

It is dangerous to put limits on wireless.

– Guglielmo Marconi, 1932.

University of Alberta

**CHANNEL ESTIMATION AND TRAINING SEQUENCE DESIGN
FOR ONE-WAY AND TWO-WAY RELAY NETWORKS**

by

Gongpu Wang

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Communications

Department of Electrical and Computer Engineering

© Gongpu Wang
Fall 2011
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Dedicated to

my beloved parents, sister and wife.

Abstract

Wireless relay networking is a highly active research field. Several relay standards have been or are being specified for next-generation mobile broadband communication systems. Channel estimates are required by wireless nodes to perform essential tasks such as precoding, beamforming and data detection. Thus this thesis focuses on channel estimation for amplify-and-forward (AF) one-way relay networks (OWRN) and two-way relay networks (TWRN).

For orthogonal frequency-division multiplexing (OFDM) based TWRNs, joint carrier frequency offset (CFO) and channel estimation is investigated. Two new zero-padding (ZP) and cyclic-prefix (CP) transmission protocols are proposed. Both protocols enable an estimator based on the nulling-based least square (NLS) algorithm and perform identically when the block length is large. A detailed performance analysis is given by proving the unbiasedness of the estimator at high signal-to-noise ratio (SNR) and by deriving the closed-form expression of the mean-square error (MSE). Since the two protocols and corresponding NLS algorithm can only estimate the convoluted channel parameters, a superimposed training strategy is proposed to estimate all the individual channel parameters. Specifically, three different algorithms that require different lengths of trainings are designed for the initial parameter estimation and an iterative algorithm is developed to refine the initial estimation results.

For TWRNs operating over time-varying fading environments, channel estimation and training sequence design are investigated. A new complex exponential basis expansion model (CE-BEM) is proposed to represent the mobile-to-mobile time-varying channel. To estimate the parameters of this model, a novel pilot symbol-aided transmission scheme is developed such that a linear approach can estimate the convoluted channels. More essentially, two algorithms are designed to extract the BEM coefficients of the individual channels. The optimal training parameters are derived by minimizing the estimation MSE.

For OWRNs operating over doubly-selective channels, estimation algorithms and training sequence design are investigated. The CE-BEM is utilized to approximate the doubly-selective channel. Since direct estimation of the CE-BEM coefficients requires large pilot overhead, an efficient estimator is developed that targets only *useful* channel parameters that could guarantee effective data detection. The training sequence design that can minimize the estimation MSE is also proposed.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Chintha Tellambura for his encouragement, guidance and invaluable support throughout my graduate studies. All my achievements have been benefited from his knowledge, insight and inspiration. These years which I have been under his supervision will be remembered as one of my “life honor”.

I am also deeply indebted to Prof. Feifei Gao, who has been my research collaborator during the past three years, for his valuable guidance and suggestions during our collaborative research. I learned and gained enormously from his knowledge and experience. It would not have been possible to write this thesis without his help.

I would also like to thank the members of my defense committee, for taking their previous time to review my thesis and giving valuable suggestions for my research.

My deepest gratitude goes to my beloved parents, my sister and my wife for their great patience and faithful support and love throughout, for which mere words cannot express my gratitude.

Much respect to my labmates. Mr. Damith Senaratne, Mr. Gayan Amarasuriya, Mr. Saman Atapattu, Mr. Kasun Hemachandra and Mr. Madushanka Soysa are my lab mates coming from Sri-Lanka. They are modest, friendly, industrious, and having good publications, which makes them excellent companions for me and others. Studying and doing research together with them these days will be a good memory of my life.

Much respect to my classmates and friends. Ms. Min Liu, Ms. Hua Shao, Mr. Rongfei Fan, Mr. Zhongshan Zhang, Mr. Bo Hu and many others. During my graduate studies, I have received tremendous help from them. They make my days meaningful and happy. The years I spent in University of Alberta is an unforgettable period of learning and working experience.

Finally, I gratefully acknowledge the financial support from China Scholarship Council.

Table of Contents

1	Introduction	1
1.1	OWRN	2
1.2	TWRN	2
1.3	Wireless Channel	3
1.3.1	Channel Model	3
1.3.2	Channel Input and Output	5
1.4	Channel Estimation in Point-to-point Systems	6
1.4.1	Estimation of Frequency-selective Channels	6
1.4.2	Estimation of Time-varying Channels	7
1.4.3	Joint CFO and Channel Estimation	7
1.5	Channel Estimation for Relay Networks	8
1.6	Superimposed Training	9
1.7	Training Sequence Design	9
1.7.1	Optimal Training Sequence Design for Point-to-point Systems	10
1.7.2	Optimal Sequence Design in OWRN and TWRN	11
1.8	Motivation	11
1.9	Contributions and Structure of the Thesis	12
2	Joint CFO and Channel Estimation for TWRN	14
2.1	Introduction	14
2.2	System Model	15
2.3	ZP-based OFDM for TWRN	16
2.3.1	OFDM Terminals	16
2.3.2	Relay Processing	16
2.3.3	Signal Reformulation at Terminals	17
2.3.4	Joint CFO and Channel Estimation	18
2.4	CP-based OFDM for TWRN	19

2.4.1	OFDM modulation at Terminals	19
2.4.2	Relay Processing	19
2.4.3	Signal Reformulation at Terminals	20
2.4.4	Joint CFO and Channel Estimation	21
2.5	Comparison between ZP-based and CP-based OFDM	22
2.5.1	Similarities	22
2.5.2	Differences	22
2.6	Nulling-based Least Square Estimation	22
2.6.1	Joint CFO and Channel Estimation Algorithm	22
2.6.2	Performance Analysis of Estimation Mean-Square Error.	23
2.7	Simulation Results	24
2.8	Conclusions	27
3	Superimposed Training Aided Estimation for TWRN	30
3.1	System Model	30
3.2	CP-Based OFDM Protocol	31
3.2.1	OFDM Modulation at Terminals	31
3.2.2	Relay Processing	31
3.2.3	Signal Reformulation at Terminals	32
3.3	Joint Estimation Algorithms	33
3.3.1	Estimation for Sufficiently Large N	33
3.3.2	Estimation with Not-so-large N	35
3.3.3	Joint Estimation with Minimum Training Length: A Special Case	36
3.3.4	Iterative Algorithm to Improve the Performance	37
3.3.5	Comparison	37
3.4	Cramér-Rao Bound	38
3.5	Simulation Results	39
3.5.1	Sufficiently Large N	39
3.5.2	Not-so-large N	40
3.5.3	Minimum N : A Special Case	42
3.6	Conclusions	43
4	Time-Varying Channel Estimation for TWRN	46
4.1	Introduction	46
4.2	System Model	47
4.2.1	Time-varying Relay Channels	48
4.2.2	Transmission Strategy	49

4.2.3	On Channel Estimation	51
4.3	Channel Estimation and Training Sequence Design	53
4.3.1	Channel Estimation Algorithm	53
4.3.2	Optimal Training Design	54
4.3.3	Parameter Selection	57
4.3.4	Doppler Shift and Transmission Efficiency	57
4.4	Recovering the Original BEM Coefficients	58
4.4.1	Time-Domain Approach	58
4.4.2	Frequency-Domain Approach	59
4.5	Simulation Results	61
4.5.1	Channel Estimation and Training Design	61
4.5.2	Data Detection	64
4.6	Conclusions	64
5	Doubly-Selective Channel Estimation for OWRN	67
5.1	Introduction	67
5.2	System model	68
5.3	Doubly-Selective Channel in OWRN	69
5.4	Estimation, Detection and Training Sequence Design	70
5.4.1	Channel Partition	71
5.4.2	Estimation Algorithm	73
5.4.3	Data Detection	76
5.4.4	Training Sequence Design	79
5.4.5	Block Parameters	80
5.5	Simulation Results	81
5.6	Conclusions	83
6	Conclusions and Future Work	85
6.1	Conclusions	85
6.2	Future Work	86
	Appendix A Proof of Lemma 2.2	97
	Appendix B Proof of Theorem 2.1	100
	Appendix C Proof of Theorem 2.2	102
	Appendix D Gradient Decent Method with Complex Variables	104

Appendix E	Proof of Theorem 4.1	105
Appendix F	Proof of Lemma 5.3	107

List of Tables

1.1	The types of wireless channels.	5
2.1	Comparison between ZP- and CP-OFDM modulated TWRN.	22
3.1	Comparison between adapted CP-OFDM and superimposed pilot aided CP-OFDM. . .	37

List of Figures

1.1	A one-way relay network	2
1.2	A two-way relay network	3
1.3	Wireless channel model	4
1.4	Traditional pilots versus superimposed pilots.	8
1.5	Data-dependent superimposed training.	10
1.6	Design criteria for optimal training sequences	11
2.1	System configuration for a three-node TWRN.	15
2.2	CFO estimation MSEs version SNR for both ZP- and CP-based OFDM modulation; $N = 16$	25
2.3	Channel estimation MSEs version SNR for both ZP- and CP-based OFDM modulation; $N = 16$	26
2.4	CFO estimation MSEs version block length for both ZP- and CP-based OFDM modulation; SNR= 10 dB.	27
2.5	Channel estimation MSEs version block length for both ZP- and CP-based OFDM modulation; SNR= 10 dB.	28
2.6	Performance SER versus SNR for both ZP- and CP-based OFDM modulation; $N = 16$	29
3.1	System configuration for two-way relay network.	31
3.2	CFO estimation MSE versus SNR: $N = 24$	40
3.3	Channel estimation MSE versus SNR: $N = 24$	41
3.4	CFO estimation MSE versus SNR: $N = 14$	42
3.5	Channel estimation MSE versus SNR: $N = 14$	43
3.6	CFO estimation MSE versus SNR: $N = 9$	44
3.7	Channel estimation MSE versus SNR: $N = 9$	45
4.1	A two-way relay network over time-varying flat-fading channels.	47
4.2	Approximating the mobile-to-mobile channels with CE-BEM.	49

4.3	Proposed transmission strategy for two-way relay network with time-varying channel.	50
4.4	Sub-block based transmission strategy.	51
4.5	Channel estimation MSEs versus SNR for \mathbf{x} .	61
4.6	Channel estimation MSEs versus SNR for λ : time-domain approach.	62
4.7	Channel estimation MSEs versus SNR for λ : frequency-domain approach.	63
4.8	Comparison between the time- and the frequency-domain approaches.	64
4.9	Channel estimation MSEs versus SNR for μ .	65
4.10	BER versus SNR for realistic mobile-to-mobile channels with different values of Q .	66
5.1	System model for AF relay network over doubly-selective channel.	68
5.2	Structure of one transmission block.	68
5.3	Partition of the matrix \mathbf{H} into \mathbf{H}_s , \mathbf{H}_b , and $\mathbf{H}_{\bar{b}}$ that are shown in dashed line on the right side of the figure.	71
5.4	Partition of the matrix \mathbf{H}_k .	72
5.5	Partition of the matrix \mathbf{G}	73
5.6	Partition of the matrix \mathbf{G}_k	74
5.7	Channel MSE versus the SNR.	81
5.8	BER versus the SNR.	82
5.9	BER versus the SNR: real channel.	83
5.10	Transmission efficiency versus SNR.	84

List of Abbreviations

Acronyms	Definition
AF	amplify and forward
AWGN	additive white Gaussian noise
BER	bit error rate
BEM	basis expansion model
BPSK	binary phase shift keying
CE-BEM	complex exponential - basis expansion model
CFO	carrier frequency offset
CP	cyclic prefix
CSI	channel state information
CRB	Cramer-Rao bound
CSCG	circularly symmetric complex Gaussian
DPS-BEM	discrete prolate spheroidal - basis expansion model
DF	decode and forward
LMMSE	linear minimum mean square error
MIMO	multiple-input multiple-output
ML	maximum-likelihood
MMSE	minimum mean square error
MSE	mean square error
OWRN	one-way relay network
OFDM	orthogonal frequency division multiplexing
PDF	probability density function
PSAM	pilot symbol assisted modulation
QPSK	Quadrature phase shift keying
SNR	signal-to-noise ratio
STC	space-time code
TDD	time division duplex
TWRN	two-way relay network

ZP

zero padding

List of Symbols

j	$\sqrt{-1}$
$(\cdot)^*$	Complex conjugation
$(\cdot)^T$	Transpose
$(\cdot)^H$	Conjugate transpose
$(\cdot)^\dagger$	Moore-Penrose pseudo-inverse
$\delta(\cdot)$	Dirac delta function
$\arg(\cdot)$	Argument
\otimes	Linear convolution
\circledast	Circular convolution
\mathbf{A}	The matrix \mathbf{A}
\mathbf{A}^{-1}	The inverse matrix of \mathbf{A}
\mathbf{a}	The vector \mathbf{a}
$a(n)$	The n th entry of the vector \mathbf{a}
$\Re(\cdot)$	Real part of the argument
$\Im(\cdot)$	Imaginary part of the argument
$E(\cdot)$	Expectation of the argument
$\ (\cdot)\ ^2$	2-norm of the argument
$\text{trace}(\mathbf{M})$	The trace of the matrix \mathbf{M}
$\mathcal{CN}(\mu, \sigma^2)$	A complex Gaussian random variable (RV) with mean μ and variance σ^2
\mathbf{I}_N	The $N \times N$ identity matrix
$\mathbf{0}_{M \times N}$	$M \times N$ all-zero matrix
$\text{diag}\{\mathbf{a}\}$	The diagonal matrix formed by vector \mathbf{a}
$\lceil a \rceil$	Integer ceiling of a
$\lfloor a \rfloor$	Integer floor of a

Chapter 1

Introduction

Wireless relay networks have been a highly active research field ever since the pioneer work [1]–[3]. Relays can improve throughput, enhance reliability [1], and increase coverage [2]. Several relay standards have been or are being specified for the next-generation mobile broadband communication systems [4] such as Long Term Evolution-Advanced (LTE-A) [71] and IEEE 802.16j [72].

High data rate for users is one of the main goal of LTE-A and IEEE 802.16j. Technologies such as orthogonal frequency-division multiplexing (OFDM) and multiple-input and multiple-output (MIMO)¹ can improve throughput but may fail at the cell edge where signal levels are lower and interference levels are typically higher. The use of relays in LTE-A can increase the data rates at the cell edge. The cell-edge relays will receive and decode the data and retransmit to the user equipment (UE)/mobile station (MS) or the evolved Node B (eNB)/base station (BS). In this way the signal quality is enhanced, and thus data transmission for remote UE or MS is enabled and coverage extended. The gain in SNR can be from 2 to 30 dB [4]. Such a relay node is defined as Type-I (or non-transparency) relay station (RS) in LTE-A. Another type of relay node defined in LTE-A and IEEE 802.16j is Type-II (or transparency) RS which is located with the coverage of an eNB/BS and has a direct link with UE/MS. Type-II RS can increase overall system capacity through transmission to the local UEs.

Currently, two-hop relay transmission has been developed in LTE-A, and IEEE 802.16j supports relay-based multihop communications in a cell. As well, two relay selection schemes are suggested in LTE-A to select and pair nearby RSs and UEs [71]. The first scheme is centralized and depends on an eNB to serve as a control node to collect required channel and location information from all RSs and UEs. The second scheme is distributed since each RS can select an appropriate UE unit in its neighborhood.

In the rest of this chapter, relay networks are briefly introduced. Wireless channel models and

¹MIMO means the use of multiple antennas at both the transmitter and receiver.

channel estimation techniques are reviewed. Channel estimation and training sequence design for relay networks are discussed. Finally, thesis contribution and organization are highlighted.

1.1 OWRN

A typical relay network consists of a source node, one relay node (or several relay nodes) and a destination node (Fig. 1.1). The transmission involves two phases. In the first phase, the source broadcasts signals to the relay node and possibly to the destination. In the second phase, the relay resends a version of the received signal in the first phase to the destination. The relay can be amplify-and-forward (AF) or decode-and-forward (DF) [2]. In the AF scheme the relay simply amplifies the received signal and then forwards to the destination. In the DF scheme the relay will decode the received signal and regenerate and retransmit to the destination. Finally, the destination combines the relayed signals and possibly the direct signal from the source to recover the data.

In the early relay networks [1]–[3], the data flow is considered unidirectional from the source to the relay and then to the destination (Fig. 1.1). Such a network is referred to as the one-way relay network (OWRN) . The other alternative is the two-way relay network (TWRN) .

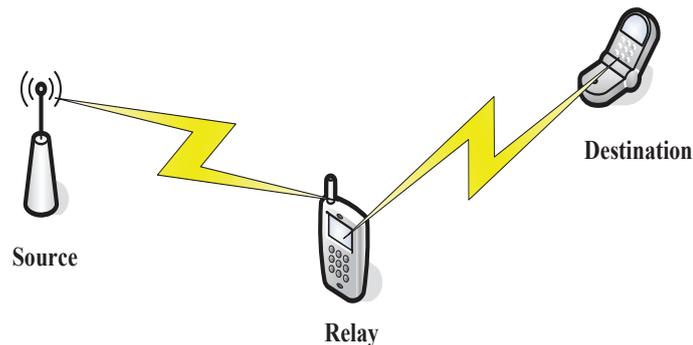


Fig. 1.1. A one-way relay network

1.2 TWRN

In a TWRN (Fig. 1.2), two terminals send information simultaneously to the relay, and the relay forwards to both terminals after a “network coding”-like process [5]. By removing the self-signal component, each terminal obtains the other terminal’s information. This two-way transmission was firstly exploited by Shannon in his early work [6] and now has drawn much attention [11], [12], [14] due to its improved spectral efficiency over OWRN. The overall communication rate between two

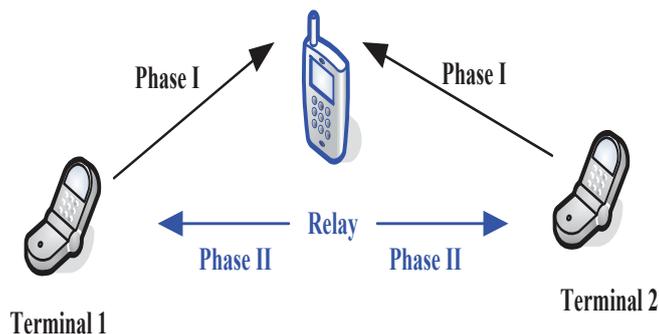


Fig. 1.2. A two-way relay network

source terminals in a TWRN is approximately twice that achieved in a OWRN [7], making a TWRN particularly attractive to bidirectional systems.

The capacity analysis and the achievable rate region for AF and DF based TWRN were explored in [16], [17]. In [18], the optimal relay mapping function that minimizes the bit-error rate (BER) was proposed. In [11], distributed space-time codes (STC)² were designed for both AF and DF TWRN. Moreover, the optimal beamforming at the multi-antenna relay that maximizes the capacity of AF-based TWRN was developed in [12].

In the next section, fundamentals of wireless channels are briefly described.

1.3 Wireless Channel

1.3.1 Channel Model

The wireless channel is described by the response $h(t, \tau)$ at time t to an impulse transmitted at time $t - \tau$. The channel consists of several independent paths. For this multipath model, the general expression can be written as [63]

$$h(\tau, t) = \sum_i a_i(t) \delta(\tau - \tau_i(t)), \quad (1.1)$$

where $a_i(t)$ is the attenuation and $\tau_i(t)$ is the delay from the transmitter to the receiver on the i -th path. An example of a wireless channel with three paths is shown in Fig. 1.3.

The general expression (1.1) is also known as a doubly-selective channel since there are several paths and the attenuations and delays are functions of time. The following two special cases for

²STC is a method to improve the transmission reliability by sending multiple, redundant copies of a data stream on multiple transmit antennas.

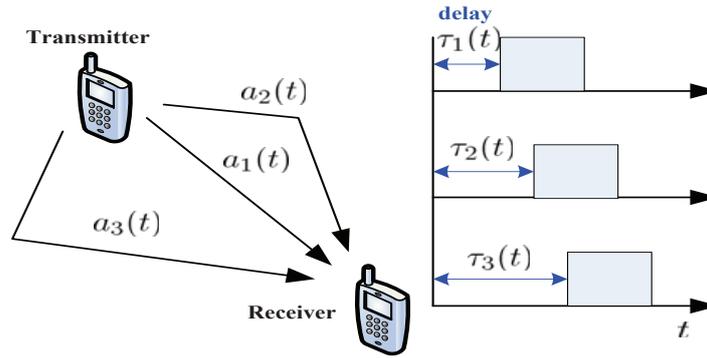


Fig. 1.3. Wireless channel model

$h(t, \tau)$ are widely used:

- time-invariant frequency-selective channel: This channel occurs when the transmitter, receiver and the environment are all stationary so that the attenuations $a_i(t)$ and propagation delays $\tau_i(t)$ do not depend on time t . However, the delays are significantly large compared to the symbol period.
- time-varying (or time-selective) flat-fading channel: The delays $\tau_i(t)$ in this case are all approximately constant and small compared to the symbol period. This channel occurs when the transmitter or the receiver is mobile and when the symbol period of the transmitted signal significantly exceeds any of all the delays.

Since the symbol period T_s decreases when the data rate increases, the channel can be flat-fading or frequency-selective depending on the data rate. Moreover, the delay spread is another relevant parameter. Delay spread T_d is defined as the difference in propagation delay between the longest and shortest path,

$$T_d = \max_{i,j} |\tau_i(t) - \tau_j(t)|. \quad (1.2)$$

When T_s is much larger than T_d , the channel is **flat-fading**. Otherwise, the channel is **frequency-selective**. For example, the typical delay spread in a wireless channel in an urban area is $5\mu\text{s}$ when the distance between transmitter and receiver is 1 km [64]. When the data rate is 1 kbps, the symbol period is 1 ms, and the channel is flat-fading since the delay is negligible compared to the symbol period. If the data rate increases to 1 Mbps, the symbol period T_s is $1\mu\text{s}$. Then the channel becomes frequency-selective due to the non-negligible delays.

Furthermore, the mobility of transmitter or receiver will induce a shift in radio frequency, which

is referred to as the Doppler shift D_s . Coherence time T_c , a parameter related to the Doppler shift, is defined as

$$T_c = \frac{1}{4D_s}. \quad (1.3)$$

If the coherence time T_c is comparable to the symbol period, the channel is **time-varying**. On the other hand, in **time-invariant** channels, the coherence time T_c is much larger than the symbol period (i.e., the channel remains constant). For example, if Doppler shift $D_s = 50$ Hz, and the transmission data rate is 1 Mbps, then the coherence time $T_c = 2.5$ ms is much larger than one symbol duration $1\mu s$. In this case the channel is time-invariant.

The types of wireless channels are depicted in Table 1.1.

TABLE 1.1
THE TYPES OF WIRELESS CHANNELS.

Types of Channel	Characteristic
Time-varying	$T_c \ll T_s$
Time-invariant	$T_c \gg T_s$
Flat-fading	$T_d \ll T_s$
Frequency-selective	$T_d \gg T_s$

1.3.2 Channel Input and Output

In terms of the wireless channel $h(t, \tau)$, the relationship between input $s(t)$ and output $y(t)$ is given by

$$y(t) = \int_{-\infty}^{+\infty} h(t, \tau)s(t - \tau)d\tau + w(t), \quad (1.4)$$

where $w(t)$ is an additive white Gaussian complex noise signal. The receiver is required to recover data signal $s(t)$ from received signal $y(t)$; this process is called data detection.

For data detection, the receiver requires the knowledge of $h(t, \tau)$, which is referred to as channel state information (CSI). To help the receiver estimate CSI, special predefined symbols may be transmitted in addition to data symbols. These symbols are called pilot symbols or training symbols. Pilot symbols are utilized by the channel estimator at the receiver to obtain CSI.

In practice, channel estimation and data detection are done by using the discrete-time baseband signals. Define the samples $y(nT_s) = y(n)$ for $n = 0, 1, \dots, N - 1$. The discrete-time baseband model equivalent to (1.4) can then be obtained as

$$y(n) = \sum_{l=0}^L h(n, l)s(n - l) + w(n), \quad (1.5)$$

where $h(n, l)$ is the sampling version of $h(t, \tau)$, i.e., $h(n, l) = h(nT_s, lT_s)$, and $s(n - l)$ is the sampling version of $s(t)$, i.e., $s(n - l) = s((n - l)T_s)$, and $L + 1$ is the number of multipaths and $w(n)$ is complex white Gaussian noise with mean zero and variance σ_w^2 .

1.4 Channel Estimation in Point-to-point Systems

1.4.1 Estimation of Frequency-selective Channels

For a frequency-selective time-invariant channel, since $h(n, l)$ does not change with time index n , i.e., $h(n, l) = h(l)$. Therefore the model (1.5) can be simplified as

$$y(n) = \sum_{l=0}^L h(l)s(n-l) + w(n). \quad (1.6)$$

Define $\mathbf{y} = [y(0), y(1), \dots, y(N-1)]^T$, $\mathbf{w} = [w(0), w(1), \dots, w(N-1)]^T$ and $\mathbf{h} = [h(0), h(1), \dots, h(L)]^T$ where N is the block length. We can write (1.6) in the following vector form

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{w}, \quad (1.7)$$

where \mathbf{S} is a $N \times (L + 1)$ circulant matrix with the first column $\mathbf{s} = [s(0), s(1), \dots, s(N - 1)]^T$. Note that sequence \mathbf{s} is the training sequence and depends on the choice of pilots and their values.

Two linear estimators are often utilized to obtain the estimate of \mathbf{h} from the received signal \mathbf{y} . The first one is least square (LS). It treats \mathbf{h} as deterministic constant and minimizes the mean square error. The LS estimate is [66]

$$\hat{\mathbf{h}} = (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{y}. \quad (1.8)$$

The second one is the linear minimum mean square error (LMMSE) estimator. It treats \mathbf{h} as a random vector and minimizes the mean square error. The LMMSE estimate is [66]

$$\hat{\mathbf{h}} = E(\mathbf{h}) + \mathbf{R}_{hy} \mathbf{R}_{yy}^{-1} (\mathbf{y} - E(\mathbf{y})), \quad (1.9)$$

where \mathbf{R}_{hy} denotes the covariance matrix of \mathbf{h} and \mathbf{y} , i.e., $\mathbf{R}_{hy} = E(\mathbf{h}\mathbf{y}^H)$ and \mathbf{R}_{yy} denotes the auto-covariance matrix of \mathbf{y} , i.e., $\mathbf{R}_{yy} = E(\mathbf{y}\mathbf{y}^H)$.

The LS estimator is simpler compared with the LMMSE estimator. But it outperforms the LS estimator in the low SNR region because it exploits the statistics of \mathbf{h} . However, at high SNR, the LS estimator achieves almost the same performance as the LMMSE estimator because the contribution of the statistics of \mathbf{h} to the estimation process is negligible at high SNR [66].

1.4.2 Estimation of Time-varying Channels

For a time-varying flat-fading channel, the model (1.5) can be simplified as

$$y(n) = h(n)s(n) + w(n). \quad (1.10)$$

Only CSI at a limited set of time instances can be estimated because the data symbols will be transmitted at other time instances. To solve this problem, time-varying channels are represented by using the Gauss-Markov model [33], which tracks channel variation through symbol-by-symbol updating, and by using the basis expansion model (BEM) [36], which decomposes the channel into the superposition of the time-varying basis functions weighted by time-invariant coefficients.

Using the BEM, we can approximate $h(n)$ as

$$h(n) = \sum_{q=0}^Q h_q u_q(n), \quad (1.11)$$

where $h_q, q = 0, 1, \dots, Q$, are the time-invariant coefficients and $u_q(n), q = 0, 1, \dots, Q$, are the basis function, and $Q + 1$ is the number of basis functions. Candidate basis functions include complex exponential (Fourier) functions [42] [36], polynomials [38], wavelet [40] and discrete prolate spheroidal (DPS) sequences [39]. It is pointed out that DPS outperform complex exponential basis in approximation [39]. The BEM can also represent doubly-selective channels [42].

1.4.3 Joint CFO and Channel Estimation

Suppose the carrier frequency of transmitter is f_1 and that of the receiver is f_2 . Carrier frequency offset (CFO) is the difference of carrier frequencies, i.e., $v = f_2 - f_1$. The CFO v may arise due to two reasons: first, a carrier frequency mismatch exists between the transmitter and receiver oscillators; second, a Doppler shift can arise due to the relative motion between the transmitter and the receiver.

In some cases CFO and CSI can be estimated jointly, which is often referred to as joint CFO and channel estimation.

With the CFO, the received signal (1.7) can be rewritten as

$$\mathbf{y} = \mathbf{\Gamma}(v)\mathbf{S}\mathbf{h} + \mathbf{w}, \quad (1.12)$$

where $\mathbf{\Gamma}(v) = \text{diag}\{1, e^{j2\pi v}, \dots, e^{j2\pi v(N-1)}\}$.

To perform joint estimation, the likelihood function of the parameters (\mathbf{h}, v) is obtained as

$$\lambda(\mathbf{y}; \mathbf{h}, v) = \frac{1}{(\pi\sigma_w^2)^N} \exp \left\{ -\frac{1}{\sigma_w^2} [\mathbf{y} - \mathbf{\Gamma}(v)\mathbf{S}\mathbf{h}]^H [\mathbf{y} - \mathbf{\Gamma}(v)\mathbf{S}\mathbf{h}] \right\}. \quad (1.13)$$

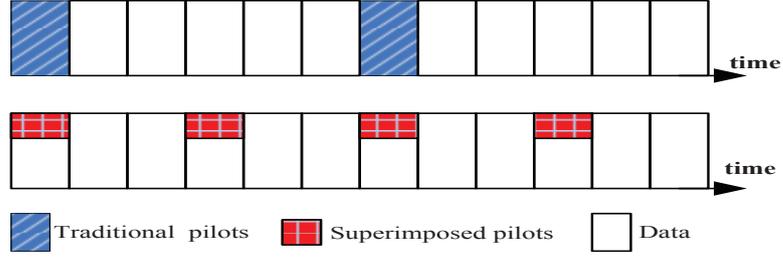


Fig. 1.4. Traditional pilots versus superimposed pilots.

Through maximizing the likelihood function (1.13), the optimal joint estimate of v and \mathbf{h} can be found. The details can be found in [58]. The maximum-likelihood (ML) joint estimates of CFO v and channel \mathbf{h} can be obtained as

$$\hat{v} = \arg \max_v \{\mathbf{y}^H \mathbf{\Gamma}(v) \mathbf{S} (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{\Gamma}^H(v) \mathbf{y}\}, \quad (1.14)$$

$$\hat{\mathbf{h}} = (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{\Gamma}^H(\hat{v}) \mathbf{y}. \quad (1.15)$$

Computational complexity analysis of this joint estimation problem is given in [58].

1.5 Channel Estimation for Relay Networks

Although the traditional estimation methods can be readily applied to DF based relay networks, there are significant differences between channel estimation for AF-based relay networks and that for traditional point-to-point networks. The main difference is that overall channels from the source node through relay node to the destination node, instead of individual channels, are to be estimated.

Flat-fading and frequency-selective fading channel estimation for OWRN was studied in [81], [20], [86], and [87]. Flat-fading and frequency-selective channel estimation of TWRN was respectively studied in [13], [14], [15], which also show that AF TWRN systems require completely different estimation techniques than those for conventional point-to-point systems. All of the above works assume time-invariant channels.

In [13], the flat-fading TWRN channels are considered as deterministic or stochastic. For deterministic channels, the nonlinear maximum-likelihood (ML)-based estimator is proposed, and for stochastic channels, the linear maximum SNR estimator is designed. In [14], the frequency-selective TWRN channels are considered and orthogonal frequency division multiplexing (OFDM) is chosen for transmission. In [15], channel estimation is investigated for MIMO TWRNs.

1.6 Superimposed Training

In general, training schemes can be classified into two categories: traditional and superimposed. Traditional pilots are time-multiplexed with data. However, superimposed pilots are added to data periodically and the sum is transmitted to the receiver. The difference between traditional pilots and superimposed pilots is shown in Fig. 1.4.

The performance of conventional pilots and superimposed pilots is studied in [29], [32]. When the total transmitted energy is constrained, superimposed pilots offer increased performance over conventional pilots for fast fading channels [32]. However, when the amplitude for each symbol is constrained, conventional training performs better in the high signal-to-noise ratio (SNR) region while at low SNR, the superimposed scheme performs better [29].

Recently, superimposed pilots have been further exploited for channel estimation [24], [26]–[28], [30], [31]. In these references two methods are applied to obtain channel estimates. The first method uses precoding [30] to separate pilots and data into orthogonal spaces. The second method utilizes first-order statistics [23] to cancel the zero-mean data and noise from the received signals.

It is often assumed that first-order statistics can only apply for zero-mean signals, and that the channel estimation performance of superimposed pilots is poorer than that of pilot symbol assisted modulation (PSAM). However, a frequency domain method is suggested for superimposed pilots [25] to obtain CSI from signals with non-zero mean noise. Moreover, the data-dependent superimposed training (DDST) proposed in [24] can shorten the gap and almost achieve the same performance as PSAM. In DDST, data sequence \mathbf{s} is distorted by another sequence \mathbf{e} so that at some frequencies, the discrete Fourier transform (DFT) transform value of the $\mathbf{s} + \mathbf{e}$ is zero, where the periodical superimposed training \mathbf{p} can be used for estimation (Fig. 1.5).

Superimposed pilots are proposed to estimate doubly-selective channels [26], [28], [31]. This technique has also been extended to relay networks [21] [89]. DDST has also been explored in AF relay networks [75] to obtain individual parameters.

1.7 Training Sequence Design

Cavers, first suggested the widely used term PSAM and proposed an analytical approach to the pilot design [45]. Optimal training sequence design has since then received much attention.

Design criteria for optimal training sequences can be classified into two categories (Fig. 1.6): information theoretic measures or detection-estimation measures. The first includes Shannon capacity and capacity upper or lower bounds [46]. The second includes estimation mean square error (MSE) [57], Cramér-Rao bound (CRB)³ [56], [58], [59], or BER [48].

³CRB is a lower bound on the variance of the estimated deterministic parameter.

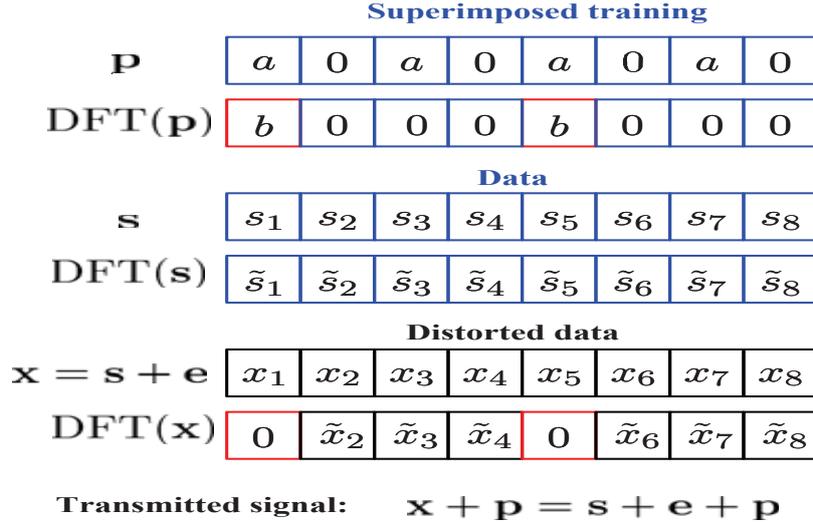


Fig. 1.5. Data-dependent superimposed training.

To reach the optimality under any of these criteria, one can design the amount and types of pilots, the power allocation between pilots and data, and the placement of the pilots for different channels and wireless systems. Many such designs have been suggested and investigated.

1.7.1 Optimal Training Sequence Design for Point-to-point Systems

Optimization of training data, power and interval length are proposed for MIMO channels to maximize a lower bound on the information theoretic capacity in [46]. This reference finds the capacity lower bound by transforming the channel estimation error into a noise, and shows that best training signal matrix must have orthonormal columns. Following this guideline, training optimization is extended to time-varying MIMO channels in [51]. Also, optimal designs of training sequences are suggested in [49] by minimizing the mean square error (MSE), and in [50] by using the cutoff rate criterion. It is pointed out in [49] that single pilot regular periodic placement can minimize the MMSE among all periodic trainings.

For frequency-selective channels, optimal designs for pilots are suggested in [52]–[54]. To minimize CRB, pilots and data symbols with higher power should be in the middle of the transmission packet [52]; to minimize the MSE, equispaced and equipowered pilot symbols are optimal [53]; to maximize channel capacity of OFDM systems, training sequences should be placed periodically in frequency [54].

For doubly-selective channels, optimal training design is discussed in [43] through maximizing a tight lower bound of channel capacity which is proved to be equivalent with minimizing MSE of channel estimation.

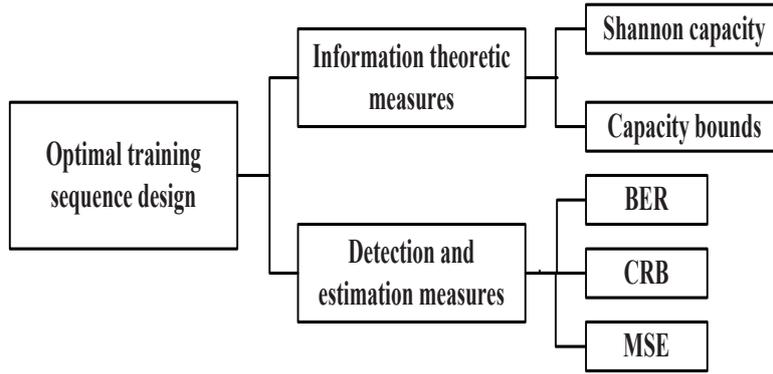


Fig. 1.6. Design criteria for optimal training sequences

For joint CFO and channel estimation, the best training sequence that minimizes the CRB is a pseudo-random white sequence [55], [58] when the CFO is perfectly estimated. By minimizing the worst-case asymptotic CRB, i.e., the large sample CRB associated with the worst case channel realization, white training sequence is proved to be optimal in [59]. One training design is suggested in [56] that can render exact CRB of CFO independent of channels. Aiming at minimizing the MSE, the authors in [57] design the optimal sequence in the case of correlated channel taps or Ricean channel responses.

1.7.2 Optimal Sequence Design in OWRN and TWRN

Training sequence designs for OWRN was studied in [20], [95]. The optimal training for OWRN can be achieved from an arbitrary sequence and a set of well designed precoding matrices for all relay nodes [20]. Training sequence designs for TWRN have been studied in [13], [14] and [15]. Training sequences should be orthogonal in order to minimize CRB or maximize SNR for flat fading channels [13]. For TWRN over frequency-selective channels, certain rules for optimal training sequences that can minimize the MSE are suggested in [14]. For MIMO flat-fading channels, optimal training sequence are proposed to minimize the estimation MSE in [15].

1.8 Motivation

Most previous TWRN works [7]–[12] assume the availability of perfect CSI at the relay and/or the source terminals. But CSI estimation in TWRN is a complicated issue. For instance, although conventional estimation methods are effective for DF-based TWRN, they fail for AF-based TWRN. The latter is more challenging because channel estimation is required not only for data detection but also for the self-data cancellation at the two terminals.

To the best of our knowledge, currently only a few papers [13]–[15] have treated the channel estimation and training sequence design issues for TWRN systems. Reference [14] developed a channel estimator and optimal training sequence for the OFDM-based TWRN. However, OFDM systems are highly sensitive to the CFO, which will destroy subcarrier orthogonality and thus require estimation and compensation. Any residual CFO will result in intercarrier interference and detection errors. The development of a joint CFO and channel estimator for OFDM-based TWRN is one of the main problems that we will address. This problem is highly challenging due to the CFO mismatch between two source terminals as well as that between the source terminals and the relay.

Existing works [13]–[15], [68], [89], [90] on TWRN channel estimation are limited to time-invariant channels. However, when the relay and the two sources are mobile, the relative motion between any two nodes doubles the Doppler spread [81]. The Doppler spread will result in time-varying channels. To our best knowledge, time-varying channel estimator for TWRN has not been developed in the previous literature. Therefore, the development of an estimator is a main problem that we will address. Moreover, the design of optimal training sequence that minimizes the MSE remains unknown. Our work thus will include the development of channel estimator, selection of optimization criterion, and pilot designs such as pilot types, power allocation among pilots or between pilots and data, and placement or intervals of pilots. Furthermore, if channels are doubly selective (both time-varying and frequency-selective), channel estimation and optimal training sequence design become even greater challenge.

Moreover, to the best of our knowledge, channel estimation techniques for OWRN or TWRN systems over doubly-selective channels have not yet been developed. Estimation of such channels is significantly more complicated than that of the time-varying channel, especially for OWRN and TWRN. The OWRN case is treated in this thesis, but the TWRN case is left as an open problem.

1.9 Contributions and Structure of the Thesis

This thesis proposes channel estimation algorithms and training sequence design for TWRNs. A doubly-selective channel estimator and training sequence design are also developed for OWRNs. The main contributions are detailed below.

Chapter 2: Joint CFO and Channel Estimation for TWRN

In this chapter, a joint CFO and channel estimator is developed for an OFDM-based TWRN. Two new zero-padding (ZP) and cyclic-prefix (CP) transmission protocols, which maintain the carrier orthogonality and ensure channel estimation, are proposed. We show that both protocols lead to the same estimation problem and the nulling-based least square (NLS) algorithm developed as the estimator. Next, a detailed performance analysis is given: the NLS estimator is shown to be unbiased and the closed-form of MSE is derived. The simulation results show the two protocols perform well

and achieve almost the same MSE when the block length is large.

Chapter 3: Superimposed Training Aided Estimation for TWRN

Although Chapter 2 develops a joint CFO and channel estimator for AF TWRNs, it only estimates the convoluted channel parameters and the mixed CFO values. In this chapter, we propose a superimposed training strategy so that all the individual frequency and channel parameters can be obtained. Specifically, the relay adds a superimposed time-domain training signal to its received signal and then transmits the combined signal to both terminals. Three different algorithms for the initial parameter estimation are developed to adapt to different available training length. Then, an iterative process is conducted to further improve the estimation accuracy. The CRB of the proposed strategy is also derived for comparison. Our simulation results show that the iteration converges in a few steps and the resultant estimation mean square errors (MSE) approaches the CRB. For the special case when the CFO between two terminals is small, the estimation MSE is very close to the CRB in the high SNR region, which indicates optimality.

Chapter 4: Time-Varying Channel Estimation for TWRN

Chapter 2 and 3 consider the time-invariant TWRN channel. In this chapter, research is extended for AF-based TWRNs in time-varying channels. For this purpose, a compact representation of time-varying channels is required. The CE-BEM is thus adapted to represent the mobile-to-mobile time-varying channels. A novel pilot symbol-aided transmission scheme is developed to facilitate the estimation and two algorithms are designed to extract the BEM coefficients of the individual channels. The optimal training sequence is derived by minimizing the channel estimation MSE.

Chapter 5: Doubly-Selective Channel Estimation for OWRN

When a OWRN operates over doubly-selective channels, the required channel estimation had been an unexplored area.

In this chapter, we choose complex exponential - basis expansion model (CE-BEM) to represent the doubly-selective OWRN channel and the estimation problem thus aims to recover the CE-BEM coefficients. However, we find that direct estimation of these coefficients requires significant pilot overhead. To solve this problem, we suggest to estimate *useful* channel parameters that could guarantee effective data detection. The optimal training sequence is also developed based on minimizing the estimation MSE.

Chapter 2

Joint CFO and Channel Estimation for TWRN

This chapter considers joint CFO and channel estimation for OFDM-based AF TWRNs. To facilitate the estimation task, new ZP and CP transmission protocols, which maintain the carrier orthogonality and ensure low estimation complexity, are proposed. Both protocols lead to the same estimator which can be implemented by the NLS algorithm and perform identically when the block length is large. We give a detailed performance analysis by proving the unbiasedness of the estimator at high SNR and by deriving the closed-form expression of the MSE.

2.1 Introduction

Most previous works [7]–[12] assume the availability of perfect CSI at the relay and/or the source terminals. But CSI estimation in TWRN is a complicated issue. Although conventional estimation methods are effective for DF-based TWRN, they do not work for AF based TWRN. For example, channel estimation for TWRN was studied in [13], [14] for frequency-flat and frequency-selective environments, respectively, and these studies show that AF TWRN systems require completely different estimation techniques, unlike those of conventional point-to-point systems.

Moreover, frequency mismatches that cause CFO require compensation before data detection [58]. The problem of CFO estimation in TWRN is even more difficult because one must consider the mismatch between two source terminals as well as that between the source terminals and relays. The estimation of CFO and channels with low complexity is, therefore, a challenging task. Moreover, maintaining carrier orthogonality in order to facilitate data detection is critically important.

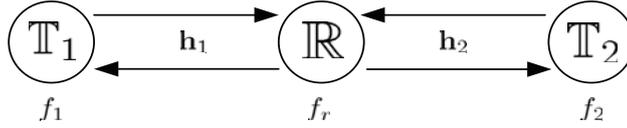


Fig. 2.1. System configuration for a three-node TWRN.

2.2 System Model

Consider a TWRN with two source nodes \mathbb{T}_1 and \mathbb{T}_2 , and one relay node \mathbb{R} (Fig. 2.1). Each node has only one half-duplex antenna. The baseband channel between \mathbb{T}_i and \mathbb{R} is denoted by $\mathbf{h}_i = [h_{i,0}, h_{i,1}, \dots, h_{i,L}]^T$, where L is the order of the corresponding channel.¹ The elements in \mathbf{h}_i are assumed to be independently zero-mean circularly symmetric complex Gaussian (CSCG) random variables. The variance of the l th element in \mathbf{h}_i is denoted by $\sigma_{i,l}^2$. Time-division-duplexing (TDD) is used, following [10], [12]. TDD leads to reciprocal channels, i.e., the channel from \mathbb{R} to \mathbb{T}_i is \mathbf{h}_i as well. When frequency division duplex (FDD) is used, the \mathbb{R} and \mathbb{T}_i adopt different frequency bands for transmission, which will result in non-reciprocal channels. That is, the channel from \mathbb{T}_i to \mathbb{R} is not the same with the channel from \mathbb{R} to \mathbb{T}_i . However, our proposed strategies may be extended to the non-reciprocal channels and more general channels. The average transmission powers of \mathbb{T}_1 , \mathbb{T}_2 , and \mathbb{R} are denoted as P_1 , P_2 , and P_r , respectively. We further denote the oscillator frequencies as f_1 , f_2 , and f_r , respectively. Moreover, perfect time synchronization is assumed [58], [59].

In [14], OFDM was adapted to TWRN under perfect time and frequency synchronization. These perfect conditions ensure that circular convolution between the two frequency-selective channels in time is equivalently converted to multiplication between two flat fading channels over different subcarriers. However, with CFOs, inter-carrier interference (ICI) will be introduced in the frequency domain at different nodes. The joint estimation problem then becomes more complicated if the same OFDM based TWRN strategy is applied [14]. Even with perfect CSI, data detection is not simple.

In this chapter, we introduce two new transmission protocols for OFDM TWRN systems. Although they incur redundancy, these protocols facilitate both joint estimation and data detection.

¹We assume the same channel length for \mathbf{h}_i for the sake of notational simplicity. Our discussion can be straightforwardly extended to the more general case.

2.3 ZP-based OFDM for TWRN

2.3.1 OFDM Terminals

Without loss of generality, we omit the block index and denote one OFDM block from \mathbb{T}_i as $\tilde{\mathbf{s}}_i = [\tilde{s}_{i,0}, \dots, \tilde{s}_{i,N-1}]^T$, where N is the block length. The corresponding time-domain signal block is obtained from the normalized inverse discrete Fourier transformation (IDFT) as

$$\mathbf{s}_i = \mathbf{F}^H \tilde{\mathbf{s}}_i = [s_{i,0}, s_{i,1}, \dots, s_{i,N-1}]^T, \quad (2.1)$$

where \mathbf{F} is the normalized DFT matrix with the (p, q) -th entry given by $\frac{1}{\sqrt{N}} e^{-j2\pi(p-1)(q-1)/N}$. To avoid the inter-block interference (IBI) in the first transmission phase, L zeros are padded at the end of \mathbf{s}_i . As in the traditional ZP-OFDM system [76], the power of $\tilde{s}_{i,n}$, $n = 0, \dots, N-1$ is $\frac{N+L}{N} P_i$, so the average power constraint P_i is kept at \mathbb{T}_i .

In Phase I, \mathbb{T}_1 and \mathbb{T}_2 up-convert the baseband signals by the carriers $e^{j2\pi f_i t}$ and send them to \mathbb{R} simultaneously. Note that the oscillator may have an initial phase, but it is omitted for brevity since the constant phase can be absorbed into the channel effects.

2.3.2 Relay Processing

The relay \mathbb{R} will down-convert the passband signal by $e^{-j2\pi f_r t}$ and the received baseband signal can be represented as

$$\mathbf{r}_{zp} = \sum_{i=1}^2 \mathbf{\Gamma}^{(N+L)}[f_i - f_r] \mathbf{H}_{zp}^{(N)}[\mathbf{h}_i] \mathbf{s}_i + \mathbf{n}_r, \quad (2.2)$$

where

$$\mathbf{\Gamma}^{(K)}[f] = \text{diag}\{1, e^{j2\pi f T_s}, \dots, e^{j2\pi f (K-1) T_s}\} \quad (2.3)$$

with T_s representing the sampling period, and

$$\mathbf{H}_{zp}^{(K)}[\mathbf{x}] = \underbrace{\begin{bmatrix} x_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ x_L & \ddots & x_0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & x_L \end{bmatrix}}_{K \text{ columns}} \quad (2.4)$$

for any vector $\mathbf{x} = [x_0, x_1, \dots, x_L]^T$. Moreover, \mathbf{n}_r is the $(N + L) \times 1$ noise vector, each entry with variance σ_n^2 .

Then, \mathbb{R} adds L zeros to the end of \mathbf{r} and scales it by the factor of

$$\alpha_{zp} = \sqrt{\frac{(N + 2L)P_r}{\mathbb{E}\{\|\mathbf{r}_{zp}\|^2\}}} = \sqrt{\frac{N + 2L}{N + L} \frac{P_r}{\sum_{i=1}^2 \sum_{l=0}^L \sigma_{i,l}^2 P_i + \sigma_n^2}} \quad (2.5)$$

to keep the average relay power constraint. The resultant $\alpha_{zp}\mathbf{r}_{zp}$ will be up-converted to passband by $e^{j2\pi f_r t}$.

2.3.3 Signal Reformulation at Terminals

Due to symmetry, we only study \mathbb{T}_1 during the second phase. After the down-conversion of the passband signal by $e^{-j2\pi f_1 t}$, the $(N + 2L) \times 1$ signal vector is expressed as

$$\begin{aligned} \mathbf{y}_{zp} &= \alpha_{zp} \mathbf{\Gamma}^{(N+2L)} [f_r - f_1] \mathbf{H}_{zp}^{(N+L)} [\mathbf{h}_1] \mathbf{r}_{zp} \\ &= \alpha_{zp} \mathbf{\Gamma}^{(N+2L)} [f_r - f_1] \mathbf{H}_{zp}^{(N+L)} [\mathbf{h}_1] \left(\sum_{i=1}^2 \mathbf{\Gamma}^{(N+L)} [f_i - f_r] \mathbf{H}_{zp}^{(N)} [\mathbf{h}_i] \mathbf{s}_i \right) \\ &\quad + \underbrace{\alpha_{zp} \mathbf{\Gamma}^{(N+2L)} [f_r - f_1] \mathbf{H}_{zp}^{(N+L)} [\mathbf{h}_1] \mathbf{n}_r + \mathbf{n}_1}_{\mathbf{n}_e}, \end{aligned} \quad (2.6)$$

where \mathbf{n}_1 is the $(N + 2L) \times 1$ noise vector at \mathbb{T}_1 with variance σ_n^2 , and \mathbf{n}_e defines the overall noise component. The covariance of \mathbf{n}_e is computed as

$$\mathbf{R}_{zp} = \sigma_n^2 (\alpha_{zp}^2 \mathbf{\Gamma}^{(N+2L)} [f_r - f_1] \mathbf{H}_{zp}^{(N+L)} [\mathbf{h}_1] (\mathbf{H}_{zp}^{(N+L)} [\mathbf{h}_1])^H (\mathbf{\Gamma}^{(N+2L)} [f_r - f_1])^H + \mathbf{I}). \quad (2.7)$$

In most practical communications,² there is $N \gg L$ and \mathbf{R}_{zp} can be approximated by its expectation over \mathbf{h}_1 :

$$\mathbf{R}_{zp} \approx \underbrace{\sigma_n^2 \left(\alpha_{zp}^2 \sum_{l=0}^L \sigma_{h_{1,l}}^2 + 1 \right)}_{\sigma_{n_e}^2} \mathbf{I}, \quad (2.8)$$

where $\sigma_{n_e}^2$ denotes the equivalent noise variance.

Lemma 2.1. The following two equalities hold for any $\mathbf{\Gamma}^{(\cdot)}[f]$ in (2.3) and $\mathbf{H}_{zp}^{(\cdot)}[\mathbf{x}]$ in (2.4), where (\cdot) represents the appropriate dimensions:

$$\mathbf{H}_{zp}^{(K)}[\mathbf{x}] \mathbf{\Gamma}^{(K)}[f] = \mathbf{\Gamma}^{(K+P)}[f] \mathbf{H}_{zp}^{(K)} \left[\mathbf{\Gamma}^{(K)}[-f] \mathbf{x} \right], \quad (2.9)$$

²In IEEE 802.11a standards [73], $N \geq 4L$ is adopted.

and conversely

$$\mathbf{\Gamma}^{(K+P)}[f]\mathbf{H}_{zp}^{(K)}[\mathbf{x}] = \mathbf{H}_{zp}^{(K)}\left[\mathbf{\Gamma}^{(P+1)}[f]\mathbf{x}\right]\mathbf{\Gamma}^{(K)}[f]. \quad (2.10)$$

Proof. Proved from the straightforward computation. \square

Lemma 2.1 says that $\mathbf{\Gamma}^{(\cdot)}[f]$ can be switched from the right (left) side of $\mathbf{H}_{zp}^{(\cdot)}[\mathbf{h}_i]$ to the left (right) side by changing the dimension of $\mathbf{\Gamma}^{(\cdot)}[f]$ and rotating \mathbf{h}_i . From Lemma 1, \mathbf{y}_{zp} can be rewritten as

$$\begin{aligned} \mathbf{y}_{zp} = & \alpha_{zp}\mathbf{H}_{zp}^{(N+L)}[\mathbf{\Gamma}^{(L+1)}[f_r - f_1]\mathbf{h}_1]\mathbf{H}_{zp}^{(N)}[\mathbf{h}_1]\mathbf{s}_1 \\ & + \alpha_{zp}\mathbf{\Gamma}^{(N+2L)}[f_2 - f_1]\mathbf{H}_{zp}^{(N+L)}[\mathbf{\Gamma}^{(L+1)}[f_r - f_2]\mathbf{h}_1]\mathbf{H}_{zp}^{(N)}[\mathbf{h}_2]\mathbf{s}_2 + \mathbf{n}_e. \end{aligned} \quad (2.11)$$

We further note that

$$\mathbf{H}_{zp}^{(N+L)}[\mathbf{x}_1]\mathbf{H}_{zp}^{(N)}[\mathbf{x}_2] = \mathbf{H}_{zp}^{(N)}[\mathbf{x}_1 \otimes \mathbf{x}_2] \quad (2.12)$$

where \otimes denotes the linear convolution between the two vectors. Hence \mathbf{y}_{zp} can be written as

$$\begin{aligned} \mathbf{y}_{zp} = & \alpha_{zp}\mathbf{H}_{zp}^{(N)}\left[\underbrace{(\mathbf{\Gamma}^{(L+1)}[f_r - f_1]\mathbf{h}_1) \otimes \mathbf{h}_1}_{\mathbf{a}_{zp}}\right]\mathbf{s}_1 \\ & + \alpha_{zp}\mathbf{\Gamma}^{(N+2L)}\left[\underbrace{f_2 - f_1}_v\right]\mathbf{H}_{zp}^{(N)}\left[\underbrace{(\mathbf{\Gamma}^{(L+1)}[f_r - f_2]\mathbf{h}_1) \otimes \mathbf{h}_2}_{\mathbf{b}_{zp}}\right]\mathbf{s}_2 + \mathbf{n}_e, \end{aligned} \quad (2.13)$$

where \mathbf{a}_{zp} , \mathbf{b}_{zp} are the $(2L + 1) \times 1$ equivalent channel vectors and v is the equivalent CFO.

Note that the same method in (2.13) can be used to find equivalent channel vectors for AF-based TWRNs with non-reciprocal channels between \mathbb{T}_i and \mathbb{R} . This extension is omitted for the sake of brevity.

2.3.4 Joint CFO and Channel Estimation

The above discussion shows that the task is to estimate \mathbf{a}_{zp} , \mathbf{b}_{zp} , and v . Assuming that \mathbf{s}_1 and \mathbf{s}_2 are the training blocks, we can rewrite (2.13) as

$$\mathbf{y}_{zp} = \mathbf{S}_1^{(N+2L)}\mathbf{a}_{zp} + \mathbf{\Gamma}^{(N+2L)}[v]\mathbf{S}_2^{(N+2L)}\mathbf{b}_{zp} + \mathbf{n}_e, \quad (2.14)$$

where $\mathbf{S}_i^{(N+2L)}$ is the $(N + 2L) \times (2L + 1)$ circulant matrix with the first column $[\alpha_{zp}\mathbf{s}_i^T, \mathbf{0}_{1 \times 2L}^T]^T$. Obviously, (2.14) is different from the conventional work [58] in that only a part of the signal component is accompanied with the CFO matrix. Moreover, $N \geq 2L + 3$ is required to estimate all

the unknown parameters. The detailed estimation algorithms will be presented in the next section.

2.4 CP-based OFDM for TWRN

Interestingly, CP-based OFDM must be modified in an *anti-symmetric* way.

2.4.1 OFDM modulation at Terminals

Each terminal \mathbb{T}_i first obtains the time domain OFDM signal \mathbf{s}_i from its frequency domain information block $\tilde{\mathbf{s}}_i$ via the normalized IDFT approach. We propose that \mathbb{T}_i adds a CP of length $2L$ in the front of \mathbf{s}_i . Define

$$\mathbf{T}_{cp}^{(P)} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_P \\ -\frac{\mathbf{I}_N}{\mathbf{I}_N} \end{bmatrix}, \quad (2.15)$$

for any $P \leq N$. Then the baseband signal sent out from \mathbb{T}_i is mathematically expressed as $\mathbf{T}_{cp}^{(2L)} \mathbf{s}_i$, which is up-converted to the passband signal by $e^{j2\pi f_i t}$.

2.4.2 Relay Processing

Relay \mathbb{R} first down-converts the passband signal by $e^{-j2\pi f_r t}$, which gives a signal block of length $N + 2L$. However, \mathbb{R} removes only the first L symbols in the CP. Define the convolution matrix

$$\mathbf{H}_{cv}^{(K)}[\mathbf{x}] = \left. \begin{bmatrix} x_P & \dots & x_0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & x_P & \dots & x_0 \end{bmatrix} \right\} K \text{ rows}, \quad (2.16)$$

for $\mathbf{x} = [x_0, x_1, \dots, x_P]^T$. The remaining signal of length $N + L$ is

$$\begin{aligned} \mathbf{r}_{cp} &= \sum_{i=1}^2 e^{j2\pi(f_i - f_r)LT_s} \mathbf{\Gamma}^{(N+L)}[f_i - f_r] \mathbf{H}_{cv}^{(N+L)}[\mathbf{h}_i] \mathbf{T}_{cp}^{(2L)} \mathbf{s}_i + \mathbf{n}_r \\ &= \sum_{i=1}^2 e^{j2\pi(f_i - f_r)LT_s} \mathbf{\Gamma}^{(N+L)}[f_i - f_r] \mathbf{T}_{cp}^{(L)} \mathbf{H}_{cp}^{(N)}[\mathbf{h}_i] \mathbf{s}_i + \mathbf{n}_r. \end{aligned}$$

where the properties

$$\mathbf{H}_{cv}^{(N+L)}[\mathbf{h}_i] \mathbf{T}_{cp}^{(2L)} = \mathbf{T}_{cp}^{(L)} \mathbf{H}_{cv}^{(N)}[\mathbf{h}_i] \mathbf{T}_{cp}^{(L)} \quad (2.17)$$

$$\mathbf{H}_{cv}^{(N)}[\mathbf{h}_i] \mathbf{T}_{cp}^{(L)} = \mathbf{H}_{cp}^{(N)}[\mathbf{h}_i] \quad (2.18)$$

are used and $\mathbf{H}_{cp}^{(N)}[\mathbf{h}_i]$ is the $N \times N$ circulant matrix with the first column $[\mathbf{h}_i^T, \mathbf{0}_{1 \times (N-L-1)}^T]^T$.

To keep the power constraint, \mathbb{R} scales \mathbf{r} by a factor

$$\alpha_{cp} = \sqrt{\frac{P_r}{\mathbb{E}\{\|\mathbf{r}_{cp}\|^2\}}} = \sqrt{\frac{P_r}{\sum_{i=1}^2 \sum_{l=0}^L \sigma_{i,l}^2 P_i + \sigma_n^2}}, \quad (2.19)$$

where the property that

$$\left(\mathbf{T}_{cp}^{(L)}\right)^H \mathbf{T}_{cp}^{(L)} = \begin{bmatrix} \mathbf{I}_{N-L} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{I}_L \end{bmatrix} \quad (2.20)$$

is used during the computation. Relay \mathbb{R} then up-converts the baseband signal $\alpha_{cp}\mathbf{r}_{cp}$ by $e^{j2\pi f_r t}$ and broadcasts it to both terminals.

2.4.3 Signal Reformulation at Terminals

After down-converting the passband signal by $e^{-j2\pi f_1 t}$, \mathbb{T}_1 obtains the baseband block of length $N + L$ and removes the first L elements; and the remaining signal is

$$\begin{aligned} \mathbf{y}_{cp} &= \alpha_{cp} e^{j2\pi(f_r - f_1)LT_s} \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cv}^{(N)}[\mathbf{h}_1] \mathbf{r}_{cp} \\ &= \alpha_{cp} \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cv}^{(N)}[\mathbf{h}_1] \mathbf{\Gamma}^{(N+L)}[f_1 - f_r] \mathbf{T}_{cp}^{(L)} \mathbf{H}_{cp}^{(N)}[\mathbf{h}_1] \mathbf{s}_1 \\ &\quad + \alpha_{cp} e^{j2\pi(f_2 - f_1)LT_s} \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cv}^{(N)}[\mathbf{h}_1] \mathbf{\Gamma}^{(N+L)}[f_2 - f_r] \mathbf{T}_{cp}^{(L)} \mathbf{H}_{cp}^{(N)}[\mathbf{h}_2] \mathbf{s}_2 \\ &\quad + \underbrace{\alpha_{cp} e^{j2\pi(f_r - f_1)LT_s} \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cv}^{(N)}[\mathbf{h}_1] \mathbf{n}_r + \mathbf{n}_1}_{\mathbf{n}_e} \end{aligned} \quad (2.21)$$

where the notation \mathbf{n}_e is slightly abused to denote the equivalent noise here, which has the covariance

$$\mathbf{R}_{cp} = \sigma_n^2 (\alpha_{cp}^2 \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cv}^{(N)}[\mathbf{h}_1] (\mathbf{H}_{cv}^{(N)}[\mathbf{h}_1])^H (\mathbf{\Gamma}^{(N)}[f_r - f_1])^H + \mathbf{I}). \quad (2.22)$$

When $N \gg L$, the following approximation can be made:

$$\mathbf{R}_n \approx \underbrace{\sigma_n^2 \left(\alpha_{cp}^2 \sum_{l=0}^L \sigma_{h_{1,l}}^2 + 1 \right)}_{\sigma_{n_e}^2} \mathbf{I}, \quad (2.23)$$

where, with slight abuse of notations, $\sigma_{n_e}^2$ is again used to denote the equivalent noise variance.

We observe that $(\mathbf{H}_{cv}^{(N)}[\mathbf{h}_i])^T$ has the same structure as $\mathbf{H}_{zp}^{(N)}[\mathbf{h}_i]$ but with the elements of \mathbf{h}_i ordered in the reverse way. Therefore, the principle of Lemma 1 can be extended to derive the following equality:

$$\mathbf{H}_{cv}^{(N)}[\mathbf{h}_i] \mathbf{\Gamma}^{(N+L)}[f] = \mathbf{\Gamma}^{(N)}[f] \mathbf{H}_{cv}^{(N)}[\mathbf{\Omega}^{(L+1)}[f] \mathbf{h}_i], \quad (2.24)$$

where

$$\mathbf{\Omega}^{(K)}[f] = \text{diag}\{e^{j2\pi f(K-1)T_s}, \dots, e^{j2\pi fT_s}, 1\}. \quad (2.25)$$

Hence, \mathbf{y}_{cp} is rewritten as

$$\begin{aligned} \mathbf{y}_{cp} &= \alpha_{cp} \mathbf{H}_{cv}^{(N)} [\mathbf{\Omega}^{(L+1)} [f_1 - f_r] \mathbf{h}_1] \mathbf{T}_{cp}^{(L)} \mathbf{H}_{cp}^{(N)} [\mathbf{h}_1] \mathbf{s}_1 \\ &\quad + \alpha_{cp} e^{j2\pi v L T_s} \mathbf{\Gamma}^{(N)} [v] \mathbf{H}_{cv}^{(N)} [\mathbf{\Omega}^{(L+1)} [f_2 - f_r] \mathbf{h}_1] \mathbf{T}_{cp}^{(L)} \mathbf{H}_{cp}^{(N)} [\mathbf{h}_2] \mathbf{s}_2 + \mathbf{n}_e \\ &= \alpha_{cp} \mathbf{H}_{cp}^{(N)} [\mathbf{\Omega}^{(L+1)} [f_1 - f_r] \mathbf{h}_1] \mathbf{H}_{cp}^{(N)} [\mathbf{h}_1] \mathbf{s}_1 \\ &\quad + \alpha_{cp} e^{j2\pi v L T_s} \mathbf{\Gamma}^{(N)} [v] \mathbf{H}_{cp}^{(N)} [\mathbf{\Omega}^{(L+1)} [f_2 - f_r] \mathbf{h}_1] \mathbf{H}_{cp}^{(N)} [\mathbf{h}_2] \mathbf{s}_2 + \mathbf{n}_e, \end{aligned} \quad (2.26)$$

where the property (2.18) is used when deriving the second equality.

We further note that

$$\mathbf{H}_{cp}^{(N)} [\mathbf{x}_1] \mathbf{H}_{cp}^{(N)} [\mathbf{x}_2] = \mathbf{H}_{cp}^{(N)} [\mathbf{x}_1 \otimes \mathbf{x}_2], \quad (2.27)$$

where \otimes denotes the N -point circular convolution between two vectors. Since $N \geq 2L + 1$ is assumed, the N -point circular convolution between \mathbf{h}_i 's coincides with the linear convolution of their non-zero part plus $N - (2L + 1)$ zeros at the end. Then \mathbf{y}_{cp} can be expressed as

$$\begin{aligned} \mathbf{y}_{cp} &= \alpha_{cp} \mathbf{H}_{cp}^{(N)} [\underbrace{[\mathbf{\Omega}^{(L+1)} [f_1 - f_r] \mathbf{h}_1] \otimes \mathbf{h}_1}_{\mathbf{a}_{cp}}] \mathbf{s}_1 \\ &\quad + \alpha_{cp} e^{j2\pi v L T_s} \mathbf{\Gamma}^{(N)} [v] \mathbf{H}_{cp}^{(N)} [\underbrace{[\mathbf{\Omega}^{(L+1)} [f_2 - f_r] \mathbf{h}_1] \otimes \mathbf{h}_2}_{\mathbf{b}_{cp}}] \mathbf{s}_2 + \mathbf{n}_e, \end{aligned} \quad (2.28)$$

where \mathbf{a}_{cp} , \mathbf{b}_{cp} are the $(2L + 1) \times 1$ equivalent channel vectors and v is the equivalent CFO.

2.4.4 Joint CFO and Channel Estimation

From previous discussion in Section 2.3.3 we know that the task is to estimate \mathbf{a}_{cp} , \mathbf{b}_{cp} , and v . Assuming that \mathbf{s}_1 and \mathbf{s}_2 are the training blocks, we can rewrite \mathbf{y}_{cp} as

$$\mathbf{y}_{cp} = \mathbf{S}_1^{(N)} \mathbf{a}_{cp} + \mathbf{\Gamma}^{(N)} [v] \mathbf{S}_2^{(N)} \mathbf{b}_{cp} + \mathbf{n}_e, \quad (2.29)$$

where $\mathbf{S}_i^{(N)}$ is the $N \times (2L + 1)$ circulant matrix with the first column \mathbf{s}_i . Moreover, $N \geq 4L + 3$ is required to estimate all the unknown parameters.

2.5 Comparison between ZP-based and CP-based OFDM

2.5.1 Similarities

Both protocols require the same amount of the redundancy,³ that is $3L$ extra samples over two phases. The transmission efficiency can be defined as $N/(2N + 3L)$. Moreover, both the final models (2.14) and (2.29) have the same structure. Thus, the same type of estimator can work for both systems.

2.5.2 Differences

The OFDM block length for the ZP approach changes from N to $N + L$ and then to $N + 2L$, while the final received signal blocks at the terminals are of the length $N + 2L$. However, the CP approach is *anti-symmetric*; that is, the block length changes from $N + 2L$ to $N + L$ and then to N , while the final received signal blocks at the terminals are of the length N . For both approaches, the number of unknowns to be estimated is $2(2L + 1) + 1$. Since ZP-based OFDM has a longer observation interval, the related estimation is expected to be more accurate.

Table 2.1 presents a detailed comparison between the two proposed OFDM protocols.

TABLE 2.1
COMPARISON BETWEEN ZP- AND CP-OFDM MODULATED TWRN.

	ZP	CP
Transmitter activity	add L zeros suffix	add $2L$ cyclic prefix
Relay activity	add L zeros suffix	remove L prefix
Destination activity	None	remove L prefix
Received signal length	$N + 2L$	N
Spectrum efficiency	$N/(2N + 3L)$	$N/(2N + 3L)$
Required Pilot Length	$N \geq 2L + 3$	$N \geq 4L + 3$

2.6 Nulling-based Least Square Estimation

Since (2.14) and (2.29) have the same structure, we will omit the superscript and the subscript, aiming to provide a unified estimation algorithm for both OFDM protocols.

2.6.1 Joint CFO and Channel Estimation Algorithm

Since \mathbf{S}_1 is a tall matrix, a matrix \mathbf{J} can be found such that $\mathbf{J}^H \mathbf{S}_1 = \mathbf{0}$. We propose to select \mathbf{J} with the orthogonal property that $\mathbf{J}^H \mathbf{J} = \mathbf{I}$, since it has the best condition number. A simple choice of \mathbf{J} is the basis of the orthogonal complement space of \mathbf{S}_1 .

³Note the CP of length L is always necessary and is not considered as a type of redundancy in our work.

Left-multiplying \mathbf{y} by \mathbf{J}^H gives

$$\mathbf{J}^H \mathbf{y} = \mathbf{0} + \underbrace{\mathbf{J}^H \mathbf{\Gamma} \mathbf{S}_2}_{\mathbf{G}} \mathbf{b} + \underbrace{\mathbf{J}^H \mathbf{n}_e}_{\mathbf{n}}, \quad (2.30)$$

where \mathbf{G} and \mathbf{n} are defined as the corresponding items. Due to the orthogonal property of \mathbf{J} , we know the statistics of \mathbf{n} remain the same as that of \mathbf{n}_e if the latter is approximated as white Gaussian noise.

The nulling-based least square (LS) estimate of \mathbf{b} can be immediately found from (2.30) as:

$$\hat{\mathbf{b}} = (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H \mathbf{J}^H \mathbf{y}. \quad (2.31)$$

The CFO is estimated from

$$\hat{v} = \arg \max_v \mathbf{y}^H \mathbf{J} \mathbf{G} (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H \mathbf{J}^H \mathbf{y}, \quad (2.32)$$

Finally, the LS estimation of channel \mathbf{a} is obtained from

$$\hat{\mathbf{a}} = (\mathbf{S}_1^H \mathbf{S}_1)^{-1} \mathbf{S}_1^H (\mathbf{y} - \hat{\mathbf{\Gamma}} \mathbf{S}_2 \hat{\mathbf{b}}), \quad (2.33)$$

where

$$\hat{\mathbf{\Gamma}} = \hat{\mathbf{\Gamma}}[\hat{v}] = \text{diag}\{1, e^{j2\pi\hat{v}T_s}, \dots, e^{j2\pi(M-1)\hat{v}T_s}\}. \quad (2.34)$$

2.6.2 Performance Analysis of Estimation Mean-Square Error.

Due to the nulling process, the estimation model (2.14) and (2.29) is more complicated than the typical model such as in [58]. Hence, the result in [58] cannot be directly extended to our scenario. We next will prove that LS estimator is unbiased at high SNR and derive the closed-form MSE expressions based on the perturbation theory. Note that asymptotic analysis is commonly used in the joint estimation problems [60].

For notation simplicity, we denote

$$\mathbf{y}_n = \mathbf{J}^H \mathbf{y}, \quad \mathbf{P}_G = \mathbf{G} (\underbrace{\mathbf{G}^H \mathbf{G}}_{\mathbf{\Phi}})^{-1} \mathbf{G}^H, \quad (2.35)$$

where $\mathbf{\Phi}$ represents the corresponding item. Let v_0 and \hat{v}_0 be the true and the estimated CFO, respectively. The LS estimator (2.32) can be written as

$$\hat{v}_0 = \arg \max_v g(v) = \arg \max_v \mathbf{y}_n^H \mathbf{P}_G \mathbf{y}_n. \quad (2.36)$$

Lemma 2.2. At high SNR, the CFO estimation error from (2.36) can be approximated by

$$\Delta v \triangleq \hat{v}_0 - v_0 \approx -\frac{\dot{g}(v_0)}{\mathbb{E}\{\dot{g}(v_0)\}}. \quad (2.37)$$

Proof. See Appendix A. □

Theorem 2.1. The MSE of the CFO estimation is

$$\mathbb{E}\{\Delta v^2\} = \frac{\sigma_{ne}^2}{2\mathbf{b}^H \dot{\mathbf{G}}^H [\mathbf{I} - \mathbf{G}(\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H] \dot{\mathbf{G}} \mathbf{b}}. \quad (2.38)$$

Proof. See Appendix B. □

Theorem 2.2. The channel estimation $\hat{\mathbf{b}}$ is unbiased at high SNR, and its MSE is

$$\text{MSE}\{\mathbf{b}\} = (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H \dot{\mathbf{G}} \mathbf{b} \mathbf{b}^H \dot{\mathbf{G}}^H \mathbf{G} (\mathbf{G}^H \mathbf{G})^{-1} \mathbb{E}\{\Delta v^2\} + \sigma_{ne}^2 (\mathbf{G}^H \mathbf{G})^{-1}. \quad (2.39)$$

Proof. See Appendix C. □

2.7 Simulation Results

In this section, we report a numerical study of the performance of our proposed joint CFO and channel estimation strategy. A four-tap channel model with the exponential delay profile $\sigma_{il}^2 = e^{-l/10}$, $i = 1, 2, l = 0, 1, 2, 3$ is assumed. The variance of the noise is taken as $\sigma_n^2 = 1$. The SNR is defined as the ratio of symbol power to the noise power, i.e., E_s/N_0 . The normalized frequencies f_1 , f_r , and f_2 are set as 0.95, 1 and 1.05, respectively. Thus the CFO is as large as 0.1. The MSE is chosen as the figure of merit, defined by

$$\begin{aligned} \text{MSE}(v) &= \frac{1}{10000} \sum_{i=1}^{10000} (\hat{v}_i - v)^2, \\ \text{MSE}(\mathbf{x}) &= \frac{1}{10000} \sum_{i=1}^{10000} \frac{1}{7} (\hat{\mathbf{x}}_i - \mathbf{x})^2, \end{aligned}$$

where \mathbf{x} represents a composite channel, such as \mathbf{a}_{zp} or \mathbf{b}_{zp} in (2.13), and \mathbf{a}_{cp} or \mathbf{b}_{cp} in (2.28). 10000 Monte-Carlo trials are used for the averaging. For each example, the ZP-based and CP-based OFDM schemes are compared fairly, i.e., the same CFO, channel, noise realization and training sequence for each Monte-Carlo run.

In the first example, $N = 16$ and the MSEs of the CFO estimation are plotted versus SNR for both ZP- and CP-based OFDM TWRN (Fig. 2.2). The theoretical MSEs are also displayed in the same figure. Fig. 2.2 reveals that for both protocols, the CFO estimation MSEs approach their

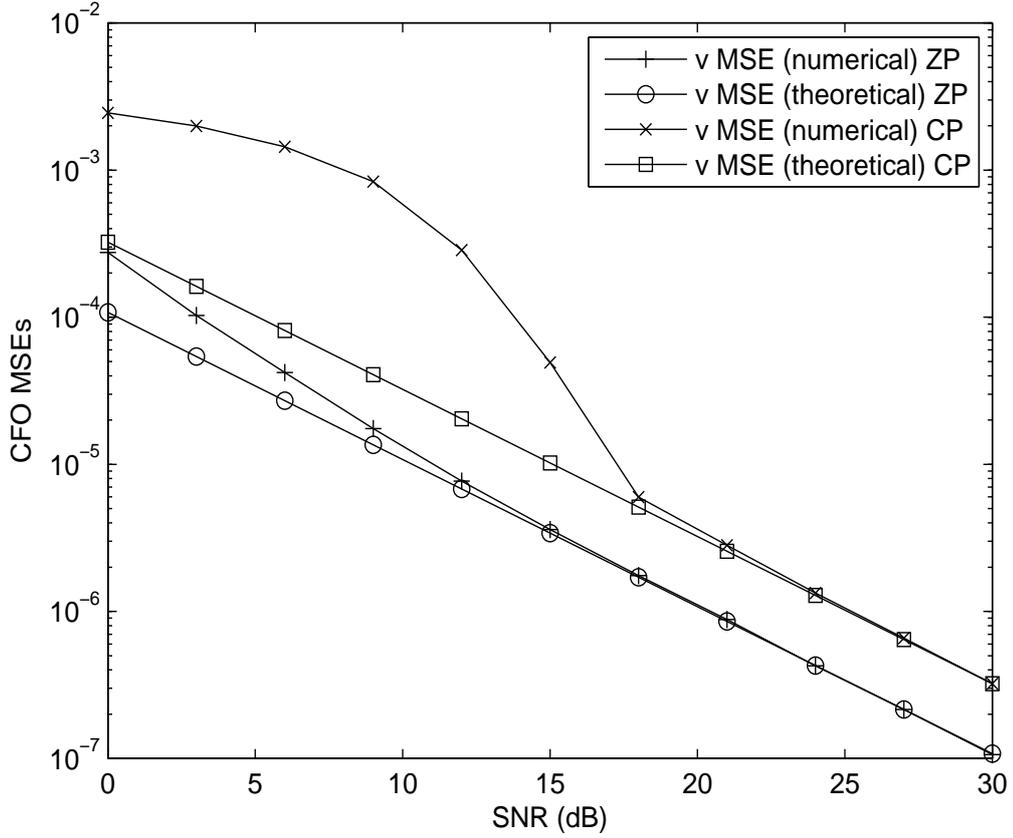


Fig. 2.2. CFO estimation MSEs version SNR for both ZP- and CP-based OFDM modulation; $N = 16$.

theoretical values at high SNR. The mismatch in the low SNR region, generally known as the *outlier* [79], [77], occurs because of the estimation ambiguity in several Monte-Carlo runs, which distorts the average performance. Moreover, as discussed in Section 2.5, CFO estimation from the ZP-based OFDM is better than that from the CP-based OFDM. The difference is observed to be 3 dB. Clearly, the reason is the longer signal block for the ZP-based OFDM.

The channel estimation MSEs versus SNR is shown in Fig. 2.3. The estimation MSEs of \mathbf{b}_{zp} and \mathbf{b}_{cp} approach their corresponding theoretical values much faster than those for CFO estimation because the errors in the estimated phase have less effect on the channel estimation but a more severe effect on the CFO estimation. Again, channel estimation of both \mathbf{a}_{zp} and \mathbf{b}_{zp} is better than that of \mathbf{a}_{cp} and \mathbf{b}_{cp} and the difference is about 1.5 dB.

The first two examples show that the ZP-based TWRN produces better estimates because of the longer observation block. Intuitively, the difference should decrease when block length N gets larger. It is then of interest to examine how the difference changes with N . Therefore, the CFO estimation MSEs versus the OFDM block length are plotted in Fig. 2.4 for a fixed SNR= 10 dB.

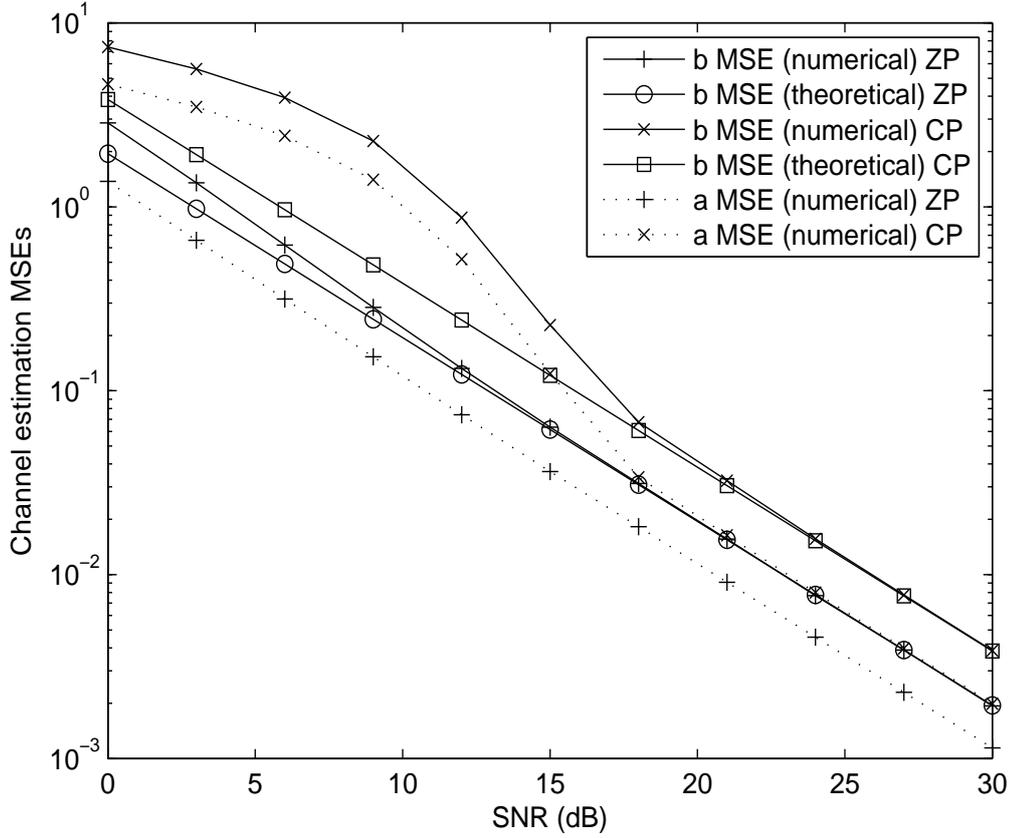


Fig. 2.3. Channel estimation MSEs version SNR for both ZP- and CP-based OFDM modulation; $N = 16$.

Fig. 2.4 reveals that with larger N , the estimations from both the ZP-based and CP-based OFDM become more accurate. Performance difference decreases as N increases and they converge for $N \geq 48$.

The channel estimation MSEs versus the OFDM block length is shown in Fig. 2.5, which suggests a similar observation made from Fig. 2.4. However, the difference between the ZP-OFDM and CP-OFDM diminishes at $N = 40$. Practically one may prefer to use CP-OFDM since it is more compatible with currently adopted standards, e.g., IEEE 802.11a [73]. One important insight gained from Fig. 2.4 and Fig. 2.5 is that the CP-OFDM will not suffer from much performance loss compared to the ZP-based OFDM when N is large. This makes our CP-OFDM protocol a good choice for practical standards, e.g., $N = 64, 1024, 2048$.

Finally, we examine the symbol decoding errors for both ZP- and CP-based OFDM protocols. To demonstrate the difference, we choose a small value of N as 16. The symbol constellations from both \mathbb{T}_1 and \mathbb{T}_2 are set as QPSK. The symbol-error rate (SER) versus SNR is shown in Fig. 2.6. We also include the detection performance with perfect synchronization and channel information for

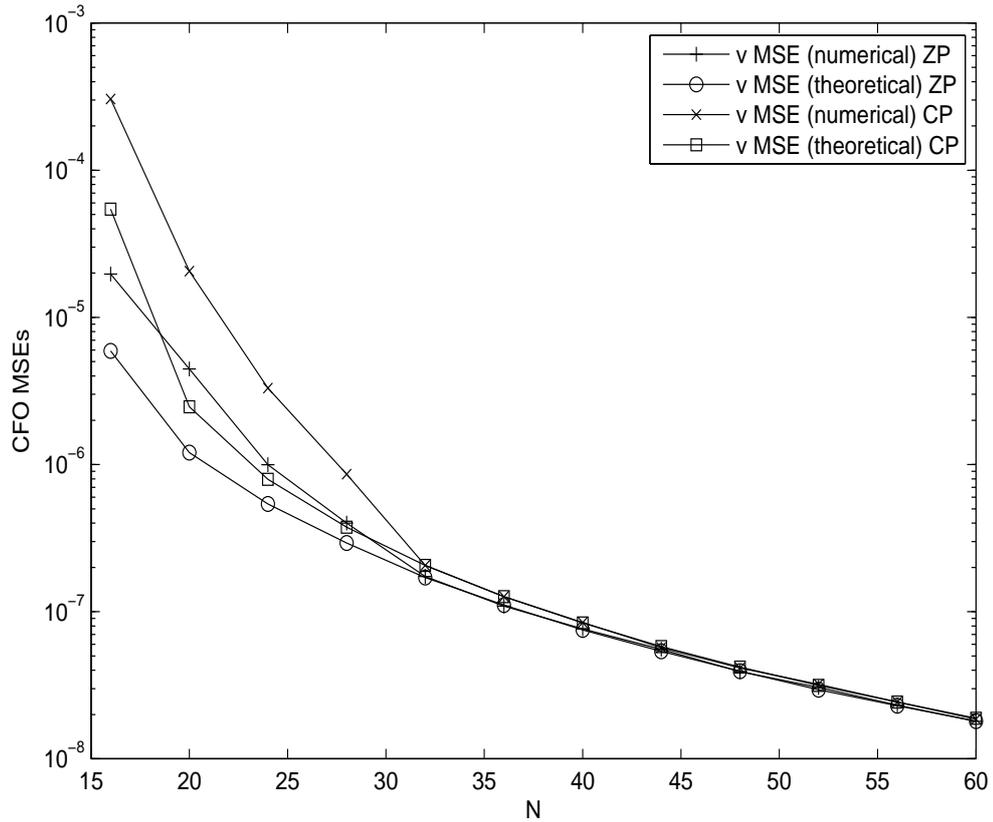


Fig. 2.4. CFO estimation MSEs version block length for both ZP- and CP-based OFDM modulation; SNR= 10 dB.

comparison. Interestingly, ZP- and CP-based OFDM exhibit similar detection performance under perfect synchronization and channel estimation. This phenomenon is similar to that observed in the point-to-point OFDM systems [78]. However, with the estimated CFO and channels, we identify a 2 dB performance gain in ZP-OFDM as compared to its CP counterpart. This gain immediately comes from the improved estimation accuracy.

2.8 Conclusions

In this chapter, we studied joint CFO and channel estimation for TWRN over frequency-selective channels. Our main contribution is the proposal and analysis of two OFDM transmission protocols for TWRN, which allow low-complexity joint estimation and data detection. The performance of the nulling-based LS estimator was studied by proving that it is unbiased at high SNR and by deriving the closed-form expressions of the MSEs. Finally, the simulation results demonstrated the effectiveness of the proposed schemes. Interestingly, although ZP-based OFDM performs better than the

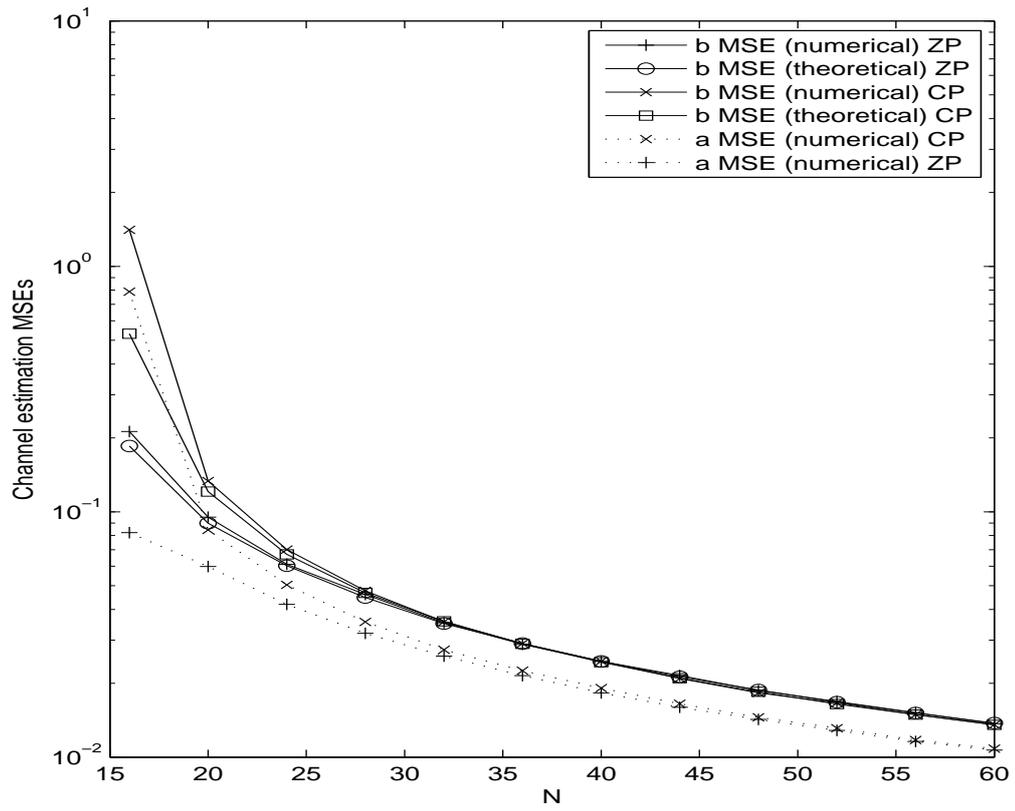


Fig. 2.5. Channel estimation MSEs version block length for both ZP- and CP-based OFDM modulation; SNR= 10 dB.

CP-based OFDM, the performance differential diminishes when the block length gets large. This practically useful finding suggests that CP-based OFDM could be preferred due to the compatibility with the existing OFDM standards.

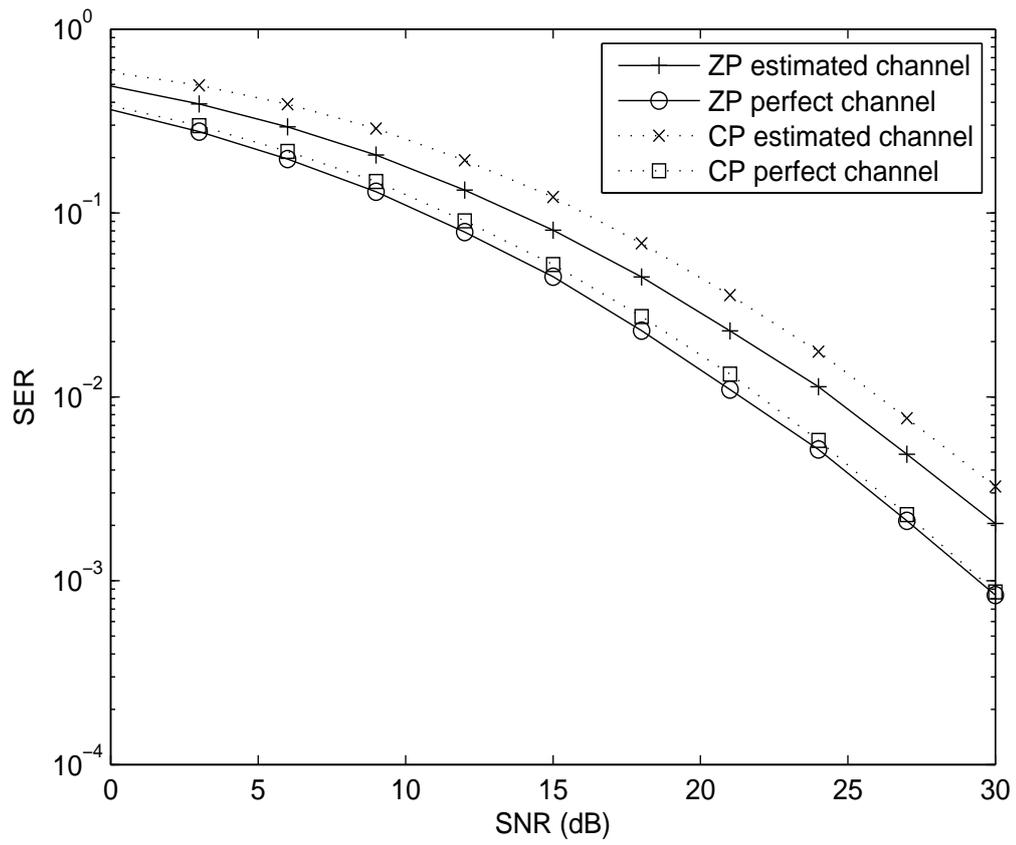


Fig. 2.6. Performance SER versus SNR for both ZP- and CP-based OFDM modulation; $N = 16$.

Chapter 3

Superimposed Training Aided Estimation for TWRN

Chapter 2 develops a joint CFO and channel estimator for TWRNs. But it can only provide the convoluted channel parameters and the mixed CFO values. In this chapter, we propose a superimposed training strategy to estimate the individual frequency and channel parameters. These parameters are useful in applications such as transmit beamforming or precoding. Depending on the available training length, we design three different algorithms for the initial parameter estimation. Then, an iterative process is proposed to further improve the estimation accuracy. To make the study complete, we also derive the estimation CRB of the proposed strategy. Our simulations show that the iteration converges in a few steps and the resultant estimation MSE approaches CRB. For the special case when the CFO between two terminals is small, the estimation MSE is very close to CRB in the high SNR region and the best estimation is expected even without iterations.

3.1 System Model

Consider a classical TWRN with two terminal nodes \mathbb{T}_j , $j = 1, 2$ and one relay node \mathbb{R} (Fig. 3.1). Each node has only one half-duplex antenna. The baseband channel from \mathbb{T}_j to \mathbb{R} is denoted as $\mathbf{h}_j = [h_{j,0}, \dots, h_{j,L}]^T$, whose elements are independent with zero means and variances $\sigma_{j,l}^2$. The channel from \mathbb{R} to \mathbb{T}_j is also \mathbf{h}_j . The training block length is set as N , which may or may not be the same as the data block length. The average powers of \mathbb{T}_j and \mathbb{R} are denoted as P_j and P_r , respectively. Furthermore, we denote the carrier frequency of \mathbb{T}_j as f_j and that of \mathbb{R} as f_r . In real applications, Doppler shifts and oscillator instabilities may result in CFOs such as $f_2 - f_1$ and $f_r - f_j$.

The target of this work is to separately estimate CFOs and channels \mathbf{h}_1 and \mathbf{h}_2 . It is important to

achieve this goal within two transmission phases such that the training is compatible with the two-phase data transmission and can be embedded in the data frame. To do so, we modify the OFDM transmission scheme and introduce superimposed training at the relay node.

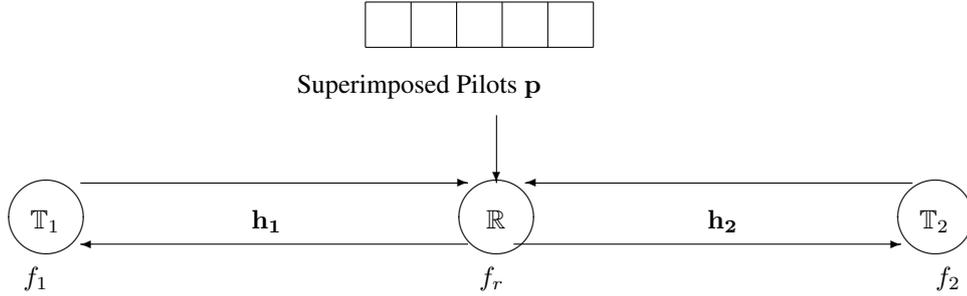


Fig. 3.1. System configuration for two-way relay network.

3.2 CP-Based OFDM Protocol

3.2.1 OFDM Modulation at Terminals

Denote one OFDM block from \mathbb{T}_i as $\tilde{\mathbf{s}}_i = [\tilde{s}_{i,0}, \dots, \tilde{s}_{i,N-1}]^T$. The corresponding time-domain signal block is obtained from the normalized inverse discrete Fourier transformation (IDFT) as

$$\mathbf{s}_i = \mathbf{F}^H \tilde{\mathbf{s}}_i = [s_{i,0}, s_{i,1}, \dots, s_{i,N-1}]^T, \quad (3.1)$$

where \mathbf{F} is the $N \times N$ normalized DFT matrix with the (p, q) -th entry given by $\frac{1}{\sqrt{N}} e^{-j2\pi(p-1)(q-1)/N}$. To maintain the subcarrier orthogonality during the overall transmission, we propose to add the cyclic prefix of length $2L$ as did in [90]. This implicitly requires $N \geq 2L$ which is nevertheless satisfied by most OFDM systems [73].

3.2.2 Relay Processing

With the assumption of the perfect time synchronization, the signals from \mathbb{T}_1 and \mathbb{T}_2 arrive at \mathbb{R} simultaneously. The relay then down-converts the received signal by the carrier $e^{-j2\pi f_r t}$. It is important to mention that \mathbb{R} removes only the first L symbols in each block. The resultant baseband signal block at \mathbb{R} is of length $N + L$ and can be expressed as

$$\mathbf{r} = \sum_{i=1}^2 e^{j2\pi(f_i - f_r)LT_s} \mathbf{\Gamma}^{(N+L)}[f_i - f_r] \mathbf{\Gamma}_{cp}^{(L)} \mathbf{H}_{cp}^{(N)}[\mathbf{h}_i] \mathbf{s}_i + \mathbf{n}_r, \quad (3.2)$$

where $\mathbf{\Gamma}_{cp}^{(L)}$ is defined in (2.15), $\mathbf{\Gamma}^{(K)}[f]$ is defined in (2.3) and $\mathbf{H}_{cv}^K[\mathbf{x}]$ is defined in (2.16).

The relay then superimposes a time-domain training signal \mathbf{p} over \mathbf{r} and obtain

$$\mathbf{t} = \alpha \mathbf{r} + \mathbf{p}, \quad (3.3)$$

where the scaling factor α and the superimposed training should satisfy the following power constraint:

$$\alpha^2 \mathbb{E}\{\|\mathbf{r}\|^2\} + \|\mathbf{p}\|^2 = \alpha^2(N+L) \sum_{j=1}^2 \sum_{l=0}^L \sigma_{j,l}^2 P_j + \alpha^2(N+L)\sigma_n^2 + \|\mathbf{p}\|^2 \leq (N+L)P_r. \quad (3.4)$$

Note that (3.4) is a constraint on the average power because the instant channel values are unknown before the estimation.

It can be easily shown that the range of α is $\left(0, \sqrt{\frac{P_r}{\sum_{j=1}^2 \sum_{l=0}^L \sigma_{j,l}^2 P_j + \sigma_n^2}}\right)$ that balances the power between the training from \mathbb{T}_j and the superimposed training from \mathbb{R} . Note that the training signal \mathbf{p} is generated from N training signals \mathbf{p}_0 and a cyclic prefix of length L . By using our definition (2.15), the training signal can be expressed as $\mathbf{p} = \mathbf{T}_{cp}^{(L)} \mathbf{p}_0$. The pilots \mathbf{p}_0 , \mathbf{s}_1 and \mathbf{s}_2 are pre-designed at both source terminals for channel estimation.

Finally, \mathbb{R} up-converts the resultant signal \mathbf{t} to passband by the carrier $e^{j2\pi f_r t}$.

3.2.3 Signal Reformulation at Terminals

Due to symmetry, we only illustrate the process at \mathbb{T}_1 . After down-converting the passband signal by $e^{-j2\pi f_1 t}$, \mathbb{T}_1 obtains the baseband block of length $N+L$. It then removes the first L elements and the remaining signal is written as

$$\begin{aligned} \mathbf{y} = & \alpha \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cv}^{(N)}[\mathbf{h}_1] \mathbf{\Gamma}^{(N+L)}[f_1 - f_r] \mathbf{T}_{cp}^{(L)} \mathbf{H}_{cp}^{(N)}[\mathbf{h}_1] \mathbf{s}_1 \\ & + \alpha e^{j2\pi(f_2 - f_1)LT_s} \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cv}^{(N)}[\mathbf{h}_1] \mathbf{\Gamma}^{(N+L)}[f_2 - f_r] \mathbf{T}_{cp}^{(L)} \mathbf{H}_{cp}^{(N)}[\mathbf{h}_2] \mathbf{s}_2 \\ & + e^{j2\pi(f_r - f_1)LT_s} \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cp}^{(N)}[\mathbf{h}_1] \mathbf{p}_0 \\ & + \underbrace{\alpha e^{j2\pi(f_r - f_1)LT_s} \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cv}^{(N)}[\mathbf{h}_1] \mathbf{n}_r + \mathbf{n}_1}_{\mathbf{n}_e}. \end{aligned} \quad (3.5)$$

The equivalent Gaussian noise \mathbf{n}_e has the covariance

$$\mathbf{R}_n = \sigma_n^2 \left(\alpha^2 \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cv}^{(N)}[\mathbf{h}_1] (\mathbf{H}_{cv}^{(N)}[\mathbf{h}_1])^H (\mathbf{\Gamma}^{(N)}[f_r - f_1])^H + \mathbf{I} \right). \quad (3.6)$$

Assuming $N > 2L$ and using the properties in Section (2.4), we can further simplify \mathbf{y} as

$$\begin{aligned} \mathbf{y} = & \alpha \mathbf{H}_{cp}^{(N)}[(\mathbf{\Omega}^{(L+1)}[f_1 - f_r] \mathbf{h}_1) \otimes \mathbf{h}_1] \mathbf{s}_1 \\ & + \alpha e^{j2\pi(f_2 - f_1)LT_s} \mathbf{\Gamma}^{(N)}[f_2 - f_1] \mathbf{H}_{cp}^{(N)}[(\mathbf{\Omega}^{(L+1)}[f_2 - f_r] \mathbf{h}_1) \otimes \mathbf{h}_2] \mathbf{s}_2 \end{aligned}$$

$$+ e^{j2\pi(f_r - f_1)LT_s} \mathbf{\Gamma}^{(N)}[f_r - f_1] \mathbf{H}_{cp}^{(N)}[\mathbf{h}_1] \mathbf{p}_0 + \mathbf{n}_e. \quad (3.7)$$

Define

$$\begin{aligned} w &= f_r - f_1, \quad \mathbf{a} = (\mathbf{\Omega}^{(L+1)}[-w] \mathbf{h}_1) \otimes \mathbf{h}_1, \\ v &= f_2 - f_1, \quad \mathbf{b} = (\mathbf{\Omega}^{(L+1)}[v - w] \mathbf{h}_1) \otimes \mathbf{h}_2, \quad \mathbf{\Gamma}_L^{(K)}[f] = e^{j2\pi f L T_s} \mathbf{\Gamma}^{(K)}[f]. \end{aligned} \quad (3.8)$$

Then (2.28) can be expressed as

$$\begin{aligned} \mathbf{y} &= \alpha \mathbf{H}_{cp}^{(N)}[\mathbf{a}] \mathbf{s}_1 + \alpha \mathbf{\Gamma}_L^{(N)}[v] \mathbf{H}_{cp}^{(N)}[\mathbf{b}] \mathbf{s}_2 + \mathbf{\Gamma}_L^{(N)}[w] \mathbf{H}_{cp}^{(N)}[\mathbf{h}_1] \mathbf{p}_0 + \mathbf{n}_e \\ &= \alpha \mathbf{S}_1 \mathbf{a} + \alpha \mathbf{\Gamma}_L^{(N)}[v] \mathbf{S}_2 \mathbf{b} + \mathbf{\Gamma}_L^{(N)}[w] \mathbf{P} \mathbf{h}_1 + \mathbf{n}_e. \end{aligned} \quad (3.9)$$

where \mathbf{S}_j is the $N \times (2L + 1)$ circulant matrix with the first column \mathbf{s}_i , and \mathbf{P} is the $N \times (L + 1)$ circulant matrix with the first column \mathbf{p}_0 .

3.3 Joint Estimation Algorithms

Based on the new signal model (3.9), the task is to estimate the individual channels \mathbf{h}_1 and \mathbf{h}_2 and the CFOs v and w . We omit all redundant superscripts and subscripts for notation simplicity and rewrite (3.9) as

$$\mathbf{y} = \alpha \mathbf{S}_1 \mathbf{a} + \alpha \mathbf{\Gamma}[v] \mathbf{S}_2 \mathbf{b} + \mathbf{\Gamma}[w] \mathbf{P} \mathbf{h}_1 + \mathbf{n}_e. \quad (3.10)$$

Note that the number of parameters to be estimated is $2L + 4$. Furthermore, \mathbf{a} is a function of w and \mathbf{h}_1 , and \mathbf{b} is a function of v, w, \mathbf{h}_1 and \mathbf{h}_2 . Depending on the number of pilots, we can thus develop three different estimation methods.

3.3.1 Estimation for Sufficiently Large N

When $N \geq 5L + 5$ ¹, there are sufficient degrees of freedom in the training signals, and $v, w, \mathbf{a}, \mathbf{b}$ and \mathbf{h}_1 can be simply treated as individual variables. That is, the above-mentioned relationships among the variables are ignored. Rewrite \mathbf{y} as

$$\mathbf{y} = \underbrace{[\alpha \mathbf{S}_1 \quad \alpha \mathbf{\Gamma}[v] \mathbf{S}_2 \quad \mathbf{\Gamma}[w] \mathbf{P}]}_{\mathbf{C}} \underbrace{\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{h}_1 \end{bmatrix}}_{\mathbf{d}} + \mathbf{n}_e, \quad (3.11)$$

¹In practical OFDM systems [73], $L = 16$ and N can be as large as 1024, 2048.

where \mathbf{C} and \mathbf{d} are defined in (3.11). From the least-squares (LS) method, the two CFO estimates are

$$\{\hat{v}, \hat{w}\} = \arg \max_{v,w} \mathbf{y}^H \mathbf{C} (\mathbf{C}^H \mathbf{C})^{-1} \mathbf{C}^H \mathbf{y}, \quad (3.12)$$

where \hat{v} and \hat{w} can be obtained either from a two dimensional search or from the alternating projection that converts the 2-dimensional maximization into a series of 1-D maximization problems. Details on the implementation of the alternating projection method can be found from [80]. Then values of \mathbf{d} is estimated as

$$\hat{\mathbf{d}} = (\mathbf{C}^H \mathbf{C})^{-1} \mathbf{C}^H \mathbf{y}, \quad (3.13)$$

where the value of \mathbf{C} is obtained by using the estimates \hat{v} and \hat{w} .

We next explore the relationships among \mathbf{a} , \mathbf{b} , and \mathbf{h}_1 to improve the quality of the estimates. From (3.8), we note that

$$\mathbf{a} = \mathbf{H}_{zp}^{(L+1)} [\boldsymbol{\Omega}^{(L+1)} [-w] \mathbf{h}_1] \mathbf{h}_1, \quad (3.14)$$

where $\mathbf{H}_{zp}^{(K)}[\mathbf{x}]$ is a tall Toeplitz matrix with the first column $[\mathbf{x}^T, \mathbf{o}_{1 \times (K-1)}^T]$. The estimate of $\mathbf{H}_{zp}^{(L+1)} [\boldsymbol{\Omega}^{(L+1)} [-w] \mathbf{h}_1]$ can be expressed as $\mathbf{H}_{zp}^{(L+1)} [\boldsymbol{\Omega}^{(L+1)} [-\hat{w}] \hat{\mathbf{h}}_1]$.

By subtracting the estimate of the second item in (3.10) from \mathbf{y} and by using (3.14), an improved estimate of \mathbf{h}_1 is obtained as

$$\hat{\mathbf{h}}_1 = (\alpha \mathbf{S}_1 \mathbf{H}_{zp}^{(L+1)} [\boldsymbol{\Omega}^{(L+1)} [-\hat{w}] \hat{\mathbf{h}}_1] + \boldsymbol{\Gamma}[w] \mathbf{P})^\dagger (\mathbf{y} - \alpha \boldsymbol{\Gamma}[v] \mathbf{S}_2 \hat{\mathbf{b}}). \quad (3.15)$$

Similarly, from (3.8) we find that

$$\mathbf{b} = \mathbf{H}_{zp}^{(L+1)} [\boldsymbol{\Omega}^{(L+1)} [v - w] \mathbf{h}_1] \mathbf{h}_2. \quad (3.16)$$

Thus \mathbf{h}_2 can be estimated as

$$\hat{\mathbf{h}}_2 = (\mathbf{H}_{zp}^{(L+1)} [\boldsymbol{\Omega}^{(L+1)} [\hat{v} - \hat{w}] \hat{\mathbf{h}}_1])^\dagger \hat{\mathbf{b}}. \quad (3.17)$$

In summary, equations (3.12), (3.15) and (3.17) provide estimates of all the parameters. These initial estimates can be further improved by the iterative estimator developed in Section 3.3.4.

3.3.2 Estimation with Not-so-large N

In order to reduce the overhead, we will use fewer pilots than before. Define \mathcal{K}_1 , \mathcal{K}_2 , and \mathcal{K}_r as the frequency domain pilot index sets from \mathbb{T}_1 , \mathbb{T}_2 , and \mathbb{R} , with cardinality K_1 , K_2 , and K_r respectively. We require $K_1 \geq L + 1$, $K_2 \geq L + 1$, $K_r \geq L + 1$ and $\mathcal{K}_1 \cup \mathcal{K}_2 \cup \mathcal{K}_r = \{1, \dots, N\}$. Here, we do not assume disjoint sets, and so $K_1 + K_2 + K_r \geq N$. From (3.10), we know that the frequency domain pilots are $\tilde{\mathbf{s}}_j = \mathbf{F}\mathbf{s}_j$, and $\tilde{\mathbf{p}}_0 = \mathbf{F}\mathbf{p}_0$. As will be seen later that pilots in \mathcal{K}_r are used to estimate \mathbf{h}_1 and those in \mathcal{K}_2 are used to estimate \mathbf{h}_2 at \mathbb{T}_1 . Due to symmetry, pilots in \mathcal{K}_1 are used to estimate \mathbf{h}_1 at \mathbb{T}_2 . This gives the above requirements on cardinality.

Let us collect non-zero pilots from \mathbb{T}_j into a $K_j \times 1$ vector $\check{\mathbf{s}}_j$ and non-zero pilots from \mathbb{R} into a $K_r \times 1$ vector $\check{\mathbf{p}}_0$. Since \mathbf{S}_j and \mathbf{P} are columnwise circulant matrices, they can be represented as

$$\mathbf{S}_j = \mathbf{F}^H \text{diag}\{\check{\mathbf{s}}_j\} \mathbf{F}_{[:,1:2L+1]} = \mathbf{F}_{[:,\mathcal{K}_j]}^H \text{diag}\{\check{\mathbf{s}}_j\} \mathbf{F}_{[\mathcal{K}_j,1:2L+1]} \quad (3.18)$$

$$\mathbf{P} = \mathbf{F}^H \text{diag}\{\check{\mathbf{p}}_0\} \mathbf{F}_{[:,1:L+1]} = \mathbf{F}_{[:,\mathcal{K}_r]}^H \text{diag}\{\check{\mathbf{p}}_0\} \mathbf{F}_{[\mathcal{K}_r,1:L+1]}. \quad (3.19)$$

Define $\bar{\mathcal{K}}_1$ as the complement set of \mathcal{K}_1 . Multiplying both sides of (3.10) with $\mathbf{F}_{[\bar{\mathcal{K}}_1,:]}$ yields

$$\mathbf{F}_{[\bar{\mathcal{K}}_1,:]} \mathbf{y} = \underbrace{\left[\alpha \mathbf{F}_{[\bar{\mathcal{K}}_1,:]} \mathbf{\Gamma}[v] \mathbf{F}_{[:,\mathcal{K}_2]}^H \text{diag}\{\check{\mathbf{s}}_2\} \quad \mathbf{F}_{[\bar{\mathcal{K}}_1,:]} \mathbf{\Gamma}[w] \mathbf{P} \right]}_{\mathbf{C}_1} \underbrace{\begin{bmatrix} \check{\mathbf{b}} \\ \mathbf{h}_1 \end{bmatrix}}_{\mathbf{d}_1} + \mathbf{F}_{[\bar{\mathcal{K}}_1,:]} \mathbf{n}_e, \quad (3.20)$$

where $\check{\mathbf{b}} = \mathbf{F}_{[\mathcal{K}_2,1:2L+1]} \mathbf{b}$ is the DFT response of \mathbf{b} on the subcarrier set \mathcal{K}_2 and \mathbf{C}_1 is an $(N - K_1) \times (K_2 + L + 1)$ matrix.

As long as $N - K_1 - K_2 - (L + 1) \geq 2$, namely when there is sufficient degree of freedom to estimate the two unknown CFOs, they can be estimated as

$$\{\hat{v}, \hat{w}\} = \arg \max_{v,w} \mathbf{y}^H \mathbf{F}_{[\bar{\mathcal{K}}_1,:]}^H \mathbf{C}_1 (\mathbf{C}_1^H \mathbf{C}_1)^{-1} \mathbf{C}_1^H \mathbf{F}_{[\bar{\mathcal{K}}_1,:]} \mathbf{y}. \quad (3.21)$$

Considering the range of K_j and K_r , the minimum number of N is $3L + 5$, when sets are disjoint and $K_1 = K_2 = L + 1$, $K_r = L + 3$. Then,

$$\hat{\mathbf{d}}_1 = (\mathbf{C}_1^H \mathbf{C}_1)^{-1} \mathbf{C}_1^H \mathbf{F}_{[\bar{\mathcal{K}}_1,:]} \mathbf{y}. \quad (3.22)$$

By definition

$$\check{\mathbf{b}} = \mathbf{F}_{[\mathcal{K}_2,1:2L+1]} \mathbf{H}_{zp}^{(L+1)} [\mathbf{\Omega}^{(L+1)} [v - w] \mathbf{h}_1] \mathbf{h}_2. \quad (3.23)$$

Then, \mathbf{h}_2 can be estimated as

$$\hat{\mathbf{h}}_2 = (\mathbf{F}_{[\mathcal{K}_2, 1:2L+1]} \mathbf{H}_{zp}^{(L+1)} [\boldsymbol{\Omega}^{(L+1)} [v - w] \mathbf{h}_1])^\dagger \check{\mathbf{b}}. \quad (3.24)$$

3.3.3 Joint Estimation with Minimum Training Length: A Special Case

In practical applications, the relay terminal is often a simple device while the two source terminals may employ high-precision synchronization circuits. Thus, it is reasonable to expect the CFO between the two source terminals to be negligible. In this case, $v \approx 0$ or $v \ll 1/N$, i.e., one sub-carrier spacing. This is also true at the CFO tracking stage when the frequency difference between two terminals are quite small. By taking advantage of the negligible CFO between the two source terminals, parameter estimation is achieved with the minimum training length $N = 2L + 3$, i.e., the same number of the unknowns variables.

Let us choose the same frequency pilot sets for \mathbb{T}_1 and \mathbb{T}_2 , i.e., $\mathcal{K}_1 = \mathcal{K}_2$. Left multiplying the received signal \mathbf{y} by $\mathbf{F}_{[\bar{\mathcal{K}}_1, :]}$ gives

$$\mathbf{F}_{[\bar{\mathcal{K}}_1, :]} \mathbf{y} = \underbrace{\alpha \mathbf{F}_{[\bar{\mathcal{K}}_1, :]} \boldsymbol{\Gamma}[v] \mathbf{F}_{[:, \mathcal{K}_2]}^H \text{diag}\{\check{\mathbf{s}}_2\} \mathbf{F}_{[\mathcal{K}_2, 1:2L+1]} \mathbf{b}}_{\approx \mathbf{0}} + \underbrace{\mathbf{F}_{[\bar{\mathcal{K}}_1, :]} \boldsymbol{\Gamma}[w] \mathbf{P} \mathbf{h}_1}_{\mathbf{C}_2} + \mathbf{F}_{[\bar{\mathcal{K}}_1, :]} \mathbf{n}_e, \quad (3.25)$$

where \mathbf{C}_2 is an $(N - K_1) \times (L + 1)$ matrix, and the first term is negligible because $v \approx 0$. As long as $(N - K_1 - L - 1) \geq 1$, the CFO between the relay and the first source terminal can be obtained as

$$\hat{w} = \arg \max_w \mathbf{y}^H \mathbf{F}_{[\bar{\mathcal{K}}_1, :]}^H \mathbf{C}_2 (\mathbf{C}_2^H \mathbf{C}_2)^{-1} \mathbf{C}_2^H \mathbf{F}_{[\bar{\mathcal{K}}_1, :]} \mathbf{y}, \quad (3.26)$$

and

$$\hat{\mathbf{h}}_1 = \mathbf{C}_2^\dagger \mathbf{F}_{[\bar{\mathcal{K}}_1, :]} \mathbf{y}. \quad (3.27)$$

Since \hat{w} can be estimated from \hat{w} and $\hat{\mathbf{h}}_1$, then \mathbf{h}_2 can be estimated as

$$\hat{\mathbf{h}}_2 = \mathbf{C}_3^\dagger (\mathbf{y} - \alpha \mathbf{S}_1 \hat{\mathbf{a}} - \boldsymbol{\Gamma}[\hat{w}] \mathbf{P} \hat{\mathbf{h}}_1), \quad (3.28)$$

where $\mathbf{C}_3 = \alpha \mathbf{S}_2 \mathbf{H}_{zp}^{(L+1)} [\boldsymbol{\Omega}^{(L+1)} [\hat{v} - \hat{w}] \hat{\mathbf{h}}_1]$.

The estimates given by (3.26), (3.27) and (3.28) need not be improved by iterations because they already achieve CRB under high SNR conditions.

3.3.4 Iterative Algorithm to Improve the Performance

With the initial estimates of all the parameters, an iterative approach can be applied to improve the estimation accuracy. Re-denote the initial estimate as $v^{\{0\}}, w^{\{0\}}, \mathbf{a}^{\{0\}}, \mathbf{b}^{\{0\}}, \mathbf{h}_1^{\{0\}}, \mathbf{h}_2^{\{0\}}$, respectively, with the superscript representing the number of the iterations. We will estimate $v^{\{1\}}, w^{\{1\}}$ simultaneously from the ML estimation process as

$$[v^{\{1\}}, w^{\{1\}}] = \arg \min_{v, w} (\mathbf{y} - \alpha \mathbf{S}_1 \mathbf{a}^{\{0\}} - \alpha \Gamma[v] \mathbf{S}_2 \mathbf{b}^{\{0\}} - \Gamma[w] \mathbf{P} \mathbf{h}_1^{\{0\}})^H \mathbf{R}_n^{-1} \times (\mathbf{y} - \alpha \mathbf{S}_1 \mathbf{a}^{\{0\}} - \alpha \Gamma[v] \mathbf{S}_2 \mathbf{b}^{\{0\}} - \Gamma[w] \mathbf{P} \mathbf{h}_1^{\{0\}}), \quad (3.29)$$

where \mathbf{R}_n^{-1} is always obtained by using the newest estimates of w and \mathbf{h}_1 in (3.6) expressed as

$$\mathbf{R}_n = \sigma_n^2 \left(\alpha^2 \Gamma^{(N)}[w^{\{0\}}] \mathbf{H}_{cv}^{(N)}[\mathbf{h}_1^{\{0\}}] (\mathbf{H}_{cv}^{(N)}[\mathbf{h}_1^{\{0\}}])^H (\Gamma^{(N)}[w^{\{0\}}])^H + \mathbf{I} \right). \quad (3.30)$$

Then we can obtain $\mathbf{h}_2^{\{1\}}$ and $\mathbf{h}_1^{\{1\}}$ from

$$\mathbf{h}_2^{\{1\}} = \arg \min_{\mathbf{h}_2} (\mathbf{y} - \alpha \mathbf{S}_1 \mathbf{a}^{\{0\}} - \Gamma[w^{\{1\}}] \mathbf{P} \mathbf{h}_1^{\{0\}} - \alpha \Gamma[v^{\{1\}}] \mathbf{S}_2 \mathbf{H}_{12}^{\{1\}} \mathbf{h}_2)^H \mathbf{R}_n^{-1} \times (\mathbf{y} - \alpha \mathbf{S}_1 \mathbf{a}^{\{0\}} - \Gamma[w^{\{1\}}] \mathbf{P} \mathbf{h}_1^{\{0\}} - \alpha \Gamma[v^{\{1\}}] \mathbf{S}_2 \mathbf{H}_{12}^{\{1\}} \mathbf{h}_2), \quad (3.31)$$

and

$$\mathbf{h}_1^{\{1\}} = \arg \min_{\mathbf{h}_1} (\mathbf{y} - \mathbf{M}_{h_1} \mathbf{h}_1)^H \mathbf{R}_n^{-1} (\mathbf{y} - \mathbf{M}_{h_1} \mathbf{h}_1), \quad (3.32)$$

where $\mathbf{M}_{h_1} = \alpha \mathbf{S}_1 \mathbf{H}_{11}^{\{0\}} + \alpha \Gamma[v^{\{1\}}] \mathbf{S}_2 \mathbf{H}_{zp}^{(L+1)}[\mathbf{h}_2^{\{1\}}] \boldsymbol{\Omega}^{L+1}[v^{\{1\}} - w^{\{1\}}] + \Gamma[w^{\{1\}}] \mathbf{P}$, $\mathbf{H}_{12}^{\{1\}} = \mathbf{H}_{zp}^{(L+1)}[\boldsymbol{\Omega}^{(L+1)}[v^{\{1\}} - w^{\{1\}}] \mathbf{h}_1^{\{0\}}]$ and $\mathbf{H}_{11}^{\{0\}} = \mathbf{H}_{zp}^{(L+1)}[\boldsymbol{\Omega}^{(L+1)}[-w^{\{1\}}] \mathbf{h}_1^{\{0\}}]$.

The interactive processing could gain the improvement from the fact that the initial estimation does not fully exploit the correlation between \mathbf{a} , \mathbf{b} , and \mathbf{h}_1 .

3.3.5 Comparison

A comparison between the adapted CP-OFDM (Section 2.4) and superimposed pilot aided CP-OFDM is given in the following table.

TABLE 3.1
COMPARISON BETWEEN ADAPTED CP-OFDM AND SUPERIMPOSED PILOT AIDED CP-OFDM.

	Minimum Pilot Length	Estimated Parameters
adapted CP-OFDM	$4L + 3$	\mathbf{a} , \mathbf{b} and v
superimposed pilot	$3L + 5$	\mathbf{h}_1 , \mathbf{h}_2 , \mathbf{a} , \mathbf{b} , v and w

3.4 Cramér-Rao Bound

In this section, we derive the CRB that defines the theoretical bound of the estimation accuracy. The CRB is an important tool to study the performance of the estimation algorithms.

Define

$$\boldsymbol{\mu} \triangleq \alpha \mathbf{S}_1 \mathbf{a} + \alpha \Gamma[v] \mathbf{S}_2 \mathbf{b} + \Gamma[w] \mathbf{P} \mathbf{h}_1, \quad (3.33)$$

$$\boldsymbol{\eta} \triangleq [v, w, \Re\{\mathbf{h}_1\}^T, \Im\{\mathbf{h}_1\}^T, \Re\{\mathbf{h}_2\}^T, \Im\{\mathbf{h}_2\}^T]^T. \quad (3.34)$$

According to [66], the (i, j) th entry of the Fisher Information Matrix (FIM) can be calculated as

$$[\mathbf{F}]_{i,j} = 2 \Re \left[\frac{\partial \boldsymbol{\mu}^H}{\partial \eta_i} \mathbf{R}_n^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \eta_j} \right] + \text{tr} \left[\mathbf{R}_n^{-1} \frac{\partial \mathbf{R}_n}{\partial \eta_i} \mathbf{R}_n^{-1} \frac{\partial \mathbf{R}_n}{\partial \eta_j} \right]. \quad (3.35)$$

After some tedious simplifications, we derive

$$\frac{\partial \boldsymbol{\mu}}{\partial v} = j \alpha \mathbf{D}_1 \Gamma[v] \mathbf{S}_2 \mathbf{b} + j \Gamma[v] \mathbf{S}_2 \mathbf{H}_{zp}^{(L+1)} [\mathbf{h}_2] \mathbf{D}_0 \boldsymbol{\Omega}^{(L+1)} [v - w] \mathbf{h}_1, \quad (3.36)$$

$$\begin{aligned} \frac{\partial \boldsymbol{\mu}}{\partial w} &= \alpha \mathbf{S}_1 \mathbf{H}_{zp}^{(L+1)} [\mathbf{h}_1] (-j \mathbf{D}_0) \boldsymbol{\Omega}^{(L+1)} [-w] \mathbf{h}_1 \\ &\quad + \alpha \Gamma[v] \mathbf{S}_2 \mathbf{H}_{zp}^{(L+1)} [\mathbf{h}_2] (-j \mathbf{D}_0) \boldsymbol{\Omega}^{(L+1)} [v - w] \mathbf{h}_1 + j \mathbf{D}_1 \Gamma[w] \mathbf{P} \mathbf{h}_1, \end{aligned} \quad (3.37)$$

$$\begin{aligned} \frac{\partial \boldsymbol{\mu}}{\partial \Re\{\mathbf{h}_1\}^T} &= \alpha \mathbf{S}_1 \mathbf{H}_{zp}^{(L+1)} [\boldsymbol{\Omega}^{(L+1)} [-w] \mathbf{h}_1] + \alpha \mathbf{S}_1 \mathbf{H}_{zp}^{(L+1)} [\mathbf{h}_1] \boldsymbol{\Omega}^{(L+1)} [-w] \\ &\quad + \alpha \Gamma[v] \mathbf{S}_2 \mathbf{H}_{zp}^{(L+1)} [\mathbf{h}_2] \boldsymbol{\Omega}^{(L+1)} [v - w] + \Gamma[w] \mathbf{P}, \end{aligned} \quad (3.38)$$

$$\frac{\partial \boldsymbol{\mu}}{\partial \Re\{\mathbf{h}_2\}^T} = \alpha \Gamma[v] \mathbf{S}_2 \mathbf{H}_{zp}^{(L+1)} [\boldsymbol{\Omega}^{(L+1)} [v - w] \mathbf{h}_1], \quad (3.39)$$

$$\begin{aligned} \frac{\partial \mathbf{R}_n}{\partial w} &= \sigma_n^2 (j \alpha^2 \mathbf{D}_N \Gamma^{(N)} [w] \mathbf{H}_{cv}^{(N)} [\mathbf{h}_1] (\mathbf{H}_{cv}^{(N)} [\mathbf{h}_1])^H \Gamma^{(N)} [-w] \\ &\quad - j \alpha^2 \Gamma^{(N)} [w] \mathbf{H}_{cv}^{(N)} [\mathbf{h}_1] (\mathbf{H}_{cv}^{(N)} [\mathbf{h}_1])^H \mathbf{D}_N \Gamma^{(N)} [-w]) \end{aligned} \quad (3.40)$$

$$\frac{\partial \mathbf{R}_n}{\partial \Re\{\mathbf{h}_{1i}\}^T} = \sigma_n^2 \alpha^2 \Gamma^{(N)} [w] \mathbf{H}_{cv}^{(N)} [\mathbf{e}_{1i}] (\mathbf{H}_{cv}^{(N)} [\mathbf{h}_1])^H \Gamma^{(N)} [-w] \quad (3.41)$$

$$\frac{\partial \mathbf{R}_n}{\partial \Re\{\mathbf{h}_{2i}\}^T} = \frac{\partial \mathbf{R}_n}{\partial v} = \mathbf{0}_{N \times N}, \quad (3.42)$$

where

$$\mathbf{D}_0 = 2\pi T_s \text{diag}\{L, (L-1), \dots, 1, 0\}, \quad (3.43)$$

$$\mathbf{D}_1 = 2\pi T_s \text{diag}\{L, \dots, (L+N-1)\}, \quad (3.44)$$

$$\mathbf{D}_N = 2\pi T_s \text{diag}\{0, 1, \dots, (N-1)\}, \quad (3.45)$$

and \mathbf{e}_i is a $(L+1) \times 1$ vector whose i^{th} element equals 1 and others 0.

The CRB of $\boldsymbol{\eta}$ is then obtained by inverting the FIM \mathbf{F} . Since

$$\begin{bmatrix} v \\ w \\ \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1, & 0, & \mathbf{0}^T, & \mathbf{0}^T, & \mathbf{0}^T, & \mathbf{0}^T \\ 0, & 1, & \mathbf{0}^T, & \mathbf{0}^T, & \mathbf{0}^T, & \mathbf{0}^T \\ \mathbf{0}, & \mathbf{0}, & \mathbf{I}_{L+1}, & j\mathbf{I}_{L+1}, & 0 \cdot \mathbf{I}_{L+1}, & 0 \cdot \mathbf{I}_{L+1} \\ \mathbf{0}, & \mathbf{0}, & 0 \cdot \mathbf{I}_{L+1}, & 0 \cdot \mathbf{I}_{L+1}, & \mathbf{I}_{L+1}, & j\mathbf{I}_{L+1} \end{bmatrix}}_{\boldsymbol{\Xi}} \times \boldsymbol{\eta}, \quad (3.46)$$

the CRB of v, w, \mathbf{h}_1 , and \mathbf{h}_2 can be expressed as

$$\text{CRB} = \boldsymbol{\Xi} \cdot \mathbf{F}^{-1} \cdot \boldsymbol{\Xi}^H. \quad (3.47)$$

3.5 Simulation Results

The performance of the proposed three estimation algorithms along with the iterative estimator is investigated. A four-tap model for both \mathbf{h}_1 and \mathbf{h}_2 is assumed, and each tap is assumed complex Gaussian with unit variance as did in [14]. The variance of the noise is taken as $\sigma_n^2 = 1$. The normalized frequencies f_1, f_r , and f_2 are set as 0.94, 1 and 1.06, respectively. The MSE is chosen as the figure of merit, defined by

$$\begin{aligned} \text{MSE}(v) &= \frac{1}{10000} \sum_{i=1}^{10000} (\hat{v}_i - v)^2, \\ \text{MSE}(w) &= \frac{1}{10000} \sum_{i=1}^{10000} (\hat{w}_i - w)^2, \\ \text{MSE}(\mathbf{x}) &= \frac{1}{10000} \sum_{i=1}^{10000} \frac{1}{4} (\hat{\mathbf{x}}_i - \mathbf{x})^2, \end{aligned}$$

where \mathbf{x} represents \mathbf{h}_1 or \mathbf{h}_2 , and 10000 is the number of the Monte-Carlo trials used for average. In all the following simulations, α is set as half the maximum value, that is, $\alpha = 0.5 \sqrt{\frac{P_r}{\sum_{j=1}^2 \sum_{l=0}^L \sigma_{j,l}^2 P_j + \sigma_n^2}}$.

3.5.1 Sufficiently Large N

In this case, we chose $N = 24$, which is greater than $5L + 5 = 20$. The received signal \mathbf{y} at \mathbb{T}_1 is generated according to (3.5). Initial CFO and channel estimates are obtained from \mathbf{y} through (3.12) and (3.13). The estimate $\hat{\mathbf{h}}_1$ of \mathbf{h}_1 is updated as (3.15). Finally, the iterative estimator in Section 3.3.4 is applied and is found to converge in three iterations.

The MSEs and CRBs of CFO estimation as a function of SNR are shown in Fig. 3.2. The iterative algorithm improves the estimation accuracy significantly. Specifically, the improvement of w , which is the CFO between \mathbb{R} and \mathbb{T}_1 , is much more significant than that of v , the CFO between

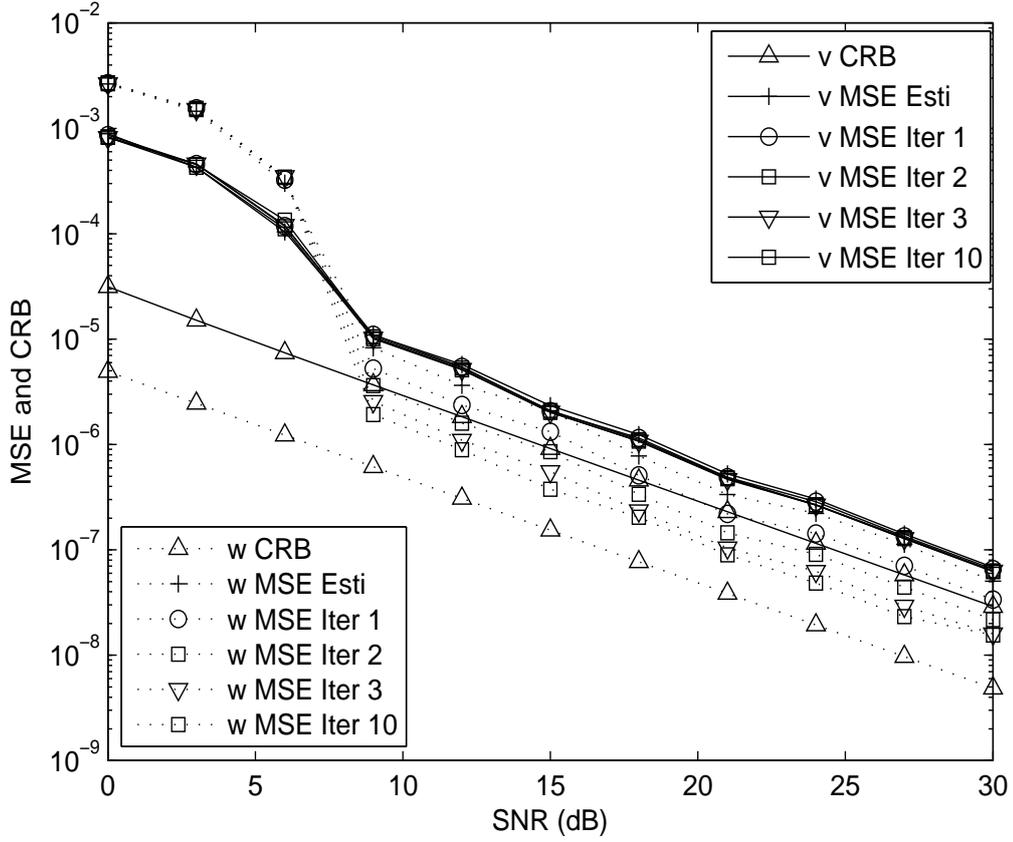


Fig. 3.2. CFO estimation MSE versus SNR: $N = 24$

\mathbb{T}_2 and \mathbb{T}_1 . The reason is due to the fact that the received signals contain more information about w than that about v . In (3.9), all components of \mathbf{y} contain information about w while only the second term of \mathbf{y} does so of v .

The channel estimation MSEs and CRBs versus SNR are shown in Fig. 3.3. It is observed that the gaps between the MSE and the CRB are smaller compared to those in the CFO estimation. The reason is that phase errors have less effect on the channel estimation than on the CFO estimation. Similarly to CFO estimation, iteration improves the estimation accuracy, and \mathbf{h}_1 improves more than \mathbf{h}_2 since most components of \mathbf{y} contain the information of \mathbf{h}_1 .

3.5.2 Not-so-large N

Next we choose $N = 3L + 5 = 14$ and $\mathcal{K}_1 = 4, \mathcal{K}_2 = 4, \mathcal{K}_r = 6$. The initial CFO and channel estimates are obtained from (3.21), (3.22) and (3.24). The estimates are iteratively updated. We find that ten iterations reach convergence. The MSEs and CRBs versus SNRs for both CFO and channel estimation are displayed in Fig. 3.4 and Fig. 3.5, respectively.

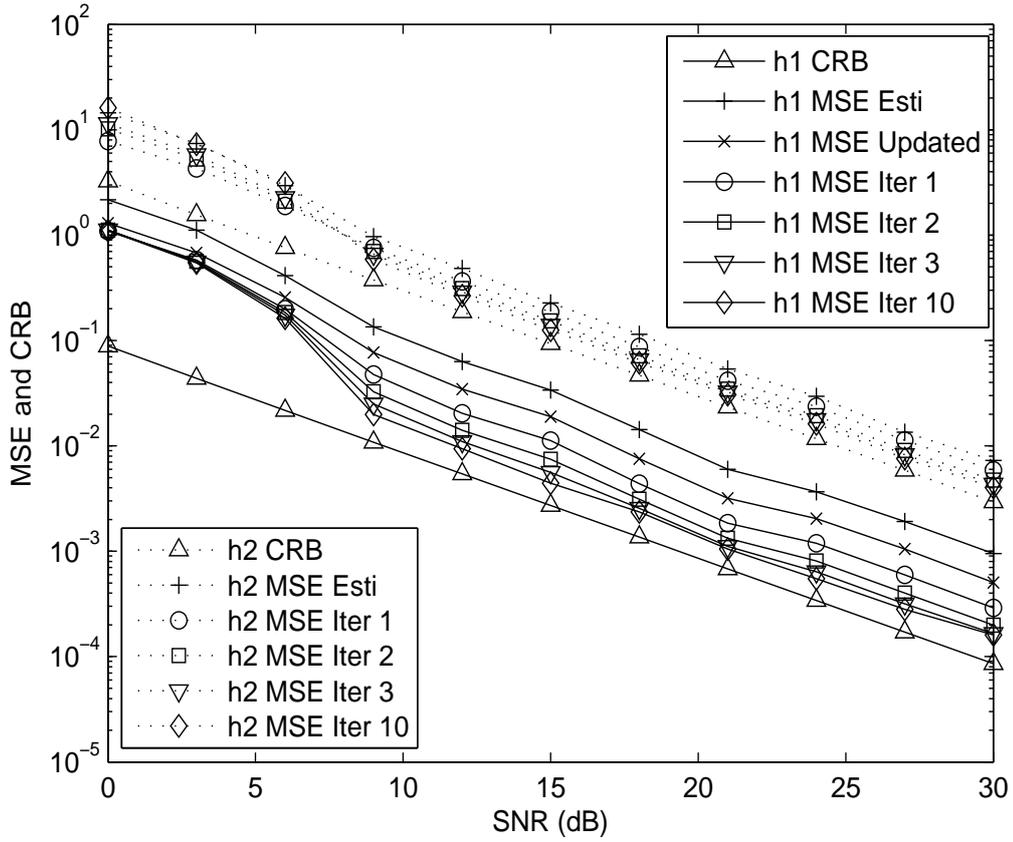


Fig. 3.3. Channel estimation MSE versus SNR: $N = 24$

Since smaller training length is applied and the average symbol power is kept the same, the performance here is a little worse than that in Fig. 3.2 and Fig. 3.3. It is observed that in the high SNR region, the MSEs approach CRBs after ten iterations. The iterative estimator improves the estimation of \mathbf{h}_1 for all SNRs while, for \mathbf{h}_2 , it only works at the high SNR region. Conversely, the iterative estimator degrades the estimation of \mathbf{h}_2 at low SNRs. A possible reason is as follows. Since only the second item of \mathbf{y} in (3.9) contains \mathbf{h}_2 and iterations require reconstruction of \mathbf{a} from the initial estimate $\hat{\mathbf{w}}$ and $\hat{\mathbf{h}}_1$, the ambiguity of $\hat{\mathbf{h}}_2$ increased at the low SNR region from errors of all factors. However, in the case of large training length, e.g., $N = 24$, \mathbf{a} is directly estimated, which avoids the error propagation from erroneous $\hat{\mathbf{w}}$ and $\hat{\mathbf{h}}_1$. Therefore similar phenomenon is not observed in Fig. 3.3.

From Fig. 3.2 to Fig. 3.5, one key conclusion is that there is room for new algorithms to improve the performance, due to the gap between MSEs and CRBs. This open question is an interesting topic for the future research.

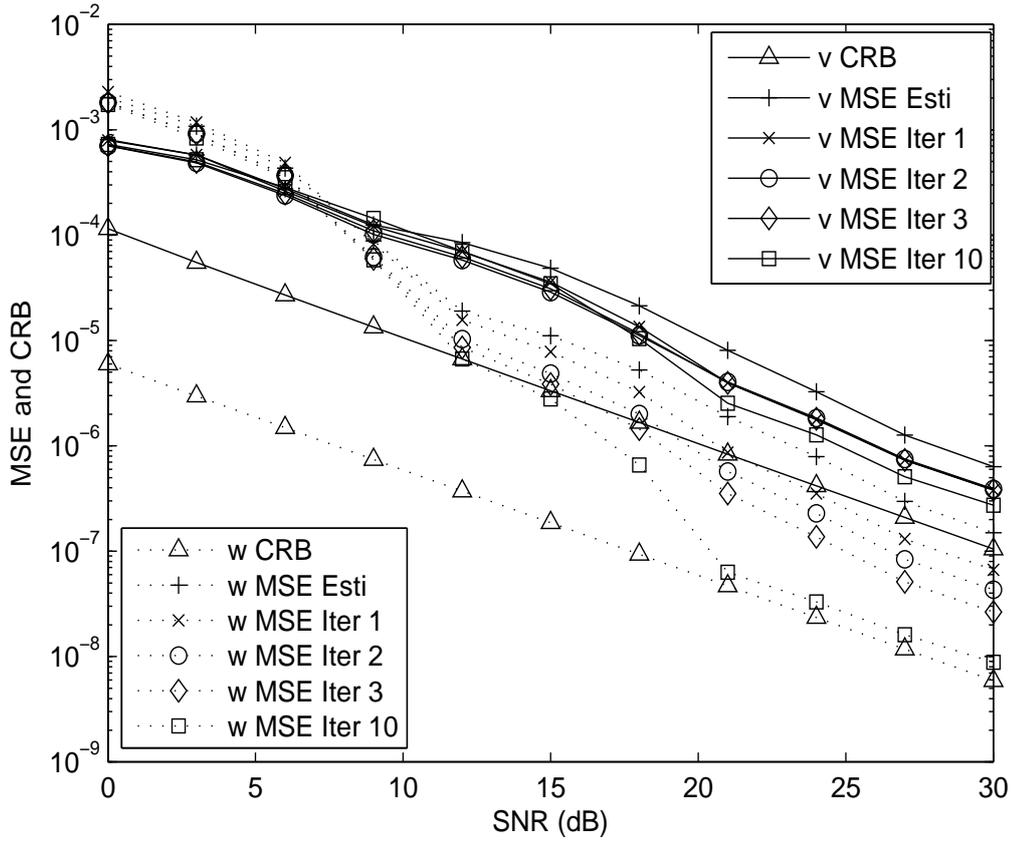


Fig. 3.4. CFO estimation MSE versus SNR: $N = 14$

3.5.3 Minimum N : A Special Case

In the last example, we set f_2 as 0.95 and 0.9401, such that the CFO between the two terminals are $v = 0.01$ and $v = 0.0001$, respectively. The minimum training length is chosen as $N = 2L + 3 = 9$. The CFO and channel estimation results can be obtained from (3.26), (3.27), and (3.28). As mentioned before, these cases do not require iterations because v is nearly zero. The estimation MSEs of w and channels, as well as their corresponding CRBs are shown in Fig. 3.6 and Fig. 3.7, respectively.

These figures show that the estimation accuracy is quite close to CRB. The reason is that the interference due to the pilots is negligible in this case as the CFO between the two source terminals is negligible. For $v = 0.01$, a relatively larger value, there exists an error floor for both CFO and channel estimation at the high SNR region. When v is as small as 0.0001, the best estimation performance can be achieved since the MSE attaches the CRB.

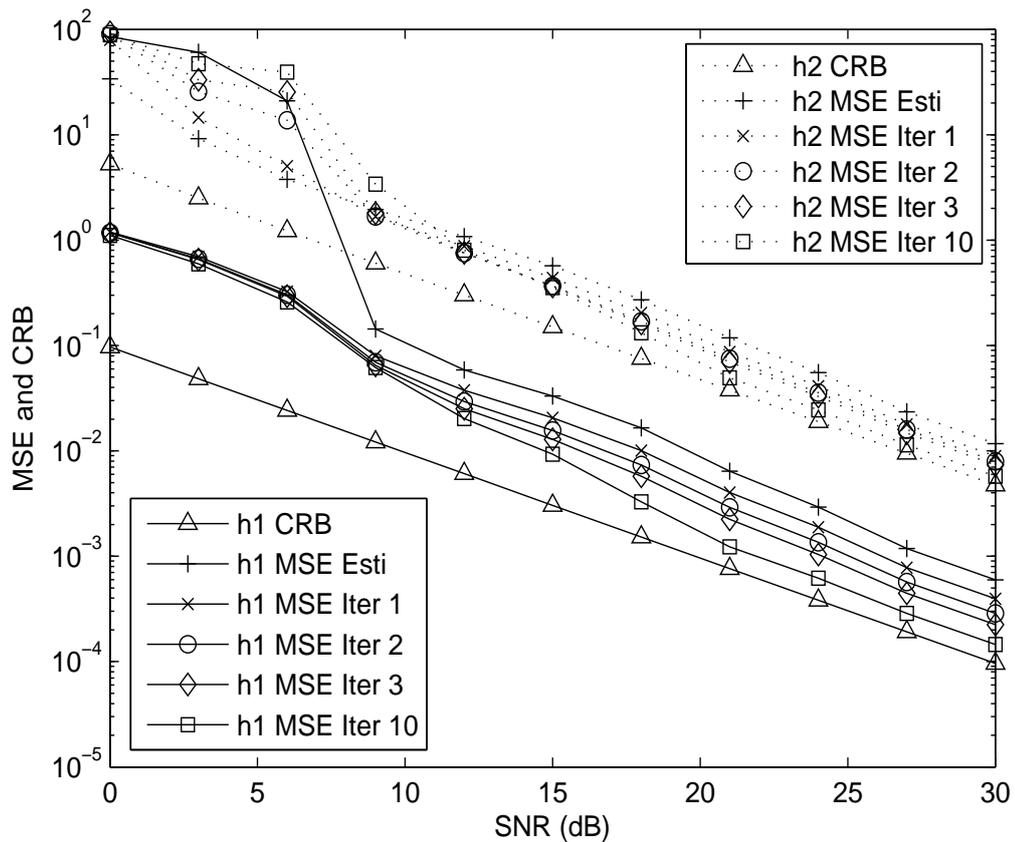


Fig. 3.5. Channel estimation MSE versus SNR: $N = 14$

3.6 Conclusions

In this chapter, superimposed pilot based CFO and channel estimation was investigated for CP-OFDM modulated TWRN. Three direct estimation algorithms as well as the iteration algorithm to improve the performance were developed. We also derived the analytical CRBs as the benchmark for the designed algorithms. With superimposed training, all the individual parameters can be estimated at all three nodes. From the simulations, it is found that although the iterative estimator improves the performance but gaps remain between the MSE and CRB, indicating room for further improving the performance.

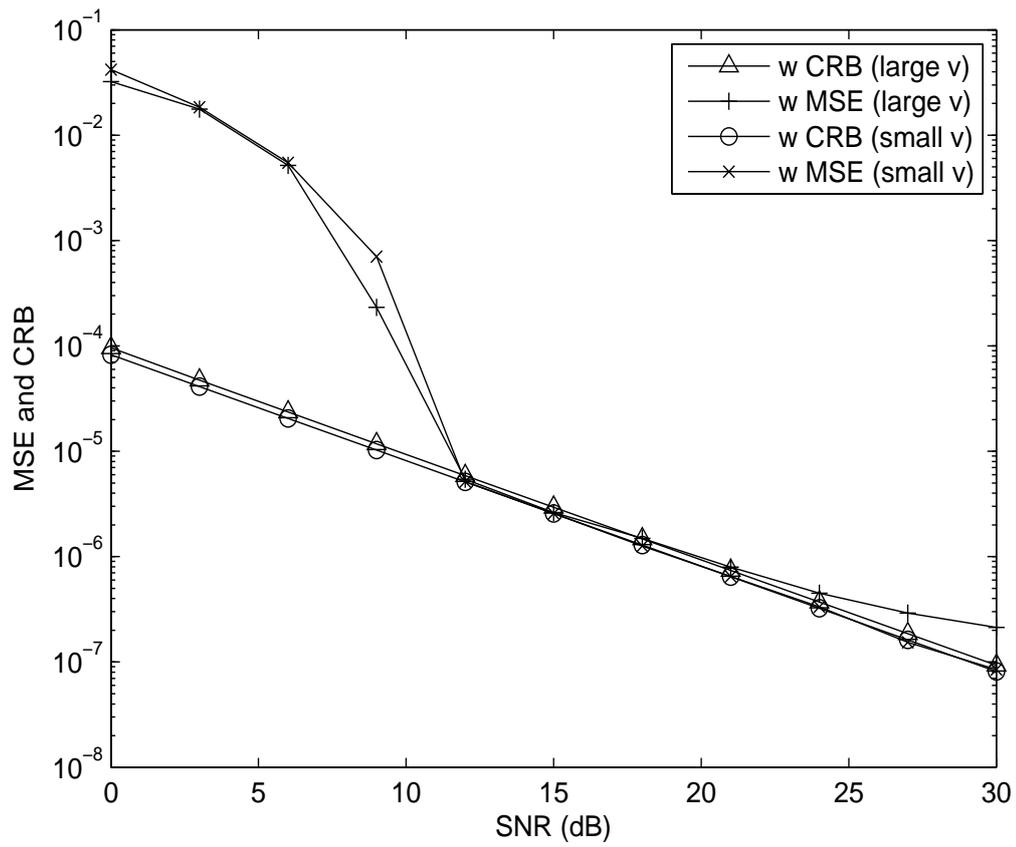


Fig. 3.6. CFO estimation MSE versus SNR: $N = 9$

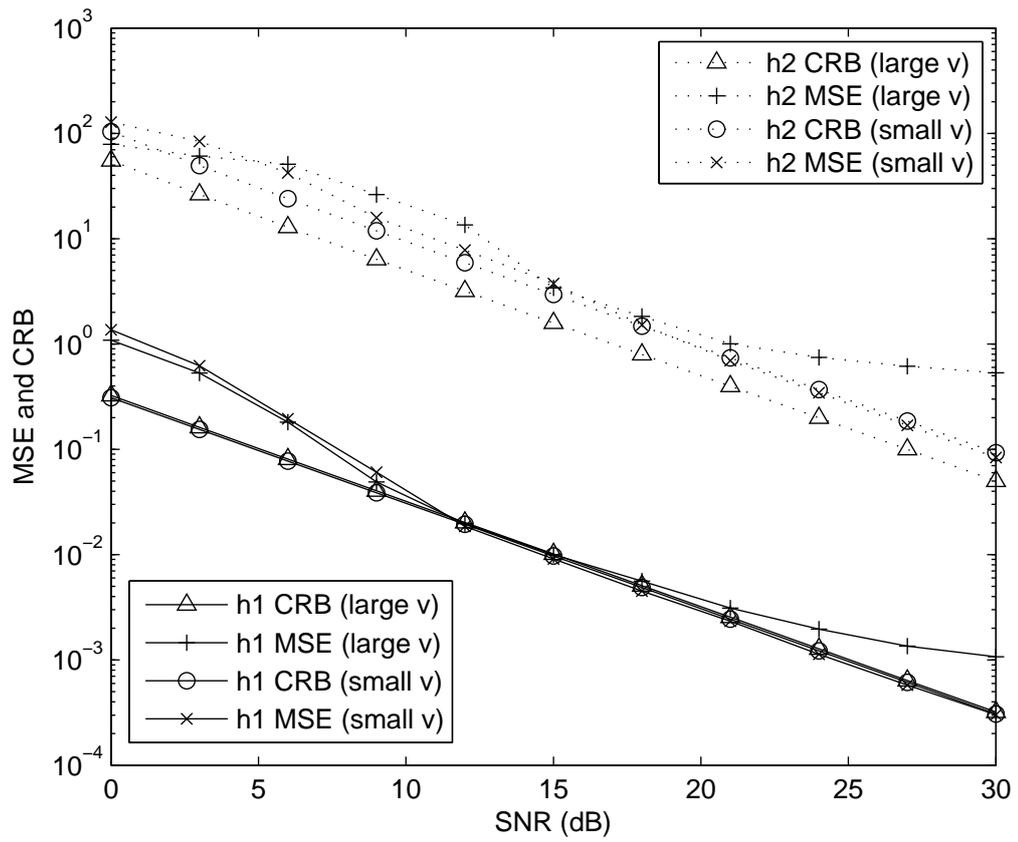


Fig. 3.7. Channel estimation MSE versus SNR: $N = 9$

Chapter 4

Time-Varying Channel Estimation for TWRN

In this chapter, channel estimation and training sequence design are considered for AF-based TWRNs in a time-varying channel. A new CE-BEM is proposed to represent mobile-to-mobile time-varying channels. To estimate such channels, a novel pilot symbol-aided transmission scheme is developed such that a linear approach can estimate the BEM coefficients of the convoluted channels. More essentially, two algorithms are designed to extract the BEM coefficients of the individual channels. The optimal training parameters, including the number of the pilot symbols, the placement of the pilot symbols, and the power allocation to the pilot symbols, are derived by minimizing the channel estimation MSE. The selections of the system parameters are thoroughly discussed in order to guide practical system design.

4.1 Introduction

Existing works [13], [14], [89], [90] about channel estimation in TWRN only consider the time-invariant environments. However, a TWRN is more susceptible to a time-varying channel because the relay and the two sources can all be mobile and the relative motion between any two nodes may double the Doppler spread [81]. This fact places additional demands on the estimation of time-varying channels, which are usually tracked by using periodic training signals, also known as PSAM [45]. To the best of our knowledge, time-varying channel estimation in TWRNs has not yet been reported. The need for such techniques motivates our current work.

Time-varying channels are typically represented in two ways: by using the Gauss-Markov model [33], which tracks channel variation through symbol-by-symbol updating, and by using the BEM [36], which decomposes the channel into the superposition of the time-varying basis functions

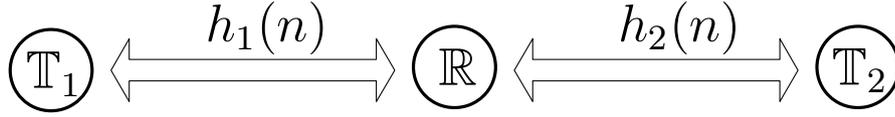


Fig. 4.1. A two-way relay network over time-varying flat-fading channels.

weighted by time-invariant coefficients. Using the Gauss-Markov channel model for a flat-fading time-varying channel, the channel estimation methods [49] and [83] minimize the MSE and maximize the data throughput criteria, respectively. The optimal training design for doubly-selective channels based on the BEM [43], [61], [62] adopted the same criteria. The equivalence between these two criteria was pointed out in [43].

To our best knowledge, currently no paper has compared the performance between the BEM and the Gauss-Markov model. Generally, the Gauss-Markov model requires sequential time-domain processing. Whereas the BEM is suitable for generating a block of channel samples [94]. For this reason, the BEM is chosen to develop our estimation algorithms.

In this chapter, we address the problem of channel estimation and training design for time-varying TWRN channels. We adopt the CE-BEM [42], which represents the time-varying channels by a finite set of parameters and Fourier bases. To handle the special features of the TWRN, we propose a new data frame structure, which enables the periodic reception of pilot symbols at the source nodes. The conventional TWRN transmission structure [13], [14] is a special case of our proposed structure. To reduce the estimation complexity, the interval between the pilot sending and pilot receiving is fixed, and the resultant BEM coefficients of the cascaded channels can be estimated linearly. The optimal training parameters, including the number of pilot symbols, the placement of pilot symbols, and the power of pilots symbols, are derived by minimizing the channel MSE criterion. Two algorithms are then designed to recover the BEM coefficients of individual channel coefficients. An iterative method is also used to refine the estimates. We also provide a thorough discussion of the system parameter selection and reveal many interesting results. Finally, simulation results are provided to corroborate our studies.

4.2 System Model

Consider a TWRN with two source nodes \mathbb{T}_1 , \mathbb{T}_2 and one relay node \mathbb{R} (Fig.4.1). Each node has only one half-duplex antenna. The baseband channel from \mathbb{T}_i , $i = 1, 2$ to \mathbb{R} is assumed to be time-varying flat-fading and is denoted by $h_i(n)$, where n is the discrete time index. Moreover, the channels are modeled as wide-sense stationary (WSS) zero mean complex Gaussian (ZMCG) random processes with variances $\sigma_{h_i}^2$. The channel from \mathbb{R} to \mathbb{T}_i is also denoted as $h_i(n)$. Perfect synchronization is assumed as in [90], [36], [42].

4.2.1 Time-varying Relay Channels

The channel statistics in a relay network depend on the mobility of the three nodes, i.e., the fixed nodes or the moving nodes [81]. Denote f_{d1} , f_{d2} and f_{dr} as the maximum Doppler shifts due to the motion of \mathbb{T}_1 , \mathbb{T}_2 , and \mathbb{R} , respectively. The discrete autocorrelation functions of $h_i(n)$'s can be represented as [44]

$$R_{h_i}(m) = \mathbb{E}\{h_i(n+m)h_i^*(n)\} = \sigma_{h_i}^2 J_0(2\pi f_{d_i} m T_s) J_0(2\pi f_{dr} m T_s), \quad i = 1, 2 \quad (4.1)$$

where $J_0(\cdot)$ is the zero-th order Bessel function of the first kind, and T_s is the symbol sampling time. The correlation function in (4.1) has been widely adopted to describe the mobile-to-mobile link, e.g., [81], [44]. If one node is fixed, i.e., if the corresponding Doppler shift is zero, then (4.1) reduces to the well-known Jakes model [63]. Meanwhile, (4.1) reveals that the power spectra of $h_1(n)$ and $h_2(n)$ span over the bandwidths $f_1 = f_{d1} + f_{dr}$ and $f_2 = f_{d2} + f_{dr}$, respectively, which indicates an increased Doppler effect for the mobile-to-mobile transmission.

The parsimonious finite-parameter BEM [36] can be applied to approximate the two time-varying channels, respectively, so that during any time interval of NT_s , $h_i(n)$'s can be modeled by

$$h_1(n) = \sum_{q=0}^{Q_1} \lambda_q w_1(q), \quad h_2(n) = \sum_{q=0}^{Q_2} \mu_q w_2(q), \quad 0 \leq n \leq N-1, \quad (4.2)$$

where λ_q 's and μ_q 's are the BEM coefficients that remain invariant within one interval of NT_s but will change in the next interval, while $w_i(q)$'s are the bases that capture the time variation and will remain the same for any interval. The number of the bases Q_i is a function of the channel bandwidth f_i and the interval length NT_s . Specific choices for $\{w_i(q)\}_{q=0}^{Q_i}$ include the polynomial [38], wavelet [40], discrete prolate spheroid sequence [39], and Fourier bases [43]. In this chapter, we choose the CE-BEM [42], a specific form of Fourier bases. Then (4.2) can be explicitly written as

$$h_1(n) = \sum_{q=0}^{Q_1} \lambda_q e^{j2\pi(q-Q_1/2)n/N}, \quad 0 \leq n \leq N-1, \quad (4.3a)$$

$$h_2(n) = \sum_{q=0}^{Q_2} \mu_q e^{j2\pi(q-Q_2/2)n/N}, \quad 0 \leq n \leq N-1. \quad (4.3b)$$

The CE-BEM (4.3) can be viewed as the Fourier series of the time-varying channels, and the number of bases Q_i should be at least $2\lceil f_i NT_s \rceil$ in order to provide sufficient degrees of freedom [43], [42]. To simplify the notation as well as the discussion, we assume $f_1 = f_2 = f_d$ and $Q_1 = Q_2 = Q$. Nonetheless, the extension to the general case is straightforward. We further denote

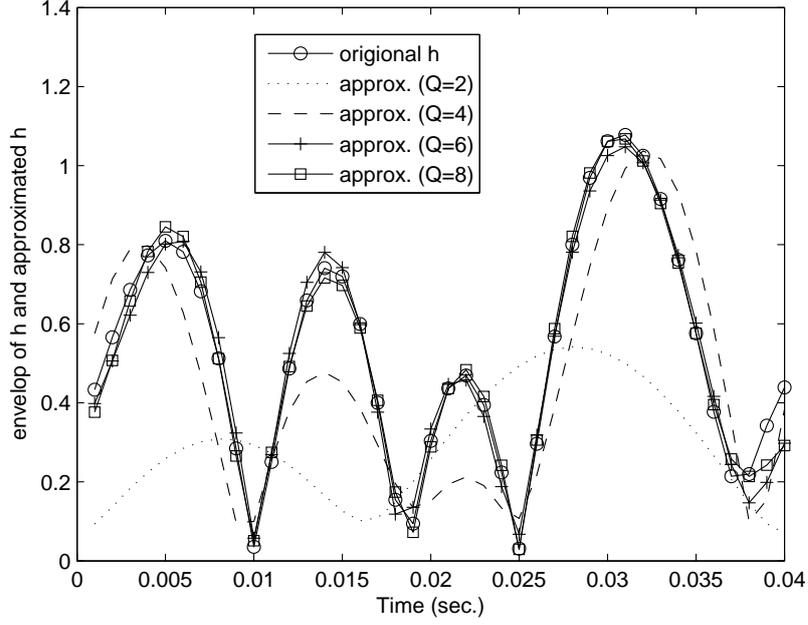


Fig. 4.2. Approximating the mobile-to-mobile channels with CE-BEM.

$\omega_q = 2\pi(q - Q/2)/N$ and define

$$\boldsymbol{\lambda} = [\lambda_0, \lambda_1, \dots, \lambda_Q]^T, \quad \boldsymbol{\mu} = [\mu_0, \mu_1, \dots, \mu_Q]^T$$

for subsequent use.

A brief example is given in Fig. 4.2, where the system parameters are taken as $f_{d1} = f_{d2} = f_{dr} = 40$ Hz, $T_s = 1$ ms, and $N = 40$. Fig. 4.2 reveals that the larger the Q is, the better the approximation will be. As pointed out in [43], [42], Q must be at least $2\lceil f_d T_s N \rceil = 4$ in order to keep the shape of the envelope, i.e., with the sufficient degrees of freedom. However, for $Q = 2$ the ambiguous estimation appears due to the lack of sufficient sampling degrees of freedom. Nonetheless, one can always use a larger Q for a better approximation.

4.2.2 Transmission Strategy

To enable the use of PSAM in our TWRN, we propose a new transmission strategy over one interval NT_s , as depicted in Fig. 4.3. Let \mathcal{D}_t and \mathcal{T}_t be the time index sets for the transmitted information symbols and the pilot symbols from \mathbb{T}_i , $i = 1, 2$, respectively. Moreover, let \mathcal{D}_r and \mathcal{T}_r be the time index sets for the received information symbols and pilot symbols at \mathbb{T}_i , respectively. These four sets are disjoint with the property that $\mathcal{D}_t \cup \mathcal{T}_t \cup \mathcal{D}_r \cup \mathcal{T}_r = \{0, 1, \dots, N - 1\}$. Let us define the cardinality of the sets as $|\mathcal{D}_t| = |\mathcal{D}_r| = D$ and $|\mathcal{T}_t| = |\mathcal{T}_r| = T$. Then, $N = 2(D + T)$ is an even

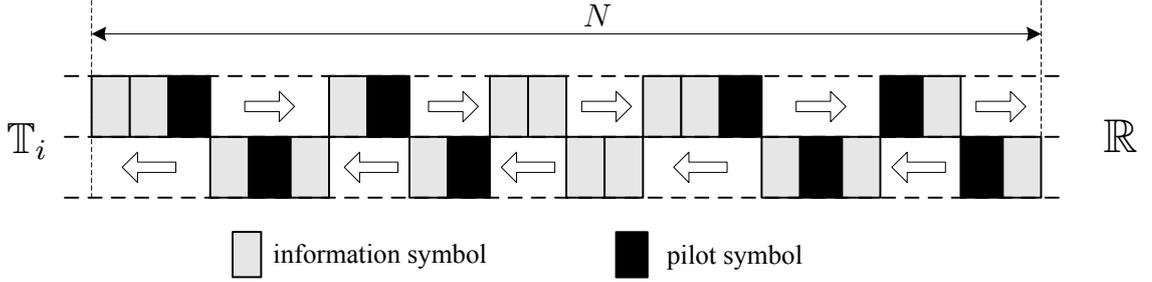


Fig. 4.3. Proposed transmission strategy for two-way relay network with time-varying channel.

integer.

We assume that the relay node \mathbb{R} forwards its received symbols on time slot $g(n)$ to both \mathbb{T}_i on time slot n ; i.e.,

$$\mathcal{D}_t \cup \mathcal{T}_t = \{g(n) | n \in \mathcal{D}_r \cup \mathcal{T}_r\}. \quad (4.4)$$

By defining the one-to-one mapping function $g(n)$, we implicitly allow for the symbols' order to be changed when \mathbb{R} forwards them back to \mathbb{T}_i . Hence, it is also possible to optimize $g(n)$ according to different criteria, i.e., data detection MSE, bit-error-rate (BER), throughput, and others.¹ A detailed discussion is beyond the scope of this chapter and will be left to future research. Note that $0 \leq g(n) < n$ is required because \mathbb{R} can only forward a symbol after receiving it. Interestingly, the conventional data transmission in a TWRN [13], [14] [90] becomes a special case of our proposed scheme if $g(n) = n - N/2$ is selected.

A special yet important case involves evenly dividing NT_s intervals into several sub-blocks, as shown in Fig. 4.4. This case corresponds to setting $g(n) = n - M$, where M divides $N/2$, and will be separately discussed later. The decision on whether to adopt the general scheme (Fig. 4.3) or the sub-block-based scheme (Fig. 4.4) depends on the synchronization requirement in practical scenarios and other design issues.

Denote the symbols sent from \mathbb{T}_i as $s_i(n)$, $n \in \mathcal{D}_t \cup \mathcal{T}_t$, of which the average power for the information symbols is P_i ; i.e., $\mathbb{E}\{|s_i(n)|^2\} = P_i, \forall n \in \mathcal{D}_t$, while the total training power² is $P_{i,t}$; i.e., $\sum_{n \in \mathcal{T}_t} |s_i(n)|^2 = P_{i,t}$. With perfect synchronization, \mathbb{R} receives

$$r(n) = h_1(n)s_1(n) + h_2(n)s_2(n) + w_r(n), \quad n \in \mathcal{D}_t \cup \mathcal{T}_t, \quad (4.5)$$

where $w_r(n)$ is the circularly symmetric complex Gaussian (CSCG) noise with the variance σ_r^2 . If

¹The dual problem of optimally re-ordering the subcarrier indices in a frequency-selective environment has been studied in [19].

²We should not consider the average power constraints for training because otherwise, the training length is trivially preferred to be as large as possible.

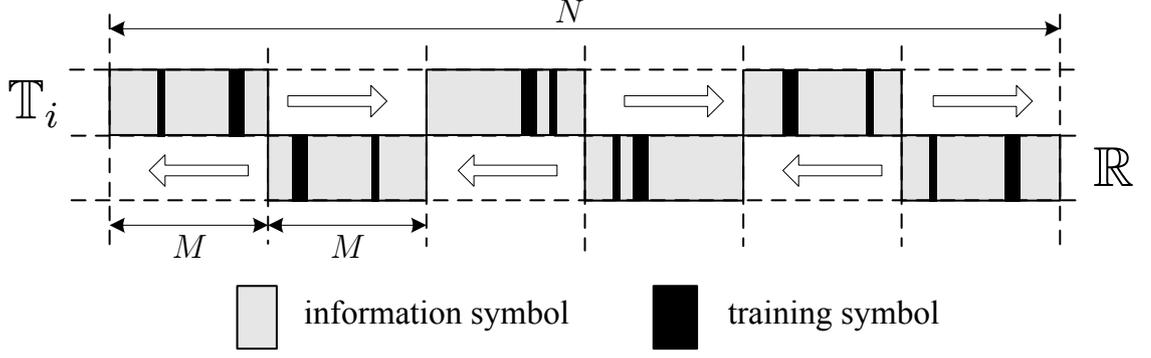


Fig. 4.4. Sub-block based transmission strategy.

the average transmit power of \mathbb{R} is P_r , then $r(n)$ will be scaled by

$$\alpha(n) = \begin{cases} \sqrt{\frac{P_r}{\sigma_{h_1}^2 P_1 + \sigma_{h_2}^2 P_2 + \sigma_r^2}} & n \in \mathcal{D}_r \\ \sqrt{\frac{P_r}{\sigma_{h_1}^2 P_{1,t}/T + \sigma_{h_2}^2 P_{2,t}/T + \sigma_r^2}} & n \in \mathcal{T}_r \end{cases} \quad (4.6)$$

before it is forwarded to \mathbb{T}_i 's to keep the power constraint.

More practical considerations should include the processing delay at \mathbb{R} as well as the path-delay between \mathbb{T}_1 and \mathbb{T}_2 . These considerations require only slightly changing the channel from $h_i(n)$ to $h_i(n + \Delta n)$, and the remaining discussion holds the same.

4.2.3 On Channel Estimation

Due to symmetry, we present only the estimation procedure at \mathbb{T}_1 , and the received signal is

$$\begin{aligned} y(n) &= \alpha(n) h_1(n) r(g(n)) + w_1(n) \\ &= \alpha(n) \underbrace{h_1(n) h_1(g(n))}_{b_1(n)} s_1(g(n)) + \alpha(n) \underbrace{h_1(n) h_2(g(n))}_{b_2(n)} s_2(g(n)) \\ &\quad + \underbrace{\alpha(n) h_1(n) w_r(g(n)) + w_1(n)}_{w(n)}, \quad n \in \mathcal{D}_r \cup \mathcal{T}_r, \end{aligned} \quad (4.7)$$

where $w_1(n)$ is the CSCG noise at \mathbb{T}_1 with variance σ_1^2 ; $w(n)$ denotes the overall noise; and $b_i(n)$, $i = 1, 2$ can be treated as the equivalent time-varying channel of $\mathbb{T}_i \rightarrow \mathbb{R} \rightarrow \mathbb{T}_1$. Obviously, if $b_i(n)$'s are known at \mathbb{T}_1 , the self-signal component $s_1(g(n))$ can be subtracted from $y(n)$ in order to detect the desired information $s_2(g(n))$.

To gain more insight into the time-varying channels, we apply (4.3) and rewrite $b_i(n)$'s as

$$b_1(n) = \sum_{p=0}^Q \sum_{q=0}^Q \lambda_p \lambda_q e^{j(\omega_p n + \omega_q g(n))}, \quad b_2(n) = \sum_{p=0}^Q \sum_{q=0}^Q \lambda_p \mu_q e^{j(\omega_p n + \omega_q g(n))}, \quad n \in \mathcal{D}_r \cup \mathcal{T}_r. \quad (4.8)$$

The new expression (4.8) indicates that in order to obtain $b_i(n)$, $0 \leq n \leq N - 1$, one needs to know either $2(Q + 1)$ parameters $\lambda_p, \mu_p, p = 0, \dots, Q$ or $2(Q + 1)^2$ parameters $\lambda_p \lambda_q, \lambda_p \mu_q, p, q = 0, \dots, Q$. For a general mapping function $g(n)$, the former approach requires a non-linear search, while the latter approach, though could be implemented from linear approach, possesses large redundancy in the number of estimated variables.

To facilitate the channel estimation, we propose to use

$$g(n) = n - M, \quad (4.9)$$

for $n \in \mathcal{T}_r$, while $g(n)$ for information transmission $n \in \mathcal{D}_r$ could still be designed from a certain optimization criterion. The condition (4.9) says that \mathbb{R} retransmits each received pilot symbol with a delay of M -symbol interval, and this interval is common for all pilot symbols.

With (4.9), the received pilot symbols at \mathbb{T}_1 can be further expressed as

$$y(n) = \alpha \sum_{m=0}^{2Q} \left(\underbrace{\sum_{q=0}^m \lambda_{m-q} \lambda_q e^{-j\omega_q M}}_{x_1(m)} \right) e^{j\theta_m n} s_1(n - M) + \alpha \sum_{m=0}^{2Q} \left(\underbrace{\sum_{q=0}^m \lambda_{m-q} \mu_q e^{-j\omega_q M}}_{x_2(m)} \right) e^{j\theta_m n} s_2(n - M) + w(n), \quad n \in \mathcal{T}_r, \quad (4.10)$$

where $\theta_m = 2\pi(m - Q)/N$, $x_i(m)$ are defined as the corresponding items, and the index n in $\alpha(n)$ is omitted for brevity. When deriving (4.10), we use the property that $\omega_p + \omega_q = \omega_{p'} + \omega_{q'}$ whenever $p + q = p' + q'$. If the sub-block transmission in Fig. 4.4 is applied, then (4.10) is also applicable for the received information symbols $n \in \mathcal{D}_r$.

Interestingly, we may treat $x_i(m)$'s as the equivalent BEM coefficients with $2Q + 1$ carriers $e^{j\theta_m n}$ for the equivalent time-varying channel $b_i(n)$, $n \in \mathcal{T}_r$. The equivalent BEM sequence $x_1(m)$ is the convolution between the original BEM λ_p and $e^{-j\omega_q M} \lambda_q$, while $x_2(m)$ is the convolution between λ_p and $e^{-j\omega_p M} \mu_q$.

Define

$$\mathbf{x}_i = [x_i(0), x_i(1), \dots, x_i(2Q)]^T, \quad \mathbf{\Gamma} = \text{diag}\{e^{-j\omega_0 M}, e^{-j\omega_1 M}, e^{-j\omega_Q M}\}$$

and define \mathbf{A} as the $(2Q + 1) \times (Q + 1)$ Toeplitz matrix with the first column $[\boldsymbol{\lambda}^T, \mathbf{0}_{1 \times Q}]^T$. We can explicitly express the convolutions as

$$\mathbf{x}_1 = \boldsymbol{\lambda} \otimes (\boldsymbol{\Gamma}\boldsymbol{\lambda}) = \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\lambda}, \quad \text{and} \quad \mathbf{x}_2 = \boldsymbol{\lambda} \otimes (\boldsymbol{\Gamma}\boldsymbol{\mu}) = \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\mu}. \quad (4.11)$$

Based on (4.10), we may estimate the equivalent BEM coefficient $x_i(m)$ (with $4Q+2$ unknowns) and recover the original BEM λ_q, μ_q (with $2Q + 2$ unknowns). Then, the equivalent time-varying channels $b_i(n)$, $n \in \mathcal{D}_r$ can be obtained from (4.8).

4.3 Channel Estimation and Training Sequence Design

Let us specify the indices in \mathcal{T}_r as $n_0 < n_1 \dots < n_{T-1}$, and define

$$\begin{aligned} \mathbf{y}_t &= [y(n_0), y(n_1), \dots, y(n_{T-1})]^T, & \mathbf{w}_t &= [w(n_0), w(n_1), \dots, w(n_{T-1})]^T, \\ \mathbf{t}_i &= [s_i(n_0 - M), s_i(n_1 - M), \dots, s_i(n_{T-1} - M)]^T, & \mathbf{T}_i &= \text{diag}\{\mathbf{t}_i\}, \quad i = 1, 2, \end{aligned}$$

where \mathbf{t}_i contains all the pilot symbols from \mathbb{T}_i . For notational simplicity, the m -th entry of \mathbf{t}_i is also denoted by $t_i(m)$, $m = 0, \dots, T - 1$.

With the aid of (4.10), we can express \mathbf{y}_t in matrix form as

$$\mathbf{y}_t = \alpha \mathbf{T}_1 \mathbf{A} \mathbf{x}_1 + \alpha \mathbf{T}_2 \mathbf{A} \mathbf{x}_2 + \mathbf{w}_t, \quad (4.12)$$

where \mathbf{A} is the $T \times (2Q + 1)$ matrix

$$\mathbf{A} = \begin{bmatrix} e^{j\theta_0 n_0} & e^{j\theta_1 n_0} & \dots & e^{j\theta_{2Q} n_0} \\ e^{j\theta_0 n_1} & e^{j\theta_1 n_1} & \dots & e^{j\theta_{2Q} n_1} \\ \vdots & \vdots & \dots & \vdots \\ e^{j\theta_0 n_{T-1}} & e^{j\theta_1 n_{T-1}} & \dots & e^{j\theta_{2Q} n_{T-1}} \end{bmatrix}. \quad (4.13)$$

4.3.1 Channel Estimation Algorithm

When $T \geq 4Q + 2$, there are sufficient observations to estimate all the unknown $x_i(m)$'s. In this case, one could choose a linear estimator, e.g., the least square (LS) or the linear minimum mean square error (LMMSE) estimator. For example in [42], [61], [62], the authors assumed that the knowledge of the statistics of the BEM coefficients were available in order to derive the LMMSE estimator. Moreover, a closed-form training design requires the assumption that the BEM coefficients are uncorrelated among themselves [42], [61], [62]. Although the same assumption can be invoked here, we would rather choose the LS estimator in order to embrace more practical scenarios

where the statistics of the BEM coefficients are not available. This choice is further justified because the LS estimator performs similarly to the LMMSE estimator at a relatively high signal-to-noise ratio (SNR).

Let us define

$$\mathbf{T} = [\mathbf{T}_1 \mathbf{A}, \mathbf{T}_2 \mathbf{A}], \quad \mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T.$$

The LS estimator of \mathbf{x} is expressed as

$$\hat{\mathbf{x}} = \frac{1}{\alpha} \mathbf{T}^\dagger \mathbf{y} = \frac{1}{\alpha} (\mathbf{T}^H \mathbf{T})^{-1} \mathbf{T}^H \mathbf{y}, \quad (4.14)$$

with the error covariance matrix given by

$$\mathbf{W} = \mathbf{T}^\dagger \begin{bmatrix} \sigma_r^2 |h_1(n_0)|^2 + \frac{\sigma_1^2}{\alpha^2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r^2 |h_1(n_{T-1})|^2 + \frac{\sigma_1^2}{\alpha^2} \end{bmatrix} (\mathbf{T}^\dagger)^H. \quad (4.15)$$

4.3.2 Optimal Training Design

The channel estimation MSE is defined as $\text{tr}(\mathbf{W})$ and is related to the instant CSI. In this case, we propose to minimize the average MSE (AMSE), which is defined as

$$\text{AMSE} = \mathbb{E}_h \{ \text{tr}(\mathbf{W}) \} = \left(\sigma_{h_1}^2 \sigma_r^2 + \frac{\sigma_1^2}{\alpha^2} \right) \text{tr}((\mathbf{T}^H \mathbf{T})^{-1}), \quad (4.16)$$

where the property $J_0(0) = 1$ is used. We further partition $(\mathbf{T}^H \mathbf{T})^{-1}$ as

$$(\mathbf{T}^H \mathbf{T})^{-1} = \begin{bmatrix} \mathbf{A}^H \mathbf{T}_1^H \mathbf{T}_1 \mathbf{A} & \mathbf{A}^H \mathbf{T}_1^H \mathbf{T}_2 \mathbf{A} \\ \mathbf{A}^H \mathbf{T}_2^H \mathbf{T}_1 \mathbf{A} & \mathbf{A}^H \mathbf{T}_2^H \mathbf{T}_2 \mathbf{A} \end{bmatrix}^{-1}. \quad (4.17)$$

The optimal training design amounts to selecting the number of the pilot symbols, their placement, and the power allocation for each pilot by minimizing the AMSE. The optimization problem is then formulated as

$$\begin{aligned} \text{(P1):} \quad & \min_{t_1, t_2, T_r} \left(\sigma_{h_1}^2 \sigma_r^2 + \frac{\sigma_1^2}{\alpha^2} \right) \text{tr}((\mathbf{T}^H \mathbf{T})^{-1}) \\ & \text{s.t.} \quad \sum_{m=0}^{T-1} |t_i(m)|^2 \leq P_{i,t}, \quad i = 1, 2. \end{aligned} \quad (4.18)$$

Since α is related only to T , we can first solve, for a given T , the following problem:

$$(P2): \quad \min_{\substack{\mathbf{t}_1, \mathbf{t}_2, \\ n_i: 0 \leq i \leq T}} \quad \text{tr}((\mathbf{T}^H \mathbf{T})^{-1}) \quad (4.19)$$

$$\text{s.t.} \quad \sum_{m=0}^{T-1} |t_i(m)|^2 \leq P_{i,t}, \quad i = 1, 2.$$

From [66], we know that

$$\text{tr}((\mathbf{T}^H \mathbf{T})^{-1}) \geq \sum_{i=0}^{4Q+1} \frac{1}{[\mathbf{T}^H \mathbf{T}]_{i,i}} = \sum_{i=1}^2 \frac{2Q+1}{\sum_{m=0}^{T-1} |t_i(m)|^2}, \quad (4.20)$$

where $[\mathbf{T}^H \mathbf{T}]_{i,i}$ is the i th diagonal elements of $\mathbf{T}^H \mathbf{T}$, and the equality holds when $\mathbf{T}^H \mathbf{T}$ is a diagonal matrix. However, this inequality does not directly show that the diagonal must hold for the optimal $\mathbf{T}^H \mathbf{T}$. Let us first formulate a new optimization problem:

$$(P3): \quad \min_{\mathbf{t}_1, \mathbf{t}_2} \quad \sum_{i=1}^2 \frac{2Q+1}{\sum_{m=0}^{T-1} |t_i(m)|^2} \quad (4.21)$$

$$\text{s.t.} \quad \sum_{m=0}^{T-1} |t_i(m)|^2 \leq P_{i,t}, \quad i = 1, 2.$$

Obviously, the optimal objective of (P3) serves as a lower bound for (P2). Since (P3) is a simple convex optimization, any training sequence satisfying $\sum_{m=0}^{T-1} |t_i(m)|^2 = P_{i,t}$ is optimal. Hence, if we can find \mathbf{t}_i 's that satisfy the equality constraints and make $\mathbf{T}^H \mathbf{T}$ diagonal, then these \mathbf{t}_i 's must also be the optimal solutions for problem (P2). In other words, the sufficient conditions for the optimal solutions to (P2) are

$$\mathbf{A}^H \mathbf{T}_i^H \mathbf{T}_i \mathbf{A} = P_{i,t} \mathbf{I}_{2Q+1}, \quad i = 1, 2, \quad (4.22a)$$

$$\mathbf{A}^H \mathbf{T}_1^H \mathbf{T}_2 \mathbf{A} = \mathbf{0}_{2Q+1}. \quad (4.22b)$$

Observing the Vandermonde structure of \mathbf{A} and the structure of θ_m , we know that if the pilot symbols are equi-powered and equi-spaced over $\{0, \dots, N-1\}$, then (4.22a) is satisfied; i.e.,

$$C1): \quad |t_i(m)|^2 = P_{i,t}/T, \quad \forall m = 0, 1, \dots, T-1, \quad i = 1, 2,$$

$$C2): \quad n_m = mL + l_0, \quad \forall l_0 \in [M, L-1], \quad \text{and } L = N/T \text{ is an integer,}$$

where we include the consideration that $n_0 \geq M$ in C2).³ Combined with C1) and C2), the following

³ $n_0 = l_0$ denotes the index of the first symbol sent by \mathbb{R} . From the adopted $g(n)$, $n_0 = l_0 \geq M$ is required.

condition can guarantee (4.22b):

$$\sum_{m=0}^{T-1} t_2^*(m)t_1(m)e^{-j\theta_u n_m} e^{-j\theta_v n_m} = 0, \quad \forall u, v = 0, 1, \dots, 2Q$$

which can be simplified as

$$\text{C3):} \quad \sum_{m=0}^{T-1} t_2^*(m)t_1(m)e^{j2\pi m k/T} = 0, \quad \forall k = -2Q, -2Q+1, \dots, 2Q.$$

One example of pilot sequences that satisfy conditions C1)–C3) is

$$\mathbf{t}_1 = \sqrt{\frac{P_{1,t}}{T}} [+1, +1, +1, \dots, +1, +1]^T, \quad (4.23a)$$

$$\mathbf{t}_2 = \sqrt{\frac{P_{2,t}}{T}} [1, e^{j2\pi v/T}, \dots, e^{j2\pi(T-1)v/T}]^T, \quad \forall v = 2Q+1, \dots, T-2Q-1. \quad (4.23b)$$

The minimum $\text{tr}((\mathbf{T}^H \mathbf{T})^{-1})$ is then $(2Q+1)(1/P_{1,t} + 1/P_{2,t})$ and does not depend on T . Hence, the optimal value of T should be independently obtained from

$$T = \arg \min_T \left(\sigma_{h_1}^2 \sigma_r^2 + \frac{\sigma_1^2}{\alpha^2} \right) = \arg \max_T \frac{P_r}{\sigma_{h_1}^2 P_{1,t}/T + \sigma_{h_2}^2 P_{2,t}/T + \sigma_r^2}. \quad (4.24)$$

The objective function (4.24) is an increasing function of T , so the optimal T should be made as large as possible. Note that this result is different from the conventional training design in point-to-point systems, where the channel estimation MSE is related only to the total training power but not to the training length.

However, increasing T would reduce the efficiency of the data transmission and, consequently, the system throughput. Besides, the constant $\sigma_{h_1}^2 \sigma_r^2$ will dominate the summation from $\left(\sigma_{h_1}^2 \sigma_r^2 + \frac{\sigma_1^2}{\alpha^2} \right)$ when T is greater than a certain threshold. Therefore, increasing T beyond a certain value cannot improve the channel estimation MSE, but the throughput will be linearly decreased. A more meaningful design of T can be obtained by maximizing the transmission throughput criterion [42], [61], [62]. In this chapter we focus only on introducing the new channel estimation strategy in TWRN, and we simply consider achieving the minimum amount of training as our optimization goal.

The selection of the minimum possible T depends on many factors and will be discussed in the next subsection. When $T = 4Q + 2$ is allowed, the optimal pilot schemes become more specific:

$$\mathbf{t}_1 = \sqrt{\frac{P_{1,t}}{4Q+2}} [+1, +1, +1, +1, \dots, +1, +1]^T,$$

$$\mathbf{t}_2 = \sqrt{\frac{P_{2,t}}{4Q+2}} [+1, -1, +1, -1, \dots, +1, -1]^T,$$

and the corresponding minimum AMSE is

$$\text{AMSE} = \left((2Q + 1)\sigma_{h_1}^2\sigma_r^2 + \frac{(\sigma_{h_1}^2 P_{1,t} + \sigma_{h_2}^2 P_{2,t} + 2\sigma_r^2(2Q + 1))\sigma_1^2}{2P_r} \right) \left(\frac{1}{P_{1,t}} + \frac{1}{P_{2,t}} \right). \quad (4.25)$$

Importantly, it can be verified that the designed optimal pilot sequences for channel estimation at \mathbb{T}_1 are also optimal at \mathbb{T}_2 . Hence, simultaneous optimal channel estimation can be achieved at both source nodes.

4.3.3 Parameter Selection

Observing C2), we know the following: (i) The pilot spacing L should at least be $M + 1$; (ii) To transmit non-zero information symbols in one NT_s interval, we need $D = \frac{LT-2T}{2} \geq 1$, so the spacing L should be at least 3;⁴ (iii) Since N must be even, either T or L should be an even integer.

The above discussion suggests the guidelines for choosing T , i.e., select the smallest integer that is greater than or equal to $4Q + 2$, divides N , and satisfies $N/T \geq 3$.

Since $T \geq 4Q + 2$ pilot symbols are needed to provide sufficient observations, and since $Q \geq 1$ for a time-varying channel, the TWRN requires that pilot symbols to be transmitted back-and-forth at least 6 times. Therefore, the conventional two-way frame transmission structure [9], [8], [14], i.e., sending and receiving the continuous data sequence only once, obviously does not work in time-varying channels. We thus re-emphasize the novelty of the proposed PSAM scheme in Fig. 4.3.

For the sub-block based frame structure in Fig. 4.4, the receiving equi-spaced pilot at \mathbb{T}_i is possible only if each sub-block contains only one pilot symbol at the same position of each sub-block.

4.3.4 Doppler Shift and Transmission Efficiency

From $N = LT \geq L(4Q + 2)$ and $Q \geq 2\lceil f_d NT_s \rceil$, a successful channel estimation requires

$$\frac{N - 2L}{8LN} \geq f_d T_s. \quad (4.26)$$

To cope with more Doppler shift, the left-hand side (LHS) of (4.26) should be as large as possible.

From $L \geq 3$, there is

$$\frac{N - 2L}{8LN} \leq \frac{N - 6}{24N} < \frac{1}{24}. \quad (4.27)$$

⁴This conclusion is also seen from the fact that if $L = 2$, then the only choice for M is 1, in which case \mathbb{T}_i alternatively transmits and receives pilot symbols while no information can be sent.

Then the proposed strategy can handle the time-varying channels with

$$f_d T_s < \frac{1}{24} = 0.0416. \quad (4.28)$$

This normalized Doppler shift, fortunately, lies in the acceptable range of most studies [43], [83], [62].

Moreover, the training requirement $4Q + 2 \leq T = N/L \leq N/3$ implies that the transmission efficiency has the range

$$\eta = \frac{N/2 - T}{N/2} \leq 1 - 16f_d T_s - \frac{4}{N} < 1 - 16f_d T_s, \quad (4.29)$$

and

$$\eta \geq \frac{N/2 - N/3}{N/2} = 1/3. \quad (4.30)$$

Therefore, the higher the $f_d T_s$ is, the less the transmission efficiency will be. This result is intuitively satisfying.

4.4 Recovering the Original BEM Coefficients

After estimating \mathbf{x}_i 's, $i = 1, 2$, we need to obtain the original BEM coefficients λ_q and μ_q in order to build the time-varying channel $b_i(n)$, $n \in \mathcal{D}_r$. This is the key difference of TWRN from OWRN, as pointed out in [13], [14]. Retrieving λ_q and μ_q from \mathbf{x}_i generally requires solving multivariate nonlinear equations. In the following, we propose two simple methods and describe them under a noise-free scenario.

4.4.1 Time-Domain Approach

Because of the structure of $x_1(m)$, a straightforward way is to estimate λ_q sequentially. Specifically, we first estimate λ_0 from

$$\lambda_0 = I_s (x_1(0)e^{j\omega_0 M})^{1/2}, \quad (4.31)$$

where $I_s = \pm 1$ denotes the sign uncertainty. By choosing any of the positive or negative signs in (4.31), λ_1 can be computed from

$$\lambda_1 = \frac{x_1(1)}{\lambda_0 e^{-j\omega_0 M} + \lambda_0 e^{-j\omega_1 M}}. \quad (4.32)$$

We then sequentially compute λ_q from $x_1(q)$ with the previous estimates of $\lambda_0, \dots, \lambda_{q-1}$. The detailed steps are straightforward and are omitted here.

The above process uses only the first $Q + 1$ entries in \mathbf{x}_1 and cannot provide satisfactory precision. Nevertheless, with this initial estimation, we can apply the gradient decent process [67] to improve the estimation accuracy. The objective is to minimize the distance between \mathbf{x} and $\mathbf{\Lambda}\mathbf{\Gamma}\boldsymbol{\lambda}$; i.e., $\zeta = \|\mathbf{x} - \mathbf{\Lambda}\mathbf{\Gamma}\boldsymbol{\lambda}\|^2$. Then, $\boldsymbol{\lambda}$ can be updated according to

$$\boldsymbol{\lambda}^{(i+1)} = \boldsymbol{\lambda}^{(i)} - \epsilon \left. \frac{\partial \zeta}{\partial \boldsymbol{\lambda}^*} \right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(i)}}, \quad (4.33)$$

where ϵ is the step size. See Appendix D for a brief illustration of the gradient decent method with complex variables. The partial differential in (4.33) can be explicitly expressed as

$$\frac{\partial \zeta}{\partial \boldsymbol{\lambda}^*} = -(\mathbf{\Lambda}\mathbf{\Gamma} + \mathbf{\Omega})^H (\mathbf{x} - \mathbf{\Lambda}\mathbf{\Gamma}\boldsymbol{\lambda}), \quad (4.34)$$

where $\mathbf{\Omega}$ is a $(2Q + 1) \times (Q + 1)$ Toeplitz matrix with the first column $[(\mathbf{\Gamma}\boldsymbol{\lambda})^T, \mathbf{0}_{1 \times Q}]^T$.

Once $\boldsymbol{\lambda}$ is obtained, $\boldsymbol{\mu}$ can be found from

$$\boldsymbol{\mu} = \mathbf{\Gamma}^H \mathbf{\Lambda}^\dagger \mathbf{x}_2. \quad (4.35)$$

Note that there exists a simultaneous sign ambiguity (SSA) in the estimated results due to step (4.31); i.e., either $\{\boldsymbol{\lambda}, \boldsymbol{\mu}\}$ or $\{-\boldsymbol{\lambda}, -\boldsymbol{\mu}\}$ is found. Nonetheless, the SSA does not affect the data detection when we reconstruct $b_i(n)$'s. A similar observation is also made in [13], [14].

4.4.2 Frequency-Domain Approach

Let $\tilde{\boldsymbol{\lambda}}$ be the Z -point discrete Fourier transform (DFT) of $\boldsymbol{\lambda}$ with $Z \geq Q + 1$, whose m th entry is defined as

$$\tilde{\lambda}_m = \sum_{q=0}^Q \lambda_q e^{-j2\pi qm/Z}, \quad m = 0, \dots, Z - 1. \quad (4.36)$$

On the other side, the m th element of the Z -point DFT of $\mathbf{\Gamma}\boldsymbol{\lambda}$ is

$$\xi_m = \sum_{q=0}^Q \lambda_q e^{-j2\pi \frac{(q-Q/2)M}{N}} e^{-j2\pi qm/Z} = e^{j\frac{\pi Q M}{N}} \sum_{q=0}^Q \lambda_q e^{-j\frac{2\pi q(\frac{ZM}{Z} + m)}{Z}}. \quad (4.37)$$

If $R \triangleq \frac{ZM}{N}$ is an integer, then (4.37) becomes $e^{j\frac{\pi Q M}{N}} \tilde{\lambda}_{\langle m+R \rangle_Z}$, where $\langle \cdot \rangle_Z$ denotes the modulo- Z operation. Then the m th element of the Z -point DFT of $\mathbf{x}_1(m)$ is

$$\tilde{x}_1(m) = \xi_m \tilde{\lambda}_m = e^{j\frac{\pi Q M}{N}} \tilde{\lambda}_m \tilde{\lambda}_{\langle m+R \rangle_Z}. \quad (4.38)$$

Our target is to retrieve Z unknown $\tilde{\lambda}_m$'s, $m = 0, 1, \dots, Z - 1$ from Z equations

$$\tilde{\lambda}_m \tilde{\lambda}_{(m+R)_Z} = \tilde{x}_1(m) e^{-j \frac{\pi Q M}{N}} \triangleq c_m, \quad \forall m = 0, \dots, Z - 1, \quad (4.39)$$

where c_m is defined as the corresponding constant.

Theorem 4.1. If Z is odd and is co-prime with R , then $\tilde{\lambda}_m$'s can be found from (4.39) as

$$\tilde{\lambda}_m = \frac{\left(\prod_{i=0}^{Z-1} c_i \right)^{1/2}}{\prod_{i=0}^{\frac{Z-3}{2}} c_{(m+(2i+1)R)_Z}}, \quad \forall m = 0, \dots, Z - 1, \quad (4.40)$$

with only a SSA.

Proof. See Appendix E. □

Note that the selection of R and Z is very important in implementing the frequency-domain approach. Let $N'/M' = N/M$ be the simplest form of the fraction; i.e., N' and M' are co-prime. The following ways to select Z are proposed:

- If N' is odd and is greater than $Q + 1$, then we can choose $Z = N'$ and $R = M'$.
- Otherwise, pick any integer κ such that $\kappa N' \geq Q$ (κ can be 1 to account for the case when N' is greater than $Q + 1$ but is even):
 - If $\kappa N'$ is even, then we choose $R = \kappa M'$ and $Z = \kappa N' + 1$, namely, $Z = RN/M + 1$. This choice will guarantee that Z is odd and is co-prime with R , while the consequence is that

$$\xi_m = e^{j \frac{\pi Q M}{N}} \sum_{q=0}^Q \lambda_q e^{-j \frac{2\pi q(R+m)}{Z}} e^{-j \frac{2\pi q M}{Z}} \approx \tilde{\lambda}_{(m+R)_Z}. \quad (4.41)$$

Note that, a similar approximation has been used in many multi-carrier systems when the channel frequency response on the adjacent carriers is assumed to be the same [84]. Our approximation is more accurate since the distortion phase of each summand is only $\frac{2\pi q M}{Z} < \frac{2\pi q}{Z}$. Moreover, we can always choose a large enough Z such that the approximation becomes sufficiently accurate.

- If $\kappa N'$ is odd, then we choose $R = \kappa M'$ and $Z = \kappa N' + 2$, namely, $Z = RN/M + 2$. This choice will guarantee that Z is odd and is co-prime with R . A similar approximation on ξ_m applies.

After obtaining $\tilde{\lambda}$, we can find λ from the first $Q + 1$ elements of the Z -point inverse Fourier transform (IDFT) of $\tilde{\lambda}$. Since the frequency-domain approach fully utilizes all the information,

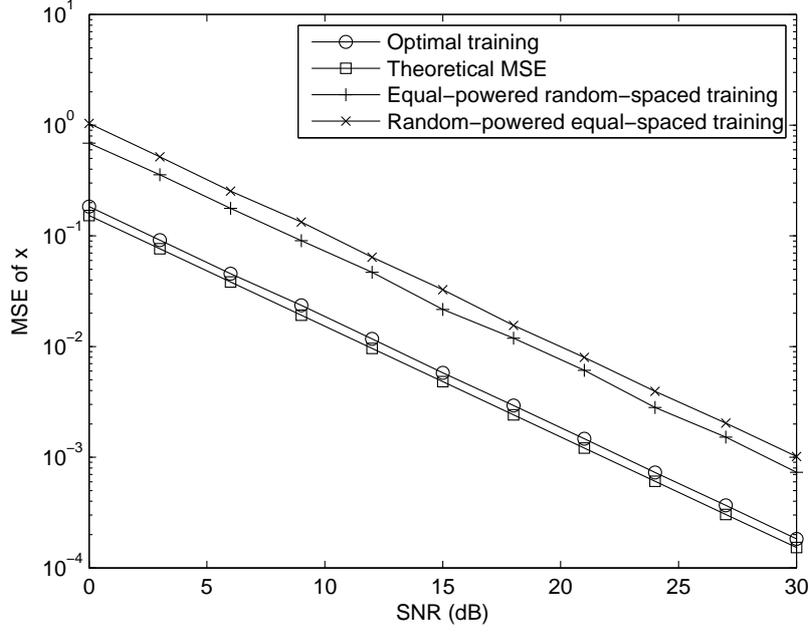


Fig. 4.5. Channel estimation MSEs versus SNR for \mathbf{x} .

the initial estimates of λ_q 's are expected more accurate than those from the time-domain approach. However in the low SNR region, (4.40) is susceptible to error enhancement due to the products in both the denominator and numerator, as will be seen in our later simulations.

The same iteration (4.33) can be then applied to improve the accuracy of λ . Finally, μ can be found from (4.35).

4.5 Simulation Results

In order to evaluate the inherent performance of our algorithms, the time-varying channels are generated directly from the BEM model (4.3). The same approach has been adopted in many papers when testing the performance of channel estimation [39], [85]. However, the real channel generated from (4.1) will be applied for data detection [83], [43], [42].

4.5.1 Channel Estimation and Training Design

The parameters for channel estimation are taken as $Q = 4$, $N = 352$, $M = 8$, and $T = 22$. A total of 10000 Monte-Carlo trials are used for averaging. Optimal training is compared with two types of random training. In the first one, all pilots are equi-powered but randomly spaced. In the second one, the pilot power levels are random, but the pilots are uniformly spaced.

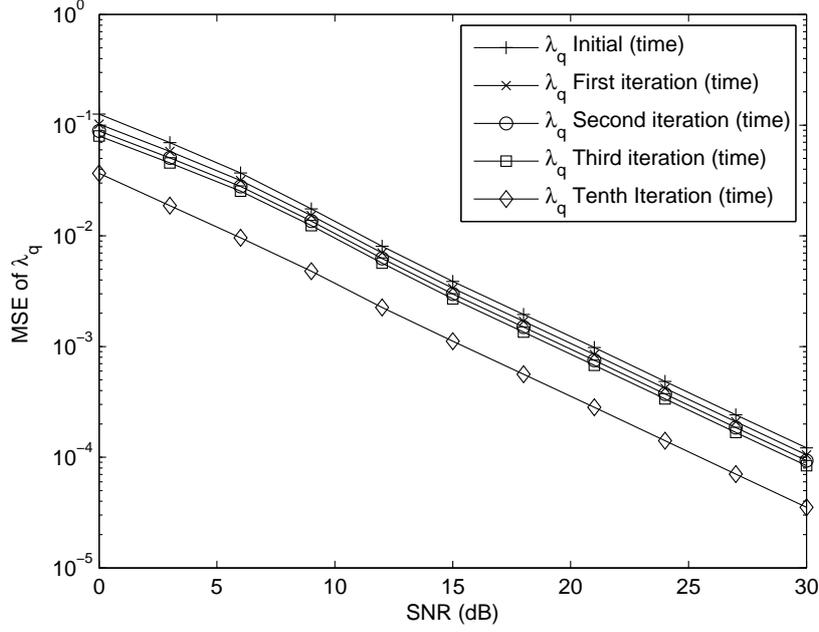


Fig. 4.6. Channel estimation MSEs versus SNR for λ : time-domain approach.

Estimation of the Equivalent BEM Coefficients

The estimation MSEs of the equivalent channel $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T$ from the three different types of training are shown in Fig. 4.5. The theoretical MSE is also displayed for comparison. Clearly, the designed optimal training sequence given in (4.23) achieves the best performance, with the MSE being close to the theoretical one.

Recovery of the Original BEM coefficients– Time-domain Approach

In this example, the coefficients λ are extracted from the estimated $\hat{\mathbf{x}}_1$. The MSEs versus the SNR for the initial estimate as well as those after several iterations are shown in Fig. 4.6, which reveals that the initial estimate is effective in the sense that the MSE curve linearly decreases with the increase of the SNR. Moreover, the iterations can significantly improve the estimation accuracy since the initial estimation utilizes only part of the observations in \mathbf{x}_1 . After the tenth iteration, the improvement is negligible.

Recovery of the Original BEM Coefficients– Frequency-domain Approach

Next we choose the frequency-domain approach to recover the coefficients λ . From (2.10), we take $R = 1, 2, 5$, respectively, and $Z = RN/M + 1$ is 45, 89, 221, respectively. The estimation MSEs versus the SNR for the initial estimation as well as those from the 10-th iteration are shown in

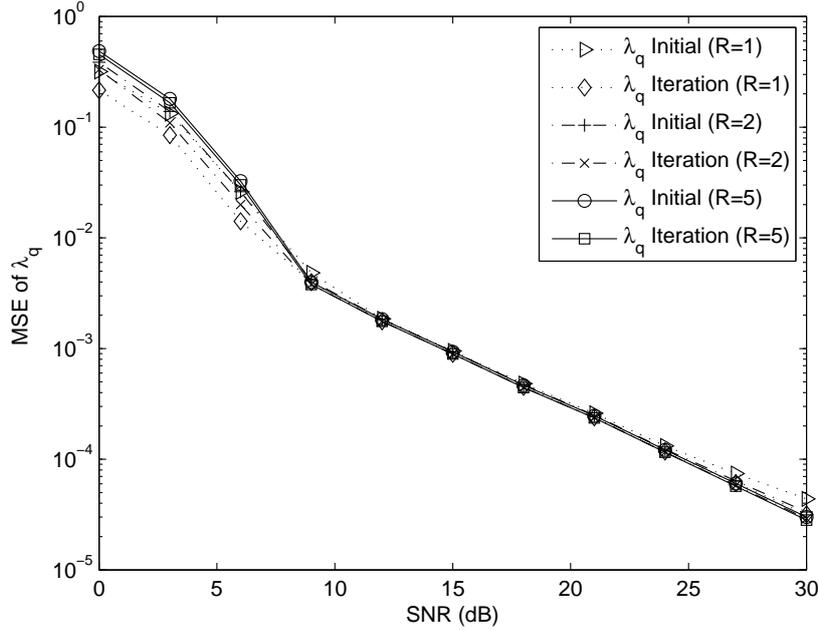


Fig. 4.7. Channel estimation MSEs versus SNR for λ : frequency-domain approach.

Fig. 4.7. The iteration in this case marginally improves the estimation accuracy because the initial frequency-domain estimate fully explores \mathbf{x}_1 . Moreover, the choice of a different R does not affect the performance significantly. As mentioned previously, the performance of the frequency-domain approach degrades at a relatively low SNR, say, 8 dB in Fig. 4.7, due to the error enhancement.

Comparing the Time- and the Frequency-domain Approaches

It is then of interest to compare the performances of the two different approaches in recovering λ . To make this comparison clear, we present the results in a new figure, and apply fifty iterations for both methods. As Fig. 4.8 indicates, the initial frequency-domain estimation outperforms the time-domain results at the high SNR region, even if the latter apply iterations. Nonetheless, the performance gap is quite small. At a relatively low SNR, say SNR= 8 dB, the time-domain approach gives a better performance.

Estimation of μ

After obtaining λ , μ can be estimated through (4.35). The corresponding MSEs versus the SNR from both time and frequency-domain approaches are shown in Fig. 4.9. It is seen that the estimated μ contains a larger error than the estimated λ in Fig. 4.8. This difference is expected since μ is obtained through the estimated λ so that the errors in λ propagate to the estimates of μ .

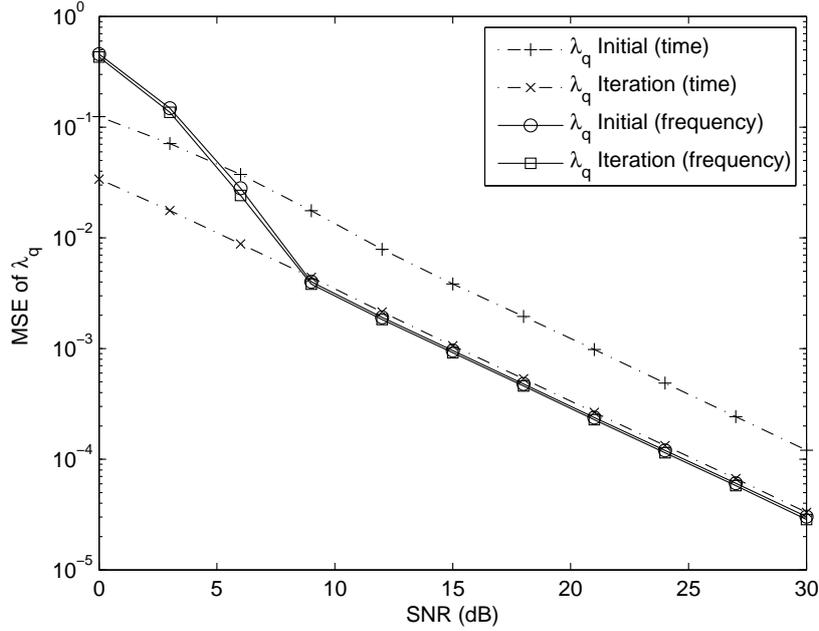


Fig. 4.8. Comparison between the time- and the frequency-domain approaches.

4.5.2 Data Detection

For data detection, the channel is generated by using the more realistic model (4.1). The bit error rate (BER) is the figure of merit. The system parameters shown in Fig. 4.2 are taken. We first apply the channel estimation method to find the BEM coefficients and to reconstruct the time-varying channels $b_i(n)$. Then, the self-signal component is canceled before the data detection. The error due to non-perfect removal of the self-signal will also affect the system performance. The time- and the frequency-domain approaches are used to estimate λ and μ . The BERs versus the SNR for different numbers of Q , and the BER under perfect channel knowledge are displayed in Fig. 4.10. Clearly, the proposed methods yield effective data detection. At high SNRs, the frequency-domain method yields better BER performance than the time-domain method since the former can provide better estimation results. An error floor is observed in the high SNR region due to the mismatch between the BEM model and the real channels. Obviously, the place where the floor begins could be improved by increasing the number of Q s.

4.6 Conclusions

In this chapter, we studied the problem of channel estimation for time-varying TWRN channels. A new PSAM scheme was designed, and the channel estimation was related to a finite number of variables by using CE-BEM. The LS estimator for the convolved BEM coefficients was derived

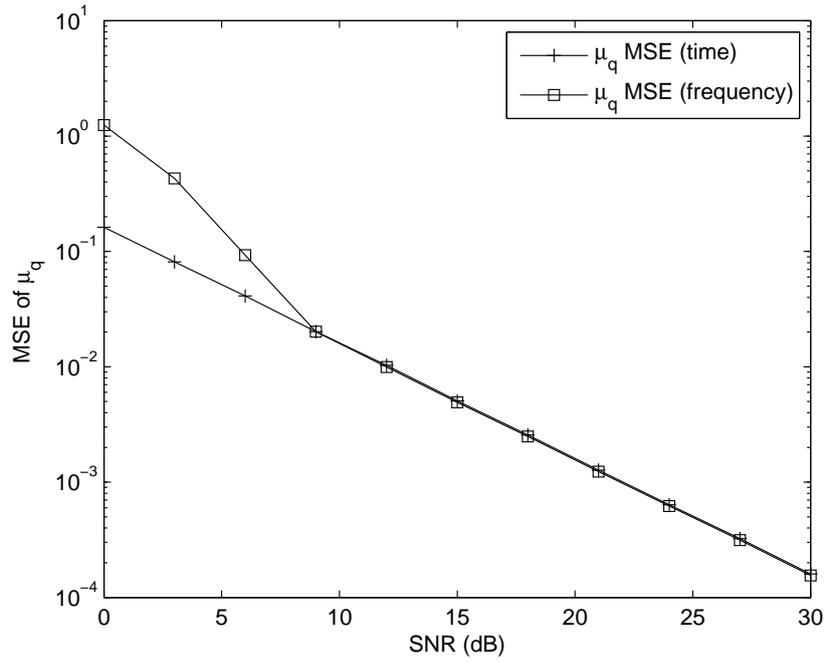


Fig. 4.9. Channel estimation MSEs versus SNR for μ .

along with the optimal training sequences. Time-domain and frequency-domain algorithms were then developed to recover the individual BEM coefficients from the convolved ones. The selection of the system parameters to guide the practical design was fully discussed. The simulation results clearly demonstrated the effectiveness of the proposed algorithms and corroborated the studies.

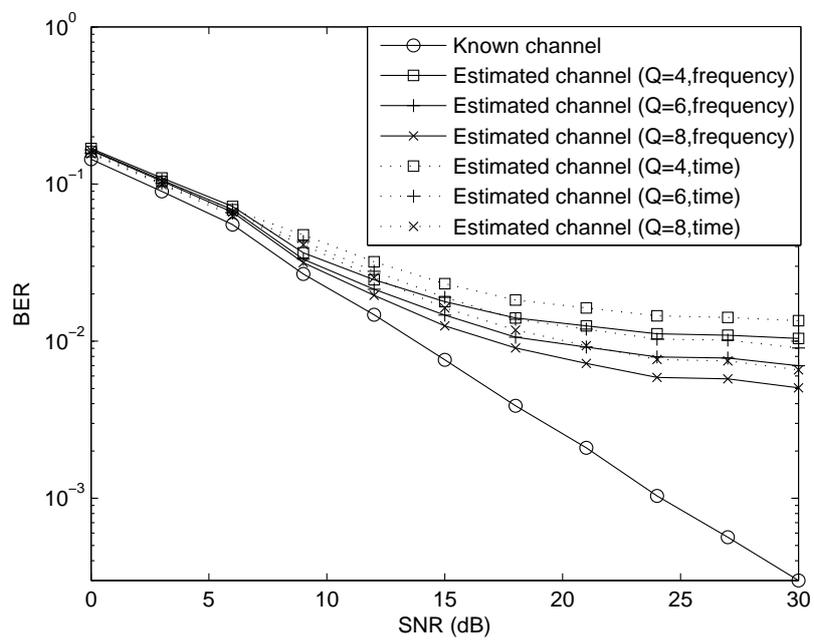


Fig. 4.10. BER versus SNR for realistic mobile-to-mobile channels with different values of Q .

Chapter 5

Doubly-Selective Channel Estimation for OWRN

In this chapter, doubly-selective channel estimation is considered for AF-based OWRN. The time-varying channel is modeled by the CE-BEM, as was done in Chapter 4. Since direct estimation of these coefficients requires significant pilot overhead, we develop an efficient estimator that only targets a set of *useful* channel parameters that could guarantee effective data detection. The training sequence design that can minimize the channel estimation MSE is also proposed.

5.1 Introduction

Assuming block fading scenarios, several channel estimation schemes were proposed for relay network with one or multiple-relay nodes. For example, [81] and [20] studied the channel estimation for relay networks and pointed out that there exist many differences in channel estimation between the AF-based relay networks and the traditional point-to-point networks. Shortly later, channel estimation under frequency-selective environment were developed in [86], [87].

However, in many practical cases the source node, the relay node and the destination node can be mobile. The relative motion between any two nodes will cause Doppler shift and thus make the channel time-varying [63]. Therefore, the relay network is expected to operate over doubly-selective channels. To our best knowledge, estimation techniques for such cases have not yet been developed. This motivates our current work.

As in Chapter 4, the CE-BEM is used to model the time-varying channel. The data frame structure is designed to adapt to the transmission in doubly-selective channels and to facilitate both channel estimation and data detection. We first develop an estimator that targets the combined channel parameters and then propose a detection algorithm. The training sequence that can minimize

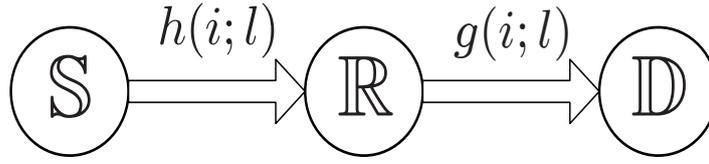


Fig. 5.1. System model for AF relay network over doubly-selective channel.

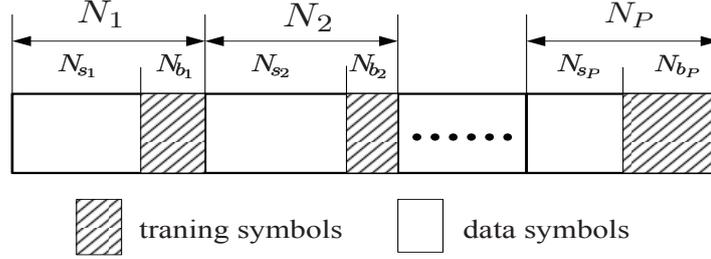


Fig. 5.2. Structure of one transmission block.

the channel estimation mean-square error is also found.

5.2 System model

Consider an AF relay network with one source node \mathbb{S} , one relay node \mathbb{R} and one destination node \mathbb{D} (Fig. 5.1). Let $h(i; l)$ denote the doubly-selective channel between the source node \mathbb{S} and the relay node \mathbb{R} , $g(i; l)$ denote the doubly-selective channel between the relay node \mathbb{R} and the destination node \mathbb{D} . There may exist a switching time at the relay node \mathbb{R} , which results in a delay Δ in the second retransmission phase. However, our model can be well adapted by setting $g(i + \Delta; l) = g(i_0; l)$ where i_0 is the new starting point for the following CE-BEM approximation. Without loss of generality, we assume that the channel length of both $h(i; l)$ and $g(i; l)$ as $L + 1$, and each tap is modeled as a zero mean complex Gaussian random process with power $\sigma_{h,l}^2$ (or $\sigma_{g,l}^2$).

We propose a new transmission scheme as shown in Fig. 5.2. Each transmission block that contains N symbols is divided into P subblocks. Assume the k th subblock contains N_k symbols, of which N_{s_k} symbols are data and are represented by \mathbf{s}_k , while N_{b_k} symbols are pilots and are represented by \mathbf{b}_k . The total number of data symbols is $N_s = \sum_{k=1}^P N_{s_k}$ and the total number of pilots is $N_p = \sum_{k=1}^P N_{b_k}$. With such a structure, we can represent the whole block as a vector

$$\mathbf{x} = [\mathbf{s}_1^T, \mathbf{b}_1^T, \dots, \mathbf{s}_P^T, \mathbf{b}_P^T]. \quad (5.1)$$

During the first phase, the relay node \mathbb{R} receives

$$r(i) = \sum_{l=0}^L h(i; l)x(i-l) + w_1(i), \quad (5.2)$$

where $w_1(i)$ is the additive complex white Gaussian noise with mean zero and variance $\sigma_{w_1}^2$, i.e., $w_1 \sim \mathcal{CN}(0, \sigma_{w_1}^2)$. During the second phase, the relay node \mathbb{R} amplifies $r(i)$ with a constant factor α and then re-transmit it to the destination node \mathbb{D} . The signal obtained by \mathbb{D} is

$$y(i) = \sum_{l=0}^L g(i; l)\alpha r(i-l) + w_2(i) \quad (5.3)$$

$$= \alpha \sum_{l=0}^L g(i; l) \left(\sum_{l=0}^L h(i; l)x(i-l) \right) + \underbrace{\alpha \sum_{l=0}^L g(i; l)w_1(i-l)}_{w(i)} + w_2(i), \quad (5.4)$$

where $w_1(i)$ is the additive complex white Gaussian noise with mean zero and variance $\sigma_{w_1}^2$, i.e., $w_1(i) \sim \mathcal{C}(0, \sigma_{w_1}^2)$ and $w(i)$ means the combined noise. Here, we consider α as a constant. Suppose the average power of the source node is \mathcal{P}_1 , i.e., $E\{|x_i(n)|^2\} = \mathcal{P}_1$ and the average power of the relay node is \mathcal{P}_r . The amplifier factor α can be chosen as

$$\alpha = \sqrt{\frac{\mathcal{P}_r}{\mathcal{P}_1 \sum_{l=0}^L \sigma_{h,l}^2 + \sigma_{w_1}^2}}. \quad (5.5)$$

5.3 Doubly-Selective Channel in OWRN

It is shown in [81], [82] that for relay networks, the channel statistics depend on the mobility of the three nodes. Denote f_{ds} , f_{dd} and f_{dr} as the maximum Doppler shifts due to the motion of \mathbb{S} , \mathbb{D} and \mathbb{R} respectively. The discrete autocorrelation functions for the l th tap of $h(i; l)$ can be represented as [44], [82]

$$R_{h,l}(m) = \sigma_{h,l}^2 E(h(n+m; l)h^*(n; l)) = \sigma_{h,l}^2 J_0(2\pi f_{ds}mT_s)J_0(2\pi f_{dr}mT_s), \quad (5.6)$$

$$R_{g,l}(m) = \sigma_{g,l}^2 E(g(n+m; l)g^*(n; l)) = \sigma_{g,l}^2 J_0(2\pi f_{dr}mT_s)J_0(2\pi f_{dd}mT_s). \quad (5.7)$$

where $J_0(\cdot)$ is the zero-th order Bessel function of the first kind, and T_s is the symbol sampling duration. If one node is fixed, i.e., corresponding Doppler shift becomes zero, then (5.6) and (5.7) reduce to the well-known Jakes model [63].

In fact, (5.6) and (5.7) reveal that the power spectra of $h(i; l)$ and $g(i; l)$ span over the bandwidth $f_{d_1} = f_{ds} + f_{dr}$ and $f_{d_2} = f_{dr} + f_{dd}$ respectively. According to the analysis of CE-BEM in [36],

[42], we can express the doubly-selective channel as

$$h(i; l) = \sum_{q=0}^{Q_1} h_q(l) e^{j2\pi(q-Q_1/2)i/N}, \quad (5.8)$$

$$g(i; l) = \sum_{q=0}^{Q_2} g_q(l) e^{j2\pi(q-Q_2/2)i/N}, \quad (5.9)$$

where $0 \leq i \leq N - 1$, $0 \leq l \leq L$, $Q_m (m = 1, 2) \triangleq 2 \lceil f_{d_m} N T_s \rceil$ is the number of basis. The CE-BEM coefficients $h_q(l)$ and $g_q(l)$ are assumed as zero-mean, complex Gaussian random variables with variance $\sigma_{h,q,l}^2$ and $\sigma_{g,q,l}^2$ respectively [36], [42], [43].

To simplify the notation as well as the following discussion, we assume $f_{d_1} = f_{d_2} = f_d$ and $Q_1 = Q_2 = Q$. We further denote $w_q = 2\pi(q - Q/2)/N$ and define

$$\mathbf{h}_q = [h_q(0), h_q(1), \dots, h_q(L)]^T, \quad (5.10)$$

$$\mathbf{g}_q = [g_q(0), g_q(1), \dots, g_q(L)]^T, \quad q \in [0, Q]. \quad (5.11)$$

Next we apply CE-BEM (5.8) and (5.9) in (5.4) for channel estimation and data detection. Our tasks are: (i) estimate the parameters such as the channel coefficients \mathbf{h}_q and \mathbf{g}_q so that the channel $h(i; l)$ and $g(i; l)$ can be recovered for each time index $i \in [0, N - 1]$, or the combined parameters that will enable successful data detection as did in [20], [90]; (ii) find the optimal training sequence that can minimize the channel estimation error; (iii) recover the data $\mathbf{s}_k, k \in [1, P]$ from the estimated channel.

5.4 Estimation, Detection and Training Sequence Design

Let us construct $N \times 1$ vectors \mathbf{r}, \mathbf{y} , and construct $N \times N$ matrices \mathbf{H}, \mathbf{G} from $g(i; l)$ in the following way:

$$\mathbf{r} = [r(0), r(1), \dots, r(N - 1)]^T, \quad (5.12)$$

$$\mathbf{y} = [y(0), y(1), \dots, y(N - 1)]^T, \quad (5.13)$$

$$\mathbf{H}_{i,j} = h(i; i - j), \quad \mathbf{G}_{i,j} = g(i; i - j), \quad (5.14)$$

for $i, j = 1, 2, \dots, N$. We can write (5.2) and (5.4) as

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{w}_1, \quad (5.15)$$

$$\mathbf{y} = \alpha \mathbf{G}\mathbf{r} + \mathbf{w}_2 = \alpha \mathbf{G}\mathbf{H}\mathbf{x} + \mathbf{w}, \quad (5.16)$$

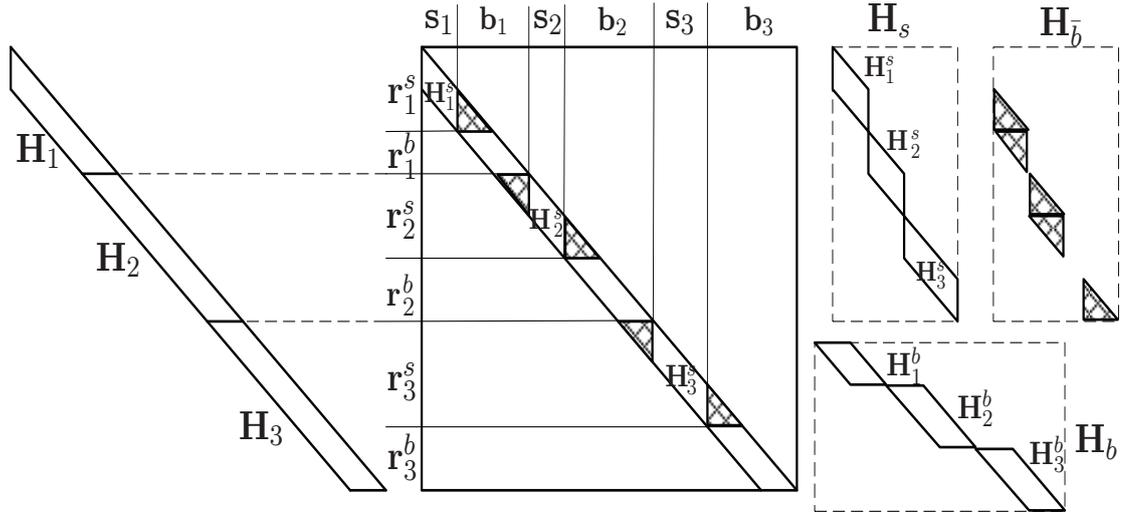


Fig. 5.3. Partition of the matrix \mathbf{H} into \mathbf{H}_s , \mathbf{H}_b , and $\mathbf{H}_{\bar{b}}$ that are shown in dashed line on the right side of the figure.

where $\mathbf{w}_i = [w_i(0), w_i(1), \dots, w_i(N-1)]^T$, $i = 1, 2$ and $\mathbf{w} = [w(0), w(1), \dots, w(N-1)]^T$.

To clearly see the problem of channel estimation and facilitate our analysis, we perform the following channel partition as our first step.

5.4.1 Channel Partition

Following the channel partition method in [43], we can split the channel matrix \mathbf{H} into three matrices, namely, \mathbf{H}_s , \mathbf{H}_b , and $\mathbf{H}_{\bar{b}}$, which are shown in Fig. 5.3. Similarly, the channel \mathbf{H}_k , the k th ($1 \leq k \leq P$) part of \mathbf{H} corresponding to the k th sub-block input of $[\mathbf{s}_k, \mathbf{b}_k]$, can also be partitioned into three matrices \mathbf{H}_k^s , \mathbf{H}_k^b and $\mathbf{H}_k^{\bar{b}}$ (Fig. 5.4). After separation of these channels, we derive two input-output relationships at the relay node

$$\mathbf{r}_s = \mathbf{H}_s \mathbf{s} + \mathbf{H}_{\bar{b}} \bar{\mathbf{b}} + \mathbf{w}_1^s, \quad (5.17)$$

$$\mathbf{r}_b = \mathbf{H}_b \mathbf{b} + \mathbf{w}_1^b, \quad (5.18)$$

where $\mathbf{r}_s = [(\mathbf{r}_1^s)^T, \dots, (\mathbf{r}_P^s)^T]^T$, $\mathbf{r}_b = [(\mathbf{r}_1^b)^T, \dots, (\mathbf{r}_P^b)^T]^T$, $\bar{\mathbf{b}}$ contains the first L and the last L entries of \mathbf{b}_k for all $1 \leq k \leq P$, and \mathbf{w}_1^s and \mathbf{w}_1^b denote the corresponding noise vectors.

Repeat the partition process for the channel \mathbf{G} and \mathbf{G}_k . That is, split \mathbf{G} into \mathbf{G}_s , $\mathbf{G}_{\bar{b}}$ and \mathbf{G}_b (Fig. 5.5), while split \mathbf{G}_k , the k th component of \mathbf{G} , into \mathbf{G}_k^s , $\mathbf{G}_k^{\bar{b}}$ and \mathbf{G}_k^b (Fig. 5.6). We obtain two input-output relationships at the destination node

$$\mathbf{y}_s = \alpha \mathbf{G}_s \mathbf{r}_s + \alpha \mathbf{G}_{\bar{b}} \mathbf{r}_{\bar{b}} + \mathbf{w}_2^s, \quad (5.19)$$

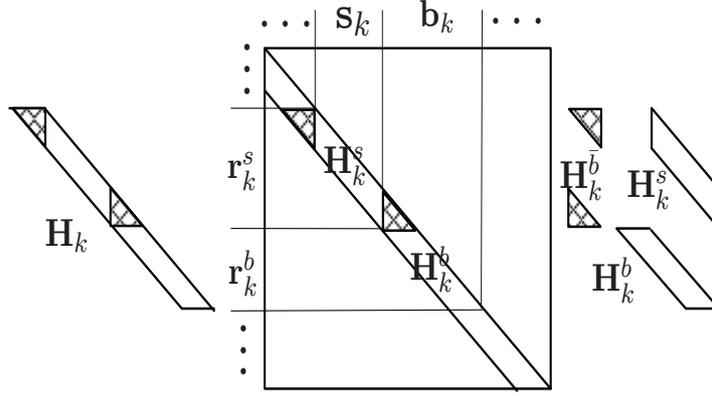


Fig. 5.4. Partition of the matrix \mathbf{H}_k .

$$\mathbf{y}_b = \alpha \mathbf{G}_b \mathbf{r}_b + \mathbf{w}_2^b, \quad (5.20)$$

where $\mathbf{y}_s = [(\mathbf{y}_1^s)^T, \dots, (\mathbf{y}_P^s)^T]^T$, $\mathbf{y}_b = [(\mathbf{y}_1^b)^T, \dots, (\mathbf{y}_P^b)^T]^T$, \mathbf{r}_b contains the first L and the last L entries of \mathbf{r}_k^b for all $1 \leq K \leq P$, \mathbf{w}_2^s and \mathbf{w}_2^b denote the corresponding noise vectors.

Combining (5.18) and (5.20) yields

$$\mathbf{y}_b = \alpha \mathbf{G}_b \mathbf{H}_b \mathbf{b} + \underbrace{\alpha \mathbf{G}_b \mathbf{w}_1^b}_{\mathbf{w}_b} + \mathbf{w}_2^b. \quad (5.21)$$

where \mathbf{w}_b is defined as the corresponding item.

It can be readily checked that (5.21) is equivalent to

$$\mathbf{y}_b = \begin{bmatrix} \mathbf{y}_1^b \\ \vdots \\ \mathbf{y}_P^b \end{bmatrix} = \begin{bmatrix} \alpha \mathbf{G}_1^b \mathbf{H}_1^b \mathbf{b}_1 \\ \vdots \\ \alpha \mathbf{G}_P^b \mathbf{H}_P^b \mathbf{b}_P \end{bmatrix} + \mathbf{w}_b. \quad (5.22)$$

Note that in (5.22) \mathbf{H}_k^b is an $(N_{b_k} - L) \times N_{b_k}$ matrix and \mathbf{G}_k^b is an $(N_{b_k} - 2L) \times (N_{b_k} - L)$ matrix. Thus, to perform channel estimation, the training length for k th sub-block should be $N_{b_k} \geq 2L + 1$

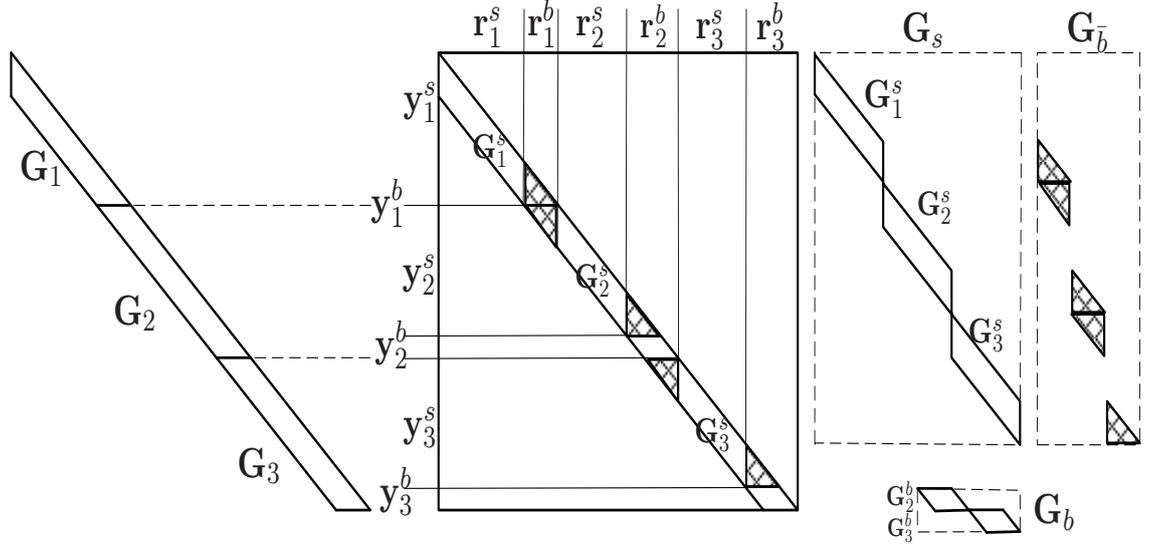


Fig. 5.5. Partition of the matrix \mathbf{G}

5.4.2 Estimation Algorithm

Let us define $\Lambda_M^{(w_q)} = \text{diag}\{1, e^{jw_q}, \dots, e^{jw_q(M-1)}\}$. For any $(L+1) \times 1$ vector $\mathbf{a} = [a_0, a_1, \dots, a_L]^T$, define an $M \times (M+L)$ Toeplitz matrix as

$$\mathbf{T}_{M+L}^{(\mathbf{a})} = \underbrace{\begin{bmatrix} a_L & \cdots & a_0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & a_L & \cdots & a_0 \end{bmatrix}}_{M+L \text{ columns}}. \quad (5.23)$$

We provide the following two Lemmas.

Lemma 5.1.

$$\mathbf{T}_{M+L}^{(\mathbf{a})} \Lambda_{M+L}^{(w_q)} = \Lambda_M^{(w_q)} \mathbf{T}_{M+L}^{(\boldsymbol{\mu}_a)}, \quad (5.24)$$

where $\boldsymbol{\mu}_a = [a_0 e^{jw_q L}, a_1 e^{jw_q(L-1)}, \dots, a_L]$.

Proof. Proved from straight calculations. \square

Lemma 5.2. For two vectors $\mathbf{a}_i = [a_{i,0}, a_{i,1}, \dots, a_{i,L}]^T, i = 1, 2$, there is

$$\mathbf{T}_{M+L}^{(\mathbf{a}_1)} \mathbf{T}_{M+2L}^{(\mathbf{a}_2)} = \mathbf{T}_{M+2L}^{(\mathbf{a}_1 \otimes \mathbf{a}_2)}, \quad (5.25)$$

where \otimes denotes linear convolution.

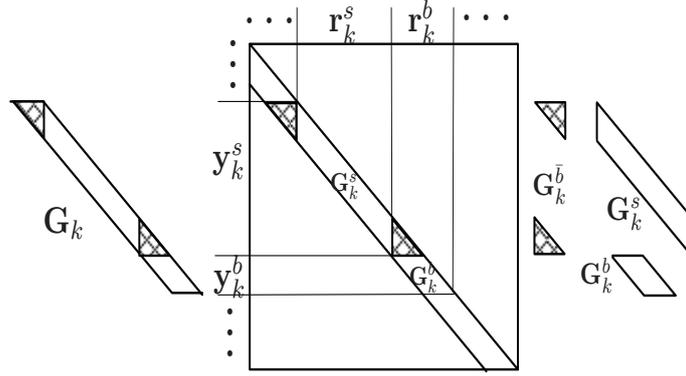


Fig. 5.6. Partition of the matrix \mathbf{G}_k

Proof. Proved from straight calculations. □

According to these definitions and (5.8), we obtain

$$\mathbf{H} = \sum_{q=0}^Q \Lambda_N^{(w_q)} \Phi_q, \quad (5.26)$$

$$\mathbf{H}_k^b = \sum_{q=0}^Q e^{jw_q(N_{s_k}+L+\sum_{i=1}^{k-1} N_i)} \Lambda_{N_{b_k}-L}^{(w_q)} \mathbf{T}_{N_{b_k}}^{(\mathbf{h}_q)}, \quad (5.27)$$

where Φ_q is a lower triangular Toeplitz matrix with the first column $[h_q(0), \dots, h_q(L), 0, \dots, 0]^T$ and $\mathbf{T}_{N_{b_k}}^{(\mathbf{h}_q)}$ is $(N_{b_k} - L) \times N_{b_k}$ Toeplitz matrix as defined in (5.23).

Similarly, based on (5.9) we can obtain

$$\mathbf{G} = \sum_{q=0}^Q \Lambda_N^{(w_q)} \Omega_q, \quad (5.28)$$

$$\mathbf{G}_k^b = \sum_{q=0}^Q e^{jw_q(N_{s_k}+2L+\sum_{i=1}^{k-1} N_i)} \Lambda_{N_{b_k}-2L}^{(w_q)} \mathbf{T}_{N_{b_k}-L}^{(\mathbf{g}_q)}, \quad (5.29)$$

where Ω_q is a lower triangular Toeplitz matrix with the first column $[g_q(0), \dots, g_q(L), 0, \dots, 0]^T$, and $\mathbf{T}_{N_{b_k}-L}^{(\mathbf{g}_q)}$ is an $(N_{b_k} - 2L) \times (N_{b_k} - L)$ Toeplitz matrix as defined in (5.23).

Combining (5.27) and (5.29) gives

$$\begin{aligned} \mathbf{G}_k^b \mathbf{H}_k^b &= \sum_{m=0}^Q e^{jw_m(N_{s_k}+2L+\sum_{i=1}^{k-1} N_i)} \Lambda_{N_{b_k}-2L}^{(w_m)} \mathbf{T}_{N_{b_k}-L}^{(\mathbf{g}_m)} \sum_{n=0}^Q e^{jw_n(N_{s_k}+L+\sum_{i=1}^{k-1} N_i)} \Lambda_{N_{b_k}-L}^{(w_n)} \mathbf{T}_{N_{b_k}}^{(\mathbf{h}_n)} \\ &= \sum_{m=0}^Q \sum_{n=0}^Q \theta_{m,n,k} \underbrace{\Lambda_{N_{b_k}-2L}^{(w_m)} \mathbf{T}_{N_{b_k}-L}^{(\mathbf{g}_m)} \Lambda_{N_{b_k}-L}^{(w_n)} \mathbf{T}_{N_{b_k}}^{(\mathbf{h}_n)}}_{\Xi_{m,n,k}} \end{aligned} \quad (5.30)$$

where

$$\theta_{m,n,k} = e^{jw_m(N_{s_k} + 2L + \sum_{i=1}^{k-1} N_i) + jw_n(N_{s_k} + L + \sum_{i=1}^{k-1} N_i)}, \quad (5.31)$$

and $\Xi_{m,n,k}$ is defined as the corresponding item. Using Lemma 5.1 and Lemma 5.2, $\Xi_{m,n,k}$ can be simplified as

$$\Xi_{m,n,k} = \Lambda_{N_{b_k}-2L}^{(w_m)} \Lambda_{N_{b_k}-2L}^{(w_n)} \mathbf{T}_{N_{b_k}-L}^{(\boldsymbol{\mu}_{g_m})} \mathbf{T}_{N_{b_k}}^{(\mathbf{h}_n)} \quad (5.32)$$

$$= \Lambda_{N_{b_k}-2L}^{(w_m+w_n)} \mathbf{T}_{N_{b_k}}^{(\boldsymbol{\lambda}_{m,n})}, \quad (5.33)$$

where

$$\boldsymbol{\mu}_{g_m} = [g_m(0)e^{jw_n L}, g_m(1)e^{jw_n(L-1)}, \dots, g_m(L)]^T, \quad (5.34)$$

$$\boldsymbol{\lambda}_{m,n} = \boldsymbol{\mu}_{g_m} \otimes \mathbf{h}_n. \quad (5.35)$$

Since $\mathbf{T}_{N_{b_k}}^{(\boldsymbol{\lambda}_{m,n})}$ is a Toeplitz matrix, we obtain

$$\begin{aligned} \mathbf{G}_k^b \mathbf{H}_k^b \mathbf{b}_k &= \sum_{m=0}^Q \sum_{n=0}^Q \theta_{m,n,k} \Lambda_{N_{b_k}-2L}^{(w_m+w_n)} \mathbf{T}_{N_{b_k}}^{(\boldsymbol{\lambda}_{m,n})} \mathbf{b}_k \\ &= \sum_{m=0}^Q \sum_{n=0}^Q \theta_{m,n,k} \Lambda_{N_{b_k}-2L}^{(w_m+w_n)} \mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} \boldsymbol{\lambda}_{m,n}, \end{aligned} \quad (5.36)$$

where $\mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)}$ is defined as

$$\mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} = \begin{bmatrix} b_k(2L), & \cdots, & b_k(0) \\ b_k(2L+1), & \cdots, & b_k(1) \\ \vdots & \vdots & \vdots \\ b_k(N_{b_k}-1), & \cdots, & b_k(N_{b_k}-2L-1) \end{bmatrix}. \quad (5.37)$$

Unfortunately, it remains challenging to estimate $\boldsymbol{\lambda}_{m,n}$ from (5.36). A direct way to estimate all $\boldsymbol{\lambda}_{m,n}$ requires N_b to be no less than $2PL + (Q+1)^2(2L+1)$, which is too large and the transmission efficiency will be reduced. To solve this problem, we choose to estimate other type of channel information that requires smaller training length but at the same time ensures effective data detection.

Let us introduce two variables $\zeta_{q,k}$ and ϖ_q defined as

$$\varpi_q = w_m + w_n = e^{j2\pi(q-Q)/N}, \quad m, n \in [0, Q], q \in [0, 2Q], \quad (5.38)$$

$$\zeta_{q,k} = e^{j\varpi_q(N_{s_k} + 2L + \sum_{i=1}^{k-1} N_i)}, \quad k \in [1, P]. \quad (5.39)$$

It can be readily checked that $\theta_{m,n,k} = \zeta_{m+n,k} e^{-jw_n L}$. Then we can combine those items that satisfy $m + n = q$ in (5.36) and obtain

$$\mathbf{G}_k^b \mathbf{H}_k^b \mathbf{b}_k = \sum_{q=0}^{2Q} \zeta_{q,k} \Lambda_{N_{b_k}-2L}^{(\varpi_q)} \mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} \boldsymbol{\eta}_q, \quad (5.40)$$

where

$$\boldsymbol{\eta}_q = \sum_{m+n=q} e^{-jw_n L} \boldsymbol{\lambda}_{m,n}. \quad (5.41)$$

Now define

$$\boldsymbol{\eta} = [\boldsymbol{\eta}_0^T, \boldsymbol{\eta}_1^T, \dots, \boldsymbol{\eta}_{2Q}^T]^T, \quad (5.42)$$

as the parameters to be estimated. Substituting (5.40) into (5.22) provides a simple model

$$\mathbf{y}_b = \alpha \boldsymbol{\Psi}_b \boldsymbol{\eta} + \mathbf{w}_b, \quad (5.43)$$

where $\boldsymbol{\Psi}_b$ is defined as

$$\boldsymbol{\Psi}_b = \begin{bmatrix} \zeta_{0,1} \Lambda_{N_{b_1}-2L}^{(\varpi_0)} \mathbf{B}_{N_{b_1}}^{(\mathbf{b}_1)}, & \dots, & \zeta_{2Q,1} \Lambda_{N_{b_1}-2L}^{(\varpi_{2Q})} \mathbf{B}_{N_{b_1}}^{(\mathbf{b}_1)} \\ \vdots & \vdots & \vdots \\ \zeta_{0,P} \Lambda_{N_{b_P}-2L}^{(\varpi_0)} \mathbf{B}_{N_{b_P}}^{(\mathbf{b}_P)}, & \dots, & \zeta_{2Q,P} \Lambda_{N_{b_P}-2L}^{(\varpi_{2Q})} \mathbf{B}_{N_{b_P}}^{(\mathbf{b}_P)} \end{bmatrix}, \quad (5.44)$$

Instead of estimating the coefficients \mathbf{h}_q and \mathbf{g}_q , we could estimate another parameter $\boldsymbol{\eta}$ from

$$\hat{\boldsymbol{\eta}} = \frac{1}{\alpha} \left(\boldsymbol{\Psi}_b^H \boldsymbol{\Psi}_b \right)^{-1} \boldsymbol{\Psi}_b^H \mathbf{y}_b. \quad (5.45)$$

Moreover, $\hat{\boldsymbol{\eta}}_q$ can be directly obtained from $\hat{\boldsymbol{\eta}}$ for each $q \in [0, 2Q]$.

5.4.3 Data Detection

Substituting (5.17) into (5.19) yields

$$\mathbf{y}_s = \alpha \mathbf{G}_s \mathbf{H}_s \mathbf{s} + \alpha \mathbf{G}_s \mathbf{H}_s \bar{\mathbf{b}} + \alpha \mathbf{G}_s \mathbf{w}_1^s + \alpha \mathbf{G}_s \mathbf{r}_b^s + \mathbf{w}_2^s. \quad (5.46)$$

Note that $\mathbf{r}_{\bar{b}}$ in (5.46) can be further decomposed as

$$\mathbf{r}_{\bar{b}} = \mathbf{H}_{\bar{b}} \check{\mathbf{b}} + \mathbf{w}_1^{\bar{b}}, \quad (5.47)$$

where $\mathbf{H}_{\bar{b}}$ contains the first L and the last L rows of every \mathbf{H}_k^b , $\check{\mathbf{b}}$ contains the first $2L$ and the last $2L$ entries of every \mathbf{b}_k , $1 \leq K \leq P$ and $\mathbf{w}_1^{\bar{b}}$ denotes the corresponding noise. Then (5.46) can be written as

$$\mathbf{y}_s = \alpha \mathbf{G}_s \mathbf{H}_s \mathbf{s} + \alpha \mathbf{G}_s \mathbf{H}_{\bar{b}} \bar{\mathbf{b}} + \alpha \mathbf{G}_{\bar{b}} \mathbf{H}_{\bar{b}} \check{\mathbf{b}} + \mathbf{w}_s, \quad (5.48)$$

where the combined noise vector $\mathbf{w}_s = \alpha \mathbf{G}_s \mathbf{w}_1^s + \alpha \mathbf{G}_{\bar{b}} \mathbf{w}_1^{\bar{b}} + \mathbf{w}_2^s$.

Lemma 5.3. Among all training choices that lead to identical covariance matrix of the channel estimation error, if the training length N_{b_k} is greater than $4L + 1$ and if the training has the first $2L$ and the last $2L$ entries equal to zero, then the interference to the data detection is minimized.

Proof. See Appendix F. □

Following Lemma 5.3, we can simplify (5.48) as

$$\mathbf{y}_s = \alpha \mathbf{G}_s \mathbf{H}_s \mathbf{s} + \mathbf{w}_s, \quad (5.49)$$

which is equivalent to

$$\mathbf{y}_s = \begin{bmatrix} \mathbf{y}_1^s \\ \vdots \\ \mathbf{y}_P^s \end{bmatrix} = \begin{bmatrix} \alpha \mathbf{G}_1^s \mathbf{H}_1^s \mathbf{s}_1 \\ \vdots \\ \alpha \mathbf{G}_P^s \mathbf{H}_1^s \mathbf{s}_P \end{bmatrix} + \mathbf{w}_s. \quad (5.50)$$

Define $\mathbf{U}_M^{(\mathbf{h}_q)}$ is a Toeplitz matrix generated by the vector \mathbf{h}_q in the following way:

$$\mathbf{U}_M^{(\mathbf{h}_q)} = \underbrace{\begin{bmatrix} h_q(0), & \cdots, & 0 \\ \vdots & \ddots & \vdots \\ h_q(L), & \ddots, & h_q(0) \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_q(L) \end{bmatrix}}_{M \text{ columns}}. \quad (5.51)$$

We have the following lemmas.

Lemma 5.4.

$$\mathbf{U}_M^{(\mathbf{h}_q)} \mathbf{\Lambda}_M^{(w_q)} = e^{-jw_q L} \mathbf{\Lambda}_{M+L}^{(w_q)} \mathbf{U}_M^{(\boldsymbol{\mu}_{h_q})}, \quad (5.52)$$

where $\boldsymbol{\mu}_{h_q} = [h_q(0)e^{jw_q L}, h_q(1)e^{jw_q(L-1)}, \dots, h_q(L)]^T$.

Proof. Proved from straight calculation. \square

Lemma 5.5.

$$\mathbf{U}_{M+L}^{(\mathbf{g}_q)} \mathbf{U}_M^{(\mathbf{h}_q)} = \mathbf{U}_M^{(\mathbf{g}_q * \mathbf{h}_q)} \quad (5.53)$$

Proof. Proved from straight calculation. \square

Then according to (5.8) and (5.9), we obtain

$$\mathbf{G}_k^s = \sum_{q=0}^Q e^{jw_q \sum_{i=1}^{k-1} N_i} \mathbf{\Lambda}_{N_{s_k}+2L}^{(w_q)} \mathbf{U}_{N_{s_k}+L}^{(\mathbf{g}_q)}, \quad (5.54)$$

$$\mathbf{H}_k^s = \sum_{q=0}^Q e^{jw_q \sum_{i=1}^{k-1} N_i} \mathbf{\Lambda}_{N_{s_k}+L}^{(w_q)} \mathbf{U}_{N_{s_k}}^{(\mathbf{h}_q)}. \quad (5.55)$$

Next it can be found

$$\mathbf{G}_k^s \mathbf{H}_k^s = \sum_{m=0}^Q \sum_{n=0}^Q \phi_{m,n,k} \mathbf{\Lambda}_{N_{s_k}+2L}^{(w_m)} \mathbf{U}_{N_{s_k}+L}^{(\mathbf{g}_m)} \mathbf{\Lambda}_{N_{s_k}+L}^{(w_n)} \mathbf{U}_{N_{s_k}}^{(\mathbf{h}_n)} \quad (5.56)$$

where $\phi_{m,n,k} = e^{j(w_m+w_n) \sum_{i=1}^{k-1} N_i}$. Using Lemma 5.4 and Lemma 5.5, it can be derived that

$$\mathbf{U}_{N_{s_k}+L}^{(\mathbf{g}_m)} \mathbf{\Lambda}_{N_{s_k}+L}^{(w_n)} \mathbf{U}_{N_{s_k}}^{(\mathbf{h}_n)} = e^{-jw_n L} \mathbf{\Lambda}_{N_{s_k}+2L}^{(w_n)} \mathbf{U}_{N_{s_k}+L}^{(\boldsymbol{\mu}_{g_m})} \mathbf{U}_{N_{s_k}}^{(\mathbf{h}_n)} \quad (5.57)$$

$$= e^{-jw_n L} \mathbf{\Lambda}_{N_{s_k}+2L}^{(w_n)} \mathbf{U}_{N_{s_k}}^{(\boldsymbol{\lambda}_{m,n})}, \quad (5.58)$$

where $\boldsymbol{\mu}_{g_m}$ and $\boldsymbol{\lambda}_{m,n}$ are defined in (5.34) and (5.35) respectively. Substituting (5.58) into (5.56), we can obtain

$$\begin{aligned} \mathbf{G}_k^s \mathbf{H}_k^s &= \sum_{m=0}^Q \sum_{n=0}^Q \phi_{m,n,k} e^{-jw_n L} \mathbf{\Lambda}_{N_{s_k}+2L}^{(w_m+w_n)} \mathbf{U}_{N_{s_k}}^{(\boldsymbol{\lambda}_{m,n})} \\ &= \sum_{q=0}^{2Q} e^{j\varpi_q \sum_{i=1}^{k-1} N_i} \mathbf{\Lambda}_{N_{s_k}+2L}^{(\varpi_q)} \mathbf{U}_{N_{s_k}}^{(\boldsymbol{\eta}_q)}. \end{aligned} \quad (5.59)$$

Clearly, given the estimates of $\boldsymbol{\eta}_q$, $\mathbf{G}_k^s \mathbf{H}_k^s$ can be reconstructed from (5.59). Hence, the data \mathbf{s}_k can be detected with the reconstructed channel information $\mathbf{G}_k^s \mathbf{H}_k^s$.

5.4.4 Training Sequence Design

The estimation error of $\boldsymbol{\eta}$ can be expressed as

$$\mathbf{e} = \hat{\boldsymbol{\eta}} - \boldsymbol{\eta} = \left(\boldsymbol{\Psi}_b^H \boldsymbol{\Psi}_b \right)^{-1} \boldsymbol{\Psi}_b^H \mathbf{w}_b. \quad (5.60)$$

The correlation matrix of \mathbf{w}_b can be found from (5.29) as

$$\mathbf{R}_{w_b} = E(\mathbf{w}_b \mathbf{w}_b^H) = \left(\sigma_{w_2}^2 \sum_{q=0}^Q \sum_{l=0}^L |g_q(l)|^2 + \sigma_{w_1}^2 \right) \mathbf{I}_{N_b - 2PL}. \quad (5.61)$$

Thus the mean square error of \mathbf{e} is

$$\sigma_e^2 = \text{tr}(E(\mathbf{e} \mathbf{e}^H)) = C_e \text{tr} \left(\boldsymbol{\Psi}_b^H \boldsymbol{\Psi}_b \right)^{-1} \quad (5.62)$$

where $C_e = \left(\sigma_{w_2}^2 \sum_{q=0}^Q \sum_{l=0}^L |g_q(l)|^2 + \sigma_{w_1}^2 \right) / \alpha^2$.

According to [88, Appendix A], we know that σ_e^2 in (5.62) is lower bounded as follows:

$$C_e \text{tr} \left(\boldsymbol{\Psi}_b^H \boldsymbol{\Psi}_b \right)^{-1} \geq \sum_m \frac{C_e}{[\boldsymbol{\Psi}_b^H \boldsymbol{\Psi}_b]_{m,m}}, \quad (5.63)$$

where the equality holds if and only if $(\boldsymbol{\Psi}_b^H \boldsymbol{\Psi}_b)$ is a diagonal matrix. We then need to design the training sequence that can diagonalize $(\boldsymbol{\Psi}_b^H \boldsymbol{\Psi}_b)$.

Based on the definition of $\boldsymbol{\Psi}_b$ (5.44), the optimal training sequence that can minimize the σ_e^2 requires the following conditions to be satisfied:

$$\sum_{k=1}^P \left(\mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} \right)^H \mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} = \mathcal{P}_b \mathbf{I}_{2L+1}, \quad (5.64)$$

$$\sum_{k=1}^P \left(\mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} \right)^H \boldsymbol{\Lambda}_{N_{b_k}-2L}^{(-\varpi_{q_1})} \zeta_{q_1,k}^H \zeta_{q_2,k} \boldsymbol{\Lambda}_{N_{b_k}-2L}^{(\varpi_{q_2})} \mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} = \mathbf{0}_{2L+1}, \quad \forall q_1 \neq q_2, q_1, q_2 \in [0, 2Q] \quad (5.65)$$

where \mathcal{P}_b is the power allocated to the training sequence.

Let us first focus on (5.64). Observing the structure of $\mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)}$, we know that (5.64) can be fulfilled if the following conditions are satisfied:

$$(C1): \quad N_{b_k} = 4L + 1, \quad \forall k \in [1, P], \quad (5.66)$$

$$(C2): \quad \mathbf{b}_k = \sqrt{\mathcal{P}_b/P} [0, \dots, 0, 1, 0, \dots, 0]^T. \quad (5.67)$$

With conditions (C1) and (C2), we can further simplify (5.65) as

$$\begin{aligned}
& \frac{\mathcal{P}_b}{P} \sum_{k=1}^P \mathbf{\Lambda}_{2L+1}^{(-\varpi_{q_1})} \zeta_{q_1,k}^H \zeta_{q_2,k} \mathbf{\Lambda}_{2L+1}^{(\varpi_{q_2})} \\
&= \frac{\mathcal{P}_b}{P} \sum_{k=1}^P e^{j \frac{2\pi}{N} (q_2 - q_1) (N_{s_k} + 2L + \sum_{i=1}^{k-1} N_i)} \mathbf{\Lambda}_{2L+1}^{(\varpi_{q_2} - \varpi_{q_1})} \\
&= \mathbf{0}_{2L+1}, \quad \forall q_1 \neq q_2, q_1, q_2 \in [0, 2Q].
\end{aligned} \tag{5.68}$$

It can be readily checked that the sufficient conditions to achieve (5.68) is

$$(C3): \quad N = P(N_{s_k} + 4L + 1), \quad N_{s_k} = N_s/P, \quad \forall k \in [1, P]. \tag{5.69}$$

Conditions (C1), (C2) and (C3) imply that the equal-spaced and equal-powered training sequence. This coincides with the optimal training sequence design in the traditional point-to-point channel [43].

5.4.5 Block Parameters

The estimator requires $N_b \geq 2PL + (2Q + 1)(2L + 1)$ and the optimal training design requires $N_b = P(4L + 1)$ to minimize the mean-square channel estimation error. Thus we know that $P \geq (2Q + 1)$ and $N \geq (N_{s_k} + 4L + 1)(2Q + 1)$.

Suppose a 3-tap channel and $N_{s_k} = 4L + 1 = 9$, we can obtain

$$2Q + 1 = 4 \lceil f_d T_s N \rceil + 1 \leq N/18. \tag{5.70}$$

It can be found

$$f_d T_s \leq \frac{1}{72} + \frac{1}{4N} \approx 0.0139 + \frac{1}{4N}. \tag{5.71}$$

Using the following parameters:

- carrier frequency $f_c = 900$ MHz and thus the wavelength $\lambda = 1/3$ m;
- data rate 20 kbps and thus the symbol period $T_s = 50 \mu s$;

Since the maximum Doppler shift f_d is V/λ , we can find the mobile speed $V \leq 66$ m/s, which can satisfy most application requirements.

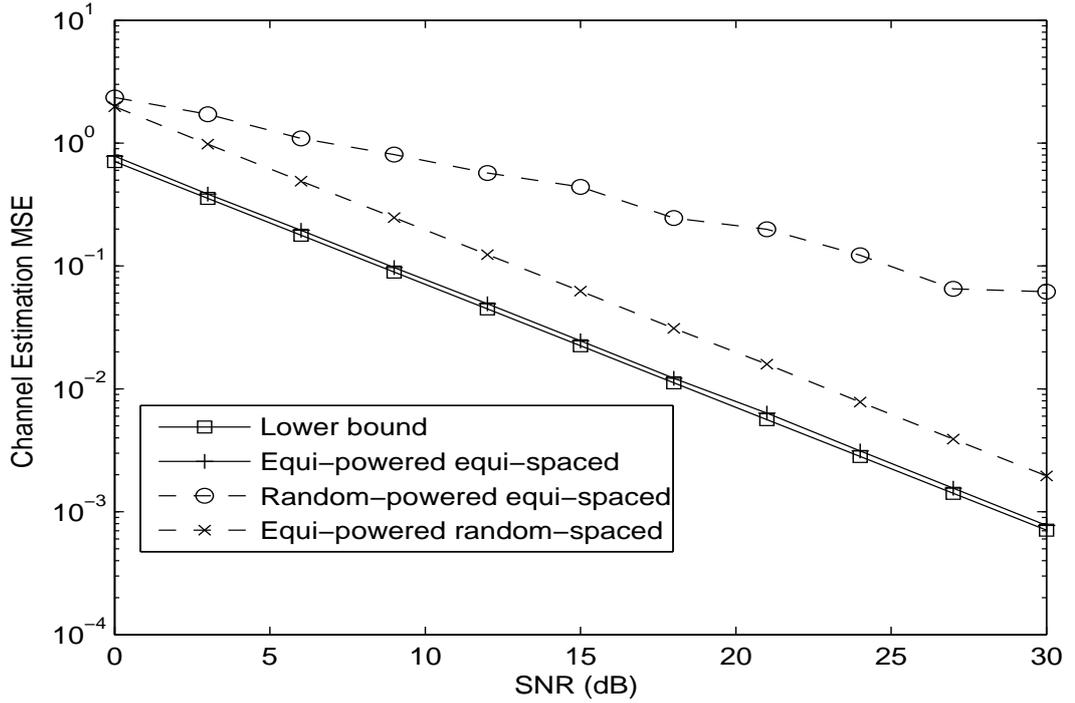


Fig. 5.7. Channel MSE versus the SNR.

5.5 Simulation Results

In order to evaluate the inherent performance of our algorithms, the doubly-selective channels are generated directly from the CE-BEM (5.8) and (5.9). The same approach has been adopted in many other papers when testing the performance of channel estimation [39], [43]. However, the real channel will be also applied for data detection.

We assume that carrier frequency $f_c = 900$ MHz, one symbol period $T_s = 50\mu s$ and the maximum mobility speed is 90 km/hour. Thus we know that the maximum Doppler shift is $f_d = 75$ Hz and $f_d T_s = 3.75 \times 10^{-3}$. Suppose one block contains 360 symbols, i.e., $N = 360$. Then $Q = 2\lceil N f_d T_s \rceil = 4$. We also assume that both doubly-selective channels $h(i; l)$ and $g(i; l)$ has 3 taps, i.e., $L = 2$. Thus we know that $P \geq (2Q + 1) = 9$ and $N_b \geq P(4L + 1) = 81$. The variance of each tap for channel $h(i; l)$ is $\sigma_{h,l}^2 = \sum_{q=0}^Q \sigma_{h,q,l}^2 = e^{-l/10}$ and that for channel $g(i; l)$ is $\sigma_{g,l}^2 = \sum_{q=0}^Q \sigma_{g,q,l}^2 = e^{-l/10}$. The variance of the noise is taken as $\sigma_{w_1}^2 = \sigma_{w_2}^2 = 1$. The SNR is defined as the ratio of symbol power to the noise power, i.e., E_s/N_0 . BPSK constellation is utilized for both training and data symbols. 1000 Monte-Carlo trials are used for the averaging.

First we set the total number of trainings $N_b = 120$ and use three types of training: equi-powered and equi-spaced (our optimal design); equi-powered but with random length; equi-spaced but with random power. For performance comparison, the total power for each types of trainings is the same.

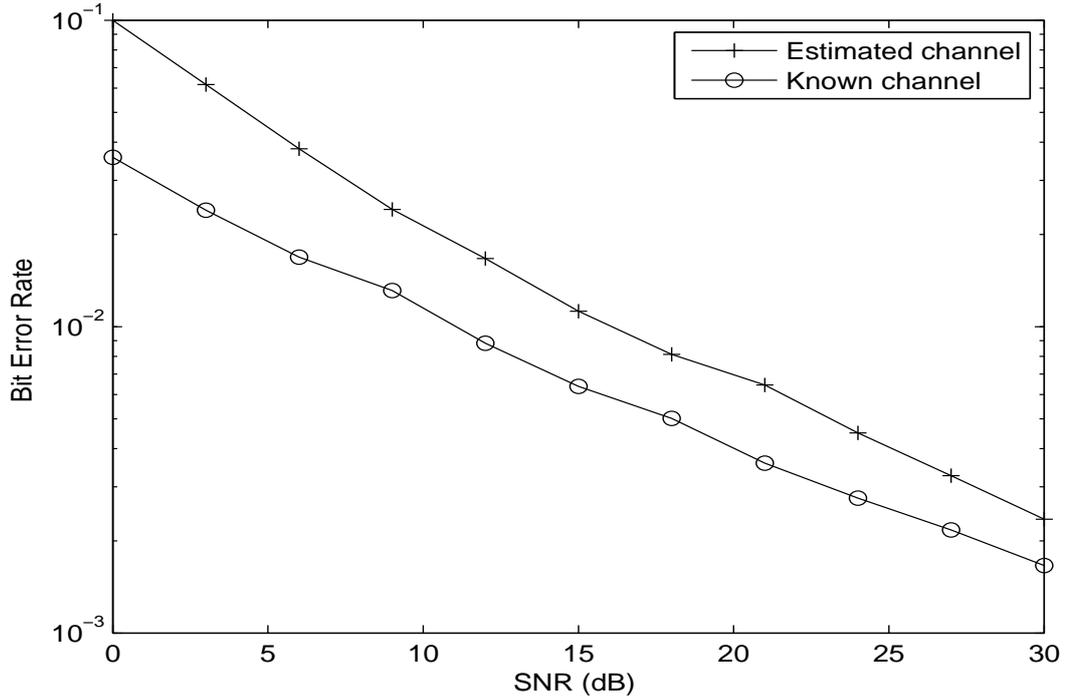


Fig. 5.8. BER versus the SNR.

For each type of training, we find the MSE of our specially defined channel $\boldsymbol{\eta}$. The estimation MSEs for each type are plotted versus SNR in Fig. 5.7. The lower bound of σ_e^2 (5.63) is also plotted for comparison. It can be seen that the equi-spaced equi-powered training achieves the minimum estimation MSE among all the three trainings and its MSE almost approaches the lower bound in (5.63).

Next we use the estimated channel $\hat{\boldsymbol{\eta}}$ to perform data detection. Define the BER as the ratio of number of successfully decoded data symbols over N_s the number of transmitted data symbols. The BER versus SNR is plotted in Fig. 5.8. The BER curve in the case of perfectly known channel $\boldsymbol{\eta}$ is also plotted for comparison. It can be seen that our detection method works well and at high SNR our BER curve approaches that of the ideal case when the channel is perfectly known at the receiver.

We also examine the performance of the proposed estimation and detection methods for real channels. That is, the channel samples are generated according to (5.6) and (5.7). We choose three different number of bases Q as 4, 6, and 8 respectively, and hence the corresponding number of data symbols N_s is 279, 243, and 207. The BER versus SNR is plotted in Fig. 5.9. For comparison, the BER curve under perfect channel knowledge at the receiver is also displayed. Clearly, the proposed methods yield effective data detection. An error floor is observed in the high SNR region due to the mismatch between the BEM model and the real channels. Obviously, the place where the floor begins could be improved by increasing the number of Q s.

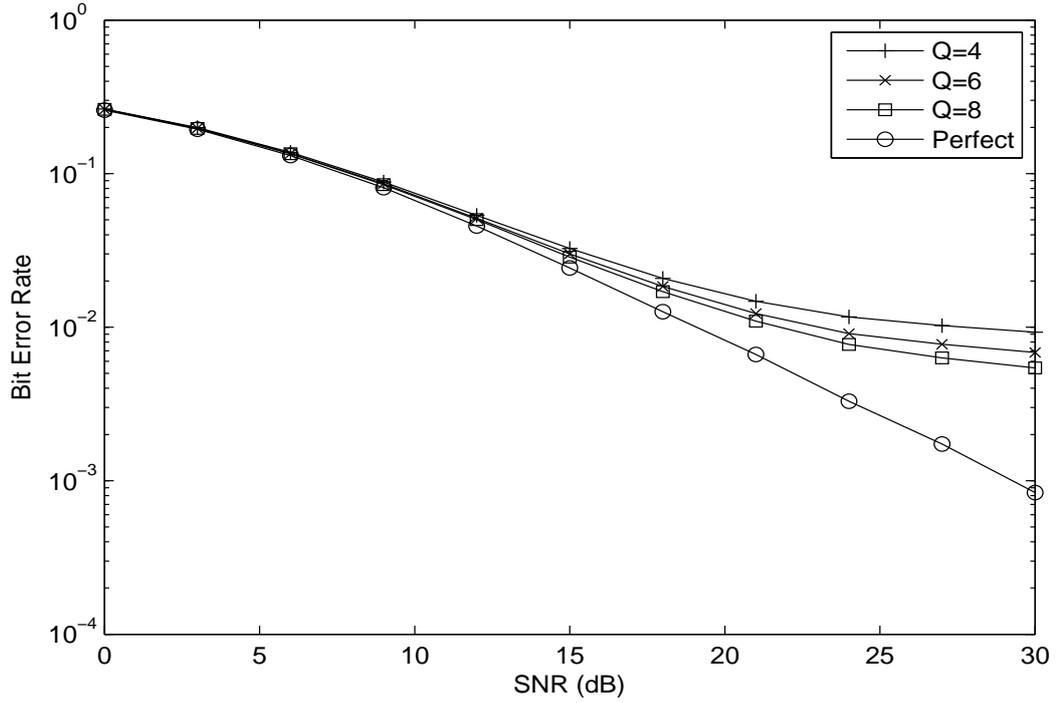


Fig. 5.9. BER versus the SNR: real channel.

In the last example, we choose three different number of subblocks P as $2Q + 1$, $2Q + 2$ and $2Q + 4$ respectively, and the space left for data transmission is $N_{s_k} = N - P(4L + 1) = 279, 270$ and 252 correspondingly. Define the transmission efficiency as the ratio of the number of successfully decoded data symbols over total number of symbols, i.e., $N_s \times \text{BER}/N$. We run the simulation process as SNR ranges from -10 dB to 30 dB. The transmission efficiency at different SNR for each P is plotted in Fig. 5.10. It is shown that when the number of subblocks P equals $2Q + 1$, the best transmission efficiency is achieved at all SNR. It can be explained that when P increases by one unit, the data loss will be $4L + 1$, which can not be compensated even if channel estimation performance can be improved by larger P .

5.6 Conclusions

In this chapter, doubly-selective channel estimation was considered for AF-based relay networks. Based on the CE-BEM, we designed an efficient method to estimate the channel coefficients and detect data symbols. The optimal training sequence that can minimize the estimation MSE was also derived. Finally, extensive numerical results are provided to corroborate the proposed studies.

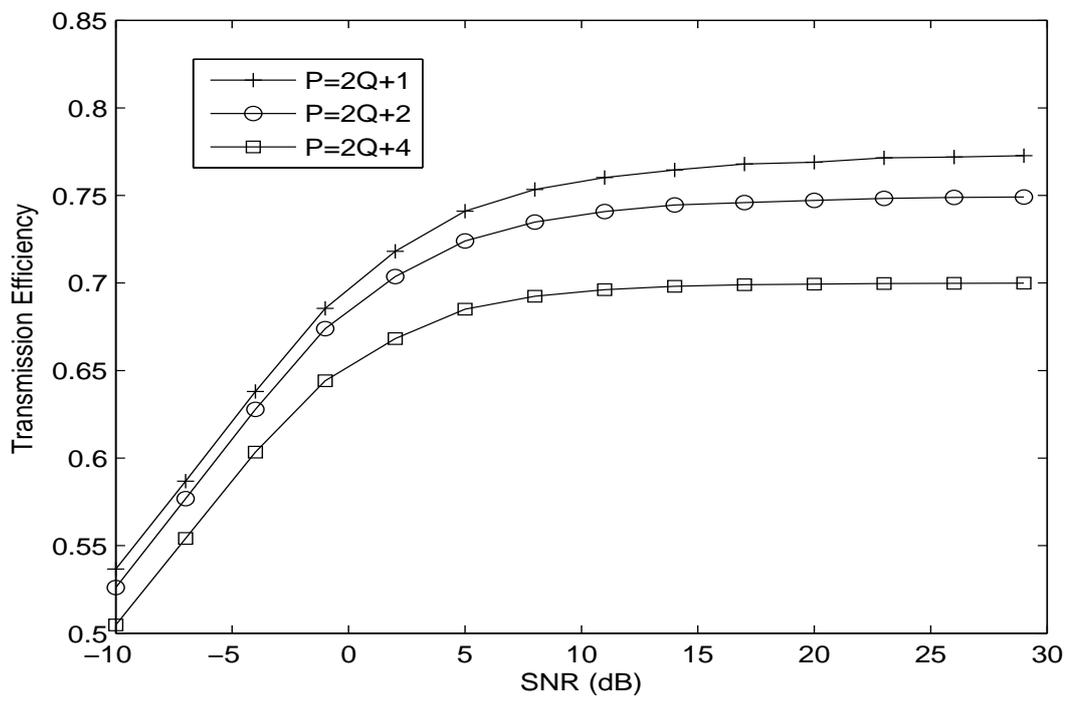


Fig. 5.10. Transmission efficiency versus SNR.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This thesis has developed new estimators and training sequence design for AF relay networks. The main contributions are given in Chapter 2 to 5.

Chapter 2 developed a joint CFO and channel estimator for OFDM-based AF TWRNs. To facilitate the estimation, two new ZP and CP transmission protocols were proposed. These protocols also maintain the carrier orthogonality. Both protocols lead to the same joint CFO and channel estimator, which was implemented by using the NLS algorithm. The performance of the estimator was analyzed in detail by proving its unbiasedness at high SNRs and by deriving the closed-form expression of its MSE.

Chapter 3 focused on estimating individual frequency and channel parameters since the method in Chapter 2 can only obtain convoluted channel parameters and the mixed CFO values. Three different algorithms were designed for the initial parameter estimation and an iterative algorithm to refine the initial estimation results. Simulations demonstrated the accuracy of the theoretical analysis. It was found that for the special case when the CFO between two terminals is small, the estimation MSE is very close to CRB in the high SNR region, and the estimates are good even without iterations.

Chapter 4 investigated the time-varying channel estimation and training sequence design for TWRNs. The CE-BEM was adapted to represent the TWRN time-varying channels. A new multi-round training scheme was suggested. Channel estimator was developed and two algorithms were suggested to recover the CE-BEM coefficients. The optimal training parameters, including the number of the pilot symbols, the placement of the pilot symbols, and the power allocation to the pilot symbols, were derived by minimizing the channel estimation MSE. It was shown that traditional one-round training scheme [14], [90] for TWRN is not optimal in the sense of minimizing the chan-

nel estimation MSE.

Chapter 5 considered the doubly-selective channel estimation and training sequence design for OWRNs. The CE-BEM was again used to represent the time-varying channel. An efficient method was proposed to estimate the channel and the optimal training sequence was derived that can minimize the MSE. Finally numerical examples were provided to corroborate the proposed studies.

6.2 Future Work

As mentioned in Chapter 1, a main challenge is to develop a doubly-selective channel estimator for TWRNs. This estimation problem is highly complicated and the CE-BEM used in this thesis for time-varying channels may not be accurate enough to represent doubly-selective channels. A potential solution is the DPS-BEM because it needs fewer basis functions to achieve the same MSE as that of CE-BEM. Developing the corresponding estimator and deriving the optimal training sequence will be a challenging research topic.

Moreover, it will be good to derive a MSE lower bound of the BEM approximation of an arbitrary time-varying channel. Such a bound will open up many problems. For example, if such a bound exists, how does one find the BEM which achieves performance close to the bound? For another instance, for the mobile-to-mobile channel in OWRN and TWRN [82], how does one find the BEM with the best performance and then develop channel estimation and data detection algorithms?

Our training sequence designs in OWRN and TWRN are based on minimizing the MSE. Optimal training sequence designs by other criteria, such as maximizing the lower bound of capacity or minimizing BER or maximizing SNR, are also worth studying. In [46], the capacity lower bound in the presence of channel estimation error is derived for the conventional point-to-point systems and then training sequence design is proposed by maximizing the lower bound. For TWRNs, the capacity lower bounds with channel estimation error remain unknown when the channel is either frequency-selective, time-varying or doubly-selective. Deriving capacity bounds for these three types of channels and designing training sequences are thus challenging problems.

In Chapter 4, to enable channel estimation, the relay forwards the received symbols with a delay but without changing the order of these symbols. In fact, the relay may reorder these symbols in order to minimize the BER or maximize the SNR. The design of an optimal mapping function is also a good choice for future work.

Furthermore, note that all of our work [41], [74], [91] about estimation of time-varying channels or doubly-selective channels for TWRNs and OWRNs are based on BEMs. However, the AR model is an alternative to approximate the time-varying channel [47]. When an AR model is chosen to represent the channel, Kalman filter is often utilized to estimate and track the channel [66], [92]. The reference [93] investigated the application of an AR model and Kalman filter for conventional

point-to-point systems. For OWRNs and TWRNs, however, the use of the AR model and Kalman filter to estimate the channel remains unexplored. The optimal training sequence design for such a scenario is also an open issue.

The comparison between the BEM and the AR model in approximating the time-varying channels is also a problem worth investigating. Which model can best approximate the channel, especially the OWRN channel and the TWRN channel? Is there a lower bound for approximation MSE? Clearly, this area remains largely unexplored.

Moreover, the optimal training sequence design for joint CFO and channel estimation for TWRNs is another challenging problem. The theoretical MSE derived in Chapter 2 may be chosen as the optimization criterion. Other criteria, such as minimizing CRB or maximizing SNR, may also be good choices and worth investigating.

Multiple-way relay networks (MWRNs) [96] is an extension of TWRNs. In a MWRN, multiple terminals want to exchange their independent information packets with the help of a relay. Channel estimation techniques for MWRN are an entirely open research area currently.

References

- [1] J. Laneman and G. Wornell, "Distributed space time block coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.
- [2] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [3] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "Multihop diversity in wireless relaying channels," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1820–1830, Oct. 2004.
- [4] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMax and LTE-advanced mobile systems", *IEEE Commun. Mag.*, vol. 47, no. 10, pp.100–105, Oct. 2009.
- [5] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp.371–381, Feb. 2003.
- [6] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, 1961, pp. 611–644.
- [7] B. Rankov and A. Wittneben, "Spectral efficient signaling for half-duplex relay channels," in *Proc. Annual Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, Oct. 2005, pp. 1066–1071.
- [8] S. Katti, S. Gollakota and D. Katabi, "Embracing wireless interference: analog network coding," Computer Science and Artificial Intelligence Laboratory Technical Report, MIT-CSAIL-TR-2007-012, Feb. 23, 2007.
- [9] S. Zhang, S.-C. Liew and P. P. Lam, "Physical-layer network coding," in *Proc. of MobiComm.*, Los Angeles, CA, Sept. 2006, pp. 358–365.
- [10] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.

- [11] T. Cui, F. Gao, T. Ho, and A. Nallanathan, "Distributed space-time coding for two-way wireless relay networks," *IEEE Trans. Signal Process.*, vol.57, no. 2, pp. 658–671, Feb. 2009.
- [12] R. Zhang, Y.-C. Liang, C. C. Chai, and S. G. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 699–712, June 2009.
- [13] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024–3033, Oct. 2009.
- [14] —, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4443–4455, Nov. 2009.
- [15] T. Pham, Y.-C. Liang, A. Nallanathan, and H. K. Garg, "Optimal training sequences for channel estimation in bi-directional relay networks with multiple antennas," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 474–479, Feb. 2010.
- [16] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. IEEE ISIT*, Seattle, USA, July 2006, pp. 1668–1672.
- [17] S. J. Kim, N. Devroye, P. Mitran, and V. Tarokh, "Achievable rate regions for bi-directional relaying," submitted to *IEEE Trans. Inf. Theory*, available at arXiv:0808.0954v1.
- [18] T. Cui, T. Ho, and J. Kliewer, "Memoryless relay strategies for two-way relay channels", *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3132–3143, Oct. 2009.
- [19] C. K. Ho, R. Zhang, and Y.-C. Liang, "Two-way relaying over OFDM: optimized tone permutation and power allocation," in *Proc. of IEEE ICC*, Beijing China, May 2008, pp. 3908–3912.
- [20] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay network," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, May 2008.
- [21] G. Wang and C. Tellambura, "Super-imposed pilot-aided channel estimation and power allocation for relay systems," in *Proc. IEEE WCNC*, Budapest, Hungary, Apr. 2009, pp. 1-5.
- [22] G. Wang, F. Gao, and C. Tellambura, "Effects of channel estimation errors on BPSK systems using superimposed pilots," in *Proc. of Int. Conf. on Wireless Commun., Networking and Inf. Security (WCNIS)*, Beijing, June, 2010, pp. 1-4.
- [23] G. T. Zhou, M. Viberg, and T. McKelvey, "First-order statistical method for channel estimation," *IEEE Signal Processing Lett.*, vol. 10, no. 3, pp. 57-60, Mar. 2003.

- [24] 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, March 2005.
- [25] J. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Commun. Lett.*, vol.7, no.9, pp.413-415, Sep. 2003.
- [26] X. Meng, J. Tugnait and S. He, "Iterative joint channel estimation and data detection using superimposed training: Algorithms and performance analysis," *IEEE Trans. Veh. Technol.*, vol.56, pp. 1873-1880, July 2007.
- [27] J. Tugnait and S. He, "Doubly-selective channel estimation using data-dependent superimposed training and exponential basis models," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 3877–3883, Nov. 2007.
- [28] S. He and J.Tugnait, "On doubly selective channel estimation using superimposed training and discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 56, pp.3214–3228, July 2008.
- [29] G. Wang, F. Gao, and C. Tellambura, "Channel Estimation with Amplitude Constraint: Superimposed Training or Conventional Training?" in *Proc. IEEE CWIT*, Kelowa, Canada, May. 2011, pp. 1-4.
- [30] V. Nguyen, H. Tuan, H.Nguyen, and N. Tran, "Optimal superimposed training design for spatially correlated fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3206-3217, Aug. 2008.
- [31] C. Carrasco-Alvarez, R. Parra-Michel, A. Orozco-Lugo, and J. Tugnait, "Enhanced channel estimation using superimposed training based on universal basis expansion," *IEEE Trans. Signal Process.*, vol. 57, pp. 1217-1222, March 2009.
- [32] M. Coldrey and P. Bohlin, "Training-based MIMO systems – Part I: performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp.5464–5476, Nov. 2006.
- [33] A. Gaston, W. Chriss, and E. Walker, "A multipath fading simulator for radio," *IEEE Trans. Veh. Technol.*, vol. VT-22, no. 4, pp. 241–244, 1973.
- [34] G. Wang, F. Gao, and C. Tellambura, "Superimposed pilots aided joint CFO and channel estimation for ZP-OFDM modulated two-way relay networks," in *Proc. IEEE Veh. Technol. Conf. (VTC) Fall*, Ottawa, Canada, Sept. 2010, pp. 1-5.

- [35] G. Wang, F. Gao, and C. Tellambura, "Superimposed pilot based joint CFO and channel estimation for CP-OFDM modulated two-way relay networks," in *Proc. IEEE Globecom*, Miami, Florida, USA, Dec.2010, pp. 1-5.
- [36] 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, pp. 1969–1986, Nov. 1998.
- [37] 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–Jun. 1978.
- [38] D. K. Borah and B. D. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, pp. 862–873, Jun. 1999.
- [39] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, pp.3597–3607, Sept. 2005.
- [40] M. Martone, "Wavelet-based separating kernels for sequences estimation with unknown rapidly time-varying channels," *IEEE Commun. Lett.* vol. 3, no. 3, pp. 78–80, Mar, 1999.
- [41] G. Wang, F. Gao, W. Chen, and C. Tellambura, "Channel estimation for two-way relay networks under time-selective environment," in *Proc. IEEE ICC*, Kyoto, Japan, June 2011, pp. 1-6.
- [42] X. Ma and G. B. Giannakis, "Maximum-diversity transmissions over doubly-selective wireless channels," *IEEE Trans. Inf. Theory*, vol.49, pp. 1832-1840, Jul. 2003.
- [43] X. Ma, G. B. Giannakis and S. Ohno, "Optimal training for block transmission over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351-1366, May 2003.
- [44] A. S. Akki and F. Haber, "A statistical model for mobile-to-mobile land communication channel," *IEEE Trans. Veh. Technol.*, vol. VT-35, no. 1, pp.2–7, Feb. 1986.
- [45] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Tech.*, vol. 40, pp. 686-693, November 1991.
- [46] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951-963, April 2003.
- [47] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Processing Mag.*, vol. 21, no. 6, pp. 12-25, 2004.

- [48] W. Zhang, X. G. Xia, and P. C. Ching, "Optimal training and pilot pattern design for OFDM systems in Rayleigh fading," *IEEE Trans. Broadcast.*, vol.52, no.4, pp. 505-514, Dec. 2006.
- [49] M. Dong, L. Tong, and B. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403-1418, May 2004.
- [50] S. Misra, A. Swami, and L. Tong, "Optimal training for time-selective wireless fading channels using cutoff rate," *EURASIP Journal on Applied Signal Processing: Special Issue on Reliable Communications over Rapidly Time-Varying Channels*, vol. 2006, pp. 1-15, 2006.
- [51] C. B. Peel, and A. L. Swindlehurst, "Throughput-optimal training for a time-varying multi-antenna channel," *IEEE Trans. Wireless Commun.*, vol. 6, no.9, Sept. 2007.
- [52] M. Dong and L. Tong, "Optimal Design and Placement of Pilot Symbols for Channel Estimation," *IEEE Trans. Signal Process.*, vol. 50, pp. 3055-3069, December 2002.
- [53] S. Ohno, G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol.50, no. 9, pp. 2138-2145, Sept. 2004.
- [54] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of known symbols for frequency selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2338-2353, August 2002.
- [55] P. Ciblat and L. Vandendorpe, "On the maximum-likelihood based data-aided frequency offset and channel estimate," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Toulouse, France, Sept. 2002, vol.1, pp.627-630.
- [56] M. Ghogho, A. Swami, "Training design for multipath channel and frequency-offset estimation in MIMO systems," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3957-3965, Oct. 2006.
- [57] P. Ciblat, P. Bianchi, and M. Ghogho, "Training sequence optimization for joint channel and frequency offset estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3424-3436, Aug. 2008.
- [58] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1580-1589, Sept. 2000.
- [59] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun.*, vol. 51, no.11, pp.1910-1917, Nov. 2003.

- [60] U. Tureli, Didem Kivanc, and H. Liu, "Experimental and analytical studies on a high resolution OFDM carrier estimator," *IEEE Trans. Veh. Tech.*, vol. 50, no. 2, pp. 629–643, Mar. 2001.
- [61] A. P. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1148–1160, Mar. 2008.
- [62] T. Whitworth, M. Ghogho, and D. McLernon, "Optimized training and basis expansion model parameters for doubly-selective channel estimation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1490–1498, Mar. 2009.
- [63] W. C. Jakes, *Microwave Mobile Communications*, New York: Wiley, 1974.
- [64] D. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge University Press, 2005.
- [65] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [66] S. M. Kay, *Fundamentals of Statistical Signal Processing*, vol. 1: Estimation Theory, Prentice Hall PTR, 1993.
- [67] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [68] G. Wang, F. Gao, and C. Tellambura, "Joint frequency offset and channel estimation methods for two-way relay networks," in *Proc. IEEE GLOBECOM*, Honolulu, Hawaii, Nov./Dec. 2009, pp. 1-5.
- [69] G. Wang, F. Gao, Y.-C. Wu, and C. Tellambura, "Joint CFO and channel estimation for CP-OFDM modulated two-way relay networks," in *Proc. IEEE ICC*, Cape Town, South Africa, May 2010, pp. 1-5.
- [70] G. Wang, F. Gao, Y.-C. Wu, and C. Tellambura, "Joint CFO and channel estimation for ZP-OFDM modulated two-way Relay networks," in *Proc. IEEE WCNC*, Sydney, Australia, Apr. 2010, , pp. 1-5.
- [71] 3GPP TR 36.814 V1.2.1, "Further advancements for EUTRA: physical layer aspects," Tech. Spec.n Group Radio Access Network Rel. 9, June 2009.
- [72] IEEE P802.16j/D9, "Draft amendment to IEEE standard for local and metropolitan area networks Part 16: air interface for fixed and mobile broadband wireless access systems: multihop relay specification," Feb. 2009.

- [73] “Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: high speed physical layer in the 5 GHz band,” IEEE802.11a, 1999.
- [74] G. Wang, F. Gao, and C. Tellambura, “BEM-based estimation for time-varying channels and training design in two-way relay networks,” in *Proc. IEEE Globecom*, Miami, Florida, USA, Dec. 2010, pp. 1-5.
- [75] G. Wang, F. Gao, G. Li, and C. Tellambura, “Channel estimation and data-dependent superimposed training sequence design in amplify-and-forward relay networks,” in *Proc. IEEE WCNC*, Cancun, Mexico, Mar. 2011, pp. 1-5.
- [76] Z. Wang and G. B. Giannakis., “Wireless multicarrier communications: where Fourier meets Shannon,” *IEEE Signal Processing Mag.*, vol. 17, no. 3, pp. 29–48, May 2000.
- [77] F. Gao, T. Cui, and A. Nallanathan, “Scattered pilots and virtual carriers based frequency offset tracking for OFDM systems: algorithms, identifiability, and performance analysis,” *IEEE Trans. Commun.*, vol. 56, no. 4, pp. 619–629, Apr. 2008.
- [78] F. Gao, Y. Zeng, A. Nallanathan, and T. S. Ng, “Robust subspace blind channel estimation for cyclic prefixed MIMO OFDM systems: algorithm, identifiability and performance analysis,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 378–388, Feb. 2008.
- [79] F. Gao and A. Nallanathan, “Blind maximum likelihood CFO estimation for OFDM systems via polynomial rooting,” *IEEE Signal Process. Lett.*, vol. 13, no. 2, pp. 73–76, Feb. 2006.
- [80] M.-O. Pun, M. Morelli, and C.-C. Jay Kuo, “Maximum-likelihood synchronization and channel estimation for OFDMA uplink transmissions,” *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 726–736, Apr. 2006.
- [81] C. S. Patel and G. L. Stuber, “Channel estimation for amplify and forward relay based cooperation diversity systems,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 3348–3356, June 2007.
- [82] C. S. Patel, G. L. Stuber and T. G. Pratt, “Statistical properties of amplify and forward relay fading channels”, *IEEE Trans. Veh. Technol.*, vol. 55, no. 1, pp. 1–9, Jan. 2006.
- [83] S. Savazzi and U. Spagnolini, “Optimizing training lengths and training intervals in time-varying fading channels,” *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1098–1112, Mar. 2009.
- [84] M. Sandell, D. McNamara, and S. Parker, “Analysis of frequency-offset tracking in MIMO OFDM systems,” *IEEE Trans. Commun.*, vol. 54, no. 8, pp. 1481–1491, Aug. 2006.

- [85] H. Liu and G. B. Giannakis, "Deterministic approaches for blind equalization of time-varying channels with antenna arrays," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3003–3013, Nov. 1998.
- [86] A. S. Lalos, A. A. Rontogiannis, K. Berberidis, "Frequency domain channel estimation for cooperative communication networks," *IEEE Trans. Signal Process.*, vol.58, no.6, pp.3400-3405, June 2010.
- [87] M. G. Song, D. Kim, G. H. Im, "Recursive channel estimation method for OFDM-based Cooperative Systems," *IEEE Commun. Lett.*, vol.14, no.11, pp.1029–1031, November 2010.
- [88] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol.50, no. 9, pp. 2138–2145, Sept. 2004.
- [89] G. Wang, F. Gao, X. Zhang, and C. Tellambura, "Superimposed training based joint CFO and channel estimation for CP-OFDM modulated two-way relay networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, ID 403936, pp. 1-9, July 2010.
- [90] G. Wang, F. Gao, Y.-C. Wu, and C. Tellambura, "Joint carrier frequency offset and channel estimation for two-way relay networks," *IEEE Trans. Wireless Commun.*, vol.10, no. 2, pp.456-465, Feb. 2011.
- [91] G. Wang, F. Gao, W. Chen, and C. Tellambura, "Channel estimation and training design for two-Way relay networks in time-selective fading environment," to appear in *IEEE Trans. Wireless Commun.* (accepted in April 2011).
- [92] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *Signal Processing*, vol. 53, no. 2/3, pp. 211–228, 1996.
- [93] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1065–1076, May 2002.
- [94] Z. Tang, R. 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.
- [95] F. Gao, B. Jiang, X. Gao, and X. Zhang, "Superimposed Training Based Channel Estimation for OFDM Modulated Amplify-and-Forward Relay Networks," *IEEE Trans. Commun.*, vol. 59, no. 7, pp.2029-2039, July 2011.

- [96] D. Gündüz, A. Yener, A. Goldsmith, and H. V. Poor, “A new achievable rate for the Gaussian parallel relay channel,” in *Proc. IEEE Int. Symposium. On. Inf. Theory (ISIT)*, Seoul, Korea, June 2009, pp. 339–408.

Appendix A

Proof of Lemma 2.2

From [77], we know that

$$\Delta v \approx -\frac{\frac{\partial g(v)}{\partial v}|_{v=v_0}}{\frac{\partial^2 g(v)}{\partial v^2}|_{v=v_0}} = -\frac{\dot{g}(v_0)}{\ddot{g}(v_0)}. \quad (\text{A-1})$$

The first order derivative of \mathbf{G} can be calculated as

$$\dot{\mathbf{G}} = \frac{\partial \mathbf{G}}{\partial v} = j\mathbf{J}^H \mathbf{D}\Gamma \mathbf{S}_2. \quad (\text{A-2})$$

Applying the equality

$$\frac{\partial \Phi^{-1}}{\partial v} = -\Phi^{-1} \frac{\partial \Phi}{\partial v} \Phi^{-1} = -\Phi^{-1} (\dot{\mathbf{G}}^H \mathbf{G} + \mathbf{G}^H \dot{\mathbf{G}}) \Phi^{-1}, \quad (\text{A-3})$$

we get

$$\begin{aligned} \dot{g}(v) = & \mathbf{y}_n^H \dot{\mathbf{P}}_G \mathbf{y}_n = \underbrace{\mathbf{y}_n^H \dot{\mathbf{G}} \Phi^{-1} \mathbf{G}^H \mathbf{y}_n}_{M_1} + \underbrace{\mathbf{y}_n^H \mathbf{G} \Phi^{-1} \dot{\mathbf{G}}^H \mathbf{y}_n}_{M_2} \\ & - \underbrace{(\mathbf{y}_n^H \mathbf{G} \Phi^{-1} \dot{\mathbf{G}}^H \mathbf{G} \Phi^{-1} \mathbf{G}^H \mathbf{y}_n + \mathbf{y}_n^H \mathbf{G} \Phi^{-1} \mathbf{G}^H \dot{\mathbf{G}} \Phi^{-1} \mathbf{G}^H \mathbf{y}_n)}_{M_3}. \end{aligned} \quad (\text{A-4})$$

It can be found that

$$\begin{aligned} E(M_1(v_0)) = & jE[(\mathbf{G}\mathbf{b} + \mathbf{n})^H \mathbf{J}^H \mathbf{D}\Gamma \mathbf{S}_2 \Phi^{-1} \mathbf{G}^H (\mathbf{G}\mathbf{b} + \mathbf{n})] \\ = & jE[\mathbf{b}^H \mathbf{G}^H \mathbf{J}^H \mathbf{D}\Gamma \mathbf{S}_2 \mathbf{b}] + jE[\mathbf{n}^H \mathbf{J}^H \mathbf{D}\Gamma \mathbf{S}_2 \Phi^{-1} \mathbf{G}^H \mathbf{n}] \end{aligned} \quad (\text{A-5})$$

$$E(M_2(v_0)) = -jE[\mathbf{b}^H \mathbf{S}_2^H \mathbf{D}\Gamma^H \mathbf{J} \mathbf{G} \mathbf{b}] - jE[\mathbf{n}^H \mathbf{G} \Phi^{-1} \mathbf{S}_2^H \mathbf{D}\Gamma^H \mathbf{J} \mathbf{n}] \quad (\text{A-6})$$

$$\begin{aligned} E(M_3(v_0)) = & E[\mathbf{b}^H (-j\mathbf{S}_2 \mathbf{D}\Gamma^H \mathbf{J} \mathbf{G} + j\mathbf{G}^H \mathbf{J}^H \mathbf{D}\Gamma \mathbf{S}_2) \mathbf{b}] \\ & + E[\mathbf{n}^H \mathbf{G} \Phi^{-1} (-j\mathbf{S}_2 \mathbf{D}\Gamma^H \mathbf{J} \mathbf{G} + j\mathbf{G}^H \mathbf{J}^H \mathbf{D}\Gamma \mathbf{S}_2) \Phi^{-1} \mathbf{G}^H \mathbf{n}]. \end{aligned} \quad (\text{A-7})$$

Combining (A-5), (A-6), (A-7) and (A-4), we get

$$\begin{aligned}
E[\dot{g}(v_0)] &= jE[\mathbf{n}^H (\mathbf{I} - \mathbf{G}\Phi^{-1}\mathbf{G}^H)\mathbf{J}^H \mathbf{D}\Gamma \mathbf{S}_2 \Phi^{-1} \mathbf{G}^H \mathbf{n}] \\
&\quad - jE[\mathbf{n}^H \mathbf{G}\Phi^{-1} \mathbf{S}_2^H \mathbf{D}\Gamma^H \mathbf{J} (\mathbf{I} - \mathbf{G}\Phi^{-1}\mathbf{G}^H) \mathbf{n}] \\
&= -2\sigma_{ne}^2 \mathfrak{S} \left\{ \underbrace{\text{tr} (\mathbf{G}^H (\mathbf{I} - \mathbf{G}\Phi^{-1}\mathbf{G}^H)\mathbf{J}^H \mathbf{D}\Gamma \mathbf{S}_2 \Phi^{-1})}_{0} \right\} = 0. \tag{A-8}
\end{aligned}$$

In order to get $\ddot{g}(v_0)$, we need to compute the first-order derivative of M_1 , M_2 , and M_3 as

$$\dot{M}_1 = \mathbf{y}_n^H \ddot{\mathbf{G}}\Phi^{-1} \mathbf{G}^H \mathbf{y}_n - \mathbf{y}_n^H \dot{\mathbf{G}}\Phi^{-1} \dot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{y}_n + \mathbf{y}_n^H \dot{\mathbf{G}}\Phi^{-1} \dot{\mathbf{G}}^H \mathbf{y}_n \tag{A-9}$$

$$\dot{M}_2 = \mathbf{y}_n^H \dot{\mathbf{G}}\Phi^{-1} \dot{\mathbf{G}}^H \mathbf{y}_n - \mathbf{y}_n^H \mathbf{G}\Phi^{-1} \dot{\Phi}\Phi^{-1} \dot{\mathbf{G}}^H \mathbf{y}_n + \mathbf{y}_n^H \mathbf{G}\Phi^{-1} \ddot{\mathbf{G}}^H \mathbf{y}_n \tag{A-10}$$

$$\begin{aligned}
\dot{M}_3 &= \mathbf{y}_n^H \dot{\mathbf{G}}\Phi^{-1} \dot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{y}_n - \mathbf{y}_n^H \mathbf{G}\Phi^{-1} \dot{\Phi}\Phi^{-1} \dot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{y}_n \\
&\quad + \mathbf{y}_n^H \mathbf{G}\Phi^{-1} \ddot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{y}_n - \mathbf{y}_n^H \mathbf{G}\Phi^{-1} \dot{\Phi}\Phi^{-1} \dot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{y}_n + \mathbf{y}_n^H \dot{\mathbf{G}}\Phi^{-1} \dot{\Phi}\Phi^{-1} \dot{\mathbf{G}}^H \mathbf{y}_n. \tag{A-11}
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
\ddot{g}(v_0) &= \dot{M}_1 + \dot{M}_2 - \dot{M}_3 - \dot{M}_4 \\
&= \mathbf{y}_n^H \ddot{\mathbf{G}}\Phi^{-1} \mathbf{G}^H \mathbf{y}_n + \mathbf{y}_n^H \mathbf{G}\Phi^{-1} \ddot{\mathbf{G}}^H \mathbf{y}_n + 2\mathbf{y}_n^H \dot{\mathbf{G}}\Phi^{-1} \dot{\mathbf{G}}^H \mathbf{y}_n \\
&\quad - 2\mathbf{y}_n^H \dot{\mathbf{G}}\Phi^{-1} \dot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{y}_n - 2\mathbf{y}_n^H \mathbf{G}\Phi^{-1} \dot{\Phi}\Phi^{-1} \dot{\mathbf{G}}^H \mathbf{y}_n \\
&\quad - \mathbf{y}_n^H \mathbf{G}\Phi^{-1} \ddot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{y}_n + 2\mathbf{y}_n^H \mathbf{G}\Phi^{-1} \dot{\Phi}\Phi^{-1} \dot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{y}_n \\
&= \mathbf{b}^H \mathbf{G}^H \ddot{\mathbf{G}}\mathbf{b} + \mathbf{n}^H \ddot{\mathbf{G}}\Phi^{-1} \mathbf{G}^H \mathbf{n} + \mathbf{b}^H \ddot{\mathbf{G}}\mathbf{G}^H \mathbf{b} + \mathbf{n}^H \mathbf{G}\Phi^{-1} \ddot{\mathbf{G}}^H \mathbf{n} \\
&\quad + 2\mathbf{b}^H \mathbf{G}^H \dot{\mathbf{G}}\Phi^{-1} \dot{\mathbf{G}}^H \mathbf{G}\mathbf{b} + 2\mathbf{n}^H \dot{\mathbf{G}}\Phi^{-1} \dot{\mathbf{G}}^H \mathbf{n} - 2\mathbf{b}^H \mathbf{G}^H \dot{\mathbf{G}}\Phi^{-1} \dot{\Phi}\mathbf{b} - \mathbf{b}^H \ddot{\Phi}\mathbf{b} \\
&\quad - 2\mathbf{n}^H \dot{\mathbf{G}}\Phi^{-1} \dot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{n} - 2\mathbf{b}^H \dot{\Phi}\Phi^{-1} \dot{\mathbf{G}}^H \mathbf{G}\mathbf{b} - 2\mathbf{n}^H \mathbf{G}\Phi^{-1} \dot{\Phi}\Phi^{-1} \dot{\mathbf{G}}^H \mathbf{n} \\
&\quad - \mathbf{n}^H \mathbf{G}\Phi^{-1} \ddot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{n} + 2\mathbf{b}^H \dot{\Phi}\Phi^{-1} \dot{\Phi}\mathbf{b} + 2\mathbf{n}^H \mathbf{G}\Phi^{-1} \dot{\Phi}\Phi^{-1} \dot{\Phi}\Phi^{-1} \mathbf{G}^H \mathbf{n} \tag{A-12}
\end{aligned}$$

After some tedious simplification, it can be obtained that

$$E[\ddot{g}(v_0)] = 2\mathbf{b}^H \dot{\mathbf{G}}^H (\mathbf{G}\Phi^{-1}\mathbf{G}^H - \mathbf{I}) \dot{\mathbf{G}}\mathbf{b}, \tag{A-13}$$

and $\ddot{g}(v_0)$ can be written as

$$\ddot{g}(v_0) = E\{\ddot{g}(v_0)\} + \mathcal{O}_2(\mathbf{n}) + \mathcal{O}_2(\mathbf{n}^2), \tag{A-14}$$

where $\mathcal{O}_2(\mathbf{n})$ and $\mathcal{O}_2(\mathbf{n}^2)$ represent the linear and quadrature functions of \mathbf{n} in $\ddot{g}(v_0)$, whose explicit forms are omitted for the sake of brevity.

Similarly, $\dot{g}(v_0)$ can be expressed as

$$\dot{g}(v_0) = \mathcal{O}_1(\mathbf{n}) + \mathcal{O}_1(\mathbf{n}^2), \quad (\text{A-15})$$

where $\mathcal{O}_1(\mathbf{n})$ and $\mathcal{O}_1(\mathbf{n}^2)$ represent the linear and quadrature functions of \mathbf{n} existing in $\dot{g}(v_0)$. Substituting (A-15) and (A-14) into (A-1) gives

$$\Delta v \approx -\frac{\mathcal{O}_1(\mathbf{n}) + \mathcal{O}_1(\mathbf{n}^2)}{\mathbf{E}\{\dot{g}(v_0)\} + \mathcal{O}_2(\mathbf{n}) + \mathcal{O}_2(\mathbf{n}^2)} \approx -\frac{\mathcal{O}_1(\mathbf{n}) + \mathcal{O}_1(\mathbf{n}^2)}{\mathbf{E}\{\dot{g}(v_0)\}} = -\frac{\dot{g}(v_0)}{\mathbf{E}\{\dot{g}(v_0)\}}. \quad (\text{A-16})$$

Appendix B

Proof of Theorem 2.1

According to Lemma (2.2), the MSE of the CFO estimation is

$$E\{\Delta v^2\} = \frac{E\{\dot{g}(v_0)^2\}}{E\{\dot{g}(v_0)\}^2}. \quad (\text{B-1})$$

The numerator can be computed as

$$\begin{aligned} E[\dot{g}(v)^2] &= E[\mathbf{y}_n^H \dot{\mathbf{P}}_G \mathbf{y}_n \mathbf{y}_n^H \dot{\mathbf{P}}_G \mathbf{y}_n] = E[\mathbf{y}_n^H \dot{\mathbf{P}}_G (\mathbf{n}\mathbf{n}^H + \mathbf{G}\mathbf{b}\mathbf{b}^H \mathbf{G}^H) \dot{\mathbf{P}}_G \mathbf{y}_n] \\ &= \sigma_{ne}^2 E[\mathbf{n}^H \dot{\mathbf{P}}_G \dot{\mathbf{P}}_G \mathbf{n}] + \sigma_{ne}^2 E[\mathbf{b}^H \mathbf{G}^H \dot{\mathbf{P}}_G \dot{\mathbf{P}}_G \mathbf{G}\mathbf{b}] \\ &\quad + E[\mathbf{n}^H \dot{\mathbf{P}}_G \mathbf{G}\mathbf{b}\mathbf{b}^H \mathbf{G}^H \dot{\mathbf{P}}_G \mathbf{n}] + E[\mathbf{b}^H \mathbf{G}^H \dot{\mathbf{P}}_G \mathbf{G}\mathbf{b}\mathbf{b}^H \mathbf{G}^H \dot{\mathbf{P}}_G \mathbf{G}\mathbf{b}], \end{aligned} \quad (\text{B-2})$$

where

$$\dot{\mathbf{P}}_G = \dot{\mathbf{G}}\Phi^{-1}\mathbf{G}^H + \mathbf{G}\Phi^{-1}\dot{\mathbf{G}}^H - \mathbf{G}\Phi^{-1}\dot{\Phi}\Phi^{-1}\mathbf{G}^H. \quad (\text{B-3})$$

At high SNR, the first term in (B-2) can be neglected, and the last term is $\mathbf{0}$ because

$$\mathbf{G}^H \dot{\mathbf{P}}_G \mathbf{G} = \mathbf{0}. \quad (\text{B-4})$$

Moreover, the second and the third term are the same. After some tedious computation, we obtain

$$\mathbf{G}^H \dot{\mathbf{P}}_G \dot{\mathbf{P}}_G \mathbf{G} = \dot{\mathbf{G}}^H [\mathbf{I} - \mathbf{G}\Phi^{-1}\mathbf{G}^H] \dot{\mathbf{G}}. \quad (\text{B-5})$$

Therefore, (B-2) can be rewritten as

$$E[\dot{g}(v)^2] = 2\sigma_{ne}^2 \mathbf{b}^H \dot{\mathbf{G}}^H [\mathbf{I} - \mathbf{G}\Phi^{-1}\mathbf{G}^H] \dot{\mathbf{G}} \mathbf{b}. \quad (\text{B-6})$$

Substituting (A-13) and (B-6) into (B-1), we proved Theorem (2.1).

Appendix C

Proof of Theorem 2.2

From (2.31), we know

$$\hat{\mathbf{b}} = (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^H \mathbf{J}^H \mathbf{y} = (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^H (\mathbf{G} \mathbf{b} + \mathbf{n}), \quad (\text{C-1})$$

where $\hat{\mathbf{G}} = \mathbf{J}^H \hat{\mathbf{\Gamma}} \mathbf{S}_2$, and $\hat{\mathbf{\Gamma}}$ is defined in (2.34).

From Taylor's expansion, we know

$$e^{jm\hat{v}_0} = e^{jmv_0} + jme^{jmv_0} \Delta v - m^2 e^{jmv_0} \Delta v^2 + \dots \quad (\text{C-2})$$

Then, (2.34) can be expressed as

$$\hat{\mathbf{\Gamma}} = \mathbf{\Gamma} + j\mathbf{D}\mathbf{\Gamma}\Delta v - \mathbf{D}^2\mathbf{\Gamma}\Delta v^2 + \dots \quad (\text{C-3})$$

At high SNR, the higher order statistics can be omitted, and $\hat{\mathbf{G}}$ can be rewritten as

$$\hat{\mathbf{G}} \approx \mathbf{J}^H (\mathbf{\Gamma} + j\mathbf{D}\mathbf{\Gamma}\Delta v) \mathbf{S}_2 = \mathbf{G} + \dot{\mathbf{G}}\Delta v. \quad (\text{C-4})$$

Substituting (C-4) into (C-1), we obtain that

$$\hat{\mathbf{b}} = \mathbf{b} - (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^H \dot{\mathbf{G}}\Delta v \mathbf{b} + (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^H \mathbf{n}. \quad (\text{C-5})$$

At high SNR, using the approximation $(\mathbf{I} + \Delta\mathbf{X})^{-1} \approx \mathbf{I} - \Delta\mathbf{X}$ for a positive semi-definite matrix [65] and omitting the higher order statistics, we obtain

$$(\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \approx (\mathbf{G}^H \mathbf{G})^{-1} - \mathbf{\Phi}^{-1} \dot{\mathbf{\Phi}} \mathbf{\Phi}^{-1} \Delta v. \quad (\text{C-6})$$

Then we can rewrite (C-5) as

$$\hat{\mathbf{b}} \approx \mathbf{b} - (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H \dot{\mathbf{G}} \Delta v \mathbf{b} + (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^H \mathbf{n}. \quad (\text{C-7})$$

Therefore,

$$\mathbb{E}\{\Delta \mathbf{b}\} = \mathbb{E}\{\hat{\mathbf{b}} - \mathbf{b}\} = \mathbb{E}\{-(\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H \dot{\mathbf{G}} \Delta v \mathbf{b} + (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^H \mathbf{n}\} = \mathbf{0}, \quad (\text{C-8})$$

$$\begin{aligned} \mathbb{E}\{\Delta \mathbf{b} \Delta \mathbf{b}^H\} &= \mathbb{E}\{(\hat{\mathbf{b}} - \mathbf{b})(\hat{\mathbf{b}} - \mathbf{b})^H\} \\ &= \mathbb{E}\{(\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H \dot{\mathbf{G}} \mathbf{b} \mathbf{b}^H \dot{\mathbf{G}}^H \mathbf{G} (\mathbf{G}^H \mathbf{G})^{-1} (\Delta v)^2\} + \sigma_{ne}^2 \mathbb{E}\{(\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1}\}. \end{aligned} \quad (\text{C-9})$$

Using (C-6), we can rewrite (C-9) as

$$\mathbb{E}\{\Delta \mathbf{b} \Delta \mathbf{b}^H\} = (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H \dot{\mathbf{G}} \mathbf{b} \mathbf{b}^H \dot{\mathbf{G}}^H \mathbf{G} (\mathbf{G}^H \mathbf{G})^{-1} \mathbb{E}\{\Delta v^2\} + \sigma_{ne}^2 (\mathbf{G}^H \mathbf{G})^{-1} \quad (\text{C-10})$$

Appendix D

Gradient Decent Method with Complex Variables

Let us define a new $2(Q + 1) \times 1$ real vector $\boldsymbol{\rho} = [\Re\{\boldsymbol{\lambda}\}^T, \Im\{\boldsymbol{\lambda}\}^T]^T$. The gradient decent method to update $\boldsymbol{\rho}$ can be directly obtained as [67]

$$\boldsymbol{\rho}^{(i+1)} = \boldsymbol{\rho}^{(i)} - \epsilon' \left. \frac{\partial \zeta}{\partial \boldsymbol{\rho}} \right|_{\boldsymbol{\rho}=\boldsymbol{\rho}^{(i)}}, \quad (\text{D-1})$$

or equivalently,

$$\begin{bmatrix} \Re\{\boldsymbol{\lambda}\}^{(i+1)} \\ \Im\{\boldsymbol{\lambda}\}^{(i+1)} \end{bmatrix} = \begin{bmatrix} \Re\{\boldsymbol{\lambda}\}^{(i)} \\ \Im\{\boldsymbol{\lambda}\}^{(i)} \end{bmatrix} - \epsilon' \left. \begin{bmatrix} \frac{\partial \zeta}{\partial \Re\{\boldsymbol{\lambda}\}} \\ \frac{\partial \zeta}{\partial \Im\{\boldsymbol{\lambda}\}} \end{bmatrix} \right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(i)}}. \quad (\text{D-2})$$

There is

$$\begin{aligned} \boldsymbol{\lambda}^{(i+1)} &= \Re\{\boldsymbol{\lambda}\}^{(i+1)} + j\Im\{\boldsymbol{\lambda}\}^{(i+1)} \\ &= \underbrace{\Re\{\boldsymbol{\lambda}\}^{(i)} + j\Im\{\boldsymbol{\lambda}\}^{(i)}}_{\boldsymbol{\lambda}^{(i)}} - \epsilon' \underbrace{\left(\frac{\partial \zeta}{\partial \Re\{\boldsymbol{\lambda}\}} + j \frac{\partial \zeta}{\partial \Im\{\boldsymbol{\lambda}\}} \right)}_{2 \frac{\partial \zeta}{\partial \boldsymbol{\lambda}^*}} \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(i)}}, \end{aligned} \quad (\text{D-3})$$

where the definition of the complex derivative [67] is used. Setting $\epsilon = 2\epsilon'$ yields (4.33).

Appendix E

Proof of Theorem 4.1

Let us first prove the following lemma:

Lemma E.1. If Z and R are co-prime, then the index set $\mathcal{I} = \{\langle m + uR \rangle_Z\}_{u=0}^{Z-1}$ is the same as the universal set $\{0, \dots, Z-1\}$, or equivalently,

$$\langle m + uR \rangle_Z \neq \langle m + vR \rangle_Z, \quad \text{for } 0 \leq u < v \leq Z-1. \quad (\text{E-1})$$

Proof. Let us first assume the contrary holds; i.e.,

$$\langle m + uR \rangle_Z = \langle m + vR \rangle_Z, \quad \exists u < v. \quad (\text{E-2})$$

Then we know

$$(v - u)R = kZ \quad (\text{E-3})$$

for some integer $k \neq 0$. Since Z and R are co-prime, their least common multiple must be ZR . However in (E-3), $(v - u) < Z$, so the equality (E-3) cannot hold. By the contradiction, we prove Lemma E.1. \square

Define a new variable

$$c = \left(\prod_{i=0}^{Z-1} c_i \right)^{1/2} = I_s \prod_{i=0}^{Z-1} \tilde{\lambda}_i. \quad (\text{E-4})$$

When Z is odd, the denominator in (4.40) can be expanded as

$$\prod_{i=0}^{\frac{Z-3}{2}} c_{\langle m+(2i+1)R \rangle_Z} = \prod_{i=0}^{\frac{Z-3}{2}} \tilde{\lambda}_{\langle m+(2i+1)R \rangle_Z} \tilde{\lambda}_{\langle m+(2i+2)R \rangle_Z} = \prod_{i=1}^{Z-1} \tilde{\lambda}_{\langle m+iR \rangle_Z} = \prod_{\substack{i=0 \\ i \neq m}}^{Z-1} \tilde{\lambda}_i, \quad (\text{E-5})$$

where Lemma 1 is applied in the last equality.

Dividing (E-4) by (E-5) proves Theorem 1, where I_s serves as SSA for all $\tilde{\lambda}_m$.

Appendix F

Proof of Lemma 5.3

The interference during the data detection can be expressed as

$$\mathbf{v} = \alpha \Delta(\mathbf{G}_s \mathbf{H}_s) \mathbf{s} + \mathbf{w}_s + \alpha \Delta(\mathbf{G}_s \mathbf{H}_{\bar{b}}) \bar{\mathbf{b}} + \alpha \Delta(\mathbf{G}_{\bar{b}} \mathbf{H}_{\check{y}}) \check{\mathbf{b}}, \quad (\text{F-1})$$

where $\Delta(\cdot)$ denotes the estimation error of the inside item. The correlation function of the interference \mathbf{v} is given by

$$\begin{aligned} \mathbf{R}_v = & \alpha^2 \mathcal{P}_s E(\Delta(\mathbf{G}_s \mathbf{H}_s) \Delta(\mathbf{G}_s \mathbf{H}_s)^H) + E(\mathbf{w}_s \mathbf{w}_s^H) \\ & + \alpha^2 E(\Delta(\mathbf{G}_s \mathbf{H}_{\bar{b}}) \bar{\mathbf{b}} \bar{\mathbf{b}}^H \Delta(\mathbf{G}_s \mathbf{H}_{\bar{b}})^H) + \alpha^2 E(\Delta(\mathbf{G}_{\bar{b}} \mathbf{H}_{\check{y}}) \check{\mathbf{b}} \check{\mathbf{b}}^H \Delta(\mathbf{G}_{\bar{b}} \mathbf{H}_{\check{y}})^H), \end{aligned} \quad (\text{F-2})$$

where \mathcal{P}_s is the power allocated to the data sequence. We need to find the training scheme that can minimize the trace of \mathbf{R}_v .

Suppose there are two training schemes with identical $E((\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})^H)$. Thus the first and the second item in (F-2) is the same for both training schemes. If the training scheme has the first $2L$ and the last $2L$ entries equal to zero, the third and fourth item in (F-2) will become zero; if the training scheme does not has such condition, then it cannot null these two semi-definite items.