; l_2) - \mathbf{h}_{\text{BEM}}(n; l_2) \right] \right\} \\ & \times \bar{c}(n-l_1)^* \bar{c}(n-l_2) \\ & = \sum_{l_1=0}^L \sum_{l_2=0}^L E_{\mathcal{H}} \left\{ E \left\{ \Phi(n) \left(\hat{\mathcal{H}}_{l_1} - \mathcal{H}_{l_1} \right) \left(\hat{\mathcal{H}}_{l_2} - \mathcal{H}_{l_2} \right)^H \Phi^H(n) \middle| \mathcal{H} \right\} \right\} \bar{c}(n-l_1) \bar{c}^*(n-l_2) \\ & = \Phi(n) E_{\mathcal{H}} \left\{ E \left\{ \mathcal{E}(n) \left(\hat{\mathcal{H}} - \mathcal{H} \right) \left(\hat{\mathcal{H}} - \mathcal{H} \right)^H \mathcal{E}^H(n) \middle| \mathcal{H} \right\} \right\} \Phi^H(n) \\ & = \frac{[(L+1)\sigma_h^2\sigma_b^2 + \sigma_v^2]}{T\sigma_c^2} \text{tr} \left\{ \left(\bar{\Phi}_{\text{OP}}^\dagger \otimes \mathbf{I}_N \right) \mathcal{E}^H(n) \Phi^H(n) \Phi(n) \mathcal{E}(n) \right\}. \end{aligned}$$

Therefore (4.43) can be written as

$$\begin{aligned} \sigma_w^2(n) & = \sigma_b^2 \sum_{l=0}^L E_{\mathcal{H}} \left\{ E \left\{ \left\| \mathbf{h}_{\text{BEM}}(n; l) - \hat{\mathbf{h}}_{\text{BEM}}(n; l) \right\|^2 \middle| \mathcal{H} \right\} \right\} + N\sigma_v^2 + O(1/T^2) \\ & + \frac{[(L+1)\sigma_h^2\sigma_b^2 + \sigma_v^2]}{T} \text{tr} \left\{ \left(\bar{\Phi}_{\text{OP}}^\dagger \otimes \mathbf{I}_N \right) \mathcal{E}^H(n) \Phi^H(n) \Phi(n) \mathcal{E}(n) \right\}. \end{aligned}$$

Taking time average of the noise power and omitting $O(1/T^2)$ terms, we have

$$\bar{\sigma}_w^2 := \frac{1}{T} \sum_{n=0}^{T-1} \sigma_w^2(n) = \sigma_b^2 \text{MSE}_1 + N\sigma_v^2 + \frac{[(L+1)\sigma_h^2\sigma_b^2 + \sigma_v^2]}{T^2} \text{tr} \left\{ \left(\bar{\Phi}_{\text{OP}}^\dagger \otimes \mathbf{I}_N \right) \mathcal{K} \right\}, \quad (4.44)$$

where

$$\mathcal{K} := \sum_{n=1}^T \mathcal{E}^H(n) \Phi^H(n) \Phi(n) \mathcal{E}(n).$$

We define the time average version of (4.40) as

$$\text{SNR}_d(\beta) = \frac{\bar{\sigma}_{xs}^2}{\bar{\sigma}_w^2}. \quad (4.45)$$

Using the constraint $\sigma_b^2 + \sigma_c^2 = P_T$, we have $\sigma_c^2 = \beta P_T$ and $\sigma_b^2 = (1 - \beta) P_T$. Incorporating these constraint-carrying variables in (4.42), (4.44), and (4.45), we have an unconstrained cost

$$\text{SNR}_d(\beta) = \frac{f_1 \beta^2 + f_2 \beta + f_3}{g_1 \beta^2 + g_2 \beta + g_3},$$

where

$$\begin{aligned} f_1 &= (L+1) \sigma_h^2 \left(N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger - T^2 \right), \\ f_2 &= -(L+1) \sigma_h^2 \left(2N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger - T^2 \right) - \frac{\sigma_v^2 N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger}{P_T}, \\ f_3 &= (L+1) \sigma_h^2 N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger + \frac{N \sigma_v^2 \text{tr} \bar{\Phi}_{\text{OP}}^\dagger}{P_T}, \\ g_1 &= (L+1) \sigma_h^2 \left(N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger - \text{tr} \left\{ \left(\bar{\Phi}_{\text{OP}}^\dagger \otimes \mathbf{I}_N \right) \mathcal{K} \right\} \right), \\ g_2 &= -(L+1) \sigma_h^2 \left(2N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger - \text{tr} \left\{ \left(\bar{\Phi}_{\text{OP}}^\dagger \otimes \mathbf{I}_N \right) \mathcal{K} \right\} \right) \\ &\quad + \frac{\sigma_v^2 \left(\text{tr} \left\{ \left(\bar{\Phi}_{\text{OP}}^\dagger \otimes \mathbf{I}_N \right) \mathcal{K} \right\} - N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger + T^2 \right)}{P_T}, \\ g_3 &= f_3. \end{aligned}$$

We seek the optimum value of β by taking the derivative

$$\frac{d[\text{SNR}_d(\beta)]}{d\beta} = \frac{(f_1g_2 - f_2g_1)\beta^2 + 2(f_1g_3 - f_3g_1)\beta + f_2g_3 - f_3g_2}{(g_1\beta^2 + g_2\beta + g_3)^2} = 0,$$

the root of which lying in $[0, 1]$ is

$$\beta_{\text{opt}} = (f_1g_2 - f_2g_1)^{-1} (f_3g_1 - f_1g_3 - \sqrt{-f_1f_2g_2g_3 - 2f_1f_3g_1g_3 - f_2f_3g_1g_2 + f_2^2g_1g_3 + f_1f_3g_2^2 + f_1^2g_3^2 + f_3^2g_1^2}) \quad (4.46)$$

Since $\mathbf{h}(n; l)$'s are mutually independent for different l 's, the calculation can be simplified if we suppose $\hat{\mathbf{h}}_{\text{BEM}}(n; l)$'s are also approximately uncorrelated for distinct l 's, i.e.,

$$E \left\{ \left[\hat{\mathbf{h}}_{\text{BEM}}(n; l_1) - \mathbf{h}(n; l_1) \right]^H \left[\hat{\mathbf{h}}_{\text{BEM}}(n; l_2) - \mathbf{h}(n; l_2) \right] \right\} \approx 0, \text{ for } l_1 \neq l_2.$$

Then (4.43) becomes (omitting $O(1/T^2)$ terms)

$$\begin{aligned} \sigma_w^2(n) &\approx \sigma_b^2 \sum_{l=0}^L E_{\mathcal{H}} \left\{ E \left\{ \left\| \hat{\mathbf{h}}_{\text{BEM}}(n; l) \right\|^2 \middle| \mathcal{H} \right\} \right\} + N\sigma_v^2 \\ &+ \sigma_c^2 \sum_{l=0}^L E \left\{ \left\| \hat{\mathbf{h}}_{\text{BEM}}(n; l) - \mathbf{h}_{\text{BEM}}(n; l) \right\|^2 \right\} \bar{c}^*(n-l) \bar{c}(n-l). \end{aligned} \quad (4.47)$$

In addition, if $\bar{c}^*(n) \bar{c}(n)$ keeps constant for all n (e.g., $\bar{c}(n) = \pm 1$), (4.47) can be further reduced to

$$\sigma_w^2(n) \approx \sum_{l=0}^L E_{\mathcal{H}} \left\{ E \left\{ \left\| \hat{\mathbf{h}}_{\text{BEM}}(n; l) \right\|^2 \middle| \mathcal{H} \right\} \right\} (\sigma_b^2 + \sigma_c^2) + \sigma_v^2.$$

Then

$$\bar{\sigma}_{x_s}^2 = \sigma_b^2 [\text{MSE}_1 + (L+1)\sigma_h^2] \text{ and } \bar{\sigma}_w^2 = \text{MSE}_1 (\sigma_b^2 + \sigma_c^2) + \sigma_v^2,$$

and

$$\text{SNR}_d(\beta) = \frac{\bar{\sigma}_{xs}^2}{\bar{\sigma}_w^2} = \frac{f_1\beta^2 + f_2\beta + f_3}{g'_1\beta + g'_2}$$

where

$$\begin{aligned} f_1 &= (L+1)\sigma^2 \left(N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger - T^2 \right), \\ f_2 &= -(L+1)\sigma^2 \left(2N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger - T^2 \right) - \frac{N\sigma_v^2 \text{tr} \bar{\Phi}_{\text{OP}}^\dagger}{P_T}, \\ f_3 &= (L+1)\sigma^2 N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger + \frac{N\sigma_v^2 \text{tr} \bar{\Phi}_{\text{OP}}^\dagger}{P_T}, \\ g'_1 &= \frac{\sigma_v^2 T^2}{P_T} - (L+1)\sigma^2 N \text{tr} \bar{\Phi}_{\text{OP}}^\dagger, \\ g'_2 &= f_3. \end{aligned}$$

Setting the first derivative of $\text{SNR}_d(\beta)$ to be zero to get the optimum β , we have

$$\beta'_{\text{opt}} = \frac{g'_2}{g'_1} \left[-1 + \sqrt{1 + \frac{g'_1(f_3g'_1 - f_2g'_2)}{g_2'^2 f_1}} \right]. \quad (4.48)$$

4.5 Bias-Variance Trade-Off

In Section 4.4, we have addressed the issue of training power allocation from an equalization viewpoint by maximizing SNR for data detection. This method can also be applied in selection of the number of basis functions Q , since SNR_d in (4.45) is also a function of Q . In this section, we give the CE-BEM-based analysis as an example to clarify our trade-off. Note that modeling error of a BEM must be considered in this issue.

Consider (4.37)–(4.39), which will be used for equalization with $\mathbf{x}_s(n)$ acting as “signal” and $\mathbf{w}(n)$ as “noise”. Since the modeling error is now considered, by (4.41) the time-average

of $\sigma_{xs}^2(n)$ now becomes (omitting $O(1/T^2)$ terms)

$$\begin{aligned}\bar{\sigma}_{xs}^2 &:= \frac{1}{T} \sum_{n=0}^{T-1} \sigma_{xs}^2(n) = \sigma_b^2 \left[\text{MSE}_1 + \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \|\mathbf{h}_{\text{BEM}}(n; l)\|^2 \right\} \right] \\ &= \sigma_b^2 \left[\text{MSE}_1 + \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \|\mathbf{h}(n; l)\|^2 \right\} - \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \left\{ \|\mathbf{e}_{\text{BEM}}(n; l)\|^2 \right\} \right] \\ &= \sigma_b^2 [\text{MSE}_1 + N(L+1)\sigma_h^2 - \text{MSE}_2].\end{aligned}$$

Next consider the power of the noise (4.43), the time-average of which (omitting $O(1/T^2)$ terms) is given by

$$\bar{\sigma}_w^2 := \frac{1}{T} \sum_{n=0}^{T-1} \sigma_w^2(n) = \sigma_b^2 \text{MSE}_c + N\sigma_v^2 + \mathcal{R}$$

where we define

$$\mathcal{R} := \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l_1=0}^L \sum_{l_2=0}^L E \left\{ \mathbf{e}_2^H(n; l_1) \mathbf{e}_2(n; l_2) \right\} c^*(n-l_1) c(n-l_2)$$

$$\mathbf{e}_1(n; l) := \hat{\mathbf{h}}_{\text{BEM}}(n; l) - \mathbf{h}_{\text{BEM}}(n; l),$$

$$\mathbf{e}_2(n; l) := \mathbf{e}_1(n; l) - \mathbf{e}_{\text{BEM}}(n; l).$$

It turns out that $\mathcal{R} = \mathcal{R}_1 + \mathcal{R}_2$, where

$$\mathcal{R}_1 := \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l_1=0}^L \sum_{l_2=0}^L E \left\{ \mathbf{e}_1^H(n; l_1) \mathbf{e}_1(n; l_2) \right\} c^*(n-l_1) c(n-l_2),$$

$$\mathcal{R}_2 := \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l_1=0}^L \sum_{l_2=0}^L E \left\{ \mathbf{e}_{\text{BEM}}^H(n; l_1) \mathbf{e}_{\text{BEM}}(n; l_2) \right\} c^*(n-l_1) c(n-l_2).$$

By **(H4.2.5)**, then $E \{ \mathbf{e}_{\text{BEM}}^H(n; l_1) \mathbf{e}_{\text{BEM}}(n; l_2) \} = 0$ for $l_1 \neq l_2$. Thus

$$\begin{aligned} \mathcal{R}_2 &= \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^L E \{ \mathbf{e}_{\text{BEM}}^H(n; l) \mathbf{e}_{\text{BEM}}(n; l) \} \sum_{m_1=0}^{P-1} \sum_{m_2=0}^{P-1} c_{m_1}^* c_{m_2} e^{j(\alpha_{m_1} - \alpha_{m_2})l} e^{-j(\alpha_{m_1} - \alpha_{m_2})n} \\ &= \sum_{l=0}^L E \left\{ \sum_{q_1=\frac{Q+1}{2}}^{T-\frac{Q+1}{2}} \sum_{q_2=\frac{Q+1}{2}}^{T-\frac{Q+1}{2}} \mathbf{h}_{q_1}^H(l) \mathbf{h}_{q_2}(l) \right\} \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{j(\omega_{q_2} - \omega_{q_1} - \alpha_{m_1} + \alpha_{m_2})n} \right] \\ &\times \sum_{m_1=0}^{P-1} \sum_{m_2=0}^{P-1} c_{m_1}^* c_{m_2} e^{j(\alpha_{m_1} - \alpha_{m_2})l}. \end{aligned}$$

We consider the correlation between $\mathbf{h}_{q_1}(l)$ and $\mathbf{h}_{q_2}(l)$,

$$E \{ \mathbf{h}_{q_1}^H(l) \mathbf{h}_{q_2}(l) \} = \frac{1}{T^2} \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} E \{ \mathbf{h}^H(n_1; l) \mathbf{h}(n_2; l) \} e^{-j\omega_{q_1} n_1} e^{j\omega_{q_2} n_2}.$$

By defining

$$R_h(n_1 - n_2; l) := E \{ \mathbf{h}^H(n_1; l) \mathbf{h}(n_2; l) \}$$

and setting $\tau := n_1 - n_2$, we have

$$E \{ \mathbf{h}_{q_1}^H(l) \mathbf{h}_{q_2}(l) \} = \frac{1}{T} \sum_{\tau=-(T-1)}^{T-1} R_h(\tau; l) e^{-j\omega_{q_1} \tau} \left[\frac{1}{T} \sum_{n_2=\max(0, -\tau)}^{\min(T-1, T-1-\tau)} e^{j(\omega_{q_2} - \omega_{q_1})n_2} \right].$$

Note that

$$\left| \frac{1}{T} \sum_{n_2=\max(0, -\tau)}^{\min(T-1, T-1-\tau)} e^{j(\omega_{q_2} - \omega_{q_1})n_2} - \frac{1}{T} \sum_{n_2=0}^{T-1} e^{j(\omega_{q_2} - \omega_{q_1})n_2} \right| \leq \frac{|\tau|}{T}.$$

Since the coherence time of the channel is limited, we can select a number T_{coh} such that

$|R_h(n_1 - n_2; l)| \approx 0$ for $|\tau| > T_{\text{coh}}$. Therefore,

$$\frac{1}{T} \sum_{n_2=\max(0, -\tau)}^{\min(T-1, T-1-\tau)} e^{j(\omega_{q_2} - \omega_{q_1})n_2} = \delta(q_1 - q_2) + O\left(\frac{1}{T}\right).$$

This fact leads to

$$E \{ \mathbf{h}_{q_1}^H(l) \mathbf{h}_{q_2}(l) \} \approx 0 \quad (4.49)$$

for $q_1 \neq q_2$ and “large” T . Omitting the $O(1/T)$ term, \mathcal{R}_2 can be rewritten as

$$\begin{aligned} \mathcal{R}_2 &= \sum_{l=0}^L E \left\{ \sum_{q_1=\frac{Q+1}{2}}^{T-\frac{Q+1}{2}} \sum_{q_2=\frac{Q+1}{2}}^{T-\frac{Q+1}{2}} \mathbf{h}_{q_1}^H(l) \mathbf{h}_{q_2}(l) \right\} \delta(q_1 - q_2) \delta(m_1 - m_2) \\ &\quad \times \sum_{m_1=0}^{P-1} \sum_{m_2=0}^{P-1} c_{m_1}^* c_{m_2} e^{j(\alpha_{m_1} - \alpha_{m_2})l} \\ &= \sum_{l=0}^L E \left\{ \sum_{q=\frac{Q+1}{2}}^{T-\frac{Q+1}{2}} \mathbf{h}_q^H(l) \mathbf{h}_q(l) \right\} \sum_{m=0}^{P-1} |c_m|^2 = \text{MSE}_2 \sum_{m=0}^{P-1} |c_m|^2. \end{aligned}$$

For the first part of \mathcal{R} , by using (4.49)

$$\begin{aligned} \mathcal{R}_1 &= \sum_{l_1=0}^L \sum_{l_2=0}^L E \left\{ \sum_{q_1=1}^Q \sum_{q_2=1}^Q \left[\hat{\mathbf{h}}_{q_1}^H(l_1) - \mathbf{h}_{q_1}^H(l_1) \right] \left[\hat{\mathbf{h}}_{q_2}(l_2) - \mathbf{h}_{q_2}(l_2) \right] \right\} \\ &\quad \times \sum_{m_1=0}^{P-1} \sum_{m_2=0}^{P-1} c_{m_1}^* c_{m_2} e^{j(\alpha_{m_1} l_1 - \alpha_{m_2} l_2)} \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{j(\omega_{q_2} - \omega_{q_1} - \alpha_{m_1} + \alpha_{m_2})n} \right] \\ &= \sum_{l_1=0}^L \sum_{l_2=0}^L E \left\{ \sum_{q_1=-\frac{Q-1}{2}}^{\frac{Q-1}{2}} \sum_{q_2=-\frac{Q-1}{2}}^{\frac{Q-1}{2}} \left[\hat{\mathbf{h}}_{q_1}^H(l_1) - \mathbf{h}_{q_1}^H(l_1) \right] \left[\hat{\mathbf{h}}_{q_2}(l_2) - \mathbf{h}_{q_2}(l_2) \right] \right\} \end{aligned}$$

Estimation variance: $L=2$, $T=399$, $T_s=25\mu\text{s}$, $TIR=0.3$, $P=7$, $f_d=0\text{Hz}$, 500 runs.

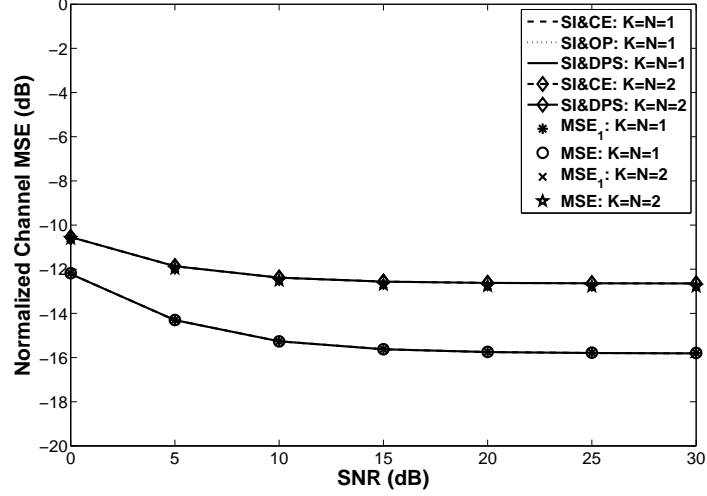


Figure 4.1: Estimation variance: NCMSE vs SNR under $f_d = 0\text{ Hz}$ (time-invariant). The curves for CE-, OP- and DPS-BEM's completely overlap, since the three basis functions are all constant for time-invariant channels ($Q = 1$). (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM; MSE_1 : defined in (4.2); MSE: defined in (4.33).)

$$\begin{aligned}
 & \times \sum_{m_1=0}^{P-1} \sum_{m_2=0}^{P-1} c_{m_1}^* c_{m_2} e^{j(\alpha_{m_1} l_1 - \alpha_{m_2} l_2)} \delta(q_1 - q_2) \delta(m_1 - m_2) \\
 & = E \left\{ \left[\sum_{l=0}^L (\hat{\mathbf{H}}_l - \mathbf{H}_l) e^{-j\alpha_m l} \right]^H \left[\sum_{l=0}^L (\hat{\mathbf{H}}_l - \mathbf{H}_l) e^{-j\alpha_m l} \right] \right\} \sum_{m=0}^{P-1} |c_m|^2.
 \end{aligned}$$

Define

$$\mathcal{B}(m) := \begin{bmatrix} 1 & e^{-j\alpha_m} & \dots & e^{-j\alpha_m L} \end{bmatrix},$$

Estimation variance: $L=2$, $T=399$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=50\text{Hz}$, 500 runs.

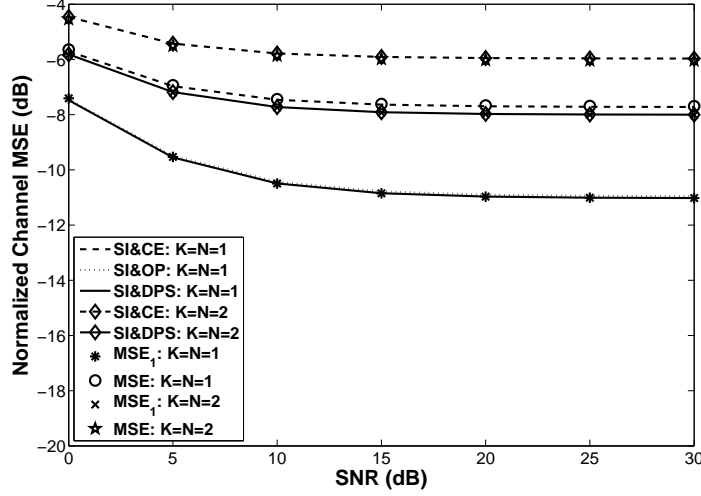


Figure 4.2: Estimation variance: NCMSE vs SNR under $f_d = 50\text{Hz}$. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM; MSE_1 : defined in (4.2); MSE: defined in (4.33).)

then \mathcal{R}_1 becomes

$$\begin{aligned} \mathcal{R}_1 &= E \left\{ \left[\mathcal{B}(m) \otimes \mathbf{I}_{NQ} \left(\hat{\mathcal{H}} - \mathcal{H} \right) \right]^H \left[\mathcal{B}(m) \otimes \mathbf{I}_{NQ} \left(\hat{\mathcal{H}} - \mathcal{H} \right) \right] \right\} \sum_{m=0}^{P-1} |c_m|^2 \\ &= E \left\{ \left(\hat{\mathcal{H}} - \mathcal{H} \right)^H \mathbf{C}^H \mathbf{C} \left(\hat{\mathcal{H}} - \mathcal{H} \right) \right\} = \frac{(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2}{T} (L+1) NQ. \end{aligned}$$

Thus

$$\mathcal{R} = \frac{(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2}{T} (L+1) NQ + \text{MSE}_2 \sigma_c^2,$$

and

$$\bar{\sigma}_w^2 = \sigma_b^2 \text{MSE}_c + N \sigma_v^2 + \frac{(L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2}{T} (L+1) NQ + \text{MSE}_2 \sigma_c^2.$$

Estimation variance: $L=2$, $T=399$, $T_s=25\mu\text{s}$, $TIR=0.3$, $P=7$, $f_d=100\text{Hz}$, 500 runs.

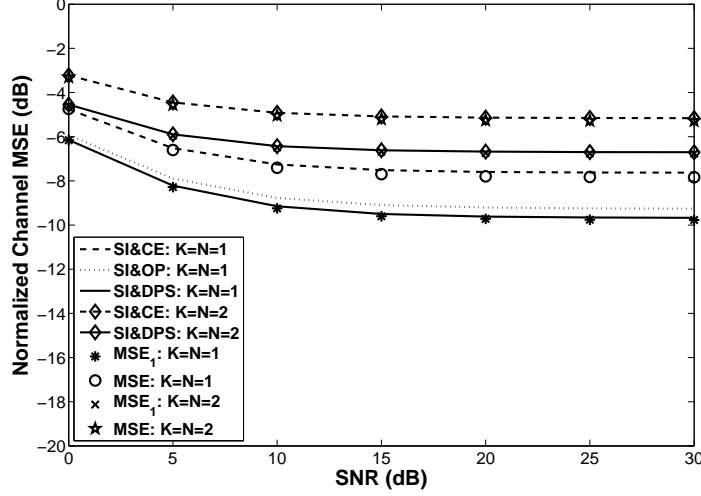


Figure 4.3: Estimation variance: NCMSE vs SNR under $f_d = 100$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM; MSE_1 : defined in (4.2); MSE: defined in (4.33).)

From an equalization viewpoint, based on (4.37)–(4.39), an equivalent SNR for data detection is defined as

$$\text{SNR}_d(Q) = \frac{\bar{\sigma}_{xs}^2(Q)}{\bar{\sigma}_w^2(Q)} = \frac{\sigma_b^2 [\text{MSE}_1 + (L+1)N\sigma_h^2 - \text{MSE}_2]}{\sigma_b^2 \text{MSE}_c + N\sigma_v^2 + \frac{(L+1)\sigma_h^2\sigma_b^2 + \sigma_v^2}{T} (L+1)NQ + \text{MSE}_2\sigma_c^2}. \quad (4.50)$$

We pick Q to maximize $\text{SNR}_d(Q)$ as we expect the detection performance to improve with increasing $\text{SNR}_d(Q)$.

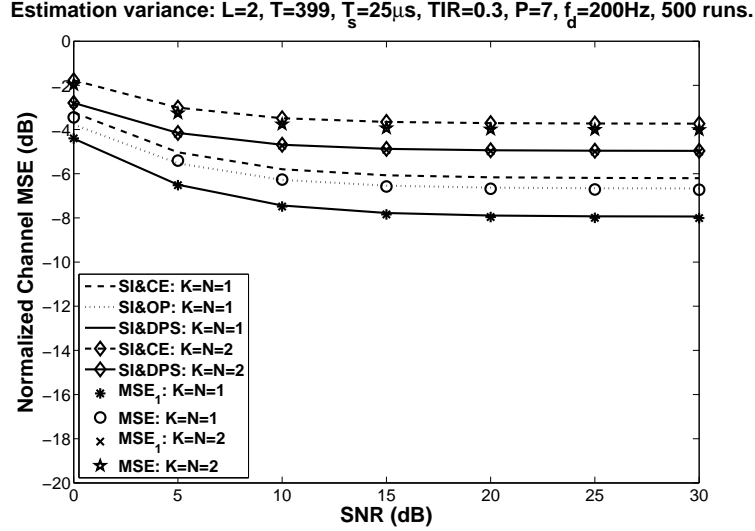


Figure 4.4: Estimation variance: NCMSE vs SNR under $f_d = 200$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM; MSE_1 : defined in (4.2); MSE: defined in (4.33).)

4.6 Simulation Examples

4.6.1 Performance Analysis for the First-Order Statistics-Based Estimator

In this example, we explore the variance of channel estimation of the first-order statistics-based estimator. Simulation results are compared with theoretical results to show the validity of our analysis.

We employ the same model for simulation as in Section 3.6, i.e., a doubly-selective Rayleigh fading channel with $L = 2$, satisfying modified Jakes' model. Performance of the channel estimation for both SISO ($K = N = 1$) and MIMO ($K = N = 2$) channels is investigated. One more time, we emphasize that BEM's are only used for processing at the receiver; random channels are generated using Jakes' model, not BEM representations.

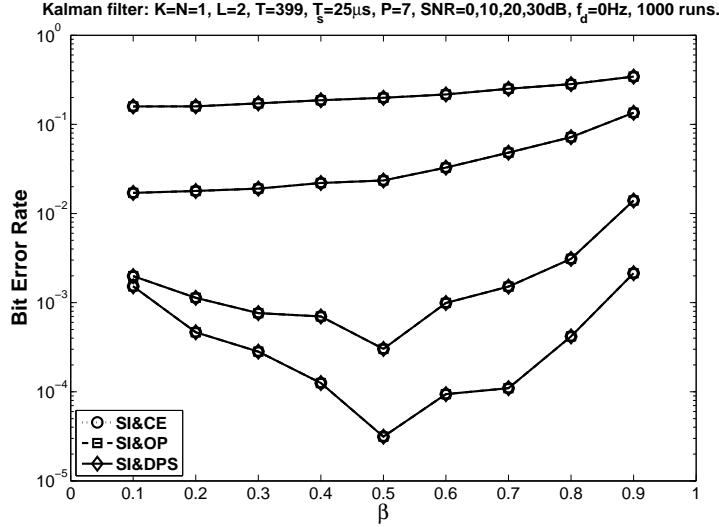


Figure 4.5: Training power allocation: BER vs β under $f_d = 0$ Hz (time-invariant). The curves for CE-, OP- and DPS-BEM's completely overlap, since the three basis functions are all constant for time-invariant channels ($Q = 1$). (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

In simulations, we pick a data record length of 399 symbols (time duration of approximately 10 ms). We consider the system operating under different Doppler spreads. For the Doppler spreads $f_d = 0, 50, 100,$ and 200 Hz, we take the number of basis functions $Q = 1, 3, 5,$ and 7 for CE-BEM, and $Q = 1, 3, 4,$ and 6 for OP- and DPS-BEM representations. The average transmitted power in $\{c(n)\}$ is 0.3 of that in $\{b(n)\}$, leading to TIR = 0.3.

In the single-user scenario, the information sequences $\{b(n)\}$ and the training sequences $\{c(n)\}$ are all BPSK modulated. The periodic training sequence $\{c(n)\}$ is generated from the m -sequence of period $P = 7$, one period of which is given by (3.82).

For the MIMO (multiple-user) case with $K = N = 2$, all the users have the same transmitted power in training and information data. The average transmitted power in

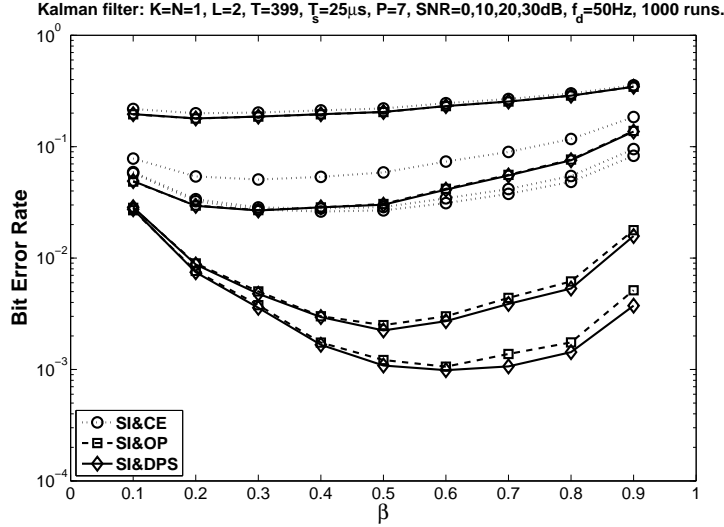


Figure 4.6: Training power allocation: BER vs β under $f_d = 50$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

$\{c_k(n)\}$ is 0.3 of that in $\{b_k(n)\}$ ($k = 1, 2, \dots, K$). The information sequences $\{b_k(n)\}$ and the training sequences $\{c_k(n)\}$ are also both BPSK modulated. The training sequence is generated from the above m -sequence of period $\tilde{P} = 7$ by the procedure we introduced in Section 3.5. The training sequences are of length $P = 14$, and the training sequence for the first user is given by (3.84).

To explore different estimators in equal conditions, we assume the additive noise $\{\mathbf{v}(n)\}$ is zero-mean (i.e., $\mathbf{m} = 0$), white complex-Gaussian, uncorrelated with $\{b(n)\}$ with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$, so that no terms are discarded. The (receiver) SNR refers to the energy per bit per user over one-sided noise spectral density with both information and superimposed training sequence counting toward the bit energy. The results for SISO and MIMO scenarios are shown in Figures 4.1–4.4, for various Doppler spreads and SNR's, based on 500 Monte Carlo runs.

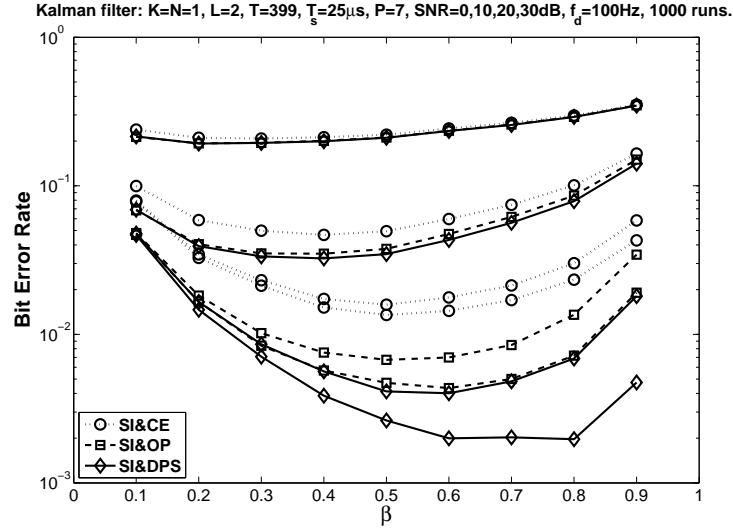


Figure 4.7: Training power allocation: BER vs β under $f_d = 100$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

In each figure, performances of CE-, OP-, and DPS-BEM-based estimators using first-order statistics are shown for the SISO case, together with CE- and DPS-BEM-based MIMO estimators. The normalized channel MSE in simulation is defined as (3.83) for the SISO channel and (3.85) for the MIMO channel. Simulation results are compared with the theoretical analysis of the “pure” estimation error MSE_1 (defined in (4.2)) as well as the “entire” error MSE_c (defined in (4.33)) counting modeling error. We plot MSE_1 based on CE- (using (4.7)), DPS- (using (4.16)), and OP-BEM (using (4.20)), for SISO and MIMO respectively, and MSE_c based only on CE-BEM (using (4.7), (4.34), and (4.35)).

Figure 4.1 exhibits the normalized channel MSE for simulation and theoretical results for Doppler spread $f_d = 0$ Hz. Note that since the channel is time-invariant, the basis sets of CE-, OP-, and DPS-BEM are the same (constant sequences), leading to the same simulated and analytical results. Since no modeling error is present, MSE_1 and MSE_c are equal. From

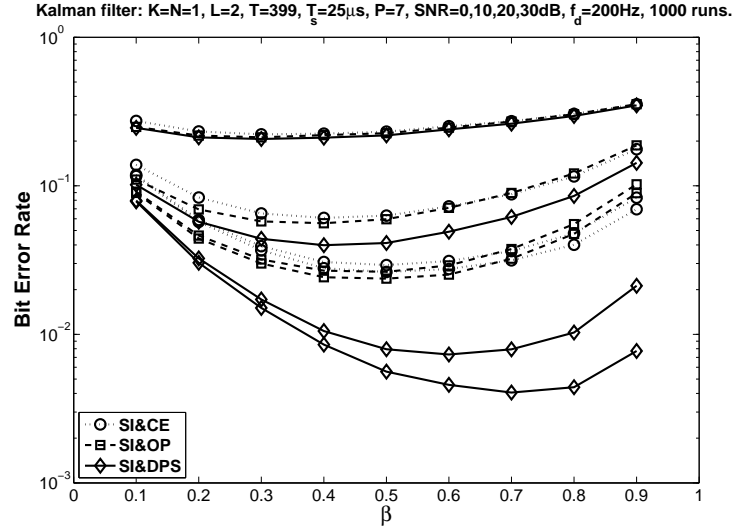


Figure 4.8: Training power allocation: BER vs β under $f_d = 200$ Hz. (SI: superimposed training; CE: CE-BEM; OP: OP-BEM; DPS: DPS-BEM.)

this figure, we can see that the simulation results and the analytical results agree very well, whether the channel is SISO or MIMO.

For time-varying channels, the modeling error cannot be eliminated. Considering Figures 4.2–4.4 corresponding to $f_d = 50, 100,$ and 200 Hz, different channel models now give distinct estimation error. The CE-BEM-based estimator has the highest estimation variance, and that of DPS-BEM is the lowest. For low Doppler spreads (slow fading channel), the OP-BEM-based solution has similar performance as that of the DPS-BEM-based one (see Figure 4.2), for the modeling errors are both tiny. As the Doppler spreads increases, however, the OP-BEM-based estimator deteriorates until reaching similar performance as the CE-BEM-based estimator for $f_d = 200$ Hz. In these three figures, the theoretically derived MSE_1 fits the simulation results of the DPS-BEM-based estimator very well, which confirms the fact that DPS-BEM offers the smallest (usually negligible) modeling error

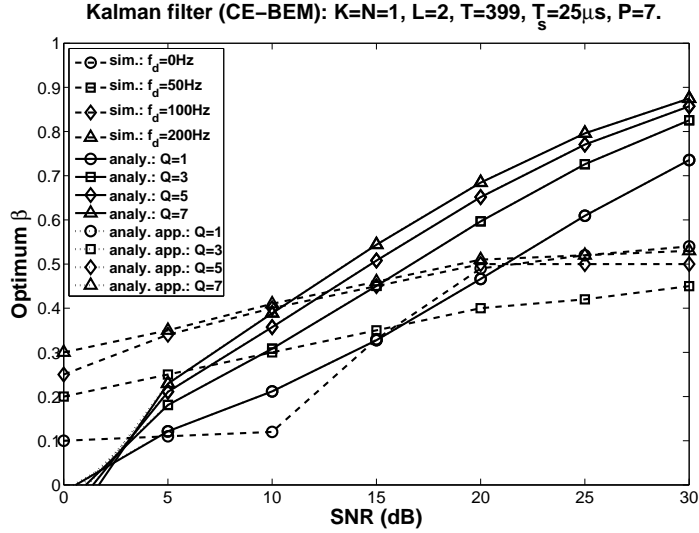


Figure 4.9: Training power allocation: optimum β vs SNR for CE-BEM. (“sim.”: simulation results; “analy.”: β_{opt} in (4.46); “analy. app.”: β'_{opt} in (4.48).)

among the three. Counting the modeling error, the curves for MSE_c (based on CE-BEM) also fit the simulation results of the CE-BEM-based estimator well.

4.6.2 Training Power Allocation

Under the same settings, we now consider the issue of training power allocation for the first-order statistics-based estimator. Only the SISO scenario is considered.

Figures 4.5–4.8 show the curves for BER versus different β 's (defined in (4.36)) that stands for the ratio of the power assigned to superimposed training to the total transmitted power, for different Doppler spreads. At the receiver, a Kalman filter is applied based on the estimated channel to detect the information sequence. All three BEM representations are studied.

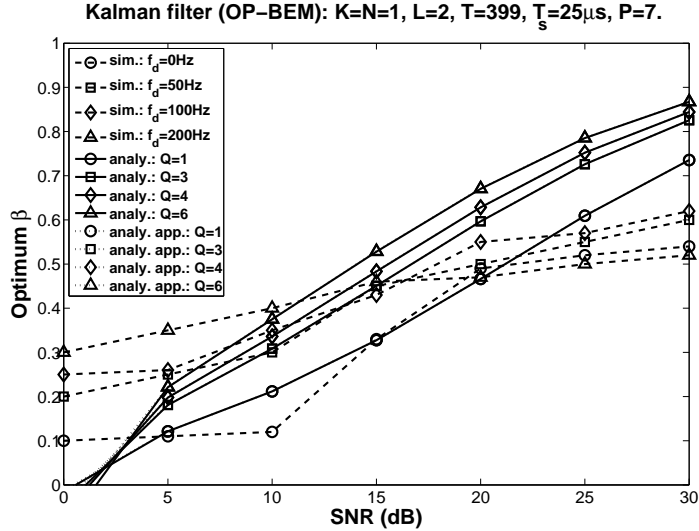


Figure 4.10: Training power allocation: optimum β vs SNR for OP-BEM. (“sim.”: simulation results; “analy.”: β_{opt} in (4.46); “analy. app.”: β'_{opt} in (4.48).)

We choose the optimal β as that corresponding to the smallest BER. As we expected, the optimal β grows with increasing SNR—at low SNR’s, more power assigned to training for better channel estimation cannot offset the effective SNR loss for information symbols, so that we should allocate more power to information for higher effective SNR; for higher SNR, noise is no longer the key factor, so that more training power for better estimation can achieve better BER’s.

Figures 4.9–4.11 compare the analytical results of the optimal β with the simulated results based on a Kalman filter. We consider three cases: the value of β minimizing the BER (denoted by “sim.” in the figures), the theoretical result β_{opt} in (4.46) (denoted by “analy.” in the figures), and an approximation of it in (4.48) (denoted by “analy. app.” in the figures). Note that we select β that minimizes the BER from Figures 4.5–4.8. The

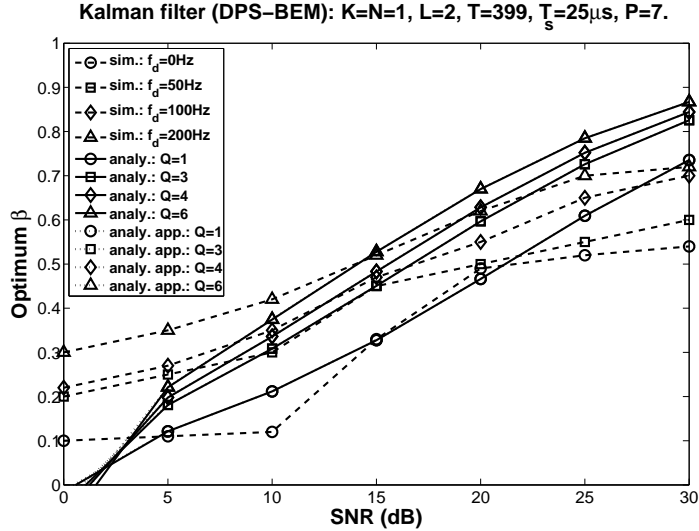


Figure 4.11: Training power allocation: optimum β vs SNR for DPS-BEM. (“sim.”: simulation results; “analy.”: β_{opt} in (4.46); “analy. app.”: β'_{opt} in (4.48).)

analytical results in (4.46) and (4.48) may produce negative solutions if the SNR is extremely low (e.g., 0dB)—we simply take $\beta = 0$ in that case. In the three figures, the approximation solution of (4.48) agrees well with that of (4.46). It is also seen that for all the curves, the optimal β grows as SNR increases.

We do not consider the modeling error in (4.46) and (4.48). In simulations, modeling error acts as a noise term in data reception, decreasing the actual $\text{SNR}_d(\beta)$, so that at high SNR’s the simulated optimal β is smaller than that of the analytical one. Due to larger modeling errors, the simulated optimal β for CE- and OP-BEM is smaller than that of the DPS-BEM. The simulation results for DPS-BEM fit the analytical solutions well for the range of SNR from 5dB to 20dB.

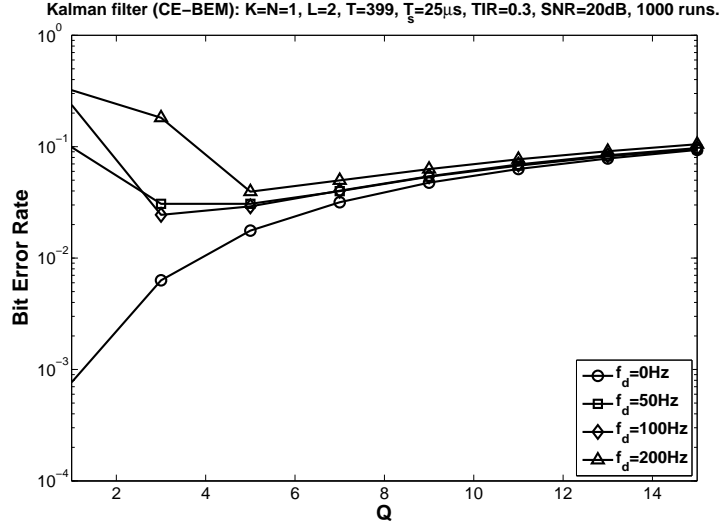


Figure 4.12: Bias-variance trade-off: BER vs Q under $\text{TIR} = 0.3$ for different f_d 's.

		$f_d = 0 \text{ Hz}$	$f_d = 50 \text{ Hz}$	$f_d = 100 \text{ Hz}$	$f_d = 200 \text{ Hz}$
$\text{TIR} = 0.3$	Analytical	1	5	3	5
	Simulation	1	5	3	5
	(2.9b)	1	3	3	5
$\text{TIR} = 1.0$	Analytical	1	7	5	7
	Simulation	1	7	5	7
	(2.9b)	1	3	3	5

Table 4.1: Selected optimal Q .

4.6.3 Bias-Variance Trade-Off

Under the same settings, we now consider the problem of bias-variance trade-off, i.e., selecting appropriate basis functions. Only CE-BEM and SISO channels are considered.

Figure 4.12 shows BER's for different Q , the number of basis functions employed in CE-BEM, for Doppler spreads $f_d = 0, 50, 100$, and 200 Hz where we choose $\text{TIR} = 0.3$. At the receiver, a Kalman filter is adopted for information symbol detection. Figure 4.14 displays those of $\text{TIR} = 1.0$. Figures 4.13 and 4.15 show the detection SNR defined in

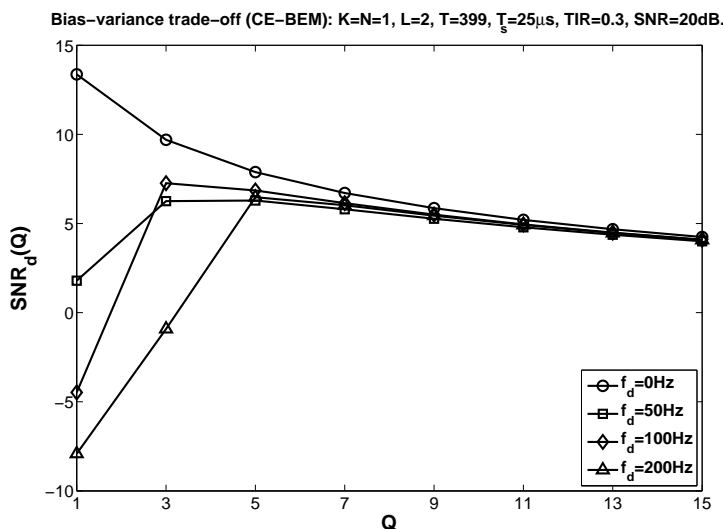


Figure 4.13: Bias-variance trade-off: $\text{SNR}_d(Q)$ (defined in (4.50)) vs Q under $\text{TIR} = 0.3$ for different f_d 's.

(4.50), as a function of Q . The agreement between the two sets of figures is very good: BER's are minimized for the same values of Q that maximize $\text{SNR}_d(Q)$.

In CE-BEM, the number of basis functions is usually given by (2.9b). Using superimposed training, TIR needs also to be considered to select Q : If more power has been assigned to superimposed training, more basis functions can be employed to get more accurate estimation, and vice versa. In Table 4.1, we compare the selected Q by simulation, and that maximize $\text{SNR}_d(Q)$ (denoted by "Analytical"), and by (2.9b). Our analytical results agree with the simulation better than that given by (2.9b).

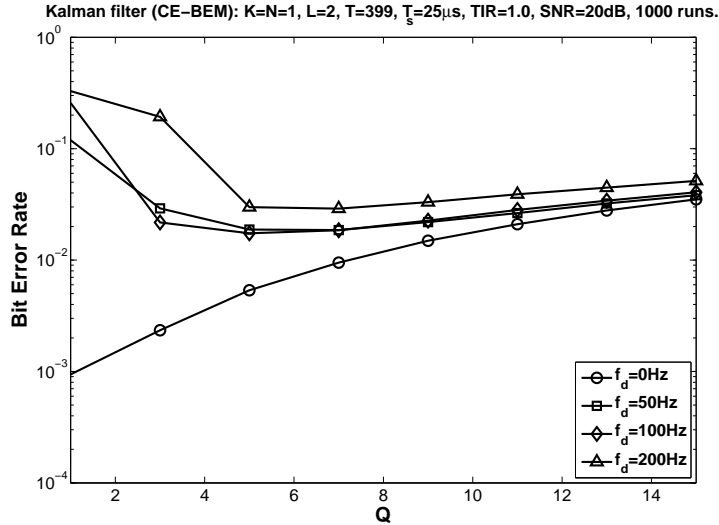


Figure 4.14: Bias-variance trade-off: BER vs Q under $\text{TIR} = 1.0$ for different f_d 's.

4.7 Conclusions

In this chapter, performance analysis of the first-order statistics-based estimator proposed in the previous chapter was discussed, under different BEM settings. Modeling error was also considered. We clearly showed that in this estimator, the major interference using superimposed training comes from the unknown information sequences. Power allocation and bias-variance trade-off of the first-order statistics-based estimator were also considered in this chapter, based on the results of performance analysis. We cast these optimization issues as maximization of an SNR for equalizer design. Numerical examples illustrated good agreement of our analytical results with the simulations.

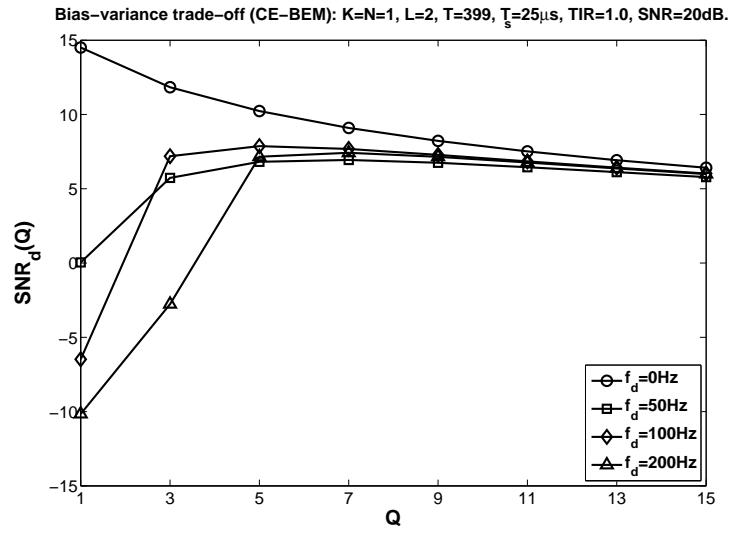


Figure 4.15: Bias-variance trade-off: $\text{SNR}_d(Q)$ (defined in (4.50)) vs Q under $\text{TIR} = 1.0$ for different f_d 's.

CHAPTER 5

DETERMINISTIC MAXIMUM LIKELIHOOD (DML) APPROACH

5.1 Introduction

By the performance analysis in Chapter 4, we can clearly see that the first-order statistics-based estimator proposed in Chapter 3 views the information sequence as interference in channel estimation, which leads to a poor received SNR. Since the training and information sequences pass through an identical channel, we exploit this fact to enhance channel estimation. Now we consider joint channel and information sequence estimation via an iterative DML approach, assuming that the noise $\mathbf{v}(n)$ is complex Gaussian. Convergence to a local extremum is guaranteed, and moreover, if the initial superimposed training-based solution is “good”, the global extremum (minimum error probability sequence detector) can be achieved by the DML approach.

We discuss the DML approach in this chapter. Section 5.2 deals with the single-user scenario and Section 5.3 considers the multiple-user case. Simulation examples illustrate our approach in Section 5.4, and Section 5.5 concludes this chapter.

5.2 DML Approach Using BEM

Consider the first-order statistics-based channel estimator described in Chapter 3. As in (3.1) and (3.2), the SIMO channel output is given by

$$\mathbf{x}(n) = \sum_{l=0}^L \mathbf{h}(n; l) s(n-l), \quad (5.1)$$

and its noisy measurement is given by

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n). \quad (5.2)$$

We make the following assumptions:

(H5.2.1) The time-varying channel $\{\mathbf{h}(n; l)\}$ satisfies a BEM representation as in (2.20)

$$\mathbf{h}(n; l) = \sum_{q=1}^Q \mathbf{h}_q(l) \psi_q(n) \quad (5.3)$$

where the basis functions $\{\psi_q(n)\}_{q=1}^Q$ are known at the receiver. Also $N \geq 1$.

(H5.2.2) The complex Gaussian noise $\{\mathbf{v}(n)\}$ may be of unknown mean $E\{\mathbf{v}(n)\} = \mathbf{m}$, white, uncorrelated with $\{b(n)\}$, and $E\{[\mathbf{v}(n + \tau) - \mathbf{m}][\mathbf{v}(n) - \mathbf{m}]^H\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$.

We collect $T - L$ samples of the observations into the vector

$$\mathbf{Y} := \begin{bmatrix} \mathbf{y}^T(T-1) & \mathbf{y}^T(T-2) & \cdots & \mathbf{y}^T(L) \end{bmatrix}^T. \quad (5.4)$$

Define

$$\mathbf{s} := \begin{bmatrix} s(T-1) & s(T-2) & \cdots & s(0) \end{bmatrix}^T,$$

and let

$$\tilde{\mathbf{v}}(n) := \mathbf{v}(n) - \mathbf{m}.$$

Given the vectors of the BEM coefficients in (3.12) and (3.15),

$$\begin{aligned}\mathbf{H}_l &:= [\mathbf{h}_1^T(l), \mathbf{h}_2^T(l), \dots, \mathbf{h}_Q^T(l)]^T, \\ \mathcal{H} &:= \begin{bmatrix} \mathbf{H}_0^T & \mathbf{H}_1^T & \dots & \mathbf{H}_L^T \end{bmatrix}^T.\end{aligned}$$

Define

$$\begin{aligned}\tilde{\mathbf{V}} &:= \begin{bmatrix} \tilde{\mathbf{v}}^T(T-1) & \tilde{\mathbf{v}}^T(T-2) & \dots & \tilde{\mathbf{v}}^T(L) \end{bmatrix}^T, \\ \mathcal{M} &:= \begin{bmatrix} \mathbf{m}^T & \mathbf{m}^T & \dots & \mathbf{m}^T \end{bmatrix}^T\end{aligned}$$

where \mathcal{M} is of the same size as $\tilde{\mathbf{V}}$, and $\mathbf{V} = \tilde{\mathbf{V}} + \mathcal{M}$ is a column-vector consisting of samples of noise $\{\mathbf{v}(n)\}$ in a manner similar to (5.4). Using (5.1)–(5.3) we have the following linear model

$$\mathbf{Y} = \mathcal{T}(\mathbf{s})\mathcal{H} + \tilde{\mathbf{V}} + \mathcal{M} \quad (5.5)$$

where $\mathcal{T}(\mathbf{s})$ is a block Hankel matrix (a block Hankel matrix has identical block entries on its antidiagonals) given by

$$\begin{aligned}\mathcal{T}(\mathbf{s}) &:= \begin{bmatrix} s(T-1)\Sigma_{T-1} & s(T-2)\Sigma_{T-1} & \dots & s(T-L-1)\Sigma_{T-1} \\ s(T-2)\Sigma_{T-2} & s(T-3)\Sigma_{T-2} & \dots & s(T-L-2)\Sigma_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ s(L)\Sigma_L & s(L-1)\Sigma_L & \dots & s(0)\Sigma_L \end{bmatrix}, \\ \Sigma_n &:= \begin{bmatrix} \psi_1(n)\mathbf{I}_N & \psi_2(n)\mathbf{I}_N & \dots & \psi_Q(n)\mathbf{I}_N \end{bmatrix}.\end{aligned}$$

Also using (5.1) and (5.2), an alternative linear model for \mathbf{Y} is given by

$$\mathbf{Y} = \mathcal{F}(\mathcal{H}) \mathbf{s} + \tilde{\mathbf{V}} + \mathcal{M} \quad (5.6)$$

where

$$\mathcal{F}(\mathcal{H}) := \begin{bmatrix} \mathbf{h}(T-1;0) & \cdots & \mathbf{h}(T-1;L) & & \\ & & \ddots & & \ddots \\ & & & \mathbf{h}(L;0) & \cdots & \mathbf{h}(L;L) \end{bmatrix}$$

is a “filtering matrix”.

Consider (5.1), (5.2), and (5.5). Under the assumption of temporally white complex Gaussian measurement noise, consider the joint estimation

$$\{\hat{\mathcal{H}}, \hat{\mathbf{s}}, \hat{\mathbf{m}}\} = \arg \min_{\mathcal{H}, \mathbf{s}, \mathbf{m}} \|\mathbf{Y} - \mathcal{T}(\mathbf{s}) \mathcal{H} - \mathcal{M}\|^2, \quad (5.7)$$

where $\hat{\mathbf{s}}$ is the estimate of \mathbf{s} . We follow a DML approach assuming no statistical model for the input sequence $\{s(n)\}$. Under a white Gaussian noise assumption, the DML estimates are obtained by the nonlinear LS optimization (5.7). Using (5.5) and (5.6), we have a separable nonlinear LS problem that can be solved sequentially as (joint optimization with respect to \mathcal{H} and \mathbf{m} can be further “separated”)

$$\begin{aligned} \{\hat{\mathcal{H}}, \hat{\mathbf{s}}, \hat{\mathbf{m}}\} &= \arg \min_{\mathbf{s}} \left\{ \min_{\mathcal{H}, \mathbf{m}} \|\mathbf{Y} - \mathcal{T}(\mathbf{s}) \mathcal{H} - \mathcal{M}\|^2 \right\} \\ &= \arg \min_{\mathcal{H}, \mathbf{m}} \left\{ \min_{\mathbf{s}} \|\mathbf{Y} - \mathcal{F}(\mathcal{H}) \mathbf{s} - \mathcal{M}\|^2 \right\}. \end{aligned}$$

The finite alphabet properties of the information sequences can also be incorporated into the DML methods. These algorithms, first proposed by [68] and also applied in [50], iterate between estimates of the channel and the input sequences. At iteration i , with an initial guess of the channel $\mathcal{H}^{(i)}$ and the mean $\mathbf{m}^{(i)}$, the algorithm estimates the input sequence $\mathbf{s}^{(i)}$ and the channel $\mathcal{H}^{(i+1)}$ and mean $\mathbf{m}^{(i+1)}$ for the next iteration by

$$\mathbf{s}^{(i)} = \arg \min_{\mathbf{s} \in \mathcal{S}} \left\| \mathbf{Y} - \mathcal{F} \left(\mathcal{H}^{(i)} \right) \mathbf{s} - \mathcal{M}^{(i)} \right\|^2, \quad (5.8a)$$

$$\mathcal{H}^{(i+1)} = \arg \min_{\mathcal{H}} \left\| \mathbf{Y} - \mathcal{T} \left(\mathbf{s}^{(i)} \right) \mathcal{H} - \mathcal{M}^{(i)} \right\|^2, \quad (5.8b)$$

$$\mathbf{m}^{(i+1)} = \arg \min_{\mathbf{m}} \left\| \mathbf{Y} - \mathcal{T} \left(\mathbf{s}^{(i)} \right) \mathcal{H}^{(i+1)} - \mathcal{M} \right\|^2 \quad (5.8c)$$

where \mathcal{S} is the (discrete) domain of \mathbf{s} . The optimizations in (5.8b) and (5.8c) are linear LS problems whereas the optimization in (5.8a) can be achieved by using the Viterbi algorithm. Since the above iterative procedure involving (5.8a)–(5.8c) decreases the cost at every iteration, one achieves a local minimum of the nonlinear LS cost (local maximum of DML function).

The maximum likelihood estimation of the noise mean in the optimization (5.8c) may be obtained by letting

$$\left. \frac{\partial \left\| \mathbf{Y} - \mathcal{T} \left(\mathbf{s}^{(i)} \right) \mathcal{H}^{(i+1)} - \mathcal{M} \right\|^2}{\partial \mathbf{m}} \right|_{\mathbf{m}=\mathbf{m}^{(i+1)}} = \mathbf{0},$$

which yields

$$\mathbf{m}^{(i+1)} = \frac{1}{T-L} \sum_{n=L}^{T-1} \left[\mathbf{y}(n) - \sum_{l=0}^L \mathbf{h}^{(i+1)}(n;l) s^{(i)}(n-l) \right].$$

We now summarize our DML approach:

1. **a.** Use the first-order statistics-based approach described in Chapter 3 to estimate the channel. Denote the estimate of the channel coefficients by $\hat{\mathcal{H}}^{(1)}$ and $\hat{\mathbf{h}}_q^{(1)}(l)$. In this method $\{c(n)\}$ is known and $\{b(n)\}$ is regarded as interference. (If we assume $\mathbf{m} = \mathbf{0}$ or known, the following steps to estimate \mathbf{m} should be skipped and simply set $\hat{\mathbf{m}} = \mathbf{0}$ or the known value.)
- b.** Estimate the mean $\hat{\mathbf{m}}^{(1)}$ as

$$\hat{\mathbf{m}}^{(1)} := \frac{1}{T} \sum_{n=0}^{T-1} \left[\mathbf{y}(n) - \sum_{l=0}^L \hat{\mathbf{h}}^{(1)}(n; l) c(n-l) \right] \quad (5.9)$$

where $\hat{\mathbf{h}}^{(1)}(n; l) := \sum_{q=1}^Q \hat{\mathbf{h}}_q^{(1)}(l) \psi_q(n)$ is given by (5.3).

- c.** Design a Viterbi sequence detector (see Appendix B.1) to estimate $\{s(n)\}$ as $\{\hat{s}(n)\}$ using the estimated channel $\hat{\mathcal{H}}^{(1)}$, mean $\hat{\mathbf{m}}^{(1)}$, and cost function in (5.8a) with $i = 1$. Note that knowledge of $\{c(n)\}$ is used in $s(n) = b(n) + c(n)$, therefore, we are in essence estimating $\{b(n)\}$.
2. **a.** Substitute $\tilde{s}(n)$ for $s(n)$ in (5.1) and use the corresponding formulation in (5.5) to estimate the channel \mathcal{H} as

$$\hat{\mathcal{H}}^{(2)} = \mathcal{T}^\dagger(\tilde{\mathbf{s}}) \left[\mathbf{Y} - \hat{\mathcal{M}}^{(1)} \right].$$

Following (5.9) the mean \mathbf{m} is estimated as $\hat{\mathbf{m}}^{(2)}$ for $i = 1$.

- b.** Design a Viterbi sequence detector using the estimated channel $\hat{\mathcal{H}}^{(2)}$, the mean $\hat{\mathbf{m}}^{(2)}$, and cost (5.8a) with $i = 2$, as in Step 1.c.

3. Step 2 provides one iteration of (5.8a)–(5.8c). Repeat a few times till any (relative) improvement in channel estimation over previous iteration is below a pre-specified threshold.

Since the Viterbi detector used in the proposed DML approach is computationally burdensome, we can replace it with a Kalman filter with hard decisions to expedite the iterations—at the expense of a little BER loss. This iterative method can follow these steps:

1. **a.** As Step 1.a of the DML approach.
- b.** As Step 1.b of the DML approach.
- c.** Design a Kalman filter (see Appendix B.2) of delay d to estimate $\{s(n)\}$ as $\{\hat{s}(n)\}$ using the estimated channel $\hat{\mathcal{H}}^{(1)}$ and mean $\hat{\mathbf{m}}^{(1)}$. Quantize $\{\hat{s}(n)\}$ into $\{\tilde{s}(n)\}$ with the knowledge of the symbol alphabet (hard decisions). Note that knowledge of $\{c(n)\}$ is used in $s(n) = b(n) + c(n)$, therefore, we are in essence estimating $\{b(n)\}$.
2. **a.** As Step 2.a of the DML approach.
- b.** Design a Kalman filter using the estimated channel $\hat{\mathcal{H}}^{(2)}$, the mean $\hat{\mathbf{m}}^{(2)}$, as in Step 1.c.
3. Step 2 provides one iteration of our proposed iterative method. Repeat a few times till any (relative) improvement in channel estimation over previous iteration is below a pre-specified threshold.

5.3 DML Approach: Multiple-User (MIMO) Channels

In this section we extend the DML approach to multiple-user (MIMO) channels corresponding to the estimator described in Section 3.5.

We collect $T - L$ samples of the observations to form the $N(T - L)$ -column vector as in (5.4)

$$\mathbf{Y} := \left[\mathbf{y}^T(T-1) \quad \mathbf{y}^T(T-2) \quad \cdots \quad \mathbf{y}^T(L) \right]^T$$

and the KT -column vector

$$\mathbf{s} := \left[s_1(T-1) \quad \cdots \quad s_K(T-1) \quad \cdots \quad s_1(0) \quad \cdots \quad s_K(0) \right]^T. \quad (5.10)$$

Define the $N \times NQ$ matrix

$$\Sigma_n := \left[\psi_1(n) \mathbf{I}_N \quad \psi_2(n) \mathbf{I}_N \quad \cdots \quad \psi_Q(n) \mathbf{I}_N \right],$$

the $N(T - L) \times NQ(L + 1)K$ matrix

$$\mathcal{T}(\mathbf{s}) := \begin{bmatrix} s_1(T-1)\Sigma_{T-1} & \cdots & s_1(T-L-1)\Sigma_{T-1} & \cdots & s_K(T-L-1)\Sigma_{T-1} \\ s_1(T-2)\Sigma_{T-2} & \cdots & s_1(T-L-2)\Sigma_{T-2} & \cdots & s_K(T-L-2)\Sigma_{T-2} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_1(L)\Sigma_L & \cdots & s_1(0)\Sigma_L & \cdots & s_K(0)\Sigma_L \end{bmatrix}$$

the $NQ(L+1)K$ -column vector (by (3.77))

$$\mathcal{H} := \begin{bmatrix} \mathcal{H}_1^T & \mathcal{H}_2^T & \cdots & \mathcal{H}_K^T \end{bmatrix}^T,$$

and the $N(T-L)$ -column vector

$$\tilde{\mathbf{V}} := \begin{bmatrix} \tilde{\mathbf{v}}^T(T-1), & \tilde{\mathbf{v}}^T(T-2), & \dots, & \tilde{\mathbf{v}}^T(L) \end{bmatrix}^T$$

where $\tilde{\mathbf{v}}(n) := \mathbf{v}(n) - \mathbf{m}$. We also have the following linear model

$$\mathbf{Y} = \mathcal{T}(\mathbf{s})\mathcal{H} + \tilde{\mathbf{V}} + \mathcal{M} \tag{5.11}$$

where $\mathcal{M} := \begin{bmatrix} \mathbf{m}^T & \mathbf{m}^T & \cdots & \mathbf{m}^T \end{bmatrix}^T$.

We further define the $N(T-L) \times K(L+1)$ matrix

$$\mathcal{F}(\mathcal{H}) := \begin{bmatrix} \mathbf{h}_1(T-1;0) & \cdots & \mathbf{h}_K(T-1;0) & \cdots & \mathbf{h}_K(T-1;L) & & \\ & & \ddots & & \ddots & & \ddots \\ & & & \mathbf{h}_1(L;0) & \cdots & \mathbf{h}_1(L;L) & \cdots & \mathbf{h}_K(L;L) \end{bmatrix}$$

and obtain another linear model as follows

$$\mathbf{Y} = \mathcal{F}(\mathcal{H})\mathbf{s} + \tilde{\mathbf{V}} + \mathcal{M}. \tag{5.12}$$

By (5.11) and (5.12), the DML approach described in Section 5.2 can be followed. Under the assumption of white complex Gaussian measurement noise, the joint estimators of the relevant parameters are given by the following nonlinear optimization problem

$$\begin{aligned} \{\hat{\mathcal{H}}, \hat{\mathbf{s}}, \hat{\mathbf{m}}\} &= \arg \left\{ \min_{\mathcal{H}, \mathbf{s}, \mathbf{m}} \|\mathbf{Y} - \mathcal{T}(\mathbf{s}) - \mathcal{M}\|^2 \right\} \\ &= \arg \left\{ \min_{\mathcal{H}, \mathbf{s}, \mathbf{m}} \|\mathbf{Y} - \mathcal{F}(\mathcal{H}) - \mathcal{M}\|^2 \right\}. \end{aligned}$$

We also follow the DML approach assuming no statistical model for the input sequences $\{s_k(n)\}$. We may separate the nonlinear LS problem sequentially as

$$\begin{aligned} \{\hat{\mathcal{H}}, \hat{\mathbf{s}}, \hat{\mathbf{m}}\} &= \arg \min_{\mathbf{s}} \left\{ \min_{\mathcal{H}, \mathbf{m}} \|\mathbf{Y} - \mathcal{T}(\mathbf{s})\mathcal{H} - \mathcal{M}\|^2 \right\} \\ &= \arg \min_{\mathcal{H}, \mathbf{m}} \left\{ \min_{\mathbf{s}} \|\mathbf{Y} - \mathcal{F}(\mathcal{H})\mathbf{s} - \mathcal{M}\|^2 \right\}. \end{aligned}$$

At iteration i , with an initial guess of the channel $\mathcal{H}^{(i)}$ and the mean $\mathbf{m}^{(i)}$, the algorithm estimates the input sequence $\mathbf{s}^{(i)}$ and the channel $\mathcal{H}^{(i+1)}$ and mean $\mathbf{m}^{(i+1)}$ for the next iteration by

$$\mathbf{s}^{(i)} = \arg \min_{\mathbf{s} \in \mathcal{S}} \|\mathbf{Y} - \mathcal{C}(\mathcal{H}^{(i)})\mathbf{s} - \mathcal{M}^{(i)}\|^2, \quad (5.13a)$$

$$\mathcal{H}^{(i+1)} = \arg \min_{\mathcal{H}} \|\mathbf{Y} - \mathcal{T}(\mathbf{s}^{(i)})\mathcal{H} - \mathcal{M}^{(i)}\|^2, \quad (5.13b)$$

$$\mathbf{m}^{(i+1)} = \arg \min_{\mathbf{m}} \|\mathbf{Y} - \mathcal{T}(\mathbf{s}^{(i)})\mathcal{H}^{(i+1)} - \mathcal{M}\|^2. \quad (5.13c)$$

We now summarize our MIMO DML approach:

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $TIR=0.3$, $P=7$, $f_d=0\text{Hz}$, 500 runs.

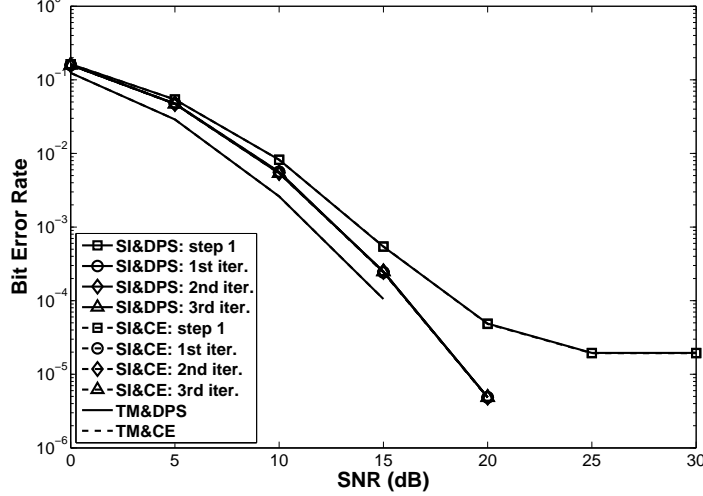


Figure 5.1: DML approach (SISO): BER vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 1$. The curves for CE- and DPS-BEM’s completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

1. a. Use (3.80) and (3.81) to estimate the channel. Denote the channel estimates by $\hat{\mathcal{H}}_k^{(1)}$ and $\hat{\mathbf{h}}_k^{(1)}(n; l)$. In this method $\{c_k(n)\}$ is known and $\{b_k(n)\}$ is regarded as interference.

- b. The noise mean \mathbf{m} is estimated as

$$\hat{\mathbf{m}}^{(1)} = \frac{1}{T} \sum_{n=1}^T \left[\mathbf{y}(n) - \sum_{k=1}^K \sum_{l=0}^L \hat{\mathbf{h}}_k^{(1)}(n; l) c_k(n-l) \right]. \quad (5.14)$$

- c. Design a Viterbi sequence detector (see Appendix B.1) to estimate $\{s_k(n)\}$ as $\{\tilde{s}_k(n)\}$ using the estimated channel $\hat{\mathcal{H}}^{(1)}$, mean $\hat{\mathbf{m}}^{(1)}$ and cost (5.13a) with $i = 1$.

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=50\text{Hz}$, 500 runs.

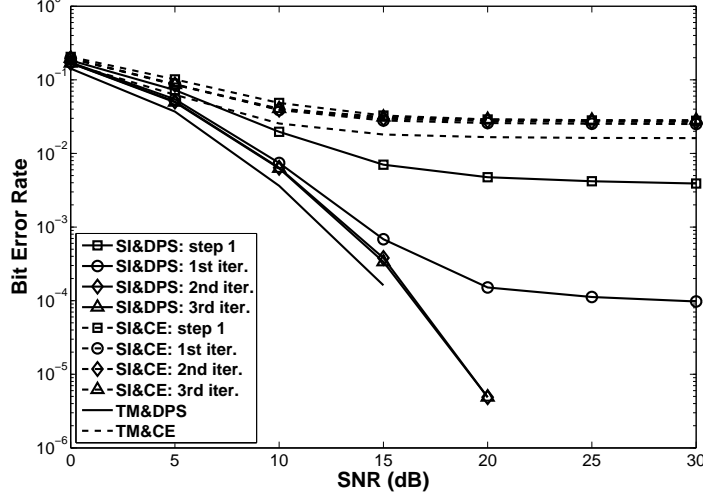


Figure 5.2: DML approach (SISO): BER vs SNR under $f_d = 50\text{Hz}$ and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

Note that knowledge of $\{c_k(n)\}$ is used in $s_k(n) = b_k(n) + c_k(n)$, therefore, we are in essence estimating $b_k(n)$ in the Viterbi detector.

2. a. Substitute $\tilde{s}_k(n)$ for $s_k(n)$ in (5.10) and use the corresponding formulation in equation (5.13b) to estimate the time-invariant channel coefficient matrix \mathcal{H} as

$$\hat{\mathcal{H}}^{(2)} = \mathcal{T}^\dagger(\tilde{\mathbf{s}}) \left[\mathbf{Y} - \hat{\mathcal{M}}^{(1)} \right]$$

and estimate the time-varying channel as $\hat{\mathbf{h}}_k^{(2)}(n; l)$ using (3.81). The mean \mathbf{m} is estimated as $\hat{\mathbf{m}}^{(2)}$ using (5.14) with $\hat{\mathbf{h}}_k^{(1)}(n; l)$ replaced with $\hat{\mathbf{h}}_k^{(2)}(n; l)$.

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $TIR=0.3$, $P=7$, $f_d=100\text{Hz}$, 500 runs.

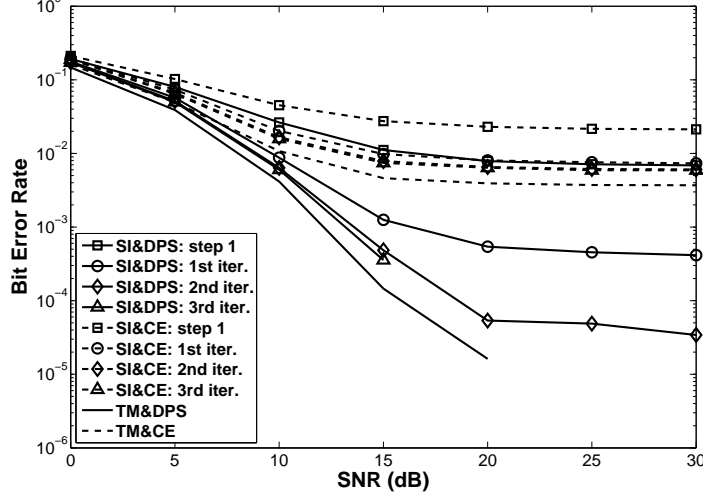


Figure 5.3: DML approach (SISO): BER vs SNR under $f_d = 100\text{Hz}$ and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

- b. Design a Viterbi sequence detector using the estimated channel $\hat{\mathcal{H}}^{(2)}$, mean $\hat{\mathbf{m}}^{(2)}$ and cost (5.13a) with $i = 2$, as in Step 1.c.
3. Step 2 provides one iteration of (5.13a)–(5.13c). Repeat a few times until reaching the desired point.

An approximation of the MIMO DML approach by replacing Viterbi detector with multiple-user Kalman filter is given by the following steps:

1. **a.** As Step 1.a of the MIMO DML approach.
- b.** As Step 1.b of the MIMO DML approach.

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=200\text{Hz}$, 500 runs.

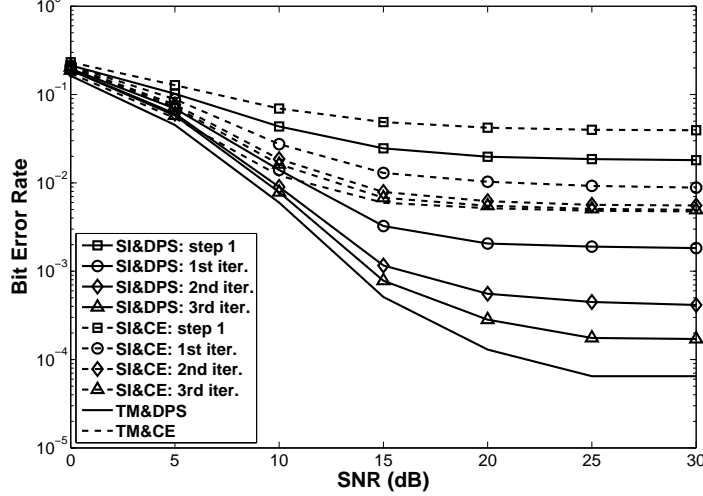


Figure 5.4: DML approach (SISO): BER vs SNR under $f_d = 200\text{Hz}$ and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

- c. Design a multiple-user Kalman filter (see Appendix B.2) of delay d to estimate $\{s_k(n)\}$ as $\{\hat{s}_k(n)\}$ using the estimated channel $\hat{\mathcal{H}}^{(1)}$ and mean $\hat{\mathbf{m}}^{(1)}$. Quantize $\{\hat{s}_k(n)\}$ into $\{\tilde{s}_k(n)\}$ with the knowledge of the symbol alphabet (hard decisions). Note that knowledge of $\{c_k(n)\}$ is used in $s_k(n) = b_k(n) + c_k(n)$, therefore, we are in essence estimating $\{b_k(n)\}$.

2. a. As Step 2.a of the DML approach.

- b. Design a multiple-user Kalman filter using the estimated channel $\hat{\mathcal{H}}^{(2)}$, the mean $\hat{\mathbf{m}}^{(2)}$, as in Step 1.c.

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $TIR=0.3$, $P=7$, $f_d=0\text{Hz}$, 500 runs.

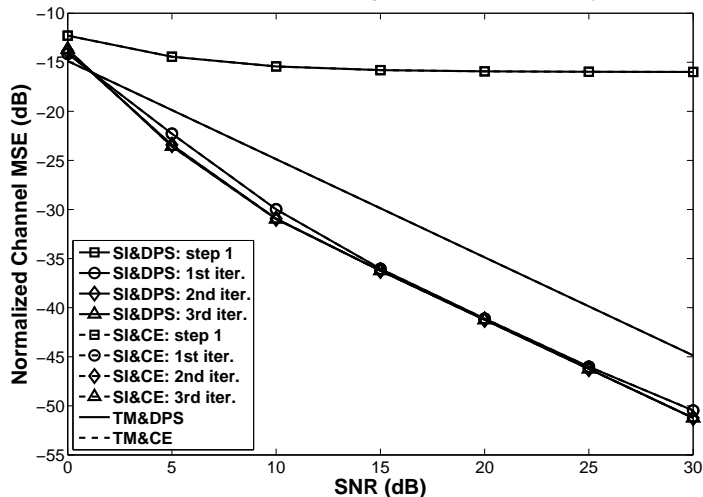


Figure 5.5: DML approach (SISO): NCMSE vs SNR under $f_d = 0\text{Hz}$ (time-invariant) and $K = N = 1$. The curves for CE- and DPS-BEM’s completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

- Step 2 provides one iteration of our proposed iterative method. Repeat a few times till any (relative) improvement in channel estimation over previous iteration is below a pre-specified threshold.

5.4 Simulation Examples

5.4.1 DML Approach: Single User

In this example, we adopt the same simulation conditions as in Section 3.6.1 to perform a comparison between the first-order statistics-based estimator in Chapter 3 and our DML

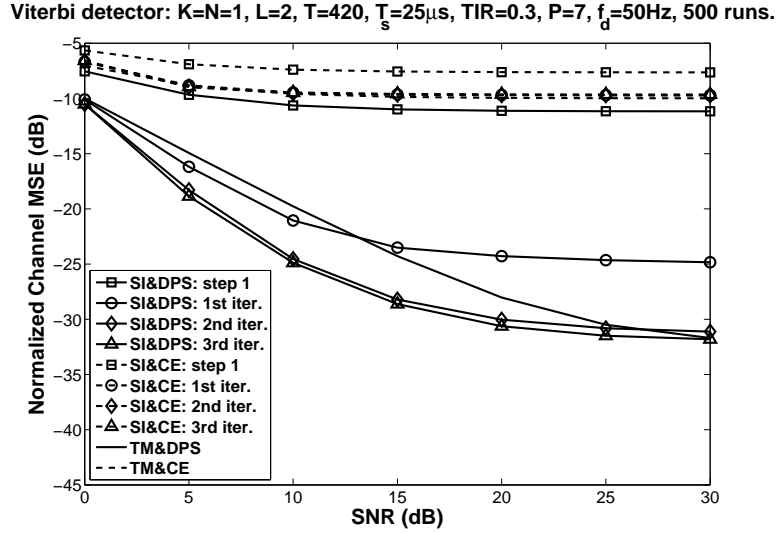


Figure 5.6: DML approach (SISO): NCMSE vs SNR under $f_d = 50\text{Hz}$ and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

approach. We generate a doubly-selective Rayleigh fading channel following Jakes’ model with $N = 1$ and $L = 2$.

In simulations, we pick a data record length of 420 symbols (time duration of approximately 10 ms). We consider the system operating under different Doppler spreads with different number of basis functions Q . For the Doppler spreads $f_d = 0, 50, 100$, and 200Hz , we take $Q = 1, 3, 5$, and 7 for the CE-BEM-based solution, and $Q = 1, 3, 4$, and 6 for the DPS-BEM representations. The average transmitted power in $\{c(n)\}$ is 0.3 of that in $\{b(n)\}$, leading to TIR of 0.3.

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=100\text{Hz}$, 500 runs.

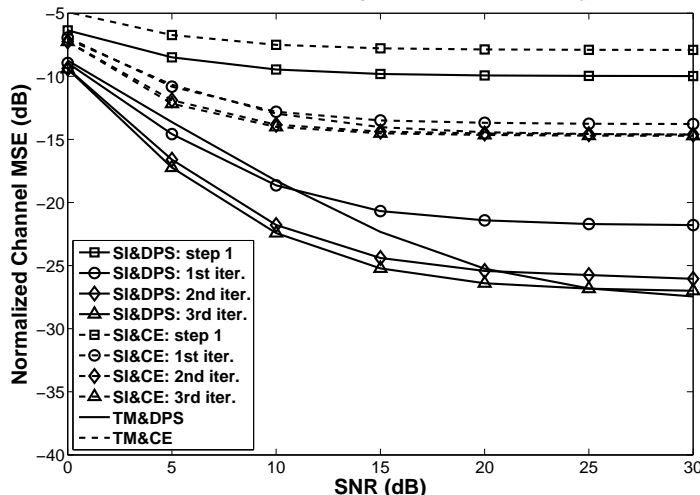


Figure 5.7: DML approach (SISO): NCMSE vs SNR under $f_d = 100\text{Hz}$ and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

We first consider a single-user scenario. The information sequences $\{b(n)\}$ and the training sequences $\{c(n)\}$ are all BPSK modulated. The periodic training sequence $\{c(n)\}$ is generated from the m -sequence of period $P = 7$, one period of which is given by (3.82).

To explore different estimators and their iterative versions under identical conditions, we assume the additive noise $\{\mathbf{v}(n)\}$ is zero-mean (i.e., $\mathbf{m} = 0$), white complex-Gaussian, uncorrelated with $\{b(n)\}$ with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$, so that no terms are discarded in the first-order statistics-based estimator, the first step of our DML approach. The (receiver) SNR refers to the energy per bit over one-sided noise spectral density with both information and superimposed training sequence counting toward the bit energy. At the receiver, a Viterbi detector is used for data reception.

Viterbi detector: $K=N=1$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=200\text{Hz}$, 500 runs.

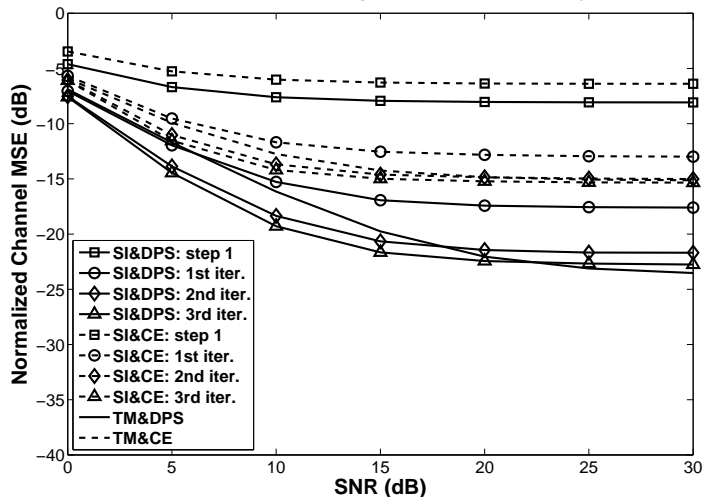


Figure 5.8: DML approach (SISO): NCMSE vs SNR under $f_d = 200\text{Hz}$ and $K = N = 1$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

The results for a record length of $T = 420$ symbols are shown in Figures 5.1–5.8 for various Doppler spreads and SNR’s. The results are based on 500 Monte Carlo runs. To compare with other possible approaches, CE- and DPS-BEM-based TM training described in Appendix A is also considered for doubly-selective channel estimation. Training sessions are periodically inserted in the transmitted symbol frame. We take a training session of length of $2L + 1 = 5$ symbols with the training sequence $\{0, 0, \sqrt{2L + 1}, 0, 0\}$, and at the receiver an LS estimation is performed. A data session of 18 symbols is inserted between two successive training sessions to form a frame of length 23 symbols. Such a frame is repeated over a record length of 418 symbols. Thus, we have a training-to-information bit and power ratio of about 0.3.

Viterbi detector: $K=N=2$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $TIR=0.3$, $P=7$, $f_d=0\text{Hz}$, 500 runs.

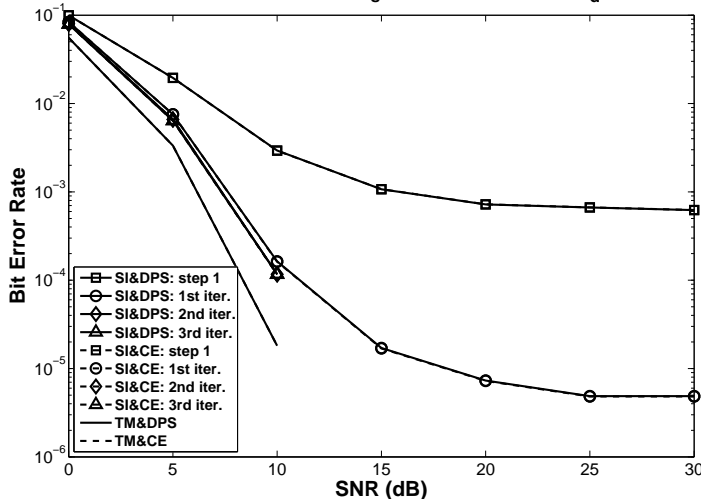


Figure 5.9: DML approach (MIMO): BER vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 2$. The curves for CE- and DPS-BEM’s completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

For comparison, we plot the results of the CE- and DPS-BEM-based superimposed training schemes (denoted as SI in the figures), including the first-order statistics-based estimator (denoted as “step 1” in the figures), and the DML approach after one, two, and three iterations (denoted as “1st iter.”, “2nd iter.”, and “3rd iter.” in the figures), and TM training approaches (denoted as TM in the figures). Figures 5.1–5.4 show BER’s for $f_d = 0, 50, 100$, and 200 respectively. Figures 5.5–5.8 show the normalized channel MSE correspondingly, which is defined as (3.83).

From the eight figures, we can see that after iterations, superimposed training-based estimation and detection performances improve a lot, because the information data that

Viterbi detector: $K=N=2$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=50\text{Hz}$, 500 runs.

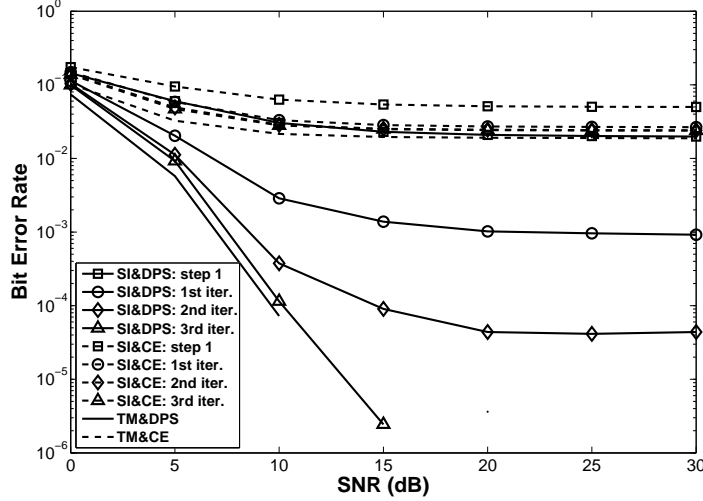


Figure 5.10: DML approach (MIMO): BER vs SNR under $f_d = 50$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

are viewed as interference by the first-order statistics-based estimator are now exploited to enhance the channel estimation for the next iteration. Therefore, the self-interference is efficiently removed after iterations. The DML approach provides comparable error performance with TM training, but at a higher data transmission rate. The valuable bandwidth resources can thus be saved by using iterative estimation, at the expense of increased computational complexity.

Since we assume the channel satisfies a BEM, the modeling error of the prescribed BEM sets a limit for the estimation performance. Except for $f_d = 0$ Hz, the DPS-BEM is much more accurate in describing a band-limited time-varying channel, so that it provides much better estimation performance than the CE-BEM. For both BER and NCMSE curves,

Viterbi detector: $K=N=2$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=100\text{Hz}$, 500 runs.

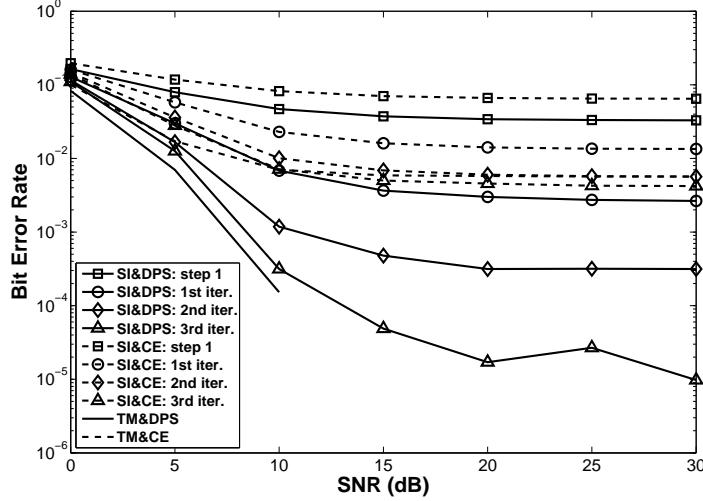


Figure 5.11: DML approach (MIMO): BER vs SNR under $f_d = 100\text{ Hz}$ and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

the error floors of the DPS-BEM-based solutions are lower than that of the CE-BEM-based ones.

As the Doppler spread f_d increases, the performance of the DML-based superimposed training deteriorates compared with the TM training. Figures 5.4 and 5.8 for $f_d = 200\text{ Hz}$ clearly show this result. It is partially because the first-order statistics-based estimator performs worse for fast fading channels, since more basis functions involved to represent the channel result in higher estimation variance. More iterations are required to approach the performance of TM training.

Viterbi detector: $K=N=2$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=200\text{Hz}$, 500 runs.

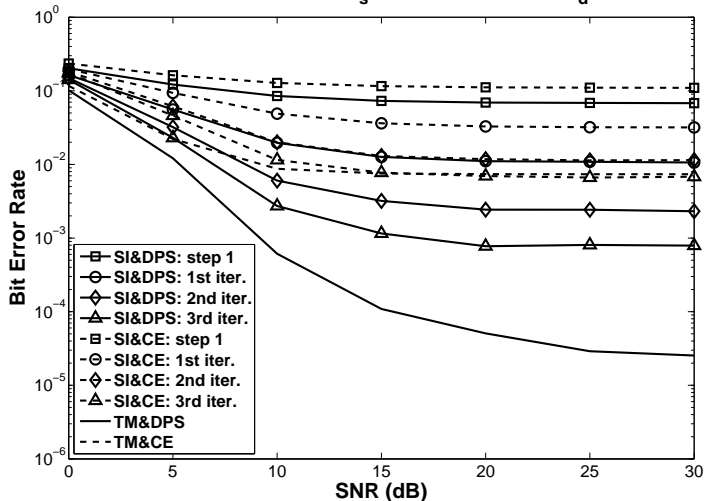


Figure 5.12: DML approach (MIMO): BER vs SNR under $f_d = 200\text{ Hz}$ and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

5.4.2 DML Approach: Multiple Users

In this example, we follow the settings in Section 5.4.1 except that a multiple-user scenario is considered. It can also be viewed as an extension of Section 3.6.2, for now the iterative DML approach based on the multiple-user channel estimator using the first-order statistics is considered.

In simulations, we assume that all the users have the same transmitted power in training and information data. The average transmitted power in $\{c_k(n)\}$ is 0.3 of that in $\{b_k(n)\}$ ($k = 1, 2, \dots, K$), leading to the same TIR as in Section 5.4.1. We consider a simple two-user scenario, i.e., $K = 2$, each user with two receive antennas, i.e., $N = 2$. The information sequences $\{b_k(n)\}$ and the training sequences $\{c_k(n)\}$ are all BPSK modulated. The training

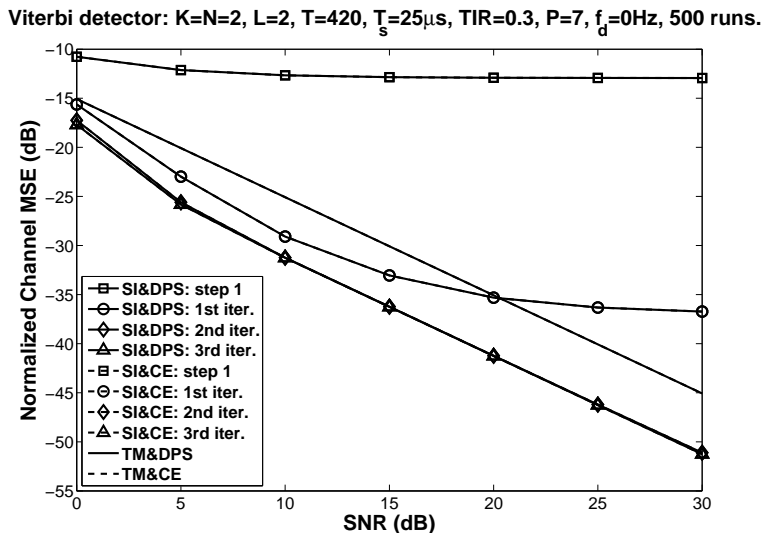


Figure 5.13: DML approach (MIMO): NCMSE vs SNR under $f_d = 0$ Hz (time-invariant) and $K = N = 2$. The curves for CE- and DPS-BEM's completely overlap, since the two basis functions are both constant for time-invariant channels ($Q = 1$). (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

sequence is generated from the m -sequence of period $\tilde{P} = 7$ by the procedure we introduced in Section 3.5.

The additive noise $\{\mathbf{v}(n)\}$ is also zero-mean, white complex-Gaussian, uncorrelated with $\{b_k(n)\}$ with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_2 \delta(\tau)$. The (receiver) SNR refers to the energy per bit per user over one-sided noise spectral density with both information and superimposed training sequence counting toward the bit energy.

At the receive end, a Viterbi detector is used for symbol detection. We consider different Doppler spreads of $f_d = 0, 50, 100$, and 200 Hz for this communications system. We also pick Q of CE-BEM as 1, 3, 5, 7 and DPS-BEM as 1, 3, 4, 6.

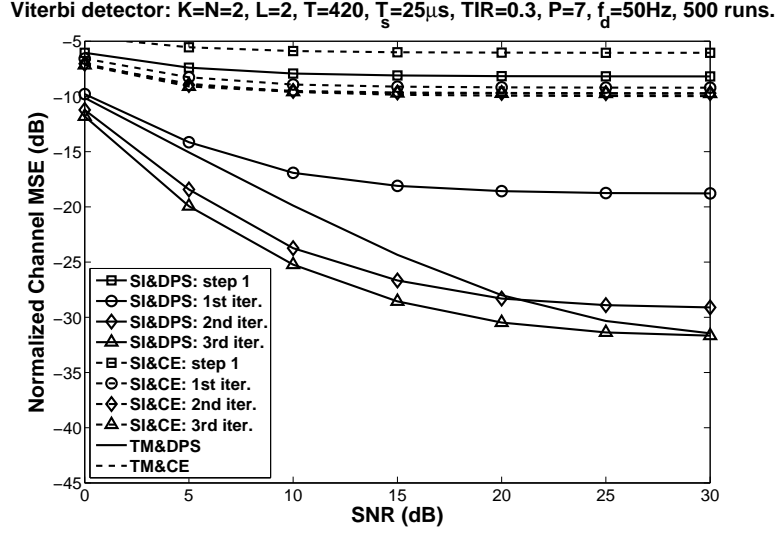


Figure 5.14: DML approach (MIMO): NCMSE vs SNR under $f_d = 50\text{Hz}$ and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

The results for a record length of $T = 420$ symbols are shown in Figures 5.9–5.16 for various Doppler spreads and SNR’s. The results are based on 500 Monte Carlo runs. For comparison, CE-BEM and DPS-BEM-based periodically placed TM training with zero-padding, as we described in Appendix A, is also considered. We take a training session of length $(K + 1)L + K = 8$ symbols with the first user’s training $\{\mathbf{0}_{1 \times 2}, \sqrt{(K + 1)L + K}, \mathbf{0}_{1 \times 5}\}$ and the second user’s $\{\mathbf{0}_{1 \times 5}, \sqrt{(K + 1)L + K}, \mathbf{0}_{1 \times 2}\}$. A data session of 27 symbols is inserted between two such training sessions to form a frame of length 35 symbols. Such a frame is repeated over a record length of 420 bits. Thus, we have a training-to-information bit and power ratio of about 0.3. At the receiver an LS estimation is performed.

Viterbi detector: $K=N=2$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $TIR=0.3$, $P=7$, $f_d=100\text{Hz}$, 500 runs.

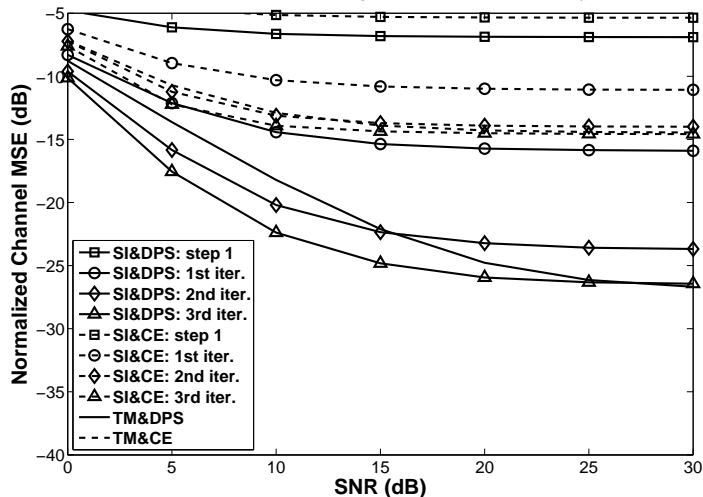


Figure 5.15: DML approach (MIMO): NCMSE vs SNR under $f_d = 100$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

Figures 5.9–5.16 show similar results as in the SISO case: The DML approach enhances the channel estimation and data detection performances significantly over the first-order statistics-based estimator for a multiple-user channel; DPS-BEM well outperforms CE-BEM, so that it appears to be a good choice to approximate the time-varying channel.

5.5 Conclusions

We explored the DML approach in this chapter. By exploiting the fact that training and information sequences pass through an identical channel, the iterative DML approach was used to jointly improve the channel and sequence estimation. Beginning with the first-order statistics-based channel estimator, the detected data symbols from the preceding

Viterbi detector: $K=N=2$, $L=2$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.3$, $P=7$, $f_d=200\text{Hz}$, 500 runs.

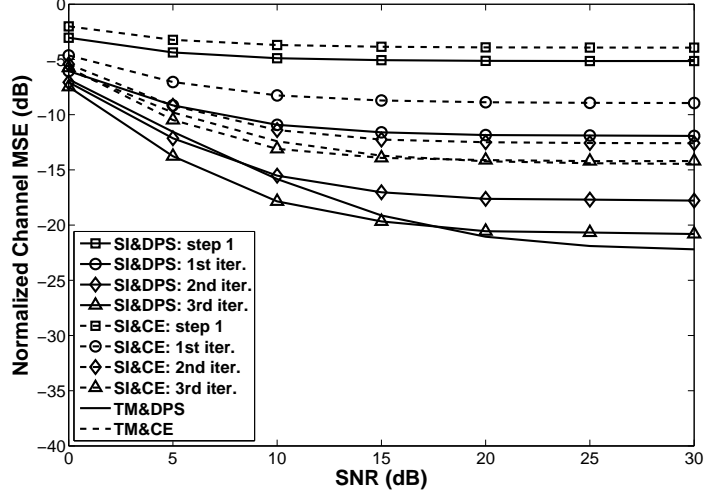


Figure 5.16: DML approach (MIMO): NCMSE vs SNR under $f_d = 200$ Hz and $K = N = 2$. (SI: superimposed training; TM: time-multiplexed training; CE: CE-BEM; DPS: DPS-BEM; “step 1”: the first-order statistics-based estimator; “1st iter.”: the first DML iteration; “2nd iter.”: the second DML iteration; “3rd iter.”: the third DML iteration.)

iteration are used to reduce the self-interference at the current iteration. A local maximum of DML function is guaranteed. Symbol detection techniques such as Kalman filtering can also be adopted instead of Viterbi algorithm to reduce the computational complexity in iterations—the method can be viewed as an approximation of the DML approach.

CHAPTER 6

DOUBLY-SELECTIVE CHANNEL ESTIMATION USING DATA-DEPENDENT SUPERIMPOSED TRAINING

6.1 Introduction

For the first-order statistics-based channel estimator proposed in Chapter 3, the information sequence acts as interference resulting in a poor training SNR. Simulation results have shown that noticeable error floors occur in BER and channel MSE curves for this estimator. Although we can employ the DML method described in Chapter 5 to reduce the interference, DML iterations add to computational complexity and delay in symbol detection at the receiver.

Inspired by the work of [20], where the training sequence is distorted according to the information data before transmission so as to eliminate the self-interference on reception, we extend this data-dependent method to time-varying channels by the aid of BEM representations.

For the first-order statistics-based estimator using CE- or DPS-BEM, (4.11) addresses the source of the estimation error (see **Remark 4.2.2.2** for a detailed discussion):

$$\hat{\mathbf{d}}_{mq} = \mathbf{d}_{mq} + \mathbf{s}_{mq} + \mathbf{w}_{mq},$$

where the information sequence's contribution, given by \mathbf{s}_{mq} , interferes with the estimation of \mathbf{d}_{mq} from $\hat{\mathbf{d}}_{mq}$, and hence with channel estimation from the observations. For the CE-BEM-based estimator,

$$\mathbf{s}_{mq} := \frac{1}{T} \sum_{n=0}^{T-1} \left\{ \sum_{l=0}^L \mathbf{h}(n; l) b(n-l) \right\} e^{-j(\omega_q + \alpha_m)n}, \quad (6.1)$$

and for the DPS-BEM-based estimator,

$$\mathbf{s}_{mq} := \sum_{n=0}^{T-1} \left[\sum_{l=0}^L \mathbf{h}(n; l) b(n-l) \right] u_q(n) e^{-j\alpha_m n} \quad (6.2)$$

as in (4.12). Our goal of the data-dependent superimposed training is to null out the influence of \mathbf{s}_{mq} on the channel estimation by transmitter-end processing.

In this chapter, we focus on transmitter-end processing techniques to reduce self-interference of superimposed training. In Section 6.2, we present the data-dependent superimposed training based on CE-BEM; this scheme is extended to channels satisfying DPS-BEM representation in Section 6.3, where the approach of partially-data-dependent superimposed training is also proposed. Our approaches are demonstrated by simulation examples in Section 6.4, and Section 6.5 concludes this chapter.

6.2 Data-Dependent Superimposed Training Using CE-BEM

We assume:

(H6.2.1) The time-varying channel satisfies CE-BEM (3.3) where the frequencies ω_q 's ($q = 1, 2, \dots, Q$) are distinct and known with $\omega_q \in [0, 2\pi)$. Also $N \geq 1$.

(H6.2.2) The information data sequence $\{b(n)\}$ is zero-mean and white, with the variance

$$E\{|b(n)|^2\} = \sigma_b^2;$$

(H6.2.3) The measurement noise $\{\mathbf{v}(n)\}$ is zero-mean, white, and uncorrelated with $\{b(n)\}$,

$$\text{with the autocorrelation } E\{\mathbf{v}(n+\tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau);$$

(H6.2.4) The superimposed training sequence $c(n) = c(n+P)$ for all n is a non-random

$$\text{periodic sequence with period } P \text{ and average power } \sigma_c^2 := \sum_{n=0}^{P-1} |c(n)|^2 / P.$$

6.2.1 Data-Dependent Processing at the Transmitter

Consider the DFT of information sequence $\{b(n)\}$ over the block $n = 0, 1, \dots, T-1$,

$$b_r := \frac{1}{T} \sum_{n=0}^{T-1} b(n) e^{-j\omega_r n}, \quad \omega_r := \frac{2\pi r}{T}, \quad (6.3)$$

for $r = 0, 1, \dots, T-1$ and $b(n) = \sum_{r=0}^{T-1} b_r e^{j\omega_r n}$. Then the interference \mathbf{s}_{mq} of (6.1) can be expressed as

$$\begin{aligned} \mathbf{s}_{mq} &= \frac{1}{T} \sum_{n=0}^{T-1} \left\{ \sum_{q_1=1}^Q \sum_{l=0}^L \mathbf{h}_{q_1}(l) e^{j\omega_{q_1} n} \sum_{r=0}^{T-1} b_r e^{j\omega_r (n-l)} \right\} e^{-j(\omega_q + \alpha_m)n} \\ &= \sum_{q_1=1}^Q \sum_{l=0}^L \sum_{r=0}^{T-1} \left[\mathbf{h}_{q_1}(l) e^{-j\omega_r l} b_r \right] \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{j(\omega_{q_1} - \omega_q + \omega_r - \alpha_m)n} \right] \\ &= \sum_{q_1=1}^Q \sum_{l=0}^L \sum_{r=0}^{T-1} \left[\mathbf{h}_{q_1}(l) e^{-j\omega_r l} b_r \right] \delta((q_1 - q + r - mK) \bmod T). \end{aligned}$$

Therefore, if we can make $b_r = 0$ for $r = q + mK - q_1$, $1 \leq q, q_1 \leq Q$, $m = 0, 1, \dots, P-1$, then $\mathbf{s}_{mq} = \mathbf{0}$. We do so by modifying $\{c(n)\}$ based on $\{b(n)\}$ (at the transmitter).

Define a set

$$\Omega := \{r : -(Q-1) + mK \leq r \leq (Q-1) + mK, m = 0, 1, \dots, P-1\}. \quad (6.4)$$

The frequency components $\{b_r : r \in \Omega\}$ of the information sequence are hence the “self-interference”. Define a “self-interference” sequence

$$b_e(n) := \sum_{r \in \Omega} b_r e^{j \frac{2\pi r n}{T}} \quad (6.5)$$

and a data-dependent superimposed training $\tilde{c}(n)$ over the block $n = 0, 1, \dots, T-1$ such that

$$\tilde{c}(n) := c(n) - b_e(n). \quad (6.6)$$

Note that $\{\tilde{c}(n)\}$ is no longer periodic with period P . At the transmitter, we transmit

$$\tilde{c}(n) + b(n) = c(n) + [b(n) - b_e(n)].$$

The model (3.1)–(3.4) holds with $c(n)$ replaced with $\tilde{c}(n)$. By construction, the DFT of $b(n) - b_e(n)$ over the block $n = 0, 1, \dots, T-1$ vanishes at frequencies in the set Ω . Also the DFT of $b(n-l) - b_e(n-l)$ over the block $n = 0, 1, \dots, T-1$ vanishes at frequencies in the set Ω provided that a cyclic prefix of length $M \geq L$ is used. A cyclic prefix of length M is added at the transmitter by choosing

$$s(-i) = s(T-i), \quad i = 1, 2, \dots, M \geq L$$

where $s(i) = \tilde{c}(i) + b(i)$. This allows linear convolution in (3.1) to be equal to circular convolution (implicit in the DFT operation) over the block length $n = 0, 1, \dots, T - 1$.

We summarize our data-dependent channel estimation solution as follows:

1. At the transmitter, we are given information sequence over a block as $\{b(n)\}$ for $n = 0, 1, \dots, T - 1$ with T chosen as $T = KP$, $K \geq Q$. Calculate the DFT by (6.3).
2. To eliminate interference with channel estimation at the receiver, we need to set b_r 's to be zero for $r \in \Omega$. Define the self-interference sequence $\{b_e(n)\}$ as in (6.5).
3. Define the data-dependent superimposed training $\tilde{c}(n)$ as in (6.6). Use a cyclic prefix of length $M \geq L$ and transmit.

The channel estimation given in (3.18) stays the same for data-dependent superimposed training, because we still use periodic $\{c(n)\}$ at the receiver, and we do not know $b_e(n)$ or $b(n)$ at the receiver. It is easily established that now there is no contribution of $\{b(n)\}$ to $\hat{\mathbf{d}}_{mq}$ for $0 \leq m \leq P - 1$ and $1 \leq q \leq Q$.

6.2.2 Data Detection

Now the “information sequence” is $\{b(n) - b_e(n)\}$ whereas we are interested in $\{b(n)\}$. We will follow an iterative solution, similar to the time-invariant results of [20]. The first step in our solution is to use the estimated channel to detect $\{b(n)\}$ via Viterbi algorithm (ignoring $b_e(n)$ but accounting for the known $\{c(n)\}$). Use the detected $\{b(n)\}$ to estimate $\{b_e(n)\}$, and iterate the detection procedure (but not channel estimation) with known $\{c(n)\}$ and estimate $\{\hat{b}_e(n)\}$ from the previous iteration. Note that although iterations

are also employed, as the DML algorithm, we do not have to re-estimate the channel in the subsequent iterations.

6.2.3 Performance Analysis

If the true channel follows the CE-BEM representation (3.3), the MSE in channel estimation is then given by (4.2). We now relax the assumption **(H6.2.3)**, i.e., the measurement noise $\{\mathbf{v}(n)\}$ may be nonzero-mean, as in **(H3.2.3)**. We further assume that

(H6.2.5) The time-varying channel $\{\mathbf{h}(n;l)\}$ is zero-mean, complex Gaussian with variance σ_h^2 , and mutually independent for distinct l 's: $E\{\mathbf{h}(n;l)\mathbf{h}^H(n;l)\} = \sigma_h^2\mathbf{I}_N$ and $E\{\mathbf{h}(n_1;l_1)\mathbf{h}^H(n_2;l_2)\} = \mathbf{0}$, for $l_1 \neq l_2$, for all n_1, n_2 , i.e., different channel taps are independent of each other and are identically distributed zero-mean complex Gaussian.

In the data-dependent superimposed training, the interference from the information sequence $\{b(n)\}$ has been canceled out, so $\mathbf{s}_{mq} = \mathbf{0}$. Then by (4.11), $\hat{\mathbf{d}}_{mq} = \mathbf{d}_{mq} + \mathbf{w}_{mq}$, so that

$$\begin{aligned} E\left\{[\hat{\mathbf{d}}_{m_1q_1} - \mathbf{d}_{m_1q_1}][\hat{\mathbf{d}}_{m_2q_2} - \mathbf{d}_{m_2q_2}]^H\right\} &= E\left\{\mathbf{w}_{m_1q_1}\mathbf{w}_{m_2q_2}^H\right\} \\ &= \frac{1}{T}\sigma_v^2\mathbf{I}_N\delta(m_1 - m_2)\delta(q_1 - q_2). \end{aligned}$$

By (3.11) and (3.16), it follows that

$$\text{cov}\{\hat{\mathcal{D}}, \hat{\mathcal{D}}\} = \frac{1}{T}\sigma_v^2\mathbf{I}_{NQ(P-1)}. \quad (6.7)$$

By (3.18), we also have

$$\begin{aligned} \text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} \right\} &:= E \left\{ \left(\hat{\mathcal{H}} - \mathcal{H} \right) \left(\hat{\mathcal{H}} - \mathcal{H} \right)^H \right\} \\ &= (\mathcal{C}^H \mathcal{C})^{-1} \mathcal{C}^H \text{cov} \left\{ \hat{\mathcal{D}}, \hat{\mathcal{D}} \right\} \mathcal{C} (\mathcal{C}^H \mathcal{C})^{-1}. \end{aligned} \quad (6.8)$$

Substitute (6.7) into (6.8), we have

$$\text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} \right\} = \frac{1}{T} \sigma_v^2 (\mathcal{C}^H \mathcal{C})^{-1}.$$

In a manner similar to (4.16), it then follows that

$$\begin{aligned} \text{MSE}_1 &= E \left\{ \sum_{l=0}^L \sum_{q=1}^Q \left\| \mathbf{h}_q(l) - \hat{\mathbf{h}}_q(l) \right\|^2 \right\} \\ &= \frac{NQ\sigma_v^2}{T} \text{tr} \left\{ \left(\mathbf{V}^H \text{diag} \left\{ |c_1|^2, |c_2|^2, \dots, |c_{P-1}|^2 \right\} \mathbf{V} \right)^{-1} \right\}. \end{aligned}$$

We note that the real channel over a block cannot be exactly equal to the CE-BEM representation. Counting the modeling error, the total channel MSE can be expressed by (4.34)

$$\text{MSE}_c = \text{MSE}_1 + \text{MSE}_2,$$

where MSE_1 comes from the estimation and MSE_2 is the mean square modeling error. Under the assumption **(H4.2.5)** and Jakes' model, the modeling error of CE-BEM is given by (4.35).

6.3 Data-Dependent Superimposed Training Using DPS-BEM

Exploiting the fact that DPS sequences are also approximately band-limited, in this section we extend the data-dependent superimposed training to DPS-BEM, so that spectral leakage arising from CE-BEM is efficiently reduced.

We follow the assumptions:

- (H6.3.1) The time-varying channel $\{\mathbf{h}(n; l)\}$ satisfies (3.25) with the DPS sequences $\{u_q(n)\}$ known at the receiver. Also $N \geq 1$.
- (H6.3.2) The information sequence $\{b(n)\}$ is zero-mean, white with $E\{|b(n)|^2\} = \sigma_b^2$.
- (H6.3.3) The measurement noise $\{\mathbf{v}(n)\}$ is zero-mean, white, uncorrelated with $\{b(n)\}$, with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$.
- (H6.3.4) The superimposed training sequence $c(n) = c(n+P)$ for all n is a non-random periodic sequence with period P .

Consider the interference \mathbf{s}_{mq} in (6.2), which can be expressed by using (6.3) as

$$\begin{aligned}
 \mathbf{s}_{mq} &= \sum_{n=0}^{T-1} \left[\sum_{l=0}^L \mathbf{h}(n; l) b(n-l) \right] u_q(n) e^{-j\alpha_m n} \\
 &= \sum_{n=0}^{T-1} \left\{ \sum_{q_1=1}^Q \sum_{l=0}^L \mathbf{h}_{q_1}(l) u_{q_1}(n) \sum_{r=0}^{T-1} b_r e^{j\omega_r(n-l)} \right\} u_q(n) e^{-j\alpha_m n} \\
 &= \sum_{q_1=1}^Q \sum_{l=0}^L \sum_{r=0}^{T-1} \left[\mathbf{h}_{q_1}(l) e^{-j\omega_r l} b_r \right] \left[\sum_{n=0}^{T-1} u_{q_1}(n) u_q(n) e^{j\frac{2\pi(r-mK)}{T}n} \right].
 \end{aligned}$$

We exploit the approximate band-limitedness of the time-limited DPS sequences, assuming that

$$\sum_{n=0}^{T-1} u_{q'}(n) u_q(n) e^{j2\pi(r-mK)n/T} \approx 0$$

for $|r - mK| \geq Q + k$, where k is an integer and $(Q + k)/T > 2f_d T_s$ ($k \geq -1$). Therefore, the information-induced interference comes from the frequency components b_r 's for those r 's belonging to a set

$$\Omega := \{r : -(Q + k) + mK \leq r \leq (Q + k) + mK, m = 0, \dots, P - 1\}. \quad (6.9)$$

The frequency components $\{b_r : r \in \Omega\}$ of the information sequence are hence the self-interference. If we set $b_r = 0$ for those $r \in \Omega$, then $\mathbf{s}_{mq} = \mathbf{0}$. We do so by modifying $\{c(n)\}$ based on $\{b(n)\}$ at the transmitter. Define a self-interference sequence

$$b_e(n) := \sum_{r=0, r \in \Omega}^{T-1} b_r e^{j\omega_r n} \quad (6.10)$$

and a data-dependent superimposed training sequence $\{\tilde{c}(n)\}_{n=0}^{T-1}$ such that

$$\tilde{c}(n) := c(n) - b_e(n).$$

All the other steps for the DPS-BEM-based channel estimator using data-dependent superimposed training just follow the steps described in Section 6.2.1. For symbol detection, we can also follow the iterative approach addressed in Section 6.2.2.

6.3.1 Partially-Data-Dependent (PDD) Superimposed Training

In the data-dependent superimposed training method, by setting $b_r = 0$ for $r \in \Omega$ before transmission, we discard the “frequency components” of the information sequence corresponding to Ω , so that information-induced self-interference is eliminated. By (6.9), however, the information contained at those $P(2Q + 2k + 1)$ (among total T) frequencies is also discarded. Though it may be partially recovered by other properties (e.g., the finite alphabet of the information sequence [22]), it can cause severe detection errors under severe frequency loss.

For the first-order statistics-based estimator we discussed in Section 3.3, the sequence $\{b_e(n)\}$ acts as self-interference in channel estimation, whereas it also bears “information” so that it should not be totally discarded. Given $\{b_e(n)\}$ as in (6.10), we now transmit

$$s(n) = c(n) + b(n) - (1 - \gamma) b_e(n) \quad (6.11)$$

at the transmitter, where $0 \leq \gamma \leq 1$ is the self-interference factor. When $\gamma = 1$, the information sequence $\{b(n)\}$ keeps intact, corresponding to the first-order statistics-based estimator in Section 3.3. When $\gamma = 0$, the interference-induced frequency components b_r 's ($r \in \Omega$) are totally annihilated, corresponding to the data-dependent solution described in Section 6.2 and Section 6.3. If $0 < \gamma \ll 1$, which is the partially-data-dependent (PDD) case, at each $r \in \Omega$ the frequency component b_r is reduced to γb_r . Then the self-interference will be “sufficiently” suppressed when conducting channel estimation, while the frequency components at $r \in \Omega$ remain “partially” intact. Note that in this PDD method, the self-interference is not completely removed, so that the channel estimation is not as accurate

as that of the data-dependent solution. But since no information-bearing frequencies are nulled out, the information contained there can be recovered in data reception.

Our PDD superimposed training-based channel estimation follows the data-dependent solution described in Section 6.2, except that we define the PDD superimposed training sequence as

$$\tilde{c}(n) := c(n) - (1 - \gamma) b_e(n)$$

instead of (6.6) with a prescribed interference factor γ .

For data detection at the receiver, the “information sequence” is $\{b(n) - (1 - \gamma) b_e(n)\}$ ($0 \leq \gamma \leq 1$), while we are interested in $\{b(n)\}$. We can first use the estimated channel to detect $\{b(n)\}$ via Viterbi algorithm (ignoring $(1 - \gamma) b_e(n)$ but accounting for the known $\{c(n)\}$). Since the training and information sequences pass through an identical channel, the DML approach described in Section 5.2 can be exploited to recover the suppressed frequency components b_r 's ($r \in \Omega$), as well as enhance the channel estimation, in an iterative way (see Section 6.3.4).

6.3.2 Performance Analysis

We wish to evaluate the MSE in channel estimation, defined in (4.2), using PDD superimposed training when the true channel follows DPS-BEM. We make the following assumption about the channel $\mathbf{h}(n; l)$:

(H6.3.5) The time-varying channel $\{\mathbf{h}(n; l)\}$ is zero-mean, complex Gaussian with variance σ_h^2 , and mutually independent for distinct l 's: $E\{\mathbf{h}(n; l) \mathbf{h}^H(n; l)\} = \sigma_h^2 \mathbf{I}_N$ and $E\{\mathbf{h}(n_1; l_1) \mathbf{h}^H(n_2; l_2)\} = \mathbf{0}$, for $l_1 \neq l_2$, for all n_1, n_2 , i.e., different channel taps are

independent of each other and are identically distributed zero-mean complex Gaussian.

Note that in PDD superimposed training, (4.11) still holds if \mathbf{s}_{mq} is revised as

$$\mathbf{s}_{mq} := \sum_{n=0}^{T-1} \left\{ \sum_{l=0}^L \mathbf{h}(n; l) [b(n-l) - (1-\gamma)b_e(n)] \right\} u_q(n) e^{-j\alpha_m n}.$$

By (3.34) and (6.10), we have

$$E \{ \mathbf{s}_{m'q'} \mathbf{s}_{mq}^H \} = \gamma^2 (L+1) \sigma_h^2 \sigma_b^2 \mathbf{I}_N \delta(m' - m) \delta(q' - q). \quad (6.12)$$

Then from (3.36)

$$\text{cov}\{\hat{\mathcal{H}}, \hat{\mathcal{H}}\} := E\{[\hat{\mathcal{H}} - \mathcal{H}][\hat{\mathcal{H}} - \mathcal{H}]^H\} = (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1} \tilde{\mathcal{C}}^H \text{cov}\{\hat{\mathcal{D}}, \hat{\mathcal{D}}\} \tilde{\mathcal{C}} (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1}. \quad (6.13)$$

Since

$$E \left\{ [\hat{\mathbf{d}}_{mq} - \mathbf{d}_{mq}] [\hat{\mathbf{d}}_{m'q'} - \mathbf{d}_{m'q'}]^H \right\} = E \{ \mathbf{s}_{m'q'} \mathbf{s}_{mq}^H \} + E \{ \mathbf{w}_{m'q'} \mathbf{w}_{mq}^H \},$$

by (4.10) and (6.12) we have

$$\text{cov}\{\hat{\mathcal{D}}, \hat{\mathcal{D}}\} = [\gamma^2 (L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2] \mathbf{I}_{NPQ}. \quad (6.14)$$

Substitute (6.14) for (6.13)

$$\text{cov}\{\hat{\mathcal{H}}, \hat{\mathcal{H}}\} = [\gamma^2 (L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2] (\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1}.$$

Using orthonormality of DPS sequences, the MSE in channel estimation is given by

$$\begin{aligned} \text{MSE}_1 &= \frac{1}{T} E \left\{ \sum_{l=0}^L \sum_{q=1}^Q \left\| \mathbf{h}_q(l) - \hat{\mathbf{h}}_q(l) \right\|^2 \right\} = \frac{1}{T} \text{tr}\{\text{cov}\{\hat{\mathcal{H}}, \hat{\mathcal{H}}\}\} \\ &= \frac{\gamma^2 (L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2}{T} \text{tr}\{(\tilde{\mathcal{C}}^H \tilde{\mathcal{C}})^{-1}\}. \end{aligned} \quad (6.15)$$

The MSE of the first-order statistics-based channel estimator in Section 3.3 is given by (6.15) for $\gamma = 1$, and that of the data-dependent solution in Section 6.3 is corresponding to $\gamma = 0$. For a PDD scheme with $0 < \gamma \ll 1$, the interference is significantly suppressed.

6.3.3 Power Allocation and Self-Interference Suppression

We consider the issues of superimposed training power allocation and self-interference suppression under the channel assumption (**H6.3.5**). Using the channel estimation variance developed in (6.15), we cast power allocation and self-interference suppression as jointly optimizing an SNR for equalizer design, as we discussed in Section 4.4. Since the channel estimate is used for equalizer design, we set up a model for the received signal, in which an estimation error-related term and the additive noise act as effective noise, while the effective signal is given by the output of the estimated channel driven by the information sequence. This SNR is maximized under a transmission power constraint.

Removing the estimated time-varying mean from the received data, define

$$\tilde{\mathbf{y}}(n) := \mathbf{y}(n) - \sum_{l=0}^L \hat{\mathbf{h}}(n; l) c(n-l).$$

We then have

$$\begin{aligned}
\tilde{\mathbf{y}}(n) &= \underbrace{\sum_{l=0}^L \hat{\mathbf{h}}(n;l) b(n-l)}_{=:\mathbf{x}_s(n)} + \\
&\quad \underbrace{\sum_{l=0}^L [\mathbf{h}(n;l) - \hat{\mathbf{h}}(n;l)] [b(n-l) + c(n-l)] - \sum_{l=0}^L \mathbf{h}(n;l) (1-\gamma) b_e(n-l) + \mathbf{v}(n)}_{=:\mathbf{w}(n)}.
\end{aligned} \tag{6.16}$$

The power of the “signal” part $\mathbf{x}_s(n)$ at time n is

$$E \left\{ \|\mathbf{x}_s(n)\|^2 \right\} = \sigma_b^2 \sum_{l=0}^L E \left\{ \left\| \hat{\mathbf{h}}(n;l) \right\|^2 \right\} + O(1/T^2) \tag{6.17}$$

where $O(1/T^2)$ term accounts for dependence between $\hat{\mathbf{h}}_q(l)$ and $\{b(n)\}$ (see the Appendix of [32] for details). Furthermore,

$$\begin{aligned}
\sum_{l=0}^L E \left\{ \left\| \hat{\mathbf{h}}(n;l) \right\|^2 \right\} &= \sum_{l=0}^L \left[E_{\mathcal{H}} \left\{ E \left\{ \left\| \hat{\mathbf{h}}(n;l) - \mathbf{h}(n;l) \right\|^2 \middle| \mathcal{H} \right\} \right\} + E \left\{ \|\mathbf{h}(n;l)\|^2 \right\} \right] \\
&= \text{MSE} + N(L+1) \sigma_h^2.
\end{aligned} \tag{6.18}$$

Taking the time-average of $E \left\{ \|\mathbf{x}_s(n)\|^2 \right\}$, by (6.17) and (6.18) we obtain (neglecting $O(1/T^2)$ term)

$$\bar{\sigma}_{x_s}^2 := \frac{1}{T} \sum_{n=0}^{T-1} E \left\{ \|\mathbf{x}_s(n)\|^2 \right\} = \sigma_b^2 [\text{MSE}_1 + N(L+1) \sigma_h^2]. \tag{6.19}$$

The time-averaged “noise” power in (6.16) is given by

$$\bar{\sigma}_w^2 := \frac{1}{T} \sum_{n=0}^{T-1} E \left\{ \|\mathbf{w}(n)\|^2 \right\} = \sigma_b^2 \text{MSE}_1 + N\sigma_v^2 + \mathcal{B} + \mathcal{R} - \mathcal{S}_1 - \mathcal{S}_2 + O(1/T^2),$$

where

$$\mathcal{B} = \frac{(1-\gamma)^2}{T} \sum_{n=0}^{T-1} \sum_{l_1=0}^L \sum_{l_2=0}^L E \left\{ \mathbf{h}^H(n; l_1) \mathbf{h}(n; l_2) \right\} E \left\{ b_e^*(n-l_1) b_e(n-l_2) \right\};$$

$$\mathcal{R} = \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l_1=0}^L \sum_{l_2=0}^L E \left\{ \left[\hat{\mathbf{h}}(n; l_1) - \mathbf{h}(n; l_1) \right]^H \left[\hat{\mathbf{h}}(n; l_2) - \mathbf{h}(n; l_2) \right] \right\}$$

$$\times c(n-l_1)^* c(n-l_2);$$

$$\mathcal{S}_1 = \frac{1-\gamma}{T} \sum_{n=0}^{T-1} \sum_{l_1=0}^L \sum_{l_2=0}^L E \left\{ \left[\mathbf{h}(n; l_1) - \hat{\mathbf{h}}(n; l_1) \right]^H \mathbf{h}(n; l_2) \right\} E \left\{ b^*(n-l_1) b_e(n-l_2) \right\};$$

$$\mathcal{S}_2 = \mathcal{S}_1^H.$$

Since

$$\begin{aligned} E \left\{ \left[\mathbf{h}(n; l_1) - \hat{\mathbf{h}}(n; l_1) \right]^H \mathbf{h}(n; l_2) \right\} &= E_{\mathcal{H}} \left\{ E \left\{ \left[\mathbf{h}(n; l_1) - \hat{\mathbf{h}}(n; l_1) \right]^H \mathbf{h}(n; l_2) \middle| \mathcal{H} \right\} \right\} \\ &= E_{\mathcal{H}} \left\{ \left[\mathbf{h}(n; l_1) - \mathbf{h}(n; l_1) \right]^H \mathbf{h}(n; l_2) \right\} = 0, \end{aligned}$$

we have $\mathcal{S}_1 = \mathcal{S}_2^H = 0$.

Note that $|\Omega| = P(2Q + 2k + 1)$. Consider

$$\begin{aligned}
E \{b_e(n-l) b_e^*(n-l)\} &= E \left\{ \sum_{r_1 \in \Omega} \sum_{r_2 \in \Omega} b_{r_1} b_{r_2}^* e^{j \frac{2\pi}{T} (r_1 - r_2)(n-l)} \right\} \\
&= \frac{1}{T^2} E \left\{ \sum_{r_1 \in \Omega} \sum_{r_2 \in \Omega} \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} b(n_1) b^*(n_2) e^{-j \frac{2\pi}{T} r_1 n_1} e^{j \frac{2\pi}{T} r_2 n_2} e^{j \frac{2\pi}{T} (r_1 - r_2)(n-l)} \right\} \\
&= \frac{\sigma_b^2}{T^2} \sum_{r_1 \in \Omega} \sum_{r_2 \in \Omega} \sum_{n_1=0}^{T-1} e^{-j \frac{2\pi}{T} (r_1 - r_2) n_1} e^{j \frac{2\pi}{T} (r_1 - r_2)(n-l)} = \frac{P(2Q + 2k + 1) \sigma_b^2}{T}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathcal{B} &= \frac{(1-\gamma)^2}{T} \sum_{n=0}^{T-1} \sum_{l_1=0}^L \sum_{l_2=0}^L N \sigma_h^2 \delta(l_1 - l_2) E \{b_e^*(n-l_1) b_e(n-l_2)\} \\
&= \frac{N(L+1)P}{T} (2Q + 2k + 1) (1-\gamma)^2 \sigma_h^2 \sigma_b^2.
\end{aligned}$$

Now we consider \mathcal{R} ,

$$\begin{aligned}
\mathcal{R} &= \frac{1}{T} \sum_{l_1=0}^L \sum_{l_2=0}^L \sum_{q_1=1}^Q \sum_{q_2=1}^Q E \left\{ \left[\hat{\mathbf{h}}_{q_1}^H(l_1) - \mathbf{h}_{q_1}^H(l_1) \right] \left[\hat{\mathbf{h}}_{q_2}(l_2) - \mathbf{h}_{q_2}(l_2) \right] \right\} \\
&\quad \times \sum_{m_1=0}^{P-1} \sum_{m_2=0}^{P-1} c_{m_1}^* c_{m_2} e^{j(\alpha_{m_1} l_1 - \alpha_{m_2} l_2)} \left[\sum_{n=0}^{T-1} u_{q_1}(n) u_{q_2}(n) e^{j(\alpha_{m_2} - \alpha_{m_1})n} \right] \\
&= \frac{1}{T} \sum_{l_1=0}^L \sum_{l_2=0}^L \sum_{q_1=1}^Q \sum_{q_2=1}^Q E \left\{ \left[\hat{\mathbf{h}}_{q_1}^H(l_1) - \mathbf{h}_{q_1}^H(l_1) \right] \left[\hat{\mathbf{h}}_{q_2}(l_2) - \mathbf{h}_{q_2}(l_2) \right] \right\} \\
&\quad \times \sum_{m_1=0}^{P-1} \sum_{m_2=0}^{P-1} c_{m_1}^* c_{m_2} e^{j(\alpha_{m_1} l_1 - \alpha_{m_2} l_2)} \delta(q_1 - q_2) \delta(m_1 - m_2) \\
&= \frac{1}{T} E \left\{ \left[\sum_{l=0}^L (\hat{\mathbf{H}}_l - \mathbf{H}_l) e^{-j\alpha_m l} \right]^H \left[\sum_{l=0}^L (\hat{\mathbf{H}}_l - \mathbf{H}_l) e^{-j\alpha_m l} \right] \right\} \sum_{m=0}^{P-1} |c_m|^2.
\end{aligned}$$

Define $\mathcal{E}(m) := \begin{bmatrix} 1 & e^{-j\alpha m} & \dots & e^{-j\alpha m L} \end{bmatrix}$, then \mathcal{R} becomes

$$\begin{aligned} \mathcal{R} &= \frac{1}{T} E \left\{ \left[\mathcal{E}(m) \otimes \mathbf{I}_{NQ} (\hat{\mathcal{H}} - \mathcal{H}) \right]^H \left[\mathcal{E}(m) \otimes \mathbf{I}_{NQ} (\hat{\mathcal{H}} - \mathcal{H}) \right] \right\} \sum_{m=0}^{P-1} |c_m|^2 \\ &= \frac{1}{T} E \left\{ (\hat{\mathcal{H}} - \mathcal{H})^H \mathcal{C}^H \mathcal{C} (\hat{\mathcal{H}} - \mathcal{H}) \right\} = \frac{1}{T} \text{tr} \left\{ \text{cov} \left\{ \hat{\mathcal{H}}, \hat{\mathcal{H}} \right\} \mathcal{C}^H \mathcal{C} \right\} \\ &= \frac{NQ(L+1)}{T} [\gamma^2 (L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2]. \end{aligned}$$

Using these expressions we obtain (also neglecting $O(1/T^2)$ term)

$$\begin{aligned} \bar{\sigma}_w^2 &= \sigma_b^2 \text{MSE} + N\sigma_v^2 + \frac{N(L+1)P}{T} (2Q + 2k + 1) (1 - \gamma)^2 \sigma_h^2 \sigma_b^2 \\ &\quad + \frac{NQ(L+1)}{T} [\gamma^2 (L+1) \sigma_h^2 \sigma_b^2 + \sigma_v^2]. \end{aligned} \quad (6.20)$$

We define the training power overhead β as

$$\beta := \frac{\frac{1}{P} \sum_{n=0}^{P-1} |c(n)|^2}{\frac{1}{P} \sum_{n=0}^{P-1} E \left\{ |b(n) + c(n)|^2 \right\}} = \frac{\sigma_c^2}{\sigma_b^2 + \sigma_c^2}. \quad (6.21)$$

For a fixed SNR or transmitted power budget, higher β implies smaller effective SNR at the receiver due to decreased power in the information sequence but higher channel estimation accuracy; similarly, higher γ in (6.11) implies less ‘‘information’’ impairment but higher self-interference remaining. A trade-off must be considered in choosing appropriate β and γ for the PDD superimposed training.

The equalization SNR of (6.16), as a function of β and γ , is (implicitly) obtained as

$$\text{SNR}_d(\beta, \gamma) := \frac{\bar{\sigma}_{x_s}^2(\beta, \gamma)}{\bar{\sigma}_w^2(\beta, \gamma)}. \quad (6.22)$$

Our objective is to maximize $\text{SNR}_d(\beta, \gamma)$ with respect to β and γ under the constraint of a fixed power: $P_T := \sigma_b^2 + \sigma_c^2$ is fixed. Then $\sigma_c^2 = P_T \beta$ and $\sigma_b^2 = P_T(1 - \beta)$. Incorporating these constraint-carrying variables in (6.22) via (6.19) and (6.20), we obtain an unconstrained cost

$$\text{SNR}_d(\beta, \gamma) = \frac{f_1 \beta^2 + f_2 \beta + f_3}{g_1 \beta^2 + g_2 \beta + g_3}$$

where with $\bar{\mathcal{C}} := \sigma_c^{-1} \tilde{\mathcal{C}}$

$$\begin{aligned} f_1 &= (L+1) \sigma_h^2 P_T \left[\gamma^2 \text{tr}\{(\bar{\mathcal{C}}^H \bar{\mathcal{C}})^{-1}\} - NT \right], \\ f_2 &= N(L+1) \sigma_h^2 P_T T - [2\gamma^2 (L+1) \sigma_h^2 P_T + \sigma_v^2] \text{tr}\{(\bar{\mathcal{C}}^H \bar{\mathcal{C}})^{-1}\}, \\ f_3 &= [\gamma^2 (L+1) \sigma_h^2 P_T + \sigma_v^2] \text{tr}\{(\bar{\mathcal{C}}^H \bar{\mathcal{C}})^{-1}\}, \end{aligned}$$

$$\begin{aligned} g_1 &= (L+1) \sigma_h^2 P_T \left[\gamma^2 \text{tr}\{(\bar{\mathcal{C}}^H \bar{\mathcal{C}})^{-1}\} - NP(2Q + 2k + 1)(1 - \gamma)^2 - N(L+1)Q\gamma^2 \right], \\ g_2 &= -[2\gamma^2 (L+1) \sigma_h^2 P_T + \sigma_v^2] \text{tr}\{(\bar{\mathcal{C}}^H \bar{\mathcal{C}})^{-1}\} \\ &\quad + N(L+1) \sigma_h^2 P_T \left[P(2Q + 2k + 1)(1 - \gamma)^2 + (L+1)Q\gamma^2 \right] + N\sigma_v^2 [Q(L+1) + T], \\ g_3 &= f_3. \end{aligned}$$

Similarly, this unconstrained cost can also be expressed as

$$\text{SNR}_d(\beta, \gamma) = \frac{f'_1 \gamma^2 + f'_2 \gamma + f'_3}{g'_1 \gamma^2 + g'_2 \gamma + g'_3}$$

where

$$\begin{aligned}
f'_1 &= (L+1) \sigma_h^2 P_T (1-\beta)^2 \text{tr}\{(\bar{\mathcal{C}}^H \bar{\mathcal{C}})^{-1}\}, \quad f'_2 = 0, \\
f'_3 &= NT(L+1) \sigma_h^2 P_T \beta (1-\beta) + \sigma_v^2 (1-\beta) \text{tr}\{(\bar{\mathcal{C}}^H \bar{\mathcal{C}})^{-1}\}, \\
g'_1 &= (L+1) \sigma_h^2 P_T (1-\beta) \\
&\quad \times [(1-\beta) \text{tr}\{(\bar{\mathcal{C}}^H \bar{\mathcal{C}})^{-1}\} + NP(2Q+2k+1)\beta + NQ(L+1)\beta], \\
g'_2 &= -2N(L+1) \sigma_h^2 P(2Q+2k+1) P_T (1-\beta) \beta, \\
g'_3 &= \sigma_v^2 \text{tr}\{(\bar{\mathcal{C}}^H \bar{\mathcal{C}})^{-1}\} (1-\beta) + N(L+1) \sigma_h^2 P_T P(2Q+2k+1) (1-\beta) \beta \\
&\quad + N\sigma_v^2 \beta [Q(L+1) + T].
\end{aligned}$$

We seek the optimum value of β and γ by setting the partial derivatives of the unconstrained cost (6.22) to zero:

$$\frac{\partial [\text{SNR}_d(\beta, \gamma)]}{\partial \beta} = 0, \text{ and } \frac{\partial [\text{SNR}_d(\beta, \gamma)]}{\partial \gamma} = 0.$$

The above quadratic in β has two roots, of which the root lying in $[0, 1]$ is

$$\begin{aligned}
\beta^o(\gamma) &= (f_1 g_2 - f_2 g_1)^{-1} (f_3 g_1 - f_1 g_3 \\
&\quad - \sqrt{-f_1 f_2 g_2 g_3 - 2f_1 f_3 g_1 g_3 - f_2 f_3 g_1 g_2 + f_2^2 g_1 g_3 + f_1 f_3 g_2^2 + f_1^2 g_3^2 + f_3^2 g_1^2}). \quad (6.23)
\end{aligned}$$

Note that this optimum β is a function of γ . The optimum γ , which is a function of β likewise, can be acquired in the same way as in (6.23) with f_i and g_i ($i = 1, 2, 3$) replaced

with f'_i and g'_i :

$$\begin{aligned} \gamma^o(\beta) = & (f'_1 g'_2 - f'_2 g'_1)^{-1} (f'_3 g'_1 - f'_1 g'_3 \\ & - \sqrt{-f'_1 f'_2 g'_2 g'_3 - 2f'_1 f'_3 g'_1 g'_3 - f'_2 f'_3 g'_1 g'_2 + f'^2_2 g'_1 g'_3 + f'_1 f'_3 g'^2_2 + f'^2_1 g'^2_3 + f'^2_3 g'^2_1}). \end{aligned} \quad (6.24)$$

The joint optimization of β and γ can be achieved by applying (6.23) and (6.24) iteratively: The result of (6.23) is substituted in (6.24) and vice-versa; there is no guarantee of global optimization. An alternative is to do a two-dimensional grid search for joint optimization of (6.22) with respect to β and γ , both restricted to $[0, 1]$, to obtain a “coarse” optimization, and then follow up with a “fine” optimization via iterative computation of (6.23) and (6.24).

6.3.4 Recovery of Suppressed Frequencies via DML Approach

We now consider joint channel and information sequence estimation via an iterative DML approach assuming that the noise $\mathbf{v}(n)$ is complex Gaussian. As noted in Section 6.3.1, for data detection at the receiver, now the “information sequence” is $\{b(n) - (1 - \gamma)b_e(n)\}$ ($0 \leq \gamma \leq 1$), while we are interested in $\{b(n)\}$. We can first use the estimated channel to detect $\{b(n)\}$ via Viterbi algorithm (ignoring $(1 - \gamma)b_e(n)$ but accounting for the known $\{c(n)\}$). Since the training and information sequences pass through an identical channel, this fact can be exploited to recover the suppressed frequency components b_r 's ($r \in \Omega$), as well as enhance the channel estimation, in an iterative way. In subsequent iterations, we regenerate $b_e(n)$ as $\hat{b}_e(n)$ using the detected $b(n)$ from the previous iteration, and so on.

To consider a general case, we assume the measurement noise $\mathbf{v}(n)$ is complex Gaussian, but may be nonzero-mean as in Section 5.2 ($E\{\mathbf{v}(n)\} = \mathbf{m}$, with \mathbf{m} unknown). We define

$$\mathbf{Y} := \begin{bmatrix} \mathbf{y}^T(T-1) & \mathbf{y}^T(T-2) & \cdots & \mathbf{y}^T(L) \end{bmatrix}^T,$$

$$\mathbf{s} := \begin{bmatrix} s(T-1) & s(T-2) & \cdots & s(0) \end{bmatrix}^T, \quad (6.25)$$

and similarly define \mathbf{b} , \mathbf{b}_e , and \mathbf{c} from $b(n)$, $b_e(n)$, and $c(n)$, respectively, following (6.25).

Therefore,

$$\mathbf{s} := \mathbf{b} + \mathbf{c} + (1 - \gamma) \mathbf{b}_e.$$

Let $\tilde{\mathbf{v}}(n) := \mathbf{v}(n) - \mathbf{m}$. We then have the linear model

$$\mathbf{Y} = \mathcal{T}(\mathbf{s}) \mathcal{H} + \underbrace{\begin{bmatrix} \tilde{\mathbf{v}}(T-1) \\ \vdots \\ \tilde{\mathbf{v}}(L) \end{bmatrix}}_{=:\tilde{\mathbf{V}}} + \underbrace{\begin{bmatrix} \mathbf{m} \\ \vdots \\ \mathbf{m} \end{bmatrix}}_{=:\mathcal{M}}, \quad (6.26)$$

where $\mathcal{T}(\mathbf{s})$ is a block Hankel matrix given by

$$\mathcal{T}(\mathbf{s}) := \begin{bmatrix} s(T-1)\Sigma_{T-1} & \cdots & s(T-L-1)\Sigma_{T-1} \\ \vdots & \ddots & \vdots \\ s(L)\Sigma_L & \cdots & s(0)\Sigma_L \end{bmatrix}$$

$$\Sigma_n := \begin{bmatrix} u_1(n)\mathbf{I}_N & u_2(n)\mathbf{I}_N & \cdots & u_Q(n)\mathbf{I}_N \end{bmatrix}.$$

Also, an alternative linear model for \mathbf{Y} is given by

$$\mathbf{Y} = \mathcal{F}(\mathcal{H}) \mathbf{s} + \tilde{\mathbf{V}} + \mathcal{M} \quad (6.27)$$

where

$$\mathcal{F}(\mathcal{H}) := \begin{bmatrix} \mathbf{h}(T-1;0) & \cdots & \mathbf{h}(T-1;L) & & \\ & & \ddots & & \ddots \\ & & & \mathbf{h}(L;0) & \cdots & \mathbf{h}(L;L) \end{bmatrix}$$

is a “filtering matrix”.

By (6.26), consider the joint estimation

$$\{\hat{\mathcal{H}}, \hat{\mathbf{s}}, \hat{\mathbf{m}}\} = \arg \min_{\mathcal{H}, \mathbf{s}, \mathbf{m}} \|\mathbf{Y} - \mathcal{T}(\mathbf{s}) \mathcal{H} - \mathcal{M}\|^2, \quad (6.28)$$

where, when we estimate \mathbf{s} , we are estimating \mathbf{b} with known or estimated values of \mathbf{c} and \mathbf{b}_e . Under the white Gaussian noise assumption, the DML estimations are obtained by the nonlinear least-squares optimization (6.28). Using (6.26) and (6.27), we have a separable nonlinear LS problem that can be solved sequentially as

$$\begin{aligned} \{\hat{\mathcal{H}}, \hat{\mathbf{s}}, \hat{\mathbf{m}}\} &= \arg \min_{\mathbf{s}} \left\{ \min_{\mathcal{H}, \mathbf{m}} \|\mathbf{Y} - \mathcal{T}(\mathbf{s}) \mathcal{H} - \mathcal{M}\|^2 \right\} \\ &= \arg \min_{\mathcal{H}, \mathbf{m}} \left\{ \min_{\mathbf{s}} \|\mathbf{Y} - \mathcal{F}(\mathcal{H}) \mathbf{s} - \mathcal{M}\|^2 \right\}. \end{aligned}$$

The finite alphabet properties of the information sequences can also be incorporated into the DML methods. These algorithms iterate between estimates of the channel and the input sequences. At iteration i , with an initial guess of the channel $\mathcal{H}^{(i)}$, self-interference

$\mathbf{b}_e^{(i)}$ and the mean $\mathbf{m}^{(i)}$, the algorithm estimates the input sequence $\mathbf{s}^{(i)}$ and the channel $\mathcal{H}^{(i+1)}$ and mean $\mathbf{m}^{(i+1)}$ for the next iteration by

$$\mathbf{s}^{(i)} = \arg \min_{\mathbf{b} \in \mathcal{B}} \left\| \mathbf{Y} - \mathcal{F} \left(\mathcal{H}^{(i)} \right) \mathbf{s} - \mathcal{M}^{(i)} \right\|^2, \quad \text{with } \mathbf{s} := \mathbf{b} + \mathbf{c} + (1 - \gamma) \mathbf{b}_e^{(i)} \quad (6.29)$$

$$\mathcal{H}^{(i+1)} = \arg \min_{\mathcal{H}} \left\| \mathbf{Y} - \mathcal{T} \left(\mathbf{s}^{(i)} \right) \mathcal{H} - \mathcal{M}^{(i)} \right\|^2, \quad (6.30)$$

$$\mathbf{m}^{(i+1)} = \arg \min_{\mathbf{m}} \left\| \mathbf{Y} - \mathcal{T} \left(\mathbf{s}^{(i)} \right) \mathcal{H}^{(i+1)} - \mathcal{M} \right\|^2 \quad (6.31)$$

where \mathcal{B} is the (discrete) domain of \mathbf{b} . The optimizations in (6.30) and (6.31) are linear LS problems having the solutions

$$\mathbf{m}^{(i+1)} = \frac{1}{T-L} \sum_{n=L}^{T-1} \left[\mathbf{y}(n) - \sum_{l=0}^L \mathbf{h}^{(i+1)}(n;l) s^{(i)}(n-l) \right].$$

$$\hat{\mathcal{H}}^{(i+1)} = \mathcal{T}^\dagger \left(\mathbf{s}^{(i)} \right) \left[\mathbf{Y} - \mathcal{M}^{(i)} \right].$$

whereas the optimization in (6.29) can be achieved by using the Viterbi algorithm (see Appendix B.1 for details).

During the “start-up” (Step 1), $\mathbf{b}_e^{(1)} = \mathbf{0}$ and $\hat{\mathcal{H}}^{(1)}$ is obtained from the end of Section 6.3.1. Since the above iterative procedure involving (6.29)–(6.31) decreases the cost at every iteration, one achieves a local minimum of the nonlinear least-squares cost (local maximum of DML function).

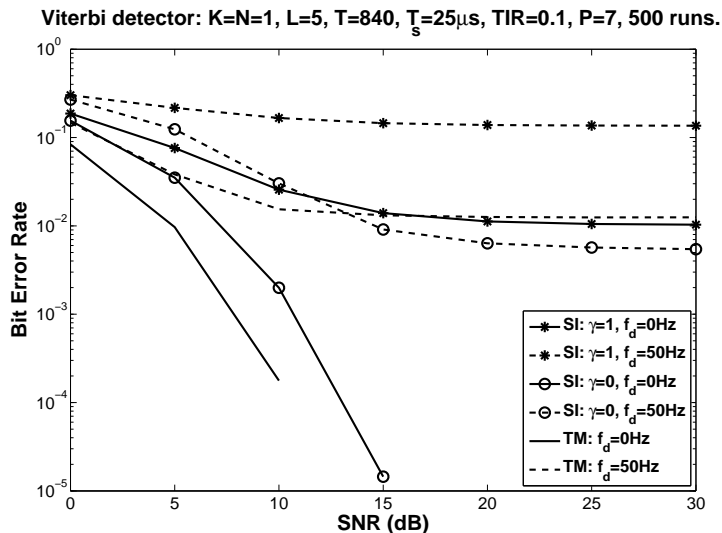


Figure 6.1: Data-dependent superimposed training (CE-BEM): BER vs SNR for non-data-dependent, data-dependent, and time-multiplexed training, under $f_d = 0$ and 50 Hz. (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)

6.4 Simulation Examples

6.4.1 Data-Dependent Superimposed Training Using CE-BEM

We first consider a doubly-selective Rayleigh fading channel following the Jakes' model with $N = 1, 2$, or 3, and $L = 5$ (6 taps). We scale $\{h(n; l)\}$ to achieve an exponential power delay profile given by $E \left\{ |h(n; l)|^2 \right\} = e^{-0.2l/(L+1)}$. We employ the communications system described in Section 2.6.3 with $T_s = 25 \mu\text{s}$. We consider the system operating under different Doppler spreads with different number of basis functions Q . For the Doppler spreads $f_d = 0, 50, 100$, and 200 Hz (corresponding to the normalized Doppler spreads $f_d T_s = 0, 0.00125, 0.0025$, and 0.005), we take $Q = 1, 5, 7$, and 11.

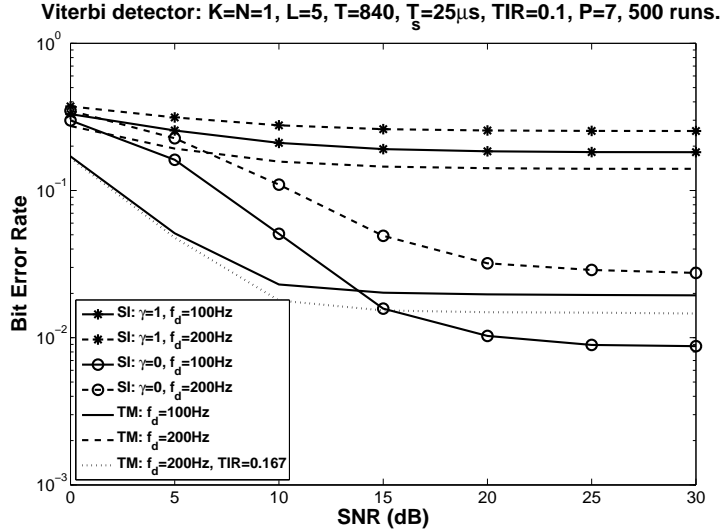


Figure 6.2: Data-dependent superimposed training (CE-BEM): BER vs SNR for non-data-dependent, data-dependent, and time-multiplexed training, under $f_d = 100$ and 200 Hz. (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)

Additive noise in each example is zero-mean complex white Gaussian. The (receiver) SNR refers to the energy per symbol over one-sided noise spectral density with both information and superimposed training sequence counting toward the symbol energy. Information sequences are BPSK. We take the superimposed training sequence period $P = 7$; it is given by $c(n) = \sigma_c e^{j\pi n(n+\nu)/P}$ where $\nu = 1$ if P is odd, and $\nu = 2$ if P is even, as in [59]. For superimposed training, the average transmitted power in $\{c(n)\}$ is 0.1 of that in $\{b(n)\}$, leading to TIR of 0.1. All the simulation results are based on 500 Monte Carlo runs.

The results for a data record length of $T = 840$ symbols are shown in Figures 6.1–6.4 for various Doppler spreads and SNR's. We use the formulation suggested in Section 3.2 by omitting terms corresponding to α_0 in (3.5c). For comparison, CE-BEM-based TM training described in Appendix A is also considered, where training sessions are periodically

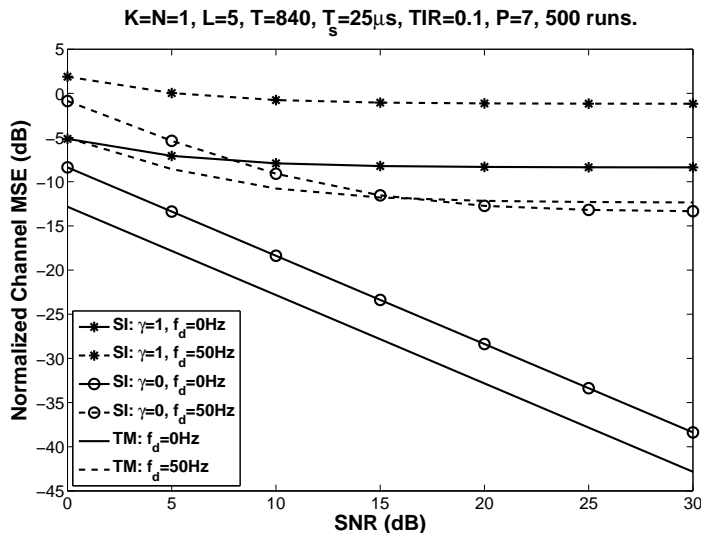


Figure 6.3: Data-dependent superimposed training (CE-BEM): NCMSE vs SNR for non-data-dependent, data-dependent, and time-multiplexed training, under $f_d = 0$ and 50 Hz. (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)

inserted between information data sessions. We take a training session of length of $2L + 1 = 11$ symbols with the training sequence $\{\mathbf{0}_{1 \times 5}, \sqrt{2L + 1}, \mathbf{0}_{1 \times 5}\}$, and at the receiver an LS estimation is performed. A data session of 110 symbols is inserted between two successive training sessions to form a frame of length 121 symbols. Such a frame is repeated over a record length of 847 symbols, i.e., a block consists $n_f = 7$ such frames. Therefore, we have a training-to-information bit and power ratio of about 0.1. The normalized channel MSE (NCMSE) in channel estimation shown in the figures is defined in (3.83). The corresponding detection results are based on Viterbi algorithm utilizing the estimated channel.

It is seen that the data-dependent superimposed training ($\gamma = 0$) yields superior results compared with the common (non-data dependent) superimposed training ($\gamma = 1$), and furthermore it is competitive with TM training without incurring the 10% training overhead

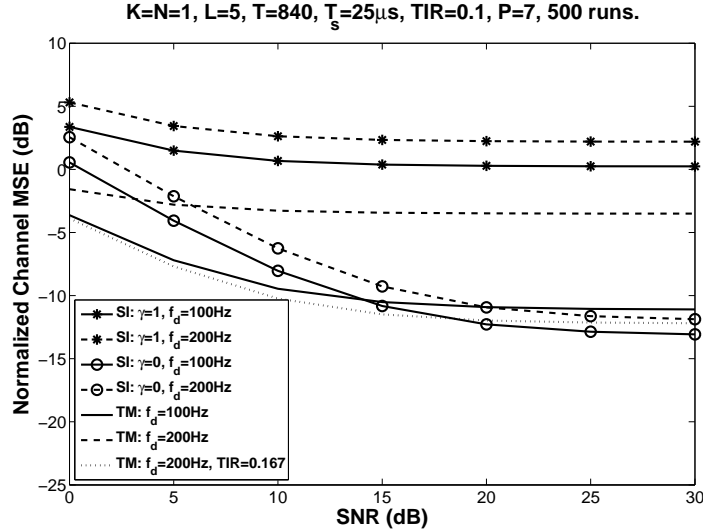


Figure 6.4: Data-dependent superimposed training (CE-BEM): NCMSE vs SNR for non-data-dependent, data-dependent, and time-multiplexed training, under $f_d = 100$ and 200 Hz. (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)

penalty resulting in a data transmission rate loss. Compared with TM training, the proposed data-dependent superimposed training yields slightly better BER for $\text{SNR} \geq 15$ dB when $f_d \neq 0$, and (slightly) worse BER for $\text{SNR} \leq 10$ dB. There is an error floor with increasing SNR for all nonzero f_d 's due to modeling errors of CE-BEM. Notice the lack of error floor for $f_d = 0$ (no modeling error) in Figures 6.1 and 6.3. When $Q = 11$ (corresponding to $f_d = 200$ Hz), for TM training we cannot satisfy training-to-information bit and power ratio of 0.1, since the number of basis functions is larger than the number of frames in one block, i.e., $Q > n_f$. We had to settle for $n_f = 7$ leading to loss of parameter identifiability (more unknown than equations). In Figures 6.2 and 6.4 we also show the results for $n_f = 11$ leading to a shorter data session with training-to-information bit and power ratio of 0.167—it is so labeled in Figures 6.2 and 6.4. The performance clearly improves and it is better

Viterbi detector: $K=1, L=5, T=840, f_d=100\text{Hz}, T_s=25\mu\text{s}, \text{TIR}=0.1, P=7, 500$ runs.

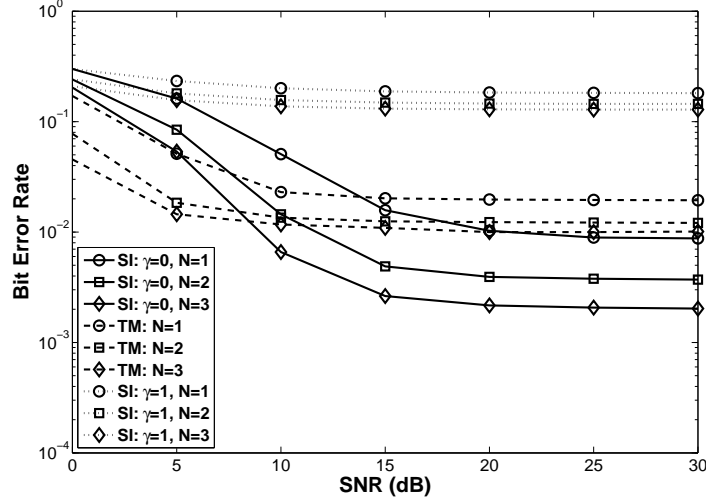


Figure 6.5: Data-dependent superimposed training (CE-BEM): BER vs SNR for non-data-dependent, data-dependent, and time-multiplexed training, under $f_d = 100$ Hz and $N = 1, 2,$ and 3 . (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)

than that of data-dependent superimposed training, but at the cost of 16.7% reduction in transmission rate.

Figure 6.5 shows the detection results (based on estimated channel and Viterbi algorithm) for multiple receivers when $f_d = 100$ Hz and $N = 1, 2,$ and 3 . Again we see that data-dependent superimposed training is better than TM training without incurring the 10% training overhead penalty (which results in a transmission rate penalty).

Now we consider a fast fading channel: $N = 1, L = 1$ (2 taps), a uniform power delay profile and $T_s = 200 \mu\text{s}$; the rest is as the above channel. Therefore, for $f_d = 100$ and 250 Hz, the normalized Doppler spreads $f_d T_s = 0.02$ and 0.05 (corresponding value of Q are 35 and 85, respectively). Here we also keep $T = 840$.

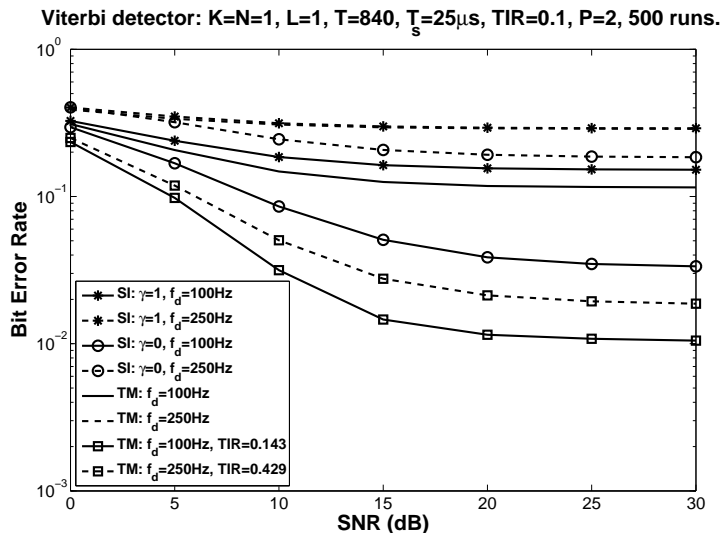


Figure 6.6: Data-dependent superimposed training (fast fading): BER vs SNR under $f_d = 100$ and 250 Hz. (SI: superimposed training; TM: time-multiplexed training; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference.)

Now the performance of all schemes is worse because of a large number of unknowns to be estimated. However, data-dependent superimposed training still outperforms TM training when we enforce the constraint training-to-information bit and power ratio of 0.1, because we cannot get $n_f \geq Q$. With the length of a training session $2L + 1 = 3$ bits, we show in Figure 6.6 the results for $n_f = 35$ and 85 for $f_d = 100$ and 250 Hz respectively, leading to reduced data sessions with training-to-information bit and power ratio of 0.143 or 0.429. The performance clearly improves and it is better than that of data-dependent superimposed training, but at the cost of 14.3% or 42.9% reduction in the transmission rate. It is also seen from Figure 6.6 that superimposed training does not perform well for $f_d = 250$ Hz because of loss of “information” due to nulling of contribution from information sequence at a “large” number of frequencies (related to Q).

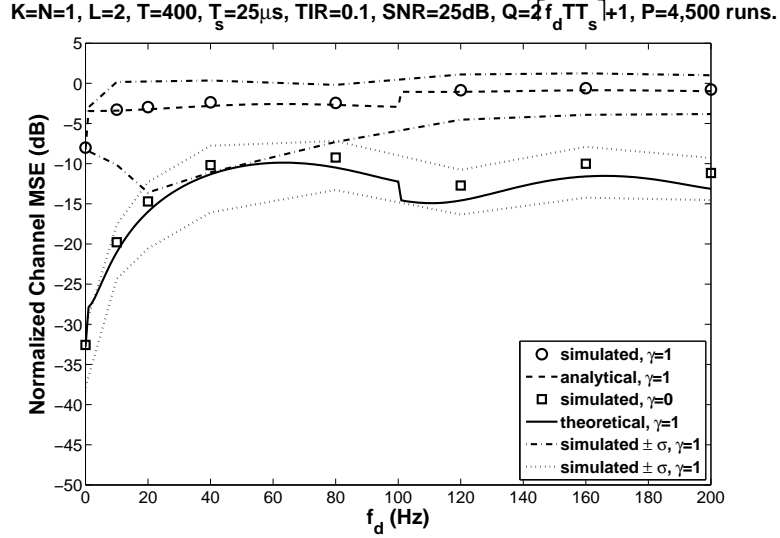


Figure 6.7: Estimation variance: NCMSE vs f_d under SNR = 25 dB for comparison between analytical and simulation-based results of non-data-dependent and data-dependent superimposed training. ($\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; σ : standard deviation.)

We next consider the performance analysis of our data-dependent superimposed training scheme. We revise the channel to be with $N = 1, L = 2$ (3 taps), a uniform power delay profile and $T_s = 25 \mu\text{s}$. We also omit α_0 in the receiver-end processing. In Figure 6.7, we show the channel MSE versus Doppler spreads where we compare our theoretical expressions with simulation-based MSE results and $\pm\sigma$ bounds. The channel MSE for non-data-dependent superimposed training is given by (4.8) and (4.35). The agreement is good between the theoretical and simulations-based results. Note that the discontinuities in the theoretical curves as a function of f_d as we picked Q values based on f_d per (2.9b).

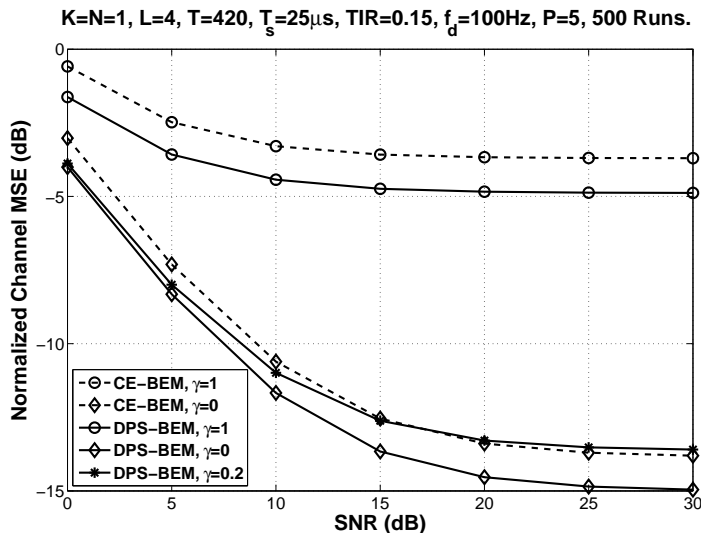


Figure 6.8: PDD superimposed training: NCMSE vs SNR for CE- and DPS-BEM-based estimators, under $f_d = 100$ Hz. ($\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)

6.4.2 (Partially) Data-Dependent Superimposed Training Using DPS-BEM

We consider a random doubly-selective Rayleigh fading channel. We take $N = 1$ and $L = 4$ (5 taps) with $\mathbf{h}(n;l)$ as in (H6.3.5) satisfying the Jakes' model.

We consider a communication system described in Section 2.6.3 with symbol interval $T_s = 25 \mu\text{s}$, record length $T = 420$ symbols, and varying Doppler spreads f_d in the range of 0 Hz to 200 Hz. For $f_d = 100$ Hz, the normalized Doppler spread $f_d T_s = 0.0025$ and for $f_d = 200$ Hz, $f_d T_s = 0.005$. We emphasize that in the simulations the DPS-BEM is used only for processing at the receiver; the random channels are generated by the Jakes' model, not the DPS-BEM. In the processing of channel estimation, we select Q , the number of basis functions, as in $Q = \lceil 2f_d T_s T \rceil + 1$. Using the estimated channel, a Viterbi detector is used for data symbol detection at the receiver. For the DPS-BEM-based estimator using

Viterbi detector: $K=N=1$, $L=4$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.15$, $f_d=100\text{Hz}$, $P=5$, 500 Runs

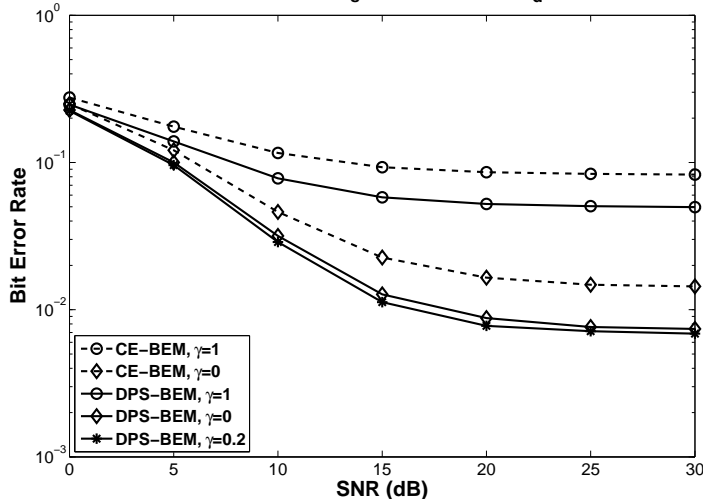


Figure 6.9: PDD superimposed training: BER vs SNR for CE- and DPS-BEM-based estimators, under $f_d = 100$ Hz. ($\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)

PDD superimposed training, we took $k = -1$ (the minimum allowed value) in (6.9), so that the information loss is comparatively mild.

The additive noise is zero-mean complex white Gaussian. The (receiver) SNR refers to the energy per bit over one-sided noise spectral density with both information and superimposed training sequence counting toward the bit energy. Information sequences are BPSK. We take the superimposed training sequence of period $P = L + 1 = 5$ as $c(n) = \sigma_c e^{j\pi n(n+\nu)/P}$ where $\nu = 1$ if P is odd and $\nu = 2$ if P is even, as in [59]. All simulated results were based on 500 Monte Carlo runs.

To show the advantage of DPS-BEM over the Fourier-based CE-BEM, we compare the performance of our proposed channel estimators using the two BEM's. Figure 6.8 shows the comparison for normalized channel MSE's (NCMSE). In an environment of $f_d = 100$ Hz,

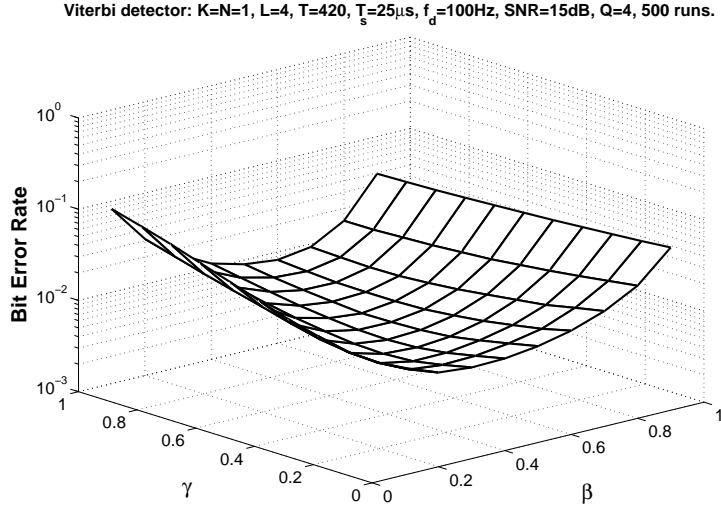


Figure 6.10: PDD superimposed training: BER vs (β, γ) under $\text{SNR} = 15\text{ dB}$ and $f_d = 100\text{ Hz}$.

we had $Q = 5$ for CE-BEM following (2.9b) and $Q = 4$ for DPS-BEM. The channel estimators have identical design parameters except for using different basis models. In the simulation, the average transmitted power σ_c^2 in $c(n)$ is 0.15 of the power in $b(n)$, leading to a training-to-information power ratio $\text{TIR} := \sigma_c^2/\sigma_b^2 = \beta/(1 - \beta) = 0.15$ (or $\beta = 0.13$). We consider the CE-BEM-based estimator using superimposed training with $\gamma = 1$ (corresponding to the first-order statistics-based estimator in Chapter 3) and $\gamma = 0$ (corresponding to the “fully” data-dependent scheme), and the DPS-BEM-based estimator using PDD superimposed training with $\gamma = 0, 0.2$, and 1. At the receiver, we follow the data detection scheme described in Section 6.2.2.

It is seen from Figure 6.8 that the data-dependent ($\gamma = 0$) training offers much lower estimation variance than that of the “non-data-dependent” ($\gamma = 1$) training, whether using CE- or DPS-BEM, since the self-interference is eliminated or greatly reduced with $\gamma = 0$.

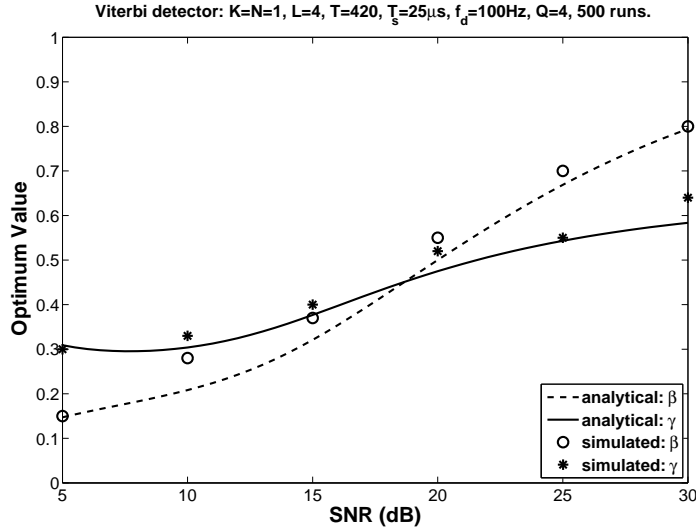


Figure 6.11: PDD superimposed training: optimum (β, γ) vs SNR under $f_d = 100$ Hz.

For the same values of γ 's, the estimator exploiting DPS-BEM provides better estimation compared with the CE-BEM-based one, because DPS-BEM has smaller modeling error than CE-BEM. These conclusions are also confirmed by the BER results in Figure 6.9. Note that for the DPS-BEM-based estimator with $\gamma = 0.2$, although the estimation cannot be as accurate as the “completely-data-dependent” ($\gamma = 0$) case, the BER result is still better since the “information” lost in the “completely-data-dependent” training is now partially retained.

Figure 6.10 depicts the BER surface as a function of β (defined in (6.21)) and γ , for SNR = 15 dB and $f_d = 100$ Hz. The BER performance varies along β - and γ -axes. We selected the coordinate point $(\tilde{\beta}^o, \tilde{\gamma}^o)$ corresponding to the minimum value at the BER surface as the simulation-based optimum value for the given SNR. In Figure 6.11 we compare the simulation-based optimum value with the “analytical” optimum value (β^o, γ^o) derived recursively (with the initial value $\gamma = 0$) by (6.23) and (6.24). In Figure 6.11, the analytical

Viterbi detector: $K=N=1$, $L=4$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.15$, $f_d=0\text{Hz}$, $P=5$, 500 Runs.

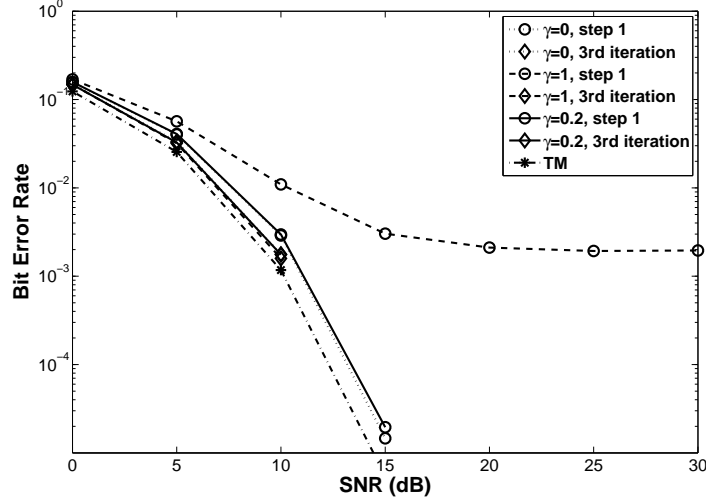


Figure 6.12: PDD superimposed training: BER vs SNR under $f_d = 0\text{Hz}$ (time-invariant). (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)

and the simulation-based results follow the same trend, and the agreement between them is good. It is also seen in Figure 6.11 that as the received signal SNR increases, the optimum β and γ increase too. Higher β implies that a higher fraction of the transmitted power is allocated to training, leading to more accurate channel estimates (with smaller estimation variance). Intuitively, for higher SNR’s, it pays to achieve more accurate channel estimates in order to achieve a lower effective noise power $\bar{\sigma}_w^2$. On the other hand, when the SNR is low, improving channel estimation does not have much effect on the effective noise power $\bar{\sigma}_w^2$. Similar comments apply to changes in optimum γ with SNR. Higher γ implies lower power in effective noise component $(1 - \gamma) b_e (n - l)$ but higher self-interference, hence higher channel estimation variance – these two effects need to be counter-balanced. Finally,

Viterbi detector: $K=N=1$, $L=4$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.15$, $f_d=100\text{Hz}$, $P=5$, 500 Runs

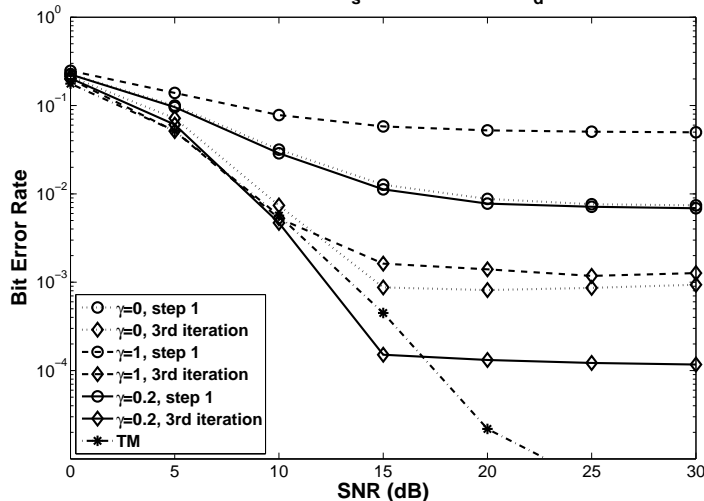


Figure 6.13: PDD superimposed training: BER vs SNR under $f_d = 100$ Hz. (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)

observe from Figure 6.10 that the “bottom” (corresponding to the neighborhood of the minimum BER point) of the BER surface is rather “flat”: the BER performance is not sensitive to changes in β and γ over a rather “wide” area around the minimum, so that the analysis described in Section 6.3.3 provides us an effective means for power allocation and interference suppression.

The DML approach is now investigated to enhance the channel estimation and data detection performance. We considered the channels with Doppler spread $f_d = 0$, 100, and 200 Hz respectively, with the corresponding number of basis functions $Q = 1$, 4, and 6. Note that the channel with $f_d = 0$ Hz is time-invariant; this case is considered because here there are no modeling errors. The “non-data-dependent” ($\gamma = 1$) superimposed training

Viterbi detector: $K=N=1$, $L=4$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.15$, $f_d=200\text{Hz}$, $P=5$, 500 Runs

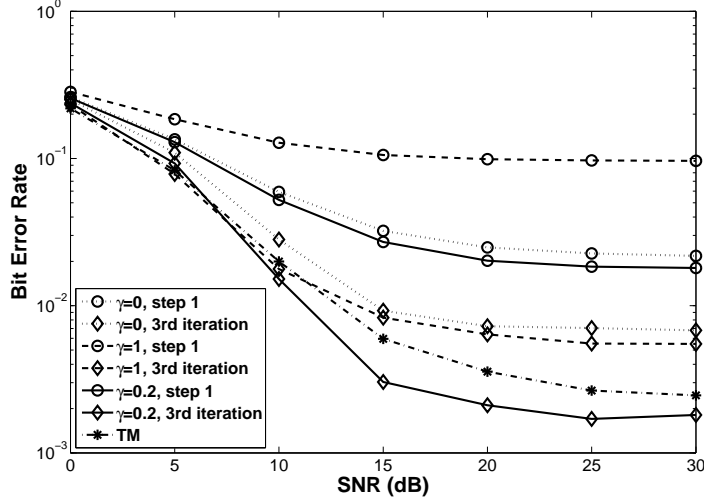


Figure 6.14: PDD superimposed training: BER vs SNR under $f_d = 200$ Hz. (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)

was compared with the “completely” ($\gamma = 0$) and “partially” ($\gamma = 0.2$) data-dependent scheme. We again set $\text{TIR} = 0.15$ for the superimposed training. At the receiver, DML iterations follow the data detection scheme we described at Section 6.2.2 (denoted by “step 1” in the figures). We show NCMSE and BER in Figures 6.12–6.17 where the results of Step 1 and the third iteration only are depicted. It is seen that the DML algorithm significantly improves the performances.

For the purpose of comparison, the TM training approach described in Appendix A, originally proposed for CE-BEM as an optimal scheme, is applied to the DPS-BEM case. We take a training block of $2L + 1 = 9$ symbols as $\{0, 0, 0, 0, \sqrt{2L + 1}, 0, 0, 0, 0\}$, which follows an information data block of length 60 leading to a frame of 69 symbols. This

Viterbi detector: $K=N=1$, $L=4$, $T=420$, $T_s=25\mu\text{s}$, $TIR=0.15$, $f_d=0\text{Hz}$, $P=5$, 500 Runs.

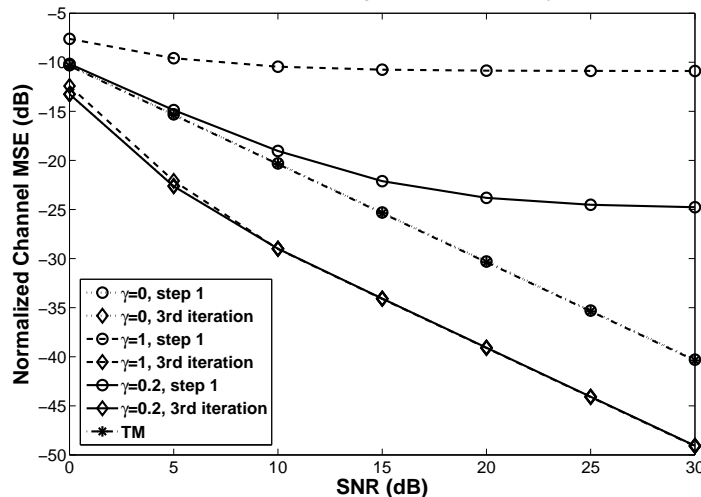


Figure 6.15: PDD superimposed training: NCMSE vs SNR under $f_d = 0\text{Hz}$ (time-invariant). (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)

subblock was repeated over a record length of 414 symbols with a total of 6 subblocks. The information data is also BPSK and have unit power. Thus, the training-to-information bit and power ratios are both 0.15 (the amplitude of the single nonzero training bit was picked to achieve this power ratio). Using the training sequence, we can uniquely determine $\mathbf{h}_q(l)$'s via an LS approach.

By (6.9) note that $|\Omega| = P(2Q + 2k + 1)$ ($k = -1$ in the simulations). For the time-invariant channel ($f_d = 0$) $Q = 1$, so that the information contained in the self-interference part of the information sequence is comparatively small. Therefore, total suppression ($\gamma = 0$) of the self-interference does not have a significant deleterious effect on the BER, while the improvement in channel estimation is significant—it is the scenario described by [20] and

Viterbi detector: $K=N=1$, $L=4$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.15$, $f_d=100\text{Hz}$, $P=5$, 500 Runs

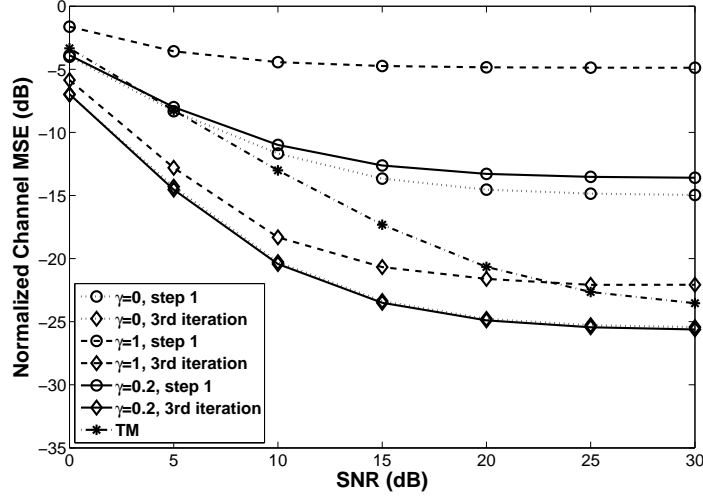


Figure 6.16: PDD superimposed training: NCMSE vs SNR under $f_d = 100$ Hz. (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)

depicted in Figures 6.12 and 6.15. Exploiting the PDD scheme ($\gamma = 0.2$) in this case does not have much impact. All the three schemes ($\gamma = 0, 0.2$, and 1) after three iterations of the DML scheme have BER performance in Figure 6.12 similar to that of the TM training.

As the Doppler spread f_d increases, we have to employ a larger Q to describe the channel and the self-interference thus grows. Now total suppression ($\gamma = 0$) is no longer a wise option, whereas the PDD scheme still performs well. In Figures 6.13 and 6.14, with $\gamma = 0.2$ the PDD scheme is superior to the other two superimposed training-based schemes in data detection, whether before or after DML iterations, since the self-interference has been greatly suppressed while the information loss has been effectively reduced. In Figures 6.16 and 6.17, although the “completely” data-dependent ($\gamma = 0$) training has the best channel

Viterbi detector: $K=N=1$, $L=4$, $T=420$, $T_s=25\mu\text{s}$, $\text{TIR}=0.15$, $f_d=200\text{Hz}$, $P=5$, 500 Runs

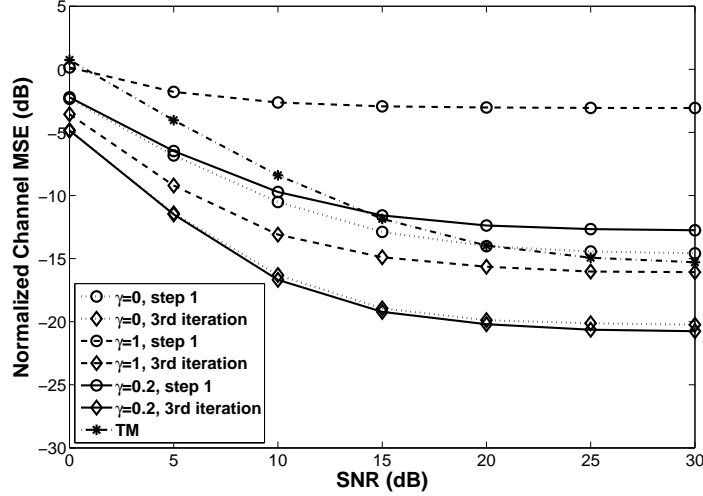


Figure 6.17: PDD superimposed training: NCMSE vs SNR under $f_d = 200$ Hz. (TM: time-multiplexed training; “step 1”: the data detection scheme in Section 6.2.2; “3rd iteration”: the third DML iteration; $\gamma = 1$: non-data-dependent training; $\gamma = 0$: total elimination of self-interference; $\gamma = 0.2$: partial elimination of self-interference at the channel estimation stage.)

estimation before iterations, the PDD scheme with $\gamma = 0.2$ yields better performance after three iterations, due to the lower BER in the DML approach. The BER performance of the PDD scheme after several DML iterations is competitive with the TM training, without incurring any training overhead penalty.

6.5 Conclusions

In this chapter, we presented a data-dependent superimposed training scheme to reduce the self-interference in channel estimation. Inspired by the work of [20], we observed that over a channel satisfying a band-limited BEM such as CE- or DPS-BEM, the periodic

superimposed training components within the received signal occur only at certain frequencies. By designing a data-dependent superimposed training sequence that suppresses the information sequence at those frequencies, the self-interference adversely affecting the channel estimation at the receiver is greatly reduced. However, the suppressed frequency components of the information sequence carry “information” as well. A PDD superimposed training method was proposed to strike a trade-off between self-interference cancelation and information integrity. Performance analysis and related optimization of parameters were also discussed. Computer simulation examples demonstrated that by using PDD superimposed training, competitive performance with TM training can be achieved but with no training overhead penalty.

CHAPTER 7

DIRECT FIR LINEAR EQUALIZATION OF DOUBLY-SELECTIVE CHANNELS BASED ON SUPERIMPOSED TRAINING

7.1 Introduction

Two powerful tools have been exploited so far in our previous chapters on channel estimation: superimposed training, providing us a means to track the temporal variation of the channel, and BEM's, reducing the problem of estimation of a time-varying channel over a period of time to estimation of time-invariant parameters.

In wireless channels, signal distortion due to multipath propagation or band-limited transmission may cause ISI at reception. An equalizer is the device to compensate ISI at the receive end. The noise-free received signal is the convolution between the transmitted symbols and the impulse response of the channel. Therefore, the equalizer, whose task is to recover the transmitted symbols, is a deconvolution device. Since a doubly-selective channel can be well described by BEM's, we expect that an equalizer, as an inverse of the channel, can also be well represented by BEM's. Given the knowledge of the time-varying channel described by CE-BEM, design of serial time-varying FIR equalizers has been discussed in [2]. Direct design of time-invariant FIR equalizers based on superimposed training, for time-invariant channels, has been investigated in [58]. In this chapter, we investigate direct design of time-varying FIR linear equalizers for doubly-selective channels using superimposed training and without first estimating the underlying channel response. We exploit the prior results of [2, 58].

In Section 7.2, the direct FIR linear equalization using superimposed training and CE-BEM is discussed. By utilizing user-specific training sequences, this direct equalizer is extended to a multiple-user wireless *ad hoc* network in Section 7.3.

7.2 Direct FIR Linear Equalization Using CE-BEM

Consider a time-varying SIMO FIR linear channel with N outputs. Let $\{s(n)\}$ denote a scalar sequence which is input to the SIMO time-varying channel with discrete-time impulse response $\{\mathbf{h}(n; l)\}$ (N -vector channel response at time n to a unit input at time $n - l$). Then the symbol-rate, channel output vector is given by (3.1)

$$\mathbf{x}(n) := \sum_{l=0}^L \mathbf{h}(n; l) s(n - l). \quad (7.1)$$

The noisy measurements of $\mathbf{x}(n)$ are given by

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n).$$

In a CE-BEM representation it is assumed that the channel follows (2.9)

$$\mathbf{h}(n; l) = \sum_{q=1}^{\tilde{Q}} \mathbf{h}_q(l) e^{j\omega_q n} \quad (7.2)$$

where N -column vectors $\mathbf{h}_q(l)$ are invariant for the whole block $n = 0, 1, \dots, T - 1$, and

$$\tilde{Q} := 2 \lceil f_d T T_s \rceil + 1,$$

$$L := \lfloor \tau_d / T_s \rfloor,$$

$$\omega_q := \frac{2\pi}{T} \left(q - \frac{1}{2} - \frac{Q}{2} \right), \quad q = 1, 2, \dots, Q.$$

In superimposed training one takes

$$s(n) = b(n) + c(n),$$

where $\{b(n)\}$ is the information sequence and $\{c(n)\}$ is the training sequence added (superimposed) at low power to the information sequence at the transmitter before modulation and transmission.

Given this channel model, the main problem considered here is: how to design an equalizer to estimate $\{b(n)\}$ when one knows only $\{c(n)\}$ but not (obviously) $\{b(n)\}$ and one does not also have (frame) synchronization with $\{c(n)\}$ at the receiver. We will design an equalizer to estimate $\{c(n)\}$ with a delay d . We will then show that this equalizer is a scaled version of the corresponding equalizer designed to estimate $\{b(n)\}$ with a delay d .

7.2.1 Time-Varying FIR Equalizers

We will restrict ourselves to serial linear equalizers instead of block linear equalizers, since as shown in [2], the latter are computationally prohibitive (compared with the former). We look for a time-varying linear equalizer $\mathbf{g}(n; l)$ ($l = 0, 1, \dots, L_e$) over the same time block as the received data with channel model (7.2). We note that for an arbitrary time-varying impulse response $\tilde{\mathbf{g}}(n; l)$, the following is always true

$$\tilde{\mathbf{g}}(n; l) = \sum_{q=1}^T \tilde{\mathbf{g}}_q(l) e^{j\omega_q n}, \quad n = 0, 1, \dots, T-1.$$

We would like to use a more parsimonious (but approximate) representation for $\tilde{\mathbf{g}}(n;l)$, denoted by $\mathbf{g}(n;l)$, given by

$$\mathbf{g}(n;l) = \sum_{q=1}^Q \mathbf{g}_q(l) e^{j\omega_q n}, \quad n = 0, 1, \dots, T-1,$$

where $Q \ll T-1$. In order to estimate the input sequence $\{s(n)\}$ (see (7.1)), we may seek a linear time-varying FIR estimator using CE-BEM to yield an estimate with equalization delay d ($0 \leq d \leq L_e$)

$$\hat{s}(n-d) = \sum_{i=0}^{L_e} \mathbf{g}^H(n;i) \mathbf{y}(n-i).$$

Existence of a zero-forcing linear equalizer has been discussed in [2]. Their conclusion is that if N is at least two, then with probability one, one has a zero-forcing solution for sufficiently large L_e and Q . For a linear MMSE solution, existence is not an issue, although MMSE equalizer performance can be expected to be “good” if zero-forcing equalizers exist [2]. Here we will seek a least squares solution $\mathbf{g}(n;l)$ to minimize a cost such as

$$J = \frac{1}{T} \sum_{n=0}^{T-1} |s(n-d) - \hat{s}(n-d)|^2.$$

7.2.2 Linear LS Equalizers Based on CE-BEM

Our algorithm is based on the following model assumptions:

- (H7.2.1) The information sequence $\{b(n)\}$ is zero-mean, i.i.d. (independent and identically distributed), with $E\{|b(n)|^2\} = \sigma_b^2$.
- (H7.2.2) The measurement noise $\{\mathbf{v}(n)\}$ is zero-mean ($E\{\mathbf{v}(n)\} = \mathbf{0}$), white, independent of $\{b(n)\}$, with $E\{[\mathbf{v}(n+\tau)][\mathbf{v}(n)]^H\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$.

(H7.2.3) The superimposed training sequence $c(n) = c(n+P)$ for all n is a non-random periodic sequence with period P . Let $\sigma_c^2 := (1/P) \sum_{n=1}^P |c(n)|^2$.

(H7.2.4) Record length T and period P satisfy $TP^{-1} > \tilde{Q}$. Moreover, $P > L + L_e - d$.

As in (3.5), the periodic training sequence of period P can be written as

$$c(n) = \sum_{m=0}^{P-1} c_m e^{j\alpha_m n}$$

where $\alpha_m := 2\pi m/P$. To design the time-varying linear equalizer to estimate a delayed version of the training sequence $c(n-d)$ ($0 \leq d \leq L_e$), we have

$$\hat{c}(n-d) = \sum_{i=0}^{L_e} \mathbf{g}_d^H(n; i) \mathbf{y}(n-i)$$

where we assume that

$$\mathbf{g}_d(n; i) = \sum_{q=1}^Q \mathbf{g}_q(i) e^{j\omega_q n}.$$

Choose $\mathbf{g}_q(i)$'s to minimize the time-averaged cost

$$\begin{aligned} J_c &:= \frac{1}{T} \sum_{n=0}^{T-1} |c(n-d) - \hat{c}(n-d)|^2 \\ &= \frac{1}{T} \sum_{n=0}^{T-1} \left| c(n-d) - \sum_{i=0}^{L_e} \sum_{q=1}^Q \mathbf{g}_q^H(i) e^{-j\omega_q n} \mathbf{y}(n-i) \right|^2. \end{aligned}$$

By taking the derivative and setting it to be zero, we have

$$\frac{\partial J_c}{\partial \mathbf{g}_{q_1}^*(i_1)} = -\frac{1}{T} \sum_{n=0}^{T-1} e^{-j\omega_{q_1} n} \mathbf{y}(n-i_1) \left[c^*(n-d) - \sum_{i=0}^{L_e} \sum_{q=1}^Q e^{j\omega_q n} \mathbf{y}^H(n-i) \mathbf{g}_q(i) \right] = \mathbf{0}$$

for $i_1 = 0, 1, \dots, L_e$ and $q_1 = 1, 2, \dots, Q$. This leads to

$$\begin{aligned} & \sum_{i=0}^{L_e} \sum_{q=1}^Q \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{j(\omega_q - \omega_{q_1})n} \mathbf{y}(n - i_1) \mathbf{y}^H(n - i) \right] \mathbf{g}_q(i) \\ &= \frac{1}{T} \sum_{n=0}^{T-1} c^*(n - d) e^{-j\omega_{q_1}n} \mathbf{y}(n - i_1) =: \mathbf{R}_c(q_1, i_1). \end{aligned} \quad (7.3)$$

To design the time-varying linear equalizer to estimate the information sequence $b(n-d)$ ($0 \leq d \leq L_e$), we have

$$\hat{b}(n-d) = \sum_{i=0}^{L_e} \bar{\mathbf{g}}_d^H(n; i) \mathbf{y}(n-i)$$

where we assume that

$$\bar{\mathbf{g}}_d(n; i) = \sum_{q=1}^Q \bar{\mathbf{g}}_q(i) e^{j\omega_q n}.$$

Choose $\bar{\mathbf{g}}_q(s)$'s to minimize

$$J_b := \frac{1}{T} \sum_{n=0}^{T-1} \left| b(n-d) - \hat{b}(n-d) \right|^2.$$

Mimicking the results for the superimposed training sequence-based equalization,

$$\begin{aligned} & \sum_{i=0}^{L_e} \sum_{q=1}^Q \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{j(\omega_q - \omega_{q_1})n} \mathbf{y}(n - i_1) \mathbf{y}^H(n - i) \right] \bar{\mathbf{g}}_q(i) \\ &= \frac{1}{T} \sum_{n=0}^{T-1} b^*(n-d) e^{-j\omega_{q_1}n} \mathbf{y}(n - i_1) =: \mathbf{R}_b(q_1, i_1). \end{aligned} \quad (7.4)$$

Comparing (7.3) and (7.4), we see that (ignoring the equalizer coefficients) the left sides of the two are identical. We now seek to establish that for large T , $\mathbf{R}_c(q_1, i_1) = \beta \mathbf{R}_b(q_1, i_1)$ for all q_1, i_1 , for some scalar β , so that $\mathbf{g}_q(i) = \beta \bar{\mathbf{g}}_q(i)$ for all i .

We first consider

$$\begin{aligned}
& \mathbf{R}_c(q_1, i_1) \\
&= \frac{1}{T} \sum_{n=0}^{T-1} c^*(n-d) e^{-j\omega_{q_1} n} \left\{ \sum_{l=0}^L \mathbf{h}(n-i_1; l) [b(n-i_1-l) + c(n-i_1-l)] + \mathbf{v}(n-i_1) \right\} \\
&= \frac{1}{T} \sum_{n=0}^{T-1} \sum_{m_1=0}^{P-1} c_{m_1}^* e^{-j\alpha_{m_1}(n-d)} e^{-j\omega_{q_1} n} \\
&\quad \times \left\{ \sum_{l=0}^L \sum_{q=1}^{\tilde{Q}} \mathbf{h}_q(l) e^{j\omega_q(n-i_1)} \left[\sum_{m_2=0}^{P-1} c_{m_2} e^{j\alpha_{m_2}(n-i_1-l)} + b(n-i_1-l) \right] + \mathbf{v}(n-i_1) \right\} \\
&= \sum_{q=1}^{\tilde{Q}} \sum_{l=0}^L \sum_{m_1=0}^{P-1} \sum_{m_2=0}^{P-1} c_{m_1}^* c_{m_2} e^{j\alpha_{m_1} d} e^{-j\alpha_{m_2}(i_1+l)} e^{-j\omega_q i_1} \mathbf{h}_q(l) A_0 \\
&\quad + \sum_{q=1}^{\tilde{Q}} \sum_{l=0}^L \sum_{m_1=0}^{P-1} c_{m_1}^* e^{j\alpha_{m_1} d} e^{-j\omega_q i_1} \mathbf{h}_q(l) A_1 + \sum_{m_1=0}^{P-1} c_{m_1}^* e^{j\alpha_{m_1} d} \mathbf{A}_2
\end{aligned}$$

where

$$\begin{aligned}
A_0 &:= \frac{1}{T} \sum_{n=0}^{T-1} e^{j(-\alpha_{m_1} + \alpha_{m_2} - \omega_{q_1} + \omega_q)n}, \\
A_1 &:= \frac{1}{T} \sum_{n=0}^{T-1} e^{j(-\alpha_{m_1} - \omega_{q_1} + \omega_q)n} b(n-i_1-l), \\
\mathbf{A}_2 &:= \frac{1}{T} \sum_{n=0}^{T-1} e^{-j(\alpha_{m_1} + \omega_{q_1})n} \mathbf{v}(n-i_1).
\end{aligned}$$

Under the condition $TP^{-1} > \tilde{Q}$ (then $(\alpha_{m_1} + \omega_{q_1}) = (\alpha_{m_2} + \omega_{q_2})$ if and only if $m_1 = m_2$ and $q_1 = q_2$), we have

$$A_0 = \delta(m_1 - m_2) \delta(q_1 - q).$$

Furthermore, we have

$$\begin{aligned} E \left\{ |A_1|^2 \right\} &= \frac{1}{T^2} \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} e^{j(-\alpha_{m_1} - \omega_{q_1} + \omega_q)(n_1 - n_2)} E \{ b(n_1 - i_1 - l) b^*(n_2 - i_1 - l) \} \\ &= \frac{1}{T^2} \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} e^{j(-\alpha_{m_1} - \omega_{q_1} + \omega_q)(n_1 - n_2)} \sigma_b^2 \delta(n_1 - n_2) = \frac{\sigma_b^2}{T}. \end{aligned}$$

Similarly, it follows that

$$\begin{aligned} E \left\{ \|\mathbf{A}_2\|^2 \right\} &= \frac{1}{T^2} \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} e^{-j(\alpha_{m_1} + \omega_{q_1})(n_1 - n_2)} \mathbf{v}^H(n_1 - i_1) \mathbf{v}(n_2 - i_1) \\ &= \frac{1}{T^2} \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} e^{-j(\alpha_{m_1} + \omega_{q_1})(n_1 - n_2)} N \sigma_v^2 \delta(n_1 - n_2) \\ &= \frac{N \sigma_v^2}{T}. \end{aligned}$$

In the mean square sense (and thus in probability), we then have the following two limits

$$\lim_{T \rightarrow \infty} A_1 \stackrel{\text{m.s.}}{=} 0, \text{ and } \lim_{T \rightarrow \infty} \mathbf{A}_2 \stackrel{\text{m.s.}}{=} \mathbf{0}.$$

Thus for “large” T , we have

$$\begin{aligned} &\lim_{T \rightarrow \infty} \mathbf{R}_c(q_1, i_1) \stackrel{\text{m.s.}}{=} \\ &\sum_{q=1}^{\tilde{Q}} \sum_{l=0}^L \sum_{m_1=0}^{P-1} \sum_{m_2=0}^{P-1} c_{m_1}^* c_{m_2} e^{j\alpha_{m_1} d} e^{-j\alpha_{m_2} (i_1 + l)} e^{-j\omega_q i_1} \mathbf{h}_q(l) \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{j(-\alpha_{m_1} + \alpha_{m_2} - \omega_{q_1} + \omega_q)n} \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{q=1}^{\tilde{Q}} \sum_{l=0}^L \sum_{m_1=0}^{P-1} \sum_{m_2=0}^{P-1} c_{m_1}^* c_{m_2} e^{j\alpha_{m_1} d} e^{-j\alpha_{m_2}(s_1+l)} e^{-j\omega_k i_1} \mathbf{h}_q(l) \delta(m_1 - m_2) \delta(q_1 - q) \\
&= \sum_{q=1}^{\tilde{Q}} \sum_{l=0}^L \sum_{m=0}^{P-1} |c_m|^2 e^{j\alpha_m(d-i_1-l)} e^{-j\omega_q i_1} \mathbf{h}_q(l) \delta(q_1 - q).
\end{aligned}$$

If the training sequence $\{c(n)\}$ is periodic white, i.e.,

$$P^{-1} \sum_{n=0}^{P-1} c(n)c^*(n-l) = \sigma_c^2 \delta(l \bmod P),$$

then

$$\sum_{m=0}^{P-1} |c_m|^2 e^{j\alpha_m(d-i_1-l)} = \sigma_c^2 \delta((d-i_1-l) \bmod P).$$

This fact then leads to

$$\lim_{T \rightarrow \infty} \mathbf{R}_c(q_1, i_1) \stackrel{\text{m.s.}}{=} \begin{cases} \sigma_c^2 e^{-j\omega_{q_1} i_1} \mathbf{h}_{q_1}((d-i_1) \bmod P) & \text{if } 1 \leq q_1 \leq \tilde{Q} \\ 0 & \text{otherwise} \end{cases} \quad (7.5)$$

for $i_1 = 0, 1, \dots, L_e$ and $q_1 = 1, 2, \dots, Q$.

Turning to (7.4), we have

$$\mathbf{R}_b(q_1, i_1) = \sum_{q=1}^{\tilde{Q}} \sum_{l=0}^L \sum_{m=0}^{P-1} c_m \mathbf{h}_q(l) e^{-j\alpha_m(i_1+l)} e^{-j\omega_k i_1} A_3 + \sum_{q=1}^{\tilde{Q}} \sum_{l=0}^L \mathbf{h}_q(l) e^{-j\omega_q i_1} A_4 + \mathbf{A}_5$$

where

$$\begin{aligned}
A_3 &:= \frac{1}{T} \sum_{n=0}^{T-1} e^{j(\alpha_m - \omega_{q_1} + \omega_q)n} \mathbf{b}^*(n-d), \\
A_4 &:= \frac{1}{T} \sum_{n=0}^{T-1} e^{j(\omega_q - \omega_{q_1})n} \mathbf{b}(n-i_1-l) \mathbf{b}^*(n-d),
\end{aligned}$$

$$\mathbf{A}_5 := \frac{1}{T} \sum_{n=0}^{T-1} e^{-j\omega_{q_1} n} \mathbf{v}(n - i_1) b^*(n - d).$$

We can show (as before) that

$$\lim_{T \rightarrow \infty} A_3 \stackrel{\text{m.s.}}{=} 0,$$

$$\lim_{T \rightarrow \infty} \mathbf{A}_5 \stackrel{\text{m.s.}}{=} \mathbf{0}.$$

Consider

$$A_6 := \frac{1}{T} \sum_{n=0}^{T-1} e^{j(\omega_q - \omega_{q_1})n} [b(n - i_1 - l) b^*(n - d) - \sigma_b^2 \delta(d - i_1 - l)].$$

It then follows that

$$\begin{aligned} & E \left\{ |A_6|^2 \right\} \\ &= \frac{1}{T^2} \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} e^{j(\omega_q - \omega_{q_1})(n_1 - n_2)} E \left\{ \left| b(n - i_1 - l) b^*(n - d) - \sigma_b^2 \delta(d - i_1 - l) \right|^2 \right\} \\ &= \frac{1}{T^2} \sum_{n_1=0}^{T-1} \sum_{n_2=0}^{T-1} e^{j(\omega_q - \omega_{q_1})(n_1 - n_2)} \left[E \left\{ |b(n)|^4 \right\} - \sigma_b^4 \right] \delta(n_1 - n_2) \delta(d - i_1 - l) \\ &= \frac{1}{T} \left[E \left\{ |b(n)|^4 \right\} - \sigma_b^4 \right] \delta(d - i_1 - l). \end{aligned}$$

Therefore, we have $\lim_{T \rightarrow \infty} A_6 \stackrel{\text{m.s.}}{=} 0$, and consequently

$$\begin{aligned} \lim_{T \rightarrow \infty} A_4 &\stackrel{\text{m.s.}}{=} \frac{1}{T} \sum_{n=0}^{T-1} e^{j(\omega_q - \omega_{q_1})n} \sigma_b^2 \delta(d - i_1 - l) \\ &= \sigma_b^2 \delta(d - i_1 - l) \delta(q_1 - q). \end{aligned}$$

Hence, for “large” T , we have

$$\lim_{T \rightarrow \infty} \mathbf{R}_b(q_1, i_1) \stackrel{\text{m.S.}}{=} \sum_{q=1}^{\tilde{Q}} \sum_{l=0}^L \mathbf{h}_q(l) e^{-j\omega_q i_1} \sigma_b^2 \delta(d - i_1 - l) \delta(q_1 - q).$$

For $i_1 = 0, 1, \dots, L_e$ and $q_1 = 1, 2, \dots, Q$ but $q_1 \leq \tilde{Q}$, we therefore have

$$\lim_{T \rightarrow \infty} \mathbf{R}_b(q_1, i_1) \stackrel{\text{m.S.}}{=} \mathbf{h}_{q_1}(d - i_1) e^{-j\omega_{q_1} i_1} \sigma_b^2. \quad (7.6)$$

If $P > L + L_e - d$, then (7.5) equals (7.6) (within a scale factor). Therefore, for “large” T , $\mathbf{R}_c(q_1, i_1) = \beta \mathbf{R}_b(q_1, i_1)$ for all q_1, i_1 with $\beta = \sigma_c^2 / \sigma_b^2$; hence $\mathbf{g}_q(i) = \beta \bar{\mathbf{g}}_q(i)$ for all i .

For the desired linear time-varying equalizer, we execute the following steps:

1. Pick L_e and d ($= \frac{L_e}{2}$ in the following simulations). Pick $Q \geq \tilde{Q}$, $P > L + L_e - d$.
2. Solve (7.3), given data $\mathbf{y}(n)$, for $\mathbf{g}_q(i)$ where $0 \leq i \leq L_e$ and $1 \leq q \leq Q$. Then

$$\mathbf{g}_d(n; i) = \sum_{q=1}^Q \mathbf{g}_q(i) e^{j\omega_q n}.$$

3. The equalized output is then given by

$$e_1(n) = \sum_{i=0}^{L_e} \mathbf{g}_d^H(n; i) \mathbf{y}(n - i) \approx \alpha_1 c(n - d) + \alpha_2 b(n - d) + \tilde{v}(n)$$

where $\tilde{v}(n)$ is the equalized noise. Estimate α_1 as

$$\hat{\alpha}_1 = \frac{\frac{1}{T} \sum_{n=0}^{T-1} e_1(n) c^*(n - d)}{\frac{1}{T} \sum_{n=0}^{T-1} |c(n - d)|^2} = \frac{\frac{1}{T} \sum_{n=0}^{T-1} e_1(n) c^*(n - d)}{\sigma_c^2}.$$

4. Define

$$e_2(n) = e_1(n) - \hat{\alpha}_1 c(n-d) \approx \alpha_2 b(n-d) + \tilde{v}(n).$$

Then we hard-quantize $e_2(n)$ to estimate $b(n-d)$.

7.3 Direct FIR Linear Equalization: Multiple Users

Multiple access schemes allow multiple users to share a common channel. Random access methods provide each user with a flexible way of gaining access to the channel whenever the user has information (packets) to be sent. In random access, typically when two packets collide, they are discarded and then have to be retransmitted. In wireless *ad hoc* networks (also known as mobile *ad hoc* networks—MANET's), absence of base stations limits the use of traditional media access control (MAC) protocols [34]. In *ad hoc* networks one needs some sort of distributed MAC requiring some form of random access which makes avoiding collisions difficult. Collisions arising from uncoordinated users decrease system throughput and worsen delay performance. Multiple packet reception (MPR) capability (or signal separation) is one way to resolve packet collisions and thereby enhance throughput, by using signal processing to separate multiple received signals [75]. Recently, wireless *ad hoc* networks with asynchronous transmissions have been considered in [15, 58, 60]. The approaches of [15, 58] use user-specific modulation induced cyclostationarity coupled with receive antenna array to achieve MPR for frequency-selective time-invariant channels. In [60] user-specific superimposed training signals (also called hidden pilots or implicit training) have been used for MPR for frequency-selective time-invariant channels. The objective of this section is to investigate approaches using user-specific superimposed training signals

for MPR in MANET's for transmissions over doubly-selective channels, with emphasis on asynchronous networks.

Consider a time-varying MIMO FIR linear channel with K inputs (users) and N outputs (receiver array with N elements at the destination node). Let $\{s_k(n)\}$ denote the k -th user's information sequence which is input to the MIMO doubly-selective channel with the k -th user's discrete-time impulse response $\{\mathbf{h}_k(n;l)\}$ (N -vector channel response at time n to a unit input at time $n-l$). Consider a typical (one-hop) MANET structure in an asynchronous mode. Assume K active users with a packet length of S symbols, in the coverage area of the node under evaluation. Each node is equipped with N (≥ 1) receive antennas and receiver node processes a data record block of size T ($\geq S$) symbols. Various packets can be located anywhere within this observation block. Using a sliding block approach (as in [15, 60]), we assume that the packet of interest is totally within the observation block. (An energy detector or related approaches can be used to ensure this [15, 60].) The noisy received (baseband-equivalent, symbol-rate) signal at the node-of-interest at time n is an N -column vector $\mathbf{y}(n)$, $n_1 \leq n \leq n_1 + T - 1$, given by (n_1 is the "initial" time of the observation block)

$$\mathbf{y}(n) = \sum_{k=1}^K \sum_{l=0}^L \mathbf{h}_k(n;l) s_k(n-l) + \mathbf{v}(n). \quad (7.7)$$

In a CE-BEM representation it is assumed that the channel for each user follows (2.9)

$$\mathbf{h}_k(n;l) = \sum_{q=1}^{\tilde{Q}} \mathbf{h}_{qk}(l) e^{j\omega_q n}, \quad k = 1, 2, \dots, K \quad (7.8)$$

where N -column vectors $\mathbf{h}_q(l)$ are invariant for the whole block $n = 0, 1, \dots, T - 1$, and

$$\begin{aligned}\tilde{Q} &:= 2\lceil f_d T T_s \rceil + 1, \\ L &:= \lfloor \tau_d / T_s \rfloor, \\ \omega_q &:= \frac{2\pi}{T} \left(q - \frac{1}{2} - \frac{Q}{2} \right), \quad q = 1, 2, \dots, Q.\end{aligned}$$

In superimposed training-based approaches, for the k -th user, one takes

$$s_k(n) = b_k(n) + c_k(n)$$

where $\{b_k(n)\}$ is the information sequence and $\{c_k(n)\}$ is a user-specific non-random periodic training sequence.

The main problem considered here is: How to design an equalizer to estimate $\{b_1(n)\}$, the information sequence of user 1 (the desired user), when one knows only $\{c_1(n)\}$ but not (obviously) $\{b_1(n)\}$ and one does not also have (frame) synchronization with $\{c_1(n)\}$ at the receiver. We will design an equalizer to estimate $\{c_1(n)\}$ with a delay d . In a manner similar to Section 7.2, we will then show that this equalizer is a scaled version of the corresponding equalizer designed to estimate $\{b_1(n)\}$ with a delay d provided that $\{c_k(n)\}$ satisfies certain properties.

7.3.1 User-Specific Training Sequences

Each user is assigned (or selects) a user-specific training sequence. The sequences

$$c_k(n) := \sigma_{ck} \exp \left[j2\pi \left(\frac{n^2}{\tilde{P}} + \alpha_k n \right) \right], \quad (7.9a)$$

$$\alpha_k := \frac{m-1}{D}, \quad m = 1, 2, \dots, D \geq K \quad (7.9b)$$

have been used in [58, 60], which are periodic with period $P = D\tilde{P}$ where D , \tilde{P} , and σ_{ck} are design parameters (D and \tilde{P} are coprime). Different users are characterized by different α_k 's and distinct sequences are mutually orthogonal and individually periodic white. (There is a common code book at each node of size D containing the possible values of α_k . During the “initial contact” period, a given node searches for all possible D signals.) In a different context, as in Section 3.5, we have proposed the user-specific sequences in (3.63)–(3.64) by using an m -sequence of periodic \tilde{P} with $P = D\tilde{P}$. These sequences are periodic with period P , mutually orthogonal and individually “nearly” periodic-white with period \tilde{P} .

Given the knowledge of the time-varying channel described by CE-BEM, we investigate direct design of time-varying FIR linear equalizers for doubly-selective channels using superimposed training and without first estimating the underlying channel response.

7.3.2 Linear LS Equalizers for the Desired User

We look for a time-varying linear equalizer $\mathbf{g}(n; l)$ ($l = 0, 1, \dots, L_e$) over the same time block as the received data with channel model (7.8). Following the discussions in Section 7.2.2, we assume

$$\mathbf{g}(n; l) = \sum_{q=1}^Q \mathbf{g}_q(l) e^{j\omega_q n}, \quad n = 0, 1, \dots, T-1.$$

In order to estimate the input sequence of the desired user (user 1, with no loss of generality) $\{s_1(n)\}$ (see (7.7)), we may seek a linear time-varying FIR estimator using CE-BEM to yield an estimate with equalization delay d

$$\hat{s}_1(n-d) = \sum_{i=0}^{L_e} \mathbf{g}^H(n; i) \mathbf{y}(n-i).$$

Similar to Section 7.2, we also seek an LS solution $\mathbf{g}(n; l)$ to minimize a cost such as

$$J = \frac{1}{T} \sum_{n=0}^{T-1} |s_1(n-d) - \hat{s}_1(n-d)|^2.$$

We first state the underlying model assumptions.

- (H7.3.1) The information sequence $\{b_k(n)\}$ is zero-mean, i.i.d. (independent and identically distributed), with $E\{|b_k(n)|^2\} = \sigma_{b_k}^2$. They are also independent across users ($k = 1, 2, \dots, K$).
- (H7.3.2) The measurement noise $\{\mathbf{v}(n)\}$ is zero-mean ($E\{\mathbf{v}(n)\} = \mathbf{0}$), white, independent of $\{b_k(n)\}$, with $E\{[\mathbf{v}(n+\tau)][\mathbf{v}(n)]^H\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$.
- (H7.3.3) The superimposed training sequence $c_k(n) = c_k(n+P)$ for all n is a non-random periodic sequence with period P . Let $\sigma_{c_k}^2 := (1/P) \sum_{n=1}^P |c_k(n)|^2$. The sequences are chosen as in (7.9), or in Section 3.5.
- (H7.3.4) Record length T and period P satisfy $TP^{-1} > \tilde{Q}$ and TP^{-1} is an integer. Moreover, $\tilde{P} > L + L_e - d$ where $d (\geq 0)$ is the desired equalization delay.

It then follows that [58, 77]

$$\frac{1}{P} \sum_{n=0}^{P-1} c_k(n) c_m^*(n - \tau) = \gamma_k(\tau) \delta(\tau \bmod \tilde{P}) \delta(k - m)$$

where

$$\gamma_k(\tau) = \begin{cases} \sigma_{c_k}^2 e^{j2\pi k\tau/D} & \text{for (7.9)} \\ \sigma_{c_k}^2 e^{j2\pi(k-1)\tau/P} & \text{for the sequences in Section 3.5.} \end{cases}$$

The periodic training sequence can be written as

$$c_k(n) = \sum_{m=0}^{P-1} c_{km} e^{j\alpha_m n}$$

where $\alpha_m := 2\pi m/P$. To design the time-varying linear equalizer to estimate a delayed version of the desired user's training sequence $c_1(n - d)$ ($0 \leq d \leq L_e$), we have

$$\hat{c}_1(n - d) = \sum_{i=0}^{L_e} \mathbf{g}_d^H(n; i) \mathbf{y}(n - i)$$

where we assume that

$$\mathbf{g}_d(n; i) = \sum_{q=1}^Q \mathbf{g}_q(i) e^{j\omega_q n}.$$

Choose $\mathbf{g}_q(i)$'s to minimize the time-averaged cost

$$J_c := \frac{1}{T} \sum_{n=0}^{T-1} |c_1(n - d) - \hat{c}_1(n - d)|^2,$$

and by taking the derivative of J_c and setting it to be zero, we have

$$\begin{aligned} 0 &= \frac{\partial J_c}{\partial \mathbf{g}_{q_1}^*(i_1)} \\ &= -\frac{1}{T} \sum_{n=0}^{T-1} e^{-j\omega_{q_1} n} \mathbf{y}(n - i_1) \left[c^*(n - d) - \sum_{i=0}^{L_e} \sum_{q=1}^Q e^{j\omega_q n} \mathbf{y}^H(n - i) \mathbf{g}_q(i) \right] \end{aligned}$$

for $i_1 = 0, 1, \dots, L_e$ and $q_1 = 1, 2, \dots, Q$. This leads to

$$\begin{aligned} &\sum_{i=0}^{L_e} \sum_{q=1}^Q \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{j(\omega_q - \omega_{q_1}) n} \mathbf{y}(n - i_1) \mathbf{y}^H(n - i) \right] \mathbf{g}_q(i) \\ &= \frac{1}{T} \sum_{n=0}^{T-1} c_1^*(n - d) e^{-j\omega_{q_1} n} \mathbf{y}(n - i_1) =: \mathbf{R}_c(q_1, i_1). \end{aligned} \quad (7.10)$$

To design the time-varying linear equalizer to estimate the desired user's information sequence $b_1(n - d)$ ($0 \leq d \leq L_e$), we have

$$\hat{b}_1(n - d) = \sum_{i=0}^{L_e} \bar{\mathbf{g}}_d^H(n; i) \mathbf{y}(n - i)$$

where we assume that

$$\bar{\mathbf{g}}_d(n; i) = \sum_{q=1}^Q \bar{\mathbf{g}}_q(i) e^{j\omega_q n}.$$

Choose $\bar{\mathbf{g}}_q(s)$'s to minimize

$$J_b := \frac{1}{T} \sum_{n=0}^{T-1} \left| b_1(n - d) - \hat{b}_1(n - d) \right|^2.$$

Mimicking the results for the superimposed training sequence-based equalization,

$$\begin{aligned}
& \sum_{i=0}^{L_e} \sum_{q=1}^Q \left[\frac{1}{T} \sum_{n=0}^{T-1} e^{j(\omega_q - \omega_{q_1})n} \mathbf{y}(n - i_1) \mathbf{y}^H(n - i) \right] \bar{\mathbf{g}}_q(i) \\
&= \frac{1}{T} \sum_{n=0}^{T-1} b_1^*(n - d) e^{-j\omega_{q_1}n} \mathbf{y}(n - i_1) =: \mathbf{R}_b(q_1, i_1). \tag{7.11}
\end{aligned}$$

Comparing (7.10) and (7.11), we see that (ignoring the equalizer coefficients) the left sides of the two are identical. We now seek to establish that for large T , $\mathbf{R}_c(q_1, i_1) = \beta \mathbf{R}_b(q_1, i_1)$ for all q_1, i_1 , for some scalar β , so that $\mathbf{g}_q(i) = \beta \bar{\mathbf{g}}_q(i)$ for all i .

Following similar derivations as in Section 7.2.2, by defining

$$\tilde{\gamma}_1(i) := \sum_{l=0}^L \gamma_1^*(d - i - l) \delta((d - i - l) \bmod \tilde{P})$$

we have

$$\lim_{T \rightarrow \infty} \mathbf{R}_c(q_1, i_1) \stackrel{\text{m.s.}}{=} \begin{cases} \tilde{\gamma}_1^*(i_1) e^{-j\omega_{q_1}i_1} \mathbf{h}_{q_1}((d - i_1) \bmod \tilde{P}) & \text{if } 1 \leq q_1 \leq \tilde{Q} \\ 0 & \text{otherwise} \end{cases} \tag{7.12}$$

for $i_1 = 0, 1, \dots, L_e$ and $q_1 = 1, 2, \dots, Q$. It is also shown that for $i_1 = 0, 1, \dots, L_e$ and $q_1 = 1, 2, \dots, Q$ but $1 \leq q_1 \leq \tilde{Q}$

$$\lim_{T \rightarrow \infty} \mathbf{R}_b(q_1, i_1) \stackrel{\text{m.s.}}{=} \mathbf{h}_{1q_1}(d - i_1) e^{-j\omega_{q_1}i_1} \sigma_{b1}^2. \tag{7.13}$$

If $\tilde{P} > L + L_e - d$, then (7.12) equals (7.13) (within a scale factor) and $\tilde{\gamma}_1(i_1) = \sigma_{c1}^2$. Therefore, for “large” T , $\mathbf{R}_c(q_1, i_1) = \beta \mathbf{R}_b(q_1, i_1)$ for all q_1, i_1 with $\beta = \sigma_{c1}^2 / \sigma_{b1}^2$; hence $\mathbf{g}_q(i) = \beta \bar{\mathbf{g}}_q(i)$ for all i .

For the desired equalizer design, we execute the following steps:

1. Pick L_e and d ($= \frac{L_e}{2}$ in the following simulations). Pick $Q \geq \tilde{Q}$, $P > L + L_e - d$.
2. Solve (7.10), given data $\mathbf{y}(n)$, for $\mathbf{g}_q(i)$ where $0 \leq i \leq L_e$ and $1 \leq q \leq Q$. Then

$$\mathbf{g}_d(n; i) = \sum_{q=1}^Q \mathbf{g}_q(i) e^{j\omega_q n}.$$

3. The equalized output is then given by

$$e_1(n) = \sum_{i=0}^{L_e} \mathbf{g}_d^H(n; i) \mathbf{y}(n-i) \approx \alpha_1 c_1(n-d) + \alpha_2 b_1(n-d) + \tilde{v}(n)$$

where $\tilde{v}(n)$ is the equalized noise plus multiple-user interference. Estimate α_1 as

$$\hat{\alpha}_1 = \frac{\frac{1}{T} \sum_{n=0}^{T-1} e_1(n) c_1^*(n-d)}{\frac{1}{T} \sum_{n=0}^{T-1} |c_1(n-d)|^2} = \frac{\frac{1}{T} \sum_{n=0}^{T-1} e_1(n) c_1^*(n-d)}{\sigma_{c_1}^2}.$$

4. Define

$$e_2(n) = e_1(n) - \hat{\alpha}_1 c_1(n-d) \approx \alpha_2 b_1(n-d) + \tilde{v}(n).$$

Then we hard-quantize $e_2(n)$ to estimate $b_1(n-d)$.

7.4 Simulation Examples

7.4.1 Direct FIR Equalization: Single User

In this example, we consider the direct FIR linear equalization using superimposed training and CE-BEM. We generate a doubly-selective SIMO Rayleigh fading channel following the Jakes' model with $N = 1, 2$, and 3 , and $L = 2$ (3 taps) in (7.1). In the simulation,

Direct equalizer: $f_d=0\text{Hz}$, $T_s=25\mu\text{s}$, $L=2$, $L_e=6$, $Q=5$, $P=15$, $T=405$, 500 runs.

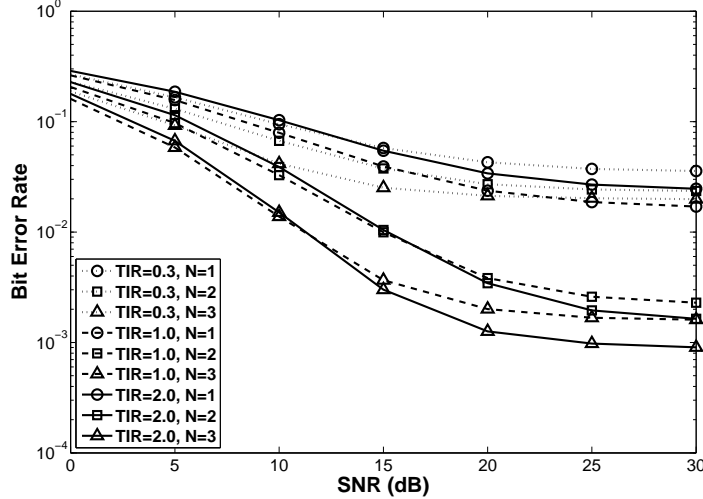


Figure 7.1: Single-user direct FIR equalization: BER vs SNR under $f_d = 0$ Hz and length of equalizer $L_e = 6$ with different TIR and number of receivers.

we pick a data record length of $T = 405$ symbols (time duration of approximately 10 ms). The communications system described in Section 2.6.3 with $T_s = 25 \mu\text{s}$ is employed. We consider the system operating under different Doppler spreads with the number of basis functions $Q = 5$. We choose TIR = 0.3, 1.0, and 2.0, so that the average transmitted power in $\{c(n)\}$ can be less, equal, or larger than the power in $\{b(n)\}$. The information sequence $\{b(n)\}$ is BPSK modulated. We take the superimposed training sequence of period $P = 15$ as $c(n) = \sigma_c e^{j\pi n(n+\nu)/P}$ where $\nu = 1$ if P is odd and $\nu = 2$ if P is even, as in [59]—this sequence is periodic white.

We first consider a single-user scenario. We assume the additive noise $\{\mathbf{v}(n)\}$ is zero-mean, white complex-Gaussian, uncorrelated with $\{b(n)\}$ with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$. The (receiver) SNR refers to the energy per bit over one-sided noise spectral

Direct equalizer: $f_d=50\text{Hz}$, $T_s=25\mu\text{s}$, $L=2$, $L_e=6$, $Q=5$, $P=15$, $T=405$, 500 runs.

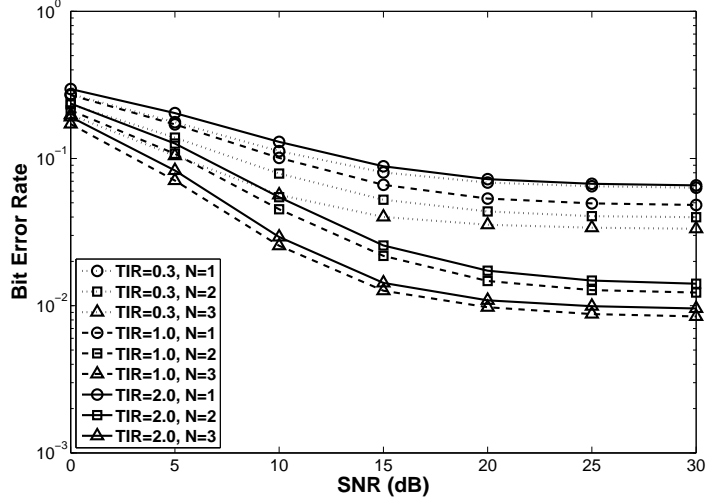


Figure 7.2: Single-user direct FIR equalization: BER vs SNR under $f_d = 50$ Hz and length of equalizer $L_e = 6$ with different TIR and number of receivers.

density with both information and superimposed training sequence counting toward the bit energy. All simulation results are based on 500 Monte Carlo runs.

The BER results for various SNR's and $f_d = 0, 50$, and 100 Hz are shown in Figures 7.1–7.3 respectively. We can clearly see that more receive antennas (larger N) will surely improve the reception, since space diversity can be exploited. However, increasing TIR does not necessarily benefit the performance, since higher TIR leads to more accurate equalizer taps as well as lower effective SNR due to less power assigned to information. Therefore, a trade-off has to be made in selecting TIR. Generally speaking, higher SNR or more receive antennas allows for more power allocated to superimposed training, due to less interference present. We also note that the error floors at the BER curves, which are partially attributed to the modeling error of the CE-BEM in approximating the inverse of the channel.

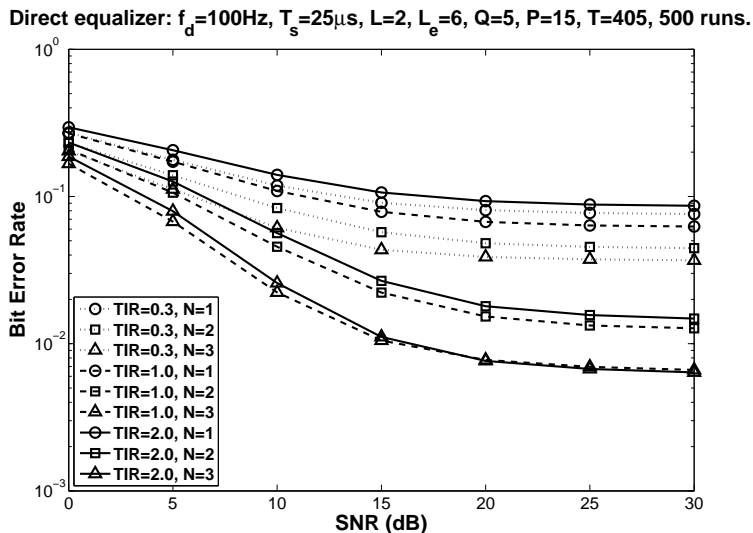


Figure 7.3: Single-user direct FIR equalization: BER vs SNR under $f_d = 100$ Hz and length of equalizer $L_e = 6$ with different TIR and number of receivers.

7.4.2 Direct FIR Equalization: Multiple Users

In this example, we consider a multiple-user scenario in a wireless *ad hoc* network. We set the number of users $K = 3$. For each user, the channel follows that described in Section 7.4.1, i.e., a doubly-selective SIMO Rayleigh fading channel following Jakes' model with $L = 2$ (3 taps). Now we take the number of receive antennas $N = 1, 2, 3$, and 4. For different users, the channel $\mathbf{h}_k(n; l)$ are mutually independent. We also employ the communications system described in Section 2.6.3 with $T_s = 25 \mu\text{s}$. In the simulation, we pick a data record length of $T = 832$ symbols. We assume the additive noise $\{\mathbf{v}(n)\}$ is zero-mean, white complex-Gaussian, uncorrelated with $\{b(n)\}$ with $E\{\mathbf{v}(n + \tau)\mathbf{v}^H(n)\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$. The (receiver) SNR refers to the energy per bit over one-sided noise spectral density with

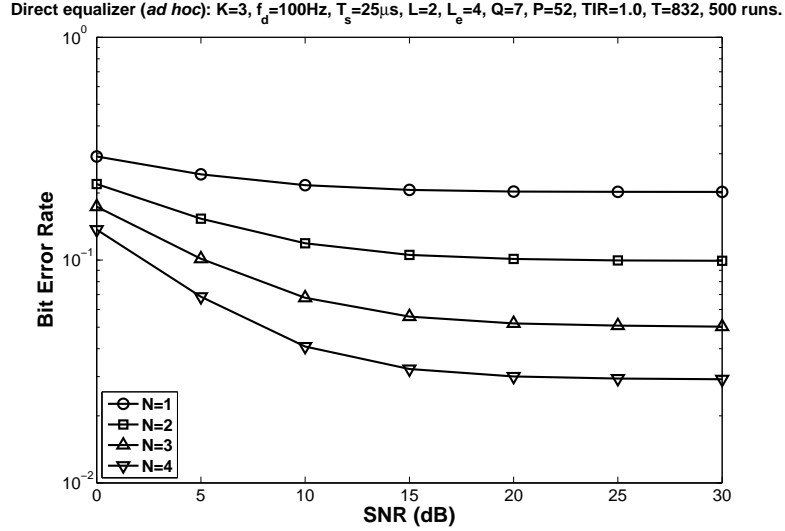


Figure 7.4: Multiple-user direct FIR equalization (*ad hoc*): BER vs SNR under $f_d = 100$ Hz and length of equalizer $L_e = 4$ with different number of receivers.

both information and superimposed training sequence counting toward the bit energy. All simulation results are based on 500 Monte Carlo runs.

Information sequences for each user are BPSK. We take the superimposed training sequences' period $P = 52$ with $D = 4$ and $\tilde{P} = 13$ in (7.9). The average transmitted power in $\{c_k(n)\}$ is equal to the power in $\{b_k(n)\}$, leading to TIR of 1.0.

In the simulation, we consider an asynchronous case where the observation window fully contains the desired user's signal and the other two interfering signals ($k = 2, 3$) occupy window $[t_k, t_k + T - 1]$ where t_k is uniformly distributed in $[-T + 1, T - 1]$; t_k changes from run to run.

Figure 7.4 shows the BER results versus SNR's for a channel with $f_d = 100$ Hz. We take the equalizer length $L_e = 4$, and the number of basis functions $Q = 7$. More receive antennas (larger N) enhance the reception significantly. However, due to the presence of

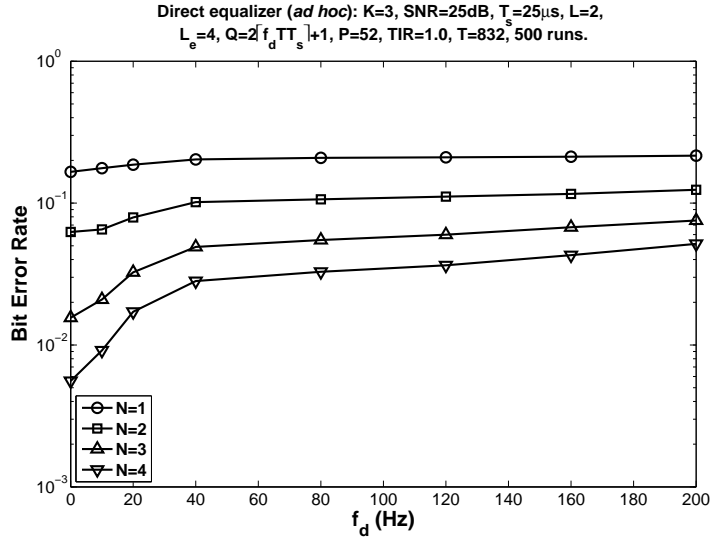


Figure 7.5: Multiple-user direct FIR equalization (*ad hoc*): BER vs Doppler spread f_d under $\text{SNR} = 25$ dB and length of equalizer $L_e = 4$ with different number of receivers.

MUI and the modeling error of CE-BEM, noticeable error floors can be observed at each curves. Figure 7.5 exhibits the BER's for various Doppler spreads. We also take $L_e = 4$, but $Q = \tilde{Q} = 2\lceil f_d T T_s \rceil + 1$, as a function of f_d . We can see that gradually BER's deteriorate with increasing Doppler spread f_d . The curves for BER's versus the equalizer length L_e for $f_d = 50$ Hz are displayed in Figure 7.6. Longer L_e may equalize the received signal better, but more taps add to estimation variance. We can see that $L_e = 4$ is in the neighborhood of the optimal length. The BER's in all these figures are rather high, due to MUI. This can be alleviated by error-correction coding.

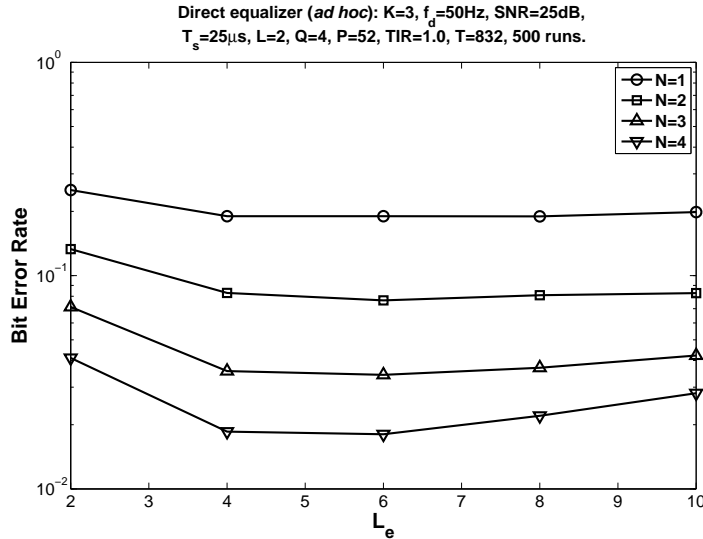


Figure 7.6: Multiple-user direct FIR equalization (*ad hoc*): BER vs length of equalizer L_e under $f_d = 50$ Hz and $\text{SNR} = 25$ dB with different number of receivers.

7.5 Conclusions

In this chapter, design of doubly-selective linear equalizers for single- and multiple-user frequency-selective time-varying channels was considered, using superimposed training and without first estimating the underlying channel response. Assuming that both the time-varying channel and the linear equalizers can be described by a CE-BEM representation, we showed that if periodic white superimposed training sequences are used, the optimal linear equalizer designed to extract the known training sequence was also a scaled version of the optimal equalizer for the information sequence. Based on this fact, a single-user direct equalizer was designed. By employing user-specific training sequences, this equalizer was extended to a multiple-user scenario, which can be used in a wireless *ad hoc* network.

CHAPTER 8

CONCLUDING REMARKS AND FUTURE WORK

This dissertation considered the issue of channel estimation and equalization and data detection, using superimposed training and BEM's, for wireless systems in doubly-selective channels. With a detailed analysis of the interference from the information sequence in superimposed training-based methods, our most important contribution is the approaches to suppress the self-interference.

Typical wireless channels are characterized by time- and frequency-selectivity: Multipath propagation and limited bandwidth result in frequency selectivity leading to ISI; temporal variation of the channel is attributed to the relative motion between the transmitter and the receiver, as well as oscillator drifts and phase noises.

In superimposed training-based estimation and equalization, at the transmitter, a periodic (non-random) training sequence is superimposed (at low power) to the information sequence, before modulation and transmission. Compared with conventional TM training, there is no loss in data transmission rate for superimposed training, but some useful power has been allocated to training sequences.

We described the doubly-selective channel, over a block of symbol intervals, by various BEM's (whereas we employed Jakes' model that is independent to BEM's to generate the "true" channel in simulations), including CE-, OP-, and DPS-BEM's, so that estimating the time-varying channel was reduced to estimating fewer time-invariant parameters.

8.1 Summary of Original Work

In Chapter 3, beginning with the first-order statistics-based channel estimator proposed in [81] that exploits superimposed training and CE-BEM, we modified this estimator to DPS- and OP-BEM's so that the spectral leakage induced by CE-BEM can be reduced. The OP-BEM-based estimator, moreover, can be applied to any BEM representation, and thus has a more general structure. By utilizing user-specific superimposed training sequences, we assigned distinct cycle frequencies of the periodic training sequences to distinct users so that channel estimation across different users is decoupled and the single-user estimator can be applied to this multiple-user scenario.

Performance analysis of the first-order statistics-based estimator proposed in the previous chapter was discussed in Chapter 4. Although the modeling error of BEM's and noise contribute to the channel estimation mismatch, performance analysis clearly shows that the major interference using superimposed training comes from the unknown information sequences. Power allocation and bias-variance trade-off of the first-order statistics-based estimator are also considered in this chapter, based on the results of performance analysis. We cast these optimization issues as maximization of an SNR for equalizer design.

How to reduce the information-induced self-interference, was the topic of the following two chapters. In Chapter 5, by exploiting the fact that training and information sequences pass through an identical channel, an iterative DML approach was proposed to jointly improve the channel and sequence estimation. Beginning with the first-order statistics-based channel estimator, we used the detected data symbols from the preceding iteration to reduce the self-interference at the current iteration. A local maximum of DML function is guaranteed by this method. To reduce the computational complexity of the ML detection,

symbol detection techniques such as Kalman filtering can also be adopted instead of the Viterbi algorithm used in the DML approach.

In contrast to the receiver-end DML approach, a data-dependent superimposed training scheme that is a transmitter-end processing technique, was proposed in Chapter 6. Inspired by the work of [20], we observed that over a channel satisfying a band-limited BEM such as CE- or DPS-BEM, the periodic superimposed training components within the received signal occur only at certain frequencies. By designing a data-dependent superimposed training sequence that suppresses the information sequence at those frequencies, the self-interference adversely affecting the channel estimation at the receiver is greatly reduced. However, the suppressed frequency components of the information sequence carry “information” as well. Therefore, a PDD superimposed training method was proposed to strike a trade-off between self-interference cancelation and information integrity. Performance analysis and related optimization of parameters were also discussed.

In Chapter 7, we considered direct equalization, without first estimating the doubly-selective channel, using superimposed training and CE-BEM. By exploiting periodic white training sequences, we showed that the optimal linear equalizer designed to extract the known training sequence was also a scaled version of the optimal equalizer for the information sequence. A direct equalizer was designed based on this fact. By employing user-specific training sequences, this direct equalizer was extended to a multiple-user scenario, which can be used in a wireless *ad hoc* network.

Computer simulation examples illustrated the performances of our approaches, and compared them with the conventional TM training schemes. The performance of the first-order statistics-based estimator is inferior to that of the TM training-based estimator, due to

the existence of information-induced self-interference. However, when the self-interference is sufficiently suppressed by the DML approach or the PDD training, superimposed training can offer competitive performance with TM training, without incurring any data rate loss, and thus provides us a promising training-based technique at a higher transmission rate.

8.2 Possible Future Directions

So far we have discussed channel estimation and equalization using superimposed training and various BEM representations. Future work may include the following areas.

First, the capacity of the system employing superimposed training should be well investigated. In other words, the fundamental question should be answered—how can superimposed training help achieve more capacity than TM training? As we discussed in Chapter 1, several researchers have explored this area (e.g., [4, 5, 8, 73], among others) and obtained important results; however, conclusive statements for more general situations (e.g., over time- or frequency-selective channel or doubly-selective channel) are still open.

Another potential direction may lie in incorporating superimposed training into other widely used techniques. For example, in orthogonal frequency division multiplexing (OFDM) systems, frequency-multiplexed training is used where equally spaced pilot (training) tones enable the receiver to achieve MMSE estimate of the channel [54]. This frequency-multiplexed training scheme, however, can be viewed as superimposed training in the time domain. Therefore, how to utilize our results of superimposed training in OFDM systems becomes an interesting topic.

We should also consider a more general training structure—affine precoding with training [89] that treats the transmitted data block as

$$\mathbf{s} = \mathbf{F}\mathbf{b} + \mathbf{c},$$

where $\mathbf{s} = \begin{bmatrix} s(0) & s(1) & \dots & s(T-1) \end{bmatrix}^T$ is a transmitted block, and \mathbf{c} is the training sequence of the same size; $\mathbf{b} = \begin{bmatrix} b(0) & b(1) & \dots & b(T_b-1) \end{bmatrix}^T$ is the information sequence of length $T_b \leq T$, and the $T \times T_b$ matrix \mathbf{F} is the affine precoder. We can clearly see that TM and superimposed training can both be viewed as special cases of affine precoding. In our data-dependent training scheme, our work is equivalent to designing a precoder \mathbf{F} that assign training and information sequences to different dimensions so as to eliminate self-interference. The information loss of this scheme is due to \mathbf{F} not being full-rank. We hence design a full-rank \mathbf{F} that corresponds to the PDD superimposed training. Therefore, affine precoding can offer us more freedom in designing the communications system, so that better performance than superimposed or TM training can be expected—it is also a hopeful area.

BIBLIOGRAPHY

- [1] E. Alameda-Hernández, D. C. McLernon, A. G. Orozco-Lugo, M. Lara, and M. Ghogho, “Synchronisation for superimposed training based channel estimation,” *Electron. Lett.*, vol. 41, no. 9, pp. 565–566, Apr. 2005.
- [2] 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.
- [3] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: information-theoretic and communications aspects,” *IEEE Trans. Wireless Commun.*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [4] P. Bohlin and M. Tapio, “Optimized data aided training in MIMO systems,” in *Proc. IEEE VTC’04-Spring*, Milan, Italy, May 17–19, 2004, pp. 679–683.
- [5] —, “Performance evaluation of MIMO communication systems based on superimposed pilots,” in *Proc. IEEE ICASSP’04*, vol. 4, Montreal, Canada, May 17–21, 2004, pp. 425–428.
- [6] D. K. Borah and B. D. Hart, “Frequency-selective fading channel estimation with a polynomial time-varying channel model,” *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 862–873, June 1999.
- [7] R. Bosisio, M. Nicoli, and U. Spagnolini, “Kalman filter of channel modes in time-varying wireless systems,” in *Proc. IEEE ICASSP’05*, vol. 3, Philadelphia, PA, Mar. 18–23 2005, pp. 785–788.
- [8] C. Budianu and L. Tong, “Channel estimation for space-time orthogonal block codes,” *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2515–2528, Oct. 2002.
- [9] N. Chen and G. T. Zhou, “What is the price paid for superimposed training in OFDM,” in *Proc. IEEE ICASSP’04*, vol. 4, Montreal, Canada, May 17–21, 2004, pp. 421–424.
- [10] —, “Superimposed training for OFDM: a peak-to-average power ratio analysis,” *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2277–2287, June 2006.
- [11] W. Chen and R. Zhang, “Estimation of time and frequency selective channels in OFDM systems: a Kalman filter structure,” in *Proc. IEEE GLOBECOM ’04*, vol. 2, Dallas, TX, Nov. 29–Dec. 3, 2004, pp. 800–803.

- [12] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, vol. 47, no. 6, pp. 957–1000, July-Aug. 1968.
- [13] T. Cui and C. Tellambura, "Superimposed pilot symbols for channel estimation in OFDM systems," in *Proc. IEEE GLOBECOM'05*, San Francisco, CA, Nov. 28–Dec. 2, 2005, pp. 2229–2233.
- [14] A. V. Dandawaté and G. B. Giannakis, "Asymptotic theory of mixed time averages and k th-order cyclic-moment and cumulant statistics," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 216–232, Jan. 1995.
- [15] R. Djapic, A.-J. van der Veen, and L. Tong, "Synchronization and packet separation in wireless *ad hoc* networks by known modulus algorithm," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 51–64, Jan. 2005.
- [16] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403–1418, May 2004.
- [17] B. Farhang-Boroujeny, "Pilot-based channel identification: proposal for semi-blind identification of communication channels," *Electron. Lett.*, vol. 31, no. 13, pp. 1044–1046, June 1995.
- [18] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [19] M. Ghogho, "Channel and DC-offset estimation using data-dependent superimposed training," *Electron. Lett.*, vol. 41, no. 22, pp. 1250–1251, Oct. 2005.
- [20] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 226–229, Mar. 2005.
- [21] —, "SISO and MIMO channel estimation and symbol detection using data-dependent superimposed training," in *Proc. IEEE ICASSP'05*, vol. 3, Philadelphia, PA, Mar. 18–23 2005, pp. 461–464.
- [22] M. Ghogho and A. Swami, "Optimal training for affine-precoded and cyclic-prefixed block transmissions," in *Proc. IEEE Workshop Statistical Signal Processing*, Bordeaux, France, July 17–20, 2005, pp. 1358–1363.
- [23] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds., *Signal Processing Advances in Wireless and Mobile Communications Volume 1: Trends in Channel Estimation and Equalization*. Englewood Cliffs, NJ: Prentice Hall, 2001.

- [24] G. B. Giannakis and C. Tepedelenlioglu, “Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels,” *Proc. IEEE*, vol. 86, no. 10, pp. 1969–1986, Oct. 1998.
- [25] A. Gorokhov and P. Loubaton, “Semi-blind second order identification of convolutive channels,” in *Proc. IEEE ICASSP’97*, vol. 5, Munich, Germany, Apr. 21–24 1997, pp. 3905–3908.
- [26] F. J. Harris, “On the use of windows for harmonic analysis with the discrete Fourier transform,” *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.
- [27] S. He and J. K. Tugnait, “Direct equalization of multiuser doubly selective channels based on superimposed training,” in *Proc. EUSIPCO’06*, Florence, Italy, Sept. 4–8, 2006.
- [28] —, “Doubly-selective channel estimation using superimposed training and discrete prolate spheroidal basis models,” in *Proc. IEEE GLOBECOM’06*, San Francisco, CA, Nov. 27–Dec. 1, 2006.
- [29] —, “On bias-variance trade-off in superimposed training-based doubly selective channel estimation,” in *Proc. 2006 Conf. Inform. Sciences & Syst.*, Princeton University, NJ, Mar. 22–24, 2006, pp. 1308–1313.
- [30] —, “Doubly-selective multiuser channel estimation using superimposed training and discrete prolate spheroidal basis expansion models,” in *Proc. IEEE ICASSP’07*, vol. 2, Honolulu, HI, Apr. 15–20, 2007, pp. 861–864.
- [31] —, “Self-interference suppression in doubly-selective channel estimation using superimposed training,” in *Proc. IEEE ICC’07*, Glasgow, UK, June 24–28, 2007.
- [32] S. He, J. K. Tugnait, and X. Meng, “On superimposed training for MIMO channel estimation and symbol detection,” *IEEE Trans. Signal Process.*, vol. 55, no. 6, June 2007.
- [33] P. Hoehner and F. Tufvesson, “Channel estimation with superimposed pilot sequence,” in *Proc. IEEE GLOBECOM’99*, Rio de Janeiro, Dec. 5–9, 2006, pp. 2162–2166.
- [34] *IEEE J. Sel. Areas Commun., Special Issue on Wireless Ad Hoc Networks*, vol. 23, Jan. 2005.
- [35] W. C. Jakes, *Microwave Mobile Communications*. New York, NY: Wiley, 1974.
- [36] C. E. Kastenholtz and W. P. Birkemeier, “A simultaneous information transfer and channel sounding modulation technique for wide-band channels,” *IEEE Trans. Commun.*, vol. 13, no. 2, pp. 162–165, June 1965.

- [37] 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.
- [38] 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, Feb. 2002.
- [39] X. Ma and G. B. Giannakis, "Maximum-diversity transmissions over doubly selective wireless channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1832–1840, July 2003.
- [40] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.
- [41] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453–466, Mar. 2005.
- [42] J. H. Manton, I. Y. Mareels, and Y. Hua, "Affine precoders for reliable communications," in *Proc. IEEE ICASSP'00*, Istanbul, Turkey, June 5–9, 2000, pp. 2749–2752.
- [43] M. Martone, "Wavelet-based separating kernels for array processing of cellular DS/CDMA signals in fast fading," *IEEE Trans. Commun.*, vol. 48, no. 6, pp. 979–995, June 2000.
- [44] F. Mazzenga, "Channel estimation and equalization for m -QAM transmission with a hidden pilot sequence," *IEEE Trans. Broadcast.*, vol. 46, no. 2, pp. 170–176, June 2000.
- [45] D. C. McLernon, A. G. Orozco-Lugo, and M. M. Lara, "On the structural equivalence of two recent algorithms for implicitly trained channel estimation," in *Proc. IEEE Int. Symposium Signal Processing and Inform. Technol.*, Rome, Italy, Dec. 18–21 2004, pp. 132–135.
- [46] X. Meng, "Estimation of wireless communications channels using superimposed training: Approaches, analysis and applications," Ph.D. dissertation, Auburn University, Auburn, AL, May 2005.
- [47] X. Meng and J. K. Tugnait, "Doubly-selective MIMO channel estimation using superimposed training," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Barcelona, Spain, July 18–21, 2004, pp. 407–411.
- [48] —, "MIMO channel estimation using superimposed training," in *Proc. IEEE ICC'04*, Paris, France, June 20–24, 2004, pp. 2663–2667.
- [49] —, "Semi-blind channel estimation and detection using superimposed training," in *Proc. IEEE ICASSP'04*, vol. 4, Montreal, Canada, May 17–21, 2004, pp. 417–420.

- [50] —, “Semi-blind time-varying channel estimation using superimposed training,” in *Proc. IEEE ICASSP’04*, vol. 3, Montreal, Canada, May 17–21, 2004, pp. 797–800.
- [51] —, “Superimposed training-based doubly-selective channel estimation using exponential and polynomial bases models,” in *Proc. 2004 Conf. Inform. Sciences & Syst.*, Princeton University, NJ, Mar. 17–19, 2004.
- [52] X. Meng, J. K. Tugnait, and S. He, “Iterative joint channel estimation and data detection using superimposed training: algorithms and performance analysis,” *IEEE Trans. Veh. Technol.*, to be published.
- [53] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River: NJ: Prentice Hall, 2000.
- [54] R. Negi and J. Cioffi, “Pilot tone selection for channel estimation in a mobile OFDM system,” *IEEE Trans. Consum. Electron.*, vol. 44, no. 3, pp. 1122–1128, Aug. 1998.
- [55] M. Niedźwiecki, *Identification of Time-Varying Processes*. New York, NY: Wiley, 2000.
- [56] S. Ohno and G. B. Giannakis, “Optimal training and redundant precoding for block transmissions with application to wireless OFDM,” *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2113–2123, Dec. 2002.
- [57] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Englewood Cliffs: NJ: Prentice Hall, 1999.
- [58] A. G. Orozco-Lugo, G. M. Galvan-Tejada, M. M. Lara, and D. C. McLernon, “A new approach to achieve multiple packet reception for *ad hoc* networks,” in *Proc. IEEE ICASSP’04*, vol. 4, Montreal, Canada, May 17–21, 2004, pp. 429–432.
- [59] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, “Channel estimation using implicit training,” *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 240–254, Jan. 2004.
- [60] A. G. Orozco-Lugo, M. M. Lara, D. C. McLernon, and H. J. Muro-Lemus, “Multiple packet reception in wireless *ad hoc* networks using polynomial phase-modulating sequences,” *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2093–2110, Aug. 2003.
- [61] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, UK: Cambridge University Press, 2003.
- [62] H. V. Poor and S. Verdú, “Probability of error in MMSE multiuser detection,” *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 858–871, May 1997.
- [63] M. F. Pop and N. C. Beaulieu, “Limitations of sum-of-sinusoids fading channel simulators,” *IEEE Trans. Commun.*, vol. 49, no. 4, pp. 699–708, Apr. 2001.

- [64] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY: McGraw-Hill, 2001.
- [65] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 1996.
- [66] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2002.
- [67] A. M. Sayeed and B. Aazhang, "Joint multipath-Doppler diversity in mobile wireless communications," *IEEE Trans. Commun.*, vol. 47, no. 1, pp. 123–132, Jan. 1999.
- [68] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Trans. Commun.*, vol. 42, no. 2-4, pp. 1000–1011, Feb.–Apr. 1994.
- [69] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty–V: the discrete case," *Bell Syst. Tech. J.*, vol. 57, no. 5, pp. 1371–1430, May-June 1978.
- [70] M. D. Srinath, P. K. Rajasekaran, and R. Viswanathan, *Introduction to Statistical Signal Processing with Applications*. Upper Saddle River, NJ: Prentice Hall, 1996.
- [71] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. Boston, MA: Kluwer, 2002.
- [72] Z. Tang, R. C. Cannizzaro, G. Leus, and P. Banelli, "Pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2226–2238, May 2007.
- [73] M. Tapio and P. Bohlin, "A capacity comparison between time-multiplexed and superimposed pilots," in *Proc. 38th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 7–10, 2004, pp. 1049–1053.
- [74] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.
- [75] L. Tong, Q. Zhao, and G. Mergen, "Multipacket reception in random access wireless networks: from signal processing to optimal medium access control," *IEEE Commun. Mag.*, vol. 39, no. 11, pp. 108–112, Nov. 2001.
- [76] F. Tsuzuki and T. Ohtsuki, "Channel estimation with selective superimposed pilot sequences under fast fading environments," in *Proc. IEEE VTC'04-Fall*, Los Angeles, CA, Sept. 26–29, 2004, pp. 62–66.
- [77] J. K. Tugnait and S. He, "Performance analysis of an MIMO channel estimator based on superimposed training and first-order statistics," in *Proc. IEEE Workshop Statistical Signal Processing*, Bordeaux, France, July 17–20, 2005, pp. 1336–1341.

- [78] —, “Direct FIR linear equalization of doubly selective channels based on superimposed training,” in *Proc. IEEE ICASSP’06*, vol. 4, Toulouse, France, May 14–19 2006, pp. 589–592.
- [79] —, “Doubly-selective channel estimation using data-dependent superimposed training and exponential bases models,” in *Proc. 2006 Conf. Inform. Sciences & Syst.*, Princeton University, NJ, Mar. 22–24, 2006, pp. 375–380.
- [80] J. K. Tugnait, S. He, and X. Meng, “On superimposed-training power allocation for time-varying channel estimation,” in *Proc. IEEE Workshop Statistical Signal Processing*, Bordeaux, France, July 17–20, 2005, pp. 1330–1335.
- [81] J. K. Tugnait and W. Luo, “On channel estimation using superimposed training and first-order statistics,” in *Proc. IEEE ICASSP’03*, vol. 4, Hong Kong, Apr. 6–10, 2003, pp. 624–627.
- [82] —, “On channel estimation using superimposed training and first-order statistics,” *IEEE Commun. Lett.*, vol. 8, no. 9, pp. 413–415, Sept. 2003.
- [83] J. K. Tugnait and X. Meng, “Synchronization of superimposed training for channel estimation,” in *Proc. IEEE ICASSP’04*, vol. 4, Montreal, Canada, May 17–21, 2004, pp. 853–856.
- [84] —, “Performance analysis and training power allocation for channel estimation using superimposed training,” in *Proc. IEEE ICASSP’05*, vol. 3, Philadelphia, PA, Mar. 18–23 2005, pp. 457–460.
- [85] —, “On superimposed training for channel estimation: performance analysis, training power allocation, and frame synchronization,” *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 752–765, Feb. 2006.
- [86] J. K. Tugnait, X. Meng, and S. He, “Doubly-selective channel estimation using superimposed training and exponential bases models,” *EURASIP J. Applied Signal Processing (Special Issue on Reliable Communications over Rapidly Time-Varying Channels)*, vol. 2006, Article ID 85303, 11 pages, 2006.
- [87] J. K. Tugnait, L. Tong, and Z. Ding, “Single-user channel estimation and equalization,” *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 17–28, May 2000.
- [88] S. Verdú, “Minimum probability of error for asynchronous Gaussian multiple-access channels,” *IEEE Trans. Inf. Theory*, vol. 32, no. 1, pp. 85–96, Jan. 1986.
- [89] A. Vosoughi and A. Scaglione, “Everything you always wanted to know about training: guidelines derived using the affine precoding framework and the CRB,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 940–954, Mar. 2006.

- [90] 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.
- [91] X. Wang and H. V. Poor, *Wireless Communication Systems: Advanced Techniques for Signal Reception*. Upper Saddle River, NJ: Prentice Hall, 2004.
- [92] J. J. Werner, J. Yang, D. D. Harman, and G. A. Dumont, "Blind equalization for broadband access," *IEEE Commun. Mag.*, vol. 37, no. 4, pp. 87–93, Apr. 1999.
- [93] Q. Yang and K. S. Kwak, "Superimposed-pilot-aided channel estimation for mobile OFDM," *Electron. Lett.*, vol. 42, no. 12, pp. 73–74, June 2006.
- [94] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, Sept. 2005.
- [95] Q. Zhao and L. Tong, "Semi-blind equalization by least squares smoothing," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1–4, 1998, pp. 645–649.
- [96] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, June 2003.
- [97] G. T. Zhou and N. Chen, "Superimposed training for doubly selective channels," in *Proc. IEEE Workshop Statistical Signal Processing*, St. Louis, MO, Sept. 28–Oct. 1, 2003, pp. 82–85.
- [98] G. T. Zhou, M. Viberg, and T. McKelvey, "A first-order statistical method for channel estimation," *IEEE Signal Process. Lett.*, vol. 10, no. 3, pp. 57–60, Mar. 2003.

APPENDICES

APPENDIX A

OPTIMAL TIME-MULTIPLEXED TRAINING FOR BLOCK TRANSMISSIONS OVER DOUBLY-SELECTIVE CHANNELS [40]

Here we summarize the optimal TM training proposed in [40], based on CE-BEM representations, which acts as a “reference” training scheme in evaluating our superimposed training-based approaches.

In [40], the authors made the following model assumptions:

- (HA.1)** The channel satisfies CE-BEM, i.e., (2.9).
- (HA.2)** The delay spread τ_d and the Doppler spread f_d are bounded, known (or at least their bounds are known), and satisfy $2f_d\tau_d < 1$.
- (HA.3)** The coefficients $\{h_q(l)\}$ are zero-mean complex Gaussian random variables, independent one another, and remain invariant per block but are allowed to change at the next block.

Under the above assumptions, the authors sought to design a TM training scheme that optimizes channel MSE and ergodic (average) capacity bounds to jointly account not only for channel estimation performance but also for transmission rate. The optimal block structure \mathbf{s} consists of sub-blocks of training and sub-blocks of information, which are transmitted alternately:

$$\mathbf{s} = \left[\mathbf{b}_0^T \quad \mathbf{c}_0^T \quad \cdots \quad \mathbf{b}_{P-1}^T \quad \mathbf{c}_{P-1}^T \right]^T$$

where \mathbf{b}_p and \mathbf{c}_p denote the p -th information and training sub-blocks respectively ($p = 0, 1, \dots, P-1$). The optimal training sequence contains an impulse guarded by zeros (silent

Parameters	Optimal Training
Placement of information symbols	Equally long information sub-blocks (length \bar{T}_b)
Placement of training symbols	Equally long training sub-blocks
Structure of training sub-blocks	$\mathbf{c}_p = [\mathbf{0}_L^T \quad c \quad \mathbf{0}_L^T]^T, \forall p$
Number of training symbols	$2L + 1$ per sub-block
Number of sub-blocks	Q training and Q information sub-blocks
Power Allocation	$\alpha = 1/(1 + \sqrt{(L + 1)/\bar{T}_b})$

Table A.1: Optimal TM training.

period). The details are shown in Table A.1, where \bar{T}_b denotes the length of the information sub-block, c is a scalar number denoting the training impulse, and α denotes the training-to-information power ratio (TIR).

We also apply this TM training structure for OP- and DPS-BEM's. It is not known whether it is optimal for these BEM's, but it can still act as a good reference when studying our superimposed training schemes.

For a channel satisfying a BEM (2.20), the noisy channel output is given by

$$y(n) = \sum_{l=0}^L \sum_{q=1}^Q h_q(l) \psi_q(n) s(n-l) + v(n),$$

where $v(n)$ is additive white Gaussian noise (AWGN) with zero-mean and variance σ_v^2 . At the time slots $n_{p,l} := \bar{T}_b + L + p(\bar{T}_b + 2L + 1) + l$ ($l = 0, 1, \dots, L$), the received signal depends only on the channel and the training impulse:

$$y(n_{p,l}) = c \sum_{q=1}^Q h_q(l) \psi_q(n_{p,l}) + v(n_{p,l}). \quad (\text{A.1})$$

We define

$$\mathbf{y}_c(l) = \begin{bmatrix} y(n_{0,l}) & y(n_{1,l}) & \cdots & y(n_{P-1,l}) \end{bmatrix}^T, \quad (\text{A.2a})$$

$$\mathbf{v}_c(l) = \begin{bmatrix} v(n_{0,l}) & v(n_{1,l}) & \cdots & v(n_{P-1,l}) \end{bmatrix}^T, \quad (\text{A.2b})$$

$$\mathbf{\Psi}_c(l) = \begin{bmatrix} \psi_1(n_{0,l}) & \cdots & \psi_Q(n_{0,l}) \\ \psi_1(n_{1,l}) & \cdots & \psi_Q(n_{1,l}) \\ \vdots & \ddots & \vdots \\ \psi_1(n_{P-1,l}) & \cdots & \psi_Q(n_{P-1,l}) \end{bmatrix}, \quad (\text{A.2c})$$

$$\mathbf{h}(l) = \begin{bmatrix} h_1(l) & h_2(l) & \dots & h_Q(l) \end{bmatrix}^T,$$

then by (A.1)

$$\mathbf{y}_c(l) = c\mathbf{\Psi}_c(l) \mathbf{h}(l) + \mathbf{v}_c(l).$$

The LS estimator of $\mathbf{h}(l)$ is given by

$$\hat{\mathbf{h}}_{\text{LS}}(l) = \frac{1}{c} \mathbf{\Psi}_c^\dagger(l) \mathbf{y}_c(l), \quad (\text{A.3})$$

and the linear MMSE estimator is given by

$$\hat{\mathbf{h}}_{\text{MMSE}}(l) = \frac{c}{\sigma_v^2} \left[\mathbf{R}_h^{-1}(l) + \frac{c^2}{\sigma_v^2} \mathbf{\Psi}_c^H(l) \mathbf{\Psi}_c(l) \right]^{-1} \mathbf{\Psi}_c^H(l) \mathbf{y}_c(l) \quad (\text{A.4})$$

that requires that $\mathbf{R}_h(l) := E \{ \mathbf{h}(l) \mathbf{h}^H(l) \}$ is known at the receiver.

This optimal TM training can be easily extended to an SIMO channel, since the above procedures can be carried out for each independent output. For a multiple-user time-invariant channel, an extension of this scheme was suggested in [41] for MIMO frequency-selective fading channels. We extend this scheme to a doubly-selective MIMO channel in a way similar to [40], even though we have no clues about its optimality.

For a multiple-user channel with K users,

$$y(n) = \sum_{k=1}^K \sum_{l=0}^L h_k(n; l) s_k(n-l) + v(n),$$

where $\{s_k(n)\}$ denotes the k -th user's information sequence and the corresponding channel is denoted by $\{h_k(n; l)\}$. We assume that the channels satisfy BEM representation, i.e.,

$$h_k(n; l) = \sum_{q=1}^Q h_{qk}(l) \psi_q(n).$$

The channel output can be expressed as

$$y(n) = \sum_{k=1}^K \sum_{l=0}^L \sum_{q=1}^Q h_{qk}(l) \psi_q(n) s_k(n-l) + v(n).$$

We design the training sub-block of the k -th user ($k = 1, 2, \dots, K$) as

$$\mathbf{c}_{k,p} = \left[\mathbf{0}_{(k-1)(L+1)+L}^T \quad c \quad \mathbf{0}_{(K-k)(L+1)+L}^T \right]^T$$

with length of $K(L+1) + L$ symbols. At the time slots

$$n_{k,p,l} := \bar{T}_b + (k-1)(L+1) + L + p [\bar{T}_b + K(L+1) + L] + l,$$

the received signal depends only on the training impulse and the k -th user's channel. The channels for different users can thus be decoupled. We have

$$\mathbf{y}_c(k, l) = c\mathbf{\Psi}_c(k, l)\mathbf{h}_k(l) + \mathbf{v}_c(k, l),$$

where $\mathbf{y}_c(k, l)$, $\mathbf{\Psi}_c(k, l)$, and $\mathbf{v}_c(k, l)$ are defined as in (A.2), only with $n_{p,l}$ replaced with $n_{k,p,l}$, and

$$\mathbf{h}_k(l) = \begin{bmatrix} h_{1k}(l) & h_{2k}(l) & \dots & h_{Qk}(l) \end{bmatrix}^T.$$

Then similar to (A.3) and (A.4), we have the LS estimator

$$\hat{\mathbf{h}}_{k,\text{LS}}(l) = \frac{1}{c}\mathbf{\Psi}_c^\dagger(k, l)\mathbf{y}_c(k, l), \quad (\text{A.5})$$

and the linear MMSE estimator

$$\hat{\mathbf{h}}_{k,\text{MMSE}}(l) = \frac{c}{\sigma_v^2} \left[\mathbf{R}_h^{-1}(k, l) + \frac{c^2}{\sigma_v^2} \mathbf{\Psi}_c^H(k, l) \mathbf{\Psi}_c(k, l) \right]^{-1} \mathbf{\Psi}_c^H(k, l) \mathbf{y}_c(k, l) \quad (\text{A.6})$$

if $\mathbf{R}_h(k, l) := E\{\mathbf{h}_k(l)\mathbf{h}_k^H(l)\}$ is known at the receiver.

APPENDIX B

SYMBOL DETECTION

The role of channel estimation is to aid in extracting the desired information data from the distorted received symbols. Two symbol detection techniques are reviewed: Viterbi detector and Kalman filter.

B.1 Maximum Likelihood Sequence Detector (Viterbi Detector) [64]

Consider an SIMO FIR linear channel with N outputs and discrete-time impulse response $\{\mathbf{h}(n;l)\}$. Let $\{s(n)\}$ denote the input sequence to the SIMO channel. The channel output is given by

$$\mathbf{x}(n) = \sum_{l=0}^L \mathbf{h}(n;l)s(n-l), \quad (\text{B.1})$$

and the noisy measurement is given by

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n) \quad (\text{B.2})$$

where $\mathbf{v}(n)$ is the AWGN. We assume:

(HB.1) The $\{\mathbf{v}(n)\}$ is uncorrelated with $\{s(n)\}$, with possible unknown mean $E\{\mathbf{v}(n)\} = \mathbf{m}$ and $E\{[\mathbf{v}(n+\tau) - \mathbf{m}][\mathbf{v}(n) - \mathbf{m}]^H\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$.

Given $\{s(n)\}$, $\{\mathbf{y}(n)\}$ is a sequence of N -dimensional Gaussian random vectors with mean $\sum_{l=0}^L \mathbf{h}(n;l)s(n-l) + \mathbf{m}$ and variance $\sigma_v^2 \mathbf{I}_N$. The joint pdf of $\mathbf{y}(n)$ given $\{s(n), s(n-$

$1), \dots, s(n-L)\}$ is

$$p(\mathbf{y}(n)|s(n), \dots, s(n-L)) = \frac{1}{(\pi\sigma_v)^N} \exp \left\{ -\frac{1}{\sigma_v^2} \left\| \mathbf{y}(n) - \sum_{l=0}^L \mathbf{h}(n;l)s(n-l) - \mathbf{m} \right\|^2 \right\}$$

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

$$\begin{aligned} & p(\mathbf{y}(0), \dots, \mathbf{y}(T-1)|s(0), \dots, s(T-1)) \\ &= \frac{1}{(\pi\sigma_v)^{NT}} \exp \left\{ -\frac{1}{\sigma_v^2} \sum_{n=0}^{T-1} \left\| \mathbf{y}(n) - \sum_{l=0}^L \mathbf{h}(n;l)s(n-l) - \mathbf{m} \right\|^2 \right\}. \end{aligned}$$

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

$$\begin{aligned} & \log p(\mathbf{y}(0), \dots, \mathbf{y}(T-1)|s(0), \dots, s(T-1)) \\ &= -NT \log(\pi\sigma_v) - \frac{1}{\sigma_v^2} \sum_{n=0}^{T-1} \left\| \mathbf{y}(n) - \sum_{l=0}^L \mathbf{h}(n;l)s(n-l) - \mathbf{m} \right\|^2. \end{aligned}$$

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

$$p(\mathbf{y}(0), \dots, \mathbf{y}(T-1) | s(0), \dots, s(T-1)),$$

or equivalently maximizes

$$\log p(\mathbf{y}(0), \dots, \mathbf{y}(T-1) | s(0), \dots, s(T-1)),$$

or minimizes the Euclidean distance

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

This ML sequence estimation (MLSE) 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 $\{\mathbf{h}(n;l)\}$ and its state at any time instance n is represented by the L most recent input symbols

$$\text{state}(n) = (s(n), s(n-1), \dots, s(n-L+1)) \quad (\text{B.3})$$

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 as the following three steps:

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

$$\sum_{n=0}^L \left\| \mathbf{y}(n) - \sum_{l=0}^L \mathbf{h}(n;l) s(n-l) - \mathbf{m} \right\|^2. \quad (\text{B.4})$$

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

$$\text{PM}_0(s(L), \dots, s(1)) = \min_{s(0)} \left\{ \sum_{n=0}^L \left\| \mathbf{y}(n) - \sum_{l=0}^L \mathbf{h}(n;l) s(n-l) - \mathbf{m} \right\|^2 \right\}. \quad (\text{B.5})$$

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

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

$$\left\| \mathbf{y}(L + n) - \sum_{l=0}^L \mathbf{h}(n; l) s(L + n - l) - \mathbf{m} \right\|^2 + \text{PM}_{n-1}(s(L + n - 1), \dots, s(n)). \quad (\text{B.6})$$

Again, the M^{L+1} sequences are divided into M^L groups corresponding to the M^L possible state $(s(L + n - 1), s(L + n - 2), \dots, s(n))$ and the most probable sequence from each group is selected while the other $M - 1$ sequences are discarded. The surviving metrics are

$$\begin{aligned} & \text{PM}_n(s(L + n), \dots, s(n + 1)) \\ &= \min_{s(n)} \left\{ \left\| \mathbf{y}(L + n) - \sum_{l=0}^L \mathbf{h}(n; l) s(L + n - l) - \mathbf{m} \right\|^2 \right. \\ & \left. + \text{PM}_{n-1}(s(L + n - 1), \dots, s(n)) \right\}. \end{aligned} \quad (\text{B.7})$$

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

In a multiple-user (of K users) context, the noisy channel input-output is given by

$$\mathbf{y}(n) = \sum_{k=1}^K \sum_{l=0}^L \mathbf{h}_k(n; l) s_k(n - l) + \mathbf{v}(n), \quad (\text{B.8})$$

where $\{s_k(n)\}$ denotes the k -th user's information sequence and $\{\mathbf{h}_k(n; l)\}$ denotes the k -th user's channel impulse response. The state at time n is now represented by the L most

recent input symbols of all the K users, i.e.,

$$\text{state}(n) = (s_1(n), \dots, s_1(n-L+1), \dots, s_2(n-L+1), \dots, s_K(n-L+1)) \quad (\text{B.9})$$

where $s_k(n) = 0$ for $n < 0$. The channel is now described by an M^{KL} -state trellis. To adapt the Viterbi algorithm in for multiple users, we simply replace the state (B.3) with (B.9), use $M' = M^K$ instead of M in all the three steps, and substitute the sum $\sum_{k=1}^K \sum_{l=0}^L \mathbf{h}_k(n;l) s_k(n-l)$ for $\sum_{l=0}^L \mathbf{h}(n;l) s(n-l)$ in (B.4)–(B.7).

B.2 Kalman Filtering

Viterbi detector (or MLSD) is the optimal receiver that provides the minimum BER. Its computational complexity, however, depends on the number of states. By (B.9), the Viterbi detector has M^{KL} states, given the M -ary input symbols, K users, and an $(L+1)$ -tap MIMO channel. The computational complexity grows exponentially with the length of channel, the number of users, and the constellation of transmitted signal. Viterbi detector may be extremely expensive to implement [64].

Kalman filter, based on MMSE criterion, offers us an alternative symbol detection technique, with much lower computational complexity than that of the optimal detector—even though at the expense of a slight sacrifice of error performance.

For the SIMO FIR linear channel with N receivers described by (B.1) and (B.2), we define a $(d+1)$ -column vector (the delay $d \geq L$) as the state vector of the Kalman filter:

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

We also define the state transition matrix

$$\mathbf{\Phi} := \begin{bmatrix} \mathbf{0}_{1 \times d} & 0 \\ \mathbf{I}_d & \mathbf{0}_{d \times 1} \end{bmatrix},$$

the control-input matrix

$$\mathbf{\Gamma} := \begin{bmatrix} 1 & \mathbf{0}_{1 \times d} \end{bmatrix}^T,$$

the observation matrix

$$\mathbf{H}(n) := \begin{bmatrix} \mathbf{h}(n;0) & \mathbf{h}(n;1) & \cdots & \mathbf{h}(n;L) & \mathbf{0}_{N \times (d-L)} \end{bmatrix},$$

and the input

$$w(n) := s(n+1).$$

Then we have the time-invariant state equation:

$$\mathbf{S}(n+1) = \mathbf{\Phi}\mathbf{S}(n) + \mathbf{\Gamma}w(n). \quad (\text{B.10})$$

By (B.1) and (B.2), the time-varying observation equation is given by

$$\mathbf{y}(n) = \mathbf{H}(n)\mathbf{S}(n) + \mathbf{v}(n). \quad (\text{B.11})$$

We assume the AWGN $\mathbf{v}(n)$ is zero-mean. The prior statistics of the above parameters are given by

$$E\{w(n)\} = 0, \quad E\{\mathbf{v}(n)\} = \mathbf{0}, \quad E\{\mathbf{S}(0)\} = \mu_s(0),$$

$$E \{w(m) w^*(n)\} = V_w(n) \delta(m-n), \quad E \{\mathbf{v}(m) \mathbf{v}^H(n)\} = \mathbf{V}_v(n) \delta(m-n),$$

$$E \{w(m) \mathbf{v}(n)^H\} = \mathbf{0}, \quad E \{w(m) \mathbf{S}^H(n)\} = \mathbf{0}, \quad E \{\mathbf{v}(m) \mathbf{S}^H(n)\} = \mathbf{0},$$

$$E \{[\mathbf{S}(0) - E\{\mathbf{S}(0)\}][\mathbf{S}(0) - E\{\mathbf{S}(0)\}]^H\} = \mathbf{V}_s(0).$$

Given the state equation (B.10), the observation equation (B.11), and the above prior statistics, the algorithm of Kalman filter is as follows [70]:

Initialization: For the time $n = 0$, $\hat{\mathbf{S}}(1 | 0) = \mu_s(0)$ and $\mathbf{V}_{\hat{s}}(1 | 0) = \mathbf{V}_s(0)$.

Filtering: For $n = 1, 2, \dots$

$$\mathbf{V}_\eta(n) = \mathbf{H}(n) \mathbf{V}_{\hat{s}}(n | n-1) \mathbf{H}^H(n) + \mathbf{V}_v(n);$$

$$\mathbf{K}(n) = \mathbf{V}_{\hat{s}}(n | n-1) \mathbf{H}^H(n) \mathbf{V}_\eta^{-1}(n);$$

$$\eta(n) = \mathbf{y}(n) - \mathbf{H}(n) \hat{\mathbf{S}}(n | n-1);$$

$$\hat{\mathbf{S}}(n | n) = \hat{\mathbf{S}}(n | n-1) + \mathbf{K}(n) \eta(n);$$

$$\hat{\mathbf{S}}(n+1 | n) = \mathbf{\Phi} \hat{\mathbf{S}}(n | n);$$

$$\mathbf{V}_{\hat{s}}(n | n) = [\mathbf{I} - \mathbf{K}(n) \mathbf{H}(n)] \mathbf{V}_{\hat{s}}(n | n-1);$$

$$\mathbf{V}_{\hat{s}}(n+1 | n) = \mathbf{\Phi} \mathbf{V}_{\hat{s}}(n | n) \mathbf{\Phi}^H + \mathbf{\Gamma} V_w(n) \mathbf{\Gamma}^H.$$

Since $\hat{\mathbf{S}}(n | n) = \begin{bmatrix} \hat{s}(n | n) & \hat{s}(n-1 | n) & \cdots & \hat{s}(n-d | n) \end{bmatrix}^T$, we extract its last term $\hat{s}(n-d | n)$ as the desired equalized output. Then hard-quantize $\hat{s}(n-d | n)$ to acquire the detected symbol $\tilde{s}(n-d)$.

For a multiple-user (MIMO) channel of (B.8) with total K users, the state vector for the k -th user is

$$\mathbf{S}_k(n) = \begin{bmatrix} s_k(n) & s_k(n-1) & \cdots & s_k(n-d) \end{bmatrix}^T$$

where we also have $d \geq L$. Then the augmented $K(d+1)$ -state vector is given by

$$\mathbf{S}(n) = \begin{bmatrix} \mathbf{S}_1^T(n) & \mathbf{S}_2^T(n) & \cdots & \mathbf{S}_K^T(n) \end{bmatrix}^T.$$

In order to apply Kalman filter, we revise the state transition matrix as

$$\mathbf{\Phi} := \mathbf{I}_K \otimes \begin{bmatrix} \mathbf{0}_{1 \times d} & 0 \\ \mathbf{I}_d & \mathbf{0}_{d \times 1} \end{bmatrix},$$

the control-input matrix as

$$\mathbf{\Gamma} := \mathbf{I}_K \otimes \begin{bmatrix} 1 & \mathbf{0}_{1 \times d} \end{bmatrix}^T,$$

the observation matrix as

$$\mathbf{H}_k(n) := \begin{bmatrix} \mathbf{h}_k(n;0) & \mathbf{h}_k(n;1) & \cdots & \mathbf{h}_k(n;L) & \mathbf{0}_{N \times (d-l)} \end{bmatrix}$$

and

$$\mathbf{H}(n) := \begin{bmatrix} \mathbf{H}_1(n) & \mathbf{H}_2(n) & \cdots & \mathbf{H}_K(n) \end{bmatrix},$$

and the input as

$$\mathbf{w}(n) := \begin{bmatrix} s_1(n+1) & s_2(n+1) & \cdots & s_K(n+1) \end{bmatrix}^T.$$

Then the state equation (B.10) and the observation equation (B.11) still hold. Then apply Kalman filtering to obtain

$$\hat{\mathbf{S}}(n|n) = \left[\hat{\mathbf{S}}_1^T(n|n) \quad \hat{\mathbf{S}}_2^T(n|n) \quad \cdots \quad \hat{\mathbf{S}}_K^T(n|n) \right]^T,$$

where $\hat{\mathbf{S}}_k^T(n|n) = \left[\hat{s}_k(n|n) \quad \hat{s}_k(n-1|n) \quad \cdots \quad \hat{s}_k(n-d|n) \right]^T$ for $k = 1, 2, \dots, K$. Finally, we hard-quantize $\hat{s}_k(n-d|n)$ as the desired equalized output for the k -th user.

APPENDIX C

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
$ a $	magnitude of a
\mathbf{a}	lower case letters in bold face for column vectors
$\ \mathbf{a}\ $	Euclidean norm of \mathbf{a}
\mathbf{A}	upper case letters in bold face for matrices
\mathbf{A}^*	complex conjugate of \mathbf{A}
\mathbf{A}^\dagger	Moore-Penrose pseudo-inverse operation
\mathbf{A}^H	complex conjugate transpose of \mathbf{A}
\mathbf{A}^T	transpose of \mathbf{A}
$[\mathbf{A}]_{n,m}$	(n, m) -th entry of \mathbf{A}
\mathcal{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
$\text{cov}\{\cdot\}$	covariance operator

$\delta(\cdot)$	Kronecker delta function, defined as
	$\delta(n) = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0, t \in \mathbb{Z} \end{cases}$
$\text{diag}\{a_1, \dots, a_N\}$	$N \times N$ diagonal matrix with $[\text{diag}\{a_1, \dots, a_N\}]_{n,n} = a_n$
$E\{\cdot\}$	expectation operator
$E_{\mathcal{H}}\{\cdot\}$	expectation operator with respect to \mathcal{H}
\mathbf{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 $x \rightarrow a$ ($a \in \mathbb{R} \cup \pm\infty$), iff $ f(x) \leq M g(x) $ as $x \rightarrow a$ for some constant $M > 0$
\mathbb{R}	real field
$\text{tr}\{\mathbf{A}\}$	trace of a square matrix \mathbf{A}
\mathbb{Z}	integer field

APPENDIX D

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
CRLB	Cramér-Rao lower bound
CSI	channel state information
DC	direct current
DFT	discrete Fourier transform
DKL-BEM	discrete Karhuen-Loève basis expansion model
DML	deterministic maximum likelihood
DPS	discrete prolate spheroidal
DPS-BEM	discrete prolate spheroidal basis expansion model
FIR	finite impulse response
FM	frequency modulation
ISI	inter-symbol interference
LMS	least mean squares
LS	least squares

MAC	media access control
MANET	mobile <i>ad hoc</i> networks
MIMO	multiple-input multiple-output
ML	maximum likelihood
MLSE	maximum likelihood sequence estimation
MSE	mean square error
MMSE	minimum mean square error
MPR	multiple packet reception
m.s.	mean-square
MUI	multiple-user interference
NCMSE	normalized channel mean square error
OFDM	orthogonal frequency division multiplexing
OP-BEM	orthogonal polynomial basis expansion model
PDD	partially-data-dependent
pdf	probability density function
PN	pseudo-noise
QAM	quadrature amplitude modulation
SIMO	single-input multiple-output
SISO	single-input single-output
SNR	signal-to-noise ratio
TIR	training-to-information power ratio
TM	time-multiplexed