

NEW ADVANCES IN SYMBOL TIMING SYNCHRONIZATION OF  
SINGLE-CARRIER, MULTI-CARRIER AND SPACE-TIME  
MULTIPLE-ANTENNA SYSTEMS

A Dissertation

by

YIK CHUNG WU

Submitted to the Office of Graduate Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of  
DOCTOR OF PHILOSOPHY

August 2005

Major Subject: Electrical Engineering

NEW ADVANCES IN SYMBOL TIMING SYNCHRONIZATION OF  
SINGLE-CARRIER, MULTI-CARRIER AND SPACE-TIME  
MULTIPLE-ANTENNA SYSTEMS

A Dissertation

by

YIK CHUNG WU

Submitted to the Office of Graduate Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Erchin Serpedin
Committee Members,	Costas N. Georghiades
	Andrew K. Chan
	Siu A. Chin
Head of Department,	Chanan Singh

August 2005

Major Subject: Electrical Engineering

## ABSTRACT

New Advances in Symbol Timing Synchronization of Single-Carrier, Multi-Carrier  
and Space-Time Multiple-Antenna Systems. (August 2005)

Yik Chung Wu, B.Eng.; M.S., The University of Hong Kong, Hong Kong

Chair of Advisory Committee: Dr. Erchin Serpedin

In this dissertation, the problem of symbol timing synchronization for the following three different communication systems is studied: 1) conventional single-carrier transmissions with single antenna in both transmitter and receiver; 2) single-carrier transmissions with multiple antennas at both transmitter and receiver; and 3) orthogonal frequency division multiplexing (OFDM) based IEEE 802.11a wireless local area networks (WLANs).

For conventional single-carrier, single-antenna systems, a general feedforward symbol-timing estimation framework is developed based on the conditional maximum likelihood principle. The proposed algorithm is applied to linear modulations and two commonly used continuous phase modulations: MSK and GMSK. The performance of the proposed estimator is analyzed analytically and via simulations.

Moreover, using the newly developed general estimation framework, all the previously proposed digital blind feedforward symbol timing estimators employing second-order statistics are cast into a unified framework. The finite sample mean-square error expression for this class of estimators is established and the best estimators are determined. Simulation results are presented to corroborate the analytical results.

Moving on to single-carrier, multiple-antenna systems, we present two algorithms. The first algorithm is based on a heuristic argument and it improves the optimum sample selection algorithm by Naguib *et al.* so that accurate timing esti-

mates can be obtained even if the oversampling ratio is small. The performance of the proposed algorithm is analyzed both analytically and via simulations.

The second algorithm is based on the maximum likelihood principle. The data aided (DA) and non-data aided (NDA) ML symbol timing estimators and their corresponding CCRB and MCRB in MIMO correlated flat-fading channels are derived. It is shown that the improved algorithm developed based on the heuristic argument is just a special case of the DA ML estimator. Simulation results under different operating conditions are given to assess and compare the performances of the DA and NDA ML estimators with respect to their corresponding CCRBs and MCRBs.

In the last part of this dissertation, the ML timing synchronizer for IEEE 802.11a WLANs on frequency-selective fading channels is developed. The proposed algorithm is compared with four of the most representative timing synchronization algorithms, one specifically designed for IEEE 802.11a WLANs and three other algorithms designed for general OFDM frame synchronization.

To my parents,  
although they received little formal education,  
they guaranteed all of their children graduated from universities.

## ACKNOWLEDGMENTS

First, I would like to express my heartfelt appreciation to my advisor, Professor Erchin Serpedin, who has been an excellent teacher and an inspiring advisor. I have to thank him for his patience in reviewing and giving precious advice on my research work. I also have to thank Professor Costas N. Georghiades, Professor Andrew K. Chan, Professor Aniruddha Datta and Professor Siu A. Chin for serving on my academic committee.

Next, I have to thank the Croucher Foundation for giving me the financial support during my three years of Ph.D. study. Without their support, this journey would be a lot rougher.

I will not forget the students from the wireless communications group for making me not feel lonely during these three years. In particular, Yan Wang, Kai Shi and Yongzhe Xie should receive special credit from me for the stimulating discussions and their generosity in sharing their knowledge.

I am very grateful to my parents for their love and support throughout my many years of education. Without them, I would not achieve what I have accomplished today.

Last but not least, thanks to my girlfriend, Ana C.J. Lam. Although we were separated by thousands of miles, she has always provided me endless understanding, love, and support, and made the impossible possible.

## TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION . . . . .	1
	A. Motivations . . . . .	1
	1. Symbol Timing Synchronization in Single-Carrier Systems with Single Antenna . . . . .	1
	2. Symbol Timing Synchronization in Single-Carrier Systems with Multiple Antennas . . . . .	2
	3. Symbol Timing Synchronization in IEEE 802.11a WLAN – A Multi-Carrier System . . . . .	3
	B. Outline and Contributions of This Dissertation . . . . .	5
	C. Commonly Used Notations . . . . .	8
II	DESIGN AND ANALYSIS OF FEEDFORWARD SYMBOL TIMING ESTIMATORS BASED ON THE CONDITIONAL MAXIMUM LIKELIHOOD PRINCIPLE . . . . .	9
	A. Introduction . . . . .	9
	B. Signal Model and the CML Function . . . . .	11
	1. Signal Model . . . . .	11
	2. The CML Function . . . . .	14
	C. Proposed Estimator . . . . .	15
	D. Relationship with the Square Nonlinearity Estimator . . . . .	19
	E. Analytical Performance Analysis . . . . .	22
	F. Simulation Results and Discussions . . . . .	25
	G. Conclusions . . . . .	33
III	UNIFIED ANALYSIS OF A CLASS OF BLIND FEEDFOR- WARD SYMBOL TIMING ESTIMATORS EMPLOYING SECOND-ORDER STATISTICS . . . . .	34
	A. Introduction . . . . .	34
	B. Unified Formulation for the Class of Symbol Timing Estimators Employing the Second-Order Statistics . . . . .	36
	1. Cyclic Correlation-Based Estimator . . . . .	37
	2. Lee’s Estimator and the Modified Estimator . . . . .	38
	3. Feedforward CML Estimator . . . . .	39

CHAPTER	Page
4. Estimators with Pre-Filter . . . . .	40
C. Performance Analysis . . . . .	41
1. Performance Bound . . . . .	41
2. MSE Expression . . . . .	42
D. Numerical Examples and Discussions . . . . .	43
E. Conclusions . . . . .	46
 IV	
SYMBOL TIMING ESTIMATION IN MIMO FLAT-FADING CHANNELS – A HEURISTIC APPROACH . . . . .	47
A. Introduction . . . . .	47
B. Signal Model . . . . .	48
C. Timing Synchronization by Optimum Samples Selection . .	49
D. Design of Training Sequences . . . . .	54
E. Timing Synchronization by Estimation . . . . .	55
F. Performance Analysis . . . . .	58
G. Simulation Results and Discussions . . . . .	60
1. Effect of Oversampling Ratio . . . . .	61
2. Effect of Length of Training Sequences . . . . .	63
3. Effect of Number of Receive Antennas . . . . .	63
4. Effect of Number of Transmit Antennas . . . . .	64
H. Conclusions . . . . .	66
 V	
SYMBOL TIMING ESTIMATION IN MIMO FLAT-FADING CHANNELS – MAXIMUM LIKELIHOOD APPROACH . . . . .	67
A. Introduction . . . . .	67
B. Signal Model . . . . .	69
C. Symbol Timing Estimation with Known Training Data . .	71
1. ML Estimator . . . . .	71
2. The CCRB and MCRB . . . . .	76
3. Optimal Orthogonal Training Sequences . . . . .	78
D. Non-Data Aided Symbol Timing Estimation . . . . .	84
1. ML Estimator . . . . .	84
2. The CCRB and MCRB . . . . .	86
E. Simulation Results and Discussions . . . . .	89
1. Effects of $N$ and $M$ . . . . .	89
2. Effects of Correlation Among Antennas . . . . .	93
3. Comparison of DA and NDA Estimators . . . . .	97
F. Conclusions . . . . .	99

CHAPTER	Page	
VI	MAXIMUM-LIKELIHOOD SYMBOL SYNCHRONIZATION FOR IEEE 802.11A WLANS IN UNKNOWN FREQUENCY- SELECTIVE FADING CHANNELS . . . . .	100
	A. Introduction . . . . .	100
	B. Received-Signal Model . . . . .	102
	1. Signal and Channel Models . . . . .	102
	2. Matrix Algebraic Formulations . . . . .	107
	C. Symbol Synchronization Performance Criterion . . . . .	110
	D. Proposed Symbol Synchronization Algorithm . . . . .	114
	1. Frame Synchronization . . . . .	114
	a. First stage . . . . .	115
	b. Second stage . . . . .	117
	2. Position of FFT Window . . . . .	118
	3. Summary and Remarks . . . . .	118
	E. Simulation Results and Discussions . . . . .	120
	1. Simulation Conditions . . . . .	120
	2. Effect of Frequency Offset . . . . .	122
	3. Performances and Comparisons with Other Algorithms . . . . .	123
	F. Conclusions . . . . .	130
VII	CONCLUSIONS AND FUTURE WORK . . . . .	132
	A. Conclusions of This Dissertation . . . . .	132
	B. Suggestions for Future Work . . . . .	133
	REFERENCES . . . . .	134
	APPENDIX A . . . . .	145
	APPENDIX B . . . . .	150
	APPENDIX C . . . . .	155
	APPENDIX D . . . . .	160
	VITA . . . . .	163

## LIST OF TABLES

TABLE		Page
I	Parameters of different feedforward timing estimation algorithms when expressed in the form of the general estimator . . . . .	44
II	Probability of synchronization failure for the proposed algorithm in channel II at SNR=25dB and different frequency offsets . . . . .	122

## LIST OF FIGURES

FIGURE	Page
1	Examples of CML function. . . . . 16
2	Block diagrams for (a) the IDFT-based CML estimator ( $K=4$ ), (b) the proposed estimator ( $K=4$ ) and (c) the squaring estimator. . . 20
3	Plots of $\mathbb{E}[(\phi - \phi^*)^2]$ and $\mathbb{E}[(\phi + \phi^*)^2]$ as a function of $E_s/N_o$ for $\varepsilon_o = 0, 0.25, 0.5$ and $0.75$ ( $g(t)$ is a square-root raised cosine pulse with $\alpha = 0.5, Q = 2, K = 4, L_o=100$ and $L_g = 3$ ). Note that all curves for different values of $\varepsilon_o$ overlap. . . . . 23
4	MSE of the proposed estimator and the IDFT-based CML esti- mator with (a) $K = 4$ , (b) $K = 8$ (QPSK, $Q=2, \alpha=0.5, L_o=100$ and $L_g=3$ ). . . . . 27
5	Comparison between analytical MSE and simulations of the pro- posed estimator (QPSK, $Q=2, K=4, \alpha=0.5, L_o=100$ and $L_g=3$ ). . . 28
6	Comparison of the MSE of the proposed estimator and the square nonlinearity estimator (QPSK, $\alpha=0.5, L_o = 100$ and $L_g=3$ ). . . . . 29
7	MSE for the proposed estimator, the algorithms in [34, pp.401], [35] and [36] (QPSK, $Q = 2, \alpha=0.5, L_o=100$ and $L_g=3$ ). . . . . 29
8	MSE of the proposed estimator ( $Q = 2, Q = 4$ and $Q = 8$ ) and the low-SNR approximated ML algorithm [37] for MSK ( $L_{\bar{g}} = 1$ and $L_o=100$ ). . . . . 30
9	MSE of the proposed estimator ( $Q = 2$ and $Q = 4$ ) and the low- SNR approximated ML algorithm [37] for GMSK with $BT = 0.5$ ( $L_{\bar{g}} = 2$ and $L_o=100$ ). . . . . 32
10	Block diagram of the general estimator. . . . . 37
11	Analytic and simulated MSEs for modified Lee's estimator and feedforward CML estimator. . . . . 45

FIGURE	Page
12	Analytic and simulated MSEs for SLN estimator with and without pre-filter. . . . . 46
13	Structure of the training sequence for symbol timing synchronization in a two transmit antenna system. . . . . 50
14	An example of $\Lambda_{ij}(k)$ with the scaled version of $ p(t) ^2$ for $-T/2 \leq t \leq T/2$ (dotted line). . . . . 53
15	Magnitude of $\mathbb{E}[(\phi - \phi^*)^2]$ and $\mathbb{E}[(\phi + \phi^*)^2]$ as a function of $E_s/N_o$ for $\varepsilon_o = -0.5, -0.25, 0, 0.25$ and $0.5$ ( $N = 2, M = 4, \alpha=0.3, Q=4, L_t=32, L_p=4$ ). Note that all curves for different values of $\varepsilon_o$ overlap. . . . . 59
16	MSE performance for different oversampling ratios $Q$ ( $N=2, M=4, L_t=32, \alpha=0.3$ ). . . . . 62
17	MSE performance for different lengths of the training sequence ( $N=2, M=4, Q=4, \alpha=0.3$ ). . . . . 63
18	MSE performance for different number of receive antennas $M$ ( $N=2, L_t=32, Q=4, \alpha=0.3$ ). . . . . 65
19	MSE performance for different number of transmit antennas $N$ ( $L_t=64, Q=4, \alpha=0.3, M=2$ ). . . . . 65
20	MSE performances comparison between the true and approximated DA ML estimators with different $L_o$ ( $M = N = 4, L_g = 4, g(t)$ being a RRC pulse with roll-off factor $\alpha = 0.3, \Phi_T = \mathbf{I}_4, \mathbf{Z} = \mathbf{Z}_{opt}$ ). 76
21	Plots of $\beta \triangleq 1/\text{tr}(\tilde{\mathbf{Z}}^H \tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o} \tilde{\mathbf{Z}} \Phi_T)$ against $\varepsilon_o$ for $\varepsilon_t = 0, 0.25, 0.5, 0.75$ ( $N = 4, L_o = 32, L_g = 4, g(t)$ being a RRC pulse with $\alpha = 0.3, \Phi_T = \mathbf{I}_4$ ). . . . . 80
22	Plots of $\text{CCRB}_{DA}(\varepsilon_o)/\text{MCRB}_{DA}(\varepsilon_o)$ against the number of transmit antennas $N$ for $\varepsilon_o = 0, 0.25, 0.5, 0.75$ ( $L_o = 32$ and $128, L_g = 4$ and $g(t)$ being a RRC pulse with $\alpha = 0.3, \Phi_T = \mathbf{I}_N, \mathbf{Z} = \mathbf{Z}_{opt}$ ). 81
23	Plots of $[\mathbf{Z}_{opt}]_{:,1}$ and $\mathbf{w}_{31}$ ( $g(t)$ being a RRC pulse with $\alpha = 0.3, L_o = 32, L_g = 4$ ). . . . . 83

FIGURE	Page
24	Plots of $[\mathbf{Z}_{opt}]_{:,2}$ and $-\mathbf{w}_{30}$ ( $g(t)$ being a RRC pulse with $\alpha = 0.3$ , $L_o = 32$ , $L_g = 4$ ). . . . . 83
25	Comparison of the MSE performances of $ML_{DA}$ with different training sequences ( $g(t)$ being a RRC pulse with $\alpha = 0.3$ , $M = N = 4$ , $L_o = 32$ , $L_g = 4$ , $\Phi_T = \Phi_R = \mathbf{I}_4$ ). . . . . 84
26	MSEs of the $ML_{DA}$ estimator and the corresponding CCRBs with different number of transmit antennas ( $\Phi_T = \mathbf{I}_N$ , $\Phi_R = \mathbf{I}_M$ , $\mathbf{Z} = \mathbf{Z}_{opt}$ ). 91
27	MSEs of the $ML_{NDA}$ estimator and the corresponding CCRBs with different number of transmit antennas ( $\Phi_T = \mathbf{I}_N$ , $\Phi_R = \mathbf{I}_M$ and the data transmitted is spatially and temporally white). . . . . 91
28	MSEs of the $ML_{DA}$ estimator and the corresponding CCRBs with different number of receive antennas ( $\Phi_T = \mathbf{I}_N$ , $\Phi_R = \mathbf{I}_M$ , $\mathbf{Z} = \mathbf{Z}_{opt}$ ). 92
29	MSEs of the $ML_{NDA}$ estimator and the corresponding CCRBs with different number of receive antennas ( $\Phi_T = \mathbf{I}_N$ , $\Phi_R = \mathbf{I}_M$ and the data transmitted is spatially and temporally white). . . . . 92
30	MSEs of the $ML_{DA}$ estimator and the corresponding CCRBs with and without fading correlation between antennas for a $4 \times 4$ system. 94
31	MSEs of the $ML_{NDA}$ estimator and the corresponding CCRBs with and without fading correlation between antennas for a $4 \times 4$ system. . . . . 94
32	MSEs of the $ML_{DA}$ estimator against the correlation coefficient $\rho$ between adjacent antennas for $E_s/N_o=10\text{dB}$ , $20\text{dB}$ and $30\text{dB}$ in a $4 \times 4$ system. . . . . 96
33	MSEs of the $ML_{NDA}$ estimator against the correlation coefficient between adjacent antennas $\rho$ for $E_s/N_o=10\text{dB}$ , $20\text{dB}$ and $30\text{dB}$ in a $4 \times 4$ system. . . . . 96
34	Comparison of MSEs of the $ML_{NDA}$ and $ML_{DA}$ and their corresponding CCRBs and MCRBs for a $4 \times 4$ system. . . . . 98
35	Packet structure for IEEE 802.11a WLANs. . . . . 103

FIGURE	Page
36	An example of $ h(iT_{sam}) ^2$ for $L_h = 6$ , the first tap of the physical channel has zero delay, other five taps have delays uniformly distributed over the interval $0 - 300\text{ns}$ , $\gamma_n$ are i.i.d., zero-mean, complex Gaussian random variables with variances following the multipath intensity profile $\phi(\tau) \sim e^{-\tau/\tau_{rms}}$ , where $\tau_{rms} = 100\text{ns}$ , $\varepsilon_o$ is a random variable uniformly distributed in $[0, 1)$ and $f(t)$ is given by (6.11) with $\alpha = 0.1$ . . . . . 108
37	OFDM symbol and FFT position. . . . . 111
38	Distributions of the detected FFT window starting position for the proposed algorithm and algorithms based on autocorrelation [14], cross-correlation [20], double correlation [21] and GAIC [24] with respect to the ideal FFT window starting position in channel I. 127
39	$P_f(0.5\text{dB})$ for the proposed algorithm and algorithms based on autocorrelation [14], cross-correlation [20], double correlation [21] and GAIC [24] as a function of SNR in channel I. . . . . 127
40	Distributions of the detected FFT window starting position for the proposed algorithm and algorithms based on autocorrelation [14], cross-correlation [20], double correlation [21] and GAIC [24] with respect to the ideal FFT window starting position in channel II. 129
41	$P_f(0.5\text{dB})$ for the proposed algorithm and algorithms based on autocorrelation [14], cross-correlation [20], double correlation [21] and GAIC [24] as a function of SNR in channel II. . . . . 129

## CHAPTER I

## INTRODUCTION

## A. Motivations

Symbol timing synchronization is a fundamental component in any communication system. Unfortunately, this is also the component that has received relatively less attention in the research community. In this dissertation, the problem of symbol timing synchronization for three different communication systems is studied. The first system refers to the conventional linearly and nonlinearly modulated single-carrier transmissions with single antenna in both the transmitter and receiver. The second system assumes linearly modulated single-carrier transmissions with multiple antennas at both the transmitter and receiver (examples are the space-time coding system and the Vertical-Bell Labs Layered Space-Time (V-BLAST) system). The third system under study is the orthogonal frequency division multiplexing (OFDM) based IEEE 802.11a wireless local area network (WLAN) – a multi-carrier system.

The motivations and objectives behind each of these three main research topics will be sketched separately in the following.

## 1. Symbol Timing Synchronization in Single-Carrier Systems with Single Antenna

In digital receivers, symbol timing synchronization can be implemented either in a feedforward or feedback mode. Although feedback schemes exhibit good tracking performance, they require a relatively long acquisition time. Therefore, for burst-mode transmissions, feedforward timing recovery schemes are more suitable. An all-digital feedforward symbol timing recovery scheme consists of first estimating the

---

The journal model is *IEEE Transactions on Automatic Control*.

timing delay from the received samples, which is the focus of this study, and then adjusting the timing using some sort of interpolation [1], [2].

Due to bandwidth efficiency considerations, non-data aided or blind symbol timing estimation schemes have attracted much attention during the last decade. Most of the feedforward timing estimators proposed in the literature exploit the cyclostationarity induced by oversampling the received signal [3]-[8]. In [3], Oerder and Meyr proposed the well-known square nonlinearity estimator. Several extensions of this square nonlinearity estimator can be found in [5]-[7]. In [8], a low-SNR approximation was applied to the maximum likelihood function in order to derive a logarithmic nonlinearity. Reference [4] reported for the first time a detailed performance analysis of the estimators based on various types of nonlinearities.

Recently, the conditional maximum likelihood (CML) principle was introduced for designing digital timing delay synchronizers by Riba, Sala and Vazquez [9], [10]. The CML solution is especially important for symbol timing synchronization because it yields self-noise free timing estimates at medium and high SNRs. However, [9], [10] concentrate on deriving a CML timing error detector (TED) so that the timing delay can only be tracked using a feedback loop.

The objective of this first study is to develop and analyze the performance of a new blind feedforward symbol timing estimator based on the CML principle that exhibits excellent performance for general linear modulations.

## 2. Symbol Timing Synchronization in Single-Carrier Systems with Multiple Antennas

Apparently, this problem might seem closely related to symbol timing estimation in single-antenna systems. However, in multiple-input multiple-output (MIMO) systems, signals from different transmit antennas are superimposed together, and the

symbol timing estimation algorithms proposed in single antenna systems may not work in the MIMO case. Furthermore, in some MIMO algorithms (e.g., space-time coding), training sequences are used to estimate the channels. This opens up two questions: how can we make use of the training sequences to perform symbol timing estimation? What kind of training sequences is beneficial to symbol timing estimation?

This problem was first studied in [11], where orthogonal training sequences are transmitted at different transmit antennas to simplify the maximization of the oversampled approximated log-likelihood function. The sample having the largest magnitude, so called the “optimal sample”, is assumed to be closest to the optimum sampling instants. However, it can be shown that the mean square error (MSE) of this algorithm is lower bounded by  $1/(12Q^2)$ , where  $Q$  is the oversampling ratio. As a result, the performance of this timing synchronization method highly depends on the oversampling ratio. In fact, relatively high oversampling ratios are required for accurate symbol-timing estimation.

There are two objectives in this study. The first one is to extend the algorithm in [11] to increase its estimation accuracy. The second objective is to develop and analyze the maximum likelihood (ML) symbol timing estimator for the MIMO communication channels. The interest in deriving the ML timing estimator is due to its statistical efficiency.

### 3. Symbol Timing Synchronization in IEEE 802.11a WLAN – A Multi-Carrier System

IEEE 802.11a WLANs, which support high-speed data transmissions up to 54Mbps, employ burst-mode transmission and OFDM as the transmission technique. Although OFDM is well known for its ability to combat the intersymbol interference (ISI) in-

roduced by multipath channels, incorrect positioning of the FFT window within an OFDM symbol reintroduces ISI during data demodulation, causing serious performance degradation [12], [13]. Symbol synchronization is therefore one of the most important tasks performed at receivers in IEEE 802.11a WLANs.

A number of methods for OFDM symbol synchronization have been proposed in the literature. Methods that exploit the periodic structure of cyclic prefixes in OFDM symbols have been proposed in [13]-[15]. Algorithms based on the use of repeated preambles have been reported in [16]-[21]. In [12] and [22], additional pilot subcarriers are used to further improve the estimation accuracy after coarse timing synchronization is established by correlation-based methods. Although the techniques of [12]-[22] (which were originally developed for general OFDM systems) may be applied to IEEE 802.11a WLANs, a higher synchronization accuracy can be obtained by using optimized algorithms that take advantage of the known preamble structure located at the beginning of a data packet.

Recently, symbol synchronization techniques that are specifically designed for IEEE 802.11a WLANs have been reported in [23] and [24]. In [23], the received signal is correlated with a known training-symbol sequence and the absence of the expected correlation peak is detected. Despite the advantage that a simple correlator can be easily implemented at the receiver, its performance is poor in dispersive channels [23], indicating that more sophisticated synchronization algorithms are required. In [24], the generalized Akaike information criterion (GAIC) is used to jointly estimate the channel and timing information. Although the reported performance is good, its complexity is extremely high.

The objective of this study is to develop the ML symbol synchronizer tailored for the IEEE 802.11a standard.

## B. Outline and Contributions of This Dissertation

### Chapters II-III

In Chapter II, a general feedforward symbol-timing estimation framework based on the CML principle is developed. The proposed timing estimator presents reduced implementation complexity and is obtained by performing an approximation on the Fourier series expansion of the CML function. The proposed algorithm is applied to linear modulations and two commonly used continuous phase modulations: MSK and GMSK. The mean-square-error (MSE) performance of the proposed estimator is analyzed both analytically and via simulations.

For the linear modulations, it is shown that the performance of the proposed estimator is very close to the theoretical limit given by the conditional Cramer-Rao bound (CCRB) and modified Cramer-Rao bound (MCRB) for signal-to-noise ratios (SNR's) in the range  $\text{SNR} \leq 30\text{dB}$ . Furthermore, the proposed estimator is shown to be asymptotically equivalent to the classic square-law nonlinearity (SLN) estimator [3] under certain conditions. In the case of MSK and GMSK modulations, although the proposed algorithm reaches the CCRB at certain SNRs, however, the CCRB is quite far away from the MCRB, and there exists an alternative algorithm whose performance comes closer to the MCRB. Therefore, it is concluded that the proposed estimator is more suitable for linear modulations rather than for MSK and GMSK modulations. Most of the results in Chapter II have been published in [25] and [26].

In Chapter III, a unifying framework that subsumes a class of blind feedforward symbol-timing estimators employing second-order statistics is proposed. The unifying expression leads to a general-purpose finite-sample MSE expression for this class of synchronizers, which is useful for systematic performance analysis and comparisons. Simulation results are also presented to corroborate the analytical results. It is found

that the feedforward CML estimator (proposed in Chapter II) and the SLN estimator with a properly designed pre-filter perform the best and their performances coincide with the asymptotic CCRB, which is the performance lower bound for the class of estimators under consideration. Most of the material in Chapter III has been published in [27].

### Chapters IV-V

In Chapter IV, a new symbol-timing estimator for MIMO systems is proposed based on a heuristic approach. It improves the optimum sample selection algorithm in [11] so that accurate timing estimates can be obtained even if the oversampling ratio is small. The increase in implementation complexity with respect to that of optimum sample selection algorithm is minimal. Both analytical and simulation results show that, for modest oversampling ratio (such as  $Q=4$ ), the MSE of the proposed estimator is significantly smaller than that of the optimum sample selection algorithm [11]. The requirements and the design procedures of the training sequences are also discussed. Most of the material in Chapter IV has been published in [28].

In Chapter V, the symbol-timing estimation problem in MIMO systems is tackled using the ML method. In particular, the data aided (DA) and non-data aided (NDA) ML symbol timing estimators and their corresponding CCRB and MCRB in MIMO correlated flat-fading channels are derived. It is shown that the improved algorithm developed in the Chapter IV is just a special case of the DA ML estimator. For the DA case, the optimal orthogonal training sequences are also derived by minimizing the MCRB. It is found that the optimal orthogonal sequences resemble the Walsh sequences, but present different envelopes. Simulation results under different operating conditions (e.g., number of antennas and correlation between antennas) are given to assess and compare the performances of the DA and NDA ML estimators with

respect to their corresponding CCRBs and MCRBs. It is found that i) the MSE of the DA ML estimator is close to the CCRB and MCRB, ii) the MSE of the NDA ML estimator is close to the CCRB but not to the MCRB, iii) the MSEs of both DA and NDA ML estimators are approximately independent of the number of transmit antennas and are inversely proportional to the number of receive antennas, iv) correlation between antennas has little effect on the MSEs of DA and NDA ML estimators, and v) DA ML estimator performs better than NDA ML estimator at the cost of lower transmission efficiency and higher implementation complexity. Most of the material in this chapter has been published in [29].

## Chapter VI

In Chapter VI, the ML symbol synchronizer for IEEE 802.11a WLANs in frequency selective fading channels is developed. A realistic channel, which includes the effects of filtering and sampling time offset in addition to the physical channel with random path delays, is considered. Furthermore, the loss in system performance due to synchronization error is used as the performance criterion [13], [30], as opposed to the requirement that the estimated symbol timing has to be within certain limits with respect to a fixed reference point. The proposed algorithm is compared with four of the most representative symbol synchronization algorithms, one of which specifically designed for IEEE 802.11a WLANs [24] and three other algorithms designed for general OFDM frame synchronization [14], [20], [21]. Simulation results indicate that in general, joint estimation of symbol position and channel (as is the case with the proposed algorithm and the algorithm based on GAIC [24]) gives better performances than the correlation based algorithms [14], [20], [21]. When compared to the GAIC algorithm [24], the proposed algorithm exhibits comparable performances, but the complexity of the proposed algorithm is much smaller than that of GAIC algorithm

due to the smaller observation length. Most of the material in this chapter has been published in [31].

### C. Commonly Used Notations

The following are the commonly used notations in this dissertation. The symbols  $\mathbf{x}^*$ ,  $\mathbf{x}^T$ ,  $\mathbf{x}^H$  and  $\|\mathbf{x}\|$  denote the conjugate, transpose, transpose conjugate and the Euclidean norm of  $\mathbf{x}$ , respectively. Notation  $\otimes$  denotes Kronecker products, and  $\text{vec}(\mathbf{H})$  denotes a vector formed by stacking the columns of  $\mathbf{H}$  one on top of each other. Notation  $\star$  stands for convolution. Notations  $\Re(x)$ ,  $\Im(x)$  and  $\mathbb{E}[x]$  denote the real part, imaginary part and expectation of  $x$ , respectively. Matrices  $\mathbf{I}_N$  and  $\mathbf{0}_N$  are the identity and the all zero matrix, respectively, and both are of dimensions  $N \times N$ , while  $\mathbf{0}_{M \times N}$  is the the all zero matrix with dimension  $M \times N$ . Notations  $\mathbf{Z}_{i,:}$ ,  $\mathbf{Z}_{:,j}$  and  $\mathbf{Z}_{ij}$  denote the  $i^{\text{th}}$  row,  $j^{\text{th}}$  column and  $(i, j)^{\text{th}}$  element of  $\mathbf{Z}$ , respectively. The symbol  $\delta(\cdot)$  stands for the Kronecker's delta. Notation  $\mathbf{j}$  is defined as  $\sqrt{-1}$ .

## CHAPTER II

### DESIGN AND ANALYSIS OF FEEDFORWARD SYMBOL TIMING ESTIMATORS BASED ON THE CONDITIONAL MAXIMUM LIKELIHOOD PRINCIPLE

#### A. Introduction

In digital receivers, symbol timing synchronization can be implemented either in a feedforward or feedback mode. Although feedback schemes exhibit good tracking performances, they require a relatively long acquisition time. Therefore, for burst-mode transmissions, feedforward timing recovery schemes are more suitable. An all-digital feedforward symbol timing recovery scheme consists of first estimating the timing delay from the received samples, which is the focus of this chapter, and then adjusting the timing using some sort of interpolation [1], [2].

Due to bandwidth efficiency considerations, non-data aided or blind symbol timing estimation schemes have attracted much attention during the last decade. Most of the feedforward timing estimators proposed in the literature exploit the cyclostationarity induced by oversampling the received signal [3]-[8]. In [3], Oerder and Meyr proposed the well-known square nonlinearity estimator. Several extensions of this square nonlinearity estimator can be found in [5]-[7]. In [8], a low-SNR approximation was applied to the maximum likelihood function in order to derive a logarithmic nonlinearity. Reference [4] reported for the first time a detailed performance analysis of the estimators based on various types of nonlinearities.

Recently, the conditional maximum likelihood (CML) principle was introduced for designing digital timing delay synchronizers by Riba, Sala and Vazquez [9], [10]. The CML solution is especially important for symbol timing synchronization because

it yields self-noise free timing estimates at medium and high SNRs. However, [9], [10] concentrate on deriving a CML timing error detector (TED) so that the timing delay can only be tracked using a feedback loop. The purpose of this chapter is to develop and analyze a feedforward timing estimator based on the CML principle. The main design and performance characteristics of CML-based feedforward symbol timing delay estimators are established for general linear modulations and two commonly used continuous phase modulations, namely, minimum shift keying (MSK) and Gaussian MSK (GMSK) [32], [33]. The performance of the timing estimators is analyzed analytically and through simulations, and compared with the conditional Cramer-Rao bound (CCRB) [9], [10], the modified Cramer-Rao bound (MCRB) [34] and other existing state-of-the-art feedforward timing delay estimators [3], [34], [35], [36], and [37].

In the proposed algorithm, an approximation is applied to the Fourier series expansion of the CML function so that the complexity of the proposed estimator is greatly reduced. Although the resulting estimator is not completely self-noise free (due to the approximation), the performances of the proposed estimator (for both linear and nonlinear modulations) are in general very close to the CCRB for signal-to-noise ratios (SNR's) smaller than 30 dB. For higher SNRs, mean square error (MSE) floor occurs, but notice that at that high SNRs, the estimation MSE achieved by the proposed estimator is already very small, so the effect of MSE floors becomes relatively less critical.

For linear modulations, it is shown that the proposed estimator is asymptotically equivalent to the well-known square nonlinearity estimator [3]. However, the proposed estimator exhibits better performance (less self-noise/jitter) than [3] when a reduced number of data samples are available. Furthermore, it is shown that the performances of the proposed estimator for linear modulations are also very close to MCRB for

SNR $\leq$ 30dB. For MSK and GMSK modulations, although the performances of the proposed estimator come very close to the CCRB at certain SNR ranges, however, the CCRB is quite far away from the MCRB, and there exists an alternative algorithm whose performance comes closer to the MCRB. Therefore, it is concluded that the proposed estimator is more suitable for linear modulations than MSK and GMSK modulations.

The rest of the chapter is organized as follows. The signal model and the CML function are first described in Section B. The proposed feedforward timing estimator is derived in Section C. The relationship between the proposed estimator and the well-known square nonlinearity estimator [3] is addressed in Section D. The MSE expressions are derived in Section E. Simulation results and discussions are then presented in Section F, and finally conclusions are drawn in Section G.

## B. Signal Model and the CML Function

### 1. Signal Model

The complex envelope of a received linear modulation is given by

$$r(t) = e^{j\theta_o} \sqrt{\frac{E_s}{T}} \sum_i d_i g(t - iT - \varepsilon_o T) + \eta(t), \quad (2.1)$$

where  $\theta_o$  is the unknown phase offset;  $E_s$  is the symbol energy;  $d_i$  stands for the zero-mean unit variance, independently and identically distributed (i.i.d.) complex valued symbols being sent;  $g(t)$  is the transmit pulse with unit energy;  $T$  is the symbol period;  $\varepsilon_o \in [0, 1)$  is the unknown symbol timing delay to be estimated and  $\eta(t)$  is the complex-valued circularly distributed white Gaussian noise with power density  $N_o$ . After passing through the anti-aliasing filter, the received signal is then sampled at the rate  $1/T_s$ , where  $T_s \triangleq T/Q$ . Note that the oversampling factor  $Q$  is determined by

the frequency span of  $g(t)$ ; if  $g(t)$  is bandlimited to  $f = \pm 1/T$  (an example of which is the square-root raised cosine pulse),  $Q = 2$  is sufficient. The received vector  $\mathbf{r}$ , which consists of  $L_o Q$  consecutive received samples (where  $L_o$  is the observation length), can be expressed as (without loss of generality, we consider the received sequence start at  $t = 0$ )

$$\mathbf{r} = [r(0), r(T_s), \dots, r((L_o Q - 1)T_s)]^T = \mathbf{A}_{\varepsilon_o} \mathbf{d}_o + \boldsymbol{\eta}, \quad (2.2)$$

where

$$\mathbf{A}_{\varepsilon} \triangleq [\mathbf{a}_{-L_g}(\varepsilon), \mathbf{a}_{-L_g+1}(\varepsilon), \dots, \mathbf{a}_{L_o+L_g-1}(\varepsilon)], \quad (2.3)$$

$$\mathbf{a}_i(\varepsilon) \triangleq [g(-iT - \varepsilon T), g(T_s - iT - \varepsilon T), \dots, g((L_o Q - 1)T_s - iT - \varepsilon T)]^T \quad (2.4)$$

$$\mathbf{d}_o \triangleq e^{j\theta_o} \sqrt{\frac{E_s}{T}} [d_{-L_g}, d_{-L_g+1}, \dots, d_{L_o+L_g-1}]^T, \quad (2.5)$$

$$\boldsymbol{\eta} \triangleq [\eta(0), \eta(1), \dots, \eta(L_o Q - 1)]^T, \quad (2.6)$$

$\eta(i) \triangleq \eta(iT/Q)$ , and  $L_g$  denotes the number of symbols affected by the inter-symbol interference (ISI) introduced by one side of  $g(t)$ .

For MSK and GMSK modulations, the complex envelope of the received signal is given by

$$r(t) = \sqrt{\frac{E_s}{T}} \exp[j\pi \sum_n a_n q(t - nT - \varepsilon_o T) + j\theta_o] + \eta(t), \quad (2.7)$$

where  $a_n$  stands for the i.i.d. binary transmitted symbols, and  $q(t)$  is the phase response of the modulator with length  $L$  and satisfies

$$q(t) = \begin{cases} 0 & t \leq 0 \\ 1/2 & t \geq LT. \end{cases} \quad (2.8)$$

The derivative of  $q(t)$  is referred to as the frequency response of the modulator, and

takes the form of a rectangular pulse or a convolution between a rectangular pulse and a Gaussian shaped pulse for MSK and GMSK modulations, respectively. According to the Laurent's expansion (LE) [38] and the fact that most of the energy of the GMSK modulation is concentrated in the first component of the expansion [37]-[39] (the MSK signal has only one component in the expansion), MSK and GMSK received signals can be approximated by

$$r(t) \cong e^{j\theta_o} \sqrt{\frac{E_s}{T}} \sum_i \bar{d}_i \bar{g}(t - iT - \varepsilon_o T) + \eta(t), \quad (2.9)$$

where

$$\bar{d}_i \triangleq \exp[j\frac{\pi}{2} \sum_{n=1}^i a_n], \quad (2.10)$$

$$\bar{g}(t) \triangleq \prod_{n=0}^{L-1} p(t + nT), \quad (2.11)$$

and

$$p(t) \triangleq \begin{cases} \sin[\pi q(t)] & 0 \leq t \leq LT \\ p(2LT - t) & LT < t \leq 2LT \\ 0 & \text{otherwise.} \end{cases} \quad (2.12)$$

Therefore, the sampled MSK and GMSK modulations can also be expressed in a form similar to (2.2). Since the pseudo-symbols (or equivalent data)  $\bar{d}_i$  are zero mean, unit variance, a single system model is sufficient to treat the linear modulations, MSK and GMSK signals within a common framework.

*Remark 1.* Notice that another formulation for the GMSK signal is to express the signal using all the  $2^{L-1}$  terms of the LE as is done in [10]. However, there is a disadvantage for doing this: including more LE terms in the formulation would significantly increase the number of pseudo-symbols. Since in CML method, the pseudo-symbols and the unknown timing delay are jointly estimated from an observation vector of

certain length, increasing the number of pseudo-symbols to be estimated would definitely degrade the overall estimation accuracy (of both pseudo-symbols and timing delay). Of course, neglecting some small LE terms (as is done in this chapter) would introduce interference and degrade the performance for the resulting estimator, but from the simulation examples to be presented in Section F, the effect of the system model approximation (2.9) is very small and occurs only at a very high SNR region (at  $\text{SNR} \geq 50\text{dB}$ ).

*Remark 2.* MSK and GMSK modulations belong to a broader class of modulation, called MSK-type modulation [32], [33]. The system model, the subsequent proposed estimator and the MSE analysis can also be applied to other members of this MSK-type modulation as long as the approximation in (2.9) is tight (e.g., 1RC, 2RC modulations). However, in this chapter, we only concentrate on two commonly used members: MSK and GMSK.

## 2. The CML Function

From (2.2), the joint maximum likelihood estimate of  $\varepsilon_o$  and  $\mathbf{d}_o$  is given by maximizing

$$p(\mathbf{r}|\varepsilon, \mathbf{d}) = \frac{1}{(\pi N_o)^{L_o Q}} \exp \left[ -\frac{(\mathbf{r} - \mathbf{A}_\varepsilon \mathbf{d})^H (\mathbf{r} - \mathbf{A}_\varepsilon \mathbf{d})}{N_o} \right], \quad (2.13)$$

or equivalently minimizing

$$J(\mathbf{r}|\varepsilon, \mathbf{d}) = (\mathbf{r} - \mathbf{A}_\varepsilon \mathbf{d})^H (\mathbf{r} - \mathbf{A}_\varepsilon \mathbf{d}), \quad (2.14)$$

where  $\varepsilon$  and  $\mathbf{d}$  are the trial values for  $\varepsilon_o$  and  $\mathbf{d}_o$ , respectively.

In the CML approach, the nuisance parameters  $\mathbf{d}_o$  are modeled as deterministic and estimated from the received vector  $\mathbf{r}$ . From the linear signal model given in (2.2), if no constraint is imposed on the possible value of  $\mathbf{d}_o$ , the maximum likelihood

estimate for  $\mathbf{d}_o$  (when  $\varepsilon$  is fixed) is [65]

$$\hat{\mathbf{d}} = (\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon)^{-1} \mathbf{A}_\varepsilon^H \mathbf{r}. \quad (2.15)$$

Plugging (2.15) into (2.14), after some straightforward manipulations and dropping the irrelevant terms, the timing delay is estimated by maximizing the following CML function [9]

$$\Lambda(\varepsilon) = \mathbf{r}^H \mathbf{A}_\varepsilon (\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon)^{-1} \mathbf{A}_\varepsilon^H \mathbf{r}. \quad (2.16)$$

In general, the maximum of the CML function can be found by plugging different values of  $\varepsilon$  into (2.16). The value that provides the maximum value of  $\Lambda(\varepsilon)$  is the CML estimate. Since  $\varepsilon$  is a continuous variable, this exhaustive search method requires a lot of computations and is impractical. Alternatively, a timing error detector (TED) [9] can be used in a feedback configuration. However, in burst mode transmissions, feedforward timing delay estimators [3]-[8] are preferred since they avoid the relatively long acquisition time and hang-up problem in feedback schemes. In the following, a new method for optimizing (2.16) is proposed so that an efficient implementation of the feedforward symbol-timing estimator results.

### C. Proposed Estimator

Fig. 1 shows some realizations of the CML function calculated using (2.16), where the true timing delay is  $\varepsilon_o = 0.25$  (for the linear modulation,  $g(t)$  is a square-root raised cosine filter with roll-off factor 0.5). It can be seen that the CML function has only one maximum. Since the CML function is *smooth*, we expect that it is not necessary to calculate the CML function for all the values of  $\varepsilon$ . It is possible that the CML function is first calculated for some  $\varepsilon$ 's and the values in between can be found by interpolation.

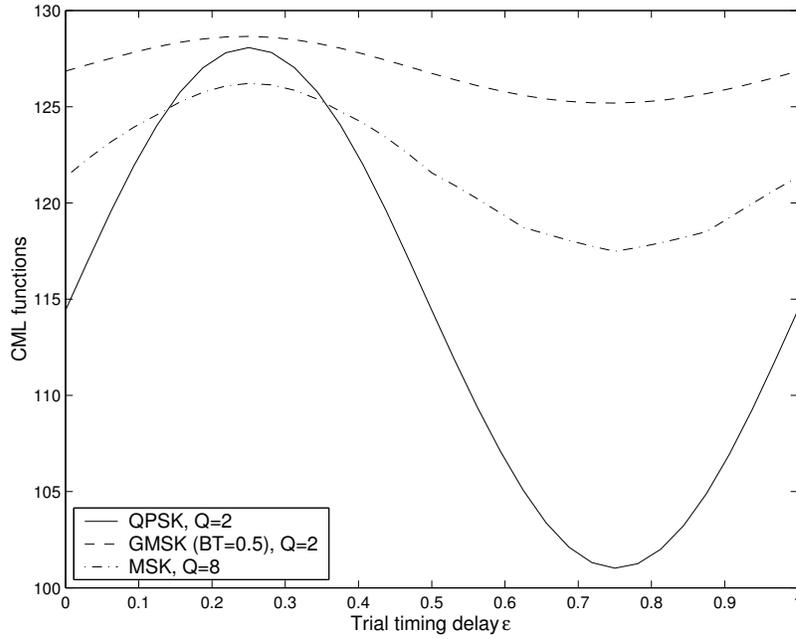


Fig. 1. Examples of CML function.

More specifically, suppose we calculated  $K$  uniformly spaced values of  $\Lambda(\varepsilon)$  using (2.16) such that a sequence  $\Lambda(k) \triangleq \Lambda(k/K)$  for  $k = 0, 1, \dots, K-1$  is obtained (without loss of generality, we consider  $K$  is even). Let construct a periodic sequence  $\tilde{\Lambda}(m)$  by periodically extending  $\Lambda(k)$ . Further, denotes  $\tilde{\Lambda}(\varepsilon)$  as the continuous and periodic function with its samples given by  $\tilde{\Lambda}(m)$ . According to the sampling theorem, as long as the sampling frequency  $1/K$  is higher than twice the highest frequency of  $\tilde{\Lambda}(\varepsilon)$ , then  $\tilde{\Lambda}(\varepsilon)$  can be represented by its samples  $\tilde{\Lambda}(m)$  without loss of information. The relationship between  $\tilde{\Lambda}(\varepsilon)$  and  $\tilde{\Lambda}(m)$  is then given by

$$\tilde{\Lambda}(\varepsilon) = \sum_{m=-\infty}^{\infty} \tilde{\Lambda}(m) \text{sinc}(\pi K(\varepsilon - m/K)) \quad (2.17)$$

where  $\text{sinc}(x) \triangleq \sin(x)/x$ . Now, expand  $\tilde{\Lambda}(\varepsilon)$  into a Fourier series

$$\tilde{\Lambda}(\varepsilon) = \sum_{\ell=-\infty}^{\infty} A_{\ell} e^{j2\pi\ell\varepsilon}, \quad (2.18)$$

where

$$A_\ell = \int_0^1 \tilde{\Lambda}(\varepsilon) e^{-j2\pi\ell\varepsilon} d\varepsilon. \quad (2.19)$$

Substituting (2.17) into (2.19) yields

$$\begin{aligned} A_\ell &= \sum_{m=-\infty}^{\infty} \tilde{\Lambda}(m) \int_0^1 \text{sinc}(\pi K(\varepsilon - m/K)) e^{-j2\pi\ell\varepsilon} d\varepsilon \\ &= \sum_{k=0}^{K-1} \Lambda(k) \sum_{l=-\infty}^{\infty} \int_0^1 \text{sinc}(\pi K(\varepsilon - l - k/K)) e^{-j2\pi\ell\varepsilon} d\varepsilon \\ &= \sum_{k=0}^{K-1} \Lambda(k) e^{-j2\pi\ell k/K} \int_{-\infty}^{\infty} \text{sinc}(\pi K\varepsilon) e^{-j2\pi\ell\varepsilon} d\varepsilon \\ &= \sum_{k=0}^{K-1} \Lambda(k) e^{-j2\pi\ell k/K} \cdot \frac{1}{K} \mathfrak{F}\{\text{sinc}(\pi\varepsilon)\}_{f=\ell/K}, \end{aligned} \quad (2.20)$$

where  $\mathfrak{F}\{\}$  denote the Fourier transform. It is clear that,

$$A_\ell = \begin{cases} \frac{1}{K} \sum_{k=0}^{K-1} \Lambda(k) e^{-j2\pi\ell k/K}, & \ell = -\frac{K}{2}, \dots, \frac{K}{2} \\ 0 & \text{otherwise.} \end{cases} \quad (2.21)$$

From (2.18), it can be seen that once the coefficients  $A_\ell$  are determined,  $\tilde{\Lambda}(\varepsilon)$  can be calculated for any  $\varepsilon \in [0, 1)$ . Then the problem of maximizing (2.16) can now be replaced by maximizing (2.18). For efficient implementation, the function  $\tilde{\Lambda}(\varepsilon)$  for  $0 \leq \varepsilon < 1$  can be approximated by a  $K'$ -point sequence ( $K' > K$ ) as follow

$$\Lambda(k') = \sum_{\ell=-K'/2}^{K'/2-1} A_\ell e^{j2\pi\ell k'/K'} \quad \text{for } k' = 0, 1, \dots, K' - 1. \quad (2.22)$$

This is equivalent to first calculating  $A_\ell$  using (2.21), then zero padding the high frequency coefficients ( $A_\ell$ ) and finally performing a  $K'$ -point inverse discrete Fourier transform (IDFT). For sufficiently large value of  $K'$ ,  $\Lambda(k')$  becomes very close to  $\tilde{\Lambda}(\varepsilon)$  for  $0 \leq \varepsilon < 1$ , and the index with the maximum amplitude can be viewed as an estimate of the unknown timing parameter  $\varepsilon_o$ . Fig. 2a shows the block diagram for

this algorithm when  $K=4$ . For the rest of the chapter, we refer to this estimator as the IDFT-based CML estimator.

To avoid the complexity in performing the  $K'$ -point IDFT, an approximation is applied to (2.18). More precisely, it can be seen from Fig. 1 that the CML function for symbol timing estimation resembles a sine function with one period in the interval  $0 \leq \varepsilon < 1$ . It is expected that the Fourier coefficient  $A_1$  is much larger than the Fourier coefficients associated with higher frequencies. Therefore, it is sufficient to approximate (2.18) as follows

$$\tilde{\Lambda}(\varepsilon) \approx A_0 + 2\Re\{A_1 e^{j2\pi\varepsilon}\} \quad \text{for } 0 \leq \varepsilon < 1, \quad (2.23)$$

where  $\Re\{x\}$  stands for real part of  $x$ . In order to maximize  $\tilde{\Lambda}(\varepsilon)$ , the following equation must hold

$$\arg(A_1) = -2\pi\varepsilon, \quad (2.24)$$

where  $\arg(x)$  denotes the phase of  $x$ . Or equivalently,

$$\hat{\varepsilon} = -\frac{1}{2\pi} \arg\left\{ \sum_{k=0}^{K-1} \Lambda(k) e^{-j2\pi k/K} \right\}. \quad (2.25)$$

The estimated delay  $\hat{\varepsilon}$  is the normalized (with respect to  $T$ ) time difference between the first sample of the received vector  $\mathbf{r}$  and the nearest optimum sampling instant. The calculation within the  $\arg(\cdot)$  operation is actually the first bin (i.e.,  $2^{nd}$  output) of a  $K$ -point discrete Fourier transform (DFT) of the sequence  $\Lambda(k)$  (or the Fourier coefficient at symbol rate  $f = 1/T$ ). Based on (2.24), it is not difficult to check that the proposed estimator (2.25) is asymptotically unbiased, a result which is independent of the approximation used in (2.23).

From a computational viewpoint, it is worth mentioning that the proposed estimator only involves the calculation of  $K$  samples of the CML function using (2.16), a

$K$ -point DFT, and an  $\arg(\cdot)$  operation. From the results to be presented, it is found that  $K = 4$  is sufficient to yield good estimates in practical applications. Therefore, the 4-point DFT in (2.25) can be computed easily without requiring any multiplications. The main complexity comes from the calculation of the 4 samples of  $\Lambda(\varepsilon)$  using (2.16). However, notice that the matrix  $\mathbf{A}_\varepsilon(\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon)^{-1} \mathbf{A}_\varepsilon^H$  can be pre-computed for  $\varepsilon = k/4$  with  $0 \leq k \leq 3$ . This greatly reduces the arithmetic complexity of implementation. Complexity can be further reduced by approximating the pre-computed  $\mathbf{A}_\varepsilon(\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon)^{-1} \mathbf{A}_\varepsilon^H$  using Sum-of-Power-of-Two (SOPOT) expressions [40], [41].

#### D. Relationship with the Square Nonlinearity Estimator

In this section, we will show that, if  $g(t)$  is a square-root raised cosine pulse, the proposed estimator in (2.25) asymptotically reduces to the well-known square nonlinearity estimator [3]. First notice that when  $g(t)$  is a square-root raised cosine pulse and in the asymptotic case (as  $L_o \rightarrow \infty$ ),  $[\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon]_{ij} \approx \delta_{ij}$  [10], where  $\delta_{ij} = 1$  if  $i = j$  and zero otherwise. Notice that the matrix  $\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon$  is of dimension  $(L_o + 2L_g) \times (L_o + 2L_g)$ . The approximation  $[\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon]_{ij} \approx \delta_{ij}$  holds very well for the central portion (of dimension  $L_o \times L_o$ ) of  $\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon$ . For the boundary of  $\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon$ , the values are smaller than 1. As  $L_o \rightarrow \infty$ , the boundary of  $\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon$  becomes insignificant and can be ignored. Putting  $[\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon]_{ij} \approx \delta_{ij}$  into (2.16), it follows that

$$\Lambda(\varepsilon) \approx \|\mathbf{A}_\varepsilon^H \mathbf{r}\|^2. \quad (2.26)$$

Now consider the  $i^{\text{th}}$  element of  $\mathbf{A}_\varepsilon^H \mathbf{r}$  ( $i = -L_g, -L_g + 1, \dots, L_o + L_g - 1$ ),

$$\begin{aligned} [\mathbf{A}_\varepsilon^H \mathbf{r}]_i &= \sum_{n=0}^{L_o Q - 1} g(nT_s - iT - \varepsilon T) r(nT_s) \\ &= \sum_{n=-\infty}^{\infty} g((i + \varepsilon)T - nT_s) \tilde{r}(nT_s) \end{aligned} \quad (2.27)$$

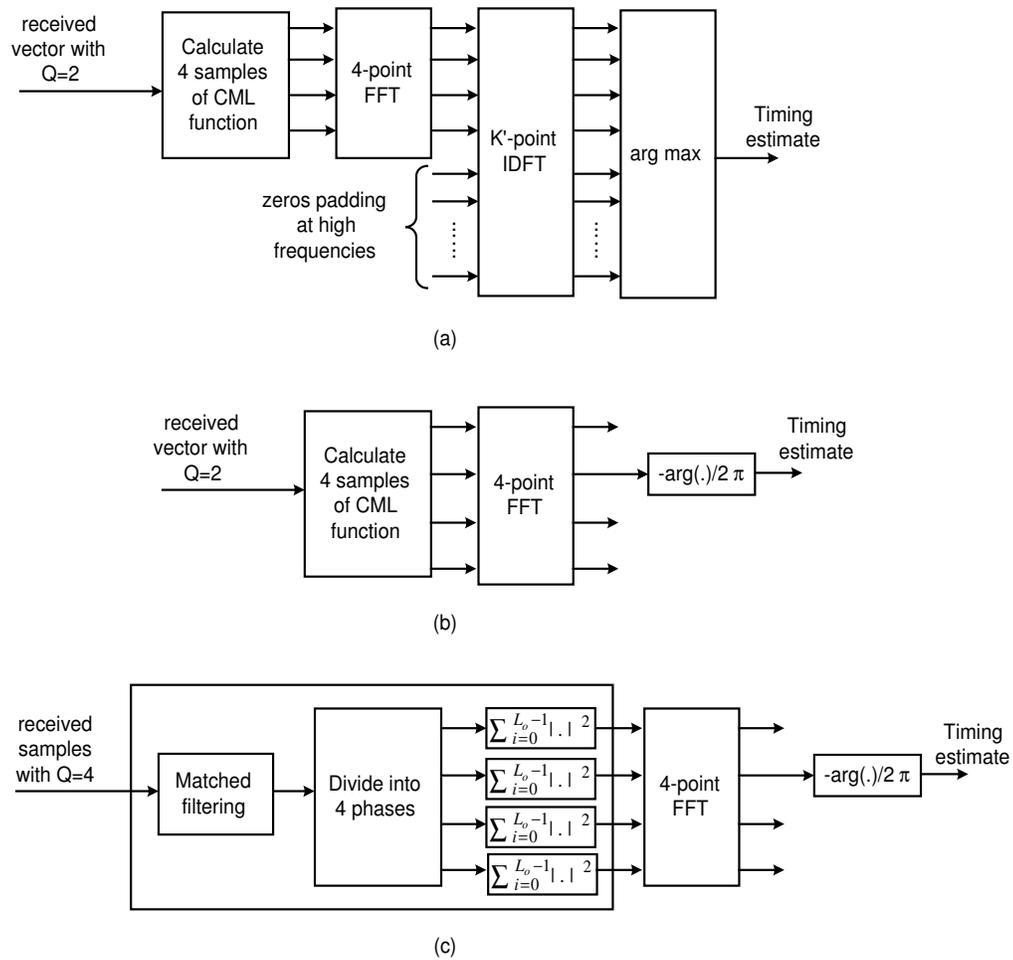


Fig. 2. Block diagrams for (a) the IDFT-based CML estimator ( $K=4$ ), (b) the proposed estimator ( $K=4$ ) and (c) the squaring estimator.

where  $\tilde{r}(t) \triangleq r(t)w(t)$ , with  $w(t)$  is a rectangular window of length  $L_oT$ . It is recognized that the summation in (2.27) is just the filtering of  $\tilde{r}(t)$ , through  $g(t)$ , followed then by sampling at  $t = (i + \varepsilon)T$ . Notice that since  $g(t)$  is a square-root raised cosine filter,  $g(t) = g(-t)$  and (2.27) actually corresponds to the sampled matched filter output. If we define  $x(t) \triangleq g(-t) \star \tilde{r}(t)$ , where  $\star$  denotes convolution, we have  $[\mathbf{A}_\varepsilon^H \mathbf{r}]_i = x((i + \varepsilon)T)$ . Plugging this result into (2.26) and noting that asymptotically, the range of  $i$  can be approximated by  $i = 0, \dots, L_o - 1$ , we have

$$\Lambda(k) \approx \sum_{i=0}^{L_o-1} |x(iT + kT/K)|^2. \quad (2.28)$$

The proposed CML feedforward timing delay estimator in (2.25) can then be rewritten as

$$\begin{aligned} \hat{\varepsilon} &= -\frac{1}{2\pi} \arg\left\{ \sum_{k=0}^{K-1} \left( \sum_{i=0}^{L_o-1} |x(iT + kT/K)|^2 \right) e^{-j2\pi k/K} \right\} \\ &= -\frac{1}{2\pi} \arg\left\{ \sum_{l=0}^{KL_o-1} |x(lT/K)|^2 e^{-j2\pi l/K} \right\}. \end{aligned} \quad (2.29)$$

Therefore, when  $L_o \rightarrow \infty$  and  $K=4$ , we have the well known squaring algorithm [3]. Figs. 2b and 2c show the block diagrams for the proposed estimator (2.25) with  $K = 4$  and the squaring algorithm. It can be seen that the structures of the proposed algorithm and the squaring algorithm are very alike. Note that both the proposed algorithm and the squaring algorithm require four samples per symbol period to form the timing estimate. For the proposed estimator, the received signal is first sampled with minimum oversampling ratio  $Q=2$  and then samples with  $K = 4$  different phases are generated by filtering (see (2.27)). For the squaring algorithm, the four different samples per symbol period are directly obtained through sampling. Notice that the squaring algorithm might work also by first sampling at  $Q = 2$  and then the intermediate (additional two) samples are computed by interpolation before sym-

bol timing estimation. Although the proposed estimator and the squaring algorithm have many characteristics in common, simulation results presented in Section F show that the proposed estimator outperforms the squaring algorithm for reduced length observation records.

### E. Analytical Performance Analysis

In this section, we derive the mean square error (MSE) expressions for the proposed estimator as a function of  $E_s/N_o$ . First, express the true timing delay  $\varepsilon_o$  as follows

$$\varepsilon_o = -\frac{1}{2\pi} \arg(e^{-j2\pi\varepsilon_o}). \quad (2.30)$$

From (2.25) and (2.30), the MSE for a specific delay is given by

$$\mathbb{E}[(\hat{\varepsilon} - \varepsilon_o)^2] = \left(\frac{1}{2\pi}\right)^2 \mathbb{E} \left[ \left( \arctan \left\{ \frac{\Im\{\phi\}}{\Re\{\phi\}} \right\} \right)^2 \right] \quad (2.31)$$

where

$$\phi \triangleq e^{j2\pi\varepsilon_o} \sum_{k=0}^{K-1} \Lambda(k) e^{-j2\pi k/K}. \quad (2.32)$$

Applying the approximation  $\arctan(x) \approx x$  for small  $x$ , we have

$$\begin{aligned} \mathbb{E}[(\hat{\varepsilon} - \varepsilon_o)^2] &\approx \left(\frac{1}{2\pi}\right)^2 \mathbb{E} \left[ \left( \frac{\phi - \phi^*}{j(\phi + \phi^*)} \right)^2 \right] \\ &\approx -\left(\frac{1}{2\pi}\right)^2 \frac{\mathbb{E}[\phi^2] - 2\mathbb{E}[\phi\phi^*] + \mathbb{E}[(\phi^*)^2]}{\mathbb{E}[\phi^2] + 2\mathbb{E}[\phi\phi^*] + \mathbb{E}[(\phi^*)^2]} \\ &= -\left(\frac{1}{2\pi}\right)^2 \frac{\Re\{\mathbb{E}[\phi^2]\} - \mathbb{E}[\phi\phi^*]}{\Re\{\mathbb{E}[\phi^2]\} + \mathbb{E}[\phi\phi^*]}. \end{aligned} \quad (2.33)$$

The last equality in (2.33) comes from the fact that  $\mathbb{E}[(\phi^*)^2] = (\mathbb{E}[\phi^2])^*$ . The second approximation in (2.33) can be justified using similar arguments as in [16]. A close examination of the fraction  $(\phi - \phi^*)^2/(\phi + \phi^*)^2$  in (2.33) illustrates that the mean of its denominator is much larger than the mean of its numerator, and the standard

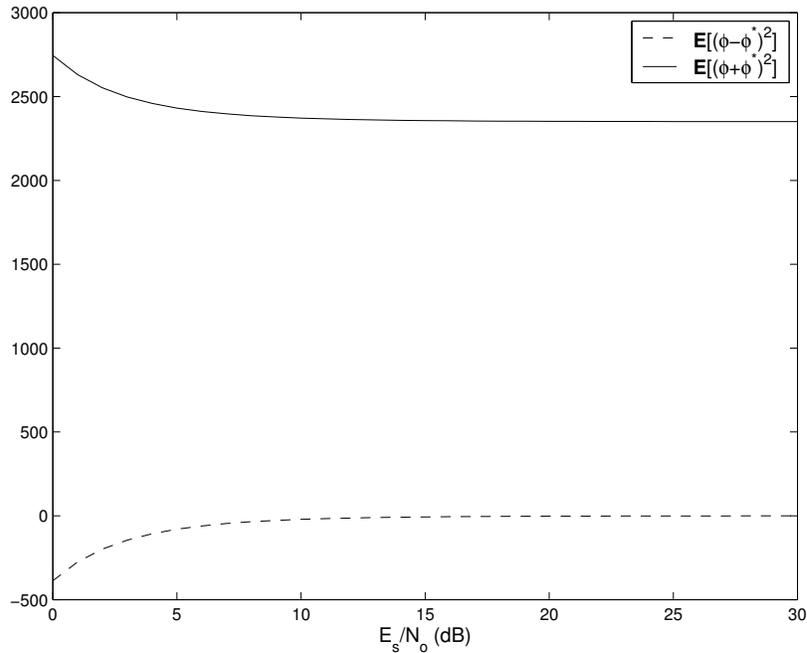


Fig. 3. Plots of  $\mathbb{E}[(\phi - \phi^*)^2]$  and  $\mathbb{E}[(\phi + \phi^*)^2]$  as a function of  $E_s/N_o$  for  $\varepsilon_o = 0, 0.25, 0.5$  and  $0.75$  ( $g(t)$  is a square-root raised cosine pulse with  $\alpha = 0.5, Q = 2, K = 4, L_o=100$  and  $L_g = 3$ ). Note that all curves for different values of  $\varepsilon_o$  overlap.

deviations of its numerator and denominator are in general much smaller than the mean of denominator. Fig. 3 plots  $\mathbb{E}[(\phi - \phi^*)^2]$  and  $\mathbb{E}[(\phi + \phi^*)^2]$  as a function of  $E_s/N_o$  when  $g(t)$  is a square-root raised cosine pulse with  $\alpha = 0.5, Q = 2, K = 4, L_o=100$  and  $L_g = 3$  for  $\varepsilon_o = 0, 0.25, 0.5$  and  $0.75$ . Note that all curves for different values of  $\varepsilon_o$  overlap. It can be seen that for  $E_s/N_o \geq 10\text{dB}$ ,  $\mathbb{E}[(\phi - \phi^*)^2]$  is much smaller than  $\mathbb{E}[(\phi + \phi^*)^2]$ . The same result can be obtained for different other pulse shapes  $g(t)$ . Also, one can check that at medium and high  $E_s/N_o$ , the standard deviations of  $(\phi - \phi^*)^2$  and  $(\phi + \phi^*)^2$  are small relative to  $\mathbb{E}[(\phi + \phi^*)^2]$ . All these considerations justify the second approximation made in (2.33).

From (2.32), we note that

$$\mathbb{E}[\phi^2] = e^{j4\pi\varepsilon_o} \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} \mathbb{E}[\Lambda(k_1)\Lambda(k_2)] e^{-j2\pi k_1/K} e^{-j2\pi k_2/K}, \quad (2.34)$$

$$\mathbb{E}[\phi\phi^*] = \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} \mathbb{E}[\Lambda(k_1)\Lambda^*(k_2)] e^{-j2\pi k_1/K} e^{j2\pi k_2/K}. \quad (2.35)$$

It is proved in Appendix A that

$$\mathbb{E}[\Lambda(k_1)\Lambda(k_2)] = \text{tr}[\mathbf{B}_{k_1}^T \mathbf{R}_{\varepsilon_o}] \text{tr}[\mathbf{B}_{k_2}^T \mathbf{R}_{\varepsilon_o}] + \text{tr}[\mathbf{B}_{k_1}^T \mathbf{R}_{\varepsilon_o} \mathbf{B}_{k_2}^T \mathbf{R}_{\varepsilon_o}] + c(k_1, k_2), \quad (2.36)$$

$$\mathbb{E}[\Lambda(k_1)\Lambda^*(k_2)] = \text{tr}[\mathbf{B}_{k_1}^T \mathbf{R}_{\varepsilon_o}] \text{tr}[\mathbf{B}_{k_2} \mathbf{R}_{\varepsilon_o}] + \text{tr}[\mathbf{B}_{k_1}^T \mathbf{R}_{\varepsilon_o} \mathbf{B}_{k_2} \mathbf{R}_{\varepsilon_o}] + c(k_1, k_2), \quad (2.37)$$

where  $\text{tr}[\cdot]$  denotes the trace of a matrix,  $\mathbf{R}_\varepsilon \triangleq \frac{E_s}{T} \mathbf{G}_\varepsilon + \frac{N_o Q}{T} \mathbf{I}_{L_o Q}$ . In (2.36) and (2.37),

$$\mathbf{B}_k \triangleq \mathbf{A}_\varepsilon (\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon)^{-1} \mathbf{A}_\varepsilon^H |_{\varepsilon=k/K}, \quad (2.38)$$

$\mathbf{G}_\varepsilon$  is an  $L_o Q \times L_o Q$  matrix with the  $(i, j)^{\text{th}}$  element ( $i, j = 0, 1, \dots, L_o Q - 1$ ) given by

$$[\mathbf{G}_\varepsilon]_{ij} \triangleq \sum_{n=-\infty}^{\infty} g^*(iT/Q - nT - \varepsilon T) g(jT/Q - nT - \varepsilon T), \quad (2.39)$$

and

$$c(k_1, k_2) \triangleq \begin{cases} \frac{E_s^2}{T^2} (m_4 - 2) \sum_{n=-\infty}^{\infty} [\mathbf{a}_n(\varepsilon_o)^H \mathbf{B}_{k_1} \mathbf{a}_n(\varepsilon_o)] [\mathbf{a}_n(\varepsilon_o)^H \mathbf{B}_{k_2} \mathbf{a}_n(\varepsilon_o)] \\ \hspace{15em} \text{for linear modulations} \\ \frac{E_s^2}{T^2} \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} (-1)^{|n_1 - n_2|} [\mathbf{a}_{n_1}(\varepsilon_o)^H \mathbf{B}_{k_1} \mathbf{a}_{n_2}(\varepsilon_o)] [\mathbf{a}_{n_1}(\varepsilon_o)^H \mathbf{B}_{k_2} \mathbf{a}_{n_2}(\varepsilon_o)] \\ \quad - 2 \sum_{n=-\infty}^{\infty} [\mathbf{a}_n(\varepsilon_o)^H \mathbf{B}_{k_1} \mathbf{a}_n(\varepsilon_o)] [\mathbf{a}_n(\varepsilon_o)^H \mathbf{B}_{k_2} \mathbf{a}_n(\varepsilon_o)] \\ \hspace{15em} \text{for MSK and GMSK} \end{cases} \quad (2.40)$$

where  $m_4 = \mathbb{E}[|d_i|^4]$  is the fourth order moment of the transmitted symbols, which is fixed for a specific constellation (e.g.,  $m_4 = 1$  for PSK and  $m_4 > 1$  for QAM).

Therefore, the MSE for a specific delay  $\varepsilon_o$  can be found by using (2.33)-(2.37). As the symbol timing delay  $\varepsilon_o$  is assumed to be uniformly distributed in  $[0, 1)$ , the average MSE is calculated by numerical integration of (2.33).

Notice that the MSE expressions in this section can only be regarded as an approximated analysis for GMSK since only the principle component of LE is taken into consideration. However, from the results to be presented in next section, excellent agreement between analytical expressions and simulations can be observed (see Fig. 9), only a small deviation occurs at very high SNRs.

## F. Simulation Results and Discussions

In this section, the performance of the proposed algorithm and other existing symbol timing estimators are assessed by Monte Carlo simulations and then compared with the analytical results derived in the last section, the CCRB<sup>1</sup> [10] and the MCRB [34]. In all the simulations, the observation length is fixed to  $L_o=100$ , and  $\varepsilon_o$  is uniformly distributed in the range  $[0, 1)$ .  $\theta_o$  is generated as a uniformly distributed random variable in the range  $[-\pi, \pi)$  and is constant in each estimation. Each point is obtained by averaging  $10^4$  simulation runs. In all figures, the CCRB and the MCRB are plotted as references.

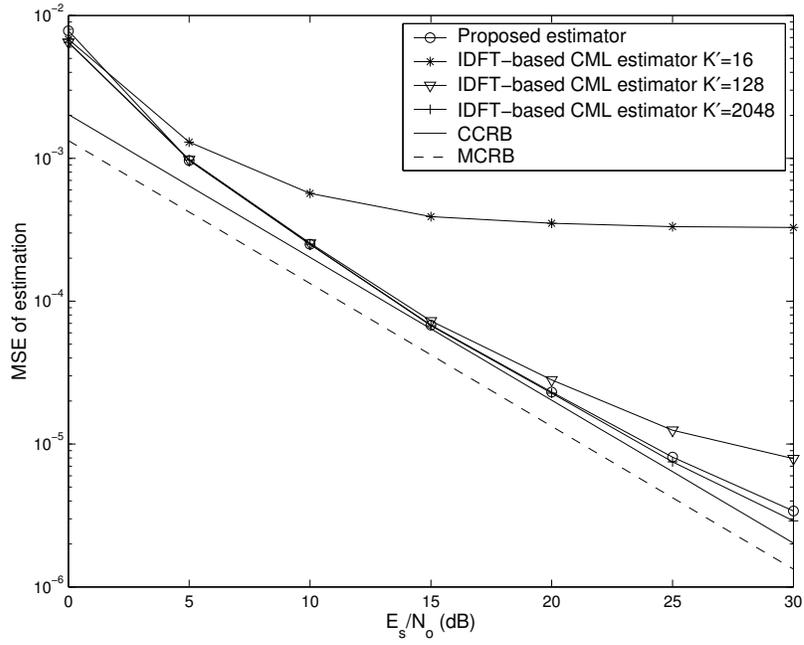
First consider the case of linear modulations. QPSK is chosen as the symbol constellation. The oversampling ratio for the proposed estimator is  $Q=2$ ,  $g(t)$  is the square-root raised cosine pulse with roll-off factor  $\alpha=0.5$ , the number of ISI symbols introduced by one side of  $g(t)$  is assumed to be  $L_g=3$ . Figs. 4a and 4b show the MSE against  $E_s/N_o$  for the proposed algorithm and the IDFT-based CML estimator for  $K = 4$  and  $K = 8$ , respectively. It can be seen (from both Figs. 4a and 4b)

---

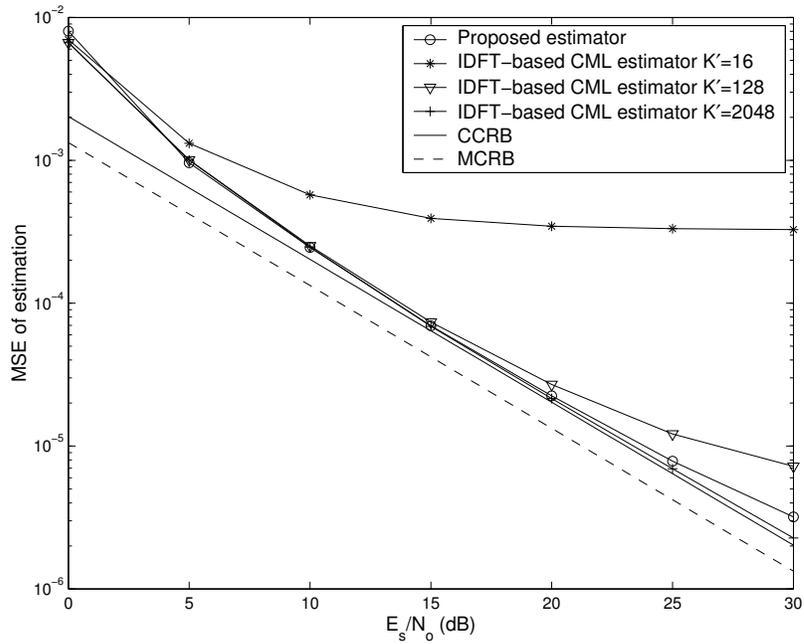
<sup>1</sup>Strictly speaking, this bound is the asymptotic CCRB. However, it is shown in the simulation results that this bound can be reached for finite observation length.

that the proposed algorithm has a performance similar to that of IDFT-based CML estimator with  $K'=2048$ . This further justifies the approximation in (2.23). Note that, for  $K = 4$ , the self noise is not completely eliminated for both IDFT-based CML estimator and the proposed estimator (as seen from the MSE departure from CCRB at high SNR in Fig. 4a). This can be explained as follows. For the IDFT-based CML estimator, the self-noise is due to the small value of  $K$  chosen, since in the derivation, it is assumed that the CML function can be *completely* represented by  $K$  samples. However, there is no guarantee that  $K = 4$  is sufficient (although  $K = 4$  results in pretty good performance). Increasing the value of  $K$  to 8 removes the self-noise of the IDFT-based CML estimator (with  $K' = 2048$ ) as shown in Fig. 4b. For the proposed estimator, although it is also required that  $K$  should be large enough such that  $\Lambda(\varepsilon)$  can be represented by its samples, the self-noise is due to another more critical factor – the approximation (2.23) in the CML function. This can be seen from the fact that the performance of the proposed estimator does not improve by increasing  $K$  from 4 to 8 (compare Figs. 4a and 4b). As  $K = 4$  is good enough for the proposed estimator,  $K = 4$  is used for the rest of the chapter.

Fig. 5 illustrates the very close match between the simulation and the analytical results derived in the last section. It is also clear that, for the SNRs under consideration, the performance of the proposed algorithm is very close to the CCRB, which means that the proposed estimator almost reaches the ultimate performance of the CML principle. Furthermore, the CCRB is close to the MCRB. Since MCRB is a lower bound on the variance of *any* unbiased estimate, this shows that the proposed algorithm is close to optimal for a wide range of  $E_s/N_o$ . Notice that at  $E_s/N_o$  around 30dB, an MSE floor begins to occur (due to the approximation (2.23) in the CML function), but at that high SNR, the estimation MSE achieved by the proposed estimator is already very small (on the order of  $10^{-6}$ ), so the effect of the MSE floor



(a)



(b)

Fig. 4. MSE of the proposed estimator and the IDFT-based CML estimator with (a)  $K = 4$ , (b)  $K = 8$  (QPSK,  $Q=2$ ,  $\alpha=0.5$ ,  $L_o=100$  and  $L_g=3$ ).

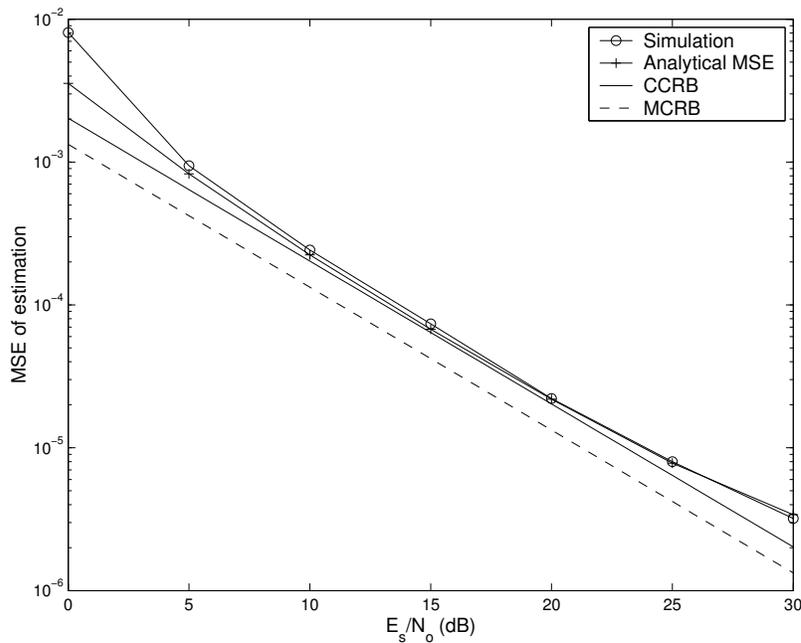


Fig. 5. Comparison between analytical MSE and simulations of the proposed estimator (QPSK,  $Q=2$ ,  $K=4$ ,  $\alpha=0.5$ ,  $L_o=100$  and  $L_g=3$ ).

becomes relatively less critical.

Fig. 6 compares the performance of the proposed estimator with that of the square nonlinearity estimator (with  $Q=4$ ) [3]. It is apparent that the proposed estimator outperforms the square nonlinearity estimator, especially at high  $E_s/N_o$ . This is because for finite observation length,  $\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \neq \mathbf{I}$ , and the self-noise is better cancelled by the matrix  $(\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon)^{-1}$  than  $\mathbf{I}$ . Fig. 7 compares the performances of the proposed algorithm with the existing state-of-the-art feedforward algorithms that require only two samples per symbol to operate: Mengali [34, pp.401], Zhu *et al.* [35] and Lee [36]. It can be seen that while the performances for different algorithms are similar at low  $E_s/N_o$ , the proposed algorithm has the smallest MSE at high  $E_s/N_o$ .

Next consider that MSK is the modulation format. Fig. 8 shows the performances of the proposed estimator (with  $Q = 2$ ,  $Q = 4$  and  $Q = 8$ ) and the low-SNR approximated maximum likelihood (ML) algorithm [37] for MSK. The number of ISI

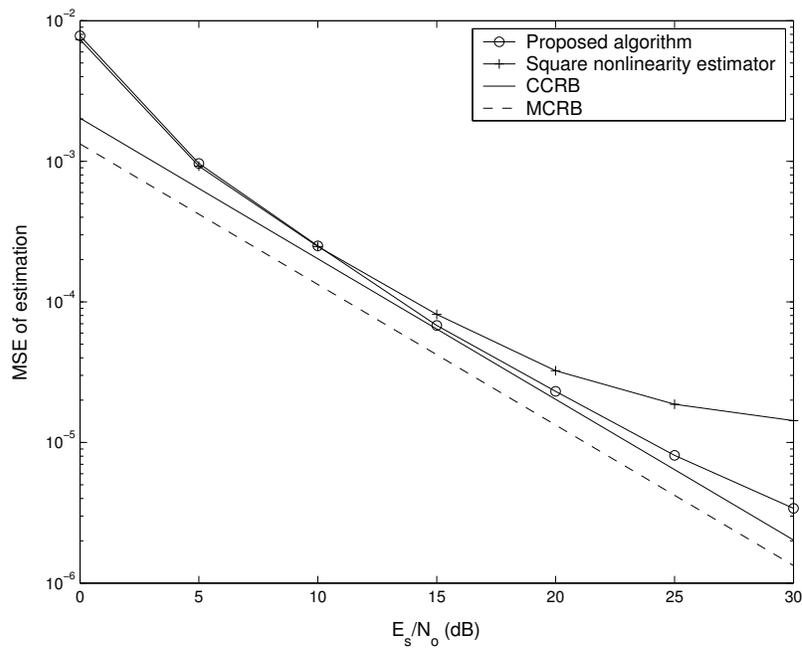


Fig. 6. Comparison of the MSE of the proposed estimator and the square nonlinearity estimator (QPSK,  $\alpha=0.5$ ,  $L_o = 100$  and  $L_g=3$ ).

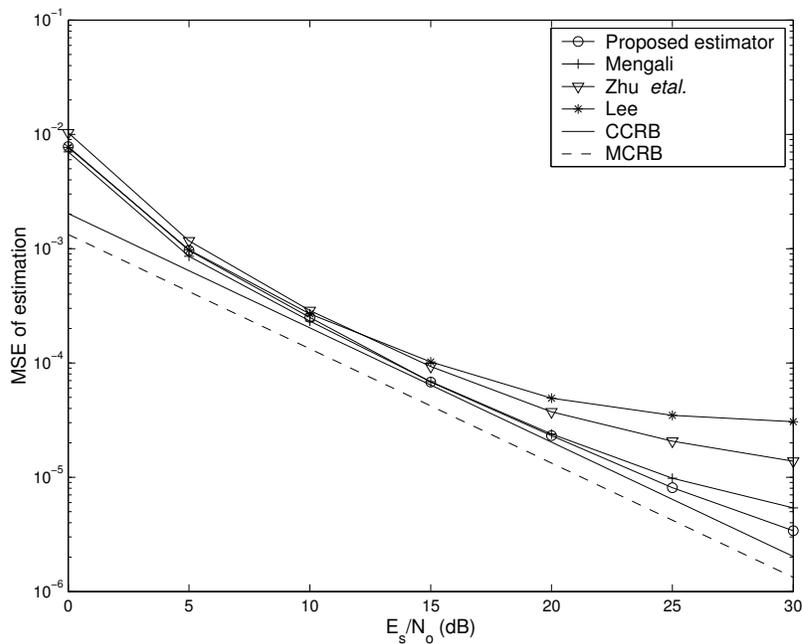


Fig. 7. MSE for the proposed estimator, the algorithms in [34, pp.401], [35] and [36] (QPSK,  $Q = 2$ ,  $\alpha=0.5$ ,  $L_o=100$  and  $L_g=3$ ).

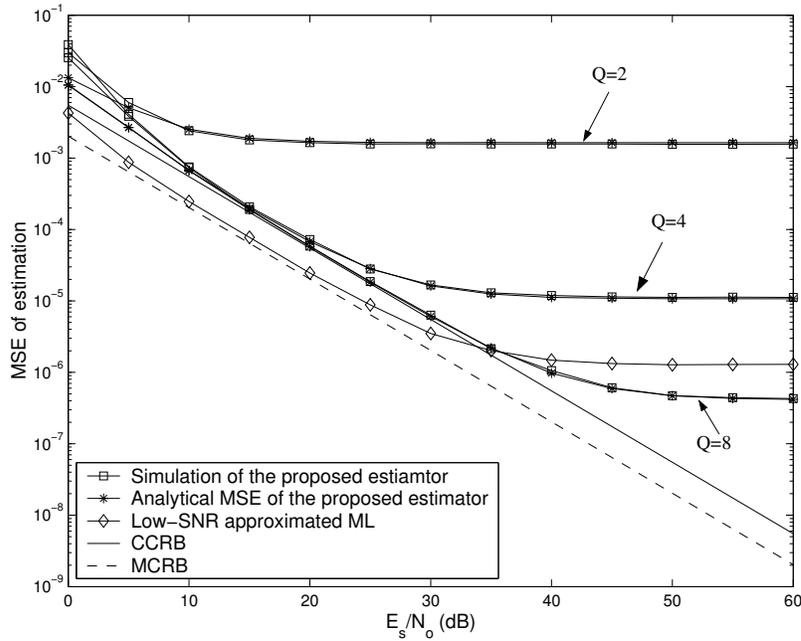


Fig. 8. MSE of the proposed estimator ( $Q = 2$ ,  $Q = 4$  and  $Q = 8$ ) and the low-SNR approximated ML algorithm [37] for MSK ( $L_{\bar{g}} = 1$  and  $L_o=100$ ).

symbols introduced by one side of  $\bar{g}(t)$  is assumed to be 1. The following observations can be inferred from Fig. 8. First, it can be seen that, for the proposed algorithm, the higher the oversampling ratio, the better the performance. This is because the pulse  $\bar{g}(t)$  is time limited [38], so its frequency response is not bandlimited; a higher oversampling ratio reduces the aliasing and thus provides better performance. Second, the theoretical MSE analysis matches very well the simulation results. Third, although a higher oversampling ratio increases the range of SNR's over which the performance of the proposed estimator comes close to the CCRB, MSE floors still occur at high SNR's due to the approximation (2.23) assumed in the derivation of estimator. Furthermore, the CCRB is far away from the MCRB and the simulation results show that the low-SNR approximated ML algorithm [37] approaches the MCRB. Therefore, direct application of the CML principle is not suitable for the MSK modulation.

Now let consider the GMSK modulation. Fig. 9 show the performances of the proposed estimator (with  $Q = 2$  and  $Q = 4$ ) and the low-SNR approximated ML algorithm [37] for GMSK with pre-modulator bandwidth  $BT = 0.5$ . The number of ISI symbols introduced by one side of  $\bar{g}(t)$  is assumed to be 2. Notice that although the proposed estimator is based on the approximated linear model (2.9), the GMSK signal in the simulation is generated according to (2.7) without approximation. The MCRB for GMSK is exact and its expression can be found in [34]. For the CCRB, it is based on the approximated linear model (2.9). Although the resulting CCRB is not exact, it is still a valid lower bound for the proposed estimator since when the proposed estimator is applied to the true GMSK signal, the ignored components in LE would become interferences and the performances would be poorer than that predicted by the CCRB, which assumes no interference from other components of LE. Note that the CCRB obtained by expressing the GMSK signal using all the LE components (as done in [10]) is not applicable here, since in that case, the resultant CCRB is *conditioned* on the fact that all the pseudo-symbols are being jointly estimated together with the unknown timing offset, while in the proposed estimator, only the pseudo-symbols related to the first LE component are estimated.

From Fig. 9, it can be seen that, for the proposed estimator, a higher oversampling ratio also results in better performance due to the same reason as in the case of MSK modulation. However, by comparing Figs. 8 and 9, if the same oversampling ratio is used, it is found that the performance of the proposed estimator for GMSK modulation is better than that corresponding to MSK. This is due to the fact that the pulse  $\bar{g}(t)$  is longer in GMSK than in MSK (although they both are time-limited), therefore, with the same oversampling ratio, the aliasing introduced in GMSK is smaller than that in MSK. Second, it is obvious that the analytical MSE expressions derived in the last section match very well with the simulation results.

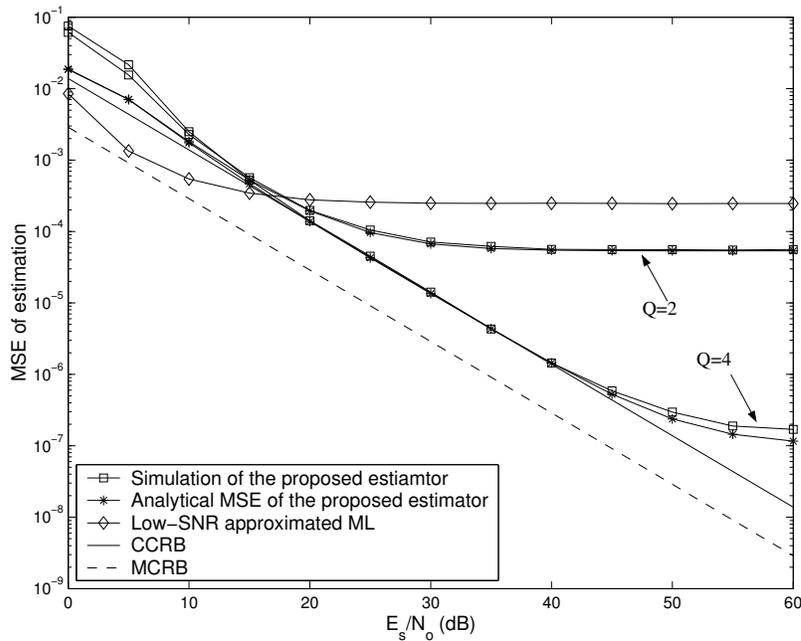


Fig. 9. MSE of the proposed estimator ( $Q = 2$  and  $Q = 4$ ) and the low-SNR approximated ML algorithm [37] for GMSK with  $BT = 0.5$  ( $L_{\bar{g}} = 2$  and  $L_o = 100$ ).

Only for the case of  $K = 4$  and at SNR=50-60 dB, the analytical MSE expressions predict a slightly better performance than simulations. Third, the performance of the proposed estimator with  $Q = 4$  comes very close to the CCRB for  $E_s/N_o \leq 40$  dB. The MSE floor, which is caused by the approximation (2.23) in the CML function, begins to occur only for  $E_s/N_o > 40$  dB. Notice that the effect of the approximation (2.9) in the system model (which results in the gap between analytical MSE and simulations) is much smaller than that of approximation (2.23) in the CML function (which causes the MSE floor). Compared to the low-SNR approximated ML algorithm [37], at low SNR's the proposed estimator exhibits poorer performance, but for medium and high SNR's, the proposed estimator performs much better.

Finally, notice that the CCRB is far away from the MCRB as in the case of MSK. Since the CCRB is a valid bound only for estimators that rely on quadratic nonlinearities [10], it is expected that algorithms exploiting higher-order ( $> 2$ ) nonlinearities

might exist with performances closer to the MCRB. An example of such an algorithm is the low-SNR approximated ML algorithm [37], for which we already demonstrated that its performance is closer to the MCRB at low SNRs. The next question is whether there is an estimator whose performance comes close to the MCRB for a larger range of SNRs. This is a subject open to future investigations.

## G. Conclusions

A new feedforward symbol-timing estimator based on the conditional maximum likelihood principle was proposed. An approximation was applied in the Fourier series expansion of the CML function so that the complexity of the proposed estimator is greatly reduced. It was shown, analytically and via simulations, that the performances of the proposed estimator for linear modulations are in general very close to the CCRB and MCRB for  $\text{SNR} \leq 30\text{dB}$ . For higher SNRs, MSE floors occur, but notice that at these high SNRs, the MSE achieved by the proposed estimator is already very small, so the effect of MSE floors becomes relatively less critical. Furthermore, for linear modulations with the transmit pulse being a square-root raised cosine pulse, the proposed estimator was shown to be asymptotically equivalent to the well-known square nonlinearity estimator [3]. However, in the presence of reduced number of samples, the proposed estimator presents better performance than [3]. For MSK and GMSK modulations, it was found that although the performances of the proposed estimator come very close to the CCRB at certain SNR ranges, however, the CCRB is quite far away from the MCRB, and there exists an alternative algorithm that come closer to the MCRB. Therefore, it was concluded that the proposed estimator is more suitable for linear modulations than MSK and GMSK modulations.

## CHAPTER III

UNIFIED ANALYSIS OF A CLASS OF BLIND FEEDFORWARD SYMBOL  
TIMING ESTIMATORS EMPLOYING SECOND-ORDER STATISTICS

## A. Introduction

The problem of digital blind feedforward symbol timing estimation assumes recovery of the timing delay of the received signal based on the oversampled and unsynchronized received samples. Many algorithms were proposed in the literature to solve this problem. The earliest one is [3], in which the well-known square law nonlinearity (SLN) estimator is proposed. Several extensions of the SLN estimator were later reported in [5], [6] and [42], in which they consider more general second-order statistics than just squaring. Recently, the pre-filtering technique was applied to the SLN estimator to improve its performances at medium and high signal-to-noise ratios (SNRs) [43]. Other than square nonlinearity, estimators based on other types of nonlinearities were also proposed. Reference [8] proposed a logarithmic nonlinearity estimator, which is obtained by applying a low SNR approximation to the maximum likelihood (ML) method; absolute value and fourth order nonlinearities-based estimators were proposed in [4]; a combination of square and fourth order nonlinearities was proposed in [7] to take advantage of the properties of both types of nonlinearities. While all the above estimators require at least 3 samples per symbol, there are also some estimators that require only two samples per symbol. Reference [36] proposed such an estimator based on an ad-hoc argument. This estimator was later modified to remove its asymptotic bias in [44], and the pre-filtering technique was also applied to this estimator in [43] to improve its performance at medium and high SNRs. On the other hand, in Chapter II, we have seen a new two samples per symbol estimator

based on the conditional maximum likelihood (CML) principle.

With so many estimators, designed using different philosophies and their performances analyzed independently under different assumptions, one would wonder whether we can have a general framework to analyze the performances of these estimators so that a fair and easy comparison can be made. This question was partially answered in [4], in which a technique for evaluating the jitter performance of symbol timing estimators employing a zero-memory, general type of nonlinearity was presented. In this chapter, we analyze a different class of estimators (from that considered in [4]) by formulating all the blind feedforward symbol timing estimators employing second-order statistics (which include the estimators in [3], [5], [6], [36], [42], [43], [44] and Chapter II) into a single estimation framework, and then by deriving the finite sample (as opposed to the asymptotic performance reported in [7]) mean-square error (MSE) expression for this class of estimators. The MSE expression for any individual estimator can be obtained from the general expression by setting suitable parameters. The analytical results are compared with the computer simulation results, and it is found that both sets of results match very well. Furthermore, it is found that within the class of estimators employing second-order statistics, the SLN estimator with a properly designed pre-filter and the feedforward CML estimator perform the best and their performances coincide with the conditional Cramer-Rao bound (CCRB) [9], which is the performance lower bound for the class of estimators under consideration.

## B. Unified Formulation for the Class of Symbol Timing Estimators Employing the Second-Order Statistics

For linear modulations transmitted through AWGN channels, the received signal can be written as

$$r(n) \triangleq r(t)|_{t=nT_s} = e^{j\theta_o} \sqrt{E_s/T} \sum_i d_i p(nT_s - iT - \varepsilon_o T) + \eta(nT_s), \quad (3.1)$$

where  $\theta_o$  is the unknown phase offset;  $E_s$  is the symbol energy;  $d_i$  stands for the zero-mean unit variance, independently and identically distributed (i.i.d.) complex valued symbols being transmitted;  $p(t) \triangleq g(t) \star g_r(t)$  is the combined response of the unit energy transmit filter  $g(t)$  and the receiving filter  $g_r(t)$ ;  $T$  is the symbol period;  $T_s \triangleq T/Q$  with  $Q$  being the oversampling ratio;  $\varepsilon_o \in [0, 1)$  is the unknown symbol timing delay to be estimated and  $\eta(nT_s)$  stands for the samples of filtered noise. It is assumed that the noise samples before receive filtering is complex-valued circularly distributed Gaussian with power density  $N_o$ .

In this chapter, we consider the class of estimators taking the following general form:

$$\hat{\varepsilon} = -\frac{1}{2\pi} \arg \left\{ \sum_{k=0}^{K-1} \Lambda(k) e^{-j2\pi k/K} \right\}, \quad (3.2)$$

where  $\Lambda(k) = \mathbf{r}^H \mathbf{B}_k \mathbf{r}$  with  $\mathbf{r} \triangleq [r(0), r(1), \dots, r(L_o Q - 1)]^T$  is the observation vector of length  $L_o$  symbols and  $\mathbf{B}_k$  is a fixed matrix of dimension  $L_o Q \times L_o Q$ . The block diagram of this general estimator is shown in Fig. 10. After the observation vector is formed,  $K$  samples of  $\Lambda(k)$  are calculated, then a  $K$ -point discrete Fourier transform (DFT) is taken, and the phase of the first bin (i.e., second output) of the DFT scaled by a constant is the timing estimate. Let us now consider some special cases.

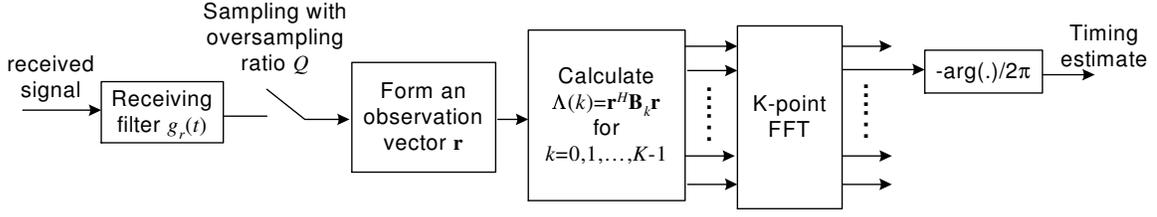


Fig. 10. Block diagram of the general estimator.

### 1. Cyclic Correlation-Based Estimator

The cyclic correlation-based estimator [5], [7] is given by

$$\hat{\varepsilon} = -\frac{1}{2\pi} \arg \left\{ \sum_{n=0}^{L_o Q - \tau - 1} r^*(n) r(n + \tau) e^{-j\pi\tau/Q} e^{-j2\pi n/Q} \right\}, \quad (3.3)$$

for  $Q \geq 3$  and some integer lag  $\tau \geq 0$ . Note that different values of  $\tau$  result in different previously proposed estimators in the literature ( $\tau = 0$  corresponds to the estimators proposed in [3] and [42],  $\tau = Q$  corresponds to the estimator in [6]).

If we decompose the summation term in (3.3) into  $Q$  polyphase components and define  $n_u(k) \triangleq \lfloor (L_o Q - \tau - 1 - k)/Q \rfloor$ , we have

$$\hat{\varepsilon} = -\frac{1}{2\pi} \arg \left\{ \underbrace{\sum_{k=0}^{Q-1} e^{-j\pi\tau/Q} \sum_{n=0}^{n_u(k)} r^*(nQ + k) r(nQ + k + \tau) e^{-j2\pi(nQ)/Q} e^{-j2\pi k/Q}}_{\Lambda_{CC}(k)} \right\}. \quad (3.4)$$

It can be seen that the cyclic correlation-based estimator takes the form of (3.2) with  $K = Q$ . Expressing  $\Lambda_{CC}(k)$  into matrix form, we have  $\Lambda_{CC}(k) = \mathbf{r}^H \mathbf{B}_k^{CC} \mathbf{r}$ , where  $\mathbf{B}_k^{CC}$  is a  $L_o Q \times L_o Q$  matrix with its  $(nQ + k, nQ + k + \tau)^{th}$  element ( $n = 0, 1, \dots, n_u(k)$ ) equal to  $e^{-j\pi\tau/Q}$  and other elements equal zero. In particular, for the well-known SLN estimator [3] (i.e.,  $\tau = 0$ ,  $Q = 4$ ), we have (for  $i = 0, 1, 2, 3$ )

$$\mathbf{B}_i^{SLN} = \mathbf{I}_{L_o} \otimes \mathbf{E}_4^{(i)} \quad (3.5)$$

where  $\otimes$  denotes the Kronecker product and  $\mathbf{E}_k^{(i)}$  is a  $k \times k$  matrix with the  $(i, i)^{th}$  element equals one and other elements equal zero.

## 2. Lee's Estimator and the Modified Estimator

A two samples per symbol estimator was proposed by Lee in [36]. Later, this estimator was modified to remove its asymptotic bias [44]. The modified version of Lee's estimator can be written as

$$\hat{\varepsilon} = -\frac{1}{2\pi} \arg \left\{ \gamma \sum_{n=0}^{L_o Q-1} |r(n)|^2 e^{jn\pi} + \sum_{n=0}^{L_o Q-2} \Re[r^*(n)r(n+1)] e^{j(n-0.5)\pi} \right\}, \quad (3.6)$$

with  $Q = 2$  and  $\gamma$  is a constant depending on the pulse shape  $p(t)$ . If  $p(t)$  is a raised cosine pulse with roll-off factor  $\alpha$ , then  $\gamma = 8 \sin(\pi\alpha/2)/(\alpha\pi(4-\alpha^2))$  [44]. The original Lee's estimator can be obtained by setting  $\gamma = 1$ . Now rewrite the expression in the  $\arg\{ \}$  of (3.6) as follows:

$$\begin{aligned} & \underbrace{\gamma \sum_{n=0}^{L_o-1} |r(nQ)|^2}_{\Lambda_{Lee}(0)} + \underbrace{\sum_{n=0}^{L_o-1} \Re[r^*(nQ)r(nQ+1)]}_{\Lambda_{Lee}(1)} e^{-j\pi/2} \\ & + \gamma \underbrace{\sum_{n=0}^{L_o-1} |r(nQ+1)|^2}_{\Lambda_{Lee}(2)} e^{-j\pi} + \underbrace{\sum_{n=0}^{L_o-2} \Re[r^*(nQ+1)r(nQ+2)]}_{\Lambda_{Lee}(3)} e^{-j3\pi/2}. \end{aligned} \quad (3.7)$$

Therefore, the estimator in (3.6) can also be expressed in the form of (3.2) with  $K = 4$ . With the fact that  $\Re(x) = (x + x^*)/2$  and expressing  $\Lambda_{Lee}(k)$  in matrix form, we have  $\Lambda_{Lee}(k) = \mathbf{r}^H \mathbf{B}_k^{Lee} \mathbf{r}$ , where

$$\begin{aligned} \mathbf{B}_0^{Lee} &= \gamma \mathbf{I}_{L_o} \otimes \mathbf{E}_2^{(0)}, & \mathbf{B}_1^{Lee} &= 0.5 \mathbf{I}_{L_o} \otimes \mathbf{J}_2, & \mathbf{B}_2^{Lee} &= \gamma \mathbf{I}_{L_o} \otimes \mathbf{E}_2^{(1)}, \\ \mathbf{B}_3^{Lee} &= \begin{bmatrix} 0 & \mathbf{0}_{1 \times 2(L_o-1)} & 0 \\ \mathbf{0}_{2(L_o-1) \times 1} & 0.5 \mathbf{I}_{L_o-1} \otimes \mathbf{J}_2 & \mathbf{0}_{2(L_o-1) \times 1} \\ 0 & \mathbf{0}_{1 \times 2(L_o-1)} & 0 \end{bmatrix} \end{aligned} \quad (3.8)$$

where  $\mathbf{J}_2$  is the counter identity matrix, constructed by flipping  $\mathbf{I}_2$  from left to right.

### 3. Feedforward CML Estimator

The feedforward symbol timing estimator based on the conditional ML principle was proposed in Chapter II. Unfortunately, the results in Chapter II cannot be directly applied here since the original estimator was derived under the assumption that the noise samples are independent of each other, but in the signal model (3.1), the noise samples are correlated due to the receiver filtering. Of course, one can always start from the probability density function of  $\mathbf{r}$ , taking into account the correlation of noise and re-derive the estimator following the idea of Chapter II. However, a faster but equivalent method is as follows. Since the correlations between noise samples are related to the receiving filter (which is known), we can whiten the filtered noise samples by pre-multiplying the observation vector  $\mathbf{r}$  with  $(\boldsymbol{\varphi}^{-1/2})^H$ , where  $\boldsymbol{\varphi}$  is the correlation matrix of the noise vector (with its elements given by  $[\boldsymbol{\varphi}]_{ij} = \int_{-\infty}^{\infty} g_r^*(t)g_r(t - (i - j)T/Q)dt$ ) and  $\boldsymbol{\varphi}^{-1/2}$  denotes any square root of  $\boldsymbol{\varphi}^{-1}$  (e.g., Cholesky decomposition) such that  $\boldsymbol{\varphi}^{-1/2}(\boldsymbol{\varphi}^{-1/2})^H = \boldsymbol{\varphi}^{-1}$ . Then the results of Chapter II can be applied readily to this transformed observation vector  $(\boldsymbol{\varphi}^{-1/2})^H \mathbf{r}$ . It turns out that the resultant feedforward CML symbol timing estimator is given by  $\hat{\varepsilon} = -\frac{1}{2\pi} \arg \left\{ \sum_{k=0}^{K-1} \Lambda_{CML}(k) e^{-j2\pi k/K} \right\}$ , where  $K \geq 3$  and  $\Lambda_{CML}(k) = \mathbf{r}^H \mathbf{B}_k^{CML} \mathbf{r}$  with

$$\mathbf{B}_k^{CML} \triangleq \boldsymbol{\varphi}^{-1} \mathbf{A}_\varepsilon (\mathbf{A}_\varepsilon^H \boldsymbol{\varphi}^{-1} \mathbf{A}_\varepsilon)^{-1} \mathbf{A}_\varepsilon^H \boldsymbol{\varphi}^{-1} \Big|_{\varepsilon=k/K}, \quad (3.9)$$

$$\mathbf{A}_\varepsilon \triangleq [\mathbf{a}_{-L_p}(\varepsilon), \mathbf{a}_{-L_p+1}(\varepsilon), \dots, \mathbf{a}_{L_o+L_p-1}(\varepsilon)], \quad (3.10)$$

$$\mathbf{a}_i(\varepsilon) \triangleq [p(-iT - \varepsilon T), p(T_s - iT - \varepsilon T), \dots, p((L_o Q - 1)T_s - iT - \varepsilon T)]^T \quad (3.11)$$

and  $L_p$  denotes the number of symbols affected by the inter-symbol interference (ISI) introduced by one side of  $p(t)$ . Notice that if  $g_r(t) = \delta(t)$ , this estimator would reduce



would reduce to that of the original estimator.

### C. Performance Analysis

#### 1. Performance Bound

In [9], the asymptotic CCRB was introduced for symbol timing estimation problem. The asymptotic CCRB is a lower bound tighter than the modified Cramer-Rao bound (MCRB), but still a valid lower bound on the variance of any consistent estimator that is quadratic with respect to the received signal (which is the class of estimators under consideration). However, the asymptotic CCRB in [9] was derived assuming white Gaussian noise samples, therefore, the whitening technique similar to that in Section B3 of this chapter has to be applied in order to include the effect of the receiving filter. Applying the results of [9] to the transformed observation vector  $(\boldsymbol{\varphi}^{-1/2})^H \mathbf{r}$ , it can be shown that for fixed  $\varepsilon_o$ ,

$$\text{CCRB}^{as}(\varepsilon_o) = \frac{1}{2\text{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^H \boldsymbol{\Omega}_{\varepsilon_o} \tilde{\mathbf{D}}_{\varepsilon_o})} \left( \frac{E_s}{N_o} \right)^{-1} \quad (3.14)$$

where  $\tilde{\mathbf{D}}_{\varepsilon} \triangleq \frac{1}{\sqrt{Q}} d\mathbf{A}_{\varepsilon}/d\varepsilon$  and  $\boldsymbol{\Omega}_{\varepsilon} \triangleq \boldsymbol{\varphi}^{-1} - \boldsymbol{\varphi}^{-1} \mathbf{A}_{\varepsilon} (\mathbf{A}_{\varepsilon}^H \boldsymbol{\varphi}^{-1} \mathbf{A}_{\varepsilon})^{-1} \mathbf{A}_{\varepsilon}^H \boldsymbol{\varphi}^{-1}$ . Since the symbol timing delay  $\varepsilon_o$  is assumed to be uniformly distributed in  $[0, 1)$ , the average asymptotic CCRB can be calculated by numerical integration of (3.14). Notice that although the value of  $Q$  appears in the expression of the asymptotic CCRB, numerical computations show that the asymptotic CCRB is independent of  $Q$  as long as  $Q \geq 2$ . This is reasonable, since the asymptotic CCRB is derived under the assumption of white noise, oversampling the signal above Nyquist rate does not provide any new information, thus the ultimate performance cannot be improved.

## 2. MSE Expression

In this section, we present the MSE expression for the general estimator (3.2). The derivation procedures follow closely to that in Chapter II. The only difference is that, the MSE expression in Chapter II was derived under the assumption of white noise, while in this chapter, the correlation of noise has to be taken into consideration. This can be easily done by modifying just a few lines of the derivations in Appendix A, and therefore only the results are presented here. It can be shown that for a true timing delay  $\varepsilon_o$ , the MSE of the general estimator (3.2) is given by

$$\text{MSE}(\varepsilon_o) \triangleq \mathbb{E}[(\hat{\varepsilon} - \varepsilon_o)^2] \cong - \left( \frac{1}{2\pi} \right)^2 \frac{\Re\{\phi_1\} - \phi_2}{\Re\{\phi_1\} + \phi_2}, \quad (3.15)$$

where

$$\phi_1 \triangleq e^{j4\pi\varepsilon_o} \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} \mathbb{E}[\Lambda(k_1)\Lambda(k_2)] e^{-j2\pi k_1/K} e^{-j2\pi k_2/K}, \quad (3.16)$$

$$\phi_2 \triangleq \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} \mathbb{E}[\Lambda(k_1)\Lambda^*(k_2)] e^{-j2\pi k_1/K} e^{j2\pi k_2/K}. \quad (3.17)$$

In the above equations,

$$\mathbb{E}[\Lambda(k_1)\Lambda(k_2)] = \text{tr}[\mathbf{B}_{k_1}^T \mathbf{R}_{\varepsilon_o}] \text{tr}[\mathbf{B}_{k_2}^T \mathbf{R}_{\varepsilon_o}] + \text{tr}[\mathbf{B}_{k_1}^T \mathbf{R}_{\varepsilon_o} \mathbf{B}_{k_2}^T \mathbf{R}_{\varepsilon_o}] + c(k_1, k_2), \quad (3.18)$$

$$\mathbb{E}[\Lambda(k_1)\Lambda^*(k_2)] = \text{tr}[\mathbf{B}_{k_1}^T \mathbf{R}_{\varepsilon_o}] \text{tr}[\mathbf{B}_{k_2} \mathbf{R}_{\varepsilon_o}] + \text{tr}[\mathbf{B}_{k_1}^T \mathbf{R}_{\varepsilon_o} \mathbf{B}_{k_2} \mathbf{R}_{\varepsilon_o}] + c(k_1, k_2), \quad (3.19)$$

where  $\text{tr}[\cdot]$  denotes the trace of a matrix,  $\mathbf{R}_{\varepsilon} \triangleq \frac{E_s}{T} \mathbf{P}_{\varepsilon} + \frac{N_o Q}{T} \boldsymbol{\varphi}$ ,

$$[\mathbf{P}_{\varepsilon}]_{ij} \triangleq \sum_{n=-\infty}^{\infty} p^*(iT/Q - nT - \varepsilon T) p(jT/Q - nT - \varepsilon T), \quad (3.20)$$

$$c(k_1, k_2) \triangleq \frac{E_s^2}{T^2} (m_4 - 2) \sum_{n=-\infty}^{\infty} [\mathbf{a}_n(\varepsilon_o)^H \mathbf{B}_{k_1} \mathbf{a}_n(\varepsilon_o)] [\mathbf{a}_n(\varepsilon_o)^H \mathbf{B}_{k_2} \mathbf{a}_n(\varepsilon_o)], \quad (3.21)$$

with  $\mathbf{a}_n(\varepsilon_o)$  defined in (3.11), and  $m_4 = \mathbb{E}[|d_i|^4]$  is the fourth order moment of the transmitted symbols, which is fixed for a specific constellation (e.g.,  $m_4 = 1$  for PSK

and  $m_4 > 1$  for QAM). As the symbol timing delay  $\varepsilon_o$  is assumed to be uniformly distributed in  $[0, 1)$ , the average MSE is calculated by numerical integration of (3.15). Notice that if we put  $g_r(t) = \delta(t)$ , then we obtain the corresponding expressions given in Chapter II.

#### D. Numerical Examples and Discussions

In this section, the general analytical MSE expression presented in the last section will be plotted as a function of  $E_s/N_o$  for different estimators. The analytic results are compared with the corresponding simulation results and the asymptotic CCRB. All the results are generated assuming i.i.d. QPSK data,  $L_o = 100$ , both  $g(t)$  and  $g_r(t)$  are square root raised cosine pulses with  $\alpha = 0.3$ ,  $L_p = 3$ , and  $\varepsilon_o$  is uniformly distributed in the range  $[0, 1)$ . The carrier phase  $\theta_o$  is generated as a uniformly distributed random variable in the range  $[-\pi, \pi)$ , and assumed constant during each estimation. Each simulation point is obtained by averaging  $10^4$  simulation runs. The asymptotic CCRB is computed assuming  $Q = 2$ . In this chapter, the results of the following representative estimators are presented:

1. Modified Lee's estimator [44]. The Lee's estimator is an algorithm obtained from an ad-hoc argument. Its modified version is used since it has a better performance than the original version.
2. Feedforward CML estimator in Chapter II. This is an algorithm derived from a well-known statistical signal processing approach, namely the ML principle.
3. SLN estimator [3]. This estimator belongs to the class of cyclic correlation-based estimator (3.3). It is chosen because it was shown in [42] and [7] that the performance of (3.3) for  $\tau = 0$  appears to be the best.

4. SLN estimator with pre-filter [43]. This estimator is included to demonstrate the efficiency of pre-filter. The pre-filter used is  $h(n) = p(t) \cos(2\pi t/T)|_{t=nT/Q}$  for  $n = -5Q, \dots, 5Q$  (i.e.,  $L_{pre} = 10Q + 1$ ) [43].

Notice that the first two estimators assume an oversampling ratio  $Q = 2$ , while the last two estimators assume an oversampling ratio  $Q = 4$ . The parameters for these four estimators when expressed in the form of the general estimator (3.2) are summarized in Table I.

Table I. Parameters of different feedforward timing estimation algorithms when expressed in the form of the general estimator

	$Q$	$\mathbf{B}_k$	$K$	remark
Modified Lee's estimator [36]	2	eqn. (3.8)	4	–
Feedforward CML estimator	2	eqn. (3.9)	4	$L_p = 3$
SLN estimator [3]	4	eqn. (3.5)	4	–
SLN estimator with pre-filter [43]	4	eqn. (3.13)	4	$h(n) = p(nT/Q) \cos(2\pi n/Q)$ for $n = -5Q, \dots, 5Q$

For the computation of  $\mathbf{B}_k^{CML}$  and  $\text{CCRB}^{as}(\varepsilon_o)$ , there is a need to calculate  $\boldsymbol{\varphi}^{-1}$ . Unfortunately, numerical calculations show that, for the  $g_r(t)$  under consideration,  $\boldsymbol{\varphi}$  is not full rank (at least to the accuracy of Matlab). A main reason for rank deficiency is that, due to the nature of  $g_r(t)$ , when  $|i - j|$  is large, the values of  $[\boldsymbol{\varphi}]_{ij}$  are very very small but not zero. A way to get around this is to replace  $\boldsymbol{\varphi}^{-1}$  by  $\bar{\boldsymbol{\varphi}}^{-1}$ , where  $[\bar{\boldsymbol{\varphi}}]_{ij} = [\boldsymbol{\varphi}]_{ij}$  if  $|i - j| < L_\varphi Q$  and zero otherwise. In this way, the matrix  $\bar{\boldsymbol{\varphi}}$  can be made full rank, but at the same time, significant part of the correlation between noise samples can still be represented accurately. Since most of the correlation induced by  $g_r(t)$  is confined to a duration of a few symbols,  $L_\varphi = 4$  is used for the rest of the chapter. Notice that the matrix  $\boldsymbol{\varphi}$  in  $\mathbf{R}_\varepsilon$  need not to be replaced by  $\bar{\boldsymbol{\varphi}}$  since no inversion is required.

Fig. 11 shows the results for the modified Lee's estimator and the feedforward CML estimator. It can be seen that the analytical and simulation results match

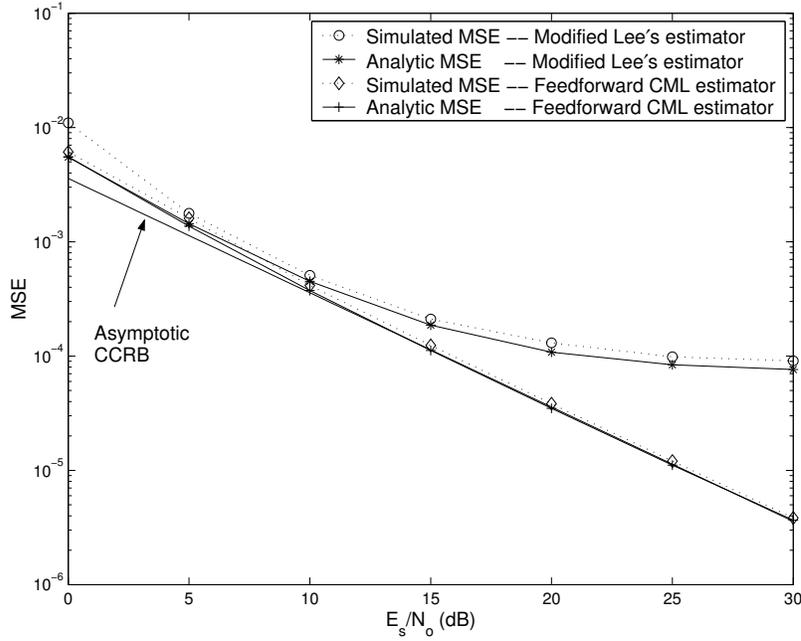


Fig. 11. Analytic and simulated MSEs for modified Lee's estimator and feedforward CML estimator.

very well. Furthermore, the feedforward CML estimator performs much better than the modified Lee's estimator at high  $E_s/N_o$  and its performance coincides with the asymptotic CCRB, meaning that the feedforward CML estimator is the best (in terms of MSE performance) within the class of symbol timing estimators employing second-order statistics. Fig. 12 shows the results for the SLN estimator with and without pre-filter. This figure also shows that the simulation results match the analytical results very well. Moreover, the figure shows that the application of a properly designed pre-filter removes the estimation error floor at high  $E_s/N_o$  and makes the performance of the resultant estimator reaches the asymptotic CCRB.

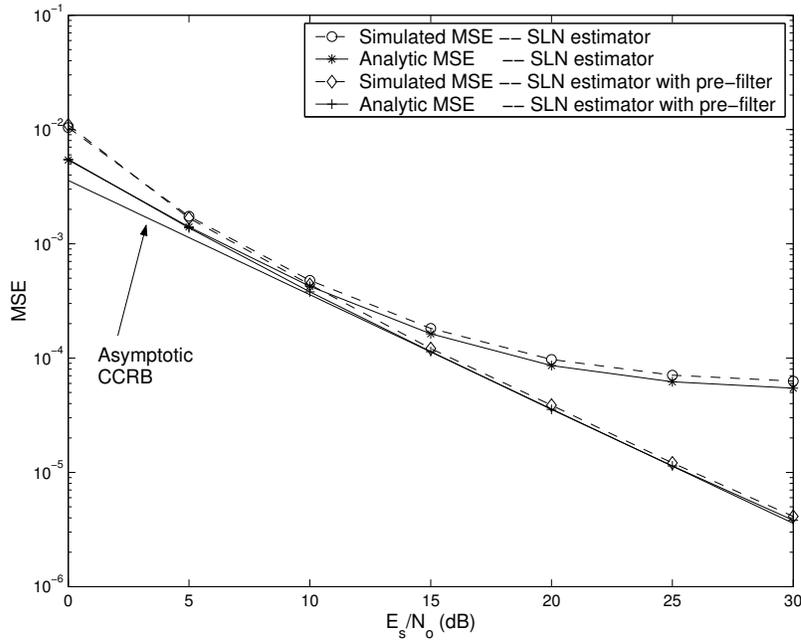


Fig. 12. Analytic and simulated MSEs for SLN estimator with and without pre-filter.

## E. Conclusions

In this Chapter, all the previously proposed feedforward symbol timing estimators employing second-order statistics were formulated into a unified framework. The finite sample mean square error (MSE) expression and the asymptotic conditional Cramer-Rao bound (CCRB) for this class of estimators were established. It was found that the analytical and simulation results match very well. Furthermore, it was found that the feedforward CML estimator and the SLN estimator with a properly designed pre-filter [43] perform the best and their performances coincide with the asymptotic CCRB, which is the performance lower bound for the class of estimators under consideration.

## CHAPTER IV

SYMBOL TIMING ESTIMATION IN MIMO FLAT-FADING CHANNELS – A  
HEURISTIC APPROACH

## A. Introduction

Communication over multiple-input-multiple-output (MIMO) channels has attracted much attention recently [45]-[55] due to the huge capacity gain over single antenna systems. While many different techniques and algorithms have been proposed to explore the potential capacity, synchronization in MIMO channels received comparatively less attention.

Symbol timing synchronization in MIMO uncorrelated flat-fading channels was first studied by Naguib *et al.* [11], where orthogonal training sequences are transmitted at different transmit antennas to simplify the maximization of the oversampled approximated log-likelihood function. The sample having the largest magnitude, so called the “optimal sample”, is assumed to be closest to the optimum sampling instants (it will be referred as the optimum sample selection algorithm in the sequel for convenience). However, it is shown in this chapter that the mean square error (MSE) of this algorithm is lower bounded by  $1/(12Q^2)$ , where  $Q$  is the oversampling ratio. As a result, the performance of this timing synchronization method highly depends on the oversampling ratio. In fact, relatively high oversampling ratios might be required for accurate symbol-timing estimation.

In this chapter, a new symbol-timing estimator for MIMO flat-fading channels is proposed. It improves the optimum sample selection algorithm in [11] so that accurate timing estimates can be obtained even if the oversampling ratio is small. The increase in implementation complexity with respect to that of optimum sample

selection algorithm is very small. The requirements and the design procedures for the training sequences are discussed. Analytical expression for MSE of the proposed estimator is also derived. Both analytical and simulation results show that, for modest oversampling ratio (such as  $Q=4$ ), the MSE of the proposed estimator is significantly smaller than that of the optimum sample selection algorithm. Furthermore, the effects of the number of transmit and receive antennas, the oversampling ratio, and the length of training sequence on the MSE are also examined.

This chapter is organized as follows. The system model of a MIMO communication system is first described in Section B. A brief overview of the optimum sample selection algorithm for symbol-timing synchronization in MIMO communication systems is given in Section C. Requirements and design of training sequences are discussed in Section D. The proposed symbol-timing estimator is then presented in Section E. Analytical MSE expressions are derived in Section F. Simulation results and discussions are then presented in Section G, and finally conclusions are drawn in Section H.

## B. Signal Model

Consider a MIMO communication system with  $N$  transmit and  $M$  receive antennas. At each receiving antenna, a superposition of faded signals from all the transmit antennas plus noise is received. It is assumed that the channel is frequency flat and quasi-static. The complex envelope of the received signal at the  $j^{\text{th}}$  receive antenna can be written as

$$r_j(t) = \sqrt{\frac{E_s}{NT}} \sum_{i=1}^N h_{ij} \sum_n d_i(n) g(t - nT - \varepsilon_o T) + \eta_j(t), \quad j = 1, 2, \dots, M \quad (4.1)$$

where  $E_s/N$  is the symbol energy;  $h_{ij}$  is the complex channel coefficient between the  $i^{\text{th}}$  transmit antenna and the  $j^{\text{th}}$  receive antenna;  $d_i(n)$  is the zero-mean complex valued symbol transmitted from the  $i^{\text{th}}$  transmit antenna;  $g(t)$  is the transmit filter with unit energy;  $T$  is the symbol duration;  $\varepsilon_o$  is the unknown timing offset, which is assumed to be uniformly distributed in the range  $[-0.5, 0.5)$ ; and  $\eta_j(t)$  is the complex-valued circularly distributed Gaussian white noise at the  $j^{\text{th}}$  receive antenna, with power density  $N_o$ . Notice that the timing offsets between all pairs of transmit and receive antennas are assumed to be the same. This assumption holds when both the transmit and receive antenna array sizes are small.

The received signal is filtered by the matched filter  $g_r(t)$  and then sampled at a rate  $Q$  times faster than the symbol rate  $1/T$ . The sampled signal at the  $j^{\text{th}}$  receive antenna is given by

$$r_j(m) = \sqrt{\frac{E_s}{NT}} \sum_{i=1}^N h_{ij} \sum_n d_i(n) p(mT/Q - nT - \varepsilon_o T) + \eta_j(m), \quad (4.2)$$

where  $r_j(m) \triangleq r_j(mT/Q)$ ,  $p(t) \triangleq g(t) \star g_r(t)$ ,  $\eta_j(m) \triangleq n_j(t) \star g_r(t)|_{t=mT/Q}$ . Throughout this chapter, it is assumed that both  $g(t)$  and  $g_r(t)$  are square root raised cosine filter with roll-off factor  $\alpha$ , and the channel coefficient  $h_{ij}$ 's are independent of each other.

### C. Timing Synchronization by Optimum Samples Selection

As proposed in [11], orthogonal training sequences can be periodically transmitted in between data symbols (as shown in Fig. 13 for a two transmit antennas system) to assist the timing synchronization. The idea is that at the receiver, if the position of the orthogonal training sequences can be correctly located, signal from any one of the transmit antennas can be extracted (and signals from other antennas are removed) by

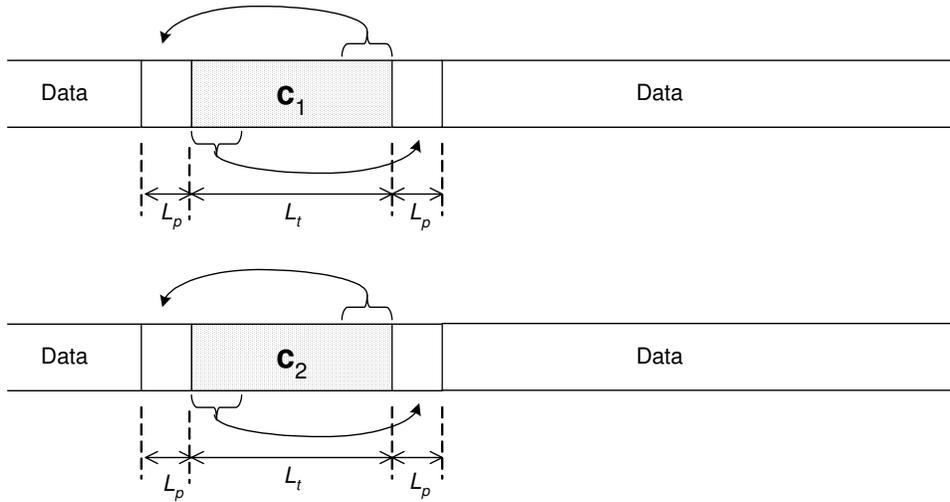


Fig. 13. Structure of the training sequence for symbol timing synchronization in a two transmit antenna system.

multiplying the received signal with the orthogonal sequence transmitted from that antenna. Note that the structure of training sequences in this chapter is different from that presented in [11]. In this chapter, a cyclic prefix and cyclic suffix, each of length  $L_p$ , are included in order to remove the inter-symbol interference (ISI) from the random data transmitted before and after the orthogonal training sequences. Since  $L_p$  is usually kept as a small number, the increase in length of training is very small, especially when the length of the orthogonal training sequences is large.

Let  $\mathbf{c}_i = [c_i(0) \ c_i(1) \dots c_i(L_t - 1)]$  be the  $i^{th}$  ( $i = 1, \dots, N$ ) orthogonal training sequence of length  $L_t$  to be transmitted from the  $i^{th}$  transmit antenna. The sampled signal at the  $j^{th}$  receive antenna can be obtained by replacing  $d_i(n)$  in (4.2) with  $c_i(n)$ . Further, let  $m = lQ + k$  ( $l = 0, 1, \dots, L_t - 1$  and  $k = k_o, k_o + 1, \dots, k_o + Q - 1$ , where  $k_o = -\lfloor(1/2 - \varepsilon_o)Q\rfloor$  and  $\lfloor x \rfloor$  denotes the nearest integer less than or equal to  $x$ ), so that each sample is indexed by the  $l^{th}$  training bit and the  $k^{th}$  phase. In order to maintain the orthogonality between the received training sequences and the local copies, the first phase is taken at  $-\lfloor(1/2 - \varepsilon_o)Q\rfloor$  such that all the  $Q$  samples for

the  $l^{\text{th}}$  training bit are taken from  $-T/2 \leq t - lT \leq T/2$ . Then the received signal  $r_j(lQ + k)$  due to the orthogonal training sequences can be rewritten as

$$r_j(lQ + k) = \sqrt{\frac{E_s}{NT}} \sum_{i=1}^N h_{ij} \sum_n c_i(n) p(kT/Q + (l - n)T - \epsilon_o T) + \eta_j(lQ + k) \quad (4.3)$$

for  $l = 0, 1, \dots, L_t - 1$ , and  $k = 0, 1, \dots, Q - 1$ ,

where  $\epsilon_o \triangleq \epsilon_o + k_o/Q$ . Note that  $k_o$  has been dropped from the index of  $\eta_j(lQ + k)$  since a fixed time shift does not affect the noise statistics. In practice, it is sufficient to estimate  $\epsilon_o$  only as it represents the time difference between the first sample of the training sequence and the next nearest optimum sampling instance. Grouping the samples with the same phase, one can form the vector  $\mathbf{r}_j(k)$  as follows:

$$\mathbf{r}_j(k) \triangleq [r_j(k) \ r_j(Q + k) \ r_j(2Q + k) \ \dots \ r_j((L_t - 1)Q + k)]^T \quad (4.4)$$

$$= \sqrt{\frac{E_s}{NT}} \sum_{i=1}^N h_{ij} \mathbf{C}_i \mathbf{p}(k) + \boldsymbol{\eta}_j(k) \quad (4.5)$$

where

$$\mathbf{C}_i \triangleq \begin{bmatrix} c_i(\text{mod}(-L_p, L_t)) & c_i(\text{mod}(-L_p + 1, L_t)) & \dots & c_i(\text{mod}(L_p, L_t)) \\ c_i(\text{mod}(-L_p + 1, L_t)) & c_i(\text{mod}(-L_p + 2, L_t)) & \dots & c_i(\text{mod}(L_p + 1, L_t)) \\ \vdots & \vdots & & \vdots \\ c_i(\text{mod}(-L_p + L_t - 1, L_t)) & c_i(\text{mod}(-L_p + L_t, L_t)) & \dots & c_i(\text{mod}(L_p + L_t - 1, L_t)) \end{bmatrix} \quad (4.6)$$

$$\mathbf{p}(k) \triangleq [p(kT/Q - L_p T - \epsilon_o T) \ p(kT/Q - (L_p - 1)T - \epsilon_o T) \ \dots \ p(kT/Q + L_p T - \epsilon_o T)]^T \quad (4.7)$$

$$\boldsymbol{\eta}_j(k) \triangleq [\eta_j(k) \ \eta_j(Q + k) \ \dots \ \eta_j((L_t - 1)Q + k)]^T. \quad (4.8)$$

Define the sequence  $\lambda_{ij}(k) \triangleq \mathbf{c}_i^H \mathbf{r}_j(k)$ , where  $\mathbf{c}_i^H$  denotes the transpose conjugate

of  $\mathbf{c}_i$ . Since  $\mathbf{c}_i$ 's are orthogonal to each other when the relative delay is zero, it follows that

$$\lambda_{ij}(k) = \sqrt{\frac{E_s}{NT}} h_{ij} p\left(\frac{kT}{Q} - \epsilon_o T\right) \|\mathbf{c}_i\|^2 + \sqrt{\frac{E_s}{NT}} \sum_{i'=1}^N h_{i'j} \mathbf{c}_i^H \tilde{\mathbf{C}}_{i'} \tilde{\mathbf{p}}(k) + \mathbf{c}_i^H \boldsymbol{\eta}_j(k), \quad (4.9)$$

for  $k = 0, \dots, Q - 1$ ,  $\|\mathbf{c}_i\| \triangleq \sqrt{\mathbf{c}_i^H \mathbf{c}_i}$  is the norm of  $\mathbf{c}_i$ , which is a constant;  $\tilde{\mathbf{C}}_i$  is the same as  $\mathbf{C}_i$  but with the  $(L_p + 1)^{th}$  column removed and  $\tilde{\mathbf{p}}(k)$  is the same as  $\mathbf{p}(k)$  but with the  $(L_p + 1)^{th}$  entry removed. The second term in (4.9) represents the ISI if the training sequences are not orthogonal when the relative delay is not zero. The last term in (4.9) is the noise term.

From (4.9), it can be observed that, if the second and third terms are very small (a training sequence design procedure that make the second term zero is discussed in the next section; the third term is small at high Signal-to-Noise Ratios (SNRs)),  $\lambda_{ij}(k)$  has the same shape as  $p(t)$  for  $-T/2 \leq t \leq T/2$ , except that it is scaled by a complex channel gain and is corrupted by additive noise. In order to remove the effect of the channel, consider the sequence  $\Lambda_{ij}(k) \triangleq |\lambda_{ij}(k)|^2$ . Now, the sequence  $\Lambda_{ij}(k)$  should have a similar shape to the function  $|p(t)|^2$  for  $-T/2 \leq t \leq T/2$ . This is illustrated in Fig. 14, where an example sequence of  $\Lambda_{ij}(k)$  is shown ( $Q=8$ ,  $L_t=32$ ,  $L_p=3$ , and in the absence of noise). Note that a scaled version of  $|p(t)|^2$  for  $-T/2 \leq t \leq T/2$  is also shown (in dotted line) for comparison. It can be seen that the optimum sampling time is at  $t=0$  and the sample with maximum amplitude is the one closest to the optimum sampling instant than the remaining samples.

A simple symbol-timing synchronization algorithm is to choose a value of  $k$  closest to the optimum sampling instants. That is, the optimum sampling phase  $k = \hat{k}$  is selected such that it maximizes  $\Lambda_{ij}(k)$ . Since the channels for different antennas are independent, the average of  $\Lambda_{ij}(k)$  over all  $i$  and  $j$  is maximized (see (4.11), where

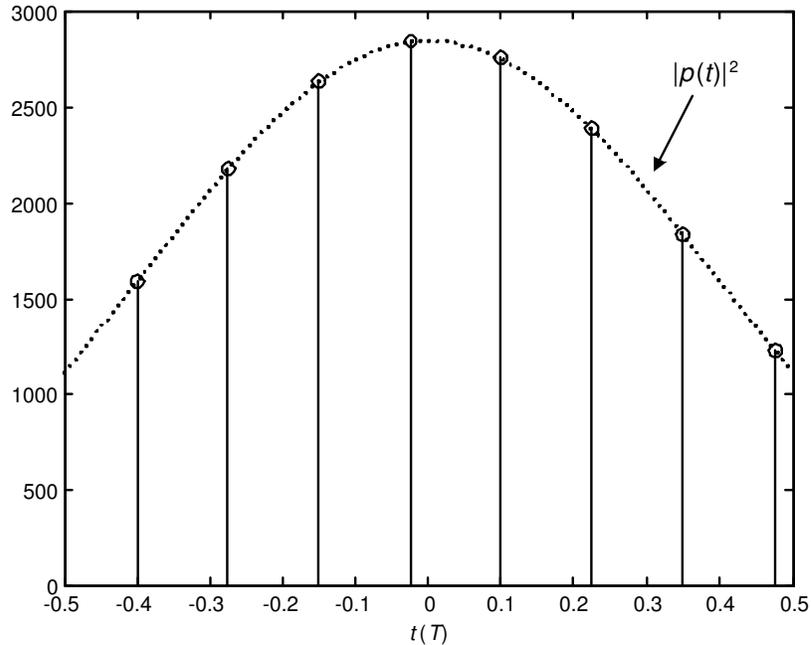


Fig. 14. An example of  $\Lambda_{ij}(k)$  with the scaled version of  $|p(t)|^2$  for  $-T/2 \leq t \leq T/2$  (dotted line).

the scaling factor  $1/MN$  is not included in order to preserve a simplified notation). As mentioned in [11], this represents the samples of an approximated log likelihood function for symbol-timing synchronization, when the ISI plus noise term in (4.9) is assumed to be Gaussian. Therefore, the optimum sampling phase is selected as [11]

$$\hat{k} = \max_{k=0,1,\dots,Q-1} \Lambda_{ML}(k) \quad (4.10)$$

with

$$\Lambda_{ML}(k) = \sum_{j=1}^M \sum_{i=1}^N \Lambda_{ij}(k). \quad (4.11)$$

Under the optimistic assumption that the samples closest to the optimum sampling positions are correctly estimated (at high SNRs), the estimation error, normalized with respect to the symbol duration, is a uniformly distributed random variable in the range  $[-1/2Q, 1/2Q]$ . Therefore, the MSE is  $1/(12Q^2)$ . Thus, a relatively high

oversampling ratio might be required in order to obtain a small MSE.

#### D. Design of Training Sequences

The performance of the proposed timing estimator is directly influenced by the presence of ISI and noise terms in (4.9). In order to minimize the contribution of the ISI term in (4.9), the training sequences need to be designed such that

$$\mathbf{c}_i^H \tilde{\mathbf{C}}_{i'} = 0, \quad (4.12)$$

for all combinations of  $i$  and  $i'$ . Combining with the fact that sequences from different antennas have to be orthogonal when the relative delay is zero, the problem of training sequences design resumes to finding  $N$  sequences such that

$$\mathbf{C}_i^H \mathbf{C}_{i'} = \begin{cases} \|\mathbf{c}_i\|^2 \mathbf{I}_{2L_p+1}, & \text{if } i = i'; \\ 0, & \text{if } i \neq i'. \end{cases} \quad (4.13)$$

This is exactly the problem of designing multiple  $(2L_p + 1)$ -perfect sequences [56]-[58], with each of length  $L_t$ . Here, we just mention the procedures for designing the training sequences, interested readers can refer to the original papers [56]-[58] for details.

1. Construct a sequence  $\mathbf{s} = [s(0) \ s(1) \ \dots \ s(L_t - 1)]$  with length  $L_t$  such that all of its out-of-phase periodic auto-correction terms are equal to zero. One example of this kind of sequence is Chu sequence [59].
2. Construct another sequence  $\mathbf{s}' = [s'(0) \ s'(1) \ \dots \ s'(L_t + 2NL_p - 1)]$  of length  $L_t + 2NL_p$  as follows

$$\mathbf{s}' = \underbrace{[s(0) \ s(1) \ \dots \ s(L_t - 1)]}_{\mathbf{s}} [s(0) \ s(1) \ \dots \ s(2NL_p - 1)]. \quad (4.14)$$

Note that  $L_t \geq 2NL_p$  must be satisfied. That is, if the number of transmit antenna  $N$  is large, we cannot use training sequences with short length.

3. The orthogonal training sequences are given by

$$\mathbf{c}_i = [s'((2i-1)L_p) \dots s'((2i-1)L_p + L_t - 1)]. \quad (4.15)$$

For example, let consider  $L_t=32$ ,  $L_p=3$ ,  $N=2$ . First we construct a Chu sequence of length 32. Then cyclically extend the Chu sequence by copying the first  $2 \times 2 \times 3 = 12$  bits and putting them at the back. Then,  $\mathbf{c}_1 = [s'(3) s'(4) \dots s'(34)]$  and  $\mathbf{c}_2 = [s'(9) s'(10) \dots s'(40)]$ .

#### E. Timing Synchronization by Estimation

In optimum samples selection algorithm, symbol timing is estimated by maximization of the oversampled approximated log-likelihood function. As the number of samples becomes very large (which requires a large oversampling ratio), the estimate could become accurate. However, noting that the approximated log likelihood function is ‘smooth’ (see Fig. 14), we expect that the maximization of the log-likelihood function can be done by interpolation based on a few samples, thus keeping the oversampling ratio at a small number.

More precisely, let us construct a periodic sequence  $\tilde{\Lambda}_{ML}(m)$  by periodically extending the approximated log likelihood sequence  $\Lambda_{ML}(k)$  in (4.11). Further, denote  $\tilde{\Lambda}_{ML}(\epsilon)$  as the continuous and periodic approximated log likelihood function with its samples given by  $\tilde{\Lambda}_{ML}(m)$ . According to sampling theorem, as long as the sampling frequency  $Q/T$  is higher than twice the highest frequency of  $\tilde{\Lambda}_{ML}(\epsilon)$ , then  $\tilde{\Lambda}_{ML}(\epsilon)$  can be represented by its samples  $\tilde{\Lambda}_{ML}(m)$  without loss of information. The relationship

between  $\tilde{\Lambda}_{ML}(\epsilon)$  and  $\tilde{\Lambda}_{ML}(m)$  is then given by

$$\tilde{\Lambda}_{ML}(\epsilon) = \sum_{m=-\infty}^{\infty} \tilde{\Lambda}_{ML}(m) \text{sinc} \left( \pi \frac{\epsilon T - mT/Q}{T/Q} \right). \quad (4.16)$$

Now, expand  $\tilde{\Lambda}_{ML}(\epsilon)$  into a Fourier series

$$\tilde{\Lambda}_{ML}(\epsilon) = \sum_{\ell=-\infty}^{\infty} A_{\ell} e^{j2\pi\ell\epsilon} \quad (4.17)$$

where

$$A_{\ell} = \int_0^1 \tilde{\Lambda}_{ML}(\epsilon) e^{-j2\pi\ell\epsilon} d\epsilon. \quad (4.18)$$

Substituting (4.16) into (4.18) yields

$$\begin{aligned} A_{\ell} &= \sum_{m=-\infty}^{\infty} \tilde{\Lambda}_{ML}(m) \int_0^1 \text{sinc} \left( \pi \frac{\epsilon T - mT/Q}{T/Q} \right) e^{-j2\pi\ell\epsilon} d\epsilon \\ &= \sum_{k=0}^{Q-1} \Lambda_{ML}(k) \sum_{l=-\infty}^{\infty} \int_0^1 \text{sinc} \left( \pi \frac{\epsilon T - lT - kT/Q}{T/Q} \right) e^{-j2\pi\ell\epsilon} d\epsilon \\ &= \sum_{k=0}^{Q-1} \Lambda_{ML}(k) e^{-j2\pi\ell k/Q} \frac{1}{Q} \mathfrak{F}\{\text{sinc}(\pi\epsilon)\}_{f=\ell/Q}, \end{aligned} \quad (4.19)$$

where  $\mathfrak{F}\{\}$  denotes the Fourier transform. It is clear that if  $Q$  is even,

$$A_{\ell} = \begin{cases} \frac{1}{Q} \sum_{k=0}^{Q-1} \Lambda_{ML}(k) e^{-j2\pi\ell k/Q} & \text{for } \ell = -Q/2 + 1, \dots, Q/2 - 1 \\ \frac{1}{2Q} \sum_{k=0}^{Q-1} \Lambda_{ML}(k) e^{-j2\pi\ell k/Q} & \text{for } \ell = -Q/2, Q/2 \\ 0 & \text{otherwise} \end{cases}, \quad (4.20)$$

and if  $Q$  is odd,

$$A_{\ell} = \begin{cases} \frac{1}{Q} \sum_{k=0}^{Q-1} \Lambda_{ML}(k) e^{-j2\pi\ell k/Q} & \text{for } \ell = -\lfloor Q/2 \rfloor, \dots, \lfloor Q/2 \rfloor \\ 0 & \text{otherwise} \end{cases}. \quad (4.21)$$

From (4.17), it can be seen that once the coefficients  $A_{\ell}$  are determined, the timing delay  $\epsilon_o$  can be estimated by maximizing  $\tilde{\Lambda}_{ML}(\epsilon)$  for  $0 \leq \epsilon \leq 1$ . Note that

$\Lambda_{ML}(k)$  only contains samples of the approximated log-likelihood function at certain delays, while  $\tilde{\Lambda}_{ML}(\epsilon)$  is a continuous function of  $\epsilon$ . Therefore, maximizing  $\tilde{\Lambda}_{ML}(\epsilon)$  provides a more accurate estimate of the timing delay than maximizing  $\Lambda_{ML}(k)$ . For efficient implementation, the maximization can be performed by Discrete Fourier Transform (DFT)-based interpolation. More specifically,  $\tilde{\Lambda}_{ML}(\epsilon)$  for  $0 \leq \epsilon \leq 1$  can be approximated by an  $K'$ -point sequence, denoted as  $\Lambda_{ML}(k')$  for  $0 \leq k' \leq K' - 1$ , by zero padding the high frequencies coefficients of  $A_\ell$  and performing a  $K'$ -point inverse Discrete Fourier Transform (IDFT). For sufficiently large value of  $K'$ ,  $\Lambda_{ML}(k')$  becomes very close to  $\tilde{\Lambda}_{ML}(\epsilon)$  for  $0 \leq \epsilon \leq 1$ , and the index with the maximum amplitude can be viewed as an improved estimate of the timing parameter  $\epsilon_o$ .

To avoid the complexity in performing the  $K'$ -point IDFT, an approximation is applied to (4.17). More precisely, extensive simulations show that  $A_{\pm 1}$  are much greater than  $A_\ell$  for  $|\ell| > 1$ , therefore,

$$\tilde{\Lambda}_{ML}(\epsilon) \approx A_0 + 2\Re\{A_1 e^{j2\pi\epsilon}\} \quad \text{for } 0 \leq \epsilon \leq 1. \quad (4.22)$$

In order to maximize the approximated log likelihood function  $\tilde{\Lambda}_{ML}(\epsilon)$ , we have

$$\arg(A_1) = -2\pi\epsilon, \quad (4.23)$$

where  $\arg(x)$  denotes the phase of  $x$ . Or equivalently,

$$\hat{\epsilon} = -\frac{1}{2\pi} \arg\left\{\sum_{k=0}^{Q-1} \Lambda_{ML}(k) e^{-j2\pi k/Q}\right\}. \quad (4.24)$$

The estimated delay  $\hat{\epsilon}$  is the time between the first sampling phase and the nearest optimum sampling instant. The calculation within the arg-operation is actually the  $2^{nd}$  output of a  $Q$ -point DFT of the sequence (or the Fourier coefficient at symbol rate  $f = 1/T$ ). Note that the increase in complexity of the proposed algorithm in (4.24) with respect to that of optimum samples selection algorithm is only a  $Q$ -

point DFT (which can be efficiently implemented using Goertzel's algorithm) and an arg-operation. From the simulation results to be presented at Section G, it is found that an oversampling factor  $Q$  of 4 is sufficient to yield good estimates in practical applications. Therefore, the 4-point DFT in (4.24) can be computed easily without any multiplications since  $\exp(-j2\pi k/4) \in \{\pm 1, \pm j\}$ . This greatly reduces the arithmetic complexity of implementation.

#### F. Performance Analysis

The MSE expression of the proposed estimator as a function of  $E_s/N_o$  is derived in this section. First, express the true delay as

$$\epsilon_o = -\frac{1}{2\pi} \arg(e^{-j2\pi\epsilon_o}). \quad (4.25)$$

Taking the difference between (4.24) and (4.25), the MSE is given by

$$\mathbb{E}[(\hat{\epsilon} - \epsilon_o)^2] = \left(\frac{1}{2\pi}\right)^2 \mathbb{E}\left[\left(\arctan\left\{\frac{\Im\phi}{\Re\phi}\right\}\right)^2\right] \quad (4.26)$$

where

$$\phi \triangleq e^{j2\pi\epsilon_o} \sum_{k=0}^{Q-1} \Lambda_{ML}(k) e^{-j2\pi k/Q}. \quad (4.27)$$

Applying the approximation  $\arctan(x) \approx x$  for small  $x$ , we have

$$\begin{aligned} \mathbb{E}[(\hat{\epsilon} - \epsilon_o)^2] &\approx \left(\frac{1}{2\pi}\right)^2 \mathbb{E}\left[\left(\frac{\phi - \phi^*}{j(\phi + \phi^*)}\right)^2\right] \\ &\approx -\left(\frac{1}{2\pi}\right)^2 \frac{\mathbb{E}[\phi^2] - 2\mathbb{E}[\phi\phi^*] + \mathbb{E}[(\phi^*)^2]}{\mathbb{E}[\phi^2] + 2\mathbb{E}[\phi\phi^*] + \mathbb{E}[(\phi^*)^2]} \\ &= -\left(\frac{1}{2\pi}\right)^2 \frac{\Re\{\mathbb{E}[\phi^2]\} - \mathbb{E}[\phi\phi^*]}{\Re\{\mathbb{E}[\phi^2]\} + \mathbb{E}[\phi\phi^*]}. \end{aligned} \quad (4.28)$$

The second approximation is justified by the fact that the mean of the denominator  $\mathbb{E}[(\phi + \phi^*)^2]$  is much larger than the mean of the numerator  $\mathbb{E}[(\phi - \phi^*)^2]$  (which is

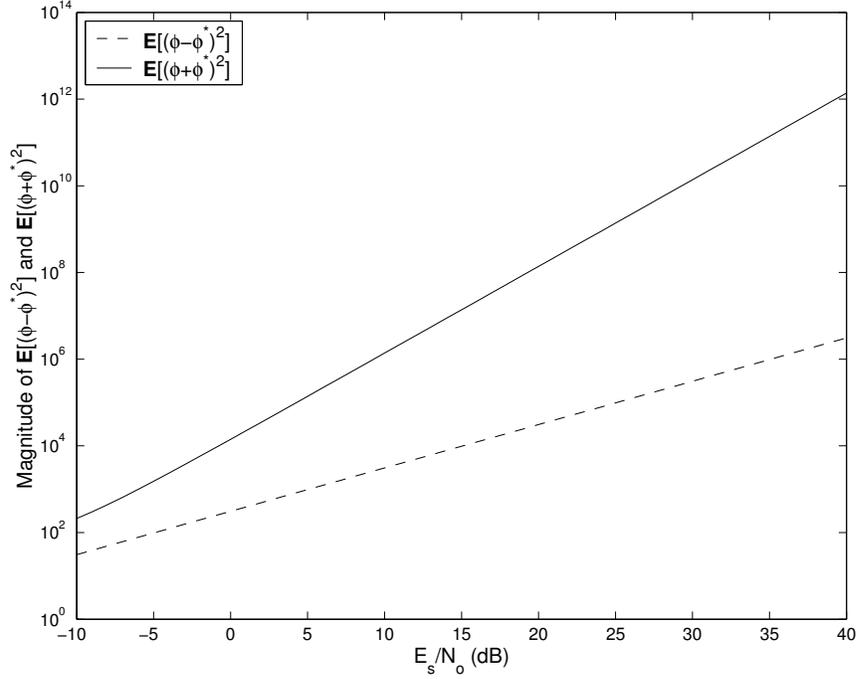


Fig. 15. Magnitude of  $\mathbb{E}[(\phi - \phi^*)^2]$  and  $\mathbb{E}[(\phi + \phi^*)^2]$  as a function of  $E_s/N_o$  for  $\varepsilon_o = -0.5, -0.25, 0, 0.25$  and  $0.5$  ( $N = 2, M = 4, \alpha=0.3, Q=4, L_t=32, L_p=4$ ). Note that all curves for different values of  $\varepsilon_o$  overlap.

illustrated in Fig. 15 for  $\varepsilon_o = -0.5, -0.25, 0, 0.25$  and  $0.5$  with  $N = 2, M = 4, \alpha=0.3, Q=4, L_t=32$  and  $L_p=4$ ) and the variance of the numerator and denominator are much smaller than the mean of the denominator (which is true for medium to high SNRs). Some additional explanations regarding this approximation can be found in [16]. The last equality of (4.28) is due to the that fact  $\mathbb{E}[(\phi^*)^2] = (\mathbb{E}[\phi^2])^*$ .

It is proved in Appendix B that

$$\mathbb{E}[\phi^2] = E_s^2 L_t^2 M N e^{j4\pi\varepsilon_o} \cdot \left[ L_t^2 \frac{(1 + MN)}{N^2 T^2} \Xi_{SS}^2 + \left( \frac{E_s}{N_o} \right)^{-1} \frac{2L_t}{NT} \Xi_{SN} + \left( \frac{E_s}{N_o} \right)^{-2} \Xi_{NN} \right], \quad (4.29)$$

$$\mathbb{E}[\phi\phi^*] = E_s^2 L_t^2 M N \cdot \left[ L_t^2 \frac{(1 + MN)}{N^2 T^2} |\Xi_{SS}|^2 + \left( \frac{E_s}{N_o} \right)^{-1} \frac{2L_t}{NT} \tilde{\Xi}_{SN} + \left( \frac{E_s}{N_o} \right)^{-2} \tilde{\Xi}_{NN} \right] \quad (4.30)$$

where

$$\Xi_{SS} \triangleq \sum_{k=0}^{Q-1} p^2(kT/Q - \epsilon_o T) e^{-j2\pi k/Q} \quad (4.31)$$

$$\Xi_{SN} \triangleq \sum_{k'=0}^{Q-1} \sum_{k''=0}^{Q-1} p(k'T/Q - \epsilon_o T) p(k''T/Q - \epsilon_o T) \cdot \varphi((k' - k'')T/Q) e^{-j2\pi k''/Q} e^{-j2\pi k'/Q} \quad (4.32)$$

$$\Xi_{NN} \triangleq \sum_{k'=0}^{Q-1} \sum_{k''=0}^{Q-1} \varphi^2((k' - k'')T/Q) e^{-j2\pi k''/Q} e^{-j2\pi k'/Q} \quad (4.33)$$

$$\tilde{\Xi}_{SN} \triangleq \sum_{k'=0}^{Q-1} \sum_{k''=0}^{Q-1} p(k'T/Q - \epsilon_o T) p(k''T/Q - \epsilon_o T) \cdot \varphi((k' - k'')T/Q) e^{j2\pi k''/Q} e^{-j2\pi k'/Q} \quad (4.34)$$

$$\tilde{\Xi}_{NN} \triangleq \sum_{k'=0}^{Q-1} \sum_{k''=0}^{Q-1} \varphi^2((k' - k'')T/Q) e^{j2\pi k''/Q} e^{-j2\pi k'/Q} \quad (4.35)$$

and

$$\varphi(\tau) \triangleq \int_{-\infty}^{\infty} g_r(t) g_r^*(t + \tau) dt \quad (4.36)$$

is the correlation between noise samples introduced by the matched filter.

Since the timing delay is assumed to be uniformly distributed, the average MSE can be calculated by numerical integration of (4.28).

## G. Simulation Results and Discussions

The performances of the synchronizers based on the optimum sample selection (4.10) and the proposed algorithm (4.24) are evaluated in this section. The MSE of the estimates are calculated using both the analytic expressions derived in the last section and Monte Carlo simulations, where each MSE value is obtained by averaging over  $10^5$  estimates. The timing offset  $\epsilon_o$  is generated to be uniformly distributed in the interval  $[-0.5, 0.5)$ . The channel coefficients  $h_{ij}$  are generated as complex Gaussian random variables with zero mean and a variance of 0.5 per dimension. The pulse

shape  $p(t)$  is a raised cosine pulse with excess bandwidth  $\alpha = 0.3$ . The training sequences are generated following the procedures in Section D with  $L_p = 4$ . Without loss of generality, we assume  $T = 1$ . In all the figures, MSE of both the proposed algorithm and the optimum sample selection algorithm are plotted against  $E_s/N_o$ , with the markers showing the simulation results, while the solid lines represent the theoretical MSE derived in the last section.

### 1. Effect of Oversampling Ratio

In (4.16), it is assumed that the sampling frequency is at least twice the highest frequency of  $\tilde{\Lambda}_{ML}(\epsilon)$ . Since  $\tilde{\Lambda}_{ML}(\epsilon)$  has the same shape as  $|p(t)|^2$  for  $-T/2 \leq t \leq T/2$ , where  $p(t)$  is a raised cosine pulse, it is natural to predict that the sampling frequency  $Q/T$  has to be greater than  $2 \times 2/T$  (i.e.,  $Q \geq 4$ ). This prediction is corroborated by Fig. 16, where the MSE are shown for  $Q=2, 4, 8$  and  $16$  in a two-transmit, four-receive antenna system with  $L_t=32$ . Several conclusions can be drawn from the figure:

1. Performances of the optimum sample selection algorithm are lower bounded by  $1/(12Q^2)$ , and are poorer than that of the proposed algorithm, for all value of  $Q$  (except  $Q=2$ , in which case the performances of both algorithms are the same).
2. It can be seen that for  $Q=2$ , the MSE of the proposed algorithm is much higher than that corresponding to other oversampling ratios. This confirms the above argument that  $Q$  has to be greater than 4 in order to represent  $\tilde{\Lambda}_{ML}(\epsilon)$  without much loss of information. For  $Q=8$  and  $Q=16$ , the performance improves at high  $E_s/N_o$ . This can be explained by the fact that  $\tilde{\Lambda}_{ML}(\epsilon)$  is a truncated version of  $|p(t)|^2$ , so  $\tilde{\Lambda}_{ML}(\epsilon)$  is no longer bandlimited. Therefore,  $\tilde{\Lambda}_{ML}(m)$  would, in general, suffer from aliasing from the neighboring spectra. Increasing  $Q$  thus reduces the aliasing and improves the performance.

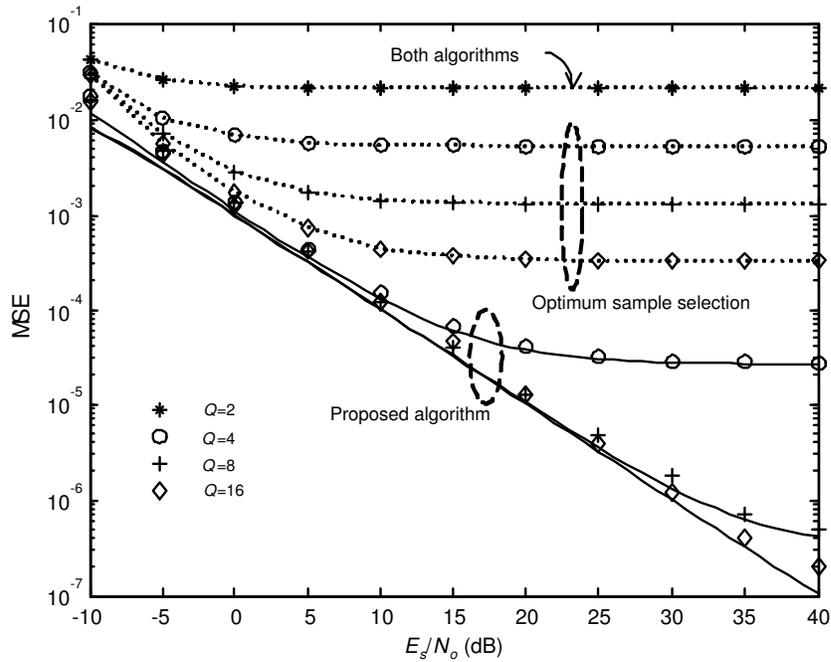


Fig. 16. MSE performance for different oversampling ratios  $Q$  ( $N=2$ ,  $M=4$ ,  $L_t=32$ ,  $\alpha=0.3$ ).

3. The analytical MSEs (solid lines in the figure) match very well with the simulation results for  $Q=4$ , 8 and 16. Note that for  $Q=2$ , the analytic MSE expression does not hold and only the simulation results have been plotted in Fig. 16.
4. Strictly speaking,  $Q$  should be at least equal to 16 in order to represent  $\tilde{\Lambda}_{ML}(\epsilon)$  using its samples  $\tilde{\Lambda}_{ML}(m)$  without loss of information. However, for  $Q=4$ , the MSE of the proposed algorithm reaches the order of  $10^{-5}$  at medium and high  $E_s/N_o$ , which is a reasonably good performance in practical applications. Because of this reason,  $Q = 4$  is used to generate the simulation results for the rest of this chapter.

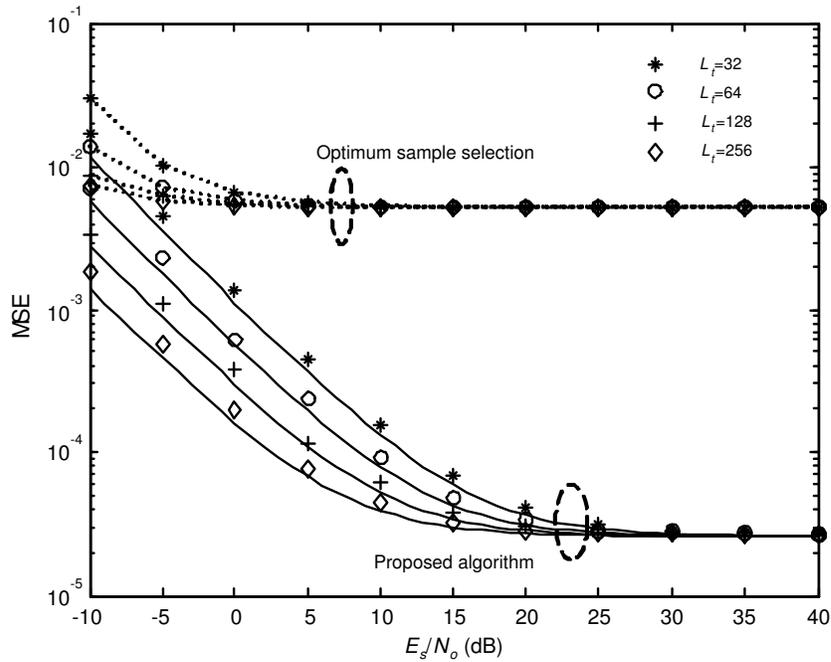


Fig. 17. MSE performance for different lengths of the training sequence ( $N=2$ ,  $M=4$ ,  $Q=4$ ,  $\alpha=0.3$ ).

## 2. Effect of Length of Training Sequences

Fig. 17 shows the MSE of a two-transmit, four-receive antenna system with different lengths ( $L_t$ ) of the training sequences. In this figure, it can be seen that increasing the length of training sequences improves the performance at low  $E_s/N_o$ . But at high  $E_s/N_o$ , the MSEs are the same for all  $L_t$ . Again, the performance of the proposed algorithm is much better than that of optimum samples selection algorithm. It is also notable that the analytic MSE expressions match the simulation results very well.

## 3. Effect of Number of Receive Antennas

Fig. 18 compares the MSE for different number of receive antennas when two transmit antennas and  $L_t = 32$  are used. We can see that increasing the number of receive antennas reduces the MSE at low  $E_s/N_o$ , but it does not help at high  $E_s/N_o$ . The

proposed algorithm exhibits much smaller MSE than the optimum sample selection algorithm. When comparing the theoretical and simulation results of the proposed algorithm, it can be seen that they match pretty well except for  $M = 1$  case. This is due to the fact that the arctan approximation in (4.28), in general, holds only for AWGN channels<sup>1</sup> but not for fading channels. In the presence of fading, the channel output may assume a large range of values and the approximation does not hold anymore. Of course, a better approximation, such as  $\arctan(x) \approx x - x^3/3 + x^5/5$  may be used, but the analysis would become extremely complicated as higher order moments are involved. Fortunately, as the number of transmit or receive antenna increases, the equivalent averaged channel across all transmit/receive antennas tends to behave like an AWGN channel and the approximation becomes valid again. This can be seen from the cases  $M = 2$  and  $M = 4$ , the theoretical and the simulation results are closer when compared with the  $M = 1$  case. For  $M = 8$  and  $M = 16$ , the theoretical and the simulation results match exactly.

#### 4. Effect of Number of Transmit Antennas

Finally, we assess the MSE when different number of transmit antennas are used with  $L_t = 64$ . The results shown in Fig. 19 illustrate that increasing the number of transmit antennas does not change the MSE performances. The theoretical and the simulation results for the proposed algorithm match very well. Once again, the proposed algorithm performs much better.

---

<sup>1</sup>Note that this approximation has been applied in similar applications [3], [60] in AWGN channels only.

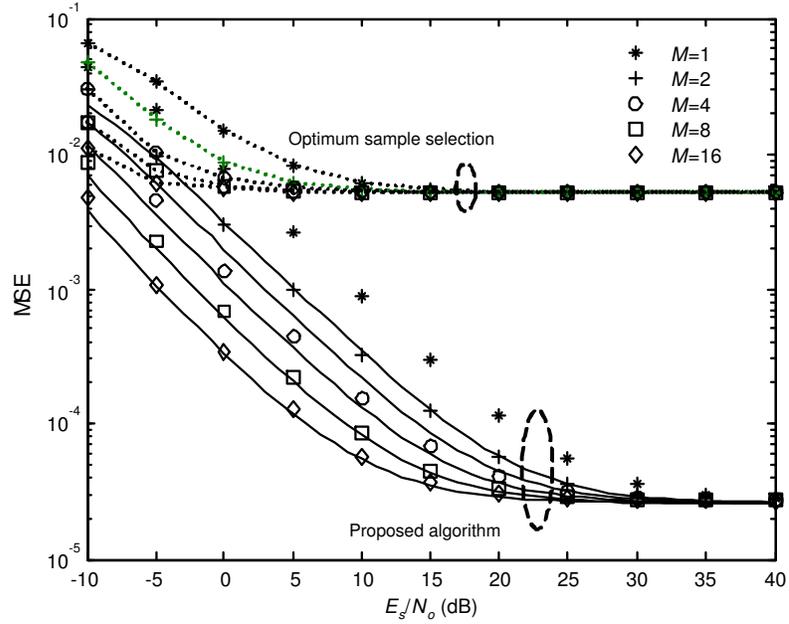


Fig. 18. MSE performance for different number of receive antennas  $M$  ( $N=2$ ,  $L_t=32$ ,  $Q=4$ ,  $\alpha=0.3$ ).

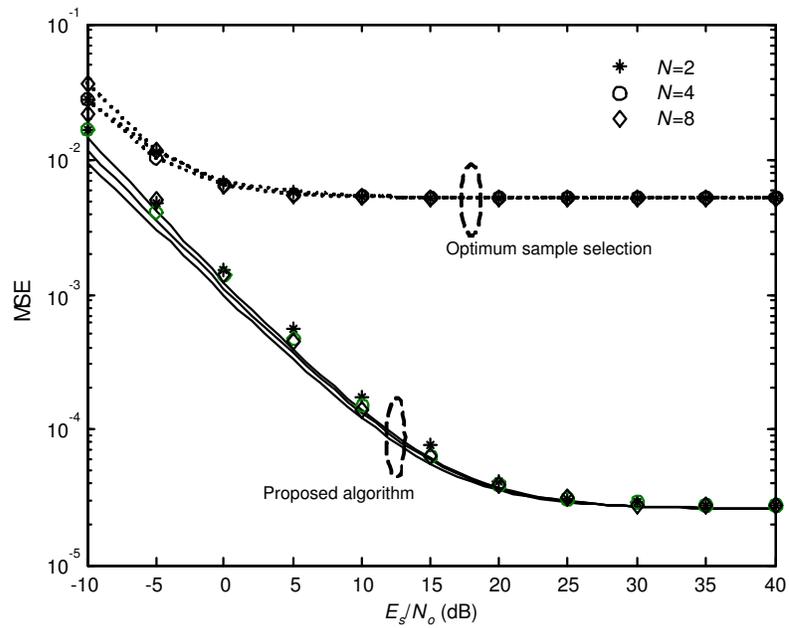


Fig. 19. MSE performance for different number of transmit antennas  $N$  ( $L_t=64$ ,  $Q=4$ ,  $\alpha=0.3$ ,  $M=2$ ).

## H. Conclusions

A new symbol-timing delay estimator for MIMO flat-fading channels has been proposed. It improves the optimum sample selection algorithm of Naguib *et al.* [11] such that accurate timing estimates are obtained even if the oversampling ratio is small. The increase in implementation complexity with respect to the optimum sample selection algorithm is very small. The requirements and the design procedure for the training sequences are discussed. Analytical expressions for MSE of the proposed estimator are derived. It is shown that the MSE analytical expressions match very well with the simulation results in most of the cases. Simulation results also show that, for modest oversampling ratios (such as  $Q=4$ ), the MSE of the proposed estimator is significantly smaller than that of the optimum sample selection algorithm. Furthermore, the performance of the proposed algorithm improves with the number of receive antennas being employed or the length of training sequences.

## CHAPTER V

SYMBOL TIMING ESTIMATION IN MIMO FLAT-FADING CHANNELS –  
MAXIMUM LIKELIHOOD APPROACH

## A. Introduction

In Chapter IV, we extended the symbol timing estimation algorithm by Naguib *et al.* [11] such that accurate timing estimates are obtained even if the oversampling ratio is small. However, the algorithms in [11] and that in Chapter IV are derived in an *ad hoc* fashion and there is no objective criteria for comparison. In this chapter, we approach the same problem using the maximum likelihood (ML) technique. Furthermore, due to a different system model employed in this chapter, the correlation between antennas can be easily taken care of.

More specifically, in this chapter, the data aided (DA) and non-data aided (NDA) maximum likelihood (ML) symbol timing estimators in MIMO correlated flat-fading channels are derived. The technique of conditional ML [9], [10], in which the nuisance parameters are treated as deterministic but unknown and are estimated together with the parameter of interest, is employed. The advantage of conditional ML method is that there is no need to know or assume the statistical properties of the nuisance parameters. It is shown that the approximated ML algorithms in [11] and Chapter IV are just a special case of the DA ML estimator; while the extended squaring algorithm in [61] is just a special case of the NDA ML estimator. For the DA case, the optimal orthogonal training sequences are also derived. It is found that the optimal orthogonal training sequences resemble Walsh sequences, but with different envelopes. Two performance bounds are derived for comparison. The first one is the conditional Cramer-Rao bound (CCRB) [62], [63], which is the Cramer-Rao bound (CRB) for

the symbol timing estimation conditioned that the nuisance parameters are treated as deterministic and are jointly estimated together with the unknown symbol timing. Therefore, the CCRB serves as a performance lower bound for the ML estimators derived. The second one is the modified CRB (MCRB) [64], which is a lower bound for *any* unbiased symbol timing estimator, irrespective of the underlying assumption about the nuisance parameters. Being easier to evaluate than CRB, MCRB serves as the ultimate estimation accuracy that may be achieved.

Simulation results under different operating conditions (e.g., number of antennas and correlation between antennas) are given to assess the performances of the DA and NDA ML estimators and compared to the corresponding CCRBs and MCRBs. It is found that i) the mean square error (MSE) of the DA ML estimator is close to the CCRB and MCRB, meaning that the DA ML estimator is almost the best estimator (in terms of the MSE performance) for the problem under consideration; ii) the MSE of the NDA ML estimator is close to the CCRB but not MCRB, meaning that NDA ML estimator is an efficient estimator conditioned that the nuisance parameters are being jointly estimated, but there *might* exist other NDA estimators with better performances; iii) the MSEs of both DA and NDA ML estimators are approximately independent of the number of transmit antennas and are inversely proportional to the number of receive antennas; iv) correlation between antennas has little effect on the MSEs of DA and NDA ML estimators unless the correlation coefficient between adjacent antennas is larger than 0.5, in which case small degradation errors occur, and v) DA ML estimator performs better than NDA ML estimator at the cost of lower transmission efficiency and higher implementation complexity.

The rest of the chapter is organized as follows. The signal model is first described in Section B. The DA symbol timing estimation problem is addressed in Section C, in which the ML estimator, the corresponding CCRB and MCRB and the optimal

orthogonal training sequences are derived. The NDA ML symbol timing estimator and the corresponding CCRB and MCRB are presented in Section D. Simulation results are then presented in Section E, and finally conclusions are drawn in Section F.

## B. Signal Model

We start with the same signal model (4.1) in Chapter IV. There is only one small difference here: the unknown timing offset  $\varepsilon_o$  is assumed to be uniformly distributed in the range  $[0, 1)$  rather than  $[-0.5, 0.5)$ , but this is only for notational convenience and does not affect the nature of the problem.

After passing through the anti-aliasing filter<sup>1</sup>, the received signal is then sampled at rate  $f_s = 1/T_s$ , where  $T_s \triangleq T/Q$ . Note that the oversampling factor  $Q$  is determined by the frequency span of  $g(t)$ ; if  $g(t)$  is bandlimited to  $f = \pm 1/T$  (an example of which is the root raised cosine (RRC) pulse), then  $Q = 2$  is sufficient. The received vector  $\mathbf{r}_j$ , which consists of  $L_o Q$  consecutive received samples ( $L_o$  is the observation length) from the  $j^{\text{th}}$  receive antenna, can be expressed as (without loss of generality, we consider the received sequence start at  $t = 0$ )

$$\mathbf{r}_j = \xi \mathbf{A}_{\varepsilon_o} \mathbf{Z} \mathbf{H}_{j,:}^T + \boldsymbol{\eta}_j, \quad (5.1)$$

where  $\xi \triangleq \sqrt{E_s/NT}$ ,

$$\mathbf{r}_j \triangleq [r_j(0) \ r_j(T_s) \ \dots \ r_j((L_o Q - 1)T_s)]^T, \quad (5.2)$$

$$\mathbf{A}_{\varepsilon_o} \triangleq [\mathbf{a}_{-L_g}(\varepsilon_o) \ \mathbf{a}_{-L_g+1}(\varepsilon_o) \ \dots \ \mathbf{a}_{L_o+L_g-1}(\varepsilon_o)], \quad (5.3)$$

$$\mathbf{a}_i(\varepsilon_o) \triangleq [g(-iT - \varepsilon_o T) \ g(T_s - iT - \varepsilon_o T) \ \dots \ g((L_o Q - 1)T_s - iT - \varepsilon_o T)]^T, \quad (5.4)$$

---

<sup>1</sup>We assume there is no matched filter here. The effect of matched filter can be taken care of using the pre-whitening technique in Chapter III Section 3.

$$\mathbf{Z} \triangleq [\mathbf{d}_1 \ \mathbf{d}_2 \ \cdots \ \mathbf{d}_N], \quad (5.5)$$

$$\mathbf{d}_i \triangleq [d_i(-L_g) \ d_i(-L_g + 1) \ \cdots \ d_i(L_o + L_g - 1)]^T, \quad (5.6)$$

$$\mathbf{H} \triangleq \begin{bmatrix} h_{11} & h_{21} & \cdots & h_{N1} \\ h_{12} & h_{22} & \cdots & h_{N2} \\ \vdots & & & \vdots \\ h_{1M} & h_{2M} & \cdots & h_{NM} \end{bmatrix}, \quad (5.7)$$

$$\boldsymbol{\eta}_j \triangleq [\eta_j(0) \ \eta_j(1) \ \dots \ \eta_j(L_o Q - 1)]^T, \quad (5.8)$$

with  $\eta_j(i) \triangleq \eta_j(iT/Q)$ , and  $L_g$  denotes the number of symbols affected by the inter-symbol interference (ISI) introduced by one side of  $g(t)$ . Stacking the received vectors from all the  $M$  receive antennas gives

$$\mathbf{r} = \xi(\mathbf{I}_M \otimes \mathbf{A}_{\varepsilon_o}) \text{vec}(\mathbf{Z}\mathbf{H}^T) + \boldsymbol{\eta}, \quad (5.9)$$

where  $\mathbf{r} \triangleq [\mathbf{r}_1^T \ \mathbf{r}_2^T \ \dots \ \mathbf{r}_M^T]^T$  and  $\boldsymbol{\eta} \triangleq [\boldsymbol{\eta}_1^T \ \boldsymbol{\eta}_2^T \ \dots \ \boldsymbol{\eta}_M^T]^T$ .

In order to include the correlation between channel coefficients, the channel transfer function is expressed as:

$$\mathbf{H} = \sqrt{\boldsymbol{\Phi}_R} \mathbf{H}_{\text{i.i.d.}} \sqrt{\boldsymbol{\Phi}_T}^T, \quad (5.10)$$

where  $\boldsymbol{\Phi}_T$  and  $\boldsymbol{\Phi}_R$  are the power correlation matrices [53] (normalized such that the diagonal elements are ones) of transmit and receive antenna arrays (which are assumed known), respectively;  $\mathbf{H}_{\text{i.i.d.}} \in \mathbb{C}^{M \times N}$  contains independently and identically distributed (i.i.d.) zero-mean, unit-variance, circular symmetric complex Gaussian entries and the matrix square roots denote Cholesky factors such that  $\sqrt{\boldsymbol{\Phi}} \sqrt{\boldsymbol{\Phi}}^H = \boldsymbol{\Phi}$ . Note that (5.10) models the correlation among transmit and receive antenna arrays independently. This model is based on the assumption that only immediate surroundings of the antenna array impose the correlation between antenna array elements and

have no impact on the correlations at the other end of the communication link. The validity of this model for narrowband nonline-of-sight MIMO channels is verified by recent measurements [50]-[53]. Substituting (5.10) into (5.9), we obtain:

$$\mathbf{r} = \xi(\mathbf{I}_M \otimes \mathbf{A}_{\varepsilon_o}) \text{vec}(\mathbf{Z} \sqrt{\Phi_T} \mathbf{H}_{\text{i.i.d.}}^T \sqrt{\Phi_R}^T) + \boldsymbol{\eta}. \quad (5.11)$$

### C. Symbol Timing Estimation with Known Training Data

#### 1. ML Estimator

In this case, the matrix  $\mathbf{Z}$  contains the known training sequences and the only unknown is  $\mathbf{H}_{\text{i.i.d.}}$ . Noting the fact that  $\text{vec}(\mathbf{A}\mathbf{Y}\mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}\mathbf{Y}$ , then (5.11) becomes

$$\begin{aligned} \mathbf{r} &= \xi(\mathbf{I}_M \otimes \mathbf{A}_{\varepsilon_o}) (\sqrt{\Phi_R} \otimes \mathbf{Z} \sqrt{\Phi_T}) \text{vec}(\mathbf{H}_{\text{i.i.d.}}^T) + \boldsymbol{\eta} \\ &= \xi(\sqrt{\Phi_R} \otimes \mathbf{A}_{\varepsilon_o} \mathbf{Z} \sqrt{\Phi_T}) \text{vec}(\mathbf{H}_{\text{i.i.d.}}^T) + \boldsymbol{\eta}, \end{aligned} \quad (5.12)$$

where the last line comes from the fact that  $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$ .

From (5.12), the joint maximum likelihood estimate of  $\varepsilon_o$  and  $\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)$  is obtained by maximizing

$$p(\mathbf{r}|\varepsilon, \mathbf{h}) = \frac{1}{(\pi N_o)^{L_o Q}} \exp \left[ -\frac{(\mathbf{r} - \bar{\mathbf{A}}_\varepsilon \mathbf{h})^H (\mathbf{r} - \bar{\mathbf{A}}_\varepsilon \mathbf{h})}{N_o} \right], \quad (5.13)$$

or equivalently minimizing

$$J_1(\mathbf{r}|\varepsilon, \mathbf{h}) = (\mathbf{r} - \bar{\mathbf{A}}_\varepsilon \mathbf{h})^H (\mathbf{r} - \bar{\mathbf{A}}_\varepsilon \mathbf{h}), \quad (5.14)$$

where  $\bar{\mathbf{A}}_\varepsilon \triangleq \xi(\sqrt{\Phi_R} \otimes \mathbf{A}_\varepsilon \mathbf{Z} \sqrt{\Phi_T})$ , and  $\varepsilon$  and  $\mathbf{h}$  are the trial values for  $\varepsilon_o$  and  $\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)$ , respectively.

Setting the partial derivatives of  $J_1(\mathbf{r}|\varepsilon, \mathbf{h})$  with respect to  $\mathbf{h}$  equal to zero, we

obtain the ML estimate for  $\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)$  (when  $\varepsilon$  is fixed) as [65]

$$\hat{\mathbf{h}} = (\bar{\mathbf{A}}_\varepsilon^H \bar{\mathbf{A}}_\varepsilon)^{-1} \bar{\mathbf{A}}_\varepsilon^H \mathbf{r}. \quad (5.15)$$

Substituting (5.15) into (5.14), after some straightforward manipulations and dropping the irrelevant terms, the timing delay is estimated by maximizing the following likelihood function

$$\Lambda_{DA}(\varepsilon) = \mathbf{r}^H \bar{\mathbf{A}}_\varepsilon (\bar{\mathbf{A}}_\varepsilon^H \bar{\mathbf{A}}_\varepsilon)^{-1} \bar{\mathbf{A}}_\varepsilon^H \mathbf{r}. \quad (5.16)$$

Using the well known properties of the Kronecker product  $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$  and  $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$  to expand  $\bar{\mathbf{A}}_\varepsilon (\bar{\mathbf{A}}_\varepsilon^H \bar{\mathbf{A}}_\varepsilon)^{-1} \bar{\mathbf{A}}_\varepsilon^H$ , we have

$$\begin{aligned} \bar{\mathbf{A}}_\varepsilon (\bar{\mathbf{A}}_\varepsilon^H \bar{\mathbf{A}}_\varepsilon)^{-1} \bar{\mathbf{A}}_\varepsilon^H &= [\sqrt{\Phi_R} (\sqrt{\Phi_R}^H \sqrt{\Phi_R})^{-1} \sqrt{\Phi_R}^H] \\ &\quad \otimes [\mathbf{A}_\varepsilon \mathbf{Z} \sqrt{\Phi_T} (\sqrt{\Phi_T}^H \mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \mathbf{Z} \sqrt{\Phi_T})^{-1} \sqrt{\Phi_T}^H \mathbf{Z}^H \mathbf{A}_\varepsilon^H] \\ &= \mathbf{I}_M \otimes \mathbf{A}_\varepsilon \mathbf{Z} (\mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_\varepsilon^H, \end{aligned} \quad (5.17)$$

where in the second equality, we used the fact that  $\sqrt{\Phi_R}$  and  $\sqrt{\Phi_T}$  are both non-singular square matrices. Substituting this result back into (5.16), the DA likelihood function is given by

$$\begin{aligned} \Lambda_{DA}(\varepsilon) &= \mathbf{r}^H (\mathbf{I}_M \otimes \mathbf{A}_\varepsilon \mathbf{Z} (\mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_\varepsilon^H) \mathbf{r} \\ &= \sum_{j=1}^M \mathbf{r}_j^H \mathbf{A}_\varepsilon \mathbf{Z} (\mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{r}_j, \end{aligned} \quad (5.18)$$

and the  $\text{ML}_{DA}$  symbol timing estimator can be written as

$$\hat{\varepsilon} = \arg \max_{\varepsilon} \Lambda_{DA}(\varepsilon). \quad (5.19)$$

We make the following remarks:

1. The maximization of the likelihood function usually involves a two-step approach. The first step (coarse search) computes  $\Lambda_{DA}(\varepsilon)$  over a grid of timing

delay  $\varepsilon_k \triangleq k/K$  for  $k = 0, 1, \dots, K - 1$ , and then the  $\varepsilon_k$  that maximizes  $\Lambda_{DA}(\varepsilon)$  is selected. The second step (fine search) finds the global maximum by using either the gradient method [63], dichotomous search [66], or interpolation [66]. In this chapter, we employ the parabolic interpolation in the second step due to its implementation simplicity. More specifically, assume that  $\Lambda_{DA}(\varepsilon_{\hat{k}})$  is identified as the maximum among all  $\Lambda_{DA}(\varepsilon_k)$  in the first step. Define  $I_1 \triangleq \Lambda_{DA}(\varepsilon_{\hat{k}-1})$ ,  $I_2 \triangleq \Lambda_{DA}(\varepsilon_{\hat{k}})$  and  $I_3 \triangleq \Lambda_{DA}(\varepsilon_{\hat{k}+1})$ , then [66]

$$\hat{\varepsilon} = \varepsilon_{\hat{k}} + \frac{I_1 - I_3}{2K(I_1 + I_3 - 2I_2)}. \quad (5.20)$$

2. The low-complexity maximization technique introduced in Chapter II Section C can be applied to maximize (5.18). However, since this low-complexity maximization technique involves an approximation in the Fourier series expansion, it introduces estimation error floor at high SNRs. Therefore, we would use the two-step maximization method in Remark 1 for the subsequent discussions.
3. The likelihood function at each receive antenna can be calculated independently and then added together to obtain the overall likelihood function.
4. The correlations in the transmit and receive antenna arrays do not appear in the estimator. That is, the  $\text{ML}_{DA}$  symbol timing estimator is independent of the antenna correlations. This is a reasonable result since another way of deriving the DA likelihood function (5.18) is not separating  $\sqrt{\mathbf{\Phi}_R}$  and  $\sqrt{\mathbf{\Phi}_T}$  from  $\mathbf{H}_{\text{i.i.d.}}$  and treat  $\text{vec}(\mathbf{H}^T)$  as deterministic unknown. Thus,  $\mathbf{\Phi}_R$  and  $\mathbf{\Phi}_T$  would not appear in the estimator.
5. In order for the estimate of  $\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)$  to hold in (5.15), it is required that  $\bar{\mathbf{A}}_\varepsilon$  is full rank [65], or equivalently  $\sqrt{\mathbf{\Phi}_R}$ ,  $\mathbf{A}_\varepsilon$ ,  $\mathbf{Z}$  and  $\sqrt{\mathbf{\Phi}_T}$  are all full rank. Note that

$\sqrt{\Phi_R}$  and  $\sqrt{\Phi_T}$  are lower triangular matrices with positive diagonal elements [67], so they are full rank. Furthermore, if  $g(t)$  being a RRC pulse (which is the most frequently used pulse shape), numerical calculations show that  $\mathbf{A}_\varepsilon$  is full rank. Finally,  $\mathbf{Z}$  can be made full rank by properly designing the training data. A sufficient condition is that parts of the training sequences from different transmit antennas are orthogonal. That is, for  $i \neq j$ ,

$$[d_i(a) \cdots d_i(b)] \cdot [d_j(a) \cdots d_j(b)]^H = 0, \quad (5.21)$$

for some  $a, b \in \{-L_g, -L_g + 1, \dots, L_o + L_g - 1\}$  with  $a < b$ .

6. For a large observation interval  $L_o$ , the  $(i, j)^{th}$  element of  $\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon$  ( $i, j = -L_g, L_g + 1, \dots, L_o + L_g - 1$ ) can be approximated by

$$[\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon]_{ij} \approx \sum_{n=-\infty}^{\infty} g^*(nT_s - iT - \varepsilon T) g(nT_s - jT - \varepsilon T) = R_{gg}((i - j)T), \quad (5.22)$$

where  $R_{gg}(\tau)$  is the continuous autocorrelation function of  $g(t)$  and the last equality is due to the fact that the sampling rate is at least at the Nyquist rate, which guarantees the equivalence between the discrete and continuous autocorrelation functions of  $g(t)$ . Therefore,  $[\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon]_{ij}$  is approximately independent of  $\varepsilon$ . Note that this approximation is very accurate for the central portion of  $\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon$ . If  $R_{gg}(\tau)$  satisfies the Nyquist condition for zero ISI (e.g.,  $g(t)$  being a RRC pulse or the class of non-bandlimited pulse shapes with  $R_{gg}(\tau)$  being time-limited to  $[-T/2, T/2]$ ), then  $[\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon]_{ij} \approx \delta_{ij}$ . Furthermore, if the training sequences from different transmit antennas are orthogonal and with the same norm (i.e.,  $\mathbf{Z}^H \mathbf{Z} = c \mathbf{I}_N$  for some constant  $c$ ), then

$$\Lambda_{DA}(\varepsilon) \approx \frac{1}{c} \sum_{j=1}^M \mathbf{r}_j^H \mathbf{A}_\varepsilon \mathbf{Z} \mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{r}_j = \frac{1}{c} \sum_{j=1}^M \sum_{i=1}^N |\mathbf{d}_i^H \mathbf{A}_\varepsilon^H \mathbf{r}_j|^2. \quad (5.23)$$

Note that  $\mathbf{A}_\varepsilon^H \mathbf{r}_j$  is the matched filtering of  $\mathbf{r}_j$  with one output sample per symbol with delay  $\varepsilon$  [10]. If the function (5.23) is sampled uniformly with  $Q$  points, this reduces to the approximated ML function in [11] and that of Chapter IV (see (4.11)).

7. An interesting question is how large  $L_o$  is sufficient for the use of (5.23) in place of (5.18) without a noticeable loss in performance. The answer depends on the signal-to-noise ratio (SNR) where the estimators work. In general, the higher the SNR, the larger the  $L_o$  is required. For example, Fig. 20 compares the MSE performances of the true ML estimator and the approximated ML estimator (the training sequences are the optimal orthogonal sequences derived later in this chapter). It can be seen that for  $\text{SNR} \leq 20\text{dB}$ ,  $L_o = 32$  is enough for both estimators to have similar performances. For  $\text{SNR} = 30\text{dB}$ ,  $L_o = 64$  is required.
8. In some space-time processing algorithms, (e.g., space-time coding [46]-[48]), it is required that the channel matrix be also estimated. It is clear that once the timing estimate  $\hat{\varepsilon}$  has been found by maximizing (5.19), the channel estimate can also be obtained readily by using (5.15). Putting  $\varepsilon = \hat{\varepsilon}$  into (5.15) and expanding it gives

$$\hat{\mathbf{h}} = \xi^{-1} ((\sqrt{\Phi_R})^{-1} \otimes (\sqrt{\Phi_T})^{-1}) (\mathbf{Z}^H \mathbf{A}_{\hat{\varepsilon}}^H \mathbf{A}_{\hat{\varepsilon}} \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_{\hat{\varepsilon}}^H \mathbf{r}. \quad (5.24)$$

If the channel coefficients are uncorrelated (i.e.,  $\Phi_T = \mathbf{I}_N$  and  $\Phi_R = \mathbf{I}_M$ ) and the training sequences from different transmit antennas are orthogonal (i.e.,  $\mathbf{Z}^H \mathbf{Z} = c \mathbf{I}_N$ ), it can be easily shown that (5.24) reduces to

$$\hat{h}_{ij} \approx \frac{1}{c\xi} \mathbf{d}_i^* \mathbf{A}_{\hat{\varepsilon}}^H \mathbf{r}_j, \quad (5.25)$$

which is the channel estimation method proposed in [11].

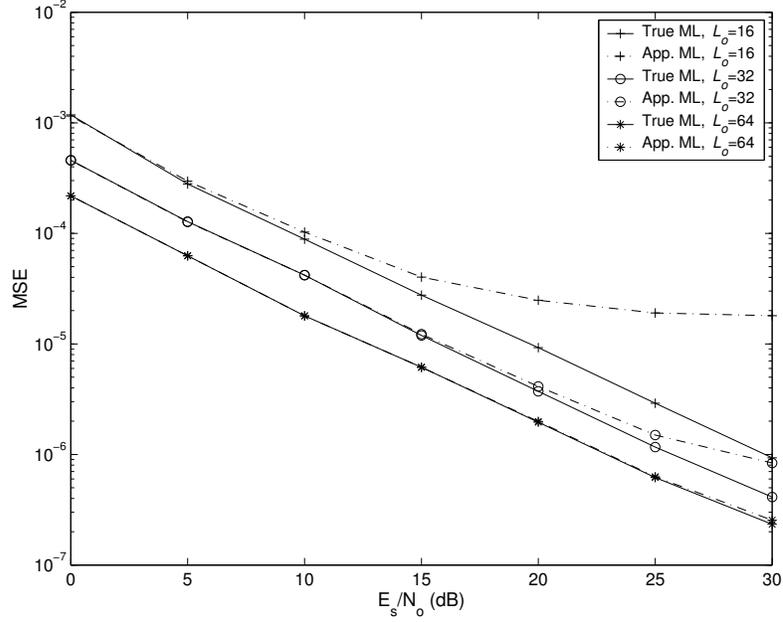


Fig. 20. MSE performances comparison between the true and approximated DA ML estimators with different  $L_o$  ( $M = N = 4$ ,  $L_g = 4$ ,  $g(t)$  being a RRC pulse with roll-off factor  $\alpha = 0.3$ ,  $\Phi_T = \mathbf{I}_4$ ,  $\mathbf{Z} = \mathbf{Z}_{opt}$ ).

## 2. The CCRB and MCRB

For the model in (5.12), it is known that for a specific timing delay  $\varepsilon_o$ , the  $\text{CCRB}_{DA}$  is given by<sup>2</sup> [9]

$$\text{CCRB}_{DA}(\varepsilon_o) = \frac{\sigma^2}{2\text{tr}(\bar{\mathbf{D}}_{\varepsilon_o}^H \mathbf{P}_{\bar{\mathbf{A}}} \bar{\mathbf{D}}_{\varepsilon_o} \mathbf{\Gamma}_h)}. \quad (5.26)$$

In (5.26),  $\sigma^2 = N_o f_s = N_o Q/T$  is the noise variance,  $\text{tr}(\cdot)$  denotes the trace of a matrix,

$$\bar{\mathbf{D}}_{\varepsilon} \triangleq \frac{d\bar{\mathbf{A}}_{\varepsilon}}{d\varepsilon} = \xi \sqrt{\Phi_R} \otimes \mathbf{D}_{\varepsilon} \mathbf{Z} \sqrt{\Phi_T}, \quad (5.27)$$

<sup>2</sup>Strictly speaking, the bound given is the asymptotic CCRB. However, it is shown in [9] that the true CCRB tends to the asymptotic CCRB when  $M, N \rightarrow \infty$

with  $\mathbf{D}_\varepsilon \triangleq d\mathbf{A}_\varepsilon/d\varepsilon$ ,  $\mathbf{P}_{\bar{\mathbf{A}}}^\perp$  is the orthogonal projector onto the null space of  $\bar{\mathbf{A}}_{\varepsilon_o}$  and is given by

$$\begin{aligned}\mathbf{P}_{\bar{\mathbf{A}}}^\perp &\triangleq \mathbf{I}_{ML_oQ} - \bar{\mathbf{A}}_{\varepsilon_o}(\bar{\mathbf{A}}_{\varepsilon_o}^H \bar{\mathbf{A}}_{\varepsilon_o})^{-1} \bar{\mathbf{A}}_{\varepsilon_o}^H \\ &= \mathbf{I}_M \otimes (\mathbf{I}_{L_oQ} - \mathbf{A}_{\varepsilon_o} \mathbf{Z} (\mathbf{Z}^H \mathbf{A}_{\varepsilon_o}^H \mathbf{A}_{\varepsilon_o} \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_{\varepsilon_o}^H) \\ &= \mathbf{I}_M \otimes \mathbf{P}_{\mathbf{AZ}}^\perp,\end{aligned}\quad (5.28)$$

where  $\mathbf{P}_{\mathbf{AZ}}^\perp \triangleq \mathbf{I}_{L_oQ} - \mathbf{A}_{\varepsilon_o} \mathbf{Z} (\mathbf{Z}^H \mathbf{A}_{\varepsilon_o}^H \mathbf{A}_{\varepsilon_o} \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_{\varepsilon_o}^H$ , and

$$\mathbf{\Gamma}_{\mathbf{h}} \triangleq \mathbb{E}[\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T) \text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)^H] = \mathbf{I}_{MN} = \mathbf{I}_M \otimes \mathbf{I}_N. \quad (5.29)$$

Substituting (5.27), (5.28) and (5.29) into (5.26), we obtain:

$$\begin{aligned}\text{CCRB}_{DA}(\varepsilon_o) &= \frac{\sigma^2}{2\xi^2 \text{tr}((\sqrt{\Phi_R} \otimes \mathbf{D}_{\varepsilon_o} \mathbf{Z} \sqrt{\Phi_T})^H (\mathbf{I}_M \otimes \mathbf{P}_{\mathbf{AZ}}^\perp) (\sqrt{\Phi_R} \otimes \mathbf{D}_{\varepsilon_o} \mathbf{Z} \sqrt{\Phi_T}) (\mathbf{I}_M \otimes \mathbf{I}_N))} \\ &= \frac{QN}{2 \text{tr}(\sqrt{\Phi_R}^H \sqrt{\Phi_R}) \text{tr}(\sqrt{\Phi_T}^H \mathbf{Z}^H \mathbf{D}_{\varepsilon_o}^H \mathbf{P}_{\mathbf{AZ}}^\perp \mathbf{D}_{\varepsilon_o} \mathbf{Z} \sqrt{\Phi_T})} \left(\frac{E_s}{N_o}\right)^{-1} \\ &= \frac{1}{2M \text{tr}(\tilde{\mathbf{Z}}^H \tilde{\mathbf{D}}_{\varepsilon_o}^H \mathbf{P}_{\mathbf{AZ}}^\perp \tilde{\mathbf{D}}_{\varepsilon_o} \tilde{\mathbf{Z}} \Phi_T)} \left(\frac{E_s}{N_o}\right)^{-1},\end{aligned}\quad (5.30)$$

where  $\tilde{\mathbf{Z}} \triangleq \mathbf{Z}/\sqrt{N}$  and  $\tilde{\mathbf{D}}_\varepsilon \triangleq \mathbf{D}_\varepsilon/\sqrt{Q}$ . In passing from the second line to the third line in (5.30), we used the fact that  $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$  and the diagonal elements of  $\Phi_R$  are all one regardless of the specific value of the correlation matrix.

For a specific timing delay  $\varepsilon_o$ ,  $\text{MCRB}_{DA}$  is given by [9]

$$\text{MCRB}_{DA}(\varepsilon_o) = \frac{\sigma^2}{2 \text{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o} \mathbf{\Gamma}_{\mathbf{h}})}, \quad (5.31)$$

and based on similar calculations with those used for  $\text{CCRB}_{DA}$ , it can be shown that

$$\text{MCRB}_{DA}(\varepsilon_o) = \frac{1}{2M \text{tr}(\tilde{\mathbf{Z}}^H \tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o} \tilde{\mathbf{Z}} \Phi_T)} \left(\frac{E_s}{N_o}\right)^{-1}. \quad (5.32)$$

The following remarks concerning the  $\text{CCRB}_{DA}$  and  $\text{MCRB}_{DA}$  are now in order:

1. Since the timing delay  $\varepsilon_o$  is assumed uniformly distributed, the average of  $\text{CCRB}_{DA}$  and  $\text{MCRB}_{DA}$  can be calculated by numerical integration of (5.30) and (5.32), respectively.
2. The  $\text{CCRB}_{DA}$  and  $\text{MCRB}_{DA}$  do not depend on the receive antenna array correlation matrix  $\Phi_R$ . Furthermore, the  $\text{CCRB}_{DA}$  and  $\text{MCRB}_{DA}$  are inversely proportional to the number of receive antennas  $M$ . Thus, the  $\text{CCRB}_{DA}$  and  $\text{MCRB}_{DA}$  will be reduced by a factor of 2 whenever the number of receive antennas  $M$  is doubled.
3. The expressions for  $\text{CCRB}_{DA}$  and  $\text{MCRB}_{DA}$  would still be given by (5.30) and (5.32) respectively even if we treat  $\text{vec}(\mathbf{H}^T)$  as deterministic unknown rather than  $\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)$  in the system model.

### 3. Optimal Orthogonal Training Sequences

Since the  $\text{CCRB}_{DA}$  can be reached asymptotically by the  $\text{ML}_{DA}$  estimator (5.19) [65], it is natural to search for optimal training sequences by minimizing the  $\text{CCRB}_{DA}$  in (5.30) with respect to  $\mathbf{Z}$ . Unfortunately, since the denominator of (5.30) is a very complicated function of  $\mathbf{Z}$ , it is difficult, if not impossible, to obtain a simple solution. On the other hand, the expression for the  $\text{MCRB}_{DA}$  in (5.32) has a much simpler dependence on  $\mathbf{Z}$ . Moreover, it will be shown later in this section that for the derived optimal training sequences, the corresponding  $\text{CCRB}_{DA}$  is actually very close to that of  $\text{MCRB}_{DA}$  (see Fig. 22). Therefore, in the following the optimal training sequences are derived by minimizing the  $\text{MCRB}_{DA}$  with respect to  $\mathbf{Z}$ .

With the constraint that the columns of  $\mathbf{Z}$  has to be orthogonal<sup>3</sup> (i.e.,  $\mathbf{Z}^H\mathbf{Z} = (L_o + 2L_g)\mathbf{I}_N$ ), it is proved in Appendix C that the matrix  $\mathbf{Z}$  that minimizes the  $\text{MCRB}_{DA}$  is given by

$$\mathbf{Z} = \sqrt{(L_o + 2L_g)}\tilde{\mathbf{U}}(\varepsilon_o)\mathbf{U}_T^H, \quad (5.33)$$

where  $\tilde{\mathbf{U}}(\varepsilon_o)$  is the matrix containing the  $N$  eigenvectors corresponding to the  $N$  largest eigenvalues of  $\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}$  as columns and  $\mathbf{U}_T$  is the unitary matrix containing all the eigenvectors of  $\Phi_T$  as columns.

In general, the optimal orthogonal training sequences depend on the unknown parameter  $\varepsilon_o$ . However, note that, following the same argument as in (5.22),  $[\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}]_{ij} \approx R_{g'g'}((i-j)T)T^2/Q$ , where  $g'(t) = dg(t)/dt$ . Therefore,  $\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}$  is approximately independent of the parameter  $\varepsilon_o$  and in practice, we can fix a nominal timing delay, say  $\varepsilon_t = 0$  (actually other values do not make a large difference as we will show), for designing the training sequences. This idea is verified by Fig. 21, where

$$\beta \triangleq \frac{1}{\text{tr}(\tilde{\mathbf{Z}}^H\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}\tilde{\mathbf{Z}}\Phi_T)} \quad (5.34)$$

is plotted against  $\varepsilon_o$  for  $\varepsilon_t = 0, 0.25, 0.5, 0.75$  with  $N = 4$ ,  $L_o = 32$ ,  $L_g = 4$ ,  $g(t)$  being a RRC pulse with roll-off factor  $\alpha = 0.3$  and  $\Phi_T = \mathbf{I}_4$ . The case of  $\varepsilon_t = \varepsilon_o$  is also shown for a reference. It is obvious that the mismatch of  $\varepsilon_t$  and  $\varepsilon_o$  does not increase the value of  $\beta$  significantly. From Fig. 21, we note that the worst case increase of  $\beta$  due to the mismatch of  $\varepsilon_t$  and  $\varepsilon_o$  is about  $2 \times 10^{-5}$  and when  $\varepsilon_t = \varepsilon_o$ ,  $\beta \approx 2.695 \times 10^{-3}$ .

Thus, the worst case relative error for the  $\text{MCRB}_{DA}$  in this example is

$$\frac{\text{MCRB}_{DA}(\varepsilon_o|\varepsilon_t \neq \varepsilon_o) - \text{MCRB}_{DA}(\varepsilon_o|\varepsilon_t = \varepsilon_o)}{\text{MCRB}_{DA}(\varepsilon_o|\varepsilon_t = \varepsilon_o)} \approx \frac{2 \times 10^{-5}}{2.695 \times 10^{-3}} = 7.42 \times 10^{-3}. \quad (5.35)$$

---

<sup>3</sup>Notice that in this chapter, the search for optimal training sequences would be confined to the class of orthogonal sequences. The question of whether there exists any non-orthogonal training sequences with better performances and how to find them is a subject open to future investigations.

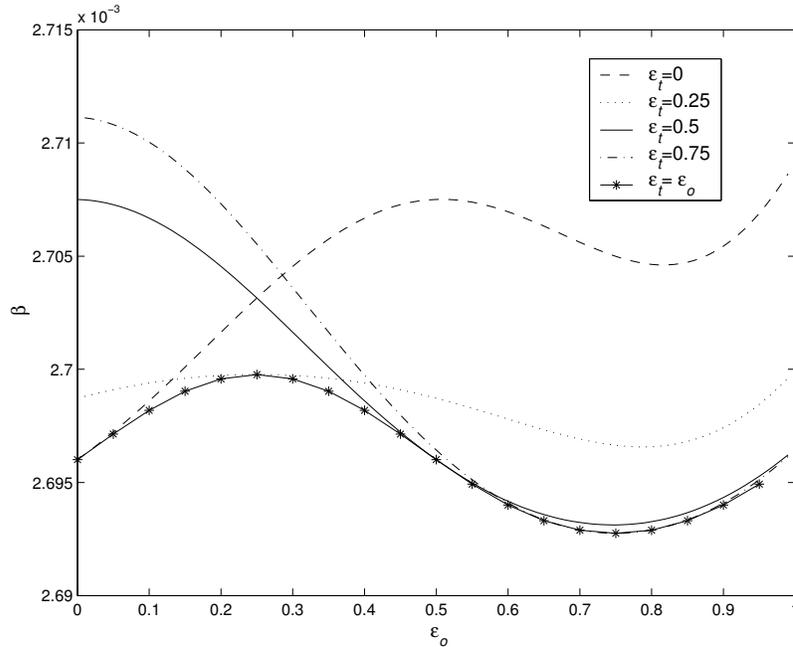


Fig. 21. Plots of  $\beta \triangleq 1/\text{tr}(\tilde{\mathbf{Z}}^H \tilde{\mathbf{D}}_{\epsilon_o}^H \tilde{\mathbf{D}}_{\epsilon_o} \tilde{\mathbf{Z}} \Phi_T)$  against  $\epsilon_o$  for  $\epsilon_t = 0, 0.25, 0.5, 0.75$  ( $N = 4, L_o = 32, L_g = 4, g(t)$  being a RRC pulse with  $\alpha = 0.3, \Phi_T = \mathbf{I}_4$ ).

The implication of the above calculation is that the worst case variation of the  $\text{MCRB}_{DA}(\epsilon_o)$  due to the mismatch between  $\epsilon_o$  and  $\epsilon_t$  is at least 100 times smaller than the value of the  $\text{MCRB}_{DA}(\epsilon_o)$  when  $\epsilon_t = \epsilon_o$ . Therefore, the optimality of the orthogonal training sequences derived is approximately independent of  $\epsilon_o$  and we can write  $\mathbf{Z}_{opt} = \sqrt{(L_o + 2L_g)} \tilde{\mathbf{U}}(0) \mathbf{U}_T^H$ .

With the optimal orthogonal training sequences  $\mathbf{Z}_{opt}$ , the ratios  $\frac{\text{CCRB}_{DA}(\epsilon_o)}{\text{MCRB}_{DA}(\epsilon_o)}$  are plotted in Fig. 22 against the number of transmit antenna  $N$  for  $\epsilon_o = 0, 0.25, 0.5, 0.75$  with  $L_o = 32$  and  $128, L_g = 4, g(t)$  being a RRC pulse with  $\alpha = 0.3$  and  $\Phi_T = \mathbf{I}_N$ . It can be seen that the ratios  $\frac{\text{CCRB}_{DA}(\epsilon_o)}{\text{MCRB}_{DA}(\epsilon_o)}$  for different  $\epsilon_o$  are close to 1 (this is true for the case  $L_o = 128$ , and for moderate number of transmit antennas when  $L_o = 32$ ). Since  $\text{MCRB}_{DA}(\epsilon_o) \leq \text{CCRB}_{DA}(\epsilon_o)$ , even there are some other orthogonal sequences actually minimize the  $\text{CCRB}_{DA}(\epsilon_o)$ , the space for performance improvement is very

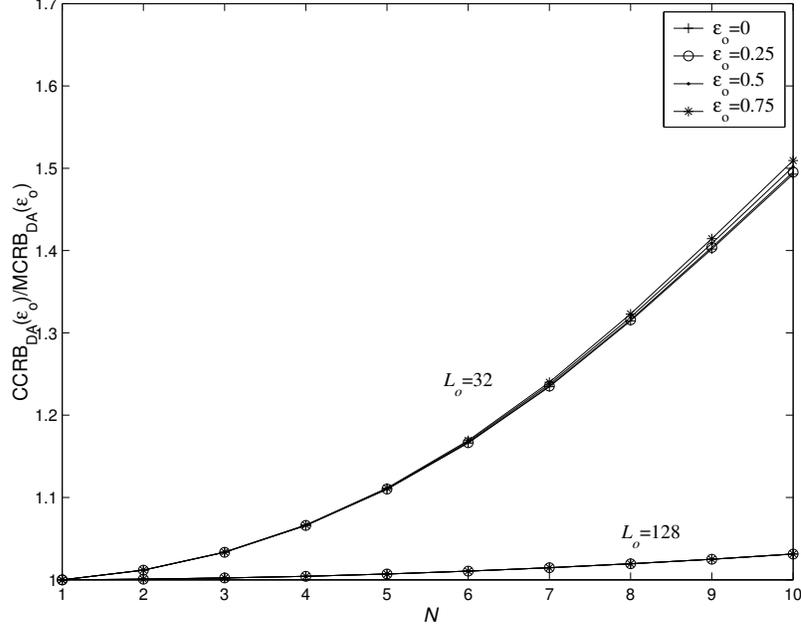


Fig. 22. Plots of  $\text{CCRB}_{DA}(\varepsilon_o)/\text{MCRB}_{DA}(\varepsilon_o)$  against the number of transmit antennas  $N$  for  $\varepsilon_o = 0, 0.25, 0.5, 0.75$  ( $L_o = 32$  and  $128$ ,  $L_g = 4$  and  $g(t)$  being a RRC pulse with  $\alpha = 0.3$ ,  $\Phi_T = \mathbf{I}_N$ ,  $\mathbf{Z} = \mathbf{Z}_{opt}$ ).

small (e.g., for  $L_o = 32$  and  $N \leq 4$ , the ratio  $\frac{\text{CCRB}_{DA}(\varepsilon_o)}{\text{MCRB}_{DA}(\varepsilon_o)}$  is smaller than 1.1, the best possible performance improvement is only  $10 \log_{10}(1.1) \approx 0.4\text{dB}$ ), not mentioning that these training sequences are difficult to find or may even do not exist. This justifies the search for optimal orthogonal training sequences by minimizing the  $\text{MCRB}_{DA}$ .

It is interesting to find that, when  $\Phi_T = \mathbf{I}_N$  and  $g(t)$  is a RRC pulse, the optimal orthogonal training sequences resemble the Walsh sequences. Let  $\mathbf{w}_n$  be the Walsh sequence with length 32 and with  $n$  sign changes. For comparison, Figs. 23 and 24 show  $[\mathbf{Z}_{opt}]_{:,1}$  and  $[\mathbf{Z}_{opt}]_{:,2}$  with  $L_o = 32$ ,  $L_g = 4$  and  $\alpha = 0.3$ , together with  $\mathbf{w}_{31}$  and  $-\mathbf{w}_{30}$  plotted from the index 5 to 36. Note that the lines are drawn for easy reading, there is no value defined in between integer indexes. It can be observed that, the values of the optimal sequences at indices 1 to 4 and 37 to 40 are very small.

Moreover, with the exception of the different envelope shapings, the sign-changing patterns of the optimal orthogonal sequences follow that of Walsh sequences (for indices 5 to 36). In general, the same relationship can be found between  $[\mathbf{Z}_{opt}]_{:,i}$  and  $\mathbf{w}_{32-i}$ . We remark also that the use of Walsh sequences with the largest number of sign changes for symbol timing estimation in space-time coding system has been initially proposed in [61].

Finally, Fig. 25 compares the performance of  $ML_{DA}$  (5.19) with different kinds of training sequences in a 4-transmit, 4-receive antenna system with  $L_o = 32$ ,  $L_g = 4$ ,  $g(t)$  being a RRC pulse with  $\alpha = 0.3$ . For simplicity, we set  $\Phi_T = \Phi_R = \mathbf{I}_4$ . Three different kinds of training sequences are considered. The first one is the optimal orthogonal training sequences derived above. The second one is the Walsh sequences  $\mathbf{w}_{31}$ ,  $\mathbf{w}_{30}$ ,  $\mathbf{w}_{29}$ ,  $\mathbf{w}_{28}$  and extended to length 40 by adding a cyclic prefix and suffix, each of length equal to 4. The final one is the perfect sequences proposed in Chapter IV, where they were derived to minimize the contribution of the ISI term in the approximated log-likelihood function (see Chapter IV Section D for detail). From Fig. 25, it can be seen that the perfect sequences perform not as well as the Walsh sequences and the optimal sequences. This is because the true ML estimator is used in simulations and the perfect sequences (which were derived based on the approximated log-likelihood function) may not have any optimality. Due to the resemblance of the optimal orthogonal sequences and the Walsh sequences, the performance of the  $ML_{DA}$  by using these two kinds of sequences are close to each other, with the case of optimal orthogonal sequences performing marginally better. For fair comparison, we mention that the perfect sequences and the Walsh sequences are constant modulus sequences while the optimal orthogonal sequences are not.

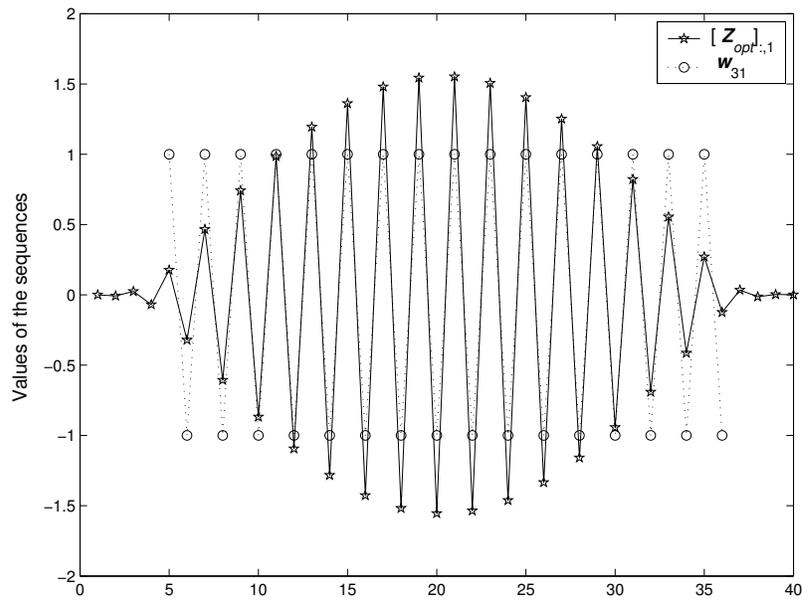


Fig. 23. Plots of  $[\mathbf{Z}_{opt}]_{:,1}$  and  $\mathbf{w}_{31}$  ( $g(t)$  being a RRC pulse with  $\alpha = 0.3$ ,  $L_o = 32$ ,  $L_g = 4$ ).

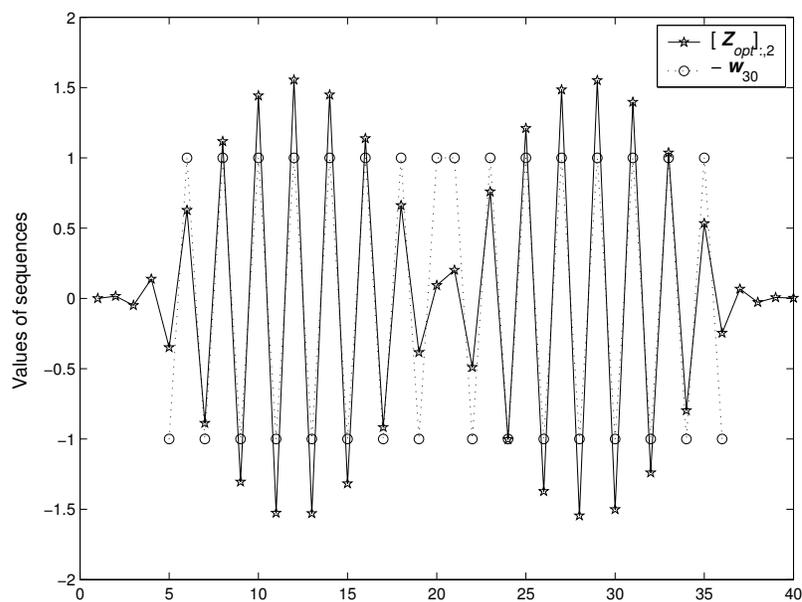


Fig. 24. Plots of  $[\mathbf{Z}_{opt}]_{:,2}$  and  $-\mathbf{w}_{30}$  ( $g(t)$  being a RRC pulse with  $\alpha = 0.3$ ,  $L_o = 32$ ,  $L_g = 4$ ).

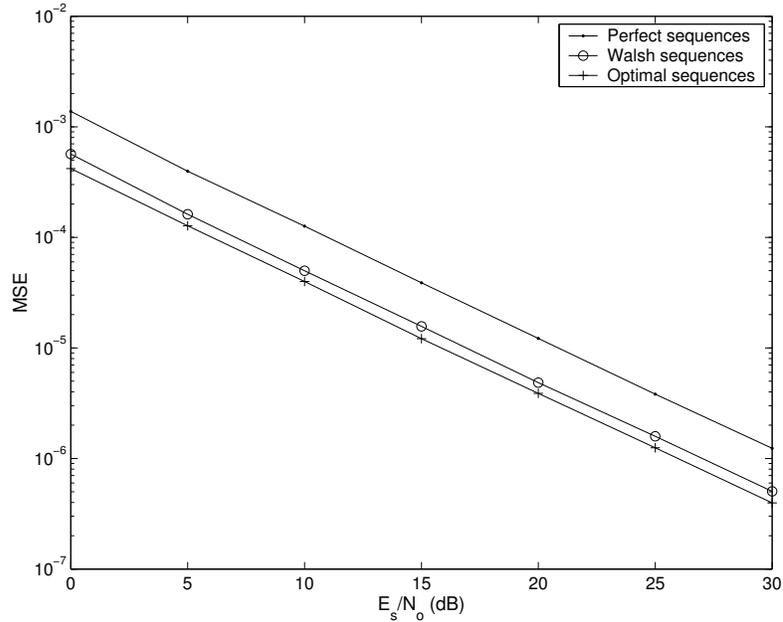


Fig. 25. Comparison of the MSE performances of  $ML_{DA}$  with different training sequences ( $g(t)$  being a RRC pulse with  $\alpha = 0.3$ ,  $M = N = 4$ ,  $L_o = 32$ ,  $L_g = 4$ ,  $\Phi_T = \Phi_R = \mathbf{I}_4$ ).

#### D. Non-Data Aided Symbol Timing Estimation

##### 1. ML Estimator

In this case, no training sequence is used and  $\mathbf{Z}$  contains real data. Now, the matrices  $\mathbf{Z}$  and  $\mathbf{H}_{i.i.d.}$  in (5.11) are unknown and (5.11) can be rewritten in the following form

$$\begin{aligned} \mathbf{r} &= \xi(\mathbf{I}_M \otimes \mathbf{A}_{\varepsilon_o})(\sqrt{\Phi_R} \otimes \mathbf{I}_{L_o+2L_g})\text{vec}(\mathbf{Z}\sqrt{\Phi_T}\mathbf{H}_{i.i.d.}^T) + \boldsymbol{\eta} \\ &= \xi(\sqrt{\Phi_R} \otimes \mathbf{A}_{\varepsilon_o})\text{vec}(\mathbf{Z}\sqrt{\Phi_T}\mathbf{H}_{i.i.d.}^T) + \boldsymbol{\eta}. \end{aligned} \quad (5.36)$$

Note that although  $\Phi_T$  is assumed to be known, it cannot be separated from  $\mathbf{Z}$  and  $\mathbf{H}_{i.i.d.}$  because the correlation in transmit antennas can be translated into correlation of unknown data or vice versa. Since the noise is white and Gaussian, the  $ML_{NDA}$

estimator resumes to the minimization of

$$J_2(\mathbf{r}|\varepsilon, \mathbf{x}) = (\mathbf{r} - \check{\mathbf{A}}_\varepsilon \mathbf{x})^H (\mathbf{r} - \check{\mathbf{A}}_\varepsilon \mathbf{x}), \quad (5.37)$$

where  $\check{\mathbf{A}}_\varepsilon \triangleq \xi(\sqrt{\check{\Phi}_R} \otimes \mathbf{A}_\varepsilon)$ ,  $\varepsilon$  and  $\mathbf{x}$  are the trial values for  $\varepsilon_o$  and  $\text{vec}(\mathbf{Z}\sqrt{\check{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T)$ , respectively.

With the linear model of (5.36), the ML estimate for  $\text{vec}(\mathbf{Z}\sqrt{\check{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T)$  (when  $\varepsilon$  is fixed) is given by

$$\hat{\mathbf{x}} = (\check{\mathbf{A}}_\varepsilon^H \check{\mathbf{A}}_\varepsilon)^{-1} \check{\mathbf{A}}_\varepsilon^H \mathbf{r}. \quad (5.38)$$

Putting (5.38) into (5.37), after some straightforward calculations and dropping the irrelevant terms, the  $\text{ML}_{NDA}$  symbol timing estimator reduces to the maximization of the following likelihood function:

$$\Lambda_{NDA}(\varepsilon) = \mathbf{r}^H \check{\mathbf{A}}_\varepsilon (\check{\mathbf{A}}_\varepsilon^H \check{\mathbf{A}}_\varepsilon)^{-1} \check{\mathbf{A}}_\varepsilon^H \mathbf{r}. \quad (5.39)$$

It can be easily shown that

$$\check{\mathbf{A}}_\varepsilon (\check{\mathbf{A}}_\varepsilon^H \check{\mathbf{A}}_\varepsilon)^{-1} \check{\mathbf{A}}_\varepsilon^H = \mathbf{I}_M \otimes \mathbf{A}_\varepsilon (\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon)^{-1} \mathbf{A}_\varepsilon^H, \quad (5.40)$$

which gives

$$\Lambda_{NDA}(\varepsilon) = \sum_{j=1}^M \mathbf{r}_j^H \mathbf{A}_\varepsilon (\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon)^{-1} \mathbf{A}_\varepsilon^H \mathbf{r}_j. \quad (5.41)$$

The  $\text{ML}_{NDA}$  symbol timing estimation can be stated as

$$\hat{\varepsilon} = \arg \max_{\varepsilon} \Lambda_{NDA}(\varepsilon) \quad (5.42)$$

and can be implemented by the two-step approach as for the  $\text{ML}_{DA}$ .

Note that the implementation of the  $\text{ML}_{NDA}$  estimator does not require the knowledge of correlation among antennas. Note also that the likelihood function in (5.41) is the sum of individual likelihood functions for each receive antenna, just

as the case of training based likelihood function in (5.18). For each of the receive antenna, the likelihood function is the same as the likelihood function for SISO systems (see (2.16)). Furthermore, applying the low-complexity maximization technique introduced in Chapter II Section C to the likelihood function (5.41) and with the approximation  $\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \approx \mathbf{I}_{L_o+2L_g}$  for Nyquist zero-ISI pulse, it can be easily shown that the  $\text{ML}_{NDA}$  (5.42) reduces to the extension of squaring algorithm proposed in [61].

## 2. The CCRB and MCRB

For the model in (5.36), the CCRB for a specific  $\varepsilon_o$  is given by [10]

$$\text{CCRB}_{NDA}(\varepsilon_o) = \frac{\sigma^2}{2\text{tr}(\check{\mathbf{D}}_{\varepsilon_o}^H \mathbf{P}_{\check{\mathbf{A}}}^\perp \check{\mathbf{D}}_{\varepsilon_o} \boldsymbol{\Gamma}_{\mathbf{x}})}, \quad (5.43)$$

where

$$\check{\mathbf{D}}_\varepsilon \triangleq \frac{d\check{\mathbf{A}}_\varepsilon}{d\varepsilon} = \sqrt{\boldsymbol{\Phi}_R} \otimes \mathbf{D}_\varepsilon, \quad (5.44)$$

$$\mathbf{P}_{\check{\mathbf{A}}}^\perp \triangleq \mathbf{I}_{ML_oQ} - \check{\mathbf{A}}_{\varepsilon_o} (\check{\mathbf{A}}_{\varepsilon_o}^H \check{\mathbf{A}}_{\varepsilon_o})^{-1} \check{\mathbf{A}}_{\varepsilon_o}^H = \mathbf{I}_M \otimes \mathbf{P}_{\mathbf{A}}^\perp, \quad (5.45)$$

with  $\mathbf{P}_{\mathbf{A}}^\perp \triangleq \mathbf{I}_{L_oQ} - \mathbf{A}_{\varepsilon_o} (\mathbf{A}_{\varepsilon_o}^H \mathbf{A}_{\varepsilon_o})^{-1} \mathbf{A}_{\varepsilon_o}^H$ , and

$$\boldsymbol{\Gamma}_{\mathbf{x}} \triangleq \mathbb{E}[\text{vec}(\mathbf{Z}\sqrt{\boldsymbol{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T) \text{vec}(\mathbf{Z}\sqrt{\boldsymbol{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T)^H]. \quad (5.46)$$

It is shown in Appendix C that

$$\boldsymbol{\Gamma}_{\mathbf{x}} = \mathbf{I}_M \otimes \boldsymbol{\Psi}, \quad (5.47)$$

where  $\boldsymbol{\Psi}$  is a Hermitian and Toeplitz matrix with elements  $[\boldsymbol{\Psi}]_{ij} \triangleq \text{tr}(\boldsymbol{\Gamma}_{\mathbf{z}}(j-i)\boldsymbol{\Phi}_T)$  and  $\boldsymbol{\Gamma}_{\mathbf{z}}(j-i) \triangleq \mathbb{E}[(\mathbf{Z}^*)_{j,:}^T (\mathbf{Z})_{i,:}]$  is the average cross-correlation matrix of the symbols transmitted with time index difference  $j-i$ .

Substituting (5.44), (5.45) and (5.47) into (5.43), we obtain:

$$\begin{aligned} \text{CCRB}_{NDA}(\varepsilon_o) &= \frac{\sigma^2}{2\xi^2 \text{tr}((\sqrt{\Phi_R} \otimes \mathbf{D}_{\varepsilon_o})^H (\mathbf{I}_M \otimes \mathbf{P}_A^\perp) (\sqrt{\Phi_R} \otimes \mathbf{D}_{\varepsilon_o}) (\mathbf{I}_M \otimes \Psi))} \\ &= \frac{1}{2M \text{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^H \mathbf{P}_A^\perp \tilde{\mathbf{D}}_{\varepsilon_o} \Psi / N)} \left( \frac{E_s}{N_o} \right)^{-1}. \end{aligned} \quad (5.48)$$

Following the same calculations as for the  $\text{CCRB}_{NDA}$ , the  $\text{MCRB}_{NDA}$  is given by

$$\begin{aligned} \text{MCRB}_{NDA}(\varepsilon_o) &= \frac{\sigma^2}{2 \text{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o} \Gamma_x)} \\ &= \frac{1}{2M \text{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o} \Psi / N)} \left( \frac{E_s}{N_o} \right)^{-1}. \end{aligned} \quad (5.49)$$

Note that the average of  $\text{CCRB}_{NDA}$  and  $\text{MCRB}_{NDA}$  can be computed by numerical integration of (5.48) and (5.49), respectively. In the following, we consider two special cases.

*Special Case 1:* The data is spatially and temporally white (e.g., Vertical-Bell Labs Layered Space-Time (V-BLAST) system<sup>4</sup> [55]). In this case,  $\Gamma_z(j-i) = \mathbf{I}_N \delta_{ij}$ , implying that  $[\Psi]_{ij} = \delta_{ij} \text{tr}(\Phi_T) = N \delta_{ij}$ . Therefore, the corresponding  $\text{CCRB}_{NDA}$  and  $\text{MCRB}_{NDA}$  are

$$\text{CCRB}_{NDA}(\varepsilon_o) = \frac{1}{2M \text{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^H \mathbf{P}_A^\perp \tilde{\mathbf{D}}_{\varepsilon_o})} \left( \frac{E_s}{N_o} \right)^{-1} \quad (5.50)$$

and

$$\text{MCRB}_{NDA}(\varepsilon_o) = \frac{1}{2M \text{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o})} \left( \frac{E_s}{N_o} \right)^{-1}, \quad (5.51)$$

respectively. Note that in this case, the  $\text{CCRB}_{NDA}$  and  $\text{MCRB}_{NDA}$  do not depend on the number of transmit antennas and the correlations among antennas.

*Special Case 2:* Space-Time Block Code (STBC) system. In general, a block of

---

<sup>4</sup>In its initial development, V-BLAST system does not employ any temporal error control code. Although temporal error control code may be applied in V-BLAST system, we assume the data is temporally white since from the point of view of the symbol synchronizer, the data appears to be uncorrelated.

space-time block coded symbols can be represented by a  $s \times N$  matrix [49]

$$\mathcal{G} = \sum_{k=1}^{rs} \Re\{b_k\} \mathbf{X}_k + j \sum_{k=1}^{rs} \Im\{b_k\} \mathbf{Y}_k, \quad (5.52)$$

where  $r$  is the rate of the STBC,  $s$  is the length of the STBC,  $b_k$ 's are the i.i.d., complex valued symbols to be encoded and  $\mathbf{X}_k, \mathbf{Y}_k$  are the fixed, real-valued elementary code matrices. Without loss of generality, we assume  $|b_k| = 1$ . It is proved in Appendix C that for the STBC system,

$$\begin{aligned} & \mathbf{\Gamma}_z(j-i) \\ &= \begin{cases} \mathbf{0}_N & \text{for } |j-i| \geq s \\ \frac{1}{2s} \sum_{n=1}^{s-\ell} (\sum_{k=1}^{rs} [\mathbf{X}_k]_{n+\ell,:}^T [\mathbf{X}_k]_{n,:} + \sum_{k=1}^{rs} [\mathbf{Y}_k]_{n+\ell,:}^T [\mathbf{Y}_k]_{n,:}) & \text{for } |j-i| = \ell, \ell < s. \end{cases} \end{aligned} \quad (5.53)$$

For example, let us consider the half-rate orthogonal space-time block code with four transmit antennas [48], in which case  $N = 4$ ,  $s = 8$ ,  $r = 1/2$  and the matrix  $\mathcal{G}$  given by

$$\mathcal{G} = \begin{pmatrix} b_1 & b_2 & b_3 & b_4 \\ -b_2 & b_1 & -b_4 & b_3 \\ -b_3 & b_4 & b_1 & -b_2 \\ -b_4 & -b_3 & b_2 & b_1 \\ b_1^* & b_2^* & b_3^* & b_4^* \\ -b_2^* & b_1^* & -b_4^* & b_3^* \\ -b_3^* & b_4^* & b_1^* & -b_2^* \\ -b_4^* & -b_3^* & b_2^* & b_1^* \end{pmatrix}. \quad (5.54)$$

Decomposing  $\mathcal{G}$  in terms of  $\mathbf{X}_k$  and  $\mathbf{Y}_k$  and using (5.53), it is found that

$$\mathbf{\Gamma}_z(j-i) = \begin{cases} \mathbf{I}_4 & \text{for } i = j, \\ \frac{1}{4} \begin{bmatrix} 0 & 2 & 0 & 1 \\ -2 & 0 & 1 & 0 \\ 0 & -1 & 0 & 2 \\ -1 & 0 & -2 & 0 \end{bmatrix} & \text{for } |j-i| = 1, \\ \frac{1}{4} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix} & \text{for } |j-i| = 3, \\ \mathbf{0}_4 & \text{otherwise.} \end{cases} \quad (5.55)$$

Then,  $\mathbf{\Psi}$  can be computed according to  $[\mathbf{\Psi}]_{ij} = \text{tr}(\mathbf{\Gamma}_z(j-i)\mathbf{\Phi}_T)$  and the  $\text{CCRB}_{NDA}$  and  $\text{MCRB}_{NDA}$  are given by (5.48) and (5.49), respectively.

### E. Simulation Results and Discussions

In this section, the mean square error (MSE) performances of the proposed symbol timing estimators  $\text{ML}_{DA}$  (5.19) and  $\text{ML}_{NDA}$  (5.42) are assessed by Monte Carlo simulations. In all the simulations,  $L_o = 32$ ,  $L_g = 4$  (i.e., the total length of training data is 40),  $Q = 2$ ,  $K = 16$ ,  $\varepsilon_o$  is uniformly distributed in the range  $[0, 1)$  and  $g(t)$  is a RRC filter with roll-off factor  $\alpha = 0.3$ . Each point is obtained by averaging  $10^4$  Monte-Carlo simulation runs. For the DA case, the optimal orthogonal sequences  $\mathbf{Z}_{opt}$  derived in Section 3 are used as training data. For the NDA case, the data format is QPSK.

#### 1. Effects of $N$ and $M$

In this section, the effects of the number of transmit and receive antennas are examined. First, let assume  $\mathbf{\Phi}_T = \mathbf{I}_N$  and  $\mathbf{\Phi}_R = \mathbf{I}_M$  for the moment. Furthermore, it is assumed there is no space-time coding in the NDA case. The effect of antenna correlation and space-time coding will be examined later. The effect of the number of transmit antennas  $N$  is shown in Figs. 26 and 27 for the DA and NDA cases,

respectively, with  $M = 4$ . From both figures, it can be seen that different numbers of transmit antennas result in similar estimation accuracies. Therefore, the MSEs are approximately independent of  $N$  for both  $\text{ML}_{DA}$  and  $\text{ML}_{NDA}$ . Next, the effect of the number of receive antennas  $M$  is shown in Figs. 28 and 29 for DA and NDA case, respectively, with  $N = 4$ . It is clear that increasing  $M$  leads to considerable MSE improvements. Since from (5.30) and (5.48), the  $\text{CCRB}_{DA}$  and  $\text{CCRB}_{NDA}$  are inversely proportional to  $M$  and from Figs. 28 and 29, the performances of  $\text{ML}_{DA}$  and  $\text{ML}_{NDA}$  are very close to their corresponding CCRBs, it can be concluded that the MSEs of  $\text{ML}_{DA}$  and  $\text{ML}_{NDA}$  estimators are approximately inversely proportional to  $M$ .

It is reasonable to have improved performances when the number of receive antennas increases since more receive antennas provides diversity gain. It is tempted to argue that using more transmit antennas should also improve the performances of symbol timing estimation since from the experience of STBC [46], [48], more transmit antennas also provides diversity gain. However, notice that the diversity gain of STBC does not come automatically by just increasing the number of transmit antennas. In STBC, the observation length for demodulating a symbol has to be increased with the number of transmit antennas. For symbol timing estimation, irrespective of the number of transmit antennas, the total transmit power and the observation length are kept constant, it is not unreasonable to have MSE performances approximately independent of  $N$ . For multiple receive antennas, although the observation length (for each receive antenna) is kept constant, the observations from different receive antennas are independent (similar to the situation of maximal-ratio receive combining scheme). These independent observations increase the *effective* observation length and performance is improved due to the longer effective observation.

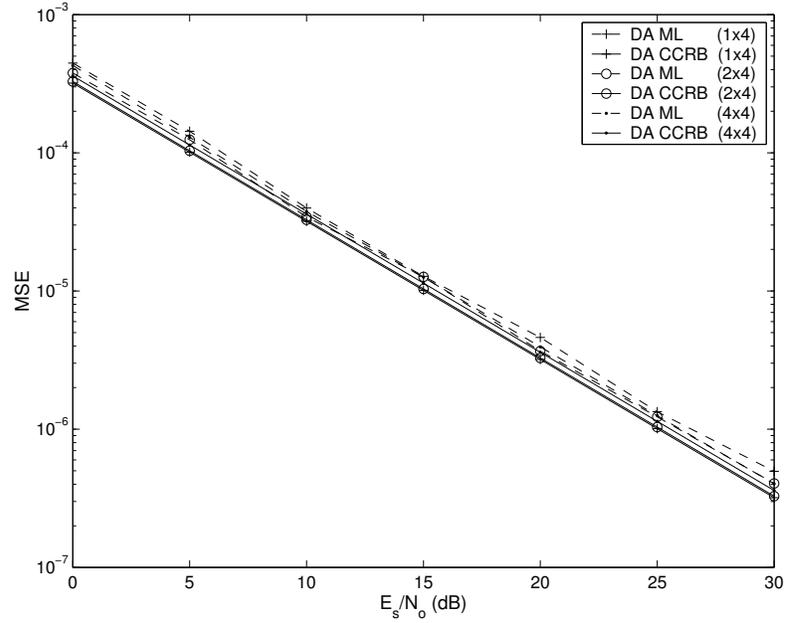


Fig. 26. MSEs of the  $ML_{DA}$  estimator and the corresponding CCRBs with different number of transmit antennas ( $\Phi_T = \mathbf{I}_N$ ,  $\Phi_R = \mathbf{I}_M$ ,  $\mathbf{Z} = \mathbf{Z}_{opt}$ ).

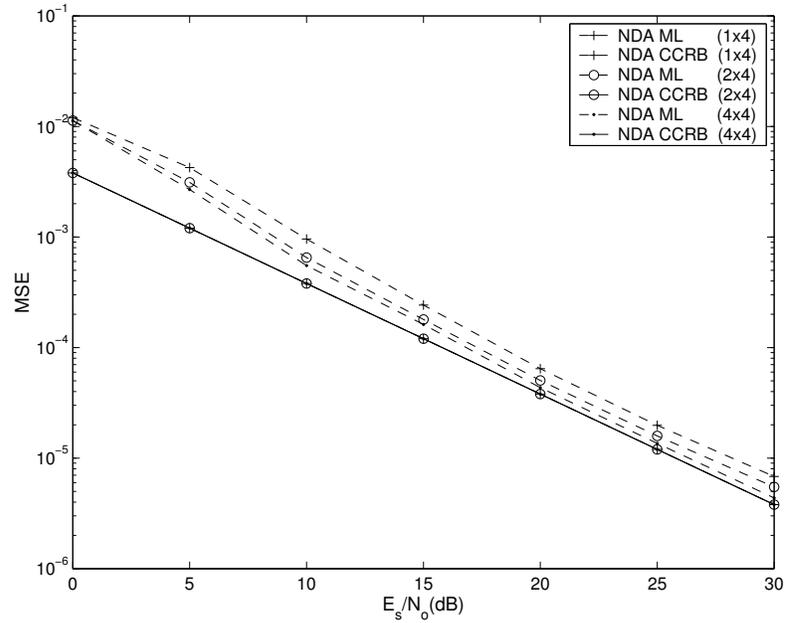


Fig. 27. MSEs of the  $ML_{NDA}$  estimator and the corresponding CCRBs with different number of transmit antennas ( $\Phi_T = \mathbf{I}_N$ ,  $\Phi_R = \mathbf{I}_M$  and the data transmitted is spatially and temporally white).

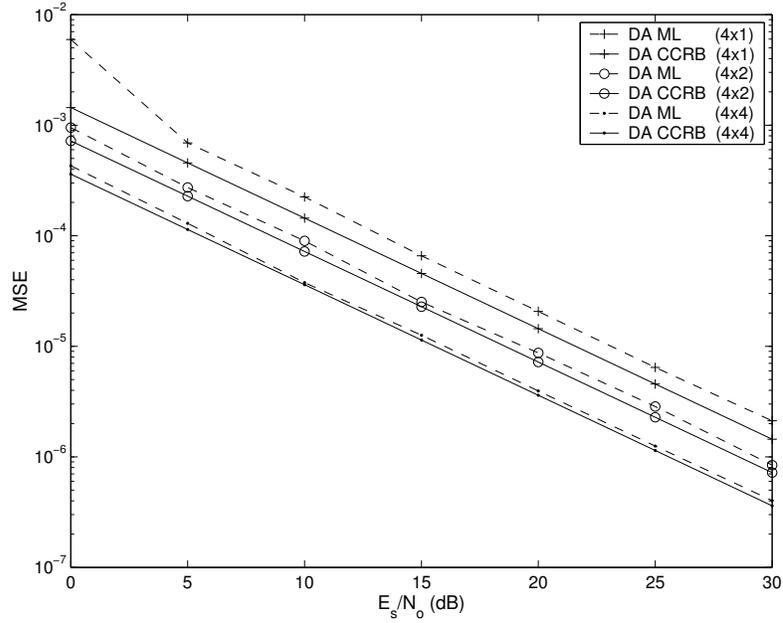


Fig. 28. MSEs of the  $ML_{DA}$  estimator and the corresponding CCRBs with different number of receive antennas ( $\Phi_T = \mathbf{I}_N$ ,  $\Phi_R = \mathbf{I}_M$ ,  $\mathbf{Z} = \mathbf{Z}_{opt}$ ).

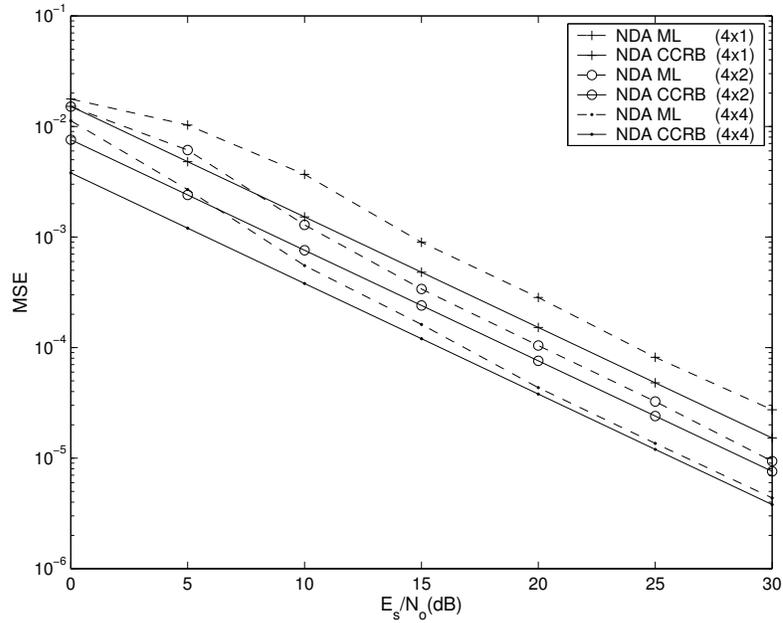


Fig. 29. MSEs of the  $ML_{NDA}$  estimator and the corresponding CCRBs with different number of receive antennas ( $\Phi_T = \mathbf{I}_N$ ,  $\Phi_R = \mathbf{I}_M$  and the data transmitted is spatially and temporally white).

## 2. Effects of Correlation Among Antennas

Figs. 30 and 31 show the MSE performances of  $ML_{DA}$  and  $ML_{NDA}$  of a  $4 \times 4$  system under the effect of correlated fading among antennas. The measured correlation matrices from Nokia [53] are used in simulations:

$$\mathbf{\Phi}_T = \begin{bmatrix} 1 & 0.4154 & 0.2057 & 0.1997 \\ 0.4154 & 1 & 0.3336 & 0.3453 \\ 0.2057 & 0.3336 & 1 & 0.5226 \\ 0.1997 & 0.3453 & 0.5226 & 1 \end{bmatrix}, \mathbf{\Phi}_R = \begin{bmatrix} 1 & 0.3644 & 0.0685 & 0.3566 \\ 0.3644 & 1 & 0.3245 & 0.1848 \\ 0.0685 & 0.3245 & 1 & 0.3093 \\ 0.3566 & 0.1848 & 0.3093 & 1 \end{bmatrix}. \quad (5.56)$$

Three cases are considered in Fig. 30 for the DA case. The first case assumes no correlation among antenna arrays, and serves as a reference and is shown by the ‘+’ markers. The second one, which is shown by ‘o’ markers, assumes that correlations exist among antennas and perfect knowledge of  $\mathbf{\Phi}_T$  is available for designing optimal training sequences. The last case, denoted by the ‘.’ markers, assumes that correlations exist among antennas but no knowledge of correlations is assumed when designing the training sequences. It can be seen that the fading correlations among antennas do not change the MSE performance of the  $ML_{DA}$  estimator or the  $CCRB_{DA}$ . Furthermore, surprisingly, the knowledge of  $\mathbf{\Phi}_T$  for designing optimal training sequences is not important as the results show that training sequences assuming no correlation perform equally well in the presence of correlation among antennas.

For the NDA case (Fig. 31), three cases are considered, too. The first one is no space-time coding and no fading correlation, which is shown using ‘+’ markers. The second one is no space-time coding but with fading correlation, which is shown by ‘o’ markers. The final one is that the data is encoded with the half rate space-time block code (5.54) and with correlated fading, which is shown by ‘.’ markers. It can be seen

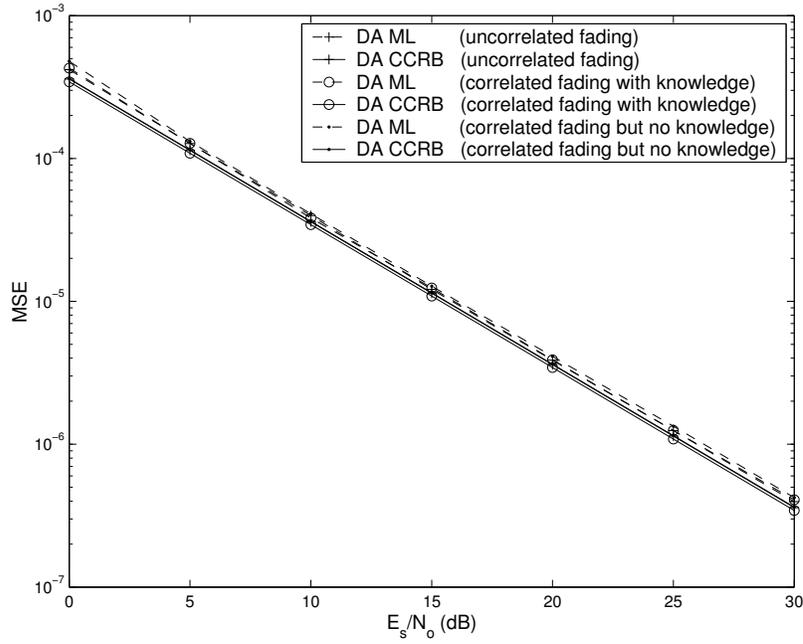


Fig. 30. MSEs of the  $ML_{DA}$  estimator and the corresponding CCRBs with and without fading correlation between antennas for a  $4 \times 4$  system.

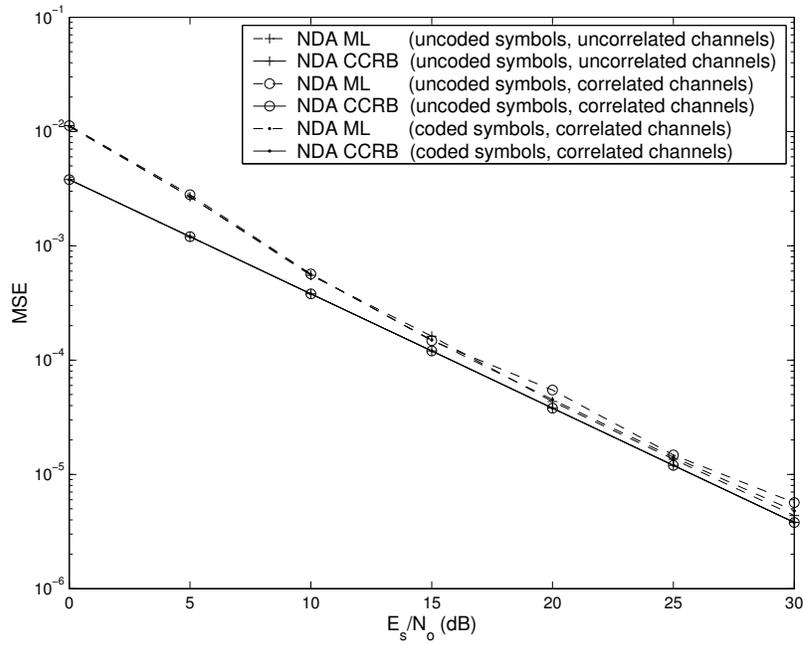


Fig. 31. MSEs of the  $ML_{NDA}$  estimator and the corresponding CCRBs with and without fading correlation between antennas for a  $4 \times 4$  system.

that the presence of correlated fading and space-time coding do not affect the MSE performances of the  $ML_{NDA}$  estimator.

In order to investigate the performance of  $ML_{DA}$  and  $ML_{NDA}$  estimators under different degree of fading correlation, we employ the following single parameter correlation model:

$$[\Phi_T]_{ij} = [\Phi_R]_{ij} = \rho^{|i-j|}, \quad (5.57)$$

where  $\rho \in [0, 1)$  is the correlation coefficient between adjacent antennas (note that  $\rho = 0$  means no correlation). Fig. 32 shows the MSEs of the  $ML_{DA}$  estimator against  $\rho$  for  $E_s/N_o=10\text{dB}$ ,  $20\text{dB}$  and  $30\text{dB}$  in a  $4 \times 4$  system. Two cases are considered. The first one assumes perfect knowledge of correlation for designing training sequences and the second one assumes no correlation when designing training sequences. It can be seen that for  $\rho \leq 0.5$ , the performance degradation due to antenna correlation is extremely small. Only when  $\rho > 0.5$ , the performance start to degrade, but with limited degree. Also, designing training sequences without knowledge of correlation results only in a slight degradation with respect to the case which assumes perfect knowledge of correlation, and this only happens when  $\rho > 0.5$ . This property facilitates the practical implementation of the proposed scheme since in practice, the correlation matrix may not be perfectly known. This also explains the results in Fig. 30 that the  $ML_{DA}$  estimator does not suffer any loss of performance since the largest measured correlation coefficient between adjacent antennas in (5.56) is about 0.5. Fig. 33 shows the MSEs of the  $ML_{NDA}$  estimator against  $\rho$  for  $E_s/N_o=10\text{dB}$ ,  $20\text{dB}$  and  $30\text{dB}$  in a  $4 \times 4$  system. Two cases are simulated. The first case is no space-time coding, while the second case is encoded by (5.54). It can be seen that, basically, the space-time coding considered in this example does not have any effect on the MSE performances of the  $ML_{NDA}$  with respect to the no coding case. Furthermore, the degradation due

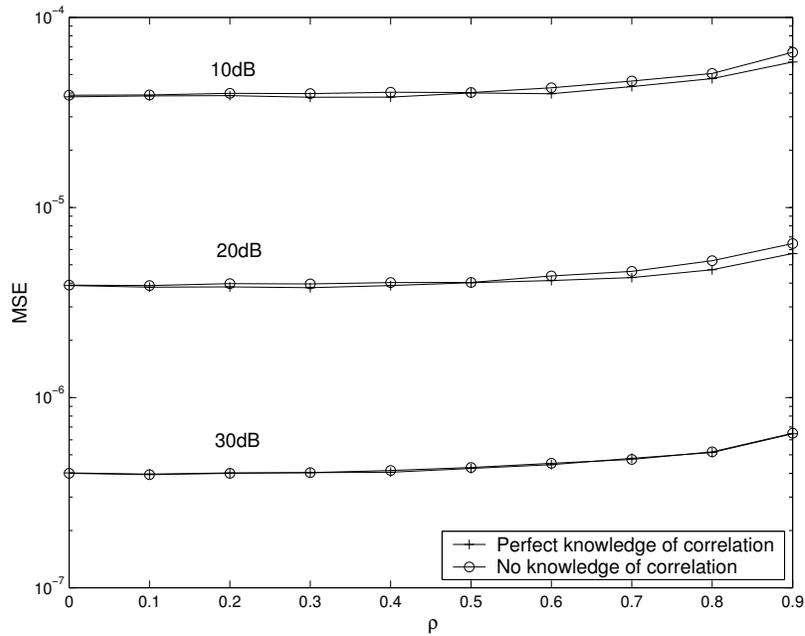


Fig. 32. MSEs of the  $ML_{DA}$  estimator against the correlation coefficient  $\rho$  between adjacent antennas for  $E_s/N_o=10\text{dB}$ ,  $20\text{dB}$  and  $30\text{dB}$  in a  $4 \times 4$  system.

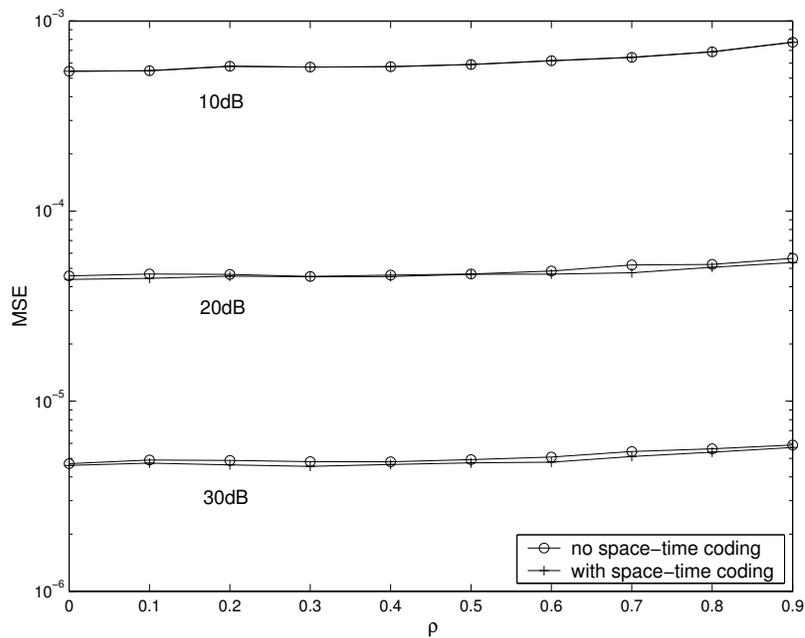


Fig. 33. MSEs of the  $ML_{NDA}$  estimator against the correlation coefficient between adjacent antennas  $\rho$  for  $E_s/N_o=10\text{dB}$ ,  $20\text{dB}$  and  $30\text{dB}$  in a  $4 \times 4$  system.

to extreme antenna correlations is very small.

The small dependence of the MSEs on correlation between antennas is due to the fact that, in this study, the nuisance parameters (i.e.,  $\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)$  for DA case and  $\text{vec}(\mathbf{Z}\sqrt{\Phi_T}\mathbf{H}_{\text{i.i.d.}}^T)$  for NDA case) are treated as deterministic unknown and are being jointly estimated together with  $\varepsilon_o$ . The correlation between antennas, can always be lumped into the nuisance parameters. Since this action does not change the dimension of the nuisance parameters and there is no constraint on the value of the nuisance parameters, the effect of correlation between antennas on the MSE of  $\hat{\varepsilon}$  would be very small.

### 3. Comparison of DA and NDA Estimators

Here, we compare the performance of the  $\text{ML}_{DA}$  and  $\text{ML}_{NDA}$  estimators with their corresponding CCRBs and MCRBs for a  $4 \times 4$  system. For simplicity, it is assumed that there is no correlation among antennas and there is no space-time coding for NDA case (since the effects of these are small as shown earlier). Fig. 34 shows the results. Note that from Fig. 34, the MSE performances of  $\text{ML}_{DA}$  and  $\text{ML}_{NDA}$  estimators are very close to their corresponding CCRBs. This means that  $\text{ML}_{DA}$  and  $\text{ML}_{NDA}$  are efficient estimators conditioned that the nuisance parameters are being jointly estimated together with the unknown timing delay. Also, note that the performance of  $\text{ML}_{DA}$  estimator is very close to the  $\text{MCRB}_{DA}$ , which implies that  $\text{ML}_{DA}$  is almost the best possible estimator under the problem at hand, regardless of how we deal with the nuisance parameters. For the NDA case, unfortunately, although the performance of  $\text{ML}_{NDA}$  estimator reaches the corresponding  $\text{CCRB}_{NDA}$ , the  $\text{CCRB}_{NDA}$  is quite far away from the  $\text{MCRB}_{NDA}$ . Notice that, according to [10], CCRB is a valid bound only for estimators that rely on quadratic nonlinearity, there is a possibility that some other NDA estimators employing higher order ( $>2$ ) nonlinearities would have

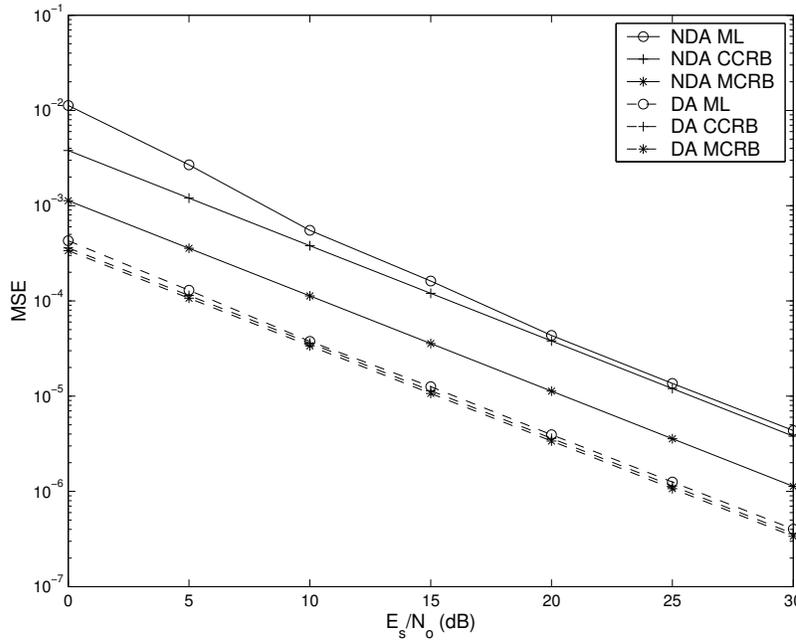


Fig. 34. Comparison of MSEs of the  $ML_{NDA}$  and  $ML_{DA}$  and their corresponding CCRBs and MCRBs for a  $4 \times 4$  system.

performances closer to the MCRB. This is subject to further investigations.

Finally, as expected,  $ML_{DA}$  estimator performs much better than the  $ML_{NDA}$  estimator. However, this comes with a price. The  $ML_{DA}$  estimator requires training sequences, resulting in lower transmission efficiency. Moreover, the estimation has to be performed at specific times when the training data is available, while  $ML_{NDA}$  can be performed at any time during transmission. This also means that, for the DA case, there is a need to synchronize the training sequences before timing estimation. This requires extra implementation complexity. In addition, degradation may occur if the positions of the training sequences are mislocated. Furthermore, the computation of the DA likelihood function (5.18) is more complicated than that of the NDA likelihood function (5.41). Therefore,  $ML_{DA}$  and  $ML_{NDA}$  provide a performance, transmission efficiency and complexity tradeoff for symbol timing estimation in MIMO channels.

## F. Conclusions

The data aided (DA) and non-data aided (NDA) maximum likelihood (ML) symbol timing estimators, their corresponding conditional CRB (CCRB) and modified CRB (MCRB) for MIMO correlated flat-fading channels have been derived in this chapter. For the DA case, the optimal orthogonal training sequences have also been derived. It was shown that the approximated ML algorithm in [11] and that in Chapter IV is just a special case of the DA ML algorithm; while the extended squaring algorithm in [61] is just a special case of the NDA ML estimator. For the optimal orthogonal training sequences, it was found that they resemble Walsh sequences but with modified envelopes. Simulation results under different operating conditions (e.g., number of antennas and correlation between antennas) were given to assess the performances of the DA and NDA ML estimators and compare them with the corresponding CCRBs and MCRBs. It was found that i) the MSE of the DA ML estimator is close to the CCRB and MCRB, meaning that the DA ML estimator is almost the best estimator (in terms of MSE performance) for the problem under consideration; ii) the MSE of the NDA ML estimator is close to the CCRB but not MCRB, meaning that NDA ML estimator is an efficient estimator conditioned that the nuisance parameters are being jointly estimated, but there *might* exist other NDA estimators with better performances; iii) the MSEs of both DA and NDA ML estimators are approximately independent of the number of transmit antennas and are inversely proportional to the number of receive antennas; iv) correlation between antennas has little impact on the MSEs of DA and NDA ML estimators unless the correlation coefficient between adjacent antennas is larger than 0.5, in which case a small degradation occurs, and vi) DA ML performs better than NDA ML estimator at the cost of lower transmission efficiency and higher implementation complexity.

## CHAPTER VI

MAXIMUM-LIKELIHOOD SYMBOL SYNCHRONIZATION FOR IEEE 802.11A  
WLANS IN UNKNOWN FREQUENCY-SELECTIVE FADING CHANNELS

## A. Introduction

IEEE 802.11a wireless local area networks (WLANs), which support high-speed data transmissions up to 54Mbps [68], employ burst-mode transmission and orthogonal frequency division multiplexing (OFDM) as the transmission technique. Although OFDM is well known for its ability to combat the intersymbol interference (ISI) introduced by multipath channels [69], [70], incorrect positioning of the FFT window within an OFDM symbol reintroduces ISI during data demodulation, causing serious performance degradation [12], [13]. Symbol synchronization is therefore one of important tasks performed at receivers in IEEE 802.11a WLANs.

A number of methods for OFDM symbol synchronization have been proposed in the literature. Methods that exploit the periodic structure of cyclic prefixes in OFDM symbols have been proposed in [13]-[15]. Algorithms based on the use of repeated preambles have been reported in [16]-[21]. In [12] and [22], additional pilot subcarriers are used to further improve the estimation accuracy after coarse timing synchronization is established by correlation-based methods. Although the techniques of [12]-[22] (which were originally developed for general OFDM systems) may be applied to IEEE 802.11a WLANs, a higher synchronization accuracy can be obtained by using optimized algorithms that take advantage of the known preamble structure located at the beginning of a data packet.

Recently, symbol synchronization techniques that are specifically designed for IEEE 802.11a WLANs have been reported in [23] and [24]. In [23], the received signal

is correlated with a known training-symbol sequence and the absence of the expected correlation peak is detected. Despite the advantage that a simple correlator can be easily implemented at the receiver, its performance is poor in dispersive channels [23], indicating that more sophisticated synchronization algorithms are required. In [24], the generalized Akaike information criterion (GAIC) is used to jointly estimate the channel and establish timing synchronization. Although the reported performance is good, its complexity is extremely high.

In this chapter, we develop a maximum-likelihood (ML) symbol synchronizer for IEEE 802.11a WLANs on frequency-selective fading channels. A realistic channel, which includes the effects of filtering and sampling time offset in addition to the physical channel with random path delays, is considered. Furthermore, the loss in system performance due to synchronization error is used as the performance criterion [13], [30], as opposed to the requirement that the estimated symbol timing has to be within certain limits with respect to a fixed reference point. The proposed algorithm is compared with four existing symbol synchronization algorithms, one of which specifically designed for IEEE 802.11a WLANs [24] and three other algorithms for general OFDM frame synchronization [14], [20], [21]. Simulation results indicate that in general, joint estimation of symbol position and channel (as is the case with the proposed algorithm and the algorithm based on GAIC [24]) gives better performances than the correlation based algorithms [14], [20], [21]. When compared to the GAIC algorithm [24], the proposed algorithm exhibits comparable performances, but the complexity of the proposed algorithm is much smaller than that of GAIC algorithm due to the smaller observation length.

We mention that there are also works on general frame synchronization for packet-based transmission over frequency-selective channels (not limited to OFDM systems) [71]-[73]. It will be shown later that part of the proposed synchronization

algorithm in this chapter reduces to the same detector proposed in [71]-[73] under certain conditions. Note also that the seemingly related synchronization technique for pilot-symbol-assisted modulation (PSAM) [74] is not applicable to the problem under consideration since the pilot-symbol insertion mechanism in PSAM is totally different from the preamble structure of a data packet in IEEE 802.11a WLAN.

This chapter is organized as follows. Section B describes the received-signal model. The symbol synchronization performance criterion is discussed in Section C. Section D derives the ML synchronization algorithms for the IEEE 802.11a WLANs. Simulation results on the synchronization performances and comparison with other algorithms are presented in Section E. Finally, conclusions are drawn in Section F.

## B. Received-Signal Model

### 1. Signal and Channel Models

Fig. 35 depicts the packet structure used in IEEE 802.11a WLANs. In each packet, the data carrying OFDM symbols are preceded by a preamble, which is used for start of packet detection, automatic gain control, symbol timing and frequency synchronization, and channel estimation. The preamble itself consists of two parts. The first part comprises 10 short training symbols,  $b(t)$ , each of length  $T_b = 800\text{ns}$ . In the second part, a cyclic prefix,  $g(t)$ , of length  $T_g = 1.6\mu\text{s}$  is followed by two long training symbols,  $c(t)$ , each of length  $T_c = 3.2\mu\text{s}$ . The baseband-equivalent model of the preamble is given by [68]

$$s(t) = \sqrt{2P} \left\{ \sum_{i=0}^9 b(t - iT_b) + g(t - 10T_b) + \sum_{i=0}^1 c(t - 10T_b - T_g - iT_c) \right\} \quad (6.1)$$

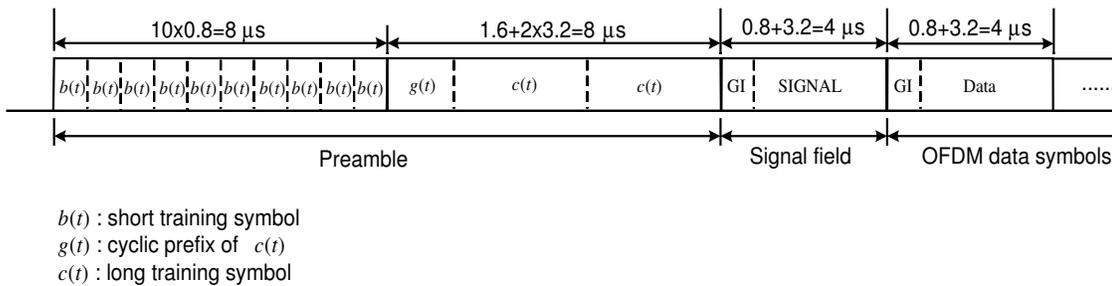


Fig. 35. Packet structure for IEEE 802.11a WLANs.

where  $P$  is the RF signal power,

$$b(t) \triangleq \Pi\left(\frac{t}{T_b}\right) \times \frac{1}{\sqrt{52}} \sum_{k=-26}^{26} S_k e^{j2\pi k \Delta_f t} \quad (6.2)$$

$$g(t) \triangleq \Pi\left(\frac{t}{T_g}\right) \times \frac{1}{\sqrt{52}} \sum_{k=-26}^{26} L_k e^{j2\pi k \Delta_f (t-T_g)} \quad (6.3)$$

$$c(t) \triangleq \Pi\left(\frac{t}{T_c}\right) \times \frac{1}{\sqrt{52}} \sum_{k=-26}^{26} L_k e^{j2\pi k \Delta_f t} . \quad (6.4)$$

In (6.2)-(6.4),  $\Pi(x)$  is a rectangular function giving a value of 1 when  $0 \leq x < 1$  and 0 otherwise,  $\Delta_f = 312.5\text{kHz}$  is the subcarrier separation,  $S_{-26:26}$  and  $L_{-26:26}$  are two training-symbol sequences given by (6) and (8) of [68], respectively. The signal model given by (6.1) is slightly different from the one specified in the standard [68] in that a raised-cosine window should be used to mask the two parts of the preamble for reducing the spectral side-lobes of transmitted signals. This windowing function is not included in the model considered here in order to simplify derivation of the synchronization algorithm. The raised-cosine window was included in the simulation model during the assessment of synchronization performances.

The signal  $s(t)$  is passed through the transmission filter  $f_T(t)$ , up-converted to high frequency and transmitted through a multipath frequency-selective fading channel. At the receiver, the signal is first passed through the RF filter and then down-

converted into baseband signal, where additional filtering occurs. Assuming the channel is static over the duration of a packet, the complex envelope of the received and filtered signal is given by

$$r(t) = e^{j2\pi vt} \left\{ s(t) \star f_T(t) \star \sum_{n=0}^{L_h-1} \gamma_n \delta(t - \tau_n) \star f_R(t) \right\} + w(t) , \quad (6.5)$$

where  $v$  is the carrier frequency offset,  $\star$  denotes continuous time convolution,  $\gamma_n$  is the complex valued channel coefficients for the  $n^{\text{th}}$  path with  $\tau_n$  as the path delay,  $L_h$  is the total number of physical paths of the multipath channel,  $f_R(t)$  is the low-pass equivalent response of the combined RF and baseband filter, and  $w(t)$  is the filtered complex-Gaussian noise. It is assumed that the channel gain is unity (i.e.,  $\mathbb{E}[\sum_{n=0}^{L_h-1} |\gamma_n|^2] = 1$ , where  $\mathbb{E}[\cdot]$  stands for expectation). Furthermore, without loss of generality, it is assumed that  $\tau_0 = 0$  since the delay of first path can be translated to a delay in sampling. Expressing the convolution in the form of integral, the received signal can be rewritten as

$$r(t) = e^{j2\pi vt} \int_{-\infty}^{\infty} s(t-u) \left[ \sum_{n=0}^{L_h-1} \gamma_n f(u - \tau_n) \right] du + w(t) , \quad (6.6)$$

where  $f(t) \triangleq f_T(t) \star f_R(t)$ .

Now, the received signal is sampled at  $t = kT_{sam} + \varepsilon_o T_{sam}$ , where  $1/T_{sam} = 20\text{MHz}$ , which is the suggested sampling rate in the standard [68], and  $\varepsilon_o \in [0, 1)$  is the unknown time offset induced by the combination of the delay of the first path of the channel and the sampling phase offset. It follows that the sampled signal is given by

$$\begin{aligned} r_k &= e^{j2\pi v(kT_{sam} + \varepsilon_o T_{sam})} \int_{-\infty}^{\infty} s(kT_{sam} + \varepsilon_o T_{sam} - u) \left[ \sum_{n=0}^{L_h-1} \gamma_n f(u - \tau_n) \right] du + w_k \\ &= e^{j2\pi v(kT_{sam} + \varepsilon_o T_{sam})} \frac{1}{T_{sam}} \int_{-\infty}^{\infty} s(kT_{sam} - u') h(u') du' + w_k , \end{aligned} \quad (6.7)$$

where  $r_k \triangleq r(kT_{sam} + \varepsilon_o T_{sam})$ ,  $w_k \triangleq w(kT_{sam} + \varepsilon_o T_{sam})$  and  $h(t)$  is the equivalent channel which includes the effects of the transmission filter, physical channel, RF and baseband filtering at the receiver, the timing delay induced by the delay of the first path of the channel and the sampling phase offset, and is defined as

$$h(t) \triangleq T_{sam} \sum_{n=0}^{L_h-1} \gamma_n f(t - \tau_n + \varepsilon_o T_{sam}) . \quad (6.8)$$

Notice that the bandwidth (one-sided baseband) of  $s(t)$  is  $BW_s = (26 + 1)\Delta_f \approx 8.44\text{MHz}$  [68] (see also equation (6.2)-(6.4)), meaning that  $BW_s < 1/2T_{sam} = 10\text{MHz}$ <sup>1</sup>. According to [75], if the bandwidth of the equivalent channel  $h(t)$  (which is equal to the bandwidth of  $f(t)$ ) satisfies<sup>2</sup>

$$BW_h < 1/T_{sam} - BW_s , \quad (6.9)$$

then by the equivalence of digital and analog filtering for band-limited signals, the sampled received signal can be expressed as

$$r_k = e^{j2\pi v(kT_{sam} + \varepsilon_o T_{sam})} \sum_{i=-\infty}^{\infty} s(kT_{sam} - iT_{sam}) h(iT_{sam}) + w_k . \quad (6.10)$$

The meaning of (6.9) is that the filter  $f(t)$  can be designed such that its bandwidth is larger than  $1/2T_{sam}$  as long as the aliasing caused by sampling lies outside the passband of signal  $s(t)$ . A simple example of  $f(t)$  that makes (6.9) satisfied is the

---

<sup>1</sup>Strictly speaking, most of the signals in communications are not bandlimited, however, notice that the frequency separation between two OFDM data channels is only 20MHz [68]; taking into account of the guard bands, we can treat the one-sided signal bandwidth to be strictly smaller than 10MHz (or at least the effect of signal outside the ‘main’ bandwidth is so small that we can ignore it).

<sup>2</sup>For practical filters, they can always be designed to have attenuation at stop band as high as possible (with the cost of complexity), therefore, we can also treat of  $h(t)$  bandlimited.

raised cosine filter

$$f_{rc}(t) = \frac{1}{T_{sam}} \cdot \frac{\sin(\pi t/T_{sam}) \cos(\pi \alpha t/T_{sam})}{\pi t/T_{sam} \sqrt{1 - 4\alpha^2 t^2/T_{sam}^2}}, \quad (6.11)$$

with  $\alpha < 0.156$  since it is required that  $BW_h = (1 + \alpha)/2T_{sam} < 1/T_{sam} - BW_s$ . The scaling factor  $1/T_{sam}$  is included in (6.11) such that the frequency response  $F_{rc}(\omega) = 1$  for  $|\omega/2\pi| < (1 - \alpha)/2T_{sam}$ . Without loss of generality, we employ both  $f_T(t)$  and  $f_R(t)$  as square root raised cosine filters such that  $f(t) = f_T(t) \star f_R(t)$  is given by (6.11) with roll-off factor  $\alpha = 0.1$  for the rest of the chapter. Because  $f_R(t)$  is a square root cosine filter, the filtered noise samples are uncorrelated with variance  $\sigma_w^2 = \mathbb{E}[|w(t)|^2]$ .

*Remark 1.* For the special case if 1) the path delays are sample spaced (i.e.,  $\tau_n = nT_{sam}$ ), 2) the timing delay  $\varepsilon_o = 0$  and 3)  $f(t)$  is a raised cosine pulse with  $\alpha < 0.156$ , then

$$\begin{aligned} h(iT_{sam}) &= T_{sam} \sum_{n=0}^{L_h-1} \gamma_n f_{rc}(iT_{sam} - nT_{sam}) \\ &= \sum_{n=0}^{L_h-1} \gamma_n \delta(i - n), \end{aligned} \quad (6.12)$$

since the values of the raised cosine pulse are zero at integer multiples of  $T_{sam}$ . Therefore, (6.10) reduces to

$$r_k = e^{j2\pi v(kT_{sam})} \sum_{n=0}^{L_h-1} \gamma_n s(kT_{sam} - nT_{sam}) + w_k, \quad (6.13)$$

which is the system model used in [24], where the physical channel is represented by the commonly used tapped delay line model with equal tap spacing [76, p.795]. Therefore, the channel model considered in this chapter is more general than that in [24].

*Remark 2.* Note that since there are ten identical short training symbols trans-

mitted at the beginning of the preamble, the frequency offset  $v$  can be easily estimated by the ML frequency estimator [77] and may be compensated before frame synchronization. To simplify the development of the synchronization algorithms,  $v = 0$  is assumed in the following derivations. Performance of the developed algorithms in the presence of residual frequency offset after compensation will be assessed by simulations.

## 2. Matrix Algebraic Formulations

From (6.10), it is apparent that the received samples depend on  $h(iT_{sam})$  for  $-\infty \leq i \leq \infty$ . However, in practice,  $h(iT_{sam})$  will have significant values only for a finite range of  $i$  since 1) the path delays occur in a finite interval and 2) the value of  $f(t)$  becomes very small when  $|t|$  is large. An example of  $|h(iT_{sam})|^2$  is shown in Fig. 36 for  $L_h = 6$ , the first tap of the physical channel has zero delay, other five taps have delay uniformly distributed over the interval  $0 - 300$ ns,  $\gamma_n$  are independent and identically distributed (i.i.d.), zero-mean, complex Gaussian random variables with variances following the multipath intensity profile  $\phi(\tau) \sim e^{-\tau/\tau_{rms}}$ , where  $\tau_{rms} = 100$ ns,  $\varepsilon_o$  is a random variable uniformly distributed in  $[0, 1)$  and  $f(t)$  is given by (6.11) with  $\alpha = 0.1$ . It can be seen that  $h(iT_{sam})$  can be well represented by a sequence with finite length. Therefore, (6.10) can be approximated by (ignoring the frequency offset)

$$r_k \cong \sum_{i=-L_1}^{L_e-L_1-1} s(kT_{sam} - iT_{sam})h(iT_{sam}) + w_k, \quad (6.14)$$

where  $L_e$  and  $L_1$  are the total number of taps and the number of taps for  $t < 0$  in the approximated equivalent channel, respectively. Note that the above approximation can be made arbitrarily accurate by using large enough values of  $L_e$  and  $L_1$ .

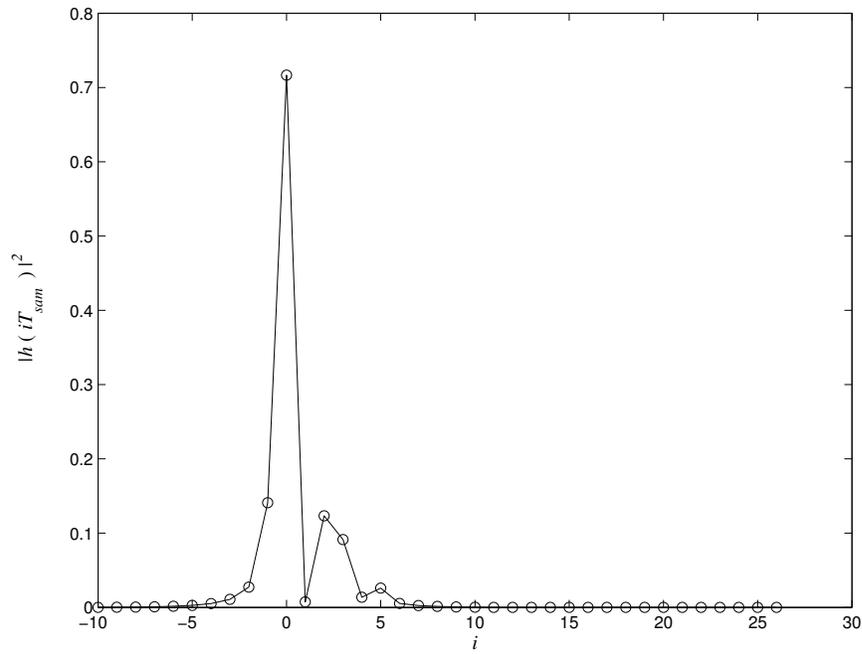


Fig. 36. An example of  $|h(iT_{sam})|^2$  for  $L_h = 6$ , the first tap of the physical channel has zero delay, other five taps have delays uniformly distributed over the interval  $0 - 300\text{ns}$ ,  $\gamma_n$  are i.i.d., zero-mean, complex Gaussian random variables with variances following the multipath intensity profile  $\phi(\tau) \sim e^{-\tau/\tau_{rms}}$ , where  $\tau_{rms} = 100\text{ns}$ ,  $\varepsilon_o$  is a random variable uniformly distributed in  $[0, 1)$  and  $f(t)$  is given by (6.11) with  $\alpha = 0.1$ .

Let  $\mathbf{r}_n$  be a received-signal vector with  $N$  received samples

$$\mathbf{r}_n = [r_n \ r_{n+1} \ \dots \ r_{n+N-1}]^T, \quad (6.15)$$

where  $N = T_b/T_{sam} = 16$  is the number of samples over the span of a short training symbol  $b(t)$ . Let  $b_n = b(nT_{sam})$  and  $g_n = g(nT_{sam})$  be the  $n^{th}$  samples of the short training symbol and of the cyclic prefix, respectively. For  $L_e - L_1 \leq n \leq 9N - L_1$ ,  $\mathbf{r}_n$  is given by

$$\mathbf{r}_n = \sqrt{2P}\mathbf{B}_{n+L_1}^{(L_e)} \mathbf{h}_o + \mathbf{w}_n, \quad (6.16)$$

where

$$\mathbf{B}_n^{(L)} \triangleq \begin{bmatrix} b_{\text{mod}(n,16)} & b_{\text{mod}(n-1,16)} & \dots & b_{\text{mod}(n-L+1,16)} \\ b_{\text{mod}(n+1,16)} & b_{\text{mod}(n,16)} & \dots & b_{\text{mod}(n-L+2,16)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{\text{mod}(n+15,16)} & b_{\text{mod}(n+14,16)} & \dots & b_{\text{mod}(n-L+16,16)} \end{bmatrix} \quad (6.17)$$

$$\mathbf{h}_o \triangleq [h(-L_1T_{sam}) \ h((-L_1+1)T_{sam}) \ \dots \ h((L_e-L_1+1)T_{sam})]^T, \quad (6.18)$$

and  $\mathbf{w}_n$  is a vector containing the noise samples with a covariance matrix  $\sigma_w^2 \mathbf{I}_N$ .

Similarly, the first received-signal vector for the cyclic prefix is given by

$$\mathbf{r}_n = \sqrt{2P}\mathbf{G}_0^{(L_e)} \mathbf{h}_o + \mathbf{w}_n \quad \text{for } n = 10N - L_1 \quad (6.19)$$

where

$$\mathbf{G}_0^{(L)} \triangleq \begin{bmatrix} g_0 & b_{15} & b_{14} & \dots & b_{16-L+1} \\ g_1 & g_0 & b_{15} & \dots & b_{16-L+2} \\ g_2 & g_1 & g_0 & \dots & b_{16-L+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{15} & g_{14} & g_{13} & \dots & g_{16-L} \end{bmatrix}. \quad (6.20)$$

For notational simplicity, we write  $\mathbf{B}_n^{(L)}$  and  $\mathbf{G}_0^{(L)}$  as  $\mathbf{B}_n$  and  $\mathbf{G}_0$ , respectively, in the

subsequent derivations.

*Remark 3.* The equivalent channel, which includes the effect of filtering and sampling offset in addition to physical channel, is in general non-causal and has length longer than the span of the physical channel. For the channel example shown in Fig. 36, the path delays of the physical channel lie within  $[0, 300\text{ns}]$ , however, the equivalent channel has non-zero taps before  $t=0$  and has length longer than 300ns (although the “head” and “tail” of the equivalent channel are very small).

### C. Symbol Synchronization Performance Criterion

Having established the system model, the next question is how to define the “beginning” of an OFDM symbol. For frequency flat fading channels, the starting position is obvious and well defined. For Rician fading channels, it is reasonable to define the symbol boundary with respect to the first path. However, in a Rayleigh multipath fading channel (e.g., the one shown in Fig. 36), the channel contains some small taps at the beginning and the starting position of the channel is not clear. It can be defined as the first non-zero tap of the channel, as the first tap with energy larger than a certain threshold, as the position of the strongest path or any other definition. Because of this, the symbol boundary of a received OFDM symbol is not well defined. Even if we choose one of the above definitions as the reference position, there is no guarantee that a certain synchronization algorithm giving estimates close to the reference position would provide good performance in OFDM systems. Moreover, in OFDM systems, due to the existence of cyclic prefix, some timing offset can be tolerated as long as the samples within the FFT window are influenced by only one transmitted OFDM symbol. Therefore, the criterion that the synchronization error has to be within certain limits of a fixed reference point is not an appropriate

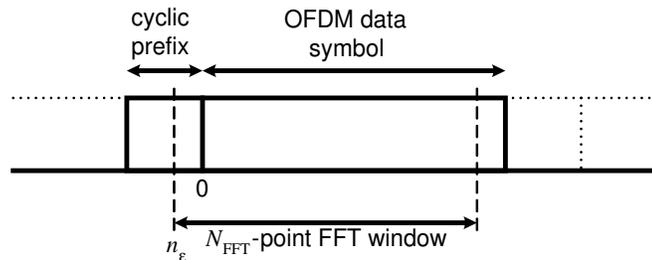


Fig. 37. OFDM symbol and FFT position.

performance measure for OFDM systems in frequency selective fading channels.

A more general and meaningful performance measure is the loss in system performance due to the synchronization error. With reference to Fig. 37, if the fast Fourier transform (FFT) window starts at position  $n_\epsilon$ , the signal at the sub-carrier  $k$  after the FFT operation,  $z_k$ , can be described as [30]

$$z_k = e^{j2\pi(k/N_{FFT})n_\epsilon} \kappa(n_\epsilon) a_k H_k + I_k + \eta_k, \quad (6.21)$$

where  $a_k$ ,  $H_k$  and  $\eta_k$  are the data, channel transfer function and noise sample at sub-carrier  $k$ , respectively,  $N_{FFT}$  is the number of FFT points in the OFDM system, which is 64 in IEEE 802.11a,  $\kappa(n_\epsilon)$  is the attenuation caused by the synchronization error, which can be well approximated by [30]

$$\kappa(n_\epsilon) = \sum_i |h(iT_{sam})|^2 \frac{N_{FFT} - \Delta\epsilon_i}{N_{FFT}}, \quad (6.22)$$

where

$$\Delta\epsilon_i \triangleq \begin{cases} n_\epsilon - i & n_\epsilon > i \\ i - N - n_\epsilon & n_\epsilon < -(N - i) \\ 0 & \text{otherwise} \end{cases}, \quad (6.23)$$

and  $I_k$  is the ISI plus inter-carrier interference (ICI) term at sub-carrier  $k$  caused by

the timing offset, which can be well approximated by Gaussian noise with power [30]

$$\sigma_\varepsilon^2(n_\varepsilon) = \sum_i |h(iT_{sam})|^2 \left( 2 \frac{\Delta\varepsilon_i}{N_{FFT}} - \left( \frac{\Delta\varepsilon_i}{N_{FFT}} \right)^2 \right). \quad (6.24)$$

For a particular channel realization, the signal-to-interference-plus-noise ratio (SINR) is given by

$$\text{SINR}(n_\varepsilon) = \frac{\kappa^2(n_\varepsilon) \mathbb{E}[|a_k H_k|^2]}{\sigma_\varepsilon^2(n_\varepsilon) + \sigma_w^2}. \quad (6.25)$$

Notice that for the special case that the equivalent channel length is smaller than the length of cyclic prefix (i.e.,  $L_e < N$ ) and  $L_e - N \leq n_\varepsilon \leq 0$ , then  $\sigma_\varepsilon^2 = 0$ , implying there is no ISI and ICI. On the other hand, if the length of the equivalent channel is larger than the length of the cyclic prefix, then  $\sigma_\varepsilon^2 > 0$ , therefore some ISI and ICI occur. In this case, the best we can do is to find a value of  $n_\varepsilon$  which maximizes the SINR. Noting that, due to (6.21), the SINR expression can be rewritten as

$$\text{SINR}(n_\varepsilon) = \frac{\mathbb{E}[|z_k|^2] - \sigma_\varepsilon^2(n_\varepsilon) - \sigma_w^2}{\sigma_\varepsilon^2(n_\varepsilon) + \sigma_w^2}. \quad (6.26)$$

Since the sum of average powers of the useful and the interference components in (6.21) is a constant [78], it follows that  $\mathbb{E}[|z_k|^2]$  is also a constant. Therefore, maximizing SINR is equivalent to minimizing  $\sigma_\varepsilon^2(n_\varepsilon)$  and the ‘‘ideal’’ symbol synchronizer should select  $n_\varepsilon$  such that  $\sigma_\varepsilon^2(n_\varepsilon)$  in (6.24) is minimized.

In practice, the ideal symbol synchronizer is not realizable since it requires the perfect knowledge of the channel realization. However, the ideal symbol synchronizer can serve as a reference to other practical synchronization algorithms. For a particular realization of channel, let  $n_\varepsilon$  be the start of FFT window estimated by a particular symbol synchronization algorithm and  $n_{id}$  be that from the ideal symbol synchronizer. Then the loss of SINR, defined as the ratio of SINR obtained from the ideal symbol

synchronizer to that from the non-ideal symbol synchronizer is given by

$$\text{SINR}_{\text{loss}}(n_\varepsilon) \triangleq \frac{\text{SINR}_{\text{id}}}{\text{SINR}(n_\varepsilon)} = \frac{\kappa^2(n_{\text{id}})[\sigma_\varepsilon^2(n_\varepsilon) + \sigma_w^2]}{\kappa^2(n_\varepsilon)[\sigma_\varepsilon^2(n_{\text{id}}) + \sigma_w^2]}. \quad (6.27)$$

For a good symbol synchronization algorithm, the loss in SINR with respect to the ideal synchronizer should be very small. Similar to [13], we define a synchronization failure as the event that the loss in SINR is greater than a tolerable system degradation. That is,

$$P_f(\Delta\gamma) \triangleq P(10 \log_{10}(\text{SINR}_{\text{loss}}) > \Delta\gamma), \quad (6.28)$$

where  $P_f(\Delta\gamma)$  is the probability of synchronization failure given that the tolerable system degradation (in dB) is  $\Delta\gamma$ , and  $P(\cdot)$  denotes the probability of an event.

*Remark 1.* Plugging (6.23) into (6.24), the power of the ISI plus ICI term can be written as

$$\begin{aligned} \sigma_\varepsilon^2(n_\varepsilon) = & \sum_{i=-\infty}^{n_\varepsilon-1} |h(iT_{\text{sam}})|^2 \left( 2 \frac{n_\varepsilon - i}{N_{\text{FFT}}} - \left( \frac{n_\varepsilon - i}{N_{\text{FFT}}} \right)^2 \right) \\ & + \sum_{i=n_\varepsilon+N+1}^{\infty} |h(iT_{\text{sam}})|^2 \left( 2 \frac{i - N - n_\varepsilon}{N_{\text{FFT}}} - \left( \frac{i - N - n_\varepsilon}{N_{\text{FFT}}} \right)^2 \right). \end{aligned} \quad (6.29)$$

The meaning of minimizing (6.29) is to find a starting position  $n_\varepsilon$  for a rectangular window of length  $N+1$  points such that the weighted sum of the energy of the channel taps outside this rectangular window is minimized. It is mentioned in [78] that, for simplicity, minimizing (6.29) can be replaced by minimizing

$$\tilde{\sigma}_\varepsilon^2(n_\varepsilon) = \sum_{i=-\infty}^{n_\varepsilon-1} |h(iT_{\text{sam}})|^2 + \sum_{i=n_\varepsilon+N+1}^{\infty} |h(iT_{\text{sam}})|^2. \quad (6.30)$$

Further notice that minimizing (6.30) is equivalent to maximizing

$$\sum_{i=n_\varepsilon}^{n_\varepsilon+N} |h(iT_{\text{sam}})|^2 \quad (6.31)$$

which is the optimal timing criterion proposed in [69, pp.90-92].

#### D. Proposed Symbol Synchronization Algorithm

For the packet structure shown in Fig. 35, determining the FFT window position for the OFDM data symbols actually involves two major steps. The first one is to identify a reference position (e.g., the transition from the short training symbol to the cyclic prefix of the long training symbol in the middle of the preamble) such that all the subsequent transitions can be predicted. We refer to this step as frame synchronization. The second step is to locate the FFT window position within an OFDM symbol such that the ISI plus ICI introduced is minimized, which we refer to it as symbol synchronization. Notice that in some cases, frame synchronization and symbol synchronization are the same process. For example, if the channel is frequency flat or the multipath channel is causal and with total length smaller than the cyclic prefix of the OFDM symbols, then the optimum position for the FFT window is at  $n_\epsilon = 0$  [78] and symbol synchronization follows naturally and immediately once frame synchronization is achieved. However, for the channel shown in Fig. 36, which is non-causal and the total length may be larger than the cyclic prefix of the OFDM symbols, symbol synchronization is essential.

##### 1. Frame Synchronization

Suppose that the arrival of the preamble can be identified by detecting the received-signal energy (e.g., using the methods in [16] or [20]), the problem of detecting the transition between the short training symbols and the cyclic prefix of the long training symbols can be decomposed into two sub-problems. Let  $\mathbf{r}_{n_1}$  be a received-signal vector within the short training interval. Since  $\mathbf{r}_{n_1}$  may not align with the beginning of a

short training symbol, the first sub-problem is to determine the current time offset with respect to the last short training symbol and hence predict the starting time of the next (expected) short training symbol. Suppose this starting time is denoted by the time index  $n_2$ . After  $n_2$  is obtained, the second sub-problem is to examine the incoming vectors  $\mathbf{r}_{n_2+qN}$ ,  $q = 0, 1, 2, \dots, 10$ , and to determine if they belong to a short training symbol or a cyclic prefix of the long training symbol. The transition point is declared at the time instant that the first  $\mathbf{r}_{n_2+qN}$  belongs to the cyclic prefix.

a. First stage

Assume that the received-signal vector  $\mathbf{r}_{n_1}$  is  $i_o$  samples ( $i_o \in \{0, 1, \dots, 15\}$ ) offset from the beginning of the current short training symbol. The probability density function (PDF) of the received-signal vector  $\mathbf{r}_{n_1}$  is

$$p(\mathbf{r}_{n_1}; \boldsymbol{\theta}, i, L) = \frac{1}{\pi^N \sigma^{2N}} \exp \left\{ -\frac{1}{\sigma^2} \|\mathbf{r}_{n_1} - \mathbf{B}_i \sqrt{2P} \mathbf{h}\|^2 \right\}, \quad (6.32)$$

where  $\boldsymbol{\theta} \triangleq [\Re(\sqrt{2P} \mathbf{h}^T), \Im(\sqrt{2P} \mathbf{h}^T), \sigma^2]^T$  with  $\mathbf{h}$ ,  $\sigma^2$  and  $L$  being the trial values of  $\mathbf{h}_o$ ,  $\sigma_w^2$  and  $L_e$ , respectively. Note that  $\mathbf{B}_i$  depends on  $i$  and  $L$  (see (6.17)). It is not possible to jointly estimate  $\mathbf{h}_o$ ,  $\sigma_w^2$ ,  $i_o$  and  $L_e$  by directly maximizing  $p(\mathbf{r}_{n_1}; \boldsymbol{\theta}, i, L)$  since (apart from the fact that the implementation complexity would be extremely high) the largest possible  $L$  is always chosen [79, p.223]. One criterion that gets around this problem is the generalized ML rule [79, p.223], in which we maximize

$$\psi(\mathbf{r}_{n_1}; i, L) = \ln p(\mathbf{r}_{n_1}; \hat{\boldsymbol{\theta}}, i, L) - \frac{1}{2} \ln \det(\mathbf{I}(\hat{\boldsymbol{\theta}}|i, L)), \quad (6.33)$$

where the second term is a penalty term that becomes more negative as  $L$  increases. In the above expression,  $\hat{\boldsymbol{\theta}}$  is the ML estimate of  $\boldsymbol{\theta}$  (given  $i$  and  $L$ ) with elements

given by [65, p.186] [79, p.222]

$$\widehat{\sqrt{2P}\mathbf{h}} = (\mathbf{B}_i^H \mathbf{B}_i)^{-1} \mathbf{B}_i^H \mathbf{r}_{n_1}, \quad (6.34)$$

$$\hat{\sigma}^2 = \frac{1}{N} \|\mathbf{r}_{n_1} - \mathbf{B}_i (\mathbf{B}_i^H \mathbf{B}_i)^{-1} \mathbf{B}_i^H \mathbf{r}_{n_1}\|^2, \quad (6.35)$$

and  $\mathbf{I}(\boldsymbol{\theta}|i, L)$  is the Fisher information matrix of  $\boldsymbol{\theta}$  (given  $i$  and  $L$ ) with elements given by [65, p.525]

$$[\mathbf{I}(\boldsymbol{\theta}|i, L)]_{mn} = \frac{1}{\sigma^4} \left[ \frac{\partial \sigma^2}{\partial \theta_m} \frac{\partial \sigma^2}{\partial \theta_n} \right] + \frac{2}{\sigma^2} \Re \left[ \frac{\partial (\mathbf{B}_i \sqrt{2P}\mathbf{h})^H}{\partial \theta_m} \frac{\partial (\mathbf{B}_i \sqrt{2P}\mathbf{h})}{\partial \theta_n} \right], \quad (6.36)$$

where  $\theta_m$  is the  $m^{\text{th}}$  element of  $\boldsymbol{\theta}$ .

As discussed in Section C, since the starting position of a frame is not well defined in multipath Rayleigh fading channels, it is necessary to clarify what is the meaning of estimated offset by maximizing (6.33). Let  $\{\hat{i}, \hat{L}\}$  be the set of values that maximize (6.33), then  $\hat{i}$  is the number of offset samples from the beginning of the current short training symbol *conditioned* that the number of paths of the channel is  $\hat{L}$  and the offset is with respect to the first estimated path. Notice that the generalized ML criterion (6.33) tends to ignore the channel paths of small energy, therefore,  $\hat{L} < L_e$ . For example, in the channel shown in Fig. 36, the estimated channel length from the generalized ML criterion is  $\hat{L} = 8$  ( $-2 \leq i \leq 5$ ) at SNR = 30dB.

It is proved in Appendix D that, the generalized ML rule (6.33) can be simplified to

$$\psi_1(\mathbf{r}_{n_1}; i, L) = (-N + L + 1) \ln \|\mathbf{r}_{n_1} - \mathbf{B}_i (\mathbf{B}_0^H \mathbf{B}_0)^{-1} \mathbf{B}_i^H \mathbf{r}_{n_1}\|^2 - \xi(L), \quad (6.37)$$

where  $\xi(L) \triangleq L \ln 2 + \ln (\det(\mathbf{B}_0^H \mathbf{B}_0))$  is a function of  $L$  only and can be pre-computed and stored in a look-up table to reduce the real-time computational complexity. The first-stage synchronization algorithm, which jointly estimates the effective channel

order  $L_e$  and the delay  $i_o$  from the received-signal vector  $\mathbf{r}_{n_1}$ , becomes

$$\hat{i}, \hat{L} = \arg \max_{\substack{i \in \{0,1,\dots,15\}, \\ L \in \{1,2,\dots,L_{max}\}}} \psi_1(\mathbf{r}_{n_1}; i, L), \quad (6.38)$$

where  $L_{max}$  is the maximum possible value of the channel order. The starting position of the next expected short training symbol is then given by  $n_2 = n_1 + 16 - \hat{i}$ .

b. Second stage

The second step is to determine the smallest value of  $q$  such that  $\mathbf{r}_{n_2+qN}$  belongs to the cyclic prefix. This problem can be handled by Neyman-Pearson (NP) detection approach [79, ch.3]. Let  $H_g$  and  $H_b$  be the hypotheses that  $\mathbf{r}_{n_2+qN}$  belongs to the cyclic prefix and the short training symbol, respectively. In each test, the probability that the received-signal vector belongs to the short training symbol is the same as the probability that it belongs to the cyclic prefix. It follows that  $n_2 + qN$  is the point of transition if the condition

$$p(\mathbf{r}_{n_2+qN}|H_g) > p(\mathbf{r}_{n_2+qN}|H_b) \quad (6.39)$$

occurs for the first time, where

$$p(\mathbf{r}_{n_2+qN}|H_g) = \frac{1}{\pi^N \sigma^{2N}} \exp \left\{ -\frac{1}{\sigma^2} \|\mathbf{r}_{n_2+qN} - \mathbf{G}_0 \sqrt{2P} \mathbf{h}\|^2 \right\} \Bigg|_{L=\hat{L}} \quad (6.40)$$

$$p(\mathbf{r}_{n_2+qN}|H_b) = \frac{1}{\pi^N \sigma^{2N}} \exp \left\{ -\frac{1}{\sigma^2} \|\mathbf{r}_{n_2+qN} - \mathbf{B}_0 \sqrt{2P} \mathbf{h}\|^2 \right\} \Bigg|_{L=\hat{L}}. \quad (6.41)$$

Taking logarithm on both sides of (6.39), putting  $\widehat{\sqrt{2P} \mathbf{h}} = (\mathbf{G}_0^H \mathbf{G}_0)^{-1} \mathbf{G}_0^H \mathbf{r}_{n_2+qN}$  into (6.40) and  $\widehat{\sqrt{2P} \mathbf{h}} = (\mathbf{B}_0^H \mathbf{B}_0)^{-1} \mathbf{B}_0^H \mathbf{r}_{n_2+qN}$  into (6.41), we find that  $n_2 + qN$  is the point of transition if

$$\psi_2(\mathbf{r}_{n_2+qN}|H_g) > \psi_2(\mathbf{r}_{n_2+qN}|H_b), \quad (6.42)$$

where

$$\psi_2(\mathbf{r}_{n_2+qN}|H_g) = \mathbf{r}_{n_2+qN}^H \mathbf{G}_0 (\mathbf{G}_0^H \mathbf{G}_0)^{-1} \mathbf{G}_0^H \mathbf{r}_{n_2+qN} \Big]_{L=\hat{L}} \quad (6.43)$$

$$\psi_2(\mathbf{r}_{n_2+qN}|H_b) = \mathbf{r}_{n_2+qN}^H \mathbf{B}_0 (\mathbf{B}_0^H \mathbf{B}_0)^{-1} \mathbf{B}_0^H \mathbf{r}_{n_2+qN} \Big]_{L=\hat{L}}. \quad (6.44)$$

## 2. Position of FFT Window

After the transition between short training symbols and the cyclic prefix of the long training symbols is detected, we can predict that the beginning of the first data carrying OFDM symbol (the OFDM symbol for the signal field) is  $n_3 \triangleq n_2 + \hat{q}N + (32 + 2 \times 64)$ , where  $32 + 2 \times 64$  is the number of samples for the long training symbols. If the equivalent channel has exactly  $\hat{L}$  paths, then the allowable range for the starting point of the FFT window is  $\{n_3 + \hat{L}, \dots, n_3 + N\}$ . However, in reality, it is expected that there are some channel taps with small energy preceding and following the  $\hat{L}$  paths. With the observation that the “head” and “tail” of the equivalent channel have energy die down more or less at the same rate, it is reasonable to set the start of the FFT window at

$$n_4 \triangleq n_3 + \hat{L} + \lfloor (N - \hat{L})/2 \rfloor. \quad (6.45)$$

For subsequent OFDM symbols, the starting points of the FFT window are then  $n_4 + \ell(N + N_{FFT})$  where  $\ell$  is a positive integer.

## 3. Summary and Remarks

The proposed symbol synchronization procedure is summarized as follows.

**Step 1)** Take a received-signal vector  $\mathbf{r}_{n_1}$  of length  $N$  at any starting position  $n_1$  after the signal is detected. Find estimated values of  $i$  and  $L$  such that

$\psi_1(\mathbf{r}_{n_1}; i, L)$  is maximized over  $i \in \{0, 1, \dots, 15\}$  and  $L \in \{1, 2, \dots, L_{max}\}$  where  $\psi_1(\mathbf{r}_{n_1}; i, L)$  is given by (6.37). Calculate  $n_2 = n_1 + 16 - \hat{i}$ .

**Step 2)** Examine  $\mathbf{r}_{n_2+qN}$ ,  $q = 0, 1, \dots, 10$ . Find the smallest value of  $q$  such that  $\psi_2(\mathbf{r}_{n_2+qN}|H_g) > \psi_2(\mathbf{r}_{n_2+qN}|H_b)$  where  $\psi_2(\mathbf{r}_{n_2+qN}|H_g)$  and  $\psi_2(\mathbf{r}_{n_2+qN}|H_b)$  are given by (6.43) and (6.44), respectively. Declare  $n_2 + \hat{q}N$  as the time of transition.

**Step 3)** The FFT window for the first OFDM symbol should start at  $n_4 \triangleq n_2 + \hat{q}N + (32 + 2 \times 64) + \hat{L} + \lfloor (N - \hat{L})/2 \rfloor$ . For subsequent OFDM symbols, the starting points of the FFT window are then  $n_4 + \ell(N + N_{FFT})$  where  $\ell$  is a positive integer.

We make the following remarks.

1. A knowledge of  $n_1$  is in general not available at the receiver, so that the best knowledge that can be obtained after accomplishing Step 1 is the difference  $n_2 - n_1$ . Despite this, it is sufficient for the receiver to locate the next expected short training symbol.
2. In practical implementation of Step 2, the receiver need not examine the entire sequence of  $\mathbf{r}_{n_2+qN}$ . The time of transition can be declared right after the condition for test is met. The rest of the  $\mathbf{r}_{n_2+qN}$  vectors can be ignored.
3. For (6.34) to hold, it is required that  $N > L$  and  $\mathbf{B}_i$  be of full rank (i.e.,  $\text{rank}(\mathbf{B}_i) = L$ ) [65, pp. 186]. Numerical calculation shows that the maximum value of  $L$  that makes  $\mathbf{B}_i$  full-rank is  $L_{max} = 12$  which corresponds to a time duration of 600ns. As WLANs are mainly used inside buildings, and measurements show that the maximum delay spreads of physical channels in different indoor environments are only about 300ns [69, pp.18-19], together with the fact

that the generalized ML criterion treats the channel taps with small energies as noise, the proposed synchronizer can handle most practical situations.

4. It is interesting to note that the form of the generalized ML rule in (6.37) is similar to the GAIC used in [24]. However, there are several differences between the algorithm proposed here and the one in [24]. First, our proposed algorithm uses the short training symbols and the cyclic prefix of the long training symbol to achieve frame synchronization whereas the algorithm in [24] uses only the long training symbol. Second, the observation-vector length used in the proposed algorithm is 16 whereas it is 64 for the one in [24]. Third, our proposed algorithm is developed based on a time-domain approach while a frequency-domain analysis is employed in [24]. Lastly, our proposed algorithm considers a more general channel model than the algorithm in [24] (see *Remark 1* of Section B). Performance and complexity comparisons between the proposed algorithm and the technique of [24] are provided in the next section.
5. If  $L$  is perfectly known, the metric in the first stage reduces to the same form as the metric in the second stage. Furthermore, in this case, the proposed first-stage detection algorithm coincides with the general frame synchronizer for packet-based transmissions in frequency selective fading channels proposed in [71]-[73] (with the frequency offset equal to zero).

## E. Simulation Results and Discussions

### 1. Simulation Conditions

Simulations are run in order to investigate the synchronization performance of the proposed algorithm. The received samples are generated according to (6.14) with

$L_1 = 15$  and  $L_e = 36$  so that the range of index  $i$  in (6.14) is  $\{-15, \dots, 20\}$ . This enables the equivalent channel to be accurately represented (see Fig. 36). The channel gains  $\gamma_n$ ,  $n = 0, 1, \dots, L_h - 1$ , were modelled as mutually-independent, circularly-symmetric, zero-mean complex-Gaussian random variables (i.e., Rayleigh fading was considered). The number of physical paths is  $L_h = 6$ . The channel dispersion was modelled by an exponential function with  $\phi(\tau) \sim e^{-\tau/\tau_{rms}}$ , where  $\tau_{rms} = 100\text{ns}$ . The combined transmit and receive filter  $f(t)$  is given by a raised cosine filter (6.11) with  $\alpha = 0.1$ . Two different models for the arrival time of the channel paths are considered. The first one (referred to as the channel I) assumes that the path delays are sample spaced (i.e.,  $\tau_n = nT_{sam}$ ,  $n = 0, 1, \dots, 5$ ) and there is no sampling phase offset (i.e.,  $\varepsilon_o = 0$ ). The second one (referred to as the channel II) is more realistic and assumes the first tap has zero delay, the other five taps present delays uniformly distributed over the interval  $[0 - 300\text{ns}]$ , and the fractional timing delay  $\varepsilon_o$  is treated as a uniform random variable over  $[0,1)$ . Note that from the measurements performed in indoor channels [69, pp.18-19], the parameters of channel II basically represent the worst case channel in indoor environments. The channel is fixed during each packet but independent from one packet to another.

As a working assumption, we follow a suggestion of the standard [68] that the last three short training symbols are used for frame synchronization although in practice it varies from one situation to another. Therefore,  $n_1$  was treated as a uniform random variable over  $[5N + 1, 6N]$  in the simulation, and a value of  $n_1$  was randomly generated in each run. For each simulation run, the loss of SINR is calculated using (6.27), where the ideal symbol synchronizer selects a starting point for the FFT window such that (6.30) is minimized (simulation results not shown in this chapter show that minimizing (6.29) or (6.30) give the same results). The noise samples are i.i.d., zero mean complex Gaussian random variable. The signal-to-noise ratio (SNR)

Table II. Probability of synchronization failure for the proposed algorithm in channel II at SNR=25dB and different frequency offsets

$v$	$P_f(0.5\text{dB})$	$P_f(1\text{dB})$
-240kHz	$3.1 \times 10^{-3}$	$1 \times 10^{-4}$
-180kHz	$3.2 \times 10^{-3}$	$2 \times 10^{-4}$
-120kHz	$2.9 \times 10^{-3}$	$1 \times 10^{-4}$
-60kHz	$2.3 \times 10^{-3}$	$1 \times 10^{-4}$
0	$3.2 \times 10^{-3}$	0
60kHz	$2.3 \times 10^{-3}$	$1 \times 10^{-4}$
120kHz	$3.5 \times 10^{-3}$	0
180kHz	$2.6 \times 10^{-3}$	$1 \times 10^{-4}$
240kHz	$4.2 \times 10^{-3}$	$1 \times 10^{-4}$

was defined as

$$\text{SNR} \triangleq \frac{\frac{1}{2}\mathbb{E}[|r(t) - w(t)|^2]}{\frac{1}{2}\mathbb{E}[|w(t)|^2]} = \frac{2P}{\sigma_w^2}. \quad (6.46)$$

Each simulation point is obtained by averaging over  $10^4$  Monte-Carlo runs.

## 2. Effect of Frequency Offset

We first examine the effect of non-zero frequency offset. It is required that the deviation of the transmitted-signal center frequency is within  $\pm 20\text{ppm}$  [68]. Assume that the receiver oscillator also has a frequency uncertainty of  $\pm 20\text{ppm}$ . Since the highest operating frequency of the WLAN is 5.8GHz [68], the worst-case frequency offset in the receiver is  $\pm 232\text{kHz}$ . Table II lists the probability of synchronization failure,  $P_f(0.5\text{dB})$  and  $P_f(1\text{dB})$ , for the proposed algorithm in channel II against different frequency offsets, under the condition of SNR = 25dB. The frequency offset is estimated as  $\hat{v} = \arg\{\mathbf{r}_n^H \mathbf{r}_{n+N}\} / 2\pi N T_{sam}$  [77] for any  $L_e - L_1 \leq n < n_1$  and compensated before frame synchronization. For the case of  $v = 0$ , no frequency offset estimation is performed and it serves as a reference. The results indicate that the presence of frequency offset does not have significant effect on the probability of synchronization failure. In generating the simulation results for the rest of this chapter, we set  $v = 0$ .

### 3. Performances and Comparisons with Other Algorithms

In this section, the performance of the proposed algorithm will be compared to an existing symbol synchronization algorithm for IEEE 802.11a WLANs [24] and three other algorithms for general OFDM symbol synchronization [14], [20], [21]. The algorithms in [14], [20] and [21] are selected for comparison since they represent different types of techniques commonly used in symbol synchronization of OFDM systems. The algorithm in [14] represents a class of techniques that exploit the periodic structure of the received signal and requires no knowledge of the preamble pattern (autocorrelation technique); the algorithm in [20] represents a class of techniques that correlate the received signal with the known training signal (cross-correlation technique) and the algorithm in [21] stands for the technique that makes use of both the knowledge of transmitted preamble and the periodic structure (double correlation technique).

The algorithm based on GAIC [24] is designed to detect the transition between  $g(t)$  and  $c(t)$  in the preamble. Due to the fact that the GAIC algorithm provides also an estimate of the channel length, the FFT window starting position for the first OFDM data symbol can be calculated in a similar way to the proposed algorithm (see (6.45)). That is, the FFT window starts at  $\hat{n}_{GAIC} + 2 \times 64 + \hat{L}_{GAIC} + \lfloor (N - \hat{L}_{GAIC})/2 \rfloor$  where  $\hat{n}_{GAIC}$  and  $\hat{L}_{GAIC}$  are the timing estimate and channel length estimate from the GAIC algorithm, respectively. For the algorithms in [14], [20] and [21], since they are not originally designed for the IEEE 802.11a standard, they have to be slightly modified. Let  $\bar{\mathbf{r}}_n \triangleq [r_n \ r_{n+1} \ \dots \ r_{n+63}]^T$  and  $\tilde{\mathbf{r}}_n \triangleq [r_n \ r_{n+1} \ \dots \ r_{n+95}]^T$  be the received-signal vectors of lengths 64 and 96, respectively (an overbar or a tilde is added to distinguish them from the received-signal vector of length 16,  $\mathbf{r}_n$ , defined in (6.15)). Furthermore, let  $\mathbf{g} \triangleq [g_0 \ g_1 \ \dots \ g_{31} \ c_0 \ c_1 \ \dots \ c_{31}]^T$ , where  $c_n \triangleq c(nT_{sam})$ , be the known training sequence of length 64, starting from the first sample of the cyclic prefix  $g(t)$ .

Then the modifications are as follows.

1. For the autocorrelation technique, the repetition structure in the long training symbols is being exploited. The transition between  $b(t)$  and  $g(t)$  is estimated by [14]

$$\hat{n}_{AC} = \arg \max_n \{ |\tilde{\mathbf{r}}_n^H \tilde{\mathbf{r}}_{n+64}| - \rho_1 (\|\tilde{\mathbf{r}}_n\|^2 + \|\tilde{\mathbf{r}}_{n+64}\|^2)/2 \} \quad (6.47)$$

where  $\rho_1 \triangleq \text{SNR}/(1 + \text{SNR})$ . Due to the structure of the long training symbols, an observation length of 96 is needed, otherwise, the cost function inside the max operation would have a plateau, leading to uncertainty for the start of the frame. Notice that similar algorithms have been proposed in [12], [13] and [36]. Once the transition between  $b(t)$  and  $g(t)$  is identified, the FFT window starting position for the first OFDM data symbol would be  $\hat{n}_{AC} + (32 + 2 \times 64) + 16 - \lambda$ , where  $\lambda$  is the pre-advancement to account for the mean shift of the estimated timing position caused by the channel dispersion [80]. As there is no channel length information, the value of pre-advancement  $\lambda$  is chosen based on the following intuitive argument. Since the length of the cyclic prefix for the OFDM symbols is 800ns, and from the channel measurements, the maximum delay spread of the indoor physical channel is smaller than 300ns [69, pp.18-19], therefore, there will be about 500ns of cyclic prefix contains very small amount of ISI and it is safest to start the FFT window in the middle of this 500ns region. That is,  $\lambda = 500\text{ns}/2T_{sam} = 5$ .

2. Define  $Q_n \triangleq \mathbf{g}^H \tilde{\mathbf{r}}_n$  as the correlation between the received vector and the known training sequence vector. Since the periodic autocorrelation property of vector  $\mathbf{g}$  resembles that of a pseudo noise (PN) sequence (this can easily be shown by numerical computations), correlation peaks are expected if the received vector starts near the transition between  $b(t)$  and  $g(t)$ . The cross-correlation based

algorithm in [20] is used to detect this correlation peak and can be stated as first finding the smallest value of  $n$  such that  $|Q_n + Q_{n+1}|^2 - \rho_2 \|\mathbf{g}\|^2 \|\bar{\mathbf{r}}_n\|^2 > 0$ , then the frame position is given by

$$\hat{n}_{CC} = \arg \max_n \{|Q_n|, |Q_{n+1}|\}, \quad (6.48)$$

where  $\rho_2 = 0.8$  as suggested by [81]. Similar to the autocorrelation algorithm, the FFT window starting position for the first OFDM data symbol would be  $\hat{n}_{CC} + (32 + 2 \times 64) + 16 - \lambda$ .

3. For the double correlation algorithm in [21],  $Q_n$  is first generated and then the conjugate of the correlation outputs 64 samples later (i.e.,  $Q_{n+64}^*$ ) is multiplied with  $Q_n$ . According to [21], the correlation peaks of the product  $|Q_n Q_{n+64}^*|$  approximately correspond to the channel tap-power for each delay. Therefore, based on the same rationale as the ideal synchronizer (see (6.31)), a sum of  $|Q_n Q_{n+64}^*|$  over a rectangular window of length  $N + 1$  should be used to locate the correct timing. Mathematically, it can be stated as

$$\hat{n}_{DC} = \arg \max_n \left\{ \sum_{i=n}^{n+N} |Q_i Q_{i+64}^*| \right\}. \quad (6.49)$$

Then the FFT window starting position for the first OFDM data symbol is given by  $\hat{n}_{DC} + (32 + 2 \times 64) + 16$ . No pre-advancement is needed since the estimator (6.49) introduces the pre-advancement implicitly.

Let us first consider the performances of different algorithms in channel I. Fig. 38 plots the distributions of the estimated FFT window positions at SNR=30dB with respect to  $n_\epsilon = 0$  (reference to Fig. 37). Notice that for channel I, as long as the FFT window starts in the interval  $n_\epsilon \in \{-10, \dots, 0\}$ , there is no performance penalty. From Fig. 38, it can be seen that the proposed algorithm and the algorithm based

on GAIC have very high estimation accuracies and all the estimated positions lie within the ISI-free region; while the other three correlation-based algorithms have larger estimation variances and present some estimated positions outside the ISI-free region (for the algorithm based on cross-correlation [20], since the threshold  $\rho_2$  is not optimal in frequency selective fading channels, about 10% of estimated positions lie outside the display of this figure). One may argue that the performance of the double correlation algorithm can be improved if all the estimates are shifted by 5 samples to the right (similarly, the results of the autocorrelation algorithm can be corrected by shifting about 3 samples to the left). However, this cannot be done in practice since the amount of mean shifts are not known in reality. Notice that the mean shifts deduced from the simulated distributions are only available if i) a lot of trials were run and ii) the optimal frame position with respect to the estimated position is known. Unfortunately, these two conditions cannot be met in practice. Furthermore, the amount of mean shifts would highly depend on the channel characteristics (delay spread and the number of physical paths), which ultimately depend on the operating environment. The amount of mean shift suitable for one environment may not be suitable for another.

Fig. 39 plots the probability of synchronization failure  $P_f(0.5\text{dB})$  as a function of SNR. It can be observed that the proposed algorithm and the algorithm based on GAIC have zero probability of synchronization failure for  $\text{SNR} \geq 10\text{dB}$ . For other correlation-based algorithms under consideration, although some perform pretty well at certain SNR regions (e.g., autocorrelation algorithm at high SNRs and double correlation algorithm at  $5\text{dB} \leq \text{SNR} \leq 10\text{dB}$ ), in general, they are not as reliable as the proposed algorithm and the algorithm based on GAIC.

Now, let us consider the performances of different algorithms in channel II. Fig. 40 plots the distributions of the estimated FFT window positions at  $\text{SNR}=30\text{dB}$  with

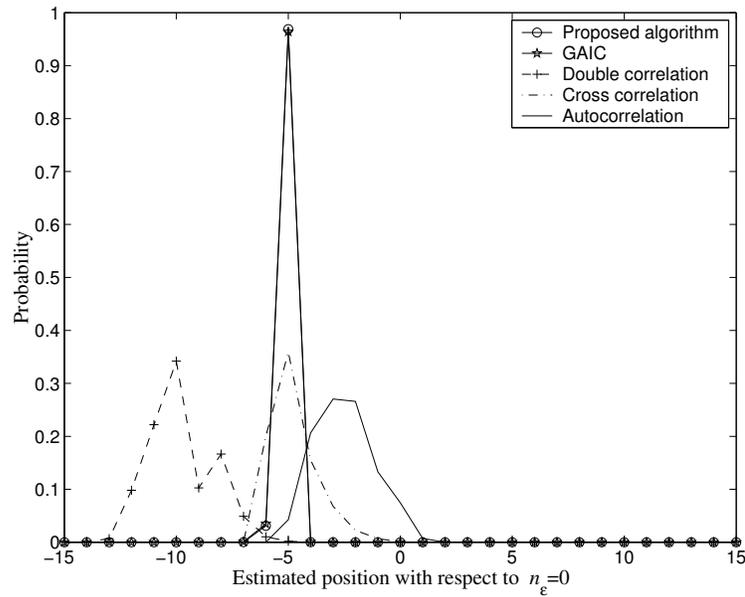


Fig. 38. Distributions of the detected FFT window starting position for the proposed algorithm and algorithms based on autocorrelation [14], cross-correlation [20], double correlation [21] and GAIC [24] with respect to the ideal FFT window starting position in channel I.

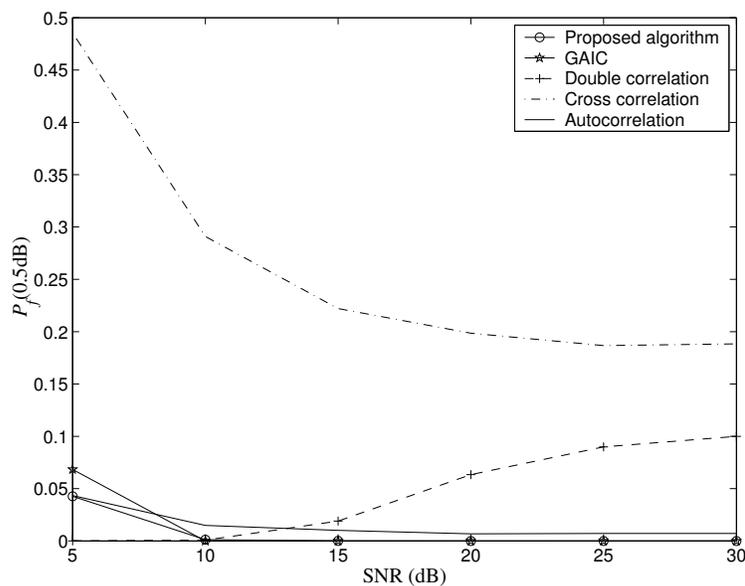


Fig. 39.  $P_f(0.5\text{dB})$  for the proposed algorithm and algorithms based on autocorrelation [14], cross-correlation [20], double correlation [21] and GAIC [24] as a function of SNR in channel I.

respect to the ideal position. Note that the ideal position is the starting point for the FFT window such that (6.30) is minimized and is different for different channel realizations. First, it can be seen that the distributions for all algorithms have larger variances compared to the case of channel I. This is because the start of the channel is not clear in channel II, small amount of noise can lead to a shift in the estimated position. Secondly, from Fig. 40, it is obvious that the proposed algorithm and the algorithm based on GAIC have estimated positions closer to the ideal position compared with the estimates from other algorithms.

Fig. 41 shows the probability of synchronization failure  $P_f(0.5\text{dB})$  as a function of SNR for channel II. Notice that the curves of  $P_f$  in general show an “U shape”. This is because at low SNRs, the estimation is not accurate due to the high level of noise, while at high SNRs, although the estimated positions can be quite accurate, a small amount of shift with respect to the ideal position leads to a large amount of loss in SINR (see (6.27)). From Fig. 41, it is clear that the proposed algorithm and the algorithm based on GAIC have similar performances and are the best compared to others (at least for  $\text{SNR} \geq 10\text{dB}$ ).

The poor performances of the correlation-based algorithms are mainly due to the fact that the signal filtered through a multipath channel would present a complicated correlation output (as opposed to a single correlation peak in AWGN channel), making the detection of the correlation peak more difficult. For the proposed algorithm and the algorithm based on GAIC, although the same pilot signal is used, they also incorporate the multipath structure of the channel into the system model, making it more robust to multipath fading. Combining the results of Fig. 39 and 41, we can conclude that at medium to high SNRs ( $\text{SNR} \geq 10\text{dB}$ ), joint estimation of the channel and timing (the proposed algorithm and the algorithm based on GAIC) gives better performance than the correlation-based algorithms, although the implementation

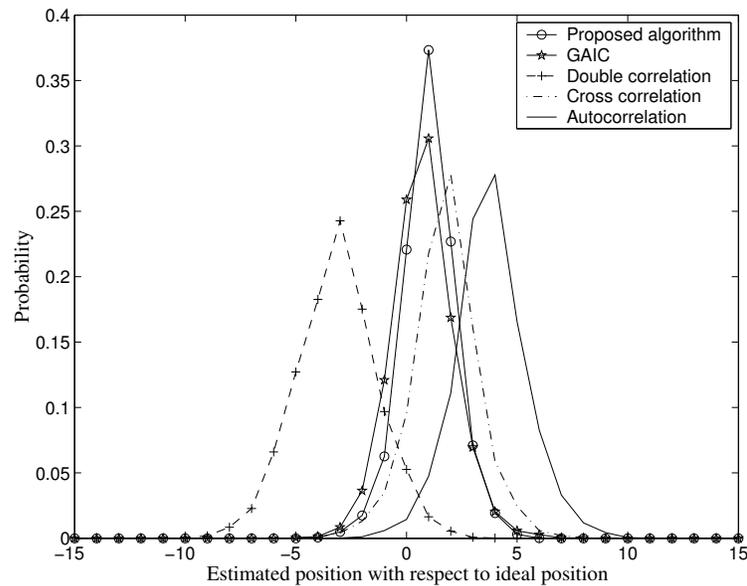


Fig. 40. Distributions of the detected FFT window starting position for the proposed algorithm and algorithms based on autocorrelation [14], cross-correlation [20], double correlation [21] and GAIC [24] with respect to the ideal FFT window starting position in channel II.

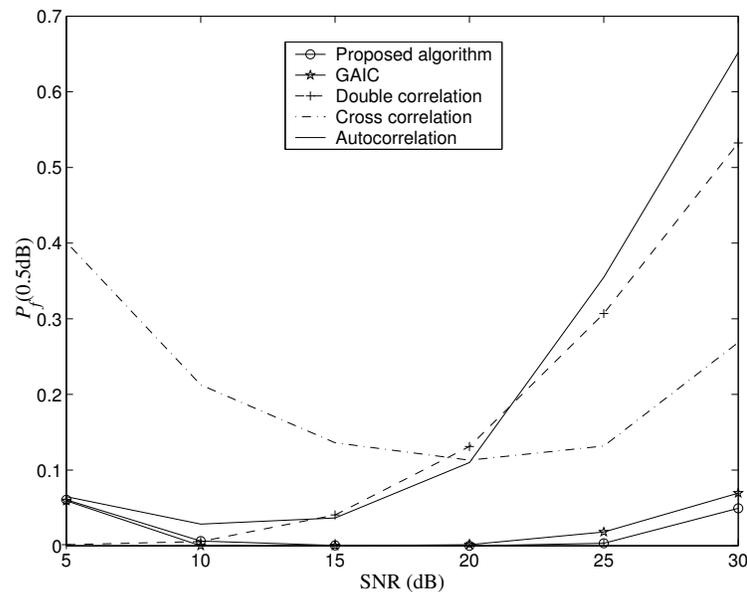


Fig. 41.  $P_f(0.5\text{dB})$  for the proposed algorithm and algorithms based on autocorrelation [14], cross-correlation [20], double correlation [21] and GAIC [24] as a function of SNR in channel II.

complexity of the correlation-based algorithms are much smaller.

Finally, we want to mention that although the performances of the proposed algorithm is comparable to that of GAIC algorithm, the complexity of the proposed algorithm is much smaller. This can be explained as follows. Both algorithms involve a least-squares fitting in the form of  $\|\mathbf{r} - \Phi\mathbf{r}\|^2$ , where  $\mathbf{r}$  is an observation vector in the time domain for the proposed algorithm and is an observation vector transformed into the frequency domain using FFT for the algorithm based on GAIC,  $\Phi$  is some square matrix depending on the parameters to be estimated (compare (6.37) of this chapter with (4) in [24]). Since the observation length is only 16 for the proposed algorithm while it is 64 for the GAIC algorithm, the number of multiplications for the proposed algorithm in each hypothesis test is 16 times less than that of GAIC algorithm. Taking into account the fact that, for the proposed algorithm, the number of hypothesis tests is smaller than that of the GAIC algorithm, and there is no need to transform the observation vector into frequency domain before least-squares fitting, the proposed algorithm is at least 16 times less complex than the GAIC algorithm.

## F. Conclusions

In this chapter, based on the maximum-likelihood principle and the preamble structure of IEEE 802.11a standard, a new symbol synchronization algorithm for IEEE 802.11a WLANs over frequency-selective fading channels is proposed. A realistic channel model was employed, which includes the effects of the physical channel, filtering and unknown sampling phase offset. Loss in system performance due to synchronization error was used as a performance criterion. Computer simulations showed that the proposed algorithm exhibits better performance than the correlation based algorithms. When compared to the algorithm based on generalized Akaike informa-

tion criterion (GAIC), the proposed algorithm has comparable performance, but with significantly reduced complexity.

## CHAPTER VII

## CONCLUSIONS AND FUTURE WORK

## A. Conclusions of This Dissertation

In this dissertation, the problem of symbol timing synchronization for three different communication systems was studied.

For single carrier systems with single antenna, a general feedforward symbol-timing estimation framework based on the CML principle was introduced. The mean-square-error (MSE) performance of the proposed estimator was analyzed both analytically and via simulations. Moreover, a unifying framework that subsumes a class of blind feedforward symbol-timing estimators employing second-order statistics was proposed and the best timing estimators were determined. Results showed that the proposed feedforward CML estimator and the SLN estimator with a properly designed pre-filter perform the best.

For single carrier systems with multiple antennas, two new symbol timing estimation algorithms were proposed. Firstly, the optimum sample selection algorithm by Naguib *et al.* was improved. Both analytical and simulation results showed that, for modest oversampling ratio (such as  $Q=4$ ), the MSE of the proposed estimator is significantly smaller than that of the optimum sample selection algorithm. Secondly, the symbol-timing estimation problem in MIMO systems was tackled using the ML principle. The DA and NDA ML symbol timing estimators and their corresponding CCRB and MCRB in MIMO correlated flat-fading channels were derived. Simulation results under different operating conditions (e.g., number of antennas and correlation between antennas) were given to assess and compare the performances of the DA and NDA ML estimators with respect to their corresponding CCRBs and MCRBs.

For the IEEE 802.11a WLANs system, the ML timing synchronizer was developed. A realistic channel, which includes the effects of filtering and sampling time offset in addition to the physical channel with random path delays, was considered. Furthermore, the loss in system performance due to synchronization error was used as the performance criterion. Results showed that the proposed algorithm performs better than correlation-based algorithms. Furthermore, the proposed algorithm performs as good as the one based on GAIC principle but has at least 16 times smaller complexity.

## B. Suggestions for Future Work

By no means, this dissertation can cover all the aspects of symbol timing synchronization. Based upon this dissertation, there are at least two directions for future research. One is to further improve the algorithms proposed in this dissertation. For example, the proposed symbol timing estimator for single-carrier single-antenna systems in Chapter II does not take into account the finite constellation of the symbols being transmitted. If this information is incorporated into the design of the estimator, it is expected that the resultant estimator would have a better performance. Another example is the synchronization problem of MIMO OFDM (multi-carrier, multi-antenna) systems. In this dissertation, the synchronization problems for MIMO and OFDM systems have been studied separately, but not together. The synchronization problem of MIMO OFDM systems thus appears to be a natural extension of this research work. The second direction is to look at the possibility of integrating the synchronization algorithms in the physical layer with those in the higher layers of general wireless communication network.

## REFERENCES

- [1] C. W. Farrow, "A continuously variable digital delay element," in *Proc. of IEEE ISCAS88*, 1998, pp. 2641-2645.
- [2] L. Erup, F. M. Gardner and R. A. Harris, "Interpolation in digital modems – part II: implementation and performance," *IEEE Trans. Commun.*, vol. 41, pp. 998-1008, June 1993.
- [3] M. Oerder and H. Meyr, "Digital filter and square timing recovery," *IEEE Trans. Commun.*, vol. 36, pp. 605-612, May 1988.
- [4] E. Panayirci and E. Y. Bar-Ness, "A new approach for evaluating the performance of a symbol timing recovery system employing a general type of nonlinearity," *IEEE Trans. Commun.*, vol. 44, pp. 29-33, Jan. 1996.
- [5] F. Gini and G. B. Giannakis, "Frequency offset and symbol timing recovery in flat-fading channels: A cyclostationary approach," *IEEE Trans. Commun.*, vol. 46, pp. 400-411, Mar. 1998.
- [6] K. E. Scott and E. B. Olasz, "Simultaneous clock phase and frequency offset estimation," *IEEE Trans. Commun.*, vol. 43, pp. 2263-2270, July 1995.
- [7] Y. Wang, E. Serpedin and P. Ciblat, "Blind feedforward cyclostationarity-based timing estimation for linear modulations," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 709-715, May 2004.
- [8] M. Morelli, A. N. D'Andrea and U. Mengali, "Feedforward ML-based timing estimation with PSK signals," *IEEE Commun. Letters*, vol. 1, pp. 80-82, May 1997.

- [9] J. Riba, J. Sala and G. Vazquez, "Conditional maximum likelihood timing recovery: estimators and bounds," *IEEE Trans. Signal Processing*, vol. 49, pp. 835-850, Apr. 2001.
- [10] G. Vazquez and J. Riba, "Non-data-aided digital synchronization," in *Signal Processing Advances in Wireless and Mobile Communications: Volume 2*, G. B. Giannakis, Y. Hua, P. Stoica and L. Tong (Eds.), Upper Saddle River, N.J.: Prentice Hall, 2001, pp. 357-402.
- [11] A. F. Naguib, V. Tarokh, N. Seshadri and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE J. Select. Areas in Commun.*, vol. 16, pp. 1459-1478, Oct. 1998.
- [12] M. Speth, F. Classen and H. Meyr, "Frame synchronization of OFDM systems in frequency selective fading channels," in *Proc. of IEEE VTC'97*, May 1997, pp. 1807-1811.
- [13] M. Speth, D. Daecke and H. Meyr, "Minimum overhead burst synchronization for OFDM based broadband transmission," in *Proc. of IEEE Globecom 98*, 1998, pp. 3227-3232.
- [14] J.-J. van de Beek, M. Sandell and P. O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 1800-1805, Jul. 1997.
- [15] D. Lee and K. Cheun, "Coarse symbol synchronization algorithms for OFDM systems in multipath channels," *IEEE Commun. Letters*, vol. 6, pp. 446-448, Oct. 2002.

- [16] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, pp. 1613-1621, Dec. 1997.
- [17] S H. Muller-Weinfurtner, "On the optimality of metrics for coarse frame synchronization in OFDM: a comparison," in *Proc. of IEEE PIMRC 98*, 1998, pp. 533-537.
- [18] M. Mizoguchi, T. Onizawa and T. Kumagai, H. Takanashi and M. Morikura, "A fast burst synchronization scheme for OFDM" in *Proc. of IEEE ICUPC 98*, 1998, pp. 125-129.
- [19] T. Keller, L. Piazzo, P. Mandarini and L. Hanzo, "Orthogonal frequency division multiplex synchronization techniques for frequency-selective fading channels," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 999-1008, June 2001.
- [20] A. J. Coulson, "Maximum likelihood synchronization for OFDM using a pilot symbol: algorithms," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 2486-2494, Dec. 2001.
- [21] F. Tufvesson, O. Edfors and M. Faulkner, "Time and frequency synchronization for OFDM using PN-sequence preambles," in *Proc. of IEEE VTC 99*, 1999, vol. 4, pp.2203-2207.
- [22] B. Yang, K. B. Letaief, R. S. Cheng and Z. Cao, "Timing recovery for OFDM transmission," *IEEE J. Select. Areas in Commun.*, vol. 18, pp. 2278-2291, Nov. 2000.
- [23] K.-W. Yip, T.-S. Ng and Y.-C. Wu, "Impacts of multipath fading on the timing synchronization of IEEE 802.11a wireless LANs," in *Proc. of IEEE ICC 2002*, May 2002, pp. 517-521.

- [24] E. G. Larsson, G. Liu, J. Li and G. B. Giannakis, "Joint symbol timing and channel estimation for OFDM based WLANs," *IEEE Commun. Letters*, vol. 5, pp. 325-327, Aug. 2001.
- [25] Y.-C. Wu and E. Serpedin, "Low-complexity feedforward symbol timing estimator using Conditional Maximum Likelihood principle," *IEEE Commun. Letters*, vol. 8, pp. 168-170, Mar. 2004.
- [26] Y.-C. Wu and E. Serpedin, "Design and analysis of feedforward symbol timing estimators based on the Conditional Maximum Likelihood principle," *IEEE Trans. Signal Processing*, vol. 53, no. 5, pp. 1908-1918, May 2005.
- [27] Y.-C. Wu and E. Serpedin, "Unified analysis of a class of blind feedforward symbol timing estimators employing second-order statistics," *IEEE Trans. Wireless Commun.*, (accepted Jan. 2005).
- [28] Y.-C. Wu, S. C. Chan and E. Serpedin, "Symbol-timing estimation in space-time coding systems based on orthogonal training sequences," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 603-613, Mar. 2005.
- [29] Y.-C. Wu and E. Serpedin, "Symbol timing estimation in MIMO correlated flat-fading channels," *Wireless Communications and Mobile Computing Journal (WCMC)*, Special Issue on Multiple-Input Multiple-Output (MIMO) Communications, vol. 4, no. 7, pp. 773-790, Nov. 2004.
- [30] M. Speth, S. A. Fechtel, G. Fock and H. Meyr, "Optimum receiver design for wireless broad-band systems using OFDM – Part I," *IEEE Trans. Commun.*, vol. 47, pp. 1668-1677, Nov. 1999.

- [31] Y.-C. Wu, K.-W. Yip, T.-S. Ng and E. Serpedin, "Maximum-likelihood symbol synchronization for IEEE 802.11a WLANs on unknown frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, (accepted Aug. 2004).
- [32] P. Galko and S. Pasupathy, "On a class of Generalized MSK," in *Proc. of IEEE ICC81*, June 1981, pp. 2.4.1-2.4.5.
- [33] O. Andrisano and M. Chiani, "The first Nyquist criterion applied to coherent receiver design for generalized MSK signals," *IEEE Trans. Commun.*, vol. 42, pp. 449-457, Feb/Mar/Apr. 1994.
- [34] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*. New York: Plenum Press, 1997.
- [35] W.-P. Zhu, M. O. Ahmad, and M. N. S. Swamy, "A fully digital timing recovery scheme using two samples per symbol," in *Proc. of IEEE Int. Symp. on Circuit and Systems*, May 2001, pp. 421-424.
- [36] S. J. Lee, "A new non-data-aided feedforward symbol timing estimator using two samples per symbol," *IEEE Commun. Letters*, vol. 6, pp. 205-207, May 2002.
- [37] M. Morelli and G. Vitetta, "Joint phase and timing recovery for MSK-type signals," *IEEE Trans. Commun.*, vol. 48, pp. 1997-1999, Dec. 2000.
- [38] P. A. Laurent, "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulse (APM)", *IEEE Trans. Commun.*, vol. 34, pp. 150-160, Feb. 1986.
- [39] Y.-C. Wu and T.-S. Ng, "Symbol timing recovery for GMSK modulations based on the square algorithm," *IEEE Commun. Letters*, vol. 5, pp. 221-223, May

2001.

- [40] H. Samueli, "An improved search algorithm for the design of multiplierless FIR filters with powers-of-two coefficients," *IEEE Trans. Circuits Syst. - II*, vol. 36, pp. 1044-1047, Jul. 1989.
- [41] C.-K.-S. Pun, Y.-C. Wu, S.-C. Chan and K.-L. Ho, "On the design and efficient implementation of the Farrow structure," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 189-192, Jul. 2003.
- [42] M. Ghogho, A. Swami, T. Durrani, "On blind carrier recovery in time-selective fading channels," in *Proc. 33rd Asilomar Conf. Signals, Systems and Computers*, 1999, vol. 1, pp.243-247.
- [43] K. Shi, Y. Wang and E. Serpedin, "On the design of digital blind feedforward jitter free timing recovery schemes for linear modulations," *IEEE Trans. Commun.*, vol. 52, pp. 1464-1469, Sept. 2004.
- [44] Y. Wang, E. Serpedin and P. Ciblat, "An alternative blind feedforward symbol timing estimator using two samples per symbol," *IEEE Trans. Commun.*, vol. 51, pp. 1451-1455, Sept. 2003.
- [45] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311-335, Mar. 1998.
- [46] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas in Commun.*, vol. 16, pp. 1451-1458, Oct. 1998.

- [47] A. F. Naguib, N. Seshadri and A. R. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Magazine*, vol. 17, pp. 76-92, May 2000.
- [48] V. Tarokh, H. Jafarkhani and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE J. Select. Areas in Commun.*, vol. 17, pp. 451-460, Mar. 1999.
- [49] E. G. Larsson, P. Stoica, and J. Li, "Orthogonal space-time block codes: maximum likelihood detection for unknown channels and unstructured interferences," *IEEE Trans. Signal Processing*, vol. 51, pp. 362-372, Feb. 2003.
- [50] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Trans. Veh. Tech.*, vol. 53, pp. 655-665, May 2004.
- [51] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Select. Areas in Commun.*, vol. 20, pp. 1211-1226, Aug. 2002.
- [52] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa and K. Huber, "Multiple-input-multiple-output measurements and modeling in Manhattan," *IEEE J. Select. Areas in Commun.*, vol. 21, pp. 321-331, Apr. 2003.
- [53] L. Schumacher, J. P. Kermoal, F. Frederiksen, K. I. Pedersen, A. Algrans and P. E. Mogensen, "MIMO channel characterisation," *Deliverable D2 V1.1 of IST-1999-11729 METRA Project*, pp. 1-57, Feb. 2001. Available online: <http://www.ist-metra.org>
- [54] D. Gesbert, M. Shafi, D.-S. Shiu, P. J. Smith and A. Naguib, "From theory to

- practice: an overview of MIMO space-time coded wireless systems,” *IEEE J. Select. Areas in Commun.*, vol. 21, pp. 281-301, Apr. 2003.
- [55] G. J. Foschini, G. D. Golden, R. A. Valenzuela and P. W. Wolniansky, “Simplified processing for high spectral efficiency wireless communication employing multi-element arrays,” *IEEE J. Select. Areas in Commun.*, vol. 17, pp. 1841-1852, Nov. 1999.
- [56] C. Fragouli, N. Al-Dhahir and W. Turin, “Finite-alphabet constant-amplitude training sequence for multiple-antenna broadband transmission,” in *Proc. of IEEE ICC 2002*, 2002, pp. 6-10.
- [57] C. Fragouli, N. Al-Dhahir and W. Turin, “Reduced-complexity training schemes for multiple-antenna broadband transmissions,” in *Proc. of IEEE WCNC 2002*, 2002, pp. 78-83.
- [58] C. Fragouli, N. Al-Dhahir and W. Turin, “Training-based channel estimation for multiple-antenna broadband transmissions,” *IEEE Trans. on Wireless Commun.*, vol. 2, pp. 384-391, Mar. 2003.
- [59] D. C. Chu, “Polyphase codes with good periodic correlation properties,” *IEEE Trans. Inform. Theory*, pp. 531-532, Jul. 1972.
- [60] A. A. D’Amico, A. N. D’Andrea and U. Mengali, “Feedforward joint phase and timing estimation with OQPSK modulation,” *IEEE Trans. Veh. Tech.*, vol. 48, pp. 824-832, May 1999.
- [61] Y.-C. Wu and S.-C. Chan, “On the symbol timing recovery in space-time coding systems,” in *Proc. of IEEE WCNC 2003*, Mar. 2003, pp. 420-424.

- [62] P. Stocia and A. Nehorai, "MUSIC, maximum likelihood and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing.*, vol. 37, pp. 720-741, May 1989.
- [63] B. Ottersten, M. Viberg, P. Stoica and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in *Radar Array Processing*, S. Haykin, J. Litva, T. J. Shepherd (Eds.), Berlin: Springer-Verlag, 1993, pp. 99-151.
- [64] A. N. D'Andrea, U. Mengali and R. Reggiannini, "The modified Cramer-Rao bound and its application to synchronization problem," *IEEE Trans. Commun.*, vol. 42, pp. 1391-1399, Feb./Mar./Apr. 1994.
- [65] S. M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory*, Englewood Cliffs, N.J.: Prentice Hall, 1993.
- [66] Y. V. Zakharov, V. M. Baronkin and T. C. Tozer, "DFT-based frequency estimators with narrow acquisition range," *IEE Proc.-Commun.*, vol. 148, pp. 1-7, Feb. 2001.
- [67] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge: Cambridge University Press, 1990.
- [68] IEEE 802.11, "Supplement to IEEE standard for information technology telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: high-speed physical layer in the 5 GHz band," IEEE Std 802.11a-1999.

- [69] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*, Boston: Artech House Publishers, 2000.
- [70] J. Heiskala and J. Terry, *OFDM Wireless LANs: A Theoretical and Practical Guide*, Indianapolis: Sams Publishing, 2002.
- [71] S. A. Fechtel and H. Meyr, "Improved frame synchronization for spontaneous packet transmission over frequency-selective radio channels," in *Proc. of IEEE PIMRC 94*, 1994, pp. 353-357.
- [72] U. Lambrette, J. Horstmannshoff and H. Meyr, "Techniques for frame synchronization on unknown frequency selective channels," in *Proc. of IEEE VTC 97*, 1997, pp. 1059-1063.
- [73] Y. Koo and Y. H. Lee, "A joint maximum likelihood approach to frame synchronization in presence of frequency offset," in *Proc. of IEEE ICC 2002*, 2002, pp. 1546-1550.
- [74] J. A. Gansman, M. P. Fitz and J. V. Krogmeier, "Optimum and suboptimum frame synchronization for pilot-symbol-assisted modulation," *IEEE Trans. Commun.*, vol. 45, pp. 1327-1337, Oct. 1997.
- [75] H. Meyr, M. Oerder and A. Polydoros, "On sampling rate, analog prefiltering, and sufficient statistics for digital receivers," *IEEE Trans. Commun.*, vol. 42, pp. 3208-3213, Dec. 1994.
- [76] J. G. Proakis, *Digital Communications*, 3rd ed., New York: McGraw-Hill, 1995.
- [77] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 42, pp. 2908-2914, Oct. 1994.

- [78] H. Steendam and M. Moeneclaey, "Analysis and optimization of the performance of OFDM on frequency-selective time-selective fading channel," *IEEE Trans. Commun.*, vol. 47, pp. 1811-1819, Dec. 1999.
- [79] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. 2: Detection Theory*, Upper Saddle River, N.J.: Prentice-Hall, 1998.
- [80] H. Minn, V. K. Bhargava and K. B. Letaief, "A robust timing and frequency synchronization for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 822-839, July 2003.
- [81] A. J. Coulson, "Maximum likelihood synchronization for OFDM using a pilot symbol: analysis," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 2495-2503, Dec. 2001.

## APPENDIX A

## PROOF OF (2.36) AND (2.37)

From the definition of  $\Lambda(\varepsilon)$  in (2.16), we have

$$\begin{aligned} \mathbb{E}[\Lambda(k_1)\Lambda(k_2)] &= \mathbb{E}[\mathbf{r}^H \mathbf{B}_{k_1} \mathbf{r} \mathbf{r}^H \mathbf{B}_{k_2} \mathbf{r}] \\ &= \sum_{l_1=0}^{L_oQ-1} \sum_{i_1=0}^{L_oQ-1} \sum_{l_2=0}^{L_oQ-1} \sum_{i_2=0}^{L_oQ-1} b_{i_1 l_1}^{(k_1)} b_{i_2 l_2}^{(k_2)} \mathbb{E}[r^*(i_1)r(l_1)r^*(i_2)r(l_2)], \end{aligned} \quad (\text{A.1})$$

where  $b_{ij}^{(k)}$  is the  $(i, j)^{th}$  element in  $\mathbf{B}_k$ . Now we concentrate on  $\mathbb{E}[r^*(i_1)r(l_1)r^*(i_2)r(l_2)]$  which is given by

$$\begin{aligned} &\mathbb{E}[r^*(i_1)r(l_1)r^*(i_2)r(l_2)] \\ &= \mathbb{E} \left[ \left( e^{-j\theta_o} \sqrt{\frac{E_s}{T}} \sum_{n_1} d_{n_1}^* g^*(i_1 T/Q - n_1 T - \varepsilon_o T) + \eta^*(i_1) \right) \right. \\ &\quad \cdot \left( e^{j\theta_o} \sqrt{\frac{E_s}{T}} \sum_{n_2} d_{n_2} g(l_1 T/Q - n_2 T - \varepsilon_o T) + \eta(l_1) \right) \\ &\quad \cdot \left( e^{-j\theta_o} \sqrt{\frac{E_s}{T}} \sum_{n_3} d_{n_3}^* g^*(i_2 T/Q - n_3 T - \varepsilon_o T) + \eta^*(i_2) \right) \\ &\quad \left. \cdot \left( e^{j\theta_o} \sqrt{\frac{E_s}{T}} \sum_{n_4} d_{n_4} g(l_2 T/Q - n_4 T - \varepsilon_o T) + \eta(l_2) \right) \right]. \end{aligned} \quad (\text{A.2})$$

First, let consider linear modulations, in which case we have

$$\mathbb{E}[d_{n_1}d_{n_2}] = 0 , \quad (\text{A.3})$$

$$\mathbb{E}[d_{n_1}d_{n_2}^*] = \delta(n_1, n_2) , \quad (\text{A.4})$$

$$\mathbb{E}[\eta(i_1)\eta(i_2)] = 0 , \quad (\text{A.5})$$

$$\mathbb{E}[\eta^*(i_1)\eta(i_2)] = \frac{N_oQ}{T}\delta(i_1, i_2) , \quad (\text{A.6})$$

$$\mathbb{E}[d_{n_1}^*d_{n_2}d_{n_3}^*d_{n_4}] = \begin{cases} 1 & \text{for } n_1 = n_2 \neq n_3 = n_4 \\ 1 & \text{for } n_1 = n_4 \neq n_2 = n_3 \\ m_4 & \text{for } n_1 = n_4 = n_2 = n_3 \\ 0 & \text{otherwise ,} \end{cases} \quad (\text{A.7})$$

and therefore, 10 out of the 16 terms which result from (A.2) vanish. With the definitions

$$G_\varepsilon(i, j) \triangleq \sum_n g^*(iT/Q - nT - \varepsilon T)g(jT/Q - nT - \varepsilon T) , \quad (\text{A.8})$$

$$Z_\varepsilon(i, j, k, l) \triangleq \sum_n g^*(iT/Q - nT - \varepsilon T)g(jT/Q - nT - \varepsilon T) \cdot g^*(kT/Q - nT - \varepsilon T)g(lT/Q - nT - \varepsilon T), \quad (\text{A.9})$$

the remaining terms can be expressed as

$$\mathbb{E}[r^*(i_1)r(l_1)r^*(i_2)r(l_2)] = S_1 + S_{12} + S_{14} + S_{23} + S_{34} + S_4, \quad (\text{A.10})$$

where

$$\begin{aligned}
S_1 &\triangleq \frac{E_s^2}{T^2} \sum_{n_1} \sum_{n_2} \sum_{n_3} \sum_{n_4} \mathbb{E}[d_{n_1}^* d_{n_2} d_{n_3}^* d_{n_4}] g^*(i_1 T/Q - n_1 T - \varepsilon_o T) g(l_1 T/Q - n_2 T - \varepsilon_o T) \\
&\quad \cdot g^*(i_2 T/Q - n_3 T - \varepsilon_o T) g(l_2 T/Q - n_4 T - \varepsilon_o T) \\
&= \frac{E_s^2}{T^2} \{G_{\varepsilon_o}(i_1, l_1) G_{\varepsilon_o}(i_2, l_2) + G_{\varepsilon_o}(i_1, l_2) G_{\varepsilon_o}(i_2, l_1) + (m_4 - 2) Z_{\varepsilon_o}(i_1, l_1, i_2, l_2)\} \quad (\text{A.11})
\end{aligned}$$

$$\begin{aligned}
S_{12} &\triangleq \frac{E_s}{T} \sum_{n_1} \sum_{n_2} \mathbb{E}[d_{n_1}^* d_{n_2}] \mathbb{E}[\eta^*(i_2) \eta(l_2)] g^*(i_1 T/Q - n_1 T - \varepsilon_o T) g(l_1 T/Q - n_2 T - \varepsilon_o T) \\
&= \frac{E_s N_o Q}{T^2} G_{\varepsilon_o}(i_1, l_1) \delta(i_2, l_2), \quad (\text{A.12})
\end{aligned}$$

$$\begin{aligned}
S_{14} &\triangleq \frac{E_s}{T} \sum_{n_1} \sum_{n_4} \mathbb{E}[d_{n_1}^* d_{n_4}] \mathbb{E}[\eta(l_1) \eta^*(i_2)] g^*(i_1 T/Q - n_1 T - \varepsilon_o T) g(l_2 T/Q - n_4 T - \varepsilon_o T) \\
&= \frac{E_s N_o Q}{T^2} G_{\varepsilon_o}(i_1, l_2) \delta(i_2, l_1), \quad (\text{A.13})
\end{aligned}$$

$$\begin{aligned}
S_{23} &\triangleq \frac{E_s}{T} \sum_{n_2} \sum_{n_3} \mathbb{E}[d_{n_2} d_{n_3}^*] \mathbb{E}[\eta^*(i_1) \eta(l_2)] g(l_1 T/Q - n_2 T - \varepsilon_o T) g^*(i_2 T/Q - n_3 T - \varepsilon_o T) \\
&= \frac{E_s N_o Q}{T^2} G_{\varepsilon_o}(i_2, l_1) \delta(i_1, l_2), \quad (\text{A.14})
\end{aligned}$$

$$\begin{aligned}
S_{34} &\triangleq \frac{E_s}{T} \sum_{n_3} \sum_{n_4} \mathbb{E}[d_{n_3}^* d_{n_4}] \mathbb{E}[\eta^*(i_1) \eta(l_1)] g^*(i_2 T/Q - n_3 T - \varepsilon_o T) g(l_2 T/Q - n_4 T - \varepsilon_o T) \\
&= \frac{E_s N_o Q}{T^2} G_{\varepsilon_o}(i_2, l_2) \delta(i_1, l_1), \quad (\text{A.15})
\end{aligned}$$

$$\begin{aligned}
S_4 &\triangleq \mathbb{E}[\eta^*(i_1) \eta(l_1) \eta^*(i_2) \eta(l_2)] \\
&= \mathbb{E}[\eta^*(i_1) \eta(l_1)] \mathbb{E}[\eta^*(i_2) \eta(l_2)] + \mathbb{E}[\eta^*(i_1) \eta^*(i_2)] \mathbb{E}[\eta(l_1) \eta(l_2)] \\
&\quad + \mathbb{E}[\eta^*(i_1) \eta(l_2)] \mathbb{E}[\eta(l_1) \eta^*(i_2)] + \mathbb{E}[\eta^*(i_1)] \mathbb{E}[\eta(l_1)] \mathbb{E}[\eta^*(i_2)] \mathbb{E}[\eta(l_2)] \\
&= \frac{N_o^2 Q^2}{T^2} (\delta(i_1, l_1) \delta(i_2, l_2) + \delta(i_1, l_2) \delta(i_2, l_1)). \quad (\text{A.16})
\end{aligned}$$

Plugging (A.10)-(A.16) back into (A.1) and expressing the summations using matrices, some straightforward calculations lead to

$$\begin{aligned} \mathbb{E}[\Lambda(k_1)\Lambda(k_2)] &= \frac{E_s^2}{T^2} \{ \text{tr}[\mathbf{B}_{k_1} \mathbf{G}_{\varepsilon_o}^T] \text{tr}[\mathbf{B}_{k_2} \mathbf{G}_{\varepsilon_o}^T] + \text{tr}[\mathbf{B}_{k_1} \mathbf{G}_{\varepsilon_o}^T \mathbf{B}_{k_2} \mathbf{G}_{\varepsilon_o}^T] \} + c(k_1, k_2) \\ &+ \frac{E_s N_o Q}{T^2} \{ \text{tr}[\mathbf{B}_{k_1} \mathbf{G}_{\varepsilon_o}^T] \text{tr}[\mathbf{B}_{k_2}] + \text{tr}[\mathbf{B}_{k_2} \mathbf{G}_{\varepsilon_o}^T \mathbf{B}_{k_1}] + \text{tr}[\mathbf{B}_{k_1} \mathbf{G}_{\varepsilon_o}^T \mathbf{B}_{k_2}] + \text{tr}[\mathbf{B}_{k_2} \mathbf{G}_{\varepsilon_o}^T] \text{tr}[\mathbf{B}_{k_1}] \} \\ &+ \frac{N_o^2 Q^2}{T^2} \{ \text{tr}[\mathbf{B}_{k_1}] \text{tr}[\mathbf{B}_{k_2}] + \text{tr}[\mathbf{B}_{k_1} \mathbf{B}_{k_2}] \}. \end{aligned} \quad (\text{A.17})$$

By expanding (2.36), it can be easily verified that (2.36) is equivalent to (A.17), thus completed the proof. A similar procedure can be used to prove (2.37).

Now, let consider MSK and GMSK. Since the pseudo-symbols in (2.10) are not circularly symmetric, (A.3) and (A.7) have to be modified accordingly. After some lengthy but straightforward calculations, it is found that

$$\begin{aligned} \mathbb{E}[d_{n_1} d_{n_2}] &= (-1)^{n_1} \delta(n_1, n_2), \quad (\text{A.18}) \\ \mathbb{E}[d_{n_1}^* d_{n_2} d_{n_3}^* d_{n_4}] &= \begin{cases} 1 & \text{for } n_1 = n_2 \neq n_3 = n_4 \\ 1 & \text{for } n_1 = n_4 \neq n_2 = n_3 \\ 1 & \text{for } n_1 = n_4 = n_2 = n_3 \\ (-1)^{|n_2 - n_1|} & \text{for } n_1 = n_3 \neq n_2 = n_4 \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.19}) \end{aligned}$$

Due to (A.18), two more cross terms in the expansion of (A.2) have to be considered.

One of them is  $S_{13}$ , which is given by

$$\begin{aligned} S_{13} &\triangleq \\ e^{-2j\theta_o} \frac{E_s}{T} \sum_{n_1} \sum_{n_3} \mathbb{E}[d_{n_1}^* d_{n_3}^*] \mathbb{E}[\eta(l_1)\eta(l_2)] g^*(i_1 T/Q - n_1 T - \varepsilon_o T) g^*(i_2 T/Q - n_3 T - \varepsilon_o T). \end{aligned} \quad (\text{A.20})$$

But thanks to the correlation property of noise samples,  $S_{13} = 0$ . The other extra term is also zero due to the same reason. For the fourth order moment in (A.19), compared to the corresponding expression for linear modulations (A.7), we notice that  $m_4 = 1$  and there is an extra non-zero fourth-order moment. Therefore, apart from setting  $m_4 = 1$  in  $S_1$ , an extra term has to be added to  $S_1$  in (A.11). The modified  $S_1$ , denoted as  $S_{1MSK}$ , can be expressed as

$$\begin{aligned}
& S_{1MSK} \\
&= S_1|_{m_4=1} + \frac{E_s^2}{T^2} \sum_{\substack{n_1 \\ n_1 \neq n_2}} \sum_{n_2} (-1)^{|n_2-n_1|} g^*(i_1T/Q - n_1T - \varepsilon_oT) g(l_1T/Q - n_2T - \varepsilon_oT) \\
&\quad \cdot g^*(i_2T/Q - n_1T - \varepsilon_oT) g(l_2T/Q - n_2T - \varepsilon_oT). \tag{A.21}
\end{aligned}$$

Plugging (A.21) into (A.1), and then expressing the multiplications using matrix notation, it can be proved that the only change is the definition of  $c(k_1, k_2)$ , which is given in (2.40).

## APPENDIX B

## PROOF OF (4.29) AND (4.30)

Since we can construct orthogonal sequences such that (4.12) is satisfied, the ISI term in (4.9) vanishes. Further with the fact that  $\|\mathbf{c}_i\|^2 = L_t$ , we have

$$\Lambda_{ML}(k) = \sum_{i=1}^N \sum_{j=1}^M |\lambda_{ij}(k)|^2 = W^2 h p^2(kT/Q - \epsilon_o T) + v(k) \quad (\text{B.1})$$

where

$$W \triangleq \sqrt{\frac{E_s}{NT}} L_t, \quad (\text{B.2})$$

$$h \triangleq \sum_{i=1}^N \sum_{j=1}^M |h_{ij}|^2, \quad (\text{B.3})$$

$$v(k) \triangleq \sum_{i=1}^N \sum_{j=1}^M \{|\mathbf{c}_i^H \boldsymbol{\eta}_j(k)|^2 + 2Wp(kT/Q - \epsilon_o T) \Re[h_{ij}(\mathbf{c}_i^H \boldsymbol{\eta}_j(k))^*]\}. \quad (\text{B.4})$$

Then (4.27) can be rewritten as

$$\phi = W^2 h e^{j2\pi\epsilon_o} \sum_{k=0}^{Q-1} p^2(kT/Q - \epsilon_o T) e^{-j2\pi k/Q} + e^{j2\pi\epsilon_o} \sum_{k=0}^{Q-1} v(k) e^{-j2\pi k/Q}. \quad (\text{B.5})$$

Before we proceed to the calculation of  $\mathbb{E}[\phi^2]$  and  $\mathbb{E}[\phi\phi^*]$ , we first calculate the mean and the second moment of  $v(k)$ . Note the following facts

$$\mathbb{E}[\eta_j(l_1 Q + k')] = 0 \quad \forall j, l_1, k' \quad (\text{B.6})$$

$$\mathbb{E}[h_{ij} \eta_j(l_1 Q + k')] = 0 \quad \forall i, j, l_1, k' \quad (\text{B.7})$$

$$\mathbb{E}[\eta_j(l_1 Q + k') \eta_{j'}(l_2 Q + k'')] = 0 \quad \forall j, j', l_1, l_2, k', k'' \quad (\text{B.8})$$

$$\mathbb{E}[\eta_j(l_1Q + k')\eta_{j'}^*(l_2Q + k'')] = N_o\varphi(((l_1 - l_2)Q + k' - k'')T/Q)\delta_{jj'} \quad (\text{B.9})$$

$$\mathbb{E}[h_{ij}h_{i'j'}] = 0 \quad \forall i, i', j, j' \quad (\text{B.10})$$

$$\mathbb{E}[h_{ij}h_{i'j'}^*] = \delta_{ii'}\delta_{jj'} , \quad (\text{B.11})$$

where  $\delta_{ii'} = 1$  if  $i = i'$  and zero otherwise. Since the matched filter is a root raised cosine filter, we also have

$$\varphi(0) = 1 \quad (\text{B.12})$$

$$\varphi(\tau) = \varphi(-\tau) \quad (\text{B.13})$$

$$\varphi(lT) = 0 \quad \text{for } l \neq 0. \quad (\text{B.14})$$

Let

$$v_{ij}(k) = |\mathbf{c}_i^H \boldsymbol{\eta}_j(k)|^2 + 2Wp(kT/Q - \epsilon_oT)\Re\{h_{ij}(\mathbf{c}_i^H \boldsymbol{\eta}_j(k))^*\} \quad (\text{B.15})$$

such that  $v(k) = \sum_{i=1}^N \sum_{j=1}^M v_{ij}(k)$ . The mean of  $v_{ij}(k)$  is

$$\begin{aligned} \mathbb{E}[v_{ij}(k)] &= \sum_{l=0}^{L_t-1} \sum_{l'=0}^{L_t-1} c_i(l')c_i^*(l)\mathbb{E}[\eta_j(lQ + k)\eta_j^*(l'Q + k)] \\ &= \sum_{l=0}^{L_t-1} |c_i(l)|^2\mathbb{E}[|\eta_j(lQ + k)|^2] \\ &= N_oL_t , \end{aligned} \quad (\text{B.16})$$

where in the first equality, we applied (B.7) and in the second equality, we applied (B.14). Therefore,  $\mathbb{E}[v(k)] = MNN_oL_t$  is a constant and independent of  $k$ .

The second moment of  $v_{ij}(k)$  is given by

$$\begin{aligned} \mathbb{E}[v_{ij}(k')v_{i'j'}(k'')] &= \mathbb{E} [|\mathbf{c}_i^H \boldsymbol{\eta}_j(k')|^2|\mathbf{c}_{i'}^H \boldsymbol{\eta}_{j'}(k'')|^2 + 4W^2p(k'T/Q - \epsilon_oT)p(k''T/Q - \epsilon_oT) \\ &\quad \cdot \mathbb{E} \{ \Re\{h_{ij}(\mathbf{c}_i^H \boldsymbol{\eta}_j(k'))^*\} \Re\{h_{i'j'}(\mathbf{c}_{i'}^H \boldsymbol{\eta}_{j'}(k''))^*\} \} ] . \end{aligned} \quad (\text{B.17})$$

Note that (B.7) makes the cross terms vanish. Considering the first term in (B.17),

$$\begin{aligned} \mathbb{E} [|\mathbf{c}_i^H \boldsymbol{\eta}_j(k')|^2 |\mathbf{c}_{i'}^H \boldsymbol{\eta}_{j'}(k'')|^2] &= \sum_{l_1=0}^{L_t-1} \sum_{l_2=0}^{L_t-1} \sum_{l_3=0}^{L_t-1} \sum_{l_4=0}^{L_t-1} c_i^*(l_1) c_i(l_2) c_{i'}^*(l_3) c_{i'}(l_4) \\ &\quad \cdot \mathbb{E}[\eta_j(l_1 Q + k') \eta_j^*(l_2 Q + k') \eta_{j'}(l_3 Q + k'') \eta_{j'}^*(l_4 Q + k'')]. \end{aligned} \quad (\text{B.18})$$

Using the fact that if  $a, b, c, d$  are jointly Gaussian, then

$$\mathbb{E}[abcd] = \mathbb{E}[ab]\mathbb{E}[cd] + \mathbb{E}[ac]\mathbb{E}[bd] + \mathbb{E}[ad]\mathbb{E}[bc] + \mathbb{E}[a]\mathbb{E}[b]\mathbb{E}[c]\mathbb{E}[d], \quad (\text{B.19})$$

and applying (B.6), (B.8) and (B.9), we have

$$\begin{aligned} &\mathbb{E}[\eta_j(l_1 Q + k') \eta_j^*(l_2 Q + k') \eta_{j'}(l_3 Q + k'') \eta_{j'}^*(l_4 Q + k'')] \\ &= N_o^2 [\varphi((l_1 - l_2)T) \varphi((l_3 - l_4)T)] \\ &\quad + N_o^2 [\varphi((l_1 - l_4)T + (k' - k'')T/Q) \varphi((l_2 - l_3)T + (k' - k'')T/Q)] \delta_{jj'}. \end{aligned} \quad (\text{B.20})$$

Plugging this result back into (B.18), we obtain

$$\begin{aligned} &\mathbb{E} [|\mathbf{c}_i^H \boldsymbol{\eta}_j(k')|^2 |\mathbf{c}_{i'}^H \boldsymbol{\eta}_{j'}(k'')|^2] \\ &= N_o^2 L_t^2 + N_o^2 \left\{ \sum_{l_1=0}^{L_t-1} \sum_{l_4=0}^{L_t-1} c_i^*(l_1) c_{i'}(l_4) \varphi((l_1 - l_4)T + (k' - k'')T/Q) \right. \\ &\quad \left. \cdot \sum_{l_2=0}^{L_t-1} \sum_{l_3=0}^{L_t-1} c_i^*(l_2) c_{i'}(l_3) \varphi((l_2 - l_3)T + (k' - k'')T/Q) \right\} \delta_{jj'}. \end{aligned} \quad (\text{B.21})$$

Consider first  $i \neq i'$ . We note that the second term in (B.21) is approximately zero since  $\varphi(\tau)$  is a decaying function of  $\tau$ . When  $l_1 = l_4$  or  $|l_1 - l_4|$  is small,  $\varphi((l_1 - l_4)T + (k' - k'')T/Q)$  has significant values. But in these cases,  $\sum_{l_1=0}^{L_t-1} \sum_{l_4=0}^{L_t-1} c_i^*(l_1) c_{i'}(l_4) = 0$  since the training sequences are designed such that they are orthogonal when the relative delay is small. When  $|l_1 - l_4|$  is large,  $\varphi((l_1 - l_4)T + (k' - k'')T/Q) \approx 0$ . The same argument applies to  $\varphi((l_2 - l_3)T + (k' - k'')T/Q)$ . For  $i = i'$ , the only case that

the second term in (B.21) is nonzero is when  $l_1 = l_4$  and  $l_2 = l_3$ . Therefore, we have

$$\mathbb{E} [|\mathbf{c}_i^H \boldsymbol{\eta}_j(k')|^2 |\mathbf{c}_{i'}^H \boldsymbol{\eta}_{j'}(k'')|^2] = N_o^2 L_t^2 (1 + \varphi^2((k' - k'')T/Q) \delta_{ii'} \delta_{jj'}). \quad (\text{B.22})$$

Now consider the second term of (B.17) (ignoring the non-random part at this moment). Expanding it out and applying (B.10) and (B.11), we note that it is zero except for the case  $i = i'$  and  $j = j'$ , in which case we have

$$\begin{aligned} & 4\mathbb{E} \left\{ \Re \left[ h_{ij} (\mathbf{c}_i^H \boldsymbol{\eta}_j(k'))^* \right] \Re \left[ h_{ij} (\mathbf{c}_i^H \boldsymbol{\eta}_j(k''))^* \right] \right\} \\ &= \mathbb{E} \left[ h_{ij} (\mathbf{c}_i^H \boldsymbol{\eta}_j(k'))^* h_{ij}^* (\mathbf{c}_i^H \boldsymbol{\eta}_j(k'')) \right] + \mathbb{E} \left[ h_{ij}^* (\mathbf{c}_i^H \boldsymbol{\eta}_j(k')) h_{ij} (\mathbf{c}_i^H \boldsymbol{\eta}_j(k''))^* \right] \\ &= N_o \mathbb{E} [|h_{ij}|^2] \sum_{l_1=0}^{L_t-1} \sum_{l_2=0}^{L_t-1} c_i(l_1) c_i^*(l_2) \varphi((l_2 - l_1)T + (k'' - k')T/Q) \\ &\quad + N_o \mathbb{E} [|h_{ij}|^2] \sum_{l_1=0}^{L_t-1} \sum_{l_2=0}^{L_t-1} c_i^*(l_1) c_i(l_2) \varphi((l_1 - l_2)T + (k' - k'')T/Q) \\ &= 2N_o L_t \varphi((k' - k'')T/Q). \end{aligned} \quad (\text{B.23})$$

Plugging (B.22) and (B.23) back into (B.17), we obtain

$$\begin{aligned} \mathbb{E}[v_{ij}(k') v_{i'j'}(k'')] &= N_o^2 L_t^2 + N_o^2 L_t^2 \varphi^2((k' - k'')T/Q) \delta_{ii'} \delta_{jj'} \\ &\quad + 2N_o L_t W^2 p(k'T/Q - \epsilon_o T) p(k''T/Q - \epsilon_o T) \varphi((k' - k'')T/Q) \delta_{ii'} \delta_{jj'}. \end{aligned} \quad (\text{B.24})$$

Finally,

$$\begin{aligned} & \mathbb{E}[v(k') v(k'')] \\ &= \sum_{j=1}^M \sum_{i=1}^N \mathbb{E}[v_{ij}(k') v_{ij}(k'')] + \sum_{j=1}^M \sum_{i=1}^N \sum_{\substack{j'=1 \\ j' \neq j}}^M \sum_{\substack{i'=1 \\ i' \neq i}}^N \mathbb{E}[v_{ij}(k') v_{i'j'}(k'')] \\ &= MN(MN + \varphi^2((k' - k'')T/Q)) N_o^2 L_t^2 \\ &\quad + 2MNN_o L_t W^2 p(k'T/Q - \epsilon_o T) p(k''T/Q - \epsilon_o T) \varphi((k' - k'')T/Q). \end{aligned} \quad (\text{B.25})$$

Now, return to the calculation of  $\mathbb{E}[\phi^2]$  and  $\mathbb{E}[\phi\phi^*]$ . From (B.5), we have

$$\begin{aligned} \mathbb{E}[\phi^2] &= W^4 \mathbb{E}[h^2] e^{j4\pi\epsilon_o} \left( \sum_{k=0}^{Q-1} p^2(kT/Q - \epsilon_o T) e^{-j2\pi k/Q} \right)^2 \\ &\quad + e^{j4\pi\epsilon_o} \mathbb{E} \left[ \left( \sum_{k=0}^{Q-1} v(k) e^{-j2\pi k/Q} \right)^2 \right]. \end{aligned} \quad (\text{B.26})$$

The cross terms vanish since  $h$  and  $v(k)$  are uncorrelated and  $\mathbb{E}[v(k)]$  is a constant and independent of  $k$ . Note that  $h$  is a central chi-square random variable with  $2MN$  degree of freedom and variance in each dimension equals 0.5, so  $\mathbb{E}[h^2] = MN(1+MN)$ .

Using (B.25), it can be easily shown that

$$\mathbb{E} \left[ \left( \sum_{k=0}^{Q-1} v(k) e^{-j2\pi k/Q} \right)^2 \right] = 2MNN_o L_t W^2 \Xi_{SN} + MNN_o^2 L_t^2 \Xi_{NN}, \quad (\text{B.27})$$

where

$$\begin{aligned} \Xi_{SN} &\triangleq \sum_{k'=0}^{Q-1} \sum_{k''=0}^{Q-1} p(k'T/Q - \epsilon_o T) p(k''T/Q - \epsilon_o T) \\ &\quad \cdot \varphi((k' - k'')T/Q) e^{-j2\pi k''/Q} e^{-j2\pi k'/Q} \end{aligned} \quad (\text{B.28})$$

$$\Xi_{NN} \triangleq \sum_{k'=0}^{Q-1} \sum_{k''=0}^{Q-1} \varphi^2((k' - k'')T/Q) e^{-j2\pi k''/Q} e^{-j2\pi k'/Q}. \quad (\text{B.29})$$

Plugging (B.27) back into (B.26), the expression for  $\mathbb{E}[\phi^2]$  can be obtained and is given by (4.29). A similar procedure can be applied to obtain the expression for  $\mathbb{E}[\phi\phi^*]$  in (4.30).

## APPENDIX C

## PROOF OF (5.33), (5.47) AND (5.53)

Proof of (5.33)

From the expression of  $\text{MCRB}_{DA}$  in (5.32), only the product inside the  $\text{tr}(\cdot)$  operator depends on  $\mathbf{Z}$ , therefore the problem of finding optimal training sequence is equivalent to maximizing  $\text{tr}(\tilde{\mathbf{Z}}^H \tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o} \tilde{\mathbf{Z}} \Phi_T)$  with respect to  $\mathbf{Z}$  with the constraints that i) the columns of  $\mathbf{Z}$  have to be independent of each other and ii)  $[\mathbf{Z}^H \mathbf{Z}]_{ii} = L_o + 2L_g$  for  $i = 1, \dots, N$ . The first constraint is for the  $\text{ML}_{DA}$  to hold and has been mentioned before. The second constraint is the power constraint and we assume that the training sequence has average unit energy on each data bit. Now, consider the eigenvector decomposition  $\tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o} = \mathbf{U}_D \Sigma_D \mathbf{U}_D^H$ , where  $\Sigma_D$  is a diagonal matrix with the eigenvalues of  $\tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o}$  located on the diagonal and  $\mathbf{U}_D$  is the unitary matrix containing all the corresponding eigenvectors as columns. Similarly, express  $\Phi_T = \mathbf{U}_T \Sigma_T \mathbf{U}_T^H$ . Then

$$\begin{aligned}
\text{tr}(\tilde{\mathbf{Z}}^H \tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o} \tilde{\mathbf{Z}} \Phi_T) &= \text{tr}(\tilde{\mathbf{Z}}^H \mathbf{U}_D \Sigma_D \mathbf{U}_D^H \tilde{\mathbf{Z}} \mathbf{U}_T \Sigma_T \mathbf{U}_T^H) \\
&= \text{tr}(\sqrt{\Sigma_T}^H \mathbf{U}_T^H \tilde{\mathbf{Z}}^H \mathbf{U}_D \Sigma_D \mathbf{U}_D^H \tilde{\mathbf{Z}} \mathbf{U}_T \sqrt{\Sigma_T}) \\
&= \text{tr}(\Xi^H \Sigma_D \Xi) \\
&= \sum_{i=1}^N [\Xi_{:,i}]^H \Sigma_D \Xi_{:,i}, \tag{C.1}
\end{aligned}$$

where  $\Xi \triangleq \mathbf{U}_D^H \tilde{\mathbf{Z}} \mathbf{U}_T \sqrt{\Sigma_T}$ . Note that, if we set  $\mathbf{Z}^H \mathbf{Z} = (L_o + 2L_g) \mathbf{I}_N$  (this is a sufficient condition that make the two constraints mentioned earlier satisfied), then the columns of  $\Xi$  are orthogonal to each other (since  $\Xi^H \Xi = \sqrt{\Sigma_T}^H \mathbf{U}_T^H \tilde{\mathbf{Z}}^H \mathbf{U}_D \mathbf{U}_D^H \tilde{\mathbf{Z}} \mathbf{U}_T \sqrt{\Sigma_T} = (L_o + 2L_g) \Sigma_T / N$ ). Therefore, by confining the training sequences to be orthogonal,

the problem then becomes to maximize  $[\boldsymbol{\Xi}_{:,i}]^H \boldsymbol{\Sigma}_D \boldsymbol{\Xi}_{:,i}$  with respect to  $\boldsymbol{\Xi}_{:,i}$  for each  $i$  with the constraints that  $[\boldsymbol{\Xi}_{:,i}]^H \boldsymbol{\Xi}_{:,i} = (L_o + 2L_g)[\boldsymbol{\Sigma}_T]_{ii}/N$  and  $[\boldsymbol{\Xi}_{:,i}]^H \boldsymbol{\Xi}_{:,j} = 0$  for  $j = 1, \dots, i - 1$ .

It is well known that for a hermitian matrix  $\mathbf{R}$ , the vector  $\mathbf{u}$  that maximizes  $\mathbf{u}^H \mathbf{R} \mathbf{u}$  subject to the constraints that  $\|\mathbf{u}\| = 1$  and  $\mathbf{u}^H \mathbf{u}_i = 0$  for  $i = 1, 2, \dots, k - 1$ , where  $\mathbf{u}_i$  is the eigenvector corresponding to the  $i^{\text{th}}$  largest eigenvalue of  $\mathbf{R}$ , is  $\mathbf{u}_k$  [67]. Setting  $\mathbf{R} = \boldsymbol{\Sigma}_D$  and with the proper power constraints, it is not difficult to see that  $\boldsymbol{\Xi}_{:,i}$  is the eigenvector corresponding to the  $i^{\text{th}}$  largest eigenvalue of  $\boldsymbol{\Sigma}_D$  scaled by the energy factor  $\sqrt{(L_o + 2L_g)[\boldsymbol{\Sigma}_T]_{ii}/N}$ . Since  $\boldsymbol{\Sigma}_D$  is a diagonal matrix, the  $i^{\text{th}}$  eigenvector is a vector of length  $L_o + 2L_g$  with a one at the  $i^{\text{th}}$  position and zero at other positions. Therefore,

$$\boldsymbol{\Xi} = \sqrt{(L_o + 2L_g)/N} \begin{bmatrix} \sqrt{\boldsymbol{\Sigma}_T} \\ \mathbf{0}_{(L_o+2L_g-N) \times N} \end{bmatrix}, \quad (\text{C.2})$$

where  $\mathbf{0}_{(L_o+2L_g-N) \times N}$  is an all zero matrix with dimensions  $(L_o + 2L_g - N) \times N$ . With  $\boldsymbol{\Xi} = \mathbf{U}_D^H \tilde{\mathbf{Z}} \mathbf{U}_T \sqrt{\boldsymbol{\Sigma}_T}$ , we have

$$\mathbf{Z} = \sqrt{(L_o + 2L_g)} \mathbf{U}_D \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}_{(L_o+2L_g-N) \times N} \end{bmatrix} \mathbf{U}_T^H = \sqrt{(L_o + 2L_g)} \tilde{\mathbf{U}}(\varepsilon_o) \mathbf{U}_T^H, \quad (\text{C.3})$$

where  $\tilde{\mathbf{U}}(\varepsilon_o)$  is the matrix containing the  $N$  eigenvectors corresponding to the  $N$  largest eigenvalues of  $\tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o}$  as columns.

Proof of (5.47)

First note that  $\mathbf{\Gamma}_x$  can be rewritten in the following form

$$\begin{aligned}
\mathbf{\Gamma}_x &\triangleq \mathbb{E}[\text{vec}(\mathbf{Z}\sqrt{\mathbf{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T)\text{vec}(\mathbf{Z}\sqrt{\mathbf{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T)^H] \\
&= \mathbb{E}[(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})\text{vec}(\sqrt{\mathbf{\Phi}_T})\text{vec}(\sqrt{\mathbf{\Phi}_T})^H(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})^H] \\
&= \mathbb{E}[(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})\mathbf{\Upsilon}(\mathbf{H}_{\text{i.i.d.}}^H \otimes \mathbf{Z}^H)], \tag{C.4}
\end{aligned}$$

where  $\mathbf{\Upsilon} \triangleq \text{vec}(\sqrt{\mathbf{\Phi}_T})\text{vec}(\sqrt{\mathbf{\Phi}_T})^H$ . The  $(i, j)^{th}$  element of  $\mathbf{\Gamma}_x$  is given by

$$\begin{aligned}
[\mathbf{\Gamma}_x]_{ij} &= \mathbb{E}[(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})_{i,:} \mathbf{\Upsilon}(\mathbf{H}_{\text{i.i.d.}}^H \otimes \mathbf{Z}^H)_{:,j}] \\
&= \mathbb{E}[\text{tr}(\mathbf{\Upsilon}(\mathbf{H}_{\text{i.i.d.}}^H \otimes \mathbf{Z}^H)_{:,j}(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})_{i,:})] \\
&= \text{tr}(\mathbf{\Upsilon} \mathbb{E}[(\mathbf{H}_{\text{i.i.d.}}^H \otimes \mathbf{Z}^H)_{:,j}(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})_{i,:}]) \tag{C.5}
\end{aligned}$$

with  $i, j = 0, 1, \dots, M(L_o + 2L_g)$ .

Let  $i = i_q(L_o + 2L_g) + i_r$  and  $j = j_q(L_o + 2L_g) + j_r$  such that  $i_q, j_q \in \{0, 1, \dots, M-1\}$  and  $i_r, j_r \in \{1, \dots, L_o + 2L_g\}$  are the quotients and remainders of divisions  $i/(L_o + 2L_g)$  and  $j/(L_o + 2L_g)$ , respectively. Also:

$$\begin{aligned}
\mathbb{E}[(\mathbf{H}_{\text{i.i.d.}}^H \otimes \mathbf{Z}^H)_{:,j}(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})_{i,:}] &= \mathbb{E}[(\mathbf{H}_{\text{i.i.d.}}^H)_{:,j_q} \otimes (\mathbf{Z}^H)_{:,j_r} ((\mathbf{H}_{\text{i.i.d.}})_{i_q,:} \otimes (\mathbf{Z})_{i_r,:})] \\
&= \mathbb{E}[(\mathbf{H}_{\text{i.i.d.}}^*)_{j_q,:}^T (\mathbf{H}_{\text{i.i.d.}})_{i_q,:}] \otimes \mathbb{E}[(\mathbf{Z}^*)_{j_r,:}^T (\mathbf{Z})_{i_r,:}] \\
&= \mathbf{I}_N \delta_{i_q j_q} \otimes \mathbf{\Gamma}_z(j_r - i_r), \tag{C.6}
\end{aligned}$$

where  $\mathbf{\Gamma}_z(j_r - i_r) \triangleq \mathbb{E}[(\mathbf{Z}^*)_{j_r,:}^T (\mathbf{Z})_{i_r,:}]$  is the average cross-correlation matrix of the symbols transmitted with the time index difference  $j_r - i_r$ . Note that  $\mathbb{E}[(\mathbf{Z}^*)_{j_r,:}^T (\mathbf{Z})_{i_r,:}]$  depends only on the time index difference but not on the absolute time index since in the NDA case we never know where the observation begins, the average cross-correlation between time indices 1 and 3 would be the same as that for time indices

5 and 7. Putting (C.6) into (C.5), we obtain:

$$[\mathbf{\Gamma}_\mathbf{x}]_{ij} = \delta_{i_q j_q} \text{tr}(\mathbf{\Upsilon}(\mathbf{I}_N \otimes \mathbf{\Gamma}_\mathbf{z}(j_r - i_r))), \quad (\text{C.7})$$

implying that

$$\mathbf{\Gamma}_\mathbf{x} = \mathbf{I}_M \otimes \mathbf{\Psi}, \quad (\text{C.8})$$

where  $\mathbf{\Psi}$  is a Hermitian, Toeplitz matrix with  $[\mathbf{\Psi}]_{ij} = \text{tr}(\mathbf{\Upsilon}(\mathbf{I}_N \otimes \mathbf{\Gamma}_\mathbf{z}(j - i)))$ . Note that  $[\mathbf{\Psi}]_{ij}$  can be simplified as

$$\begin{aligned} [\mathbf{\Psi}]_{ij} &= \text{tr}(\text{vec}(\sqrt{\mathbf{\Phi}_T}) \text{vec}(\sqrt{\mathbf{\Phi}_T})^H (\mathbf{I}_N \otimes \mathbf{\Gamma}_\mathbf{z}(j - i))) \\ &= \text{tr}(\text{vec}(\sqrt{\mathbf{\Phi}_T})^H (\mathbf{I}_N \otimes \mathbf{\Gamma}_\mathbf{z}(j - i)) \text{vec}(\sqrt{\mathbf{\Phi}_T})) \\ &= \text{tr}((\sqrt{\mathbf{\Phi}_T})^H \mathbf{\Gamma}_\mathbf{z}(j - i) (\sqrt{\mathbf{\Phi}_T})) \\ &= \text{tr}(\mathbf{\Gamma}_\mathbf{z}(j - i) \mathbf{\Phi}_T). \end{aligned} \quad (\text{C.9})$$

### Proof of (5.53)

First note that the observation interval usually involves more than one independent space-time encoded block, each given by the form (5.52), therefore  $\mathbf{\Gamma}_\mathbf{z}(j - i) = \mathbf{0}_N$  for  $|j - i| \geq s$ . Furthermore, since  $\mathbf{\Gamma}_\mathbf{z}(j - i) = \mathbf{\Gamma}_\mathbf{z}^*(i - j)$ , it is sufficient to concentrate on  $\mathbf{\Gamma}_\mathbf{z}(j - i)$ , for  $j - i = \ell$  with  $\ell = 0, 1, \dots, s - 1$ ,

$$\mathbf{\Gamma}_\mathbf{z}(\ell) = \frac{1}{s} \sum_{n=1}^{s-\ell} \mathbb{E}[(\mathcal{G}_{n+\ell,:})^H (\mathcal{G}_{n,:})] \quad (\text{C.10})$$

where the factor  $1/s$  exists because in NDA estimation, the probability that the observation start at a particular row of the matrix  $\mathcal{G}$  is  $1/s$ . Putting (5.52) into

(C.10), we obtain:

$$\begin{aligned}
& \mathbf{\Gamma}_{\mathbf{z}}(\ell) \\
&= \frac{1}{s} \sum_{n=1}^{s-\ell} \mathbb{E} \left[ \left( \sum_{k=1}^{rs} \Re(b_k) \mathbf{X}_k + j \sum_{k=1}^{rs} \Im(b_k) \mathbf{Y}_k \right)_{n+\ell,:}^H \left( \sum_{k'=1}^{rs} \Re(b_{k'}) \mathbf{X}_{k'} + j \sum_{k'=1}^{rs} \Im(b_{k'}) \mathbf{Y}_{k'} \right)_{n,:} \right] \\
&= \frac{1}{s} \sum_{n=1}^{s-\ell} \left( \sum_{k=1}^{rs} \mathbb{E} [\Re(b_k) \Re(b_k)] [\mathbf{X}_k]_{n+\ell,:}^T [\mathbf{X}_k]_{n,:} + \sum_{k=1}^{rs} \mathbb{E} [\Im(b_k) \Im(b_k)] [\mathbf{Y}_k]_{n+\ell,:}^T [\mathbf{Y}_k]_{n,:} \right),
\end{aligned} \tag{C.11}$$

where we have used the i.i.d. property of  $b_k$ ,  $\mathbb{E}[\Re(b_k) \Re(b_{k'})] = 0$ ,  $\mathbb{E}[\Im(b_k) \Im(b_{k'})] = 0$  for  $k \neq k'$  and  $\mathbb{E}[\Re(b_k) \Im(b_{k'})] = 0$  for all combination of  $k$  and  $k'$ . Further note that,  $\mathbb{E}[\Re(b_k) \Re(b_k)] = \mathbb{E}[\Im(b_k) \Im(b_k)] = 1/2$ , then we have for  $j - i = \ell$  with  $\ell = 0, 1, \dots, s - 1$ ,

$$\mathbf{\Gamma}_{\mathbf{z}}(j - i) = \frac{1}{2s} \sum_{n=1}^{s-\ell} \left( \sum_{k=1}^{rs} [\mathbf{X}_k]_{n+\ell,:}^T [\mathbf{X}_k]_{n,:} + \sum_{k=1}^{rs} [\mathbf{Y}_k]_{n+\ell,:}^T [\mathbf{Y}_k]_{n,:} \right). \tag{C.12}$$

Finally, note that since  $\mathbf{X}_k$  and  $\mathbf{Y}_k$  are real-valued,  $\mathbf{\Gamma}_{\mathbf{z}}(j - i)$  would also be real-valued and  $\mathbf{\Gamma}_{\mathbf{z}}(j - i) = \mathbf{\Gamma}_{\mathbf{z}}(i - j)$ . Therefore, it can be concluded that (C.12) is true for  $|j - i| = \ell$  ( $\ell = 0, 1, \dots, s - 1$ ).

## APPENDIX D

## PROOF OF (6.37)

The first term of (6.33) can be easily obtained by plugging (6.34) and (6.35) into (6.32) followed by taking logarithm, and it gives

$$\ln p(\mathbf{r}_{n_1}; \hat{\boldsymbol{\theta}}, i, L) = N \ln N - N \ln \pi - N \ln \|\mathbf{r}_{n_1} - \mathbf{B}_i(\mathbf{B}_i^H \mathbf{B}_i)^{-1} \mathbf{B}_i^H \mathbf{r}_{n_1}\|^2 - N. \quad (\text{D.1})$$

Consider the second term of (6.33). Performing the differentiations in (6.36), we have

$$\mathbf{I}(\boldsymbol{\theta}|i, L) = \frac{1}{\sigma^2} \begin{bmatrix} 2\Re(\mathbf{B}_i^H \mathbf{B}_i) & -2\Im(\mathbf{B}_i^H \mathbf{B}_i) & 0 \\ 2\Im(\mathbf{B}_i^H \mathbf{B}_i) & 2\Re(\mathbf{B}_i^H \mathbf{B}_i) & 0 \\ 0 & 0 & 1/\sigma^2 \end{bmatrix}. \quad (\text{D.2})$$

It follows that

$$\det(\mathbf{I}(\boldsymbol{\theta}|i, L)) = \frac{2^{2L}}{\sigma^{2(2L+2)}} \det \left( \begin{bmatrix} \Re(\mathbf{B}_i^H \mathbf{B}_i) & -\Im(\mathbf{B}_i^H \mathbf{B}_i) \\ \Im(\mathbf{B}_i^H \mathbf{B}_i) & \Re(\mathbf{B}_i^H \mathbf{B}_i) \end{bmatrix} \right). \quad (\text{D.3})$$

Using the result

$$\det \left( \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \right) = \det(A_{11}) \det(A_{22} - A_{21}A_{11}^{-1}A_{12}), \quad (\text{D.4})$$

we note that for any square matrix  $\boldsymbol{\Sigma}$ ,

$$\det \left( \begin{bmatrix} \Re(\boldsymbol{\Sigma}) & -\Im(\boldsymbol{\Sigma}) \\ \Im(\boldsymbol{\Sigma}) & \Re(\boldsymbol{\Sigma}) \end{bmatrix} \right) = \det(\Re(\boldsymbol{\Sigma})) \det(\Re(\boldsymbol{\Sigma}) + \Im(\boldsymbol{\Sigma})\Re(\boldsymbol{\Sigma})^{-1}\Im(\boldsymbol{\Sigma})). \quad (\text{D.5})$$

Since

$$\begin{aligned}
& \Re(\boldsymbol{\Sigma}) + \Im(\boldsymbol{\Sigma})\Re(\boldsymbol{\Sigma})^{-1}\Im(\boldsymbol{\Sigma}) \\
&= (\Re(\boldsymbol{\Sigma}) - j\Im(\boldsymbol{\Sigma}))\Re(\boldsymbol{\Sigma})^{-1}(\Re(\boldsymbol{\Sigma}) + j\Im(\boldsymbol{\Sigma})) \\
&= \boldsymbol{\Sigma}^*\Re(\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma},
\end{aligned} \tag{D.6}$$

it follows that

$$\det \left( \begin{bmatrix} \Re(\boldsymbol{\Sigma}) & -\Im(\boldsymbol{\Sigma}) \\ \Im(\boldsymbol{\Sigma}) & \Re(\boldsymbol{\Sigma}) \end{bmatrix} \right) = \frac{\det(\Re(\boldsymbol{\Sigma})) \det(\boldsymbol{\Sigma}^*) \det(\boldsymbol{\Sigma})}{\det(\Re(\boldsymbol{\Sigma}))} = (\det(\boldsymbol{\Sigma}))^2. \tag{D.7}$$

Plugging (D.7) into (D.3), and then taking the logarithm, we have

$$\ln \det(\mathbf{I}(\hat{\boldsymbol{\theta}}|i, L)) = 2L \ln 2 - 2(L+1) \ln \hat{\sigma}^2 + 2 \ln (\det(\mathbf{B}_i^H \mathbf{B}_i)). \tag{D.8}$$

Combining the results of (D.1) and (D.8), and dropping the terms irrelevant to optimization in (6.33), the generalized ML rule follows:

$$\psi_1(\mathbf{r}_{n_1}; i, L) = (-N + L + 1) \ln \|\mathbf{r}_{n_1} - \mathbf{B}_i(\mathbf{B}_i^H \mathbf{B}_i)^{-1} \mathbf{B}_i^H \mathbf{r}_{n_1}\|^2 - L \ln 2 - \ln (\det(\mathbf{B}_i^H \mathbf{B}_i)). \tag{D.9}$$

Note that the columns of  $\mathbf{B}_i$  are constructed by cyclically shifting the sequence  $\mathbf{b}_0 \triangleq [b_0 \ b_1 \ \dots \ b_{15}]^T$  with different numbers of shifts. Denote  $T_{lc}^m(\mathbf{b}_0)$  as the cyclic left shift of  $\mathbf{b}_0$  by  $m$  places (e.g.,  $T_{lc}(\mathbf{b}_0) = [b_1 \ \dots \ b_{15} \ b_0]^T$ ). It follows that the  $(l, k)^{th}$  element of  $\mathbf{B}_i^H \mathbf{B}_i$  is given by  $(l, k = 0, 1, \dots, L_e - 1)$

$$\begin{aligned}
[\mathbf{B}_i^H \mathbf{B}_i]_{l,k} &= [T_{lc}^{i+l}(\mathbf{b}_0)]^H [T_{lc}^{i+k}(\mathbf{b}_0)] \\
&= \mathbf{b}_0^H [T_{lc}^{k-l}(\mathbf{b}_0)] \\
&\triangleq R(k-l)
\end{aligned} \tag{D.10}$$

where  $R(\tau)$  is the periodic autocorrelation function of the sequence  $\mathbf{b}_0$  with relative offset  $\tau$  and is independent of  $i$ . Therefore, the matrix  $\mathbf{B}_i^H \mathbf{B}_i$  depends only on  $L$  and we can set  $i = 0$  in  $\mathbf{B}_i^H \mathbf{B}_i$ . With this result, the generalized ML rule simplifies to

$$\psi_1(\mathbf{r}_{n_1}; i, L) = (-N + L + 1) \ln \|\mathbf{r}_{n_1} - \mathbf{B}_i(\mathbf{B}_0^H \mathbf{B}_0)^{-1} \mathbf{B}_i^H \mathbf{r}_{n_1}\|^2 - \xi(L) \quad (\text{D.11})$$

where  $\xi(L) \triangleq L \ln 2 + \ln (\det(\mathbf{B}_0^H \mathbf{B}_0))$ .

## VITA

Yik Chung Wu obtained the B.Eng. (EEE) degree in 1998 and the M.S. degree, specializing in wireless communications, in 2001 from The University of Hong Kong (HKU). After graduating with a master's degree, he worked as a research assistant in the same university. He started his Ph.D. degree program in 2002 at Texas A&M University and obtained the Ph.D. degree in 2005. During his study at Texas A&M University, he was supported by the Croucher Foundation scholarship. His research interests are in the general area of signal processing for communication systems, and in particular receiver algorithm design, synchronization techniques, software radio and space-time processing. He has published one book chapter, 11 journal papers, 17 conference papers and has one U.S. patent pending on various topics of communications and signal processing.

He can be reached at the email address [wuyikchung@yahoo.com](mailto:wuyikchung@yahoo.com) or at the following permanent address:

2503, Ping Hay House, Tai Ping Estate, Sheung Shui, Hong Kong.