# Maximum Likelihood CFO Estimation for OFDM/OFDMA Systems with Periodic Preambles

Student: Hung-Tao Hsieh  Advisor: Dr. Wen-Rong Wu

Institute of Communications Engineering

National Chiao Tung University

July 1, 2011

# 具週期性前導序列之正交分頻多工/正交分頻多工存取系統中最大似然載波頻率偏移估計

研究生: 謝弘道　　　　　　　　　　　指導教授: 吳文榕 博士

國立交通大學

電信工程研究所博士班

## 摘要

正交分頻多工（orthogonal frequency division multiplexing；OFDM）是一多載波調變技術，以其高頻譜效率而廣為人知，目前此技術已被使用於許多無線通訊系統中。OFDM 之所以有高的頻譜使用效率是因為將頻譜切成多個子載波,且子載波間相互重疊且正交。然而當載波頻率偏移(carrier frequency offset; CFO)存在時,正交性會遭到破壞,造成子載波間的相互干擾而降低通訊的品質。因此在真正的數據傳輸前,我們必須先估計並補償 CFO。本論文旨在研究 OFDM 相關系統之 CFO 最大似然(Maximum likelihood; ML)估計,我們的系統假設在接收端可以接收到未知的週期性訊號。本論文分別就 OFDM、正交多頻分工存取(orthogonal frequency division multiple access; OFDMA)上行 (uplink)和合作式放大傳輸(amplify-and-forward; AF) OFDM 系統,提出其對應之 ML 解。

在傳統的正交分頻多工系統中,ML 解需計算相關矩陣(correlation matrix)的反矩陣,但隨著子載波數的增加,其複雜度將大到無法實現。為避免直接計算該反矩陣,本

論文提出一個新的 ML 法來直接求得 CFO 的閉合解(closed-form solution)，此法的優點是計算複雜度極低，該 ML 法亦可延伸至同時估計 CFO 和符元時間偏移(symbol timing offset)。為評估在 CFO 的估計效能，我們導出該系統之 CFO 的理論下限也就是 CRB (Cramér rao bound)。本論文接著探討交錯型(interleaved) OFDMA 系統上行中的 CFO 估計問題。由於系統特性使然，此系統接收的訊號一定是週期性的，因此無需額外的訓練訊號。跟 OFDM 系統一樣我們需計算一相關矩陣的反矩陣，但 OFDMA 系統因牽涉到多使用者的 CFO 估計，所以此反矩陣很難計算。我們因此提出使用級數展開法(series expansion) 將此反矩陣展開，並保留適當的低階項後，我們可以得到一個閉合解，然後以解根的方式解得 CF0。最後我們推導出相對應的 CRB，實驗發現無論是在傳統的 OFDM 系統或 OFDMA 系統中，本論文所提的 ML 解皆可逼近 CRB。

最後，本論文探討 AF-OFDM 系統中的 CFO 估計和功率分配問題，在此系統中雜訊為有色雜訊(color noise)，然而此特性並未在之前的研究中被討論，因此現有的 ML 解皆非最佳，現有的 CRB 也不適用。因為 CFO 的 ML 解變的相當複雜無法導出閉合解，本論文提出使用梯度下降法(gradient descent method)來求得 CFO 的 ML 解。此外，有色雜訊使得此系統中的 CRB 過於複雜以致於無法得到一簡單的閉合解。本論文做了一些近似來推導出 CRB 的閉合解。實驗發現此近似的 CRB 是準確的且我們所提出的 ML 解可以逼近此 CRB。我們接者提出兩個功率分配法將 CRB 最小化，並利用梯度下降法求得其解。模擬發現，本論文提出的功率分配法不但可以大幅改善 CFO 估計的準確性，也可提升了系統的訊雜比(signal-to-noise ratio)。

# Maximum Likelihood CFO Estimation for OFDM/OFDMA Systems with Periodic Preambles

Student:   Hung-Tao Hsieh                    Advisor:   Dr. Wen-Rong Wu

National Chiao Tung University

Institute of Communication Engineering

## Abstract

Orthogonal frequency division multiplexing (OFDM), being a multicarrier modulation technique, is well known for its high spectral efficiency and has been adopted in many wireless systems. The high efficiency of OFDM comes from the fact that the spectrums of its subcarriers are overlapped and orthogonal each other. However, when the carrier frequency offset (CFO) is present, the orthogonality between the subcarriers is lost and inter carrier interference (ICI) is induced causing performance degradation. As a result, CFO must be estimated and compensated before the actual transmission can be conducted. In this dissertation, we study the maximum-likelihood (ML) methods for CFO estimation in OFDM-based systems, assuming that unknown periodic received sequences are available at the receivers. Specifically, we solve the ML CFO estimation problems in conventional OFDM, orthogonal frequency division multiple access (OFDMA) uplink, and cooperative amplify-and-forward (AF) OFDM systems.

In conventional OFDM systems, the ML CFO estimator requires the inversion of an correla-

tion matrix. When the number of subcarriers is large, the computational complexity can become prohibitively high. We then develop a new ML method that can yield a closed-form solution without the inversion. The advantage of the proposed method is that the required computational complexity is low. The proposed method is further extended to a joint CFO and symbol timing offset (STO) estimation. Theoretical Cramér-Rao lower bounds (CRBs) are also derived to verify the optimality of the proposed approaches. We then investigate the CFO estimation problem in interleaved OFDMA uplink systems. Since the periodicity is inherent in OFDMA systems, no training sequences are required. As previously, there is an correlation matrix in the likelihood function to be inverted. Since multi-users are involved, the CFOs in the likelihood function become intractable after the matrix inversion. We propose using a series expansion method to express the inverted matrix. By properly truncating the expansion, we can obtain a closed-form expression, solve the optimum CFOs with a root-finding method, and derive the corresponding CRB. Simulations show that the performance of the proposed method can approach the CRB.

We finally consider the ML CFO estimation and the power allocation problem in cooperative AF-OFDM systems. In this scenario, the noise at the destination becomes colored. The colored-noise problem has not been considered before. Thus, the existing ML methods are not optimal and the existing CRBs are not valid. Since the likelihood function is a complicated function of the CFO, we then propose a gradient-descent method to solve the problem. The expression for the CRB for the CFO estimation in AF-OFDM systems is even more complicated and a simple solution cannot be obtained. We then propose an approximation method such that a closed-form solution can also be derived. Simulations show that the approximated CRB is accurate and the performance of the proposed gradient-descent method can approach the CRB. Minimizing the approximated CRB, we further propose two power allocation algorithms (PAA), implemented with constrained gradient-based method, for the source and relays. Simulations show that not only the performance of the CFO estimation is greatly enhanced, but also the signal-to-noise ratio (SNR) between source and destination is improved.
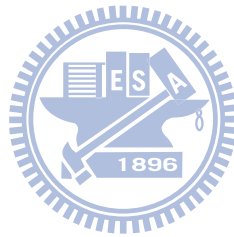
# Acknowledgements

During the Ph.D. program, I would like to show my gratitude to many people. First, I would like to thank my advisor, Prof. Wen-Rong Wu, for his kindly guidance. He spends a lot of time in discussing the problems I encounter in my research, providing valuable suggestions, and teaching me how to write technical papers. In addition to academic research, he provides lots of resources in improving our English. Besides regarding the professional area, he also guide me the attitude for human life. It is really helpful to me. Under his enthusiastic instruction, I have learned not only how to do a research or a project but also the optimistic attitude. At this moment, I have to say Prof. Wu is the key person whom I am most grateful to in my studying life.

Second, I am grateful to all the members in Prof. Wu's lab. for their valuable discussions and help in academic research including Chun-Fang Lee, Chao-Yuan Hsu, Fan-Shuo Tseng, Chun-Tao Lin, Bruice, and Chi-Han Lee. Especially, I deeply appreciate Chao-Yuan Hsu and Chun-Tao Lin of their encouragement and suggestions during the period of my Ph.D. program. Also, I would like to thank all my friends who ever encourage or help me. Furthermore, thanks the junior members in our lab., Sandy a pretty girl, Binsung, Minchung and Yu-Lang who had hold a tiny exercise club for me.

Finally, I would like to show my deep gratitude to my family, my best-loved wife, my lovely son, my cute sister, my dear parents, my parents in law, and my grandparents. Thank you all for your economic and spirit support in the period. Without your supports and encouragements, I

cannot finish my Ph.D. program. Besides, to my upcoming little daughter, I am so glad to have you in the end of pursuing my Ph.D degree.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

**O**Rthogonal frequency division multiplexing (OFDM) is known to be a promising modulation technique [1]. It can provide high transmission data rate, resist multi-path channel fading, achieve high spectrum efficiency, and have efficient implementation architectures [2]. This technique is suitable for the high-speed wireless communication. As a matter of fact, many wireless standards such as Wi-Max, IEEE802.11a, DVB, LTE [3], have adopted the OFDM modulation. The main idea behind OFDM is to split a wide band channel into narrow band subchannels such that the subcarrier in each subchannel can experience flat-fading channel effect. The high efficiency of the OFDM system comes from the fact that the spectrums of the subcarriers are overlapped and orthogonal each other. However, if the carrier frequency offset (CFO) is present [4], the orthogonality will be destroyed and inter carrier interference(ICI) is induced [5] [6] [7], degrading the performance of OFDM systems [8]. CFO arises due to the Doppler shift and the frequency mismatch between the transmitter and the receiver. It has been shown that the bit error rate (BER) of a OFDM system [9] is proportional to the magnitude of CFO. In real-world applications, CFO must be estimated and compensated before the actual transmission can be conducted. CFO estimation has been an important and active research subject in past decades.

A well-known scheme for CFO estimation is to let the transmitter emit a preamble with

repeated periods such that the receiver can use a correlator conducting the CFO estimation. The correlating operation is simple but not optimal. In this dissertation, we investigate the maximum likelihood (ML) CFO estimation for OFDM-based systems with periodic signals. We assume that the received signal for CFO estimation is periodic but unknown. Also, the statistical channel state information (CSI) is available. In the proposed algorithms, only do the periodicity and CSI be used. Thus, the proposed algorithms can be applied in the systems with periodic preambles or those with inherent periodical data structure.

## § 1.1   Conventional and Proposed Methods

Many works have considered the CFO estimation in OFDM-based systems, including standard OFDM systems [10]- [27], orthogonal frequency division multiple access (OFDMA) [28]- [35] systems, and cooperative OFDM systems [36]- [38]. The estimation methods can be classified into two categories: methods using or not using training sequences. Methods using training sequences insert a known preamble in front of each data packet, facilitating CFO estimation at the receiver [28]- [31]. For this kind of methods, the transmitted data should be known at the receiver. Methods without using training sequences, also referred to blind methods, manipulate some priori information such that the CFO can be estimated. For example, the cyclic prefix (CP) is known to be periodic in an OFDM symbol. In this dissertation, we consider the CFO estimation with a periodic receive sequence. We will solve the ML CFO estimation problems in OFDM systems, OFDMA uplink systems, and cooperative amplify-and-forward (AF) OFDM systems. In the OFDM and AF-OFDM systems, the proposed methods require periodic preambles and they are training-based while in the OFDMA system, the receive signal is inherently periodic and the proposed method is blind.

## § 1.1.1 *CFO Estimation in OFDM Systems*

CFO estimation for OFDM systems has been proposed in [10]- [27]. The methods in [10]- [16] exploit the periodic structure of CPs to accomplish the estimation task. In [18]- [27], a periodic preamble or pilot-symbol is inserted in front of each data packet such that it can be easily used by the receiver to conduct CFO estimation. CFO usually consists of a fractional part and an integer part. Most researchers focus on how to estimate the fractional part, which is also the focus of this dissertation. Integer part estimation was specifically considered in [18] and [19]. An ML CFO estimator using a preamble with two identical pilot symbols was first proposed in [20]. Using the same periodic preamble and taking null subcarriers into consideration, [21] proposes a method that is able to estimate both fractional and integer CFOs. In order to avoid the extra overhead required in [21], [22] introduces a preamble composed of two OFDM symbols: the first one has two identical periods used to estimate the fractional CFO and symbol timing offset (STO), and the second one has a special correlation with the first one used to estimate the integer CFO. To improve the performance, [23] extends the scenario to treat preambles with periodicity of greater than two. Using the approach in [23], one can remove the second pilot symbol required in [22]. As an improved version, [24] proposes a CFO estimation based on the best linear unbiased estimation (BLUE) principle. Note that [23] and [24] still use the same STO estimator as that in [22]. When the number of periods is greater than two, the method in [20] is no longer optimal. An ML CFO estimator for this problem was proposed in [25]. However, the required computational complexity is high. In order to alleviate this problem, a low complexity approach was then proposed in [26]. Another simplified algorithm was also proposed in [27]. However, due to excessive approximation in the likelihood function, the performance of the CFO estimation in [27] does not approach the Cramér-Rao bound (CRB) [41].

We consider a preamble with more than two periods. The ML CFO estimation for the system has been considered in [26]. The method in [26] is essentially a two-step approach; it first estimates the received preamble with a least-squares (LS) method, and then maximizes the cor-

responding likelihood function. In addition to regular computations, this method requires an extra procedure to solve the roots of the derivative of the likelihood function. Thus, its computational complexity is higher, and the cost for real-world implementations is also increased. We then develop a new ML method that solves the likelihood function directly. Our method yields a closed-form ML solution, and the root-finding procedure is not required. As a result, the computational complexity and the implementation cost are lower than those in [26], while the performance of the proposed method is either equal to or better than that in [26]. The proposed method is further extended to solve a joint CFO/STO estimation problem. The corresponding CRBs for CFO estimations are also derived. Note that the performance bound for STO estimation has not previously been addressed in the literature.

## § 1.1.2 *CFO Estimation in OFDMA Systems*

OFDMA, emerging as a promising technology for next generation broadband wireless network [3], has received considerable amount of research interest recently [28]- [35]. An appealing feature of OFDMA is that the transmission signals of different users are orthogonal and multiple access interference (MAI) can be avoided. However, if carrier frequency offsets (CFOs) between the transmitters and the receivers are not properly estimated and compensated, the orthogonality will be destroyed and ICI/MAI will arise. In the OFDMA downlink systems, the signals for different users are multiplexed by the same transmitter, and the receiver of each user can estimate and compensate its own CFO easily. In such a scenario, methods for CFO estimation in single-user OFDM systems [10] can be directly applied. However, in OFDMA uplink systems, all users's CFOs have to be simultaneously estimated at the base-station (BS) receiver, which is considered to be a more challenging problem.

In OFDMA uplink systems, methods without using training sequences, also referred to blind methods, manipulate the subcarrier assignment scheme such that the CFO for each user can be individually or jointly estimated [32]- [35] at the receiver. With proper subcarrier assignment, subband-based or interleaved-based estimators can be applied. The subband-based CFO esti-

mators ( [32], [33]) require that each user is assigned with some consecutive subcarriers and the subcarrier sets for different users are well separated in the frequency domain. With the scheme, a filter bank can be used to extract each user's signal, and then the conventional CFO estimation methods can be exploited. In interleave-based OFDMA systems, each user's time-domain signal is known to be periodical. With the property, the CFOs for all users can be jointly estimated by the multiple signal classification (MUSIC) [34] or the estimation of signal parameters via rotational invariance technique (ESPRIT) methods [35]. Although the computational complexity of these methods is low, the CRB cannot be achieved and the solutions are not optimal.

To solve the problem mentioned above, we then investigate the blind CFO estimation problem in the interleaved OFDMA uplink system. Our objective is to develop a low complexity ML CFO estimation method. The main obstacle in the ML method is that there is an inverted correlation matrix in the likelihood function. The CFOs in the likelihood function become intractable after the matrix inversion. Using the matrix inversion lemma, we first transform the correlation matrix into a matrix with smaller size. Then, we express the matrix with a series expansion. By properly truncating the expansion, we can obtain a closed-form expression and solve the optimum CFOs with a root-finding method. Also, the CRB can be derived. Simulations show that the performance of the proposed method can approach the CRB. The computational complexity of the proposed algorithm is as low as that of ESPRIT.

### § 1.1.3 *CFO Estimation in AF-OFDM Systems*

More recently, there is an growing interest in wireless communication systems employing cooperative relay networks [57]. The cooperative relaying system forms a virtual antenna array and allows wireless devices to achieve higher transmit diversity. The use of relays can lead to expanded coverage, system-wide power saving, and better immunity against signal fading. Two relaying protocol are well known, namely, AF and decode-and-forward (DF). In AF, the relays only retransmit a linearly amplified version of the received signal (from the source) to the destination. In DF, the relays, on the other hand, demodulate the received signal and remodulate

and retransmit the resultant signal.

Incorporating the advantages of multicarrier transmission, OFDM-based cooperative systems have been proposed as an emerging transmit technique for future wireless networks [58] [59]. Similar to OFDM systems, the orthogonality in OFDM-based cooperative systems is also critical. How to reduce the impact of CFO remains an important research subject [36]- [38]. In [36], the Alamouti code was employed to mitigate the CFO-induced ICI in AF OFDM-based cooperative systems (AF-OFDM). CFO estimation in DF OFDM-based cooperative systems (DF-OFDM) was considered in [37]. In their approach, a training sequence is designed for a relay and the sequence is transmitted on a set of subcarriers. As a result, the system can be considered as an OFDMA system in its training phase. In [38], a scheme, allocating the transmitted power on the source and a relay with the largest composite channel gain in AF-OFDMA systems, is proposed. The objective of the power allocation is to minimize the CRB in the CFO estimation. It has been shown that the optimum power allocation can effectively reduces the variance of the CFO estimation [38]. The BER performance of the whole system can then be improved accordingly [40].

One problem with the approach in [38] is that the CFO estimation and the corresponding CRB is derived under the assumption of white Gaussian noise. In reality, however, the noise observed at the destination is colored. The other problem is that multi-relay systems are not analyzed. To solve the problems, we then consider the ML CFO estimation and the power allocation problem in multi-relay AF-OFDM systems. The distinct feature of the proposed method is that the noise at the destination is considered as colored. In the colored scenario, the inversion operation in the correlation matrix becomes much more complex and a closed-form solution is very difficult to obtain. Due to this difficulty, we then propose a gradient-based method to solve the ML problem. For this scenario, the CRB derived for the white noise environment is not valid anymore. Let the CRB for the CFO estimation in AF-OFDM systems be denoted by $CRB_r$. The expression of $CRB_r$ contains an expectation operation on the source-to-relay and rely-to-destination channels [9]. To obtain $CRB_r$, we have to conduct Monte

Carlo simulations which are not efficient for our optimization problem. We then approximate the correlation matrix with a tri-diagonal matrix and propose an approximation method such that a closed-form solution can be derived for $CRB_r$. The approximated CRB, denoted as, $CRB_a$, is a function of the expected CSI. As a result, the value of $CRB_a$ can be evaluated efficiently and the corresponded power allocation scheme can be derived. When the noise at the destination is assumed to be white, $CRB_a$ is degenerated to the conventional CRB considered in the white noise environment which is denoted as $CRB_w$. We then consider the power allocation problem in AF-OFDM systems, allocating power for the source and the relays. As that in [38], our objective is to minimize the CRB. Using $CRB_w$ and $CRB_a$, we propose two power allocation algorithms (PAAs), implemented with constrained gradient-descent methods, to solve the problem. Simulations show that the proposed PAAs not only significantly improves the performance of the CFO estimation, but also greatly enhance the received SNR improving the overall system capacity.
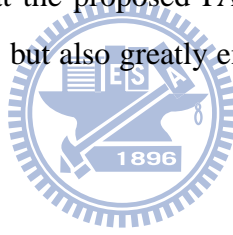
## § 1.2   Organization of the Dissertation

This dissertation is organized as follows. In Chapter 2, we propose a new ML method that solves the likelihood function directly for the CFO estimation problem in OFDM systems. In Chapter 3, we propose a blind ML CFO estimation algorithm in interleaved OFDMA uplink systems. In Chapter 4, we consider CFO estimation in AF-OFDM systems. We propose a method to solve the ML solution and two PAAs for minimizing the CRB. Finally, we draw conclusions in Chapter 5.

For convenience, the notations used in the dissertation are defined below: $(\cdot)^H$ and $(\cdot)^*$ denote the Hermitian and complex conjugate operation of a matrix respectively, $\text{Im}(.)$ and $\text{Re}(.)$ the image-part-taking and the real-part-taking operation respectively, $E\{\cdot\}$ the expectation operation, $\delta(\cdot)$ a dirac delta function $(\cdot)_p$, the $p$th element of a vector, $(\cdot)_{p,q}$ the $(p,q)$th element of a matrix, $\mathbf{X}_{p,q}$ the $(p,q)$th submatrix of a large matrix $\mathbf{X}$, $\mathbf{0}_l$ an $l \times 1$ column zero vector, $*$ the

convolution operation, $\mathbf{I}$ an identity matrix, $\text{diag}(\mathbf{x})$ a diagonal matrix with diagonal entries of $\mathbf{x}$, $\det(\cdot)$ the determinant of a matrix, $\|\cdot\|$ the Frobenius norm of a matrix, and $\text{tr}(\cdot)$ the trace of a matrix.

# Chapter 2

# Maximum Likelihood Timing and Carrier Frequency Offset Estimation for OFDM Systems with Periodic Preambles

In this chapter, we consider the CFO problem for OFDM systems. As mentioned, we consider the system with a periodic preamble placed at the beginning of a data packet. CFO estimation has been extensively studied for the case of a two-period preamble [20]. In some applications, however, a preamble with more than two periods is available. A typical example is the IEEE802.11a/g wireless local area network (LAN) systems, which features a ten-period preamble. Recently, researchers have proposed an ML CFO estimation method for such systems [26]. This approach first estimates the received preamble using a least-squares method, and then maximizes the corresponding likelihood function. In addition to the standard calculations, this method requires an extra procedure to solve the roots of a polynomial function, which is disadvantageous for real-world implementations. In this chapter, we propose a new ML method to solve the likelihood function directly. Our method can obtain a closed-form ML solution, without a need for the root-finding step. We further extend the proposed method to address the joint CFO/STO estimation problem, and also derive a lower bound on its estimation perfor-

mance. Section 2.1 briefly reviews the CFO estimation method in [26]. Sections 2.2 and 2.3 detail the proposed CFO and STO estimation algorithm. Section 2.4 derives a lower bound on STO estimation performance. Section 2.5 reports simulation results and discussions.

## § 2.1  Existing Approach

In this section, we briefly review the algorithm proposed in [26]. Let the preamble in the OFDM system be periodic with period $N$ and length $QN$. Denote the preamble signal as $s(k)$, where $k = 0, 1, ..., QN - 1$. The preamble is placed at the beginning of a packet and is subsequently transmitted through a wireless channel. Denote the channel response as $h(k)$ and the output signal as $x(k)$. Then, we have $x(k) = s(k) * h(k)$ where $*$ denotes the convolution operation. Assume that the maximum channel delay is $N$. Then, we can discard the first received $N$ samples and retain the periodic property of the preamble, $x(k)$. Thus, the received preamble can be expressed as [10]

$$y(k) = e^{\frac{j2\pi\varepsilon k}{N}} x(k) + w(k), \tag{2.1}$$

where $k = N, N + 1, \ldots, QN - 1$, $\varepsilon$ is CFO and $w(k)$ represents additive white Gaussian noise (AWGN) with a variance of $\sigma_w^2$. We can perform an index transformation by letting $k = mN + n$, where $m = 1, \cdots, Q$ and $n = 0, \cdots, N - 1$ such that $x(k) = x(mN + n)$. For notational simplicity, we further let $x_m(n) = x(mN + n)$, denoting the $n$th sample of the $m$th period of $x(k)$. Due to periodicity, we have $x_p(n) = x_q(n)$ for $p, q \in \{1, \cdots, Q\}$. Similarly, we can define $y_m(n) = y(mN + n) = y(k)$, and $w_m(n) = w(mN + n) = w(k)$. Let $K = Q - 1$ and

$$\begin{aligned}
\mathbf{y}(n) &= [\ y_1(n)\ \ y_2(n)\ \ \cdots\ \ y_K(n)\ ]^T, \\
\mathbf{x}(n) &= [\ x_1(n)\ \ x_2(n)\ \ \cdots\ \ x_K(n)\ ]^T, \\
\mathbf{w}(n) &= [\ w_1(n)\ \ w_2(n)\ \ \cdots\ \ w_K(n)\ ]^T.
\end{aligned} \tag{2.2}$$

In addition, we define four matrices as follows:

$$
\begin{aligned}
\mathbf{Y} &= \begin{bmatrix} \mathbf{y}(0) & \mathbf{y}(1) & \cdots \mathbf{y}(N-1) \end{bmatrix}, \\
\mathbf{X} &= \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1)e^{\frac{j2\pi\varepsilon}{N}} & \cdots \mathbf{x}(N-1)e^{\frac{j2\pi\varepsilon(N-1)}{N}} \end{bmatrix}, \\
\mathbf{W} &= \begin{bmatrix} \mathbf{w}(0) & \mathbf{w}(1) & \cdots \mathbf{w}(N-1) \end{bmatrix}, \\
\mathbf{A} &= \begin{bmatrix} e^{j2\pi\varepsilon} & 0 & \cdots & 0 \\ 0 & e^{j2\pi\varepsilon\cdot2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{j2\pi\varepsilon\cdot K} \end{bmatrix}.
\end{aligned}
\tag{2.3}
$$

The received preamble in (2.1) can then be rewritten as

$$
\mathbf{Y} = \mathbf{AX} + \mathbf{W}.
\tag{2.4}
$$

The method in [26] uses a two-step approach; it first estimates $\mathbf{X}$ using a LS method, and then estimates CFO by maximizing the likelihood function. Since the noise is a Gaussian random variable, $\mathbf{y}(n)$ is a Gaussian random vector with a covariance matrix of $\sigma_w^2\mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix. For a given $\mathbf{A}$, the LS estimate of $\mathbf{X}$ can be expressed as $\mathbf{X}_{LS} = \frac{1}{K}\mathbf{A}^H\mathbf{Y} \equiv \mathbf{A}^+\mathbf{Y}$, where $(\cdot)^H$ denotes the Hermitian operation. Substituting $\mathbf{X}_{LS}$ back into (2.4), we can obtain the log-likelihood function as $\Lambda(\mathbf{A}) = \sum_{n=1}^{N} \|\mathbf{y}(n) - \mathbf{AA}^+\mathbf{y}(n)\|^2 = N \cdot \text{trace}((\mathbf{I} - \mathbf{AA}^+)\mathbf{R}_Y)$, where $\mathbf{R}_Y = E[\mathbf{YY}^H]$. The $(p, q)$th entry of $\mathbf{R}_Y$ is $\frac{1}{N} \sum_{n=0}^{N-1} y_p(n)y_q^*(n)$, $p, q \in [1, K]$ [44]. The desired CFO estimation can then be derived as

$$
\begin{aligned}
\hat{\varepsilon} &= \arg\{\min_{\varepsilon} \text{ trace}((\mathbf{I} - \mathbf{AA}^+)\mathbf{R}_Y)\} \\
&= \arg\{\max_{\varepsilon} \mathbf{a}^H\mathbf{R}_Y\mathbf{a}\},
\end{aligned}
\tag{2.5}
$$

where $\mathbf{a}$ is a vector consisting of the diagonal elements of $\mathbf{A}$. It was shown in [42] that

$$
\mathbf{a}^H\mathbf{R}_Y\mathbf{a} = \sum_{m=-(K-1)}^{K-1} b(m)e^{j2\pi m\varepsilon},
\tag{2.6}
$$

11

where $b(m) = \sum_{q-p=m} \frac{1}{N} \sum_{n=0}^{N-1} y_p(n) y_q^*(n)$. Taking the derivative of (2.6) with respect to $\varepsilon$ and letting the result be zero, we obtain

$$\sum_{m=1}^{K-1} mb(m)z^m = \sum_{m=1}^{K-1} mb(-m)z^{-m}, \tag{2.7}$$

where $z = e^{j2\pi\varepsilon}$. Equation (2.7) can be rewritten as

$$\text{Im}(\sum_{m=1}^{K-1} mb(m)z^m) = 0, \tag{2.8}$$

where $\text{Im}(.)$ is an operator that isolates the imaginary part of a scalar value. Denote the set containing the roots of (2.8) by $\Omega$. The CFO can then be estimated following [26] as

$$\hat{\varepsilon} = \frac{1}{j2\pi} ln(\hat{z}) \tag{2.9}$$

where $\hat{z} = arg\{\max_{z \in \Omega}(\Lambda(z))\}$ and $|\hat{z}| = 1$.

The procedure for CFO estimation in [26] can now be summarized as

1. Construct the correlation matrix $\mathbf{R}_Y$.

2. Calculate the coefficient of (2.8) using $\mathbf{R}_Y$.

3. Find the nonzero roots of (2.8).

4. Substitute the roots into (2.6), find the maximum root, and calculate $\hat{\varepsilon}$ using (2.9).

As we see, (2.8) requires a root-finding operation. Thus, a set of suboptimum algorithms to address this issue were proposed in [26]. Unfortunately, these suboptimum methods cannot effectively reduce the computational complexity while still maintaining good performance.

## § 2.2   Proposed ML CFO Estimation

In this section, we develop a new CFO estimation method that solves the likelihood function directly. The signal model we use is the same as that in (2.4). We assume that each data packet

is transmitted through a slow-fading channel with an impulse response of $h(k)$, $k = 0, \cdots, L-1$. Here, the $h(k)$'s have Rayleigh distributions, and they are statistically independent. Note that the time-domain preamble signal is obtained from the discrete Fourier transform of the frequency-domain preamble signal, and the frequency-domain preamble signal is generally a white sequence. From the central limit theorem, the time-domain preamble signal can then be approximated as a white Gaussian sequence. Thus, the channel output, $x(k)$, which equals $\sum_{l=0}^{L-1} h(l)s(k-l)$, and the received preamble $y(k)$ in (2.1) can be approximated as Gaussian sequences. Let the variance of the time-domain preamble signal, i.e., $s(k)$ be $\sigma_s^2$. Then, the variance of $x(n)$ equals $\sigma_x^2 = E\{\sum_{j=0}^{L-1} \sum_{l=0}^{L-1} h(j)s(k-j)h(l)^*s(k-l)^*\} = \sigma_s^2 \sum_{l=0}^{L-1} |h(l)|^2 = \sigma_s^2 \sigma_h^2$, and that of $y(k)$ equals $\sigma_x^2 + \sigma_w^2$. Note that $s(k)$ can be a psuedo noise (PN) sequence. In such a case, $\sigma_s^2$ indicates the averaged preamble power of $s(k)$.

Let $f(.)$ be a probability density function. Then, we explicitly write out the log-likelihood function of $\varepsilon$ following [10] as

$$
\begin{aligned}
\Lambda(\varepsilon) &= \ln\{\prod_{n\in\tilde{I}} f(\mathbf{y}(n))\} \\
&= \ln\{\frac{\prod_{n\in\tilde{I}} f(\mathbf{y}(n))}{\prod_{m\in[1,K]} \prod_{n\in\tilde{I}} f(y_m(n))} \cdot \prod_{m\in[1,K]} \prod_{n\in\tilde{I}} f(y_m(n))\} \\
&= \ln\{\prod_{n\in\tilde{I}} \frac{f(\mathbf{y}(n))}{f(y_1(n))\cdots f(y_K(n))} \cdot \prod_{m\in[1,K]} \prod_{n\in\tilde{I}} f(y_m(n))\}. \quad (2.10)
\end{aligned}
$$

It is clear that the last term in (2.10), $\prod_{m\in[1,K],\ n\in\tilde{I}} f(y_m(n))$, is independent of $\varepsilon$ [10]. As a result, this term can be dropped. Let

$$
\mathbf{u}(n) = e^{\frac{j2\pi\varepsilon n}{N}} \left[ \begin{array}{ccc} x_1(n)e^{j2\pi\varepsilon} & \cdots & x_K(n)e^{j2\pi\varepsilon\cdot K} \end{array} \right]^T. \quad (2.11)
$$

We then rewrite (2.4) as $\mathbf{Y} = \mathbf{U} + \mathbf{W}$ where $\mathbf{U} = \mathbf{AX} = [\mathbf{u}(0), \mathbf{u}(1), \cdots, \mathbf{u}(N-1)]$. Then,

13

$\mathbf{y}(n) = \mathbf{u}(n) + \mathbf{w}(n)$. Define $\mathbf{R}_u = E[\mathbf{u}(n)\mathbf{u}^H(n)]$ and $\mathbf{R}_y = E[\mathbf{y}(n)\mathbf{y}^H(n)]$. Then,

$$\mathbf{R}_u = \sigma_x^2 \begin{bmatrix} 1 & e^{-j2\pi\varepsilon} & \cdots & e^{-j2\pi(K-1)\varepsilon} \\ e^{j2\pi\varepsilon} & 1 & \cdots & e^{-j2\pi(K-2)\varepsilon} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j2\pi(K-1)\varepsilon} & e^{j2\pi(K-2)\varepsilon} & \cdots & 1 \end{bmatrix}, \tag{2.12}$$

and $\mathbf{R}_y = \mathbf{R}_u + \sigma_w^2 \mathbf{I}$ where $\mathbf{I}$ is an identical matrix. Thus, we can express $f(\mathbf{y}(n))$ as [39] [40]

$$f(\mathbf{y}(n)) = (\pi^K \det(\mathbf{R}_y))^{-1} \exp\left[-\mathbf{y}(n)^H \mathbf{R}_y^{-1} \mathbf{y}(n)\right]. \tag{2.13}$$

According to the matrix inversion lemma [17], we derive the inverse of $\mathbf{R}_y$ as

$$\mathbf{R}_y^{-1} = \sigma_w^{-2}\mathbf{I} - \frac{\sigma_w^{-4}\mathbf{R}_u}{1+\sigma_w^{-2}E\{\mathbf{u}^H\mathbf{u}\}}. \tag{2.14}$$

Note that for $n \in \tilde{I}$, we have

$$E\{y_p(n)y_q^*(n)\} = \begin{cases} \sigma_x^2 + \sigma_w^2 & \text{if } q - p = 0 \\ \sigma_x^2 e^{-j2\pi\varepsilon(q-p)} & \text{if } q - p \neq 0 \end{cases} \tag{2.15}$$

where $p, q \in [1, K]$. As a result, $\mathbf{R}_y^{-1} = \sigma_w^{-2}\mathbf{I} - \frac{\mathbf{R}_u}{\sigma_w^4 + K\sigma_w^2\sigma_x^2}$ and

$$f(y_p(n)) = \frac{\exp(-\frac{y_p(n)y_p^*(n)}{\sigma_x^2+\sigma_w^2})}{\pi(\sigma_x^2+\sigma_w^2)}, \tag{2.16}$$

where $p \in [1, K]$. Thus, the exponential term in (2.13) becomes

$$\begin{aligned} \mathbf{y}(n)^H \mathbf{R}_y^{-1} \mathbf{y}(n) &= \sigma_w^{-2} \sum_{p=1}^{K} y_p(n)y_p^*(n) \\ &\quad - C_0 \sum_{p=1}^{K} \sum_{q=1}^{K} y_p(n)y_q^*(n)e^{j2\pi(q-p)\varepsilon} \\ &= (\sigma_w^{-2} - C_0) \sum_{p=1}^{K} y_p(n)y_p^*(n) \\ &\quad - 2C_0 \text{Re}\{\sum_{p=1}^{K-1}\sum_{q>p}^{K} y_p(n)y_q^*(n)e^{j2\pi(q-p)\varepsilon}\} \end{aligned} \tag{2.17}$$

14

where $C_0 = \sigma_x^2/(\sigma_w^4 + K\sigma_w^2\sigma_x^2)$, and Re$\{\cdot\}$ denotes the operation that isolates the real part of the indicated complex variable. Dropping the superfluous terms and substituting (2.13)-(2.17) into (2.10), we finally express the log-likelihood function as

$$\Lambda(\varepsilon) = \sum_{n=0}^{N-1} \ln \left\{ \frac{(\sigma_x^2+\sigma_w^2)^K \exp\left[-\mathbf{y}(n)^H \mathbf{R}_y^{-1} \mathbf{y}(n)\right]}{\det(\mathbf{R}_y) \exp\left[-\frac{\sum_{p=1}^{K} y_p(n) y_p^*(n)}{\sigma_x^2+\sigma_w^2}\right]} \right\} \tag{2.18}$$

$$= C_1 + C_2 \phi + C_3 \sum_{p=1}^{K-1} \sum_{q>p}^{K} |\gamma_{pq}| \cos(\psi_{pq}), \tag{2.19}$$

where

$$\gamma_{pq} = \sum_{n=0}^{N-1} y_p(n) y_q^*(n) \ \ (q \geq p \text{ and } p \geq 1), \tag{2.20}$$

$$\psi_{pq} = 2\pi\varepsilon(q-p) + \angle\gamma_{pq}, $$

$$\phi = \sum_{p=1}^{K} \gamma_{pp}, \tag{2.21}$$

$$C_1 = N \cdot \ln\left(\frac{(\sigma_x^2+\sigma_w^2)^K}{\det(\mathbf{R}_y)}\right), \tag{2.22}$$

$$C_2 = (1-K)\frac{\rho^2}{\sigma_w^2(1+(K-1)\rho)}, \tag{2.23}$$

$$C_3 = \frac{2C_2}{(1-K)\rho}, \tag{2.24}$$

$$\rho = \frac{\sigma_x^2}{\sigma_x^2+\sigma_w^2}. \tag{2.25}$$

Note that $\phi$ is the received signal energy and $\det(\mathbf{R}_y)$ is a constant, independent of $\varepsilon$. The detailed derivation of (2.19) is provided in Appendix A.1. Ignoring unrelated terms, we obtain the log-likelihood as

$$\Lambda(\varepsilon) \propto \sum_{p=1}^{K-1} \sum_{q>p}^{K} |\gamma_{pq}| \cos(\psi_{pq}). \tag{2.26}$$

To maximize the function, we first take a derivative of $\Lambda(\varepsilon)$ with respect to $\varepsilon$ and obtain

$$\frac{\partial}{\partial\varepsilon}\Lambda(\varepsilon) = -\sum_{p=1}^{K-1} \sum_{q>p}^{K} 2\pi(q-p)|\gamma_{pq}| \sin(\psi_{pq}). \tag{2.27}$$

15

Thus, we have an alternative expression to that in (2.8). Now, the problem is how to solve (2.27). Since (2.27) involves a nonlinear sine function, a closed-form solution will be difficult to calculate. Here, we use a simple approximation method to overcome the problem. Using (2.20) and (2.1), we obtain

$$
\begin{aligned}
\gamma_{pq} &= e^{j2\pi\varepsilon(p-q)} \sum_{n=0}^{N-1} |x_1(n)|^2 + \sum_{n=0}^{N-1} w_p(n)w_q^*(n) \\
&+ e^{j2\pi\varepsilon(pN+p)} \sum_{n=0}^{N-1} x_1(n)w_p^*(n) \\
&+ e^{j2\pi\varepsilon(pN-q)} \sum_{n=0}^{N-1} x_1^*(n)w_q(n).
\end{aligned}
\tag{2.28}
$$

In (2.28), we have used the periodic property that $x_1(n) = x_p(n) = x_q(n)$. Now, if the noise level is low, the noise related terms in (2.28) can be ignored. We then have

$$
\angle \gamma_{pq} \approx 2\pi\varepsilon(p-q).
\tag{2.29}
$$

From (2.29), we write

$$
\psi_{pq} \approx 2\pi\varepsilon(q-p) + 2\pi\varepsilon(p-q) = 0.
\tag{2.30}
$$

From (2.30), we can then assume that $\sin(\psi_{pq}) \approx \psi_{pq}$, and approximate the expression in (2.27) by

$$
\frac{\partial}{\partial \varepsilon} \Lambda(\varepsilon) \simeq - \sum_{p=1}^{K-1} \sum_{q>p}^{K} 2\pi(q-p)|\gamma_{pq}|(\psi_{pq}).
\tag{2.31}
$$

Setting the result in (2.31) to zero, we can estimate CFO as

$$
\hat{\varepsilon} = - \frac{\displaystyle\sum_{p=1}^{K-1} \sum_{q>p}^{K} |\gamma_{pq}|(q-p)\angle\gamma_{pq}}{2\pi \displaystyle\sum_{p=1}^{K-1} \sum_{q>p}^{K} |q-p|^2 |\gamma_{pq}|}.
\tag{2.32}
$$

Note that the approximation in (2.30) will become exact if noise is not present and if $\varepsilon$ is the true CFO. In other words, (2.27) and (2.31) will have a same zero-crossing point although the two

16

functions are different, indicating that (2.31) and (2.27) will yield the same optimum solution. If noise is present, however, (2.31) and (2.27) will not have the same optimum solution. The accuracy of the solution in (2.31) depends on the signal-to-noise ratio in (2.28). We define the signal-to-noise ratio in (2.28) as $\mathrm{SNR}_\gamma$, and that in (2.1) as SNR. Then, $\mathrm{SNR} = \sigma_x^2/\sigma_w^2$, as typically defined. From (2.28), it is simple to see that

$$\mathrm{SNR}_\gamma = \frac{N^2\sigma_x^4}{N\sigma_w^4 + 2N\sigma_x^2\sigma_w^2} = \frac{N \cdot \mathrm{SNR}^2}{1 + 2\mathrm{SNR}} \tag{2.33}$$

From (2.33), we can see that $\mathrm{SNR}_\gamma$ can be much larger than SNR as long as $N$ is reasonably large and SNR is not very low. Subsequently, the approximation in (2.31) will introduce only a small error for a wide SNR range. As a simple example, let $N = 16$ and $\mathrm{SNR} = 0$ dB. From (2.33), we obtain $\mathrm{SNR}_\gamma = 7.27$ dB, which is much higher than SNR.

Note that the proposed estimate requires that we extract the phase from $\gamma_{pq}$. It is simple to see that the result is only unambiguous when $|\angle\gamma_{pq}| < \pi$. For a particular combination of $p$ and $q$, the estimation range for CFO is $|\varepsilon| \leq 1/[2(q-p)]$. Since the maximum value for $q-p$ is $K-1$, the estimation range for CFO is $|\varepsilon| \leq 1/[2(K-1)]$. When $K$ is large, the range becomes small. In the following, we propose a method to remedy this problem. The basic idea is to apply the phase unwrapping procedure. We first calculate the phase angle for each $\gamma_{pq}$. Then, for each $p$, we calculate the phase difference of $\angle\gamma_{pq}$, $q = p+1, p+2, \ldots, K$. Let $d_{r,s}$ denote the phase difference, i.e., $d_{r,s} = \angle\gamma_{rs} - \angle\gamma_{r(s-1)}$, $r = 1, 2, \ldots, K-2$ and $s = r+2, r+3, \ldots, K$. Since the maximum value of $|d_{r,s}|$ is $\pi$, whenever $|d_{r,s}| > \pi$, the phase need to be unwrapped. This can be performed with the following operation:

$$d_{r,s} = \begin{cases} d_{r,s} - 2\pi & \text{if } d_{r,s} > \pi \\ d_{r,s} + 2\pi & \text{if } d_{r,s} < -\pi \end{cases}. \tag{2.34}$$

For a value of $r$, the $d_{r,s}$ values should have the same signs. We can use this property to further correct occasional errors. Let $g$ be the sum of all $d_{r,s}$ values, i.e., $g = \sum_{r=1}^{K-2} \sum_{s=r+2}^{K} d_{r,s}$. Then, we use the sign of $g$ to determine the sign of $d_{r,s}$ and to evaluate $\angle\gamma_{pk}$, $k = p+1 \cdots K$. Finally,

17

the unwrapped $\angle\gamma_{pq}$ can be written (with $q \geq p+2$) as:

$$\angle\gamma_{pq} = \angle\gamma_{p(p+1)} + \sum_{s=p+2}^{q} d_{p,s}. \tag{2.35}$$

Substituting (2.35) into (2.32), we can estimate CFO. Using our proposed procedure, the CFO estimation range can be greatly extended up to $|\varepsilon| < 1/2$.

Now, the procedure for our proposed ML CFO estimation can be summarized as follows:

1. Construct all $\gamma_{pq}$'s where $p \in [1, K-1]$, $q \in [p+1, K]$, and calculate their amplitude.

2. Use the phase unwrapping scheme to estimate the phase of $\gamma_{pq}$.

3. Substitute the results into (2.32) and calculate the ML estimate.

Clearly, the proposed estimate does not require the root-finding procedure, and this, in turn, effectively reduces the computational complexity. Step 1) above is similar to the calculation of **R** in Section 2.1. However, our method is easier since we only have to compute $\gamma_{pq}$ for $q > p$.

In this paragraph, we compare the computational complexity of the proposed ML estimate with that of the algorithm in [26]. Three algorithms are proposed in [26], referred to as Algorithm A, A', and B. While Algorithm A is optimal, Algorithms A' and B are suboptimal. Table 2.1 summarizes this result. In Table 2.1, MUL, ADD, LN, ABS, PH, and DIV denote the operations of multiplication, addition, natural logarithm, absolute value, phase derivation, and division, respectively. In addition, the algorithm proposed in this section is referred to as proposed Algorithm I, and the one in the next section is termed proposed Algorithm II. For the proposed algorithms, we consider the worst case in which all the phase differences, $d_{r,s}$, need to be unwrapped. Figure 2.1 shows several examples of how $Q$ and $N$ affect the complexity. Note that the computational complexity for the root-finding procedure in [26] is not included here. For convenience, we treat all operations other than addition as multiplications. As we can see, the computational complexity for the proposed algorithm is slightly lower than that for algorithms A and B in [26], and Algorithm A' in [26] is the lowest. However, algorithm

A' truncates the polynomials with order higher than two in (2.6), i.e., $\Lambda(z) = \sum\limits_{m=-2}^{2} b(m)z^m$. This impacts the estimation accuracy. Note that we can always truncate the summation terms in (2.32), and thereby reduce the computational complexity of proposed Algorithm I. Since suboptimum approaches are not our focus, we will not consider the details here. We will now discuss the computational complexity of the root-finding procedure. As shown in [43], [45], the root-finding procedure requires $O(K^3)$ multiplications. Table 2.1 shows that the computational complexity of Algorithm A is $O(NK^2)$. Thus, the computational complexity of the root-finding procedure will be high when $K$ is large. Also, its implementation cost will also be higher, since we may need dedicated electronic circuitry to implement this function.

It is well known that the performance of an unbiased estimator is bounded by the CRB [41]. If the variance of an unbiased estimator reaches the CRB, we consider the estimator efficient. Following the procedure to derive performance bounds in [41], we can calculate the CRB for our CFO estimation procedure. Let $\hat{\varepsilon}$ be an estimate of $\varepsilon$. The CRB for our CFO estimation is then

$$
\begin{aligned}
CRB(\hat{\varepsilon}) &= -\frac{1}{E[\frac{\partial^2}{\partial \varepsilon^2}\Lambda(\varepsilon)]} \\
&= \frac{(8\pi^2\rho)^{-1}\sigma_w^2(1+(K-1)\rho)}{E[\sum\limits_{p=1}^{K-1}\sum\limits_{q>p}^{K}(q-p)^2\mathrm{Re}\{\gamma_{pq}e^{j2\pi\varepsilon(q-p)}\}]} \\
&= \frac{\sigma_w^2(1+(K-1)\rho)}{8\pi^2\rho N\sigma_x^2\sum\limits_{p=1}^{K-1}\sum\limits_{q>p}^{K}(q-p)^2} \\
&= \frac{1+K\cdot\mathrm{SNR}}{8\pi^2 N\cdot\mathrm{SNR}^2}\frac{1}{\sum\limits_{p=1}^{K-1}\sum\limits_{q>p}^{K}(q-p)^2}
\end{aligned}
\tag{2.36}
$$

where $E[\cdot]$ denotes the expectation and SNR indicates the signal-to-noise ratio. For a special case, $K = 2$, the CRB is the same as that in [46].

19

# § 2.3   Proposed Joint ML STO and CFO Estimation

In this section, we extend the method developed in the previous section to solve the STO estima-
tion problem. The core idea is to apply a sliding data window for the received $(Q+1)N$ samples;
each window covers the preamble in the context of a particular timing offset. We perform the
ML CFO estimation for data in each window, and store the estimated CFO and the correspond-
ing maximum log-likelihood. Thereafter, the estimated CFO with the largest log-likelihood is
selected as the ML CFO estimate. The corresponding window position is taken as the ML STO
estimate. Let the window size be $QN$, and define the set $V_i = \{y(i), y(i+1), \ldots, y(i+QN-1)\}$
to be the received data in window $i$. Since the maximum delay is shorter than $N$, it is clear that
$0 \le i \le N-1$. If we let the STO be $\theta$, $V_\theta$ will cover the complete preamble. In Appendix A.2,
we show that the log-likelihood function for $V_i$ can be expressed by

$$\Lambda^i(\varepsilon) = C_1^i + C_2^i \phi^i + C_3^i \sum_{p=0}^{Q-2} \sum_{q>p}^{Q-1} |\gamma_{pq}^i| \cos(\psi_{pq}^i) \tag{2.37}$$

where the superscript $i$ indicates that all the variables are calculated within $V_i$, and $C_1^i$, $C_2^i$ and
$C_3^i$ can be treated as window-independent. Thus, we can simplify the above log-likelihood
function using

$$\Lambda^i(\varepsilon) \approx C_2 \phi^i + C_3 \sum_{p=0}^{Q-2} \sum_{q>p}^{Q-1} |\gamma_{pq}^i| \cos(\psi_{pq}^i), \tag{2.38}$$

where $C_2$ and $C_3$ are the same as those in (2.23) and (2.24). Since the received signal power,
$\phi^i$, is independent of CFO, we can estimate CFO using (2.32) as

$$\hat{\varepsilon}^i = -\frac{\displaystyle\sum_{p=0}^{Q-2} \sum_{q>p}^{Q-1} |\gamma_{pq}^i|(q-p)\angle\gamma_{pq}^i}{2\pi \displaystyle\sum_{p=0}^{Q-2} \sum_{q>p}^{Q-1} |q-p|^2|\gamma_{pq}^i|}. \tag{2.39}$$

Note that the upper bound in the summation terms of (2.39) is $Q$ instead of $K$. The estimated
STO is then

$$\hat{\theta} = arg\{\max_i(\Lambda^i(\hat{\varepsilon}^i))\} = i_{opt}. \tag{2.40}$$

Now, the procedure for the proposed joint ML STO and CFO estimation can be summarized as follows:

1. Calculate $\gamma_{pq}^i$ and its amplitude, where $i \in [1, N]$, $p \in [0, Q-2]$, and $q \in [p+1, Q-1]$.

2. Use the phase unwrapping procedure outlined above to calculate $\angle \gamma_{pq}^i$.

3. Substitute the results into (2.38) and (2.39), and calculate $\Lambda^i(\hat{\varepsilon}^i)$ and $\hat{\varepsilon}^i$.

4. Find $i_{\text{opt}}$ such that $\Lambda^{i_{\text{opt}}}(\hat{\varepsilon}^{i_{\text{opt}}}) > \Lambda^i(\hat{\varepsilon}^i)$, $i \neq i_{\text{opt}}$.

5. The ML STO estimate is $i_{\text{opt}}$ and the ML CFO estimate is then $\hat{\varepsilon}^{i_{\text{opt}}}$.

As we can see from the above procedure, the computational complexity of the algorithm will be $N$ times higher than that in Section 2.2. Note also that the upper limit of $p$ is $Q-2$ instead of $K-2$. In other words, we have an extra period for CFO estimation. By leveraging the sliding window structure, we can effectively reduce the computational complexity in calculating $\gamma_{pq}^i$. Similar to the definition of $\gamma_{pq}$, we obtain $\gamma_{pq}^i = \sum\limits_{n=i}^{i+N-1} y_p(n)[y_q(n)]^*$. Then, it is simple to show that

$$
\begin{aligned}
\gamma_{pq}^i &= \gamma_{pq}^{i-1} + y_p(i+N-1)[y_q(i+N-1)]^* \\
&\quad - y_p(i-1)[y_q(i-1)]^*.
\end{aligned}
\tag{2.41}
$$

From (2.41), we can see that except for $i = 0$, the calculation of $\gamma_{pq}^i$ requires only two complex multiplications and two complex additions. This will greatly reduce the required computational complexity in the scenario of joint STO and CFO estimation. The required computational complexity has been summarized in Table 2.1.

We can also obtain the CRB for the CFO estimate. All we have to do is to replace $K$ with $Q$ in (2.36). Since $Q = K + 1$, the CRB is lower than that in (2.36). Note that the STO is a discrete value. No performance lower bounds have been reported to date in the literature. In the next section, we will derive a lower bound to address this omission.

21

## § 2.4  Performance Analysis of STO Estimation

In this section, we analyze the performance of the proposed STO estimation method. We first redefine (2.38) as $\Lambda^i(\varepsilon) = C_2\phi^i + C_3\xi^i$ where

$$
\begin{aligned}
\phi^i &= \sum_{p=0}^{K}\sum_{n=i}^{i+N-1} y_p(n)y_p^*(n) \\
&= \sum_{p=0}^{K}\sum_{n=0}^{N-1} x_p(n)x_p^*(n) + w_p(n)w_p^*(n) \\
&\quad + 2\mathrm{Re}\{x_p(n)w_p^*(n)\exp(j2\pi\varepsilon\frac{pN+n}{N})\},
\end{aligned}
\tag{2.42}
$$

and

$$
\begin{aligned}
\xi^i &= \sum_{p=0}^{K-1}\sum_{q>p}^{K} |\gamma_{pq}^i|\cos(\psi_{pq}^i) \\
&= \sum_{p=0}^{K-1}\sum_{q>p}^{K}\sum_{n=0}^{N-1} x_p(n)w_q^*(n)\exp(j2\pi\varepsilon\frac{qN+n}{N}) \\
&\quad + w_p(n)x_q^*(n)\exp(-j2\pi\varepsilon\frac{pN+n}{N}) \\
&\quad + w_p(n)w_q^*(n)\exp(j2\pi\varepsilon(q-p)) \\
&\quad + x_p(n)x_q^*(n).
\end{aligned}
\tag{2.43}
$$

Note here that $\phi^i$ and $\xi^i$ are random variables. The mean value of $\Lambda^i(\varepsilon)$, denoted by $\mu_\Lambda^i$, is equal to $C_2\mu_\phi^i + C_3\mu_\xi^i$, where $\mu_\phi^i$ and $\mu_\xi^i$ are the mean of $\phi^i$ and $\xi^i$, respectively. The variance of $\Lambda^i$ can be expressed by $\nu_\Lambda^i = C_2^2\nu_\phi^i + C_3^2\nu_\xi^i + 2C_2C_3\kappa_{\phi\xi}^i$, where $\nu_\phi^i$ and $\nu_\xi^i$ denote the variance of $\phi^i$ and $\xi^i$, respectively, and $\kappa_{\phi\xi}^i$ the covariance between $\phi^i$ and $\xi^i$. The whole set of $V_i$, $0 \le i \le N-1$, has $(Q+1)N$ samples and it may cover three regions. The first region consists of the noise samples, the second region the periodic preamble samples, and the third region the data samples. We denote these regions by $I_N$, $I_P$, and $I_D$. Thus, the signal variance in $I_N$ is $\sigma_w^2$, that in $I_P$ is $\sigma_x^2 + \sigma_w^2$ and that in $I_D$ is $\sigma_d^2 + \sigma_w^2$, where $\sigma_d^2$ represents the variance of data samples. Recall that $\theta$ is the actual STO in the system. Using $\theta$ as a reference, we can have three cases

22

for the value of $i$: $i = \theta$, $i < \theta$, and $i > \theta$ ($0 \leq i \leq N - 1$). The statistics of $\phi^i$ and $\xi^i$ are different across these three cases. In Appendix A.3, we provide a detailed derivation of $\mu_\phi^i$, $\mu_\xi^i$, $\nu_\phi^i$, $\nu_\xi^i$, and $\kappa_{\phi\xi}^i$.

For the proposed STO estimation algorithm, an error occurs when $i_{opt} \neq \theta$. Thus, we can define the error probability of STO estimation as $P(\cup_{i,i\neq\theta}\{\Lambda^\theta < \Lambda^i\})$, where $P(.)$ denotes the probability of a certain event. Note that the evaluation of $P(\Lambda^\theta < \Lambda^i)$ only requires one-dimensional integration. If the log-likelihood functions for all $i$'s are independent and identically distributed (i.i.d.), we have $P(\cup_{i,i\neq\theta}\{\Lambda^\theta < \Lambda^i\}) = \sum_{i,i\neq\theta} P(\Lambda^\theta < \Lambda^i)$. Unfortunately, the log-likelihood functions are not independent. As a result, we have to conduct multi-dimensional integration, which is both complex and difficult. Therefore, we propose a simple alternative to overcome the problem. Instead of the exact error probability, we attempt to derive a lower bound.

As shown in [22], the likelihood function is approximately Gaussian. We denote the distribution of $\Lambda^i$ using $G(\mu_\Lambda^i, \nu_\Lambda^i)$, where $G(\cdot)$ denotes the Gaussian distribution. Consider the joint density function of $\Lambda^i$ and $\Lambda^j$. Using the Gaussian assumption, we write the bivariate Gaussian distribution as

$$
\begin{aligned}
P(\Lambda^i, \Lambda^j) &= \frac{1}{2\pi \cdot \nu_\Lambda^i \cdot \nu_\Lambda^j \cdot \sqrt{1 - C_c(i,j)}} \\
&\quad \cdot \exp\left(-\frac{z_{ij}}{2(1 - C_c(i,j))}\right)
\end{aligned} \tag{2.44}
$$

where $1 \leq i, j \leq N$,

$$
\begin{aligned}
z_{ij} &= \frac{(\Lambda^i - \mu_\Lambda^i)^2}{\nu_\Lambda^i} + \frac{(\Lambda^j - \mu_\Lambda^j)^2}{\nu_\Lambda^j} \\
&\quad - \frac{2C_c(i,j)(\Lambda^i - \mu_\Lambda^i)(\Lambda^j - \mu_\Lambda^j)}{\sqrt{\nu_\Lambda^i \cdot \nu_\Lambda^j}},
\end{aligned} \tag{2.45}
$$

and

$$
C_c(i,j) = \frac{E\{\Lambda^i(\Lambda^j)^*\} - \mu_\Lambda^i \mu_\Lambda^{j^*}}{\sqrt{\nu_\Lambda^i \cdot \nu_\Lambda^j}}. \tag{2.46}
$$

Note that $C_c(i, j)$ is the corresponding correlation coefficient. The numerator of $C_c(i, j)$ is expressed as

$$
\begin{aligned}
E\{\Lambda^i(\Lambda^j)^*\} &= \mu_\Lambda^i \mu_\Lambda^{j*} + C_2^2 \kappa_{\phi\phi}^{ij} + C_3^2 \kappa_{\xi\xi}^{ij} \\
&\quad + C_2 C_3 \kappa_{\phi\xi}^{ij} + C_2 C_3 \kappa_{\xi\phi}^{ij},
\end{aligned}
\tag{2.47}
$$

where $\kappa_{ab}^{ij}$ denotes the covariance of $a^i$ and $b^{j*}$ ($a^i, b^j \in \{\phi^i, \phi^j, \xi^i, \xi^j\}$). The main idea here is only to calculate $P(\Lambda^\theta > \Lambda^i)$ for all $i$'s (except for $i = \theta$), and then use the result to derive a lower bound. Thus, we only have to consider $C_c(i, \theta)$ as

$$
\kappa_{\phi\phi}^{i\theta} = 2\sigma_x^2 \sigma_w^2 (QN - |i - \theta|),
\tag{2.48}
$$

$$
\kappa_{\xi\xi}^{i\theta} = Q(Q-1)\sigma_x^2 \sigma_w^2 \left[\frac{N}{3}(2Q-1) - \frac{1}{2}|i - \theta|\right]
$$

$$
+ \frac{1}{2}QN(Q-1)\sigma_w^4,
\tag{2.49}
$$

and

$$
\begin{aligned}
\kappa_{\phi\xi}^{i\theta} &= \kappa_{\xi\phi}^{i\theta} = (Q-1)^2 N \sigma_x^2 \sigma_w^2 \\
&\quad + (Q-1)(N - |i - \theta|)\sigma_x^2 \sigma_w^2.
\end{aligned}
\tag{2.50}
$$

Substituting (2.45)-(2.50) into (2.44), we can then evaluate $P(\Lambda^\theta > \Lambda^i)$. Given this definition, we have $P(\Lambda^\theta > \Lambda^i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\Lambda^\theta} P(\Lambda^i, \Lambda^\theta) d\Lambda^i d\Lambda^\theta$. Simulations have been conducted to evaluate the validity of our theoretical results. Using the scenario depicted in Section 2.5, we compare the theoretical and simulated $P(\Lambda^\theta > \Lambda^i)$ in Figure 2.2. From the figure, we see that the theoretical $P(\Lambda^\theta > \Lambda^i)$ is close to the simulated result. If we let $P_{min} = \min_{i \neq \theta} P(\Lambda^\theta > \Lambda^i)$, we can then treat $P_{min}$ as an upper bound for the correct probability of STO estimation (i.e., $i_{opt} = \theta$). Thus, we can then have a lower bound for the error probability of STO estimation (LBSTO) as $1 - P_{min}$.

# § 2.5  Simulations and Discussions

In this section, we report our simulation results, where these evaluate the performance of the proposed algorithms. We adopt a Rayleigh multipath channel with an exponential power decay and five channel taps. The preamble, generated from a frequency-domain BPSK modulated signal, has 10 periods and each period has 16 samples. The data following the preamble are transmitted using a 16-QAM scheme. The mean square error (MSE) of the estimated CFO is used as a performance measure. We first consider the CFO-only estimation problem. In this case, the first received $N$ samples are discarded. As previously mentioned, we term the proposed approach for this scenario as Algorithm I (as described in Section 2.2). We compare the proposed ML estimator with that in [26]. One optimum algorithm (Algorithm A) and two suboptimum algorithms (Algorithm A' and B) in [26] are simulated. Figure 2.3 shows the simulation result for SNR at 10dB. From the figure, we can see that the performance of Algorithms A' and B are poorer. Algorithm A and the proposed algorithm offer a similar level of performance that is very close to the CRB. To evaluate the impact of CFO on system performance, we conduct simulations for systems with and without CFO. For the system with CFO, we first use the proposed method to estimate CFO, and then conduct CFO compensation. Figure 2.4 shows the BER comparison for $\varepsilon = 0.2$. As we can see from the figure, the BER performance degrades slightly when CFO is present.

We then consider the case of the joint STO and CFO estimation process. In this case, discarding the first received $N$ samples is not necessary. As a result, one additional preamble is available. This means that the proposed method may offer better performance compared to the previous scenario. However, the price we pay for the additional STO estimation is the increase in computational complexity. As mentioned, we name this approach proposed Algorithm II (as explained in Section 2.3). Using a similar approach, the method in [26] can also be used to estimate STO. However, its computational complexity increases much more than our method. Figure 2.5 shows the simulation result for the CFO estimate. The proposed method offers good

performance. Only when CFO is very close to $\pm 0.5$, does the performance of the proposed algorithms degrade. Figure 2.6 shows the CFO estimation result for various SNRs. From the figure, we see that the proposed method still works well for SNRs as low as -5 dB. The algorithms in [26] perform well until SNR reaches -7 dB, somewhat better than the proposed algorithms. However, when SNR falls below -8 dB, the proposed algorithms again outperform those in [26]. This may be because the correlation matrix in (2.6) is very noisy, and the roots therefore cannot be solved reliably. Figure 2.7 shows the error probability for the STO estimation. We observe that the derived lower bound for the STO estimation is tight when the SNR is high. Note that the error probability we defined is only relevant to performance evaluation. If the channel response is shorter than the CP (which is the typical case), we can always has some tolerance for the STO estimation. Thus, there is no need to calculate the exact channel delay. In real-world applications, it is a common practice to reduce the estimated STO by a couple of samples when conducting STO compensation. Another property is that STO estimation performance is not particularly impacted when CFO is closed to 0.5. In the literature, there exist a number of STO estimation methods. We select the two algorithms proposed in [22] and [27], for comparison. Figure 2.8 shows the MSE curves for these approaches and for the proposed algorithms ($\theta = 8$). The figure confirms that the proposed method offers the best performance.
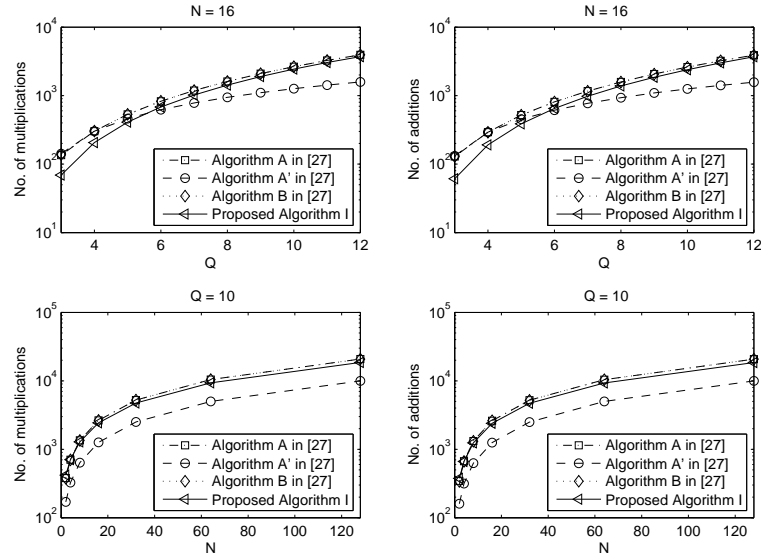
Figure 2.1: Computational complexity comparison for the algorithm in [26] and proposed Algorithm I. Note that the complexity of the root-finding procedure is not considered in [26].
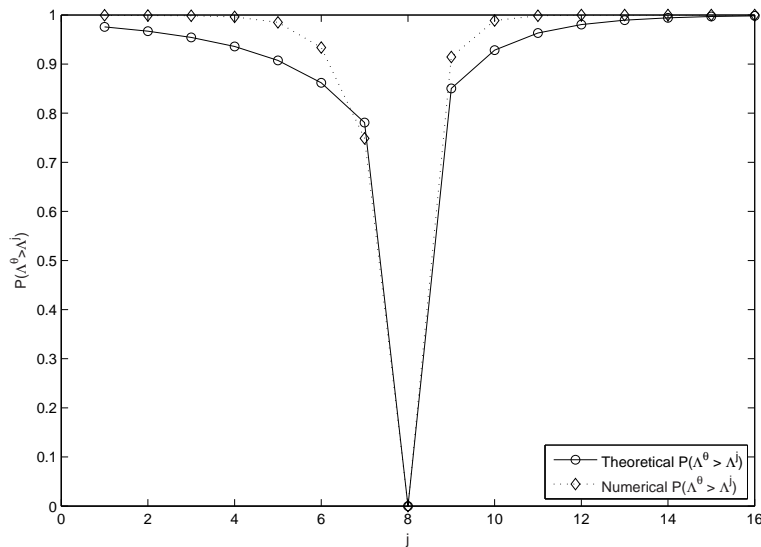


Figure 2.2: Comparison of simulated and theoretical $P(\Lambda_\theta > \Lambda_j)$.
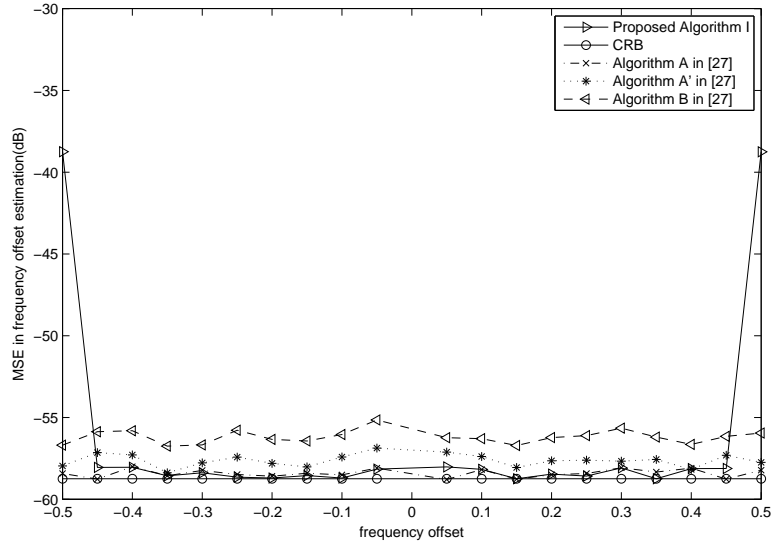
Figure 2.3: Performance comparison of CFO estimation, the algorithm in [26] and proposed algorithm I; SNR = 10dB.
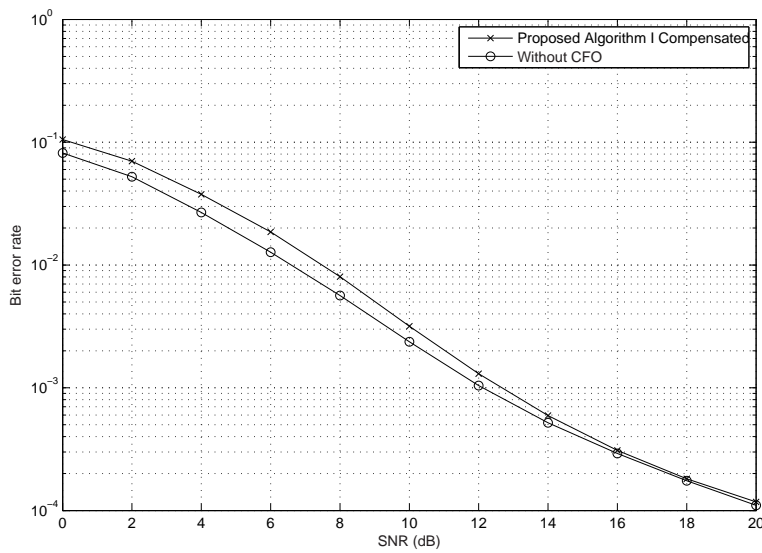


Figure 2.4: BER comparison for systems with and without CFO.

28
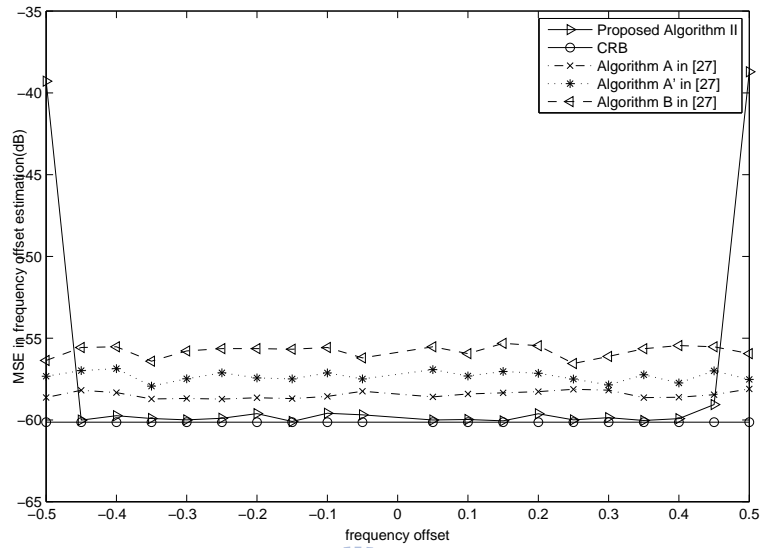
Figure 2.5: Performance comparison for CFO estimation; SNR = 10dB.
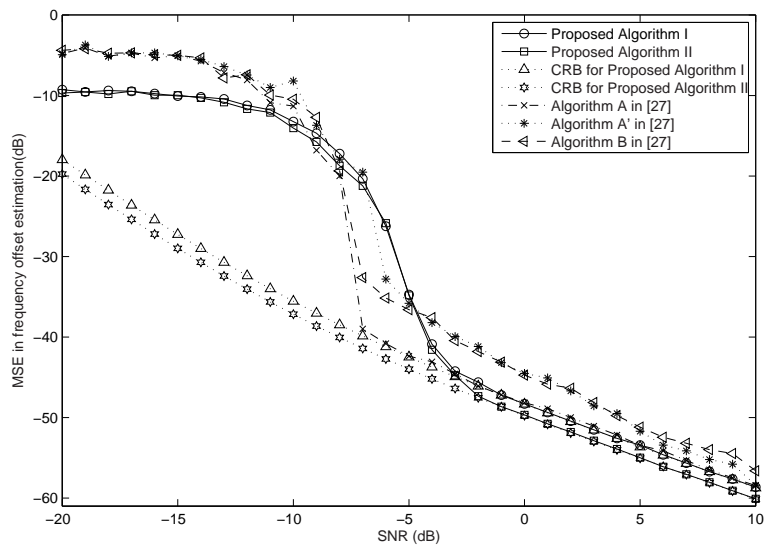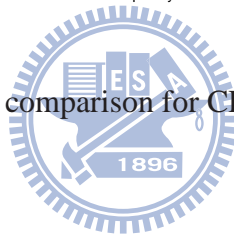


Figure 2.6: Performance comparison for CFO estimation; N = 16, Q = 10.
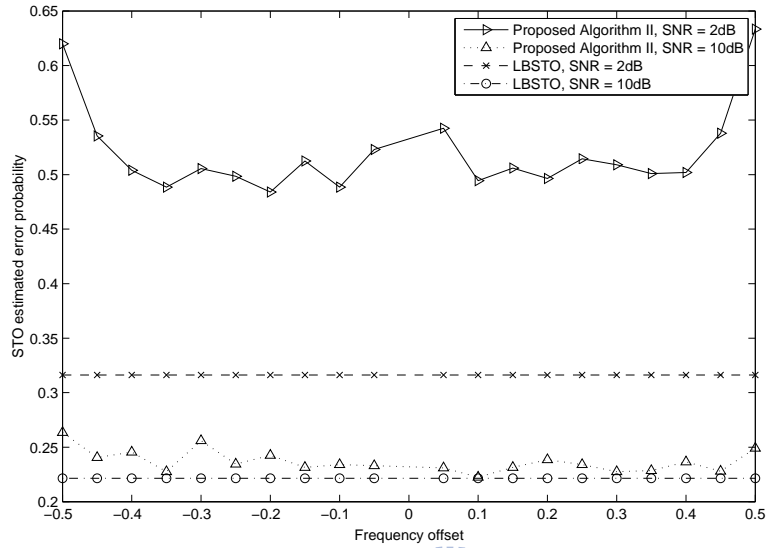
Figure 2.7: Error probability of STO estimation (proposed Algorithm II).
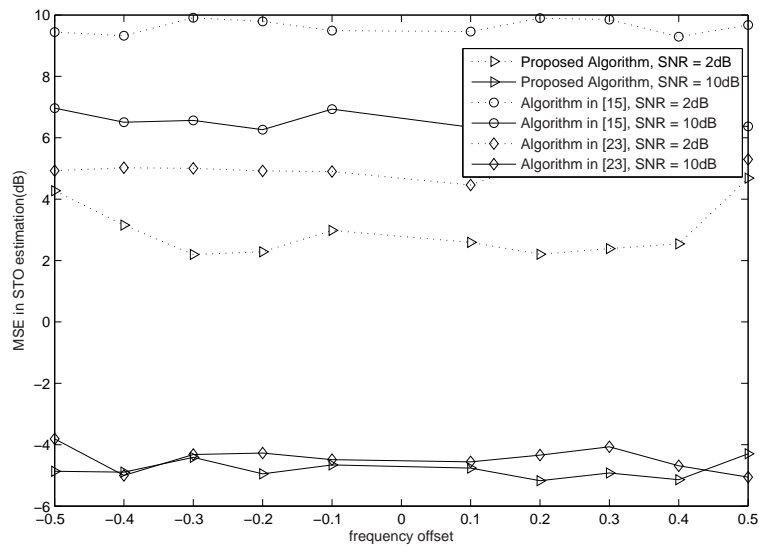


Figure 2.8: Performance comparison for STO estimation (SNR = 2dB and 10dB).

Table 2.1: Computational complexity comparison for the algorithm in [26] and for the proposed algorithms.

| | Algorithm A in [26] | Algorithm A' in [26] | Algorithm B in [26] | Proposed algorithm I | Proposed algorithm II |
|---|---|---|---|---|---|
| No. of MUL | $2NK^2 + 7K - 10$ | $2N(5K - 6) + 11$ | $2K(NK + 3) - 6$ | $(2NK + \frac{K}{2} + 3)(K-1)-2$ | $NQ(2N + \frac{7Q}{2} + \frac{1}{2}) - 4$ |
| No. of ADD | $K(2NK + 3) - 5$ | $2N(5K - 6) + 4$ | $2NK^2 + 2K - 3$ | $K(K - 1)(2N + \frac{K}{6} - \frac{1}{3}) + K - 5$ | $QN(2N + \frac{Q^2}{6} + 4Q - 6) - (Q^2 - Q + 8) + 5N$ |
| No. of LN | 1 | 1 | 1 | 0 | 0 |
| No. of ABS | 0 | 0 | 0 | $\frac{K(K - 1)}{2}$ | $\frac{NQ(Q - 1)}{2}$ |
| No. of PH | 0 | 0 | 0 | $\frac{K(K - 1)}{2}$ | $\frac{NQ(Q - 1)}{2}$ |
| No. of DIV | 1 | 1 | 2 | 1 | N |

31

# Chapter 3

# Blind Maximum-Likelihood Carrier-Frequency-Offset Estimation for Interleaved OFDMA Uplink Systems
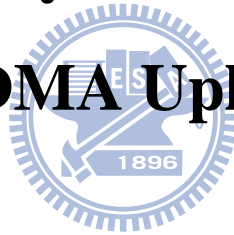
In this chapter, we will study the CFO estimation problem in interleaved OFDMA systems [34]. Blind ML CFO estimation is considered to be difficult in interleaved OFDMA uplink systems. This is because multiple CFOs have to be simultaneously estimated (each corresponding to a user's carrier), and an exhaustive multi-dimensional search is often required. The computational complexity of the search may be prohibitively high. Methods such as MUSIC [42] and ESPRIT [35] have been proposed as alternatives. However, these methods cannot maximize the likelihood function and the performance is not optimal. In this chapter, we propose a new method to solve the problem. With our formulation, the likelihood function can be maximized and the optimum solution can be obtained by solving a polynomial function. Compared to the exhausted search, the computational complexity of the proposed algorithm can be reduced dramatically. Section 3.1 briefly describes the system model and derives the proposed CFO estimation method. Section 3.2 shows the performance analysis for the proposed method. Finally, Section 3.3 evaluates the performance of the proposed method and analyze computational

complexity.

# § 3.1   The Propposed CFO Estimation Method

### § 3.1.1   *Signal Model for Interleaved OFDMA Uplink System*

In an OFDMA system, let $M$ users share the $N_s$ subcarriers of an OFDM symbol and the $M$ users simultaneously transmit their data streams. The subcarriers are divided into $Q$ subchannels and each subchannel has $N = N_s/Q$ subcarriers. Each user occupies a specific subchannel, and the subcarriers assigned to user $m$ are denoted as $s_m^k$'s where $k \in \Upsilon_m$. Here, $\Upsilon_m$ denotes a subset of subcarrier indices. For an interleaved OFDMA system [34] [35], the subset for the $m$th user is defined as : $\Upsilon_m = \{q_m, Q + q_m, \cdots, q_m + (N-1)Q\}$ where $q_m$ is the subchannel index and $q_m \in \{0, 1, \cdots, Q-1\}$. In the system, it is assumed that $\Upsilon_m \bigcap \Upsilon_k = \phi$ for $m \neq k$ where $\phi$ denotes the empty set. In our system, we assume that the sequence each user transmits is unknown to the BS and $M < Q$.

Consider a specific OFDMA symbol and denote the frequency domain signal that user $m$ transmits as an $N_s \times 1$ vector, $\mathbf{u}_m$. Note that the elements of $\mathbf{u}_m$ are nonzero only in designated subcarriers, i.e., $\Upsilon_m$. Taking the inverse discrete Fourier transform (IDFT) of $\mathbf{u}_m$, we can obtain the time domain signal for user $m$, denoting as $\bar{\mathbf{s}}_m = [\bar{s}_m(0), \cdots, \bar{s}_m(N_s - 1)]^T$. Inserting a cyclic prefix (CP) of length $L$ at the beginning of the symbol, user $m$ can then serially transmit the resultant signal through a wireless channel. Let the channel response from user $m$ to the BS receiver be denoted as $h_m(l), l = 0, \ldots, L_m - 1$, where $L_m$ is the channel length and $L_m \leq L$. Also, let the normalized CFO for user $m$ be denoted as $\varepsilon_m$. Then, the CP-removed received OFDMA symbol at the BS can be expressed as

$$
\begin{aligned}
y(k) &= \sum_{m=1}^{M} \exp(j2\pi\varepsilon_m k/N_s) \\
&\quad \sum_{l=0}^{L_m - 1} h_m(l)\bar{s}_m(k - l) + \eta(k),
\end{aligned}
\tag{3.1}
$$

34

where $k = 0, \cdots, N_s - 1$ and $\eta(k)$ represents additive white Gaussian noise (AWGN) with a variance of $\sigma_\eta^2$.

As mentioned, subchannel $q_m$ is assigned to user $m$ in the interleaved OFDMA system. It is equivalent to say that user $m$ is assigned to subchannel zero and an CFO of $q_m$ is introduced. So the received noiseless symbol from user $m$ can be re-written as

$$
\begin{aligned}
\bar{x}_m(k) &= \exp(j2\pi\varepsilon_m k/N_s) \sum_{l=0}^{L_m-1} h_m(l)\bar{s}_m(k-l) \\
&= \exp(j2\pi(\varepsilon_m + q_m)k/N_s) \sum_{l=0}^{L_m-1} h_m(l)s_m(k-l) \\
&= w^{\varepsilon_{e,m}\cdot k}x_m(k),
\end{aligned}
\tag{3.2}
$$

where $w = \exp(j2\pi/N_s)$, $\varepsilon_{e,m} = \varepsilon_m + q_m$, $x_m(k) = \sum_{l=0}^{L_m-1} h_m(l)s_m(k-l)$, and $s_m(k)$ is the transmitted signal of user $m$ if subchannel zero is assigned. The term $\varepsilon_{e,m}$ denotes the effective CFO for user $m$. It includes the virtual CFO caused by the subchannel $q_m$. Note that the periodicity of the transmitted sequence still remains after it is passed through the channel. Since the time domain signal has a period of $N$, we can make an index transformation by letting $k = (p-1)N + n$, where $p = 1, \cdots, Q$ and $n = 0, \cdots, N-1$. With the transformation, we can convert the $k$th sample of a signal into the $n$th sample in the $p$th period. The $n$th sample in each period, corresponding to a signal, can then be extracted to form a vector. Then, we have

$$
\mathbf{y}(n) = \mathbf{U}\mathbf{D}(n)\mathbf{x}(n) + \boldsymbol{\eta}(n),
\tag{3.3}
$$

where $\mathbf{y}(n) = [y(n), y(n+N), \ldots, y(n+(Q-1)N)]^T = [y_1(n), y_2(n), \ldots, y_Q(n)]^T$, $\mathbf{U}$ is a $Q$-by-$M$ matrix and $(\mathbf{U})_{p,q} = w^{\{\varepsilon_{e,q}\cdot(p-1)N\}}$, $\mathbf{D}(n) = \mathrm{diag}([w^{n\cdot\varepsilon_{e,1}}, \ldots, w^{n\cdot\varepsilon_{e,M}}]^T)$, $\mathbf{x}(n) = [x_1(n), \ldots, x_M(n)]^T$, and $\boldsymbol{\eta}(n) = [\eta(n), \eta(n+N), \ldots, \eta(n+(Q-1)N)]^T = [\eta_1(n), \eta_2(n), \ldots, \eta_Q(n)]^T$. We will use (3.3) as our signal model in the derivation of the ML CFO estimate.

## § 3.1.2  *Proposed Method*

To the best of our knowledge, blind ML CFO estimation has not been studied before in OFDMA uplink systems. Here, we propose a method to solve the problem. For interleaved OFDMA

uplink systems, the transmitted time-domain signal is obtained from the IDFT of its frequency-domain signal. From the central limit theorem, we know that if the number of subcarriers is reasonably large, the corresponding time-domain signal can be approximated as a white Gaussian sequence. Similar to [28], we assume that each users is under perfect power control, so signals arrive at the BS with equal average power. If we further assume that each channel tap experiences independently Rayleigh fading, and all users's signals are white and independent of each other, the received sequence $y(k)$ in (3.1) can also be approximated as a Gaussian sequence (see Chapter 2) with a variance of $M\sigma_x^2 + \sigma_\eta^2$, where $\sigma_x^2 = E\{|x_m(n)|^2\}$. Let $-0.5 < \varepsilon_m < 0.5$ and $f(.)$ be a probability density function. Then, we can explicitly write out the log-likelihood function, shown in Chapter 2, as

$$\Lambda(\boldsymbol{\varepsilon}) = \ln\{\prod_{n=0}^{N-1} f(\mathbf{y}(n))\}. \tag{3.4}$$

Define $\mathbf{R}_y = E\{\mathbf{y}(n)\mathbf{y}^H(n)\}$ and

$$(\mathbf{R}_y)_{p,q} = \sigma_\eta^2 \delta(p-q) + \sigma_x^2 \Gamma(p,q), \tag{3.5}$$

where

$$\Gamma(p,q) = \sum_{m=1}^{M} w^{(\varepsilon_{e,m})N(p-q)}. \tag{3.6}$$

Thus, we can express $f(\mathbf{y}(n))$ as [39] [40]

$$f(\mathbf{y}(n)) = (\pi^Q \det(\mathbf{R}_y))^{-1} \exp\left[-\mathbf{y}(n)^H \mathbf{R}_y^{-1} \mathbf{y}(n)\right]. \tag{3.7}$$

The log-likelihood function can be expressed as

$$\Lambda(\boldsymbol{\varepsilon}) = \sum_{n=0}^{N} \{-Q \cdot \ln(\pi) - \ln(\det(\mathbf{R}_y)) - \mathbf{y}(n)^H \mathbf{R}_y^{-1} \mathbf{y}(n)\}. \tag{3.8}$$

Let $\mathbf{u}(n) = \mathbf{U}\mathbf{D}(n)\mathbf{x}(n)$. Then, $\mathbf{y}(n) = \mathbf{u}(n) + \mathbf{w}(n)$. As assumed, the transmitted sequences are independent of each other, i.e., $\mathbf{R}_y = \sigma_x^2 \mathbf{U}\mathbf{U}^H + \sigma_\eta^2 \mathbf{I}$. Note that $\mathbf{U}$ is a $Q$-by-$M$ matrix. In order to use (3.8) and solve the $M$ unknown CFOs, $\mathbf{U}$ must be a full-rank tall matrix. From

(3.3), we see that $\mathbf{U}$ is a Vandermonde matrix ($\varepsilon_{e,m} \neq \varepsilon_{e,n}$ if $m \neq n$) [52]. Since we assume $M < Q$, the full-rank property then holds. As a result, (3.8) can be applied.

In order to find the maximum of the log-likelihood function for $i$th user, we take a derivative with respect to $\varepsilon_{e,i}$ [51]:

$$
\begin{aligned}
\frac{\partial}{\partial \varepsilon_{e,i}} \Lambda(\boldsymbol{\varepsilon}) &= -N \cdot \mathrm{tr}[\mathbf{R}_y^{-1} \frac{\partial}{\partial \varepsilon_{e,i}} \mathbf{R}_y] \\
&\quad - \sum_{n=0}^{N} \{\mathbf{y}(n)^H [(\frac{\partial}{\partial \varepsilon_{e,i}} \mathbf{R}_y^{-1}) \mathbf{y}(n)] \}.
\end{aligned}
\tag{3.9}
$$

We use the matrix inversion lemma [17] to write the inverse of $\mathbf{R}_y$ as

$$
\begin{aligned}
\mathbf{R}_y^{-1} &= \sigma_\eta^{-2} \mathbf{I} - \sigma_\eta^{-4} \mathbf{U}(\sigma_x^{-2}\mathbf{I} + \sigma_\eta^{-2}\mathbf{U}^H\mathbf{U})^{-1}\mathbf{U}^H \\
&= \sigma_\eta^{-2}\mathbf{I} - \sigma_\eta^{-4}\mathbf{U}(\mathbf{R}_s)^{-1}\mathbf{U}^H,
\end{aligned}
\tag{3.10}
$$

where $\mathbf{R}_s = \sigma_x^{-2}\mathbf{I} + \sigma_\eta^{-2}\mathbf{U}^H\mathbf{U}$. With (3.10), we only need the inverse of an $M$-by-$M$ matrix $\mathbf{R}_s$ rather than a $Q$-by-$Q$ matrix $\mathbf{R}_y$.

However, $(\mathbf{R}_s)^{-1}$ is difficult to obtain. Even it can, the relationship between the likelihood function and the CFOs may not be tractable after the inversion. To solve the problem, we propose using the Neumann series to expand $(\mathbf{R}_s)^{-1}$ [49]. Let $\mathbf{S}$ be a nonsingular matrix and $\beta(\mathbf{S})$ be its maximum absolute eigenvalue. Then, the series $\sum_{k=0}^{\infty} \mathbf{S}^k$ will converge to $(\mathbf{I} - \mathbf{S})^{-1}$ [52] if $\beta(\mathbf{S}) < 1$ [50]. However, the condition of $\beta(\mathbf{S}) < 1$ is not always satisfied for a nonsingular $\mathbf{S}$. This problem can be overcome by dividing $\mathbf{S}$ by a real parameter $\lambda > 0$ and expanding the resultant matrix. It is simple to show that there always exist a $\lambda$ such that $\beta(\mathbf{R}_s/\lambda) < 1$. Now, we can rewrite $\mathbf{R}_s$ as

$$
\begin{aligned}
\mathbf{R}_s &= \sigma_x^{-2}\mathbf{I} + \sigma_\eta^{-2}\mathbf{U}^H\mathbf{U} \\
&= \lambda(\mathbf{I} + \mathbf{B}),
\end{aligned}
\tag{3.11}
$$

where $\mathbf{B}$ is obtained as $(1/\lambda)\mathbf{R}_s - \mathbf{I}$, and its $(p,q)$th element is

$$
(\mathbf{B})_{p,q} = (\frac{1}{\lambda\sigma_x^2} - 1)\delta(p - q) + \frac{1}{\lambda\sigma_\eta^2} \sum_{k=1}^{Q} w^{(-\varepsilon_{e,p} + \varepsilon_{e,q})N(k-1)}.
\tag{3.12}
$$

From the Neumann series shown above, the inverse of $\mathbf{R}_s$ can be expanded as

$$
\begin{aligned}
(\frac{1}{\lambda}\mathbf{R}_s)^{-1} &= (\mathbf{I} + \mathbf{B})^{-1} \\
&= \sum_{k=0}^{\infty}(-1)^k \mathbf{B}^k
\end{aligned}
\tag{3.13}
$$

For simplicity, we can retain the first three and truncate high order terms, i.e.,

$$
\mathbf{R}_s^{-1} \approx \frac{1}{\lambda}\sum_{k=0}^{2}(-1)^k \mathbf{B}^k.
\tag{3.14}
$$

The determination of the optimum $\lambda$ and the analysis of the truncation error will be discussed in the next section. From (3.12), we can find the $(p,q)$th element of $\mathbf{B}^2$ as

$$
\begin{aligned}
(\mathbf{B}^2)_{p,q} &= (\frac{1}{\lambda\sigma_x^2} - 1)^2\delta(p - q) \\
&+ \frac{2}{\lambda\sigma_\eta^2}(\frac{1}{\lambda\sigma_x^2} - 1)\Gamma_0(p,q) \\
&+ (\frac{1}{\lambda\sigma_\eta^2})^2\sum_{k=1}^{M}\Gamma_0(p,k)\Gamma_0(k,q),
\end{aligned}
\tag{3.15}
$$

where $\Gamma_0(p,q) = \sum_{n=1}^{Q} w^{(-\varepsilon_{e,p}+\varepsilon_{e,q})N(n-1)}$. Substituting (3.12) and (3.15) into (3.14), we then obtain

$$
\begin{aligned}
(\mathbf{R}_s^{-1})_{p,q} &= \frac{1}{\lambda}[(2 - \frac{1}{\sigma_x^2\lambda} + (\frac{1}{\sigma_x^2\lambda} - 1)^2)\delta(p - q) \\
&+ \frac{1}{\sigma_\eta^2\lambda}(\frac{2}{\sigma_x^2\lambda} - 3)\Gamma_0(p,q) \\
&+ (\frac{1}{\sigma_\eta^2\lambda})^2\sum_{k=1}^{M}\Gamma_0(p,k)\Gamma_0(k,q)].
\end{aligned}
\tag{3.16}
$$

38

Using (3.16) in (3.10), we can approximate the inverse of $\mathbf{R}_y$ as

$$
\begin{aligned}
(\mathbf{R}_y^{-1})_{p,q} &= \sigma_\eta^{-2}\delta(p-q) - C_0\Gamma(p,q) \\
&\quad - C_1 \sum_{a=1}^{M} w^{(\varepsilon_{e,a})N(p-1)} \\
&\quad \cdot \sum_{b=1}^{M} \Gamma_0(a,b)w^{(-\varepsilon_{e,b})N(q-1)} \\
&\quad - C_2 \sum_{a=1}^{M} w^{-(\varepsilon_{e,a})N(q-1)} \sum_{b=1}^{M} w^{(\varepsilon_{e,b})N(p-1)} \\
&\quad \cdot \sum_{k=1}^{M} \Gamma_0(b,k)\Gamma_0(k,a),
\end{aligned}
\tag{3.17}
$$

where
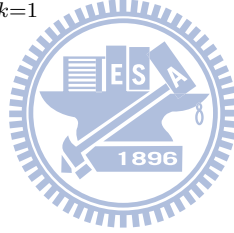
$$
\begin{aligned}
C_0 &= \frac{1}{\sigma_\eta^4\lambda}(2 - \frac{1}{\sigma_x^2\lambda} + (\frac{1}{\sigma_x^2\lambda} - 1)^2) \\
C_1 &= \frac{1}{\sigma_\eta^6\lambda^2}(\frac{2}{\sigma_x^2\lambda} - 3) \\
C_2 &= \frac{1}{\sigma_\eta^8\lambda^3}.
\end{aligned}
\tag{3.18}
$$

Note that $\Gamma_0(.,.)$ can not be directly estimated. However, it can be combined with some variables in (3.17) and converted to $\Gamma(.,.)$ as defined in (3.6). The value of $\Gamma(p,q)$ can be estimated from that of the $(p-q)$th diagonal term of $\mathbf{R}_y$ as:

$$
\Gamma(p,q) = \begin{cases} \dfrac{\sum\limits_{p=1}^{Q-m}[\mathbf{R}_y]_{p,p+m}}{(Q-m)\sigma_x^2} & \text{, if } m = q - p \geq 0 \\ \dfrac{\sum\limits_{q=1}^{Q-m}[\mathbf{R}_y]_{q+m,q}}{(Q-m)\sigma_x^2} & \text{, if } m = p - q > 0. \end{cases}
\tag{3.19}
$$

39

The second and third terms in (3.17) can be re-written as

$$\sum_{a=1}^{M}\sum_{b=1}^{M} \Gamma_0(a,b)w^{(\varepsilon_{e,a})N(p-1)}w^{(-\varepsilon_{e,b})N(q-1)}$$

$$= \sum_{n=1}^{Q}\sum_{a=1}^{M}\sum_{b=1}^{M}w^{(-\varepsilon_{e,a}+\varepsilon_{e,b})N(n-1)}$$

$$\cdot \; w^{(\varepsilon_{e,a})N(p-1)}w^{(-\varepsilon_{e,b})N(q-1)}$$

$$= \sum_{n=1}^{Q}\Gamma(p,n)\Gamma(n,q) \tag{3.20}$$

and

$$\sum_{a=1}^{M} w^{-(\varepsilon_{e,a})N(q-1)} \sum_{b=1}^{M} w^{(\varepsilon_{e,b})N(p-1)} \sum_{k=1}^{M} \Gamma_0(b,k)\Gamma_0(k,a)$$

$$= \sum_{m=1}^{Q}\sum_{n=1}^{Q}\Gamma(p,m)\Gamma(m,n)\Gamma(n,q). \tag{3.21}$$

Substituting (3.17)-(3.21) into (3.9), we can obtain

$$\frac{\partial}{\partial \varepsilon_{e,i}}\Lambda(\varepsilon) = \frac{j2\pi N}{N_s}\{\sum_{p=1}^{Q}\sum_{q=1}^{Q}\{N\sigma_x^2(q-p)x^{q-p}$$

$$[C_0\Gamma(p,q) + C_1\sum_{n=1}^{Q}\Gamma(p,n)\Gamma(n,q)$$

$$+C_2\sum_{m=1}^{Q}\sum_{n=1}^{Q}\Gamma(p,m)\Gamma(m,n)\Gamma(n,q)]$$

$$+\gamma(p,q)[C_0(p-q)x^{p-q}$$

$$+C_1\sum_{n=1}^{Q}((n-q)x^{n-q}\Gamma(p,n)$$

$$+(p-n)x^{p-n}\Gamma(n,q))$$

$$+C_2\sum_{m=1}^{Q}\Gamma(p,m)\sum_{n=1}^{Q}\Gamma(n,q)$$

$$((m-n)x^{m-n}$$

$$+(n-q)x^{n-q}\Gamma(p,m)\Gamma(m,n)$$

$$+(p-m)x^{p-m}\Gamma(m,n)\Gamma(n,q))]\}\} \tag{3.22}$$

40

where

$$\gamma(p, q) = \sum_{n=0}^{N-1} y^*(n + (p-1)N)y(n + (q-1)N) \tag{3.23}$$

and

$$x = \exp(j2\pi \frac{N\varepsilon_{e,i}}{N_s}). \tag{3.24}$$

The detailed derivation of (3.22) is provided in Appendix B.1. Setting (3.22) to zero , we can solve all the possible $2(Q-1)$ roots, $\hat{x}$'s. The effective CFO can then be obtained by

$$\hat{\varepsilon}_{e,i} = \frac{N_s}{N}(\frac{ln(\hat{x})}{j2\pi}). \tag{3.25}$$

As defined in (3.2), the true CFO for $i$th user is then given by

$$\hat{\varepsilon}_i = -q_i + \frac{N_s}{N}(\frac{ln(\hat{x})}{j2\pi}), \tag{3.26}$$

where $q_i$ is the subchannel index for user $i$. It is apparent that after adding $-q_i$, there will be only one root falling into the range of subchannel 0, and the root is the estimated CFO for user $i$.

A direct method for solving the roots in (3.22) is via an exhaustive grid search over the interval spanned by $\varepsilon_{e,i}$. However, the computational complexity is high. Taking a closer look at (3.22), we find that (3.22) is a polynomial function of $x$, i.e.,

$$\frac{\partial}{\partial \varepsilon_{e,i}} \Lambda(\boldsymbol{\varepsilon}) = \sum_{k=1}^{Q-1} \alpha_p(k)x^k + \sum_{k=1}^{Q-1} \alpha_n(k)x^{-k} = 0. \tag{3.27}$$

The detailed derivation for $\alpha_p(k)$ and $\alpha_n(k)$ is provided in Appendix B.2. Using (3.27), we can then use a more efficient root-finding method to obtain the roots.

## § 3.2  Performance Analysis

### § 3.2.1  *Truncation Error in (3.14)*

As we can see, the series in (3.13) is infinite and truncation has to be conducted. In the previous section, we retain the first three terms in the series. One may be curious about how large the

error will be. In this subsection, we analyze the truncation error in (3.14).

For a positive-definite Hermitian matrix $\mathbf{R}$ with rank $K$, we can have its eigen-decomposition as

$$\mathbf{R} = \mathbf{V}\mathbf{G}\mathbf{V}^H, \tag{3.28}$$

where $\mathbf{G} = \text{diag}[g_1, \cdots, g_K]$ is a diagonal matrix, $g_i$'s being positive are the eigenvalues of $\mathbf{R}$ with descending order, i.e., $g_1 > \cdots > g_K$, and $\mathbf{V}$ is an unitary matrix consisting of the eigenvectors. As shown in section 3.1, $\mathbf{R}$ also can be expressed as

$$(\frac{\mathbf{R}}{\lambda})^{-1} = (\mathbf{I} - \mathbf{A})^{-1} = \sum_{k=0}^{\infty} \mathbf{A}^k, \tag{3.29}$$

where $\lambda$ is a real number ensuring that the maximum absolute eigenvalue of $\mathbf{R}/\lambda$ is smaller than one, and $\mathbf{A}$ is a matrix to be determined. Substitute (3.28) and $\mathbf{V}\mathbf{V}^H = \mathbf{I}$ into (3.29), we can obtain that $\mathbf{A} = \mathbf{V}(\mathbf{I} - \mathbf{G}/\lambda)\mathbf{V}^H$ and

$$\mathbf{A}^k = \mathbf{V}(\mathbf{I} - \frac{\mathbf{G}}{\lambda})^k\mathbf{V}^H \tag{3.30}$$

From (3.30), it is simple to see that for the convergence of (3.29), $|1 - g_i/\lambda|$, $i = 1, 2, \ldots, K$, has to be smaller than one. Also, the smaller the value of $|1 - g_i/\lambda|$, the faster the convergence we can have. Since the values of $g_i$'s may be different, the convergent rate of each $|1 - g_i/\lambda|$ (referred to as a mode) may be different. As a result, the overall convergence is dominated by the mode with the maximum $|1 - g_i/\lambda|$. To have the fastest convergence, we then want to find a $\lambda$ minimizing the maximum $|1 - g_i/\lambda|$ ($1 \leq i \leq K$). This yields a min-max optimization problem as:

$$\min_{\lambda} \max_{i=1,\cdots,K} |1 - g_i/\lambda| \tag{3.31}$$

subject to the constraints

$$|1 - g_i/\lambda| < 1, \tag{3.32}$$

where $i = 1, 2, \ldots, K$. The optimum value of $\lambda$ has been shown to be [54]

$$\lambda = \frac{g_1 + g_K}{2}. \tag{3.33}$$

42

Substituting (3.33) into $|1 - g_i/\lambda|$, we find that there is a same maximum value yielded by $g_1$ and $g_K$. Denote the value as the slowest convergence rate (SCR) of $\mathbf{R}$, i.e.,

$$
\begin{aligned}
\mathcal{M}(\mathbf{R}) &= |1 - g_1/\lambda| = |1 - g_K/\lambda| = \frac{g_1 - g_K}{g_1 + g_K} \\
&= \frac{\mathcal{S}(\mathbf{R}) - 1}{\mathcal{S}(\mathbf{R}) + 1},
\end{aligned}
\tag{3.34}
$$

where $\mathcal{S}(\mathbf{R}) = g_1/g_K$ is the eigenvalue spread (EVS) of $\mathbf{R}$. It is obviously that a smaller EVS yields a smaller SCR. Furthermore, if the SCR is smaller, the convergence of the series of (3.29) will be faster and the truncation error will be smaller. However, a closed-form expression for the truncation error is difficult to obtain. Instead of the exact value of the error, we will try to derive an upper bound. Let the number of the terms retained in (3.29) be $\mathcal{L}$ and power of the truncation error be $\mathcal{E}$. Then, we have

$$
\begin{aligned}
\mathcal{E} &= \| \sum_{k=0}^{\infty} \mathbf{A}^k - \sum_{k=0}^{\mathcal{L}-1} \mathbf{A}^k \| = \| \sum_{k=\mathcal{L}}^{\infty} \mathbf{A}^k \| \\
&\leq \sum_{k=\mathcal{L}}^{\infty} \|\mathbf{A}^k\| \leq \sum_{k=\mathcal{L}}^{\infty} \|\mathbf{A}\|^k \\
&\leq \sum_{k=\mathcal{L}}^{\infty} \mathcal{M}^k(\mathbf{R})
\end{aligned}
\tag{3.35}
$$

where $\mathcal{M}(\mathbf{R})$ is the maximum diagonal value of $\mathbf{I} - \frac{\mathbf{G}}{\lambda}$ in (3.30). If $\mathcal{M}(\mathbf{R}) \neq 0$, then

$$
\mathcal{E} \leq \frac{\mathcal{M}(\mathbf{R})^{\mathcal{L}}}{1 - \mathcal{M}(\mathbf{R})} = \frac{(\mathcal{S}(\mathbf{R}) - 1)^{\mathcal{L}}}{2(\mathcal{S}(\mathbf{R}) + 1)^{\mathcal{L}-1}}.
\tag{3.36}
$$

It is simple to see that when $\mathcal{S}(\mathbf{R}) = 1$ and $\mathcal{M}(\mathbf{R}) = 0$, $\mathcal{E} = 0$ giving the fastest convergence of (3.29). In this case, $\mathbf{R}$ is diagonal and only one term is required in (3.29).

We now compare the EVSs of $\mathbf{R}_y$ and $\mathbf{R}_s$ in (3.10), and show that $\mathcal{S}(\mathbf{R}_s) < \mathcal{S}(\mathbf{R}_y)$. As shown in Section 3.1, $\mathbf{R}_y = \sigma_x^2 \mathbf{U}\mathbf{U}^H + \sigma_\eta^2 \mathbf{I}$ and $\mathbf{R}_s = \sigma_x^{-2}\mathbf{I} + \sigma_\eta^{-2}\mathbf{U}^H\mathbf{U} = \sigma_x^{-2}\mathbf{I} + \sigma_\eta^{-2}\mathbf{R}_u$. Let $\{g_{u,1}, \ldots, g_{u,Q}\}$ be the eigenvalues of $\mathbf{U}\mathbf{U}^H$ and $g_{u,1} \geq \ldots \geq g_{u,Q}$. Since the rank of $\mathbf{U}$, a $Q$-by-$M$ matrix, is $M$, the smallest $Q-M$ eigenvalues of $\mathbf{U}\mathbf{U}^H$ are zero, i.e. $g_{u,M+1} = \cdots = g_{u,Q} = 0$. And, the non-zero eigenvalues of $\mathbf{U}\mathbf{U}^H$ and $\mathbf{U}^H\mathbf{U}$ are the same. This indicates that we can obtain

the eigenvalues of $\mathbf{R}_s$ from $\mathbf{R}_y$ as:

$$\text{eig}(\mathbf{R}_s) = \sigma_x^{-2} + \sigma_\eta^{-2}\text{eig}(\mathbf{R}_u), \tag{3.37}$$

where $\text{eig}(\mathbf{R}_u)$ denotes the first $M$ eigenvalues of $\mathbf{R}_u$. So, the EVSs of $\mathbf{R}_y$ and $\mathbf{R}_s$ can be easily obtained as:

$$
\begin{aligned}
\mathcal{S}(\mathbf{R}_y) &= (\sigma_x^2 g_{u,1} + \sigma_\eta^2)/(\sigma_\eta^2) \\
&= \rho \cdot g_{u,1} + 1,
\end{aligned}
\tag{3.38}
$$

and

$$
\begin{aligned}
\mathcal{S}(\mathbf{R}_s) &= (\sigma_x^{-2} + \sigma_\eta^{-2} g_{u,1})/(\sigma_x^{-2} + \sigma_\eta^{-2} g_{u,M}) \\
&= (\rho \cdot g_{u,1} + 1)/(\rho \cdot g_{u,M} + 1),
\end{aligned}
\tag{3.39}
$$

where $\rho = \sigma_x^2/\sigma_\eta^2$. Note that the received SNR is defined as SNR$= M\sigma_x^2/\sigma_\eta^2 = M\rho$. It is easy to see that the EVS of $\mathbf{R}_s$ is smaller than that of $\mathbf{R}_y$. Therefore, the SCR of $\mathbf{R}_s$ is smaller than that of $\mathbf{R}_y$. Thus, the matrix inversion lemma used in (3.10) not only reduces the computational complexity, but also reduces the truncation error in (3.14).

From (3.39), it is also simple to see that for low SNR, the EVS of $\mathbf{R}_s$ approaches one, so the truncation error in (3.14) can be ignored. For high SNR, the EVS of $\mathbf{R}_s$ approaches to that of $\mathbf{R}_u$. The EVS of $\mathbf{R}_u = \mathbf{U}^H\mathbf{U}$ depends on the subchannel assignment since the $(a,b)$th entry of $\mathbf{U}$ is $w^{\{\varepsilon_{e,b}\cdot(a-1)N\}}$ and $\varepsilon_{e,b} = \varepsilon_b + q_b$. To analyze the truncation error, we first have to analyze the EVS of $\mathbf{U}^H\mathbf{U}$. Unfortunately, a general closed-form for the EVS is difficult to obtain. Here, we study two special cases to show that the EVS of $\mathbf{U}^H\mathbf{U}$ is low and the truncation error in (3.14) can be small in our applications.

Define a $M_{max}$-user system as a system which can simultaneously handle $M_{max}$ users at most. The first case we consider is a 2-user system. For the system, the EVS of $\mathbf{U}^H\mathbf{U}$ can be solved from (3.3) in a closed-form as:

$$\mathcal{S}(\mathbf{U}^H\mathbf{U}) = \frac{Q + \vartheta_{1,2}}{Q - \vartheta_{1,2}}, \tag{3.40}$$

44

where

$$
\begin{aligned}
\vartheta_{1,2} &= \{[1 - \cos(2\pi\delta_\varepsilon) - \cos(2\pi\delta_\varepsilon/Q) \\
&\quad + \cos(2\pi\delta_\varepsilon(Q-1)/Q)]^{0.5} \\
&\quad \cdot \ [1 - \cos(2\pi\delta_\varepsilon/Q)]^{-0.5}.
\end{aligned}
$$

As we can see, the EVS varies with $\delta_\varepsilon = |\varepsilon_{e,1} - \varepsilon_{e,2}| = |q_1 - q_2 + \varepsilon_1 - \varepsilon_2|$. Therefore, $\mathcal{S}(\mathbf{U}^H\mathbf{U})$ is dependent on $\Delta\varepsilon = |\varepsilon_1 - \varepsilon_2|$ and $\Delta q = |q_1 - q_2|$. Note that $\Delta q$ is the difference of the neighbor subchannel indices. We now use an example to exam the EVS of $(\mathbf{U}^H\mathbf{U})$. Let $N_s = 128$ and $N = 8$. Then, there are $N_s/N = 16$ subchannels and $1 \leq \Delta q \leq 15$. Figure 3.1 shows the result. From the figure, we can see that the EVSs for $\Delta q = 1$ and $\Delta q = 15$ are much larger than those for $2 \leq \Delta q \leq 14$. Figure 3.2 shows the EVSs for $2 \leq \Delta q \leq 14$. It is clear that all the values are smaller than 1.5. This indicates that the truncation error will be small as long as adjacent subchannels are not used simultaneously, i.e., $|\mathrm{mod}(\Delta q - i \cdot Q, Q)| \neq 1$ where $i$ is an integer.

The second case we consider is a 4-user system. For the system, the closed-form solution of the EVS is not obtainable. Simulations are then conducted to obtain numerical results. Note that in this case, the EVS is a function of $\{q_1, \ldots, q_4, \varepsilon_1, \ldots, \varepsilon_4\}$. It is difficult to exam the behavior of the EVS in terms of these variables. For convenience, we still define two variables: $\Delta q = |q_1 - q_2| = |q_2 - q_3| = |q_3 - q_4|$ and $\Delta\varepsilon = \sum_{i=1}^{M-1} |\varepsilon_{i+1} - \varepsilon_i|$. Note that the definition and the implication of $\Delta\varepsilon$ are different from those in the 2-user case. Using the same simulation setting as that in the 2-user system, we obtain the EVS versus $\Delta\varepsilon$ in Figure 3.3. Here, we assume that adjacent subchannels are not used. From the figure, we can find that all the EVSs are smaller than 1.7. We can expect the truncation error in (3.14) will be small. From Figure 3.2 and 3.3, we can also find that the smallest truncation error can be obtained when $\Delta q = 8$ and $\Delta q = 4$ for the 2-user and 4-user systems, respectively. However, a large $\Delta q$ will result in a smaller $M_{max}$, the maximum number of the users. Thus, the selection of $\Delta q$ is a tradeoff between $M_{max}$ and the SCR of $\mathbf{R}_u$. In this study, we let the smallest $\Delta q$ be two.

Using (3.33) and (3.37), we can see that the optimum $\lambda$ in (3.11) is equal to $\sigma_x^{-2} + \sigma_\eta^{-2}(g_{u,1} + g_{u,M})/2$. Note that the eigenvalues of $\mathbf{R}_u$ can be estimated as $\text{eig}(\sigma_x^{-2}(\mathbf{R}_y - \sigma_\eta^2 \mathbf{I}))$ and $\mathbf{R}_y$ can be estimated as $\sum_{n=0}^{N-1} \mathbf{y}(n)\mathbf{y}^H(n)$. So, the optimum $\lambda$ can then be calculated. To evaluate the performance of the proposed expansion, we define a normalized truncation error as (with the optimum $\lambda$):

$$\mathcal{E}_n = E_{\boldsymbol{\varepsilon}}\{\|\mathbf{R}_s/\lambda \cdot [(\mathbf{R}_s/\lambda)^{-1} - \sum_{k=0}^{2}(-1)^k \mathbf{B}^k]\|\}, \tag{3.41}$$

where $\boldsymbol{\varepsilon}$ is the set for all possible $\varepsilon_{e,m}$ ($m = 1, \cdots, M$). Using the result in (3.36), we can also define a normalized upper bound as

$$\begin{aligned}\mathcal{E}_n &\leq& E_{\boldsymbol{\varepsilon}}\{\|\mathbf{R}_s/\lambda\| \cdot \|(\mathbf{R}_s/\lambda)^{-1} - \sum_{k=0}^{2}(-1)^k \mathbf{B}^k\|\} \\ &\leq& E_{\boldsymbol{\varepsilon}}\{\|\mathbf{R}_s/\lambda\| \frac{(\mathcal{S}(\mathbf{R}_s/\lambda) - 1)^3}{2(\mathcal{S}(\mathbf{R}_s/\lambda) + 1)^2}\}. \end{aligned} \tag{3.42}$$

We now use some examples to evaluate the normalized truncation error and its upper bound in (3.42). The result is shown in Table 3.1. It is simple to see that the normalized truncation error increases with the decrease of $\Delta q$ and with the increase of SNR. Also, the deviation of the upper bound from the actual error is small; the upper bound overestimates the error by one to two dB. From Table 3.1, we can also see that even with $\Delta q = 2$, the truncation error is still quite small, i.e., $-20$dB. Section 3.3 gives more results to show the property.

## § 3.2.2 *CRB Analysis*

For the training-based method, only the AWGN is considered as a random variable, and the CRB for the CFO estimation can then be derived [29]. In the blind method, the transmit symbol is treated as an additional random variable, and the CRB for the blind CFO estimation can also be derived in Chapter 2. Here, we generalize the result in Chapter 2 (for single-user OFDM systems) to derive the CRB in OFDMA systems. From (3.4), the $(p, q)th$ entry of the Fisher information matrix $\mathbf{F}$ is given by

$$(\mathbf{F})_{p,q} = -E\{\frac{\partial^2 \ln(\Lambda(\boldsymbol{\varepsilon}))}{\partial \varepsilon_p \partial \varepsilon_q}\} = -E\{\frac{\partial^2 \ln(\Lambda(\boldsymbol{\varepsilon}))}{\partial \varepsilon_{e,p} \partial \varepsilon_{e,q}}\}, \tag{3.43}$$

where $1 \leq p, q \leq M$. Substituting (3.8) into (3.43) yields

$$
\begin{aligned}
(\mathbf{F})_{p,q} &= N \cdot \text{tr}(-\mathbf{R}_y^{-1}\frac{\partial \mathbf{R}_y}{\partial \varepsilon_p}\mathbf{R}_y^{-1}\frac{\partial \mathbf{R}_y}{\partial \varepsilon_q} + \mathbf{R}_y^{-1}\frac{\partial^2 \mathbf{R}_y}{\partial \varepsilon_p \partial \varepsilon_q}) \\
&\quad -E\{\sum_{n=0}^{N-1} \mathbf{y}^H(n)\mathbf{J}\mathbf{y}(n)\} \\
&= N \cdot \text{tr}(\mathbf{R}_y^{-1}\frac{\partial \mathbf{R}_y}{\partial \varepsilon_p}\mathbf{R}_y^{-1}\frac{\partial \mathbf{R}_y}{\partial \varepsilon_q}),
\end{aligned}
\tag{3.44}
$$

where

$$
\begin{aligned}
\mathbf{J} &= -\mathbf{R}_y^{-1}\frac{\partial \mathbf{R}_y}{\partial \varepsilon_p}\mathbf{R}_y^{-1}\frac{\partial \mathbf{R}_y}{\partial \varepsilon_q}\mathbf{R}_y^{-1} \\
&\quad +\mathbf{R}_y^{-1}(\frac{\partial^2 \mathbf{R}_y}{\partial \varepsilon_p \partial \varepsilon_q})\mathbf{R}_y^{-1} \\
&\quad -\mathbf{R}_y^{-1}\frac{\partial \mathbf{R}_y}{\partial \varepsilon_p}\mathbf{R}_y^{-1}\frac{\partial \mathbf{R}_y}{\partial \varepsilon_q}\mathbf{R}_y^{-1}.
\end{aligned}
\tag{3.45}
$$

From [51], we see that

$$
\begin{aligned}
E\{\sum_{n=0}^{N-1} \mathbf{y}^H(n)\mathbf{J}\mathbf{y}(n)\} &= \sum_{n=0}^{N-1} E\{\mathbf{y}^H(n)\mathbf{J}\mathbf{y}(n)\} \\
&= \sum_{n=0}^{N-1} E\{\text{tr}(\mathbf{J}\mathbf{y}(n)\mathbf{y}^H(n))\} \\
&= N \cdot \text{tr}(\mathbf{J}\mathbf{R}_y).
\end{aligned}
\tag{3.46}
$$

Finally, the CRB for the $\varepsilon_i$ estimate is obtained as

$$
CRB(\varepsilon_i) = (\mathbf{F}^{-1})_{i,i}.
\tag{3.47}
$$

We average the diagonal terms of (3.47) to have a single index for performance comparison.

## § 3.2.3  *Computational Complexity*

Here, the computational complexity of the proposed method is assessed and compared with that of the existing schemes. For the proposed method, there are three operation steps. In the

47

first step, we need to calculate the correlation matrix in (3.5) and its eigen-decomposition to obtain (3.37). The computational complexity for this step is $O(Q^3 + Q^2N)$. In the second step, we need to calculate the coefficients in (3.27). From the received signal and the known correlation matrix, we can obtain $\gamma(p,q)$ in (3.23) and $\Gamma(p,q)$ in (3.19). Let $\Gamma(p,q)$ and $\gamma(p,q)$ be the $(p,q)th$ entries of two matrices $\mathbf{\Gamma}$ and $\boldsymbol{\gamma}$, respectively. The computational complexity for constructing these matrices is $O(Q^2N)$. Note that not all coefficients in (3.27) (please see (5.45) and (5.46) in Appendix B.2) are required to calculate. Some coefficient pairs in (5.45) and in (5.46) are complex conjugate each other. Also, some terms in (5.45) and (5.46) appear repeatedly. For example, $F_1(p,q)$ in (5.45), the 6th term of (5.45), and the 7th term of (5.45) all include $\sum_m \Gamma(p,m)\Gamma(m,n)$. Without the redundant computations, the complexity for calculating the coefficients in (5.45) or (5.46) is found to be $O(6Q^3)$. Then, we have evaluated the computational complexity for calculating all the coefficients in the polynomial (3.27). The last step is the root-searching process in (3.27) and the CFOs sorting for each user in (3.26). Since there are $2Q - 1$ terms in (3.27), the roots can be solved with the complexity of $O(8Q^3)$ (see Section 2.2). Compared with the root-searching process, the complexity in calculating (3.26) is small and can be ignored. Adding all together and taking only dominant terms, we can have the entire complexity for the proposed method is $O(15Q^3 + 2Q^2N)$. For the ESPRIT frequency estimator, the complexity has shown to be $O(5Q^3 + Q^2N)$ [35]. Therefore, the computational complexity of the proposed method is on the same complexity order as that of ESPRIT.

Next, we evaluate the computational complexity of the training-based schemes. For APFE, the total complexity has shown to be $O(MN_cN_w(L^3 + LN_s^2))$ [28] [29], where $N_c$ denotes the number of iterations and $N_w$ the number of grid points used for each iteration. For simplified AAPFE, the computational complexity is $O(MN_cN_wN_s^2K)$ [29] [30]. Since the computational complexity of APFE algorithm is high, suboptimum training-based schemes were then proposed [28] [30] [31]. For the method in [28], the computational complexity is $O(2MTN_s^2 + TN_s(MN)^2)$ where $T$ is the number of the Monte Carlo runs finding a mean likelihood [48].

The computational complexity for the method in [30] is $O(N_c(M^2N^2 + 1/2N_s\log_2(N_s) + 3/2M^3N^3 + 3/8M^2N^2N_s))$ while that in [31] is $O(N_c(N_s(MN)^2 + M^3))$ for $\mathcal{L} = 1$. Here, $\mathcal{L}$ is the number of terms retained in an infinite series [31]. Note that for $\mathcal{L} = 1$, the method has the worse performance but the lowest computational complexity. Also note that the simulation results in [31] indicates that $\mathcal{L}$ should be at least three for acceptable performance. However, as shown in [31], the computational complexity order for $\mathcal{L} = 3$ is difficult to evaluate.

## § 3.3 Simulations

In this section, we report simulation results demonstrating the effectiveness of the proposed method. In the first set of simulations, we compare the performance of the proposed method with existing blind methods. Note that ESPRIT is known to be better than the MUSIC algorithm [35]. Thus, we only conduct simulations for ESPRIT [35]. In the second set of simulations, we compare the performance of the proposed method with existing ML methods such as APFE and AAPFE. Note that existing ML schemes require training sequences. Finally, we compare the computational complexity of all schemes.

### § 3.3.1 *System Setup*

In our simulations, the channel response used for each user is generated according to the HIPER-LAN/2 channel model [47]. The channel response, having 6 taps, follows an exponential power decay profile. Each tap coefficient is modeled as an independent complex Gaussian random variable with zero mean. The CFO of each user is generated with a uniform distribution in the interval $(-0.5, 0.5)$. The symbols used for CFO estimation are modulated with a binary phase-shift keying (BPSK) scheme while those for data transmission with a 16-QAM scheme. The interleaved OFDMA system used in our simulations has $N_s = 128$ subcarriers. Since there are multiple CFOs to be estimated, the mean square error (MSE) is used as the performance index

defined as

$$MSE = \sum_{m=1}^{M} (\hat{\varepsilon}_m - \varepsilon_m)^2. \tag{3.48}$$

All the simulation results are obtained by averaging $1000$ Monte Carlo runs.

As shown in Section 3.2, $\Delta q$, the difference of the neighbor subchannel indices, influences the truncation error in (3.14) and $M_{max}$, the maximum number of the users. As shown in Section 3.2, the larger the $\Delta q$, the smaller the truncation error and the smaller the $M_{max}$. We compare the results for $\Delta q = 2$ and $\Delta q = 4$ in the following.

## § 3.3.2 *Performance Assessment for $\Delta q = 2$*

Firstly, we let $\Delta q = 2$, the smallest $\Delta q$ we use. In this case, the SCR will be maximal and $M_{max}$ is also maximal. It corresponds to the worst case in the proposed method. An important design parameter for the proposed method is the number of subcarriers for each user, $N$. The number of subchannels is then $Q = N_s/N$. Observing (3.27) and (3.47), we see that a higher $Q$ gives better performance, but requires higher complexity. To see the impact of $N$, we let $M = 4$ and observe the CRB for different $N$. The result is shown in Figure 3.4. We can see that the CRB is almost the same for $N = 4$ and $N = 8$. To reduce computations, we choose $N = 8$ for the simulations we conducted. Without loss of generality, we assume that the CP length is $N$ which is larger than the channel length. For the first set of simulations, we compare the performance of the proposed algorithm with that of ESPRIT. Figure 3.5 shows the result for $M = 2$, while Figure 3.6 for $M = 4$. As expected, the performance of the proposed algorithm is significantly better than that of ESPRIT since the proposed method conducts the ML estimation. We can also see that the proposed method can approach the CRB. At high SNR regions, the performance of the proposed algorithm slightly deviates from the CRB. This is due to our approximation used in (3.14). When the number of users is larger, the deviation is also larger.

In the second set of simulations, we compare the performance of the proposed algorithm with that of other ML algorithms. The simulation setup is the same as that of the first set of simulations. The ML problem can be directly solved by using an exhaustive grid-search over the

multi-dimensional space spanning $\{\varepsilon_{e,1}, \cdots, \varepsilon_{e,M}\}$. To reduce the computational complexity, we use the APFE and AAPFE schemes. The AAPFE is a suboptimum solution of APFE and it also truncates the Neumann series to approximate an inverse matrix (with an order $K$) [53]. In each iteration, only one user's CFO is updated while other users' CFOs remain unchanged. The CFO update is conducted by a grid-search method. Note that the purpose of the expansion is different from ours. Also, AAPFE does not use the optimum $\lambda$ to achieve the best result. In our simulations, we let $N_c = 2$ and $N_w = 100$. We have tried $K = 1$ and $K = 2$. Figure 3.7 shows the result for $M = 2$ and Figure 3.8 for $M = 4$. We can see that the APFE performs the best and the AAPFE with $K = 1$ performs the worst. Note that the conventional APFE and AAPFE have to use training sequences. From the figure, we see that the performance gaps between the APFE, the AAPFE with $K = 2$, and the proposed blind algorithm are very small. Also note that all these algorithms tend to deviate from the CRB when SNR is high.

## § 3.3.3 *Performance Assessment for $\Delta q = 4$*

Figure 3.10 shows the computational complexity of the schemes we consider. Figure 3.10(a) shows the complexity versus the number of subcarriers for the 2-user case. Figure 3.10(b) shows the complexity versus the number of users for a fixed total number of subcarriers, $N_S = 128$ and $Q = 8$. From the figures, we find that the computational complexity of the proposed method is similar to that of the conventional blind methods. However, the proposed method outperforms the conventional methods by 10dB (see Figure 3.5 and Figure 3.6). Compared to the training-based methods, the proposed blind method can have similar performance (see Figure 3.7 and Figure 3.8), but much lower computational complexity (see Figure 3.10).

Table 3.1: Normalized truncation error versus SNR

| SNR | 0dB | 3dB | 6dB | 9dB | 20dB | 30dB |
|---|---|---|---|---|---|---|
| M= 2, $\Delta q = 2$ (dB) | -26.6867 | -26.2986 | -26.0997 | -25.9989 | -25.9049 | -25.8976 |
| Upper bound in (3.42) | -25.2057 | -24.7714 | -24.5481 | -24.4348 | -24.3291 | -24.3209 |
| M= 2, $\Delta q = 3$ (dB) | -32.4019 | -32.0139 | -31.8150 | -31.7141 | -31.6202 | -31.6128 |
| Upper bound in (3.42) | -31.5097 | -31.0943 | -30.8811 | -30.7729 | -30.6720 | -30.6642 |
| M= 2, $\Delta q = 4$ (dB) | -35.8326 | -35.4446 | -35.2457 | -35.1449 | -35.0509 | -35.0436 |
| Upper bound in (3.42) | -35.1642 | -34.7558 | -34.5462 | -34.4399 | -34.3408 | -34.3331 |
| M= 4, $\Delta q = 2$ (dB) | -20.0507 | -19.6616 | -19.4621 | -19.3610 | -19.2667 | -19.2594 |
| Upper bound in (3.42) | -18.0176 | -17.5638 | -17.3303 | -17.2117 | -17.1010 | -17.0924 |
| M= 4, $\Delta q = 3$ (dB) | -25.1858 | -24.7970 | -24.5978 | -24.4967 | -24.4026 | -24.3952 |
| Upper bound in (3.42) | -23.8575 | -23.4277 | -23.2068 | -23.0947 | -22.9901 | -22.9820 |
| M= 4, $\Delta q = 4$ (dB) | -27.2568 | -26.8681 | -26.6689 | -26.5679 | -26.4737 | -26.4664 |
| Upper bound in (3.42) | -26.1414 | -25.7185 | -25.5013 | -25.3910 | -25.2882 | -25.2802 |
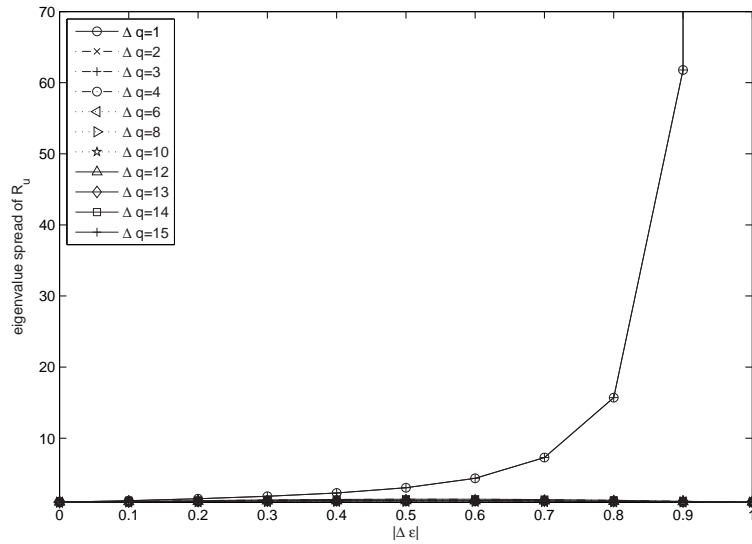


Figure 3.1: Eigenvalue spread of $\mathbf{R}_u$ for $M_{max} = 2$ ($1 \leq \Delta q \leq 15$).
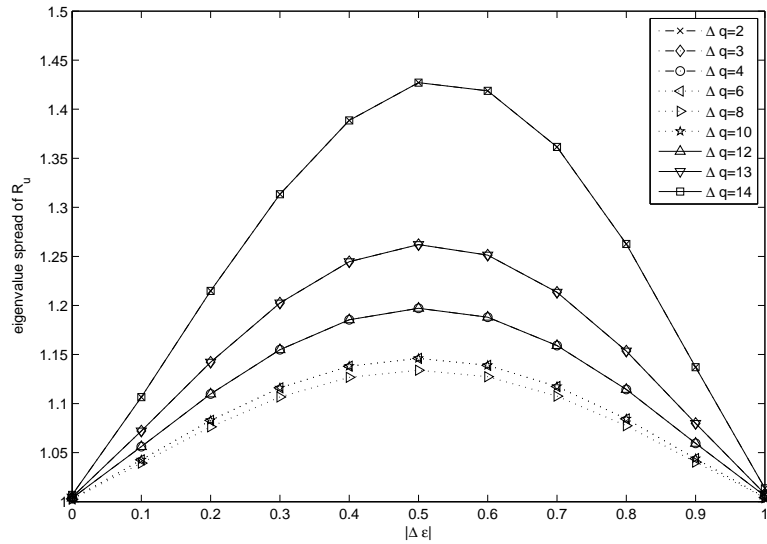
Figure 3.2: Eigenvalue spread of $\mathbf{R}_u$ for $M_{max} = 2$ ($2 \leq \Delta q \leq 14$).
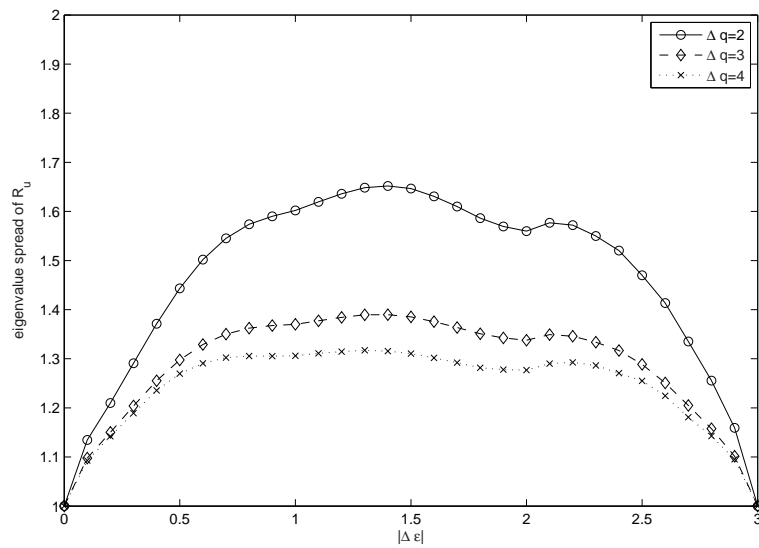


Figure 3.3: Eigenvalue spread of $\mathbf{R}_u$ for $M_{max} = 4$ ($2 \leq \Delta q \leq 4$).
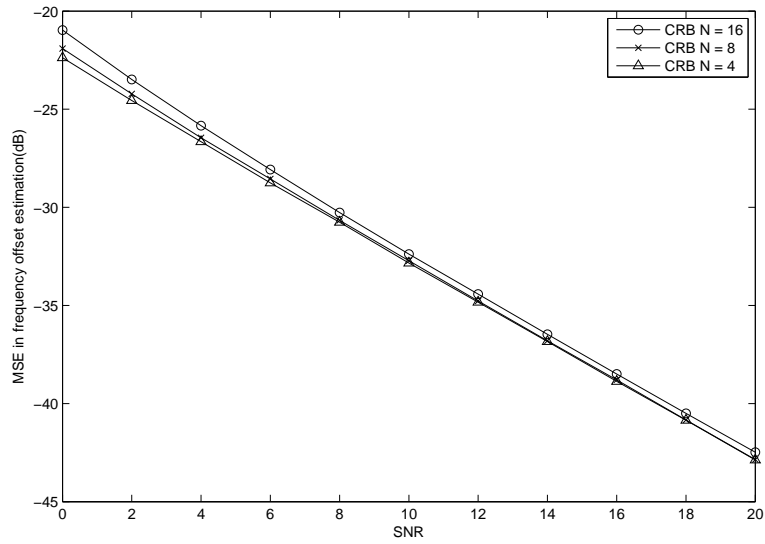
Figure 3.4: CRB comparison for various $N$.
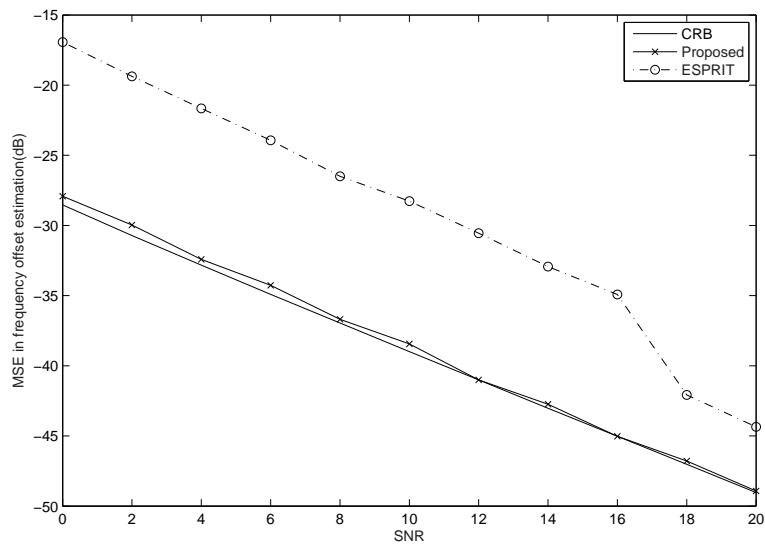


Figure 3.5: Performance comparison for ESPRIT and proposed algorithm ($M = 2$).

Figure 3.6: Performance comparison for ESPRIT and proposed algorithm ($M = 4$).



Figure 3.7: Performance comparison for training-based and proposed algorithm ($M = 2$).

Figure 3.8: Performance comparison for training-based and proposed algorithm ($M = 4$).
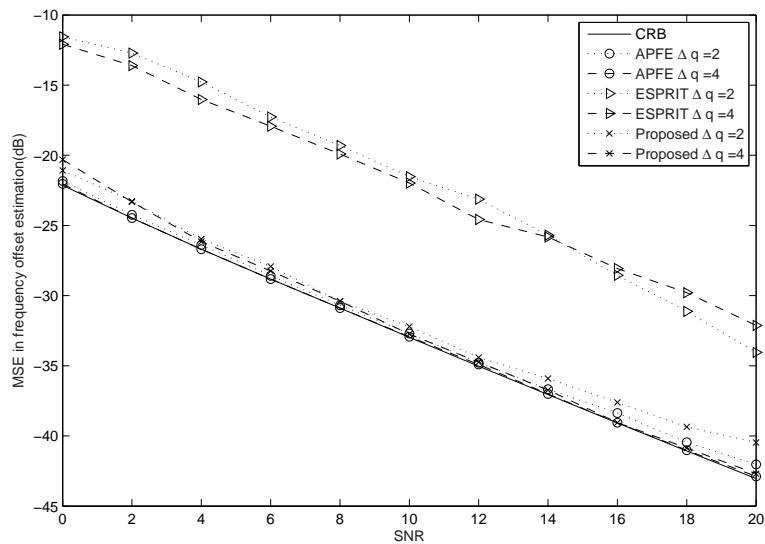


Figure 3.9: Performance comparison for all algorithms with $\Delta q = 2$ and $\Delta q = 4$ ($M = 4$).

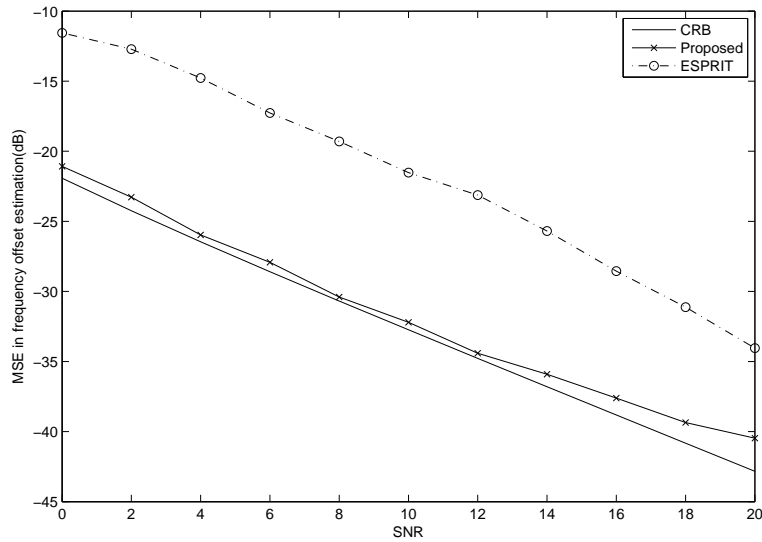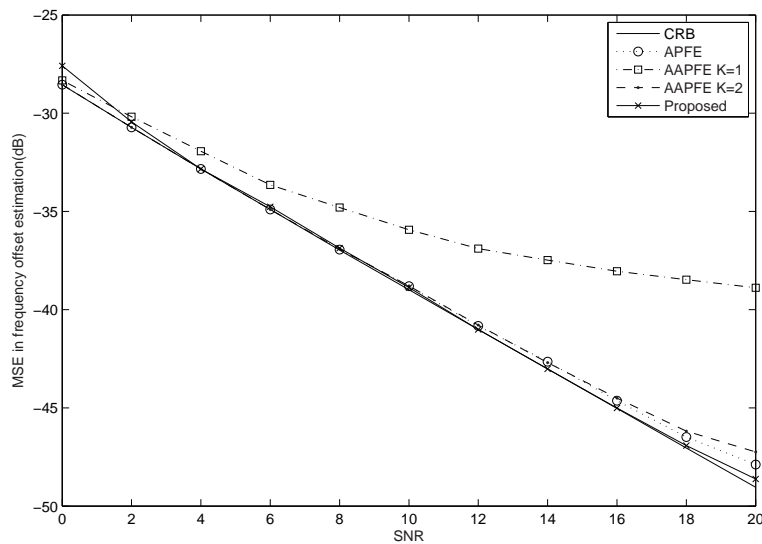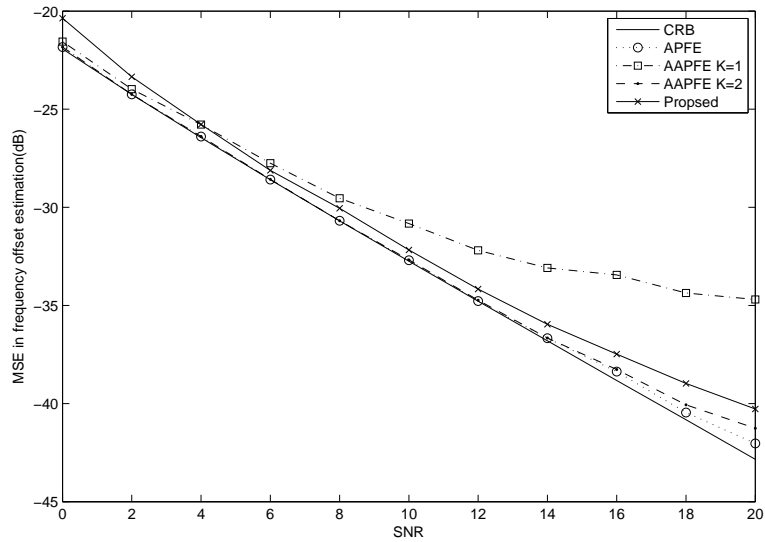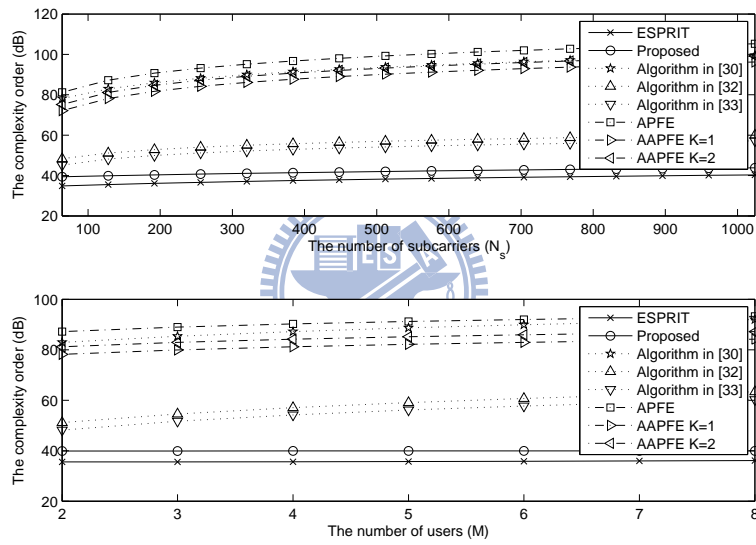Figure 3.10: Computational complexity comparison for the proposed algorithm and conventional algorithms.

# Chapter 4

# CFO Estimation and Power Allocation in Amplify-and-Forward Cooperative OFDM Systems
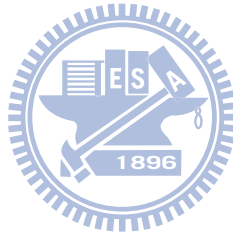
In previous chapters, we focus on the non-cooperative OFDM/OFDMA systems. Recently, there is an growing interest in wireless communication systems employing cooperative relay networks [57]. The cooperative relaying system allows wireless devices to achieve higher transmit diversity, expand coverage, yield system-wide power saving, or have better immunity against channel fading. As mentioned, two relaying protocol are well known, namely, AF and DF. In this chapter, we consider the AF-OFDM systems. One major impact of the system is that the noise at the destination is colored. To obtain the ML CFO estimation in AF-OFDM systems, the inverse of the correlation matrix has to be conducted. Unlike the correlation matrices in previous two chapters, the off-diagonal sub-matrices in the correlation matrix are not zero due to the correlation existing in the adjacent received symbols. The closed-form expression of the ML solution is much more difficult to obtain. We then propose using a gradient-descent method to solve the problem. To evaluate the performance of the proposed method, the CRB for the CFO estimation in AF-OFDM systems has to be derived. Since the expression of the

CRB is highly nonlinear and contains an expectation operation on the source-to-relay and rely-to-destination channels [9]. Again, a closed-form expression cannot be derived. To solve the problem, we then propose an approximation method to derive a closed-form solution for the CRB. The approximated CRB is a function of the expected channel state information (CSI) and it can be evaluated efficiently. Finally, we consider the power allocation problem in the AF-OFDM system. As that in [38], our objective is to minimize the CRB. Using the approximated CRB, we propose two constrained gradient-based methods achieving the optimum power allocation. The rest of the chapter is organized as follows. Section 4.1 briefly describes the system model we consider. Section 4.2 derives the ML CFO estimation and Section 4.3 derives CRB bounds for AF-OFDM systems. Section 4.4 describes the power allocation schemes minimizing the various CRBs. Section 4.5 shows the simulation results.

## § 4.1   System Model

An AF cooperative wireless communication system, shown in Figure 4.1, consists of $M + 2$ nodes, one source node (S), $M$ relay nodes ($R_1$, $R_2$,$\cdots$, $R_M$), and one destination node (D). All nodes are equipped with a single antenna. In the system, the source transmits signal to the destination with the assistance of $M$ relays. The transmission from the source to the destination takes place over two phases, i.e., broadcasting phase (BP) and relaying phase (RP). In BP, we assume that the signal-to-noise ration (SNR) between S and D is low and D cannot received the signal broadcasted from S. In RP, each relay normalizes the received signal, amplifies the received signals, and then re-transmit the resultant signal to the destination. We assume that the statistical CSIs between S and $R_m$, and those between $R_m$ and D are known at the destination (but the instantaneous CSIs between S and D are unknown). Also, the preamble sequence of an OFDM frame is periodical and is unknown to the relays and the destination.

## § 4.1.1  *Channel Model*

The channel impulse responses for S-$R_m$ and $R_m$-D links are denoted by $\mathbf{h}_{SR_m} = [h_{SR_m}(0)$ $, \cdots, h_{SR_m}(L_{SR_m} - 1)]^T$, and $\mathbf{h}_{R_m D} = [h_{R_m D}(0), \cdots, h_{R_m D}(L_{R_m D} - 1)]^T$, where $L_{SR_m}$ and $L_{R_m D}$ are the length of the channels. The channels are modeled as quasi-static frequency-selective Rayleigh fading with an exponential power decay profile. Without loss of generality, we assume that all the channels are independent each other. Also, the channel tap $h_{SR_m}(l)$, $1 \leq m \leq M$ and $0 \leq l \leq L_{SR_m} - 1$, is modeled as a complex Gaussian random variable with zero mean and a variance of

$$E\{|h_{SR_m}(l)|^2\} = \lambda_{SR_m} \cdot e^{-2\alpha_{SR_m} l} \tag{4.1}$$

where $\alpha_{SR_m}$ is the decaying factor of the exponential profile and $\lambda_{SR_m}$ a power scaling factor. The gain for $\mathbf{h}_{SR_m}$, denoted as $G_{SR_m}$, is then $G_{SR_m} = \lambda_{SR_m} \sum_{l=0}^{L_{SR_m}-1} e^{-2\alpha_{SR_m} l}$. Similar definition can be applied for $h_{R_m D}(l)$, i.e., $E\{|h_{R_m D}(l)|^2\} = \lambda_{R_m D} \cdot e^{-2\alpha_{R_m D} l}$ and the gain for $\mathbf{h}_{R_m D}$, denoted as $G_{R_m D}$, is then $G_{R_m D} = \lambda_{R_m D} \sum_{l=0}^{L_{R_m D}-1} e^{-2\alpha_{R_m D} l}$.

## § 4.1.2  *Signal Model*

Let the preamble signal in an OFDM frame be periodic with a period of $N$ and a length of $N_s = QN$. Also, let the cyclic prefix (CP) length be $L_{cp}$ and $L_{cp} \geq \max\{L_{SR_m} + L_{R_m D}\}$ (for all $m$). We assume that the preamble signal transmitted from S, denoted as $s(k)$, is unknown to $R_m$ and D. Thus, We can treat $s(k)$ as a Gaussian random variable ($k = 0, 1, ..., N_s - 1$) with variance $\sigma_s^2$. Without loss of generality, we also assume that $\sigma_s^2 = 1$. In the BP, information bits are first modulated into complex symbols, and blocks of modulated symbols are fed to an OFDM modulator with $N_s$ subcarriers and then transmitted. The received signal at $m$th relay can be expressed as

$$y_{R_m}(k) = \omega^{\varepsilon_{SR_m} k} \sqrt{g_0} \sum_{v=0}^{L_{SR_m}-1} h_{SR_m}(v) s(k - v) + \eta_m(k), \tag{4.2}$$

61

where $\eta_m(k)$ is the additive white Gaussian noise (AWGN) with variance $\sigma_{\eta_m}^2$, $\omega = \exp(j2\pi/N_s)$, $k = [0, \cdots, N_s - 1]$, $\sqrt{g_0}$ the gain factor of the transmitted amplifier at S, and $\varepsilon_{SR_m}$ the normalized CFO between S and $R_m$.

Since an automatic gain control (AGC) device is usually equipped at the receiver front end, the received signal for each relay is first normalized with a factor of $N_m = G_{SR_m}g_0 + \sigma_{\eta_m}^2$. Each relay then amplifies the received signal with a gain of $\sqrt{g_m}$ ($1 \le m \le M$), and forwards the resultant signal to the destination.

After the reception of the preamble signals transmitted from the relays, D discards the CP to restore its periodicity. The received signal at D during the RP is given by

$$y_D(k) = \sum_{m=1}^{M} \omega^{\varepsilon_{R_m D} k} \sqrt{\frac{g_m}{N_m}} \sum_{u=0}^{L_{R_m D}-1} h_{R_m D}(u) y_{R_m}(k-u) + \eta(k), \qquad (4.3)$$

where $\eta(k)$ is the AWGN with a variance of $\sigma_\eta^2$ and $\varepsilon_{R_m D}$ the normalized CFO between $R_m$ and D. Substituting (4.2) into (4.3), we can have

$$
\begin{aligned}
y_D(k) &= \sum_{m=1}^{M} \omega^{\varepsilon_{R_m D} k} \sqrt{g_m/N_m} \sum_{u=0}^{L_{R_m D}-1} h_{R_m D}(u)\{\omega^{\varepsilon_{SR_m}(k-u)} \cdot \\
&\qquad \sqrt{g_0} \sum_{v=0}^{L_{SR_m}-1} h_{SR_m}(v) s(k-u-v) + \eta_m(k-u)\} + \eta(k) \\
&= \sum_{m=1}^{M} \omega^{(\varepsilon_{R_m D}+\varepsilon_{SR_m})k} \sqrt{g_m/N_m} \sum_{u=0}^{L_{R_m D}-1} \{h_{R_m D}(u) \\
&\qquad \omega^{\varepsilon_{SR_m}(-u)} \sqrt{g_0} \sum_{v=0}^{L_{SR_m}-1} [h_{SR_m}(v) s(k-u-v)]\} \\
&\qquad + \omega^{\varepsilon_{R_m D} k} \sqrt{g_m/N_m} \sum_{u=0}^{L_{R_m D}-1} [h_{R_m D}(u)\eta_m(k-u)] + \eta(k) \\
&= \sum_{m=1}^{M} \omega^{(\varepsilon_{SR_m}+\varepsilon_{R_m D})k} \sqrt{g_m/N_m} \hat{h}_{R_m D}(k) * \hat{s}_m(k) + \xi(k) + \eta(k), \qquad (4.4)
\end{aligned}
$$

62

where

$$\hat{s}_m(k) = \sqrt{g_0}h_{SR_m}(k) * s(k),$$

$$\hat{h}_{R_m D}(k) = h_{R_m D}(k)\omega^{-\varepsilon_{SR_m}k},$$

$$\xi(k) = \sum_{m=1}^{M}\omega^{\varepsilon_{R_m D}k}\sqrt{g_m/N_m}h_{R_m D}(k) * \eta_m(k). \quad (4.5)$$

Apparently, the noise at the destination, $\xi(k)$ plus $\eta(k)$, is colored. The power allocation method proposed in [38] ignores this effect and the derived result is then not optimal. It is interesting to note that the CFO between $S$ and $D$ is the same as the summation of the CFO between $S$ and $R$ and that between $R$ and $D$, i.e., $\varepsilon_{SD} = \varepsilon_{SR_m} + \varepsilon_{R_m D}$ [38]. Then, the received signal at D can be rewritten as

$$y_D(k) = \omega^{\varepsilon_{SD}k}x(k) + \xi(k) + \eta(k), \quad (4.6)$$

where

$$x(k) = \sum_{m=1}^{M}\sqrt{g_m/N_m}\hat{h}_{R_m D}(k) * \hat{s}_m(k).$$

Rewriting the above equation with a vector form, we can obtain the following result:

$$\mathbf{y}_D = \mathbf{\Phi}\mathbf{x} + \boldsymbol{\xi} + \boldsymbol{\eta}, \quad (4.7)$$

where

$$\mathbf{y}_D = [\ y_D(0)\ \ y_D(1)\ \ \cdots\ \ y_D(N_s - 1)\ ]^T,$$

$$\mathbf{x} = [\ x(0)\ \ x(1)\ \ \cdots\ \ x(N_s - 1)\ ]^T,$$

$$\boldsymbol{\xi} = [\ \xi(0)\ \ \xi(1)\ \ \cdots\ \ \xi(N_s - 1)\ ]^T,$$

$$\boldsymbol{\eta} = [\ \eta(0)\ \ \eta(1)\ \ \cdots\ \ \eta(N_s - 1)\ ]^T,$$

$$\mathbf{\Phi} = \begin{bmatrix} \omega^{\varepsilon_{SD}\cdot 0} & 0 & \cdots & 0 \\ 0 & \omega^{\varepsilon_{SD}\cdot 1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega^{\varepsilon_{SD}\cdot(N_s-1)} \end{bmatrix}.$$

63

From above equations, we can find that $x(k)$ is the output of an effective channel, $h_{eff,m}(k) = \hat{h}_{R_mD}(k) * h_{SR_m}(k)$. As mentioned, the received signal is colored and the conventional approaches for estimating CFO is not valid.

## § 4.2  ML CFO Estimation

Since $h_{SR_m}(k)$ is unknown at D, the effective channel, $h_{eff,m}(k)$ is a random variable at D. As defined, the preamble signal in frequency domain is unknown at the relays and the destination. For OFDM-based systems, the time-domain noiseless signal is obtained from the inverse discrete-Fourier-transform (IDFT) of its frequency-domain signal. From the central limit theorem, we know that if the number of subcarriers is reasonably large, the time-domain transmitted signal and the output of the unknown effective channels can then be approximated as a white Gaussian sequence (see Section 2.2). So, the received signal, $y_D(k)$ is composed with a white Gaussian signal $x(k)$, a colored noise $\xi(k)$ and AWGN $\eta(k)$.

Note that $x(k)$, $\xi(k)$, and $\eta(k)$ are independent each other since the transmitted signal at S and the noise at $R_m$ and D are independent each other. Let $h_{R_mD}(l)$'s be given and fixed. Also, let the received signal power at D be denoted as $\sigma_y^2$ and $\sigma_y^2 = \sigma_x^2 + \sigma_n^2$ where

$$
\begin{aligned}
\sigma_x^2 &= \sum_{m=1}^{M} g_0 G_{SR_m} \frac{g_m}{N_m} \sum_{l=0}^{L_{R_mD}-1} h_{R_mD}(l) h_{R_mD}^*(l) \\
&= \sum_{m=1}^{M} g_0 G_{SR_m} \frac{g_m}{N_m} \gamma_m(0) \tag{4.8} \\
\sigma_n^2 &= \sigma_\xi^2 + \sigma_\eta^2, \tag{4.9} \\
\sigma_\xi^2 &= \sum_{m=1}^{M} \sigma_{\eta_m}^2 \frac{g_m}{N_m} \gamma_m(0) \tag{4.10} \\
\gamma_m(k) &= \omega^{-\varepsilon_{R_mD}k} \sum_{l=0}^{L_{R_mD}-k-1} h_{R_mD}(l) h_{R_mD}^*(l+k), \\
&= \omega^{-\varepsilon_{SD}k} \omega^{\varepsilon_{SR_m}k} \sum_{l=0}^{L_{R_mD}-k-1} h_{R_mD}(l) h_{R_mD}^*(l+k), \tag{4.11}
\end{aligned}
$$

64

and $k = 0, \cdots, L_{R_mD} - 1$. Note that $\sigma_y^2$, $\sigma_x^2$, and $\sigma_n^2$ are functions of $h_{R_mD}(l)$'s.

The log-likelihood function (LLF) for the received signal at D can then be expressed as

$$
\begin{aligned}
\Lambda(\varepsilon_{SD}|h_{RD}) &= \ln[f(\mathbf{y}_D|h_{RD})] \\
&= -\ln[\pi] - \ln[\det(\tilde{\mathbf{R}}_{y|h})] - \mathbf{y}_D^H \tilde{\mathbf{R}}_{y|h}^{-1} \mathbf{y}_D,
\end{aligned}
\tag{4.12}
$$

where $(.|h_{RD})$ denotes the event conditioning on $h_{R_mD}(l)$'s, $f(.)$ is a probability density function, and $\tilde{\mathbf{R}}_{y|h} = E\{\mathbf{y}_D \mathbf{y}_D^H | h_{RD}\}$ is a $N_s \times N_s$ correlation matrix. The correlation matrix in (4.12) can be decomposed as $Q^2$ $N \times N$ submatrices as:

$$
\tilde{\mathbf{R}}_{y|h} = \begin{bmatrix}
\tilde{\mathbf{R}}_1 & \tilde{\mathbf{R}}_2 & \tilde{\mathbf{R}}_3 & \cdots & \tilde{\mathbf{R}}_Q \\
\tilde{\mathbf{R}}_2^H & \tilde{\mathbf{R}}_1 & \tilde{\mathbf{R}}_2 & \cdots & \tilde{\mathbf{R}}_{Q-1} \\
\tilde{\mathbf{R}}_3^H & \tilde{\mathbf{R}}_2^H & \tilde{\mathbf{R}}_1 & \cdots & \tilde{\mathbf{R}}_{Q-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\tilde{\mathbf{R}}_Q^H & \tilde{\mathbf{R}}_{Q-1}^H & \tilde{\mathbf{R}}_{Q-2}^H & \cdots & \tilde{\mathbf{R}}_1
\end{bmatrix},
\tag{4.13}
$$

where

$$
(\tilde{\mathbf{R}}_1)_{p,q} = \begin{cases}
\sigma_x^2 + \sigma_\xi^2 + \sigma_\eta^2 & \text{, if } p = q \\
\gamma(k) & \text{, if } q - p = k, 0 < k < \max(L_{R_mD}) \\
\gamma(k)^* & \text{, if } p - q = k, 0 < k < \max(L_{R_mD})
\end{cases}
$$

$$
(\tilde{\mathbf{R}}_2)_{p,q} = \begin{cases}
\sigma_x^2 \omega^{-\varepsilon_{SD}N} & \text{, if } p = q \\
\gamma(N + q - p) & \text{, if } N - (p - q) \leq L_{R_mD} - 1
\end{cases}
$$

$$
(\tilde{\mathbf{R}}_k)_{p,q}|_{k \geq 3} = \begin{cases}
\sigma_x^2 \omega^{-\varepsilon_{SD}N(k-1)} & \text{, if } p = q \\
0 & \text{, otherwise}
\end{cases}
$$

and

$$
\gamma(k) = \sum_{m=1}^{M} \sigma_{\eta_m}^2 \frac{g_m}{N_m} \gamma_m(k).
\tag{4.14}
$$

As we can see, the LLF in (4.12) is very complicated and a closed-form solution for the CFO estimation is difficult to derive. To solve the problem, we propose using the gradient-descent

method in the ML estimation problem. Note that $\gamma_m(k)$ is the autocorrelation function of the received signal when the relays transmit white sequences, and it is required in the estimation. There are a couple of scenarios that $\gamma_m(k)$ can be estimated. For example, the relays can transmit white sequences to D in the BP and the destination can use the received signal to estimate $\gamma_m(k)$. It is important to see that the receiver does not have to know the transmit sequences since only the autocorrelation function is needed not the channel responses. For other scenarios, D may have priori information about $\gamma_m(k)$. For example, if the relays are base stations [58] communicating with D, $h_{R_m D}$ and $\varepsilon_{R_m D}$ can be known at D, so is $\gamma_m(k)$. When a new user joins the network (treated as S), $\gamma_m(k)$ is then known as a priori for the user. For these reasons, we then assume that $\gamma_m(k)$ is available as a priori.

Taking a derivative with respect to $\varepsilon_{SD}$, we have

$$\frac{\partial \Lambda(\varepsilon_{SD}|h_{RD})}{\varepsilon_{SD}} = \mathrm{tr}(\tilde{\mathbf{R}}_{y|h}^{-1} \frac{\partial \tilde{\mathbf{R}}_{y|h}}{\partial \varepsilon_{SD}}) - \mathbf{y}_D^H \tilde{\mathbf{R}}_{y|h}^{-1} \frac{\partial \tilde{\mathbf{R}}_{y|h}}{\partial \varepsilon_{SD}} \tilde{\mathbf{R}}_{y|h}^{-1} \mathbf{y}_D. \tag{4.15}$$

For the gradient-descent method, the update equation is given by

$$\hat{\varepsilon}_{SD}(k+1) = \hat{\varepsilon}_{SD}(k) + \mu \frac{\partial \Lambda(\varepsilon_{SD}|h_{RD})}{\partial \varepsilon_{SD}}, \tag{4.16}$$

where $\mu$ is a step size. Here, we only focus on the estimation of the fractional CFO, i.e., $|\varepsilon_{SD}(k)| < 0.5$. The operations of the gradient-descent method are summarized as follows:

1. Initialization : Set $k = 0$ and initialize the $\hat{\varepsilon}_{SD}(k)$ as zero.

2. Use (4.16) to obtain $\varepsilon_{SD}(k+1)$.

3. If $|\varepsilon_{SD}(k+1)| > 0.5$, let $|\varepsilon_{SD}(k+1)| = 0.5$.

4. Check if $\mu(\partial \Lambda(\varepsilon_{SD}|h_{RD})/\partial \varepsilon_{SD})$ is small than a preset $\epsilon$. If yes, the algorithm stops and output $\hat{\varepsilon}_{SD}^{(r)}$. Otherwise, set $k$ as $k+1$ and go to Step 2.

To evaluate the performance of the estimation, we define a mean square error (MSE), denoted by $\mathrm{MSE}_r$, as

$$\mathrm{MSE}_r = \frac{1}{N_e} \sum_{i=1}^{N_t} (\hat{\varepsilon}_{SD,i}^{(r)} - \varepsilon_{SD})^2, \tag{4.17}$$

where $\hat{\varepsilon}_{SD,i}^{(r)}$ indicates the $i$th estimation of CFO and $N_e$ the total number of estimations.

Note that the ML CFO estimation in AWGN environment has been solved in Section 2.2, and the result is given by

$$\hat{\varepsilon}_{SD}^{(w)} = -\frac{\sum\limits_{p=1}^{Q-1} \sum\limits_{q>p}^{Q} |\Upsilon_{pq}|(q-p)\angle\Upsilon_{pq}}{2\pi \sum\limits_{p=1}^{Q-1} \sum\limits_{q>p}^{Q} |q-p|^2|\Upsilon_{pq}|}, \qquad (4.18)$$

where

$$\Upsilon_{pq} = \sum_{n=0}^{N-1} y_D((p-1)N+n)y_D^*((q-1)N+n)$$

and $\angle\Upsilon_{pq}$ is the phase of $\Upsilon_{pq}$. One can ignore the color property of the receive noise and use (4.18) to conduct the CFO estimation. To evaluate the performance, we define the MSE for this approach, denoted by $\text{MSE}_w$, as

$$\text{MSE}_w = \frac{1}{N_e} \sum_{i=1}^{N_t} (\hat{\varepsilon}_{SD,i}^{(w)} - \varepsilon_{SD})^2, \qquad (4.19)$$

where $\hat{\varepsilon}_{SD,i}^{(w)}$ means the $i$th CFO estimation. In simulations, we will use $\text{MSE}_r$ and $\text{MSE}_w$ for performance comparison.

## § 4.3   CRB Analysis

In this section, we will derive the CRB for the ML estimation of the CFO. Since the channel is fading, we then take an expectation on the LLF in (4.12) and obtain the expected LLF as

$$\Lambda(\varepsilon_{SD}) = E_h\{\Lambda(\varepsilon_{SD}|h_{RD})\} \qquad (4.20)$$

where the expectation is conducted on $h_{RD}$. Using the LLF, we can have the CRB as [51]

$$\text{CRB}_r \geq \frac{1}{F}, \qquad (4.21)$$

where

$$F = -E\{\frac{\partial^2}{\partial \varepsilon_{SD}^2}\Lambda(\varepsilon_{SD})\}$$
$$= E_h\{\text{tr}(\tilde{\mathbf{R}}_{y|h}^{-1}\frac{\partial\tilde{\mathbf{R}}_{y|h}}{\partial\varepsilon_{SD}}\tilde{\mathbf{R}}_{y|h}^{-1}\frac{\partial\tilde{\mathbf{R}}_{y|h}}{\partial\varepsilon_{SD}})\}. \tag{4.22}$$

From (4.22), we can see that a simple closed-form expression for the LLF is difficult to obtain since matrix multiplications/inversions and the expectation operations are involved. To solve the problem, we propose using some approximations. Define a Rayleigh exponential decay channel as $h_{R_mD}(l) = \check{h}_{R_mD}(l)\exp(-\alpha_{R_mD}l)$, where $\check{h}_{R_mD}(l)$'s are independent and identically distributed random variables with zero mean and a variance of $\lambda_{R_mD}$. Then we can rewrite (4.14) as

$$\gamma(k) = \sum_{m=1}^{M}\omega^{-\varepsilon_{R_mD}k}\sigma_{\eta_m}^2 g_m/N_m \sum_{l=0}^{L_{R_mD}-k-1}\check{h}_{R_mD}(l)\check{h}_{R_mD}^*(l+k)e^{-\alpha_{R_mD}(2l+k)}$$
$$= \sum_{m=1}^{M}\omega^{-\varepsilon_{R_mD}k}\sigma_{\eta_m}^2 g_m/N_m e^{-\alpha_{R_mD}k} \cdot$$
$$\sum_{l=0}^{L_{R_mD}-k-1}\check{h}_{R_mD}(l)\check{h}_{R_mD}^*(l+k)e^{-2\alpha_{R_mD}l}. \tag{4.23}$$

From (4.23), we see that $\gamma(k)$ is a summation of independent random variables, $\gamma_m(k)$ for different $m$'s. To facilitate the derivation, we assume that $\gamma(k)$ is a complex Gaussian random variable. From the central limit theorem, the approximation error will be smaller when $M$ is larger. It is simple to see that the expectation of $\gamma(k)$ is zero when $k \geq 1$ and that of $\gamma(0)$ is $\sum_{m=1}^{M}\sigma_{\eta_m}^2 g_m/N_m G_{R_mD}$. The variance of $\gamma(k)$ for $k \geq 1$ can be derived as

$$\bar{\sigma}_\gamma^2(k) = E_h\{\gamma(k)\gamma(k)^*\} = \sum_{m=1}^{M}\sigma_{\eta_m}^4(g_m/N_m)^2\lambda_{R_mD}^2\frac{e^{-2\alpha_{R_mD}k}(1 - e^{-4\alpha_{R_mD}(L_{R_mD}-k)})}{1 - e^{-4\alpha_{R_mD}}}. \tag{4.24}$$

68

The variance of $\gamma(0)$ (as defined in (4.10) and (4.14), $\gamma(0) = \sigma_\xi^2$) is

$$
\begin{aligned}
E_h\{\gamma(0)\gamma(0)^*\} &= E_h\{|\gamma(0)|^2\} = E_h\{\sum_{a=1}^{M}\sum_{b=1}^{M}\sigma_{\eta_a}^2\frac{g_a}{N_a}\gamma_a(0)\sigma_{\eta_b}^2\frac{g_b}{N_b}\gamma_b(0)\} \\
&= \sum_{a=1}^{M}\sum_{b=1}^{M}\sigma_{\eta_a}^2\frac{g_a}{N_a}\sigma_{\eta_b}^2\frac{g_b}{N_b} \\
&\quad \sum_{u=0}^{L_{R_aD}-1}\sum_{v=0}^{L_{R_aD}-1} E_h\{\check{h}_{R_aD}(u)\check{h}_{R_aD}^*(u)\check{h}_{R_bD}(v)\check{h}_{R_bD}^*(v)\}e^{-2\alpha_{R_aD}u}e^{-2\alpha_{R_bD}v} \\
&\geq \sum_{a=1}^{M}\sum_{b=1}^{M}\sigma_{\eta_a}^2\frac{g_a}{N_a}\sigma_{\eta_b}^2\frac{g_b}{N_b}\sum_{u=0}^{L_{R_aD}-1} E_h\{\check{h}_{R_aD}(u)\check{h}_{R_aD}^*(u)\} \\
&\quad \sum_{v=0}^{L_{R_aD}-1} E_h\{\check{h}_{R_bD}(v)\check{h}_{R_bD}^*(v)\}e^{-2(\alpha_{R_aD}u+\alpha_{R_bD}v)} \\
&= \sum_{a=1}^{M}\sum_{b=1}^{M}\sigma_{\eta_a}^2\frac{g_a}{N_a}\sigma_{\eta_b}^2\frac{g_b}{N_b}\lambda_{R_aD}\lambda_{R_bD}\sum_{u=0}^{L_{R_aD}-1}\sum_{v=0}^{L_{R_aD}-1}e^{-2(\alpha_{R_aD}u+\alpha_{R_bD}v)} \\
&= \sum_{a=1}^{M}\sum_{b=1}^{M}\sigma_{\eta_a}^2\frac{g_a}{N_a}\sigma_{\eta_b}^2\frac{g_b}{N_b}G_{R_aD}G_{R_bD} = \bar{\sigma}_\xi^4 \\
&> \sum_{a=b=1}^{M}(\sigma_{\eta_a}^2\frac{g_a}{N_a})^2\lambda_{R_aD}^2\sum_{u=v=0}^{L_{R_aD}-1}e^{-4\alpha_{R_aD}u} = \bar{\sigma}_\gamma^2(0), \qquad (4.25)
\end{aligned}
$$

where

$$
\bar{\sigma}_\xi^2 = E_h\{\sigma_\xi^2\} = \sum_{m=1}^{M}\frac{g_m}{N_m}\sigma_{\eta_m}^2 G_{R_mD}. \qquad (4.26)
$$

So, $\bar{\sigma}_\xi^2 > \sqrt{\bar{\sigma}_\gamma^2(0)} > \sqrt{\bar{\sigma}_\gamma^2(1)} > \cdots > \sqrt{\bar{\sigma}_\gamma^2(L_{R_mD}-1)}$ for all $m$ since $\bar{\sigma}_\gamma^2(0) > \bar{\sigma}_\gamma^2(1) > \cdots > \bar{\sigma}_\gamma^2(L_{R_mD}-1)$ for all $m$ (which can be observed in (4.24)). Note that $\gamma(k)$'s ($k > 0$) appear in the off-diagonal terms of the sub-matrices in (4.13) and $\sigma_\xi^2$ the diagonal terms of $\tilde{\mathbf{R}}_1$ in (4.13). Comparing with $\sigma_\xi^2$, the variance of the off-diagonal terms of the sub-matrices is small, especially for large $k$. Thus, some off-diagonal terms of the sub-matrices in (4.13) can be ignored when $k$ is large.

We now use the HIPERLAN/2 channel model [47] [60] to evaluate the magnitude of $\bar{\sigma}_\gamma^2(k)$. Assume that $\alpha_{R_mD} = \alpha = 0.5$ and $L_{R_mD} = 8$ for all $m$. Define the terms depending on $k$ in

(4.24) as $\kappa_m(k) = e^{-2\alpha k}(1 - e^{-4\alpha(L_{RmD}-k)})$ and divide $\kappa_m(k)$ with $\kappa_m(0)$ (the minimum of (4.25)). We then have

$$f_m(\alpha, k) = \frac{e^{-2\alpha k}(1 - e^{-4\alpha(L_{RmD}-k)})}{(1 - e^{-4\alpha(L_{RmD})})}.$$

It can be treated as the ratio between $\gamma(k)$ and $\bar{\sigma}_\xi^2$ in (4.13) since $\bar{\sigma}_\gamma^2(0) < \bar{\sigma}_\xi^4$. Figure 4.2 shows the simulation results for $f_m(\alpha, k)$. From the figure, we see that $f_m(\alpha, k)$ descends almost 90% when $k \geq 2$. In other words, the random variable $\gamma(k)$ approaches its zero mean when $k \geq 2$. To simplify the problem, we then assume that $\gamma(k)$ is zero for $k \geq 2$. With the assumption, $\tilde{\mathbf{R}}_1$ becomes a tri-diagonal matrix and the off-diagonal entries of $\tilde{\mathbf{R}}_2$ are zero except for $(\tilde{\mathbf{R}}_2)_{N,1} = \gamma(1)$. To obtain a closed-form Fisher information, we then assume that $\tilde{\mathbf{R}}_2$ is a diagonal matrix and $\gamma(1)$ only appears in $\tilde{\mathbf{R}}_1$. Note that the assumption is more valid when $\alpha$ is larger, an environment for large open space [64]. As a result, all the sub-matrices in (4.13) are diagonal except for those in the main diagonal. However, the inverse of $\tilde{\mathbf{R}}_{y|h}$ and the expectation operation in (4.22) are still difficult to derive since all the submatrices in (4.22) are non-zero. To solve the problem, we re-arrange the elements of the received signal vector $\mathbf{y}_D$ as

$$\mathbf{y}_D = [y_D(0), y_D(N), \cdots, y_D((Q-1)N), \cdots, y_D(N-1), \cdots, y_D(QN-1)]^T. \quad (4.27)$$

Then, its correlation matrix can be re-expressed as a matrix composed of $N^2$ $Q \times Q$ sub-matrices:

$$\begin{aligned}
\mathbf{R}_{y|h} &= E\{\mathbf{y}_D\mathbf{y}_D^H|h_{RD}\} \\
&= \begin{bmatrix}
\mathbf{R}_1 & \mathbf{R}_2 & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{R}_2^H & \mathbf{R}_1 & \mathbf{R}_2 & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{R}_2^H & \mathbf{R}_1 & \cdots & \mathbf{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_1
\end{bmatrix},
\end{aligned} \quad (4.28)$$

where $\mathbf{0}$ indicates a $Q \times Q$ matrix composed of all zeros elements,

$$\mathbf{R}_1 = \sigma_n^2\mathbf{I}_Q + \sigma_x^2\mathbf{u}\mathbf{u}^H, \quad (4.29)$$

$$\mathbf{R}_2 = \gamma(1)\mathbf{I}_Q,$$

and

$$\mathbf{u} = [1, \omega^{\varepsilon_{SD}N}, \cdots, \omega^{\varepsilon_{SD}N(Q-1)}]^T. \tag{4.30}$$

The correlation matrix of (4.28) has a better structure than that in (4.13) since most submatrices are zero and $\mathbf{R}_2$ is a diagonal matrix. As a result, $\mathbf{R}_{y|h}$ is a tri-diagonal matrix. Note that the inverse of a tri-diagonal matrix has been derived in [61] [62]. Using the results in [61] [62], we can obtain the inverse of a tri-diagonal matrix as follows. Define $\mathbf{T} = \mathbf{R}_{y|h}^{-1}$. Then, the $(p, q)$th submatrix ($Q \times Q$) of $\mathbf{T}_{p,q}$ can then be expressed as

$$\mathbf{T}_{p,q} = \begin{cases} (-1)^{p+q}(\gamma(1))^{q-p}\mathbf{P}_{p-1}\mathbf{Q}_{q+1}(\mathbf{P}_N)^{-1} & , \text{if } 1 \le p \le q \le N \\ (-1)^{p+q}(\gamma(1)^*)^{p-q}\mathbf{P}_{q-1}\mathbf{Q}_{p+1}(\mathbf{P}_N)^{-1} & , \text{if } 1 \le q < p \le N \end{cases} \tag{4.31}$$

where

$$\mathbf{P}_i = \mathbf{R}_1\mathbf{P}_{i-1} - |\gamma(1)|^2\mathbf{P}_{i-2}, \text{ for } i = 2, \cdots, N \tag{4.32}$$

where $C_i(a)$ denotes the coefficients of the polynomial, $\lfloor b \rfloor$ the nearest integer less than or equal to $b$, and $\lfloor b \rfloor \ge 0$. Here, $\mathbf{P}_0 = \mathbf{I}$ and $\mathbf{P}_1 = \mathbf{R}_1$ and

$$\mathbf{Q}_i = \mathbf{P}_{N+1-i}, \text{ for } i = 2, \cdots, N+1. \tag{4.33}$$

Taking a derivative with respect to $\varepsilon_{SD}$, we can have (4.28)

$$\begin{aligned} \mathbf{R}'_{y|h} &= \frac{\partial \mathbf{R}_{y|h}}{\partial \varepsilon_{SD}} \\ &= \begin{bmatrix} \mathbf{R}'_1 & \mathbf{R}'_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{R}'^H_2 & \mathbf{R}'_1 & \mathbf{R}'_2 & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}'^H_2 & \mathbf{R}'_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}'_1 \end{bmatrix}, \end{aligned} \tag{4.34}$$

where

$$(\mathbf{R}'_1)_{p,q} = \begin{cases} 0 & , \text{if } p = q \\ j\frac{2\pi(p-q)}{Q}\sigma_x^2\omega^{\varepsilon_{SD}(p-q)N} & , \text{otherwise} \end{cases}, \tag{4.35}$$

$$\mathbf{R}'_2 = \gamma'(1)\mathbf{I}_Q,$$

71

and

$$\gamma'(1) = \frac{\partial \gamma(1)}{\partial \varepsilon_{SD}} = -j\frac{2\pi}{N_s}\sum_{m=1}^{M}\sigma_{\eta_m}^2\gamma_m(1). \tag{4.36}$$

Since the subcarrier length $N_s$ is usually much larger than $2\pi$, we can further assume that $\gamma'(1) = 0$. Thus, (4.34) is reduced to be diagonal. From (4.28) and (4.34), we can rewrite the Fisher information in (4.22) as

$$\begin{aligned} F &= E_h\{\mathrm{tr}(\mathbf{R}_{y|h}^{-1}\frac{\partial \mathbf{R}_{y|h}}{\partial \varepsilon_{SD}}\mathbf{R}_{y|h}^{-1}\frac{\partial \mathbf{R}_{y|h}}{\partial \varepsilon_{SD}})\} \\ &= E_h\{\mathrm{tr}(2\sum_{p=1}^{N-1}\sum_{q>p}^{N}\mathbf{W}_{p,q} + \sum_{p=1}^{N}\mathbf{W}_{p,p})\}, \end{aligned} \tag{4.37}$$

where

$$\mathbf{W}_{p,q} = (|\gamma(1)|^2)^{q-p}(\mathbf{P}_{p-1}\mathbf{Q}_{q+1}\mathbf{P}_N^{-1}\mathbf{R}_1')^2. \tag{4.38}$$

The detailed derivations of (4.37) is given in Appendix C.1. Substituting (4.33) into (4.38), we can have

$$E_h\{\mathrm{tr}(\mathbf{W}_{p,q})\} = E_h\{(|\gamma(1)|^2)^{q-p}\mathrm{tr}(E_h(\{\mathbf{P}_{p-1}\mathbf{P}_{N-q}\mathbf{P}_N^{-1}\mathbf{R}_1'\})^2)\} \tag{4.39}$$

As we can see the expectation operation on the right hand side of (4.39) is difficult to conduct. To continue the derivation, we then use the following approximation:

$$E_h\{\mathrm{tr}(\mathbf{W}_{p,q})\} \approx \mathrm{tr}((E_h\{|\gamma(1)|^2\})^{q-p}(E_h\{\mathbf{P}_{p-1}\}E_h\{\mathbf{P}_{N-q}\}E_h\{\mathbf{P}_N^{-1}\}E_h\{\mathbf{R}_1'\})^2). \tag{4.40}$$

Let $\bar{\mathbf{W}}_{p,q}$ be the matrix inside the trace operator in the right hand side of (4.40), i.e., $\bar{\mathbf{W}}_{p,q} = E_h(\mathbf{W}_{p,q}) = (E_h\{|\gamma(1)|^2\})^{q-p}(E_h\{\mathbf{P}_{p-1}\}E_h\{\mathbf{P}_{N-q}\}E_h\{\mathbf{P}_N^{-1}\}E_h\{\mathbf{R}_1'\})^2$. We can then have

$$\bar{\mathbf{W}}_{p,q} = (\bar{\sigma}_\gamma^2(1))^{q-p}(\bar{\mathbf{P}}_{p-1}\bar{\mathbf{Q}}_{q+1}\bar{\mathbf{P}}_N^{-1}\bar{\mathbf{R}}_1')^2, \tag{4.41}$$

where

$$\bar{\mathbf{R}}_1 = E_h\{\mathbf{R}_1\} = \bar{\sigma}_x^2\mathbf{u}\mathbf{u}^H + \bar{\sigma}_n^2\mathbf{I}_Q \tag{4.42}$$

$$(\bar{\mathbf{R}}_1')_{p,q} = E_h\{(\mathbf{R}_1')_{p,q}\} = \begin{cases} 0 & \text{, if } p = q \\ j\frac{2\pi(p-q)}{Q}\bar{\sigma}_x^2\omega^{\varepsilon_{SD}(p-q)N} & \text{, otherwise} \end{cases}$$

$$\bar{\mathbf{P}}_i = E_h\{\mathbf{P}_i\} = \bar{\mathbf{R}}_1\bar{\mathbf{P}}_{i-1} - \bar{\sigma}_\gamma^2(1)\bar{\mathbf{P}}_{i-2}, \text{ for } i = 2, \cdots, N \tag{4.43}$$

72

and

$$\bar{\sigma}_x^2 \quad = \quad E_h\{\sigma_x^2\} = \sum_{m=1}^{M} g_m/N_m g_0 G_{R_m D} G_{SR_m} \tag{4.44}$$

$$\bar{\sigma}_n^2 \quad = \quad \bar{\sigma}_\xi^2 + \sigma_\eta^2 \tag{4.45}$$

with $\bar{\mathbf{P}}_0 = \mathbf{I}$, $\bar{\mathbf{P}}_1 = \bar{\mathbf{R}}_1$, and $\bar{\mathbf{Q}}_i = \bar{\mathbf{P}}_{N+1-i}$ for $i = 2, \cdots, N+1$. Substituting (4.41) into (4.37), we can obtain an approximated Fisher information, denoted by $F_a$, as

$$F_a = \mathrm{tr}(2 \sum_{p=1}^{N-1} \sum_{q>p}^{N} \bar{\mathbf{W}}_{p,q} + \sum_{p=1}^{N} \bar{\mathbf{W}}_{p,p}) \tag{4.46}$$

and a closed-form solution for the CRB, denoted by $\mathrm{CRB}_a$, as

$$\mathrm{CRB}_a = \frac{1}{F_a}. \tag{4.47}$$

If the noise forwarded from relays is treated as white, we can let $\gamma(k)$ be equal to zero in $\mathrm{CRB}_a$ and $\mathrm{CRB}_r$. In this case, $\mathrm{CRB}_a$ and $\mathrm{CRB}_r$ become identical and the CRB is reduced to

$$
\begin{aligned}
\mathrm{CRB}_w \quad &= \quad \frac{1}{F} \Big|_{\gamma(k)=0} = \frac{1}{F_a} \Big|_{\gamma(k)=0} \\
&= \quad \frac{Q^2 \bar{\sigma}_n^2 (\bar{\sigma}_n^2 + Q \cdot \bar{\sigma}_x^2)}{8\pi^2 N \cdot \bar{\sigma}_x^4 \sum_{p=1}^{Q-1} \sum_{q>p}^{Q} (q-p)^2}
\end{aligned} \tag{4.48}
$$

which is the same as the CRB in Chapter 2. Note that the approximation error will become zero if the colored noise is not present. This also indicates that the $\mathrm{CRB}_a$ approaches $\mathrm{CRB}_r$ when the colored noise is small.

## § 4.4 Power Allocation Algorithms

In this section, we consider the power allocation problem in AF-OFDM systems. Our approach is to minimize the CRB of the CFO estimation. As we have seen from (4.21), (4.47), and (4.48), the CRBs are functions of the transmitted power of S and that of $R_m$, denoted as $g_0$ and $g_m$. The

optimization problem can then be formulated as

$$\min_{g_0} \min_{\mathbf{g}} \mathrm{CRB}_r \tag{4.49}$$

$$\text{s.t. } \mathscr{E} = g_0 + \sum_{m=1}^{M} g_m = P,$$

where $\mathbf{g} = [g_1, \cdots, g_M]^T$ and $P$ is the total transmission power. Since a closed-form expression for $\mathrm{CRB}_r$ is not available, we then use $\mathrm{CRB}_a$ and $\mathrm{CRB}_w$ as approximations. It is clear that (4.49) is a constrained optimization problem, and many methods can be used to obtain the solution. However, since the functions in $\mathrm{CRB}_a$ and $\mathrm{CRB}_w$ are highly nonlinear and not convex, advanced methods may not be efficient to apply. For simplicity, we propose using the conventional gradient-descent method to solve the problem. From (4.47) and (4.48), we can see that matrix operations are involved in $\mathrm{CRB}_a$ while scalar operations in $\mathrm{CRB}_w$. Thus, the computational complexity with $\mathrm{CRB}_a$ will be higher than that with $\mathrm{CRB}_w$. As we will see in the next section, however, power allocation designed with $\mathrm{CRB}_a$ will have better performance.

In a relay network, the roles of the source and the relays are different in a relay system. For this reason, we denote the power allocated to the source as $\beta P$ and that for the relays as $(1 - \beta)P$ where $0 < \beta < 1$. In other words, $g_0 = \beta P$ and $g_m = w_m(1 - \beta)P$ where $w_m$ corresponds to a normalization factor such that $\sum_{m=1}^{M} w_m = 1$. To simplify the expression, we define

$$
\begin{aligned}
A_m &= G_{R_m D} G_{SRm} \\
B_m &= \sigma_{\eta_m}^2 G_{R_m D}
\end{aligned}
$$

and substitute them into (4.44) and (4.45). Then the variance of $x(k)$ and $\xi(k) + \eta(k)$ can then express as

$$\bar{\sigma}_x^2 = \sum_{m=1}^{M} \frac{A_m w_m \beta(1 - \beta)P^2}{N_m} \tag{4.50}$$

$$\bar{\sigma}_n^2 = \bar{\sigma}_\eta^2 + \sum_{m=1}^{M} \frac{B_m w_m(1 - \beta)P}{N_m} \tag{4.51}$$

where

$$N_m = G_{SR_m}\beta P + \sigma_{\eta_m}^2. \tag{4.52}$$

First, we solve the power allocation problem by minimizing $\text{CRB}_w$. The optimization problem can now be formulated as

$$\min_{\beta} \min_{\mathbf{w}} \text{CRB}_w \tag{4.53}$$

$$\text{s.t.} \quad 0 < \beta < 1,$$
$$\sum_{m=1}^{M} w_m = 1,$$

where $\mathbf{w} = [w_1, \cdots, w_M]$. To facilitate the derivation of the solution, we split the parameters into two groups, $\beta$ and $w_m$'s, and solve the minimization problem alternatively. The alternative optimization method [68] is an iterative method approximating a multidimensional optimization with a series of one-dimensional optimizations. This technique is now exploited to simplify the multi-dimensional minimization in (4.53).

We first optimize $\beta$ for the source, and then $w_m$'s for the relays. The first optimization problem can be expressed as

$$\min_{\beta(\mathbf{w})} \text{CRB}_w \tag{4.54}$$

$$\text{s.t. } 0 < \beta < 1.$$

Denote the value of $\mathbf{w}$ at $k$th iteration as $\mathbf{w}(k)$ and $\mathbf{w}(k) = [w_1(k), \cdots, w_M(k)]^T$. Letting $\mathbf{w}(k)$ be a constant vector and taking a derivative of (4.54) with respect to $\beta$, we can have

$$\mathbb{D}_{1,b} = \frac{\partial}{\partial \beta} \text{CRB}_w = \frac{Q^2}{8\pi^2 N \sum_{p=1}^{Q} \sum_{q>p}^{Q} (q-p)^2} \frac{Q\bar{\sigma}_x^2 + 2\bar{\sigma}_n^2}{\bar{\sigma}_x^6} [\bar{\sigma}_x^2 \frac{\partial}{\partial \beta} \bar{\sigma}_n^2 - \bar{\sigma}_n^2 \frac{\partial}{\partial \beta} \bar{\sigma}_x^2], \tag{4.55}$$

where

$$\frac{\partial}{\partial \beta} \bar{\sigma}_x^2 = \sum_{m=1}^{M} w_m(k) P^2 A_m \left( \frac{1 - 2\beta}{N_m} - \frac{\beta(1-\beta)G_{SR_m}P}{N_m^2} \right) \tag{4.56}$$

$$\frac{\partial}{\partial \beta} \bar{\sigma}_n^2 = \frac{\partial}{\partial \beta} \bar{\sigma}_\xi^2 = -\sum_{m=1}^{M} w_m(k) P B_m \left( \frac{1}{N_m} + \frac{(1-\beta)G_{SR_m}P}{N_m^2} \right). \tag{4.57}$$

75

Setting $\mathbb{D}_{1,b} = 0$, we can the solve the problem, i.e.,

$$\beta(k+1) = \mathcal{R}(\mathbb{D}_{1,b} = 0) \tag{4.58}$$

where $\mathcal{R}(\cdot)$ indicates the root of a polynomial and $\beta(k+1)$ is the $\beta$ value obtained at the $(k+1)$th iteration. Note that by definition $\beta$ must be greater than zero and smaller than one. From (4.55), we can see that $\mathbb{D}_{1,b} < 0$ when $\beta = 0$ since $\bar{\sigma}_x^2 = 0$ and $\partial\bar{\sigma}_x^2/\partial\beta > 0$. When $\beta = 1$, $\mathbb{D}_{1,b} > 0$ since $\partial\bar{\sigma}_n^2/\partial\beta = 0$ ($\bar{\sigma}_n^2 = \bar{\sigma}_\eta^2$) and $\partial\bar{\sigma}_x^2/\partial\beta < 0$. We then conclude that the root in (4.58) must be in the range of $(0,1)$, and the constraint of $\beta$ is automatically satisfied.

After $\beta(k+1)$ is solved, we substitute the solution into (4.53) and we obtain the second optimization problem as

$$\min_{\mathbf{w}(\beta(k+1))} \text{CRB}_w \tag{4.59}$$
$$\text{s.t.} \sum_{m=1}^{M} w_m = 1.$$

Taking the gradient with respect to $w_m$ in (4.53), we obtain

$$\mathbb{D}_{1,w_m} = \frac{\partial}{\partial w_m}\text{CRB}_w = \frac{Q^2}{8\pi^2 N \sum_{p=1}^{Q}\sum_{q>p}^{Q}(q-p)^2}[\frac{(\bar{\sigma}_n^2 + Q\bar{\sigma}_x^2)\frac{\partial}{\partial w_m}\bar{\sigma}_n^2}{\bar{\sigma}_x^4} + \frac{\bar{\sigma}_n^2(\frac{\partial}{\partial w_m}\bar{\sigma}_n^2 + Q\frac{\partial}{\partial w_m}\bar{\sigma}_x^2)}{\bar{\sigma}_x^4}$$
$$- 2\frac{\bar{\sigma}_n^2(\bar{\sigma}_n^2 + Q\bar{\sigma}_x^2)\frac{\partial}{\partial w_m}\bar{\sigma}_x^2}{\bar{\sigma}_x^6}], \tag{4.60}$$

where

$$\frac{\partial}{\partial w_m}\bar{\sigma}_x^2 = \frac{P^2 A_m \beta(1 - \beta(k+1))}{N_m}, \tag{4.61}$$
$$\frac{\partial}{\partial w_m}\bar{\sigma}_n^2 = \frac{(1 - \beta(k+1))PB_m}{N_m}. \tag{4.62}$$

To solve the optimization problem in (4.59), we use a gradient descent method in conjunction with a constrained projection [17]:

$$\mathbf{w}(k+1) = [\mathbf{w}(k) + \mu_w \mathbb{D}_{1,\mathbf{w}}]_{\mathscr{C}_w=1}^+, \tag{4.63}$$

76

where $\mu_w$ is the step size, $\mathbb{D}_{1,\mathbf{w}} = [\mathbb{D}_{1,w_1}, \cdots, \mathbb{D}_{1,w_M}]^T$, $\mathscr{E}_w = \sum_{m=1}^{M} w_m$, and $[\cdot]_{\mathscr{E}_w=1}^+$ indicates the projection of $\mathbf{w}(k+1)$ (with power of $\mathscr{E}_w$) onto a feasible solution. As shown in [63], the projection can be easily implemented by an normalization operation (normalizing the sum of $w_m(k+1)$'s). Based on the projected $w_m(k+1)$'s, the total transmitted power is normalized to satisfy the power constraint. Here, we summarize the operation of the gradient method as follows:

1. Initialization : Set $k = 0$ and initialize the $\mathbf{w}$ by a uniform power allocation method : $\beta(0) = 1/(M+1)$ and $w_m(0) = 1/M$ for all $m$.

2. Substitute $\beta(k)$ and $\mathbf{w}(k)$ into (4.55), set the result as zero, and obtain $\beta(k+1)$.

3. Use (4.60) to update $\mathbf{w}(k)$: $\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\mathbb{D}_{1,w}$.

4. If $\sum_{m=1}^{M} w_m(k+1) \neq 1$ , conduct the projection as:
   $\mathbf{w}(k+1) = \mathbf{w}(k+1)/\mathscr{E}_w$, where $\mathscr{E}_w = \sum_{m=1}^{M} w_m(k+1)$.

5. Check if $k$ reach the maximum number of iteration. If yes, the algorithm stops. Otherwise, set $k$ as $k+1$ and go to Step 2.

For convenience, we refer to this algorithm as proposed power allocation Algorithm I (PPAA-I).

Using the similar approach, we can obtain the power allocation scheme minimizing $\text{CRB}_a$. Invoking the alternative optimization method, we can have the first problem for minimizing $\text{CRB}_a$ as

$$\min_{\beta(\mathbf{w})} \text{CRB}_a \tag{4.64}$$

$$\text{s.t. } 0 < \beta < 1$$

and then the second problem as

$$\min_{\mathbf{w}(\beta)} \text{CRB}_a \tag{4.65}$$

$$\text{s.t. } \sum_{m=1}^{M} w_m = 1.$$

As that in PPAA-I, we solve $\beta$ first. Taking derivative of $\text{CRB}_a$ with respect to $\beta$, we have [from (4.47)]

$$\mathbb{D}_{2,b} = \frac{\partial}{\partial \beta} \text{CRB}_a = -\frac{1}{F_a^2} \frac{\partial F_a}{\partial \beta}. \tag{4.66}$$

From (4.46), the approximated Fisher information can be rewritten as $F_a = tr(\bar{\mathbf{R}}_y \bar{\mathbf{R}}_y' \bar{\mathbf{R}}_y \bar{\mathbf{R}}_y')$, where

$$\bar{\mathbf{R}}_y = \begin{bmatrix} \bar{\mathbf{R}}_1 & \bar{\mathbf{R}}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \bar{\mathbf{R}}_2^H & \bar{\mathbf{R}}_1 & \bar{\mathbf{R}}_2 & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{R}}_2^H & \bar{\mathbf{R}}_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \bar{\mathbf{R}}_1 \end{bmatrix}, \tag{4.67}$$

$$\bar{\mathbf{R}}_2 = \sqrt{\bar{\sigma}_\gamma^2(1)}\mathbf{I}_Q$$

$$\bar{\mathbf{R}}_y' = \frac{\partial \bar{\mathbf{R}}_y}{\partial \varepsilon_{SD}}$$

$$= \begin{bmatrix} \bar{\mathbf{R}}_1' & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{R}}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \bar{\mathbf{R}}_1' & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \bar{\mathbf{R}}_1. \end{bmatrix}, \tag{4.68}$$

The matrix $\bar{\mathbf{R}}_1'$ is shown in (4.42). Let $\mathbf{Z} = \bar{\mathbf{R}}_y^{-1} \bar{\mathbf{R}}_y'$ and we have $F_a = \text{tr}(\mathbf{ZZ})$. Taking a derivative of $F_a$ with respect to $\beta$, we have [51] [65]

$$\begin{aligned} \frac{\partial F_a}{\partial \beta} &= \text{tr}([\frac{\partial \text{tr}(\mathbf{ZZ})}{\partial \mathbf{Z}}]^T \frac{\partial \mathbf{Z}}{\partial \beta}) \\ &= 2 \cdot \text{tr}(\mathbf{Z} \frac{\partial \mathbf{Z}}{\partial \beta}) \\ &= -2 \cdot \text{tr}(\bar{\mathbf{R}}_y^{-1} \bar{\mathbf{R}}_y' \bar{\mathbf{R}}_y^{-1} \frac{\partial \bar{\mathbf{R}}_y}{\partial \beta} \bar{\mathbf{R}}_y^{-1} \bar{\mathbf{R}}_y') + 2 \cdot \text{tr}(\bar{\mathbf{R}}_y^{-1} \bar{\mathbf{R}}_y' \bar{\mathbf{R}}_y^{-1} \frac{\partial \bar{\mathbf{R}}_y'}{\partial \beta}) \end{aligned} \tag{4.69}$$

where

$$
(\frac{\partial \bar{\mathbf{R}}_1}{\partial \beta})_{p,q} = \begin{cases} \frac{\partial \bar{\sigma}_x^2}{\partial \beta} + \frac{\partial \bar{\sigma}_n^2}{\partial \beta} & \text{, if } p = q \\ \frac{\partial \bar{\sigma}_x^2}{\partial \beta} \omega^{\varepsilon_{SD}(p-q)N} & \text{, otherwise} \end{cases},
$$

$$
\frac{\partial \bar{\mathbf{R}}_2}{\partial \beta} = \frac{\partial \sqrt{\bar{\sigma}_\gamma^2(1)}}{\partial \beta} \mathbf{I}_Q,
$$

$$
(\frac{\partial \bar{\mathbf{R}}_1'}{\partial \beta})_{p,q} = \begin{cases} 0 & \text{, if } p = q \\ j\frac{2\pi(p-q)}{Q}\frac{\partial \bar{\sigma}_x^2}{\partial \beta}\omega^{\varepsilon_{SD}(p-q)N} & \text{, otherwise} \end{cases},
$$

and

$$
\frac{\partial \sqrt{\bar{\sigma}_\gamma^2(1)}}{\partial \beta} = \sum_{m=1}^M -\frac{w_m^2(1-\beta)P^2\sigma_{\eta_m}^4\lambda_{R_mD}^2}{\sqrt{\bar{\sigma}_\gamma^2(1)}}e^{-2\alpha}\frac{1-e^{-4\alpha(L-1)}}{1-e^{-4\alpha}}(\frac{N_m+(1-\beta)G_{SR_m}P}{N_m^3}).
$$

Unlike PPAA-I, the $\beta$ solved in $\mathcal{R}(\mathbb{D}_{2,b} = 0)$ cannot be guaranteed to meet the constraint $(0 < \beta < 1)$. When the constraint is not met, we then take the minimum $\beta$ value in the admissible interval, i.e.,

$$
\beta(k+1) = \min_{0<\beta<1} |\mathbb{D}_{2,b}|. \tag{4.70}
$$

To solve the second problem in (4.65), we take a derivative of $\text{CRB}_a$ with respect to $\mathbf{w}$:

$$
\mathbb{D}_{2,w_m} = \frac{\partial}{\partial w_m}\text{CRB}_a = -\frac{1}{F_a^2}\frac{\partial F_a}{\partial w_m}. \tag{4.71}
$$

Note that the result is similar to that in (4.69) except that $\beta$ in (4.69) is replaced with $w_m$. Thus, we have

$$
(\frac{\partial \bar{\mathbf{R}}_1}{\partial w_m})_{p,q} = \begin{cases} \frac{\partial \bar{\sigma}_x^2}{\partial w_m} + \frac{\partial \bar{\sigma}_n^2}{\partial w_m} & \text{, if } p = q \\ \frac{\partial \bar{\sigma}_x^2}{\partial w_m}\omega^{\varepsilon_{SD}(p-q)N} & \text{, otherwise} \end{cases},
$$

$$
\frac{\partial \bar{\mathbf{R}}_2}{\partial w_m} = \frac{\partial \sqrt{\bar{\sigma}_\gamma^2(1)}}{\partial w_m}\mathbf{I}_Q,
$$

$$
(\frac{\partial \bar{\mathbf{R}}_1'}{\partial w_m})_{p,q} = \begin{cases} 0 & \text{, if } p = q \\ j\frac{2\pi(p-q)}{Q}\frac{\partial \bar{\sigma}_x^2}{\partial w_m}\omega^{\varepsilon_{SD}(p-q)N} & \text{, otherwise} \end{cases},
$$

and

$$\frac{\partial \sqrt{\bar{\sigma}_\gamma^2(1)}}{\partial w_m} = \frac{\sigma_{\eta_m}^4 \lambda_{R_m D}^2 \beta (1-\beta)^2 P^2 w_m}{N_m^2 \sqrt{\bar{\sigma}_\gamma^2(1)}} e^{-2\alpha} \frac{1 - e^{-4\alpha(L-1)}}{1 - e^{-4\alpha}}.$$

Let $\mathbb{D}_{2,\mathbf{w}} = [\mathbb{D}_{2,w_1}, \cdots, \mathbb{D}_{2,w_M}]$ and we can obtain the update of $\mathbf{w}$ as:

$$\mathbf{w}(k+1) = [\mathbf{w}(k) + \mu_w \mathbb{D}_{2,\mathbf{w}}]_{\mathscr{E}_w = 1}^+ \tag{4.72}$$

The gradient method can be summarized as follows :

1. Initialization : set $k = 0$ and initialize the $\mathbf{w}$ by a uniform power allocation method: $\beta(0) = 1/(M+1)$ and $w_m(0) = 1/M$ for all $m$.

2. Substitute $\beta(k)$ and $\mathbf{w}(k)$ into (4.70) and obtain $\beta(k+1)$.

3. Use (4.72) to obtain $\mathbf{w}(k+1)$.

4. If $\sum_{m=1}^M w_m(k+1) \neq 1$ , let $\mathbf{w}(k+1) = \mathbf{w}(k+1)/\mathscr{E}$, where $\mathscr{E} = \sum_{m=1}^M w_m(k+1)$.

5. Check if $k$ reach the maximum number of iteration. If yes, the algorithm stops. Otherwise, set $k \leftarrow k+1$ and go to Step 2.

For convenience, we refer to this algorithm as the proposed power allocation algorithm II (PPAA-II).

## § 4.5 Simulation Results

In this section, we report simulation results demonstrating the effectiveness of the proposed ML CFO estimator and the proposed power allocation algorithms. Here, we consider an AF-OFDM system with the symbol size of 128 and the preamble period of 64. The AF-OFDM system has two relays and the total transmitted power is 10 ($P = 10$). The channel model of the HIPERLAN/2 is used and the number of the taps is 8. To simply the scenario, we let the

decay factors for S-$R_m$ and $R_m$-D links are the same, i.e. $\alpha_{SR_m} = \alpha_{R_mD} = \alpha = 0.5$ for all $m$. The noise variance at each relay and that at the destination are set as one.

In the first set of simulations, we compare the performance of CFO estimation for the cases when the colored property of the noise is taken into account or not. The step size for the proposed gradient-descent method is set as $0.01$. We consider a scenario that $\text{SNR}_{SR_m} = \text{SNR}_{SR}$ and $\text{SNR}_{R_mD} = \text{SNR}_{RD}$ for all $m$, where

$$\text{SNR}_{SR_m} = G_{SR_m} g_0 / \sigma^2_{\eta_m}$$

and

$$\text{SNR}_{R_mD} = g_m \frac{G_{R_mD}}{\bar{\sigma}^2_\eta}.$$

We also assume that $\gamma_m(k)$'s in (4.11) are perfectly known. Figure 4.3 and Figure 4.4 show the simulation results for $\text{SNR}_{RD}$ is 10 and 20 dB, respectively. There are a couple of observations we can derive from the figures. First, the proposed ML estimate taking the effect of colored noise into account outperforms the conventional algorithm ignoring the effect. Second, the MSE of the proposed gradient-descent method is close to $\text{CRB}_r$, indicating its optimum performance. Third, we see that the proposed $\text{CRB}_a$ is close to $\text{CRB}_r$. This is to say that the approximation error of (4.40) is small. As defined, $\text{SNR}_{RD}$ in Figure 4.4 is higher than that in Figure 4.3. Thus, the colored-noise effect in Figure 4.4 is more severe than that in Figure 4.3, especially when $\text{SNR}_{SR}$ is low. This is the reason why the gap between $\text{MSE}_w$ and $\text{CRB}_r$ is larger than that in Figure 4.4 when $\text{SNR}_{SR}$ is low. Finally, we can see that the gap between $\text{CRB}_a$ and $\text{CRB}_r$ becomes smaller when $\text{SNR}_{SR}$ is higher. This is because in higher $\text{SNR}_{SR}$, the colored-noise effect becomes smaller. The observations match the analysis derived in Section 4.3.

In the previous set of simulations, we assume that $\gamma_m(k)$'s are perfectly known and use $\text{MSE}_r$ to evaluate the estimation performance. One may wander how the estimation errors of $\gamma_m(k)$'s will affect the CFO estimation. We then consider the scenario that the relays transmit white sequences in the BP and the receiver at D conducts estimation of $\gamma_m(k)$'s with the received signal. Let the number of OFDM symbols transmitted be $K$. Denote $\text{MSE}_c$ as the estimation

81

performance by using the estimated $\gamma_m(k)$'s, instead of perfectly known $\gamma_m(k)$'s. Figure 4.5 and Figure 4.6 show the simulation results. It is seen that with only two training symbols, the performance of the proposed algorithm with estimated $\gamma_m(k)$'s is very close to that with known $\gamma_m(k)$'s. When $\mathrm{SNR}_{RD}$ is high, the proposed algorithm can provide good performance even with one training symbol.

We then investigate the convergence behavior of proposed PAAs. Note that the SNR is not an appropriate measure in evaluating the performance of a PAA since the transmitted power for each transmitter will be different for different PAAs. For this reason, we use the channel gain to reflect the channel quality. Here, we let the channel gains for the S-$R_m$ link and the $R_1$-D link be set as: $\lambda_{SR_1} = 0.1$, $\lambda_{SR_2} = \lambda_{R_1D} = 5$, and $\lambda_{R_2D} = 5$. Since the proposed algorithms are gradient-based, the choice of the step size is critical. The larger the step size, the faster the algorithm will converge. However, if the step size is too large, the algorithm may become unstable. The fastest convergence rate a gradient-descent method can achieve depends on the characteristic of the optimization problem. In our simulations, we let the step size be 100, obtained by trial-and-errors. Figure 4.7 shows the learning curves of the proposed PAAs. As we can see, PPAA-I converges around 20 iterations while PPAA-II around 25. This indicates both PPAA-I and PPAA-II converge fast.

In the third set of simulations, we evaluate the performance of proposed PAAs. To see the effect of the colored noise, we let $\lambda_{R_2D}$ be varied. Figure 4.8 shows the achieved CRB for proposed and uniform power allocation algorithms. As we can see, the proposed PPAs have much lower CRBs. Also, PPAA-II is always outperform PPAA-I. It is clear that if $\lambda_{R_2D}$ is large, the colored-noise effect in the destination will be more severe. From the figure, we also find that the difference between the all PAAs is smaller when $\lambda_{R_2D}$ is small since the S-$R_1$ and $R_2$-D links are too week to transmit data. When the level of $\lambda_{R_2D}$ is increased, the performance gap is increased. When $\lambda_{R_1D} = \lambda_{R_2D} = 5$, all the relay diversity can be exploited since the $R_m$-D channel gain is the same for all $m$ and the performance gap between PPAA-I and PPAA-II is largest. When $\lambda_{R_2D} \gg \lambda_{R_1D}$, the PAA between relays will be reduced to a relay selection

problem and the performance gap between PPAA-I and PPAA-II becomes small again.

As shown, the proposed PAAs are designed to minimize the CRB of the CFO estimation. We are curious that if the PAAs are also helpful for signal detection. It is simple to see that if $\text{SNR}_{SD}$ is high, the signal recovery will be easier. To evaluate the SNR performance of the proposed PAAs, we use the PAA proposed in [67] as a reference. Figure 4.9 shows $\text{SNR}_{SD}$ for various PAAs. From the figure, we find that proposed PAAs provide much higher $\text{SNR}_{SD}$ than the uniform power allocation. Also, the performance of proposed PAAs is very close to that of the PAA in [67], especially when $\text{SNR}_{SD}$ is high. Note that the PAA in [67] is designed to maximize the overall capacity of a relay system and that in the proposed algorithm is to minimize the CRB of the CFO estimate. In the following, we explain why these two different criteria lead to the same result. The capacity of the source-to-destination channel can be expressed as

$$\mathscr{R} = W \log_2(1 + \text{SNR}_{SD}), \tag{4.73}$$

where $W$ is the bandwidth and $\text{SNR}_{SD} = \bar{\sigma}_x^2/\bar{\sigma}_n^2$. Using $\text{SNR}_{SD}$ in $\text{CRB}_w$, we can rewrite (4.48) as

$$\text{CRB}_w = \frac{Q^2}{8\pi^2 N \cdot \sum\limits_{p=1}^{Q-1} \sum\limits_{q>p}^{Q} (q-p)^2} \frac{(1 + Q \cdot \text{SNR}_{SD})}{\text{SNR}_{SD}^2}. \tag{4.74}$$

When $\text{SNR}_{SD}$ is high, the constant 1 in (4.73) and (4.74) can be ignored. Thus, (4.73) and (4.74) can be rewritten as

$$\mathscr{R} \propto \log_2(\text{SNR}_{SD}), \tag{4.75}$$

and

$$\text{CRB}_w \propto \frac{1}{\text{SNR}_{SD}}. \tag{4.76}$$

From (4.75) and (4.76), we can see that maximizing $\mathscr{R}$ is equivalent to maximizing $\text{SNR}_{SD}$ and then minimizing $\text{CRB}_w$. A direct link between $\text{SNR}_{SD}$ and $\text{CRB}_a$ is difficult to derive. However, as shown in (4.48), $\text{CRB}_a$ and $\text{CRB}_w$ are similar when the colored noise is small. We

can then conjec                                    r to

Figure 4.8, PP/                                    sed

PAAs can not o                                     ing

proposed PAAs                                      n is
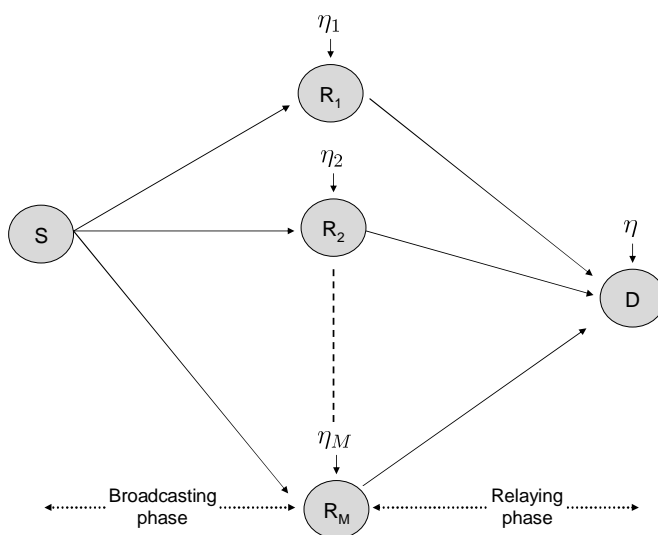
initiated. This v



Figure 4.1: Cooperative system with one source-destination pair and $M$ relay nodes. The noise for relay nodes and destination node are AWGN.
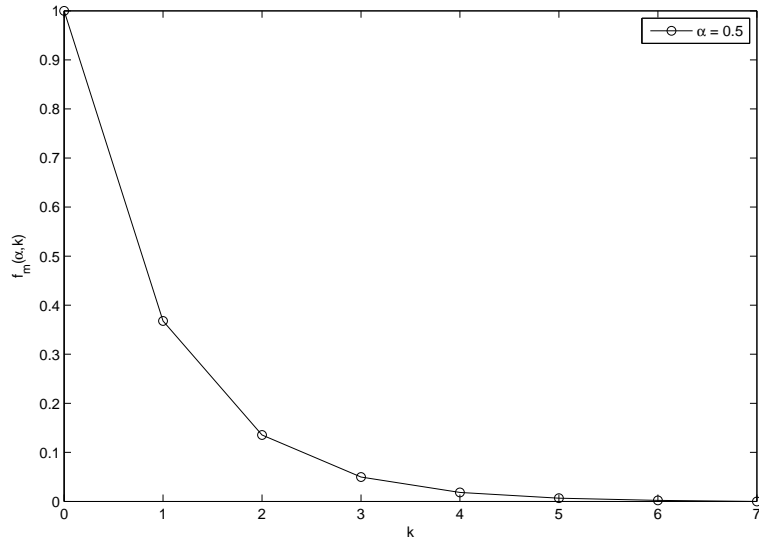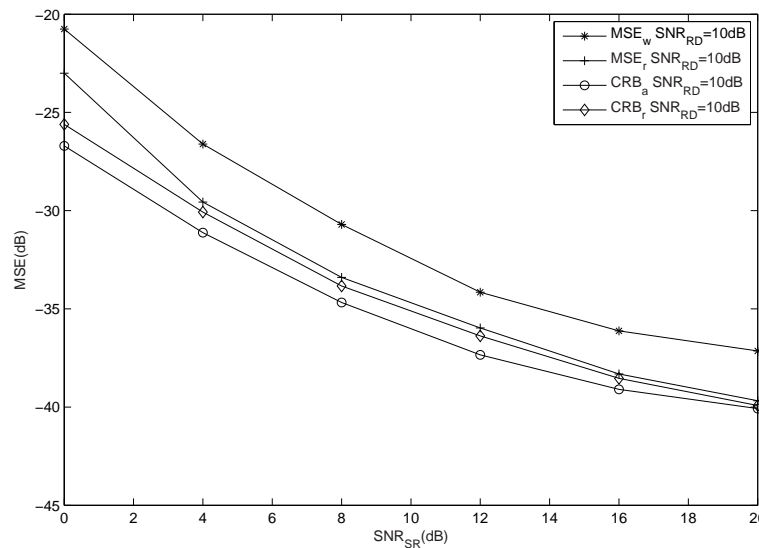
1

Figure 4.2: Normalized exponential decay factor.



Figure 4.3: Performance evaluation with perfect $\bar{\gamma}(l)$ for low $\text{SNR}_{RD}$ case.
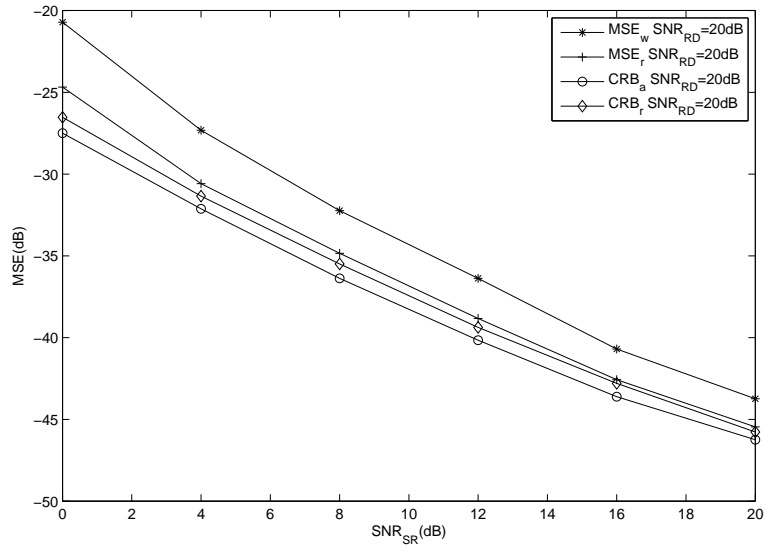
Figure 4.4: Performance evaluation with perfect $\bar{\gamma}(l)$ for high SNR$_{RD}$ case.
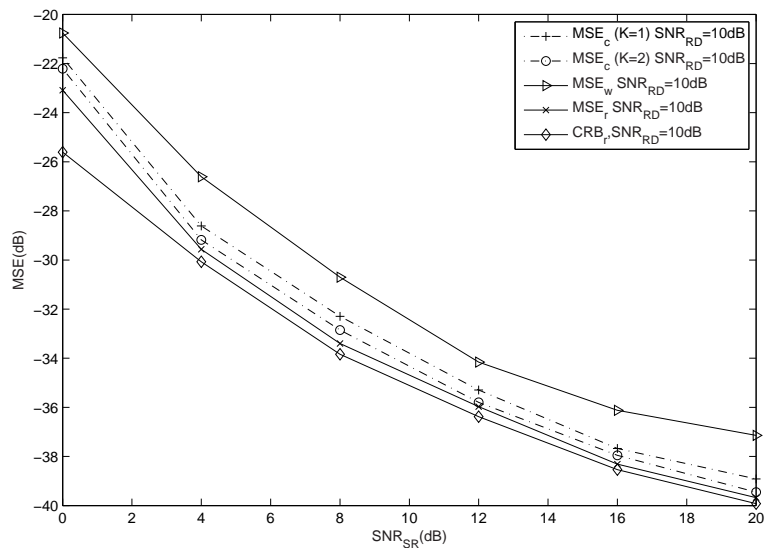


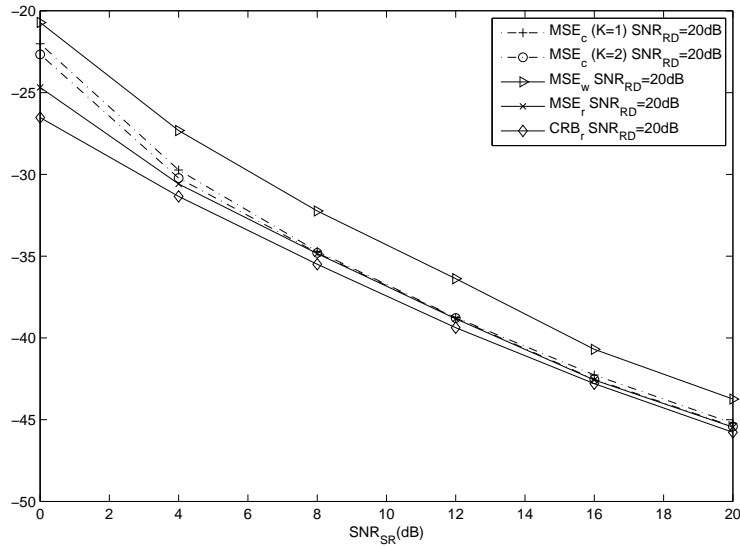Figure 4.5: Performance evaluation with estimated $\bar{\gamma}(l)$ for low SNR$_{RD}$ case.

86

Figure 4.6: Performance evaluation with estimated $\bar{\gamma}(l)$ for high SNR$_{RD}$ case.
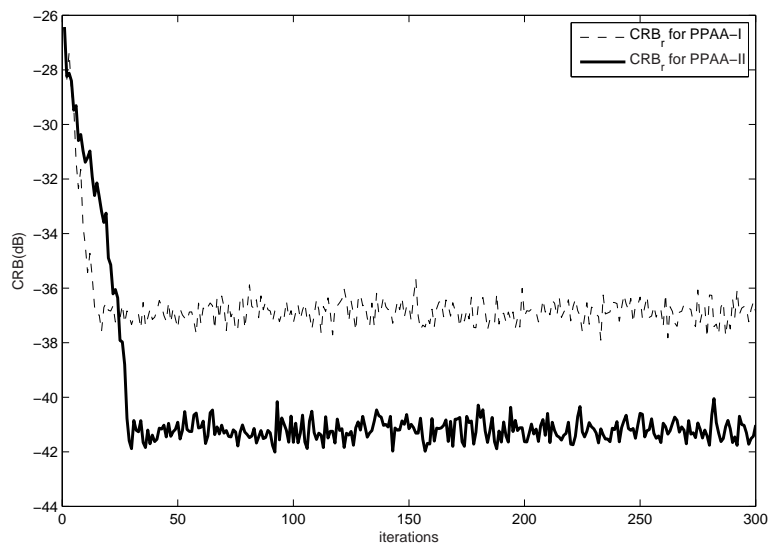


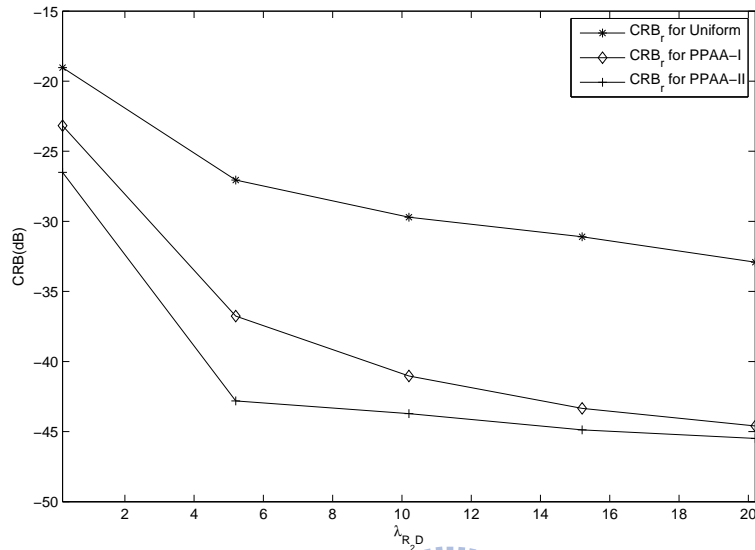Figure 4.7: Evolution of the proposed algorithms

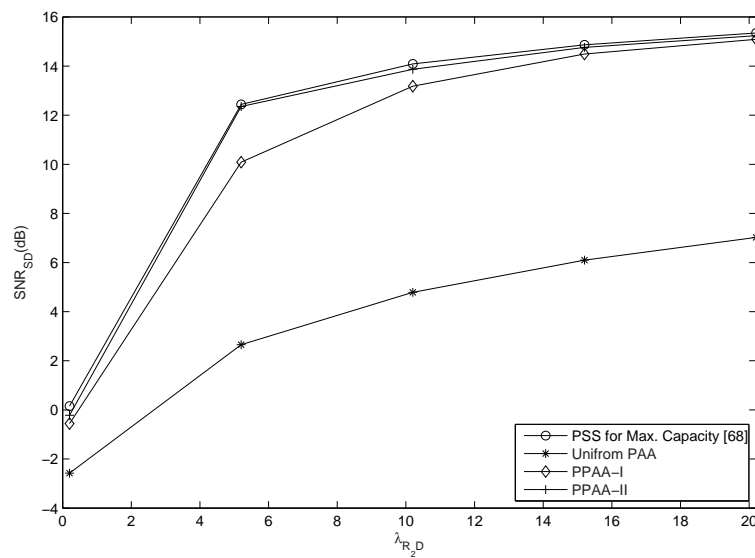Figure 4.8: The performance comparison between proposed algorithms.



Figure 4.9: The corresponding average SNR$_{SD}$ for proposed algorithms.

# Chapter 5

# Conclusions

In this dissertation, we focus on the ML CFO estimation in OFDM, OFDMA uplink, and AF-OFDM systems. For OFDM systems, some conventional CFO estimations are not ML-based and the performance cannot approach the CRB. The others are ML-based requiring the inversion of a large correlation matrix, and the computational complexity is usually very high, precluding the real-world applications of these methods. This motivates us to study the low-complexity ML CFO estimation algorithm. To the best of our knowledge, there are no blind ML algorithms for estimating CFO in OFDMA uplink systems. To fill the gap, we then further study blind ML CFO estimations for OFDMA uplink systems. Simulation results show that while the computational complexity of the proposed algorithms are low, the performance can approach the CRBs. Finally, we consider the ML CFO estimation problem in cooperative AF-OFDM systems. For the systems, all conventional ML CFO estimations are not optimal. This is because the receive noise at the destination is colored, a case not considered before. The colored property complicates the calculation of the correlation-matrix inverse, making the closed-form ML solution and the corresponding CRB difficult to derive. To solve the problem, we then propose using a gradient-descent method to obtain the ML solution and a new method to derive two closed-form expression for the CRBs. We also propose two PAAs to minimize the CRBs of the CFO estimation. From simulations, we see that the proposed PAAs not only improve the CFO

estimation, but also improve the received SNR. The distinct feature of the algorithms proposed in this dissertation is that only the periodicity of the preamble is assumed. The exact values of the preamble sequence are not required. It turns of that two random variables result in the likelihood functions, i.e. noise and the received signal. This is different from the conventional ML CFO estimation in which only the noise is considered as the random variable.

In concluding the dissertation, we suggest some possible topics for future research.

1. In this dissertation, we only consider a single-antenna scenario. Nowadays, multi-input-multi-output (MIMO) OFDMs are widely used. The ML CFO estimation in MIMO-OFDM can then serve an interesting problem for further study. Note that there may be correlations between antennas in the transmitters or receivers.

2. Although the proposed method in OFDMA uplink systems is simple, it cannot be used in a full-loaded scenario (the number of users is the same as the number of subchannels in OFDMA systems). How to extend the proposed method to such a scenario deserve further study. An possible approach is to use an expectation-maximization (EM) algorithm referred to as iterative space alternating generalized EM (SAGE) [55], [56]. However, the complexity of the SAGE algorithm can be very high for large $N_s$. Note that in real-world applications, only a number of users will be activated at a specific time [56]. Thus, only the CFOs of the newly activated users have to be estimated, and the knowledge of the previously estimated CFOs can be exploited in each new estimation. It is interesting to incorporate the SAGE algorithm into the proposed method, which may serve as a topic for further research.

3. In AF-OFDM systems, we solve the ML CFO estimation with the gradient-descent method. Though difficult, it may be worth pursuing the closed-form expression for the ML solution. Similarly, the closed-form solution for the PAAs minimizing CRB in AF-OFDM system can also be investigated.

4. The OFDM modulation scheme is widely used in real-world communication systems. It

is then desirable to consider the subcarrier power allocation in AF-OFDM or MIMO-AF-OFDM relay systems. The design problem is obviously much complicated since one extra dimension, the subcarrier, is added. Note that the number of subcarriers is often large in real-world applications..

5. In this dissertation, we only study AF-OFDM systems with one source node, multiple relay nodes, and one destination node. In cooperative systems, other scenarios are possible. For example, there may be multiple source or destination nodes. In such case, the number of CFO is more than one and the estimation problem will becomes more challenging and also deserve for further study. Also, we have not considered the direct link between the source node and the destination node, which is another scenario can be investigated.
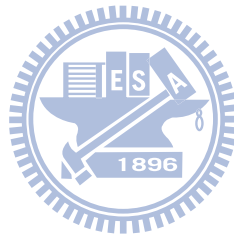
# Appendix A

## § A.1   Derivation of (2.19)

The likelihood function in (2.18) can be rewritten as

$$
\begin{aligned}
\Lambda(\varepsilon) &= \sum_{n=0}^{N-1} \ln \left\{ \frac{(\sigma_x^2 + \sigma_w^2)^K \exp\left[ -\mathbf{y}(n)^H \mathbf{R}_y^{-1} \mathbf{y}(n) \right]}{\det(\mathbf{R}_y)\exp\left[ -\frac{\sum\limits_{p=1}^{K} y_p(n)y_p^*(n)}{\sigma_x^2+\sigma_w^2} \right]} \right\} \\
&= \sum_{n=0}^{N-1} \{\ln[(\sigma_x^2 + \sigma_w^2)^K (\det(\mathbf{R}_y))^{-1}] + (\sum_{p=1}^{K} y_p(n)y_p^*(n))/(\sigma_x^2 + \sigma_w^2) \\
&\quad -\mathbf{y}(n)^H \mathbf{R}_y^{-1} \mathbf{y}(n)\}.
\end{aligned}
\tag{5.1}
$$

Then, substituting (2.17) into (5.1), we derive the log-likelihood function as

$$
\begin{aligned}
\Lambda(\varepsilon) &= \sum_{n=0}^{N-1} \{\ln[\frac{(\sigma_x^2 + \sigma_w^2)^K}{\det(\mathbf{R}_y)}] + \frac{\sum\limits_{p=1}^{K} y_p(n)y_p^*(n)}{\sigma_x^2 + \sigma_w^2} - (\sigma_w^{-2} - C_0)\sum_{p=1}^{K} y_p(n)y_p^*(n) \\
&\quad +2C_0 \mathrm{Re}\{\sum_{p=1}^{K-1}\sum_{q>p}^{K} y_p(n)y_q^*(n)e^{j2\pi(q-p)\varepsilon}\}\}
\end{aligned}
$$

93

$$
= \sum_{n=0}^{N-1}\{\ln[(\sigma_x^2+\sigma_w^2)^K(\det(\mathbf{R}_y))^{-1}]+[\frac{1}{\sigma_x^2+\sigma_w^2}-(\sigma_w^{-2}-C_0)]\sum_{p=1}^{K}y_p(n)y_p^*(n)
$$

$$
+2C_0\mathrm{Re}\{\sum_{p=1}^{K-1}\sum_{q>p}^{K}y_p(n)y_q^*(n)e^{j2\pi(q-p)\varepsilon}\}\}
$$

$$
= \frac{(1-K)\sigma_x^4}{\sigma_w^2(\sigma_x^2+\sigma_w^2)(K\sigma_x^2+\sigma_w^2)}\sum_{p=1}^{K}\sum_{n=0}^{N-1}y_p(n)y_p^*(n)
$$

$$
+2C_0\mathrm{Re}\{\sum_{p=1}^{K-1}\sum_{q>p}^{K}\sum_{n=0}^{N-1}y_p(n)y_q^*(n)e^{j2\pi(q-p)\varepsilon}\}
$$

$$
+N\{\ln[(\sigma_x^2+\sigma_w^2)^K(\det(\mathbf{R}_y))^{-1}]\} \tag{5.2}
$$

where $C_0=\sigma_x^2/(\sigma_w^4+K\sigma_w^2\sigma_x^2)$. Substituting (2.21) and (2.25) into (5.2), we can express (5.2) as

$$
\Lambda(\varepsilon) = C_1+C_2\sum_{p=1}^{K}\gamma_{pp}+C_3\mathrm{Re}\{\sum_{p=1}^{K-1}\sum_{q>p}^{K}\gamma_{pq}e^{j2\pi(q-p)\varepsilon}\}
$$

$$
= C_1+C_2\sum_{p=1}^{K}\gamma_{pp}+C_3\mathrm{Re}\{\sum_{p=1}^{K-1}\sum_{q>p}^{K}(|\gamma_{pq}|e^{j\angle\gamma_{pq}})e^{j2\pi(q-p)\varepsilon}\}
$$

$$
= C_1+C_2\phi+C_3\sum_{p=1}^{K-1}\sum_{q>p}^{K}|\gamma_{pq}|\cos(\psi_{pq}) \tag{5.3}
$$

, where $\psi_{pq}$, $\phi$, $C_1$, $C_2$, $C_3$ are defined as (2.22)-(2.24).

## § A.2  Derivation of (2.38)

We assume that the channel noise, the received preamble, and the received data are statistically uncorrelated one another. We define three column vectors $\mathbf{y}_1(n)=[y_0(n),\cdots,y_{Q-1}(n)]^T$, $\mathbf{y}_2(n)=[y_1(n),\cdots,y_{Q-1}(n)]^T$, and $\mathbf{y}_3(n)=[y_0(n),\cdots,y_{Q-2}(n)]^T$ and their correlation matrix as $\mathbf{R}_{yk}$ for $k=1,2$ and 3. Note that $i$ is the window index of (2.37) and $\theta$ is the real STO. So (2.19) can be derived for two cases, $i\leq\theta$ and $i>\theta$. Using the approach taken to derive

(2.18), we obtain the log-likelihood function for the first case as

$$
\begin{aligned}
\Lambda^{i \leq \theta}(\varepsilon) &= \ln\left\{ \prod_{n=i}^{n=i+N-1} \frac{f(\mathbf{y}(n))}{\prod_{k=1}^{Q} f(y_{k-1}(n))} \right\} \\
&= \sum_{n=i}^{\theta-1} \ln\left\{ \frac{f(\mathbf{y}_2(n))}{f(y_1(n)) \cdots f(y_{Q-1}(n))} \right\} \\
&\quad + \sum_{n=\theta}^{i+N-1} \ln\left\{ \frac{f(\mathbf{y}_1(n))}{f(y_0(n)) \cdots f(y_{Q-1}(n))} \right\} \\
&= \frac{\theta - i}{N} C_{12} + \frac{i + N - \theta}{N} C_1' \\
&\quad + \sum_{n=i}^{\theta-1} C_2 \sum_{p=1}^{Q-1} y_p(n) y_p^*(n) \\
&\quad + \sum_{n=i}^{\theta-1} C_3 \mathrm{Re}\left\{ \sum_{p=1}^{Q-2} \sum_{q>p}^{Q-1} y_p(n) y_q^*(n) e^{j 2\pi (q-p)\varepsilon} \right\} \\
&\quad + \sum_{n=\theta}^{i+N-1} C_3' \mathrm{Re}\left\{ \sum_{p=0}^{Q-2} \sum_{q>p}^{Q-1} y_p(n) y_q^*(n) e^{j 2\pi (q-p)\varepsilon} \right\} \\
&\quad + \sum_{n=\theta}^{i+N-1} C_2' \sum_{p=0}^{Q-1} y_p(n) y_p^*(n),
\end{aligned}
$$
(5.4)
(5.5)

where

$$
C_{12} = N \cdot \ln\left( \frac{(\sigma_x^2 + \sigma_w^2)^K}{\det(\mathbf{R}_{y2})} \right),
\tag{5.6}
$$

$$
C_1' = N \cdot \ln\left( \frac{(\sigma_x^2 + \sigma_w^2)^Q}{\det(\mathbf{R}_{y1})} \right),
\tag{5.7}
$$

$$
C_2' = (1 - Q) \frac{\rho^2}{\sigma_w^2 (1 + (Q-1)\rho)},
\tag{5.8}
$$

$$
C_3' = \frac{2 C_2'}{(1 - Q)\rho}.
\tag{5.9}
$$

Similarly, we can derive the log-likelihood function for $i > \theta$ as

$$
\begin{aligned}
\Lambda^{i>\theta}(\varepsilon) &= \sum_{n=\theta+N}^{i+N-1} \ln\left\{\frac{f(\mathbf{y}_3(n))}{f(y_0(n))\cdots f(y_{Q-2}(n))}\right\} \\
&\quad + \sum_{n=i}^{\theta+N-1} \ln\left\{\frac{f(\mathbf{y}_1(n))}{f(y_0(n))\cdots f(y_{Q-1}(n))}\right\} \quad (5.10) \\
&= \frac{i-\theta}{N}C_{13} + \frac{\theta+N-i}{N}C_1' \\
&\quad + \sum_{n=\theta+N}^{i+N-1} C_2 \sum_{p=0}^{Q-2} y_p(n)y_p^*(n) + \\
&\quad \sum_{n=\theta+N}^{i+N-1} C_3 \mathrm{Re}\{\sum_{p=0}^{Q-3}\sum_{q>p}^{Q-2} y_p(n)y_q^*(n)e^{j2\pi(q-p)\varepsilon}\} \\
&\quad + \sum_{n=i}^{\theta+N-1} C_2' \sum_{p=0}^{Q-1} y_p(n)y_p^*(n) + \\
&\quad \sum_{n=i}^{\theta+N-1} C_3' \mathrm{Re}\{\sum_{p=0}^{Q-2}\sum_{q>p}^{Q-1} y_p(n)y_q^*(n)e^{j2\pi(q-p)\varepsilon}\}
\end{aligned}
$$

$$(5.11)$$

where

$$
C_{13} = N \cdot \ln\left(\frac{(\sigma_x^2+\sigma_w^2)^K}{\det(\mathbf{R}_{y3})}\right).
$$

Since $\mathbf{y}_2(n)$, $i \le n \le \theta - 1$, in (5.4) and $\mathbf{y}_3(n)$, $\theta + N \le n \le i + N - 1$, in (5.10) contain $Q - 1$ periods of the preamble, $\det(\mathbf{R}_{y2})$ and $\det(\mathbf{R}_{y3})$, will be the same as $\det(\mathbf{R}_y)$ (in (2.22)). Consequently, $C_{12} = C_{13} = C_1$. From (2.22)-(2.24), we see that $C_1$, $C_2$, and $C_3$ can be calculated by replacing $Q$ and $\mathbf{R}_{y1}$ with $K$ and $\mathbf{R}_y$, respectively, in (5.7)-(5.9). When $Q$ is reasonably large, we obtain $C_1' \approx C_1$, $C_2' \approx C_2$, and $C_3' \approx C_3$. Thus, we rewrite (5.5) and

(5.11) as

$$
\begin{aligned}
\Lambda^{i \leq \theta}(\varepsilon) \quad \simeq \quad & \sum_{n=i}^{\theta-1} C_3 \mathrm{Re}\{\sum_{p=1}^{Q-2} \sum_{q>p}^{Q-1} y_p(n) y_q^*(n) e^{j2\pi(q-p)\varepsilon}\} \\
& + \quad \sum_{n=\theta}^{i+N-1} C_3 \mathrm{Re}\{\sum_{p=0}^{Q-2} \sum_{q>p}^{Q-1} y_p(n) y_q^*(n) e^{j2\pi(q-p)\varepsilon}\} \\
& + \quad C_1 + \sum_{n=i}^{\theta-1} C_2 \sum_{p=1}^{Q-1} y_p(n) y_p^*(n) \\
& + \quad \sum_{n=\theta}^{i+N-1} C_2 \sum_{p=0}^{Q-1} y_p(n) y_p^*(n), \quad\quad\quad (5.12)
\end{aligned}
$$

and

$$
\begin{aligned}
\Lambda^{i > \theta}(\varepsilon) \quad \simeq \quad & \sum_{n=i}^{\theta+N-1} C_3 \mathrm{Re}\{\sum_{p=0}^{Q-2} \sum_{q>p}^{Q-1} y_p(n) y_q^*(n) e^{j2\pi(q-p)\varepsilon}\} \\
& + \quad \sum_{n=\theta+N}^{i+N-1} C_3 \mathrm{Re}\{\sum_{p=0}^{Q-3} \sum_{q>p}^{Q-2} y_p(n) y_q^*(n) e^{j2\pi(q-p)\varepsilon}\} \\
& + \quad C_1 + \sum_{n=\theta+N}^{i+N-1} C_2 \sum_{p=0}^{Q-2} y_p(n) y_p^*(n) \\
& + \quad \sum_{n=i}^{\theta+N-1} C_2 \sum_{p=0}^{Q-1} y_p(n) y_p^*(n). \quad\quad\quad (5.13)
\end{aligned}
$$

We now approximate $\sum_{p=1}^{Q-1} y_p(n) y_p^*(n)$ and $\sum_{p=1}^{Q-2} \sum_{q>p}^{Q-1} y_p(n) y_q^*(n) e^{j2\pi(q-p)\varepsilon}$ in (5.12) with $\sum_{p=0}^{Q-1} y_p(n)$ $y_p^*(n)$ and $\sum_{p=0}^{Q-2} \sum_{q>p}^{Q-1} y_p(n) y_q^*(n) e^{j2\pi(q-p)\varepsilon}$, respectively. Similarly, we also approximate $\sum_{p=0}^{Q-2} y_p(n)$ $y_p^*(n)$ and $\sum_{p=0}^{Q-3} \sum_{q>p}^{Q-2} y_p(n) y_q^*(n) e^{j2\pi(q-p)\varepsilon}$ in (5.13) with $\sum_{p=0}^{Q-1} y_p(n) y_p^*(n)$, and $\sum_{p=0}^{Q-2} \sum_{q>p}^{Q-1} y_p(n) y_q^*(n)$ $e^{j2\pi(q-p)\varepsilon}$, respectively. Given these approximations, $\Lambda^{i \leq \theta}(\varepsilon)$ and $\Lambda^{i > \theta}(\varepsilon)$ can be identically written as

$$
\begin{aligned}
\Lambda^i(\varepsilon) \quad \simeq \quad & C_1 + C_2 \sum_{n=i}^{i+N-1} \sum_{p=0}^{Q-1} y_p(n) y_p^*(n) \quad\quad\quad (5.14) \\
& + \quad C_3 \sum_{n=i}^{i+N-1} \mathrm{Re}\{\sum_{p=0}^{Q-2} \sum_{q>p}^{Q-1} y_p(n) y_q^*(n) e^{j2\pi(q-p)\varepsilon}\}.
\end{aligned}
$$

Using the approach that similar to that in Appendix A.1, we finally obtain

$$\Lambda^i(\varepsilon) \simeq C_1 + C_2\phi^i + C_3\text{Re}\{\sum_{p=0}^{Q-2}\sum_{q>p}^{Q-1}|\gamma_{pq}^i|\cos(\psi_{pq}^i)\} \tag{5.15}$$

where $\gamma_{pq}^i = \sum_{n=i}^{i+N-1} y_p(n)y_q^*(n)$, $\phi^i = \sum_{p=0}^{Q-1}\gamma_{pp}^i$, and $\psi_{pq}^i = 2\pi\varepsilon(q-p) + \angle\gamma_{pq}^i$. Note that the approximations we made are equivalent to adding $|\theta - i|$ samples (noise or data) in calculating the likelihood functions. Since the number of samples in the $i_{th}$ sliding data window, $QN$, is usually much larger than the number of added samples, $|\theta - i|$, the added samples will not change the likelihood functions too much. The approximation errors also depend on the distance between the window position and the actual STO, i.e., $|\theta - i|$. When the distance is larger, the error is larger. However, if the distance is larger, the likelihood function tends to be smaller and a larger error is then tolerable. Finally, we note that the added samples, whether noise or data, are uncorrelated with the preamble samples.

## § A.3   Derivations of $\mu_\phi^i$, $\mu_\xi^i$, $\nu_\phi^i$, $\nu_\xi^i$, and $\kappa_{\phi\xi}^i$

We first note that $\theta$ is the real STO in the system. Using $\theta$ as a reference, we have three cases for the value of $i$: $i = \theta$, $i < \theta$, and $i > \theta$ ($0 \le i \le N - 1$). For the first case, the window covers the preamble data only ($I_P$). Thus, (2.42) and (2.43) can be simplified to

$$\phi^\theta = \sum_{p=0}^{Q-1}\sum_{n=\theta}^{\theta+N-1} x_p(n)x_p^*(n) + w_p(n)w_p^*(n)$$
$$+ 2\text{Re}\{x_p(n)w_p^*(n)\exp(j2\pi\varepsilon\frac{pN + n}{N})\}, \tag{5.16}$$

98

and

$$\xi^{\theta} = \sum_{p=0}^{Q-2}\sum_{q>p}^{Q-1}\sum_{n=\theta}^{\theta+N-1} x_p(n)x_q^*(n)$$
$$+x_p(n)w_q^*(n)\exp(j2\pi\varepsilon\frac{qN+n}{N})$$
$$+w_p(n)x_q^*(n)\exp(-j2\pi\varepsilon\frac{pN+n}{N})$$
$$+w_p(n)w_q^*(n)\exp(j2\pi\varepsilon(q-p)). \tag{5.17}$$

The mean values of $\phi^i$ and $\xi^i$ for the first case are then

$$\mu_{\phi,1}^{\theta} = QN(\sigma_x^2 + \sigma_w^2), \tag{5.18}$$
$$\mu_{\xi,1}^{\theta} = \frac{QN(Q-1)}{2}\sigma_x^2. \tag{5.19}$$

The corresponding variance values are

$$\nu_{\phi,1}^{\theta} = 2QN(\sigma_x^2\sigma_w^2), \tag{5.20}$$
$$\nu_{\xi,1}^{\theta} = \frac{QN(Q-1)}{2}\sigma_w^4$$
$$+\sigma_x^2\sigma_w^2\frac{QN(Q-1)(2Q-1)}{3}. \tag{5.21}$$

The corresponding covariance value is

$$\kappa_{\phi\xi,1}^{\theta} = QN(Q-1)(\sigma_x^2\sigma_w^2). \tag{5.22}$$

Here, $\kappa^i_{\phi\xi,j}$ denotes $\kappa^i_{\phi\xi}$ in the $j$th case discussed. For the second case, the window covers the sets $I_N$ and $I_P$. Thus, $\phi^i$ and $\xi^i$ can be expressed as

$$
\begin{aligned}
\phi^i &= \sum_{n=i}^{\theta-1} w_0(n)w_0^*(n) \\
&+ \sum_{n=\theta}^{i+N-1} x_0(n)x_0^*(n) + w_0(n)w_0^*(n) \\
&+ 2\mathrm{Re}\{x_0(n)w_0^*(n)\exp(j2\pi\varepsilon\frac{n}{N})\} \\
&+ \sum_{p=1}^{Q-1}\sum_{n=i}^{i+N-1} x_p(n)x_p^*(n) + w_p(n)w_p^*(n) \\
&+ 2\mathrm{Re}\{x_p(n)w_p^*(n)\exp(j2\pi\varepsilon\frac{pN+n}{N})\}
\end{aligned}
\tag{5.23}
$$

and

$$
\begin{aligned}
\xi^i &= \sum_{q>0}^{Q-1}\sum_{n=i}^{\theta-1} w_0(n)x_q^*(n)\exp(-j2\pi\varepsilon\frac{n}{N}) + w_0(n)w_q^*(n)\exp(j2\pi\varepsilon q) \\
&+ \sum_{q>0}^{Q-1}\sum_{n=\theta}^{N+i-1} x_0(n)x_q^*(n) + x_0(n)w_q^*(n)\exp(j2\pi\varepsilon\frac{qN+n}{N}) \\
&+ w_0(n)x_q^*(n)\exp(-j2\pi\varepsilon\frac{n}{N}) + w_0(n)w_q^*(n)\exp(j2\pi\varepsilon q) \\
&+ \sum_{p=1}^{Q-2}\sum_{q>p}^{Q-1}\sum_{n=i}^{i+N-1} x_p(n)x_q^*(n) + x_p(n)w_q^*(n)\exp(j2\pi\varepsilon\frac{qN+n}{N}) \\
&+ w_p(n)x_q^*(n)\exp(-j2\pi\varepsilon\frac{pN+n}{N}) + w_p(n)w_q^*(n)\exp(j2\pi\varepsilon(q-p)).
\end{aligned}
\tag{5.24}
$$

Their mean values are

$$
\begin{aligned}
\mu^i_{\phi,2} &= (QN+i-\theta)(\sigma_x^2+\sigma_w^2) + (i-\theta)\sigma_x^2, \tag{5.25} \\
\mu^i_{\xi,2} &= (Q-1)(N+i-\theta)\sigma_x^2 + \frac{(Q-1)(Q-2)}{2}N\sigma_x^2, \tag{5.26}
\end{aligned}
$$

100

the corresponding variance values are

$$
\begin{aligned}
\nu_{\phi,2}^{i} &= 2\sigma_x^2\sigma_w^2[QN + i - \theta], && (5.27) \\
\nu_{\xi,2}^{i} &= \frac{QN(Q-1)}{2}\sigma_w^4 + \sigma_x^2\sigma_w^2 N(Q-1) \\
& \quad \cdot [(Q-1)(2 + \frac{i-\theta}{N}) + \frac{(Q-2)(2Q-3)}{3}], && (5.28)
\end{aligned}
$$

and the covariance value is

$$
\kappa_{\phi\xi,2}^{i} = (QN + i - \theta)(Q-1)(\sigma_x^2\sigma_w^2). \tag{5.29}
$$

For the third case, the window covers sets $I_P$ and $I_D$, and we write $\phi^i$ and $\xi^i$ as

$$
\begin{aligned}
\phi^i \quad &= \sum_{p=0}^{Q-2}\sum_{n=i}^{i+N-1}\{x_p(n)x_p^*(n) + w_p(n)w_p^*(n) \\
&+ \quad 2\mathrm{Re}[x_p(n)w_p^*(n)\exp(j2\pi\varepsilon\tfrac{pN+n}{N})]\} \\
&+ \quad \sum_{n=i}^{\theta+N-1}\{x_{Q-1}(n)x_{Q-1}^*(n) + w_{Q-1}(n)w_{Q-1}^*(n) \\
&+ \quad 2\mathrm{Re}[x_{Q-1}(n)w_{Q-1}^*(n)\exp(j2\pi\varepsilon\tfrac{(Q-1)N+n}{N})]\} \\
&+ \quad \sum_{n=\theta+N}^{i+N-1}\{x_{Q-1}(n)x_{Q-1}^*(n) + w_{Q-1}(n)w_{Q-1}^*(n) \\
&+ \quad 2\mathrm{Re}[x_{Q-1}(n)w_{Q-1}^*(n)\exp(j2\pi\varepsilon\tfrac{(Q-1)N+n}{N})]\} \tag{5.30}
\end{aligned}
$$

and

$$
\begin{aligned}
\xi^i \;=\; & \sum_{p=0}^{Q-3}\sum_{q>p}^{Q-2}\sum_{n=i}^{i+N-1} x_p(n)x_q^*(n) + x_p(n)w_q^*(n)\exp(j2\pi\varepsilon\frac{qN+n}{N}) \\
& + w_p(n)x_q^*(n)\exp(-j2\pi\varepsilon\frac{pN+n}{N}) + w_p(n)w_q^*(n)\exp(j2\pi\varepsilon(q-p)) \\
& + \sum_{p=0}^{Q-2}\sum_{n=i}^{\theta+N-1} x_p(n)x_{Q-1}^*(n) \\
& + x_p(n)w_{Q-1}^*(n)\exp(j2\pi\varepsilon\frac{(Q-1)N+n}{N}) \\
& + w_p(n)x_{Q-1}^*(n)\exp(-j2\pi\varepsilon\frac{pN+n}{N}) \\
& + w_p(n)w_{Q-1}^*(n)\exp(j2\pi\varepsilon(Q-p-1)) \\
& + \sum_{p=0}^{Q-2}\sum_{n=\theta+N}^{i+N-1} x_p(n)x_{Q-1}^*(n) \\
& + x_p(n)w_{Q-1}^*(n)\exp(j2\pi\varepsilon\frac{(Q-1)N+n}{N}) \\
& + w_p(n)x_{Q-1}^*(n)\exp(-j2\pi\varepsilon\frac{pN+n}{N}) \\
& + w_p(n)w_{Q-1}^*(n)\exp(j2\pi\varepsilon(Q-p-1)).
\end{aligned} \tag{5.31}
$$

Thus, the mean values are

$$
\begin{aligned}
\mu_{\phi,3}^i \;=\;& (QN+\theta-i)(\sigma_x^2+\sigma_w^2) \\
& +(i-\theta)(\sigma_d^2+\sigma_w^2), 
\end{aligned} \tag{5.32}
$$

$$
\begin{aligned}
\mu_{\xi,3}^i \;=\;& (Q-1)(N-i+\theta)\sigma_x^2 \\
& +\frac{(Q-1)(Q-2)}{2}N\sigma_x^2, 
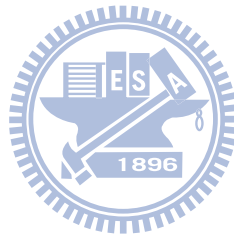\end{aligned} \tag{5.33}
$$

the variance values are

$$
\nu_{\phi,3}^i \;=\; 2\sigma_x^2\sigma_w^2(QN-i+\theta) + 2(i-\theta)\sigma_d^2\sigma_w^2, \tag{5.34}
$$

$$
\begin{aligned}
\nu_{\xi,3}^i \;=\;& \sigma_x^2\sigma_w^2 N(Q-1)[(2Q-3)(1+\frac{\theta-i}{N}) \\
& +\frac{(Q-2)(2Q-3)}{3}+1] + \frac{QN(Q-1)}{2}\sigma_w^4 \\
& +(Q-1)^2(i-\theta)(\sigma_d^2+\sigma_w^2), 
\end{aligned} \tag{5.35}
$$

and the covariance value is

$$\kappa^i_{\phi\xi,3} = (Q-1)\sigma_x^2\sigma_w^2[QN + 2(\theta - i)]. \qquad (5.36)$$

# Appendix B

## § B.1   Derivation of (3.22)

Taking the derivative of $(\mathbf{R}_y)_{p,q}$ with respect to $\varepsilon_{e,i}$, we have the result as

$$(\frac{\partial}{\partial \varepsilon_{e,i}} \mathbf{R}_y)_{p,q} = \frac{j2\pi N \sigma_x^2}{N_s}(p-q)w^{(\varepsilon_{e,i})N(p-q)}. \tag{5.37}$$

Then the first term in the right-hand-side of (3.9) can be derived as

$$
\begin{aligned}
(\mathbf{R}_y^{-1}\frac{\partial}{\partial \varepsilon_{e,i}} \mathbf{R}_y)_{p,q} &= \sum_{k=1}^{Q}(\mathbf{R}_y^{-1})_{p,k}\frac{j2\pi N \sigma_x^2}{N_s}(k-q)w^{(\varepsilon_{e,i})N(k-q)} \\
&= \frac{j2\pi N \sigma_x^2}{N_s}\{\sigma_\eta^{-2}w^{(\varepsilon_{e,i})N(p-q)} - C_0 \sum_{k=1}^{Q}(k-q)x^{k-q}\Gamma(p,k) \\
&\quad -C_1 \sum_{k=1}^{Q}\sum_{b=1}^{Q}(k-q)x^{k-q}\Gamma(p,b)\Gamma(b,k) - C_2 \sum_{k=1}^{Q}(k-q)x^{k-q} \\
&\quad \cdot \sum_{a=1}^{Q}\sum_{b=1}^{Q}\Gamma(p,a)\Gamma(a,b)\Gamma(b,k)\},
\end{aligned}
\tag{5.38}
$$

where

$$x = \exp(j2\pi\frac{N\varepsilon_{e,i}}{N_s}). \tag{5.39}$$

In order to obtain the second term in (3.9), the derivative of the third and fourth terms in the right-hand-side of (3.17) can be first found as

$$
\begin{aligned}
\frac{\partial}{\partial \varepsilon_{e,i}}\{\sum_{n=1}^{Q}\Gamma(n,q)\Gamma(p,n)\} &= \frac{j2\pi N}{N_s}[\sum_{n=1}^{Q}(n-q)\Gamma(p,n)w^{(\varepsilon_{e,i})N(n-q)} \\
&\quad +(p-n)\Gamma(n,q)w^{(\varepsilon_{e,i})N(p-n)}]
\end{aligned}
\tag{5.40}
$$

and

$$\frac{\partial}{\partial \varepsilon_{e,i}} \quad \{\sum_{m=1}^{Q} \sum_{n=1}^{Q} \Gamma(m,n)\Gamma(n,q)\Gamma(p,m)\}$$

$$= \frac{j2\pi N}{N_s}[\sum_{m=1}^{Q} \Gamma(p,m) \sum_{n=1}^{Q} \Gamma(n,q)$$

$$\cdot (m-n)w^{(\varepsilon_{e,i})N(m-n)}$$

$$+(n-q)\Gamma(p,m)\Gamma(m,n)w^{(\varepsilon_{e,i})N(n-q)}$$

$$+(p-m)\Gamma(m,n)\Gamma(n,q)w^{(\varepsilon_{e,i})N(p-m)}]. \tag{5.41}$$

Thus, we can have the second term in (3.9) as

$$\frac{\partial}{\partial \varepsilon_{e,i}}(\mathbf{R}_y^{-1})_{p,q} = \frac{j2\pi N}{N_s}\{-C_0(p-q)w^{(\varepsilon_{e,i})N(p-q)}$$

$$-C_1 \sum_{n=1}^{Q}(n-q)w^{(\varepsilon_{e,i})N(n-q)}\Gamma(p,n) - C_1 \sum_{n=1}^{Q}(p-n)w^{(\varepsilon_{e,i})N(p-n)}\Gamma(n,q)$$

$$-C_2 \sum_{m=1}^{Q} \sum_{n=1}^{Q}[(m-n)w^{(\varepsilon_{e,i})N(m-n)}\Gamma(p,m)\Gamma(n,q)$$

$$+(n-q)w^{(\varepsilon_{e,i})N(n-q)}\Gamma(p,m)\Gamma(m,n)$$

$$+(p-m)w^{(\varepsilon_{e,i})N(p-m)}\Gamma(m,n)\Gamma(n,q)]\}. \tag{5.42}$$

Substituting (3.3), (3.5), (5.38), and (5.42) into (3.9), we can rewrite the log-likelihood function
as

$$\frac{\partial}{\partial \varepsilon_{e,i}}\Lambda(\boldsymbol{\varepsilon}) = \frac{j2\pi N}{N_s}\{N\sigma_x^2 \sum_{p=1}^{Q} \sum_{q=1}^{Q}(q-p)x^{q-p}[C_0\Gamma(p,q)$$

$$+C_1 \sum_{n=1}^{Q} \Gamma(p,n)\Gamma(n,q)$$

$$+C_2 \sum_{m=1}^{Q} \sum_{n=1}^{Q} \Gamma(p,m)\Gamma(m,n)\Gamma(n,q)]$$

106

$$+\gamma(p,q)[C_0(p-q)x^{p-q}$$

$$+C_1\sum_{n=1}^{Q}((n-q)x^{n-q}\Gamma(p,n)$$

$$+(p-n)x^{p-n}\Gamma(n,q))$$

$$+C_2\sum_{m=1}^{Q}\sum_{n=1}^{Q}((m-n)x^{m-n}\Gamma(p,m)\Gamma(n,q)$$

$$+(n-q)x^{n-q}\Gamma(p,m)\Gamma(m,n)$$

$$+(p-m)x^{p-m}\Gamma(m,n)\Gamma(n,q))]\}. \tag{5.43}$$
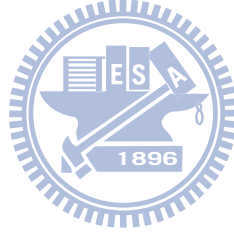
# § B.2   Derivation of (3.27)

Rewrite (3.22) as

$$
\begin{aligned}
\frac{\partial}{\partial \varepsilon_{e,i}} \Lambda(\varepsilon) &= \sum_{p=1}^{Q}\sum_{q=1}^{Q}\{N\sigma_x^2(q-p)x^{q-p}[C_0\Gamma(p,q) + C_1\sum_{n=1}^{Q}\Gamma(p,n)\Gamma(n,q) \\
&\quad + C_2\sum_{m=1}^{Q}\sum_{n=1}^{Q}\Gamma(p,m)\Gamma(m,n)\Gamma(n,q)] + \gamma(p,q)[C_0(p-q)x^{p-q} \\
&\quad + C_1\sum_{n=1}^{Q}((n-q)x^{n-q}\Gamma(p,n) + (p-n)x^{p-n}\Gamma(n,q)) \\
&\quad + C_2\sum_{m=1}^{Q}\sum_{n=1}^{Q}((m-n)x^{m-n}\Gamma(p,m)\Gamma(n,q) \\
&\quad + (n-q)x^{n-q}\Gamma(p,m)\Gamma(m,n) \\
&\quad + (p-m)x^{p-m}\Gamma(m,n)\Gamma(n,q))]\} \\
&= \sum_{p=1}^{Q}\sum_{q=1}^{Q}F_1(p,q)x^{q-p} + \sum_{p=1}^{Q}\sum_{q=1}^{Q}F_2(p,q)x^{p-q} \\
&\quad + \sum_{p=1}^{Q}\sum_{q=1}^{Q}\{\gamma(p,q)[C_1\sum_{n=1}^{Q}((n-q)x^{n-q}\Gamma(p,n) \\
&\quad + (p-n)x^{p-n}\Gamma(n,q)) \\
&\quad + C_2\sum_{m=1}^{Q}\sum_{n=1}^{Q}((m-n)x^{m-n}\Gamma(p,m)\Gamma(n,q) \\
&\quad + (n-q)x^{n-q}\Gamma(p,m)\Gamma(m,n) \\
&\quad + (p-m)x^{p-m}\Gamma(m,n)\Gamma(n,q))]\} \\
&= 0, \tag{5.44}
\end{aligned}
$$

where $F_1(p,q) = N\sigma_x^2(q-p)[C_0\Gamma(p,q) + C_1\sum_{n=1}^{Q}\Gamma(p,n)\Gamma(n,q) + C_2\sum_{m=1}^{Q}\sum_{n=1}^{Q}\Gamma(p,m)\Gamma(m,n)$ $\Gamma(n,q)]$ and $F_2(p,1) = C_0(p-q)\gamma(p,q)$. Conducting some variable transformation, we can express the power of $x$ as a single variable of $k$. We can then collect all the items with positive

$k$ into one expression and have

$$\{\frac{\partial}{\partial \varepsilon_{e,i}}\Lambda(\boldsymbol{\varepsilon})\}_+ = \sum_{k=1}^{Q-1}\{\sum_{p=1,q=p+k}^{Q-k}F_1(p,q)x^k + \sum_{q=1,p=q+k}^{Q-k}F_2(p,q)x^k$$

$$+C_1\sum_{p=1}^{Q}\sum_{q=1,n=q+k}^{Q-k}\gamma(p,q)\Gamma(p,n)kx^k$$

$$+C_1\sum_{q=1}^{Q}\sum_{n=1,p=n+k}^{Q-k}\gamma(p,q)\Gamma(n,q)kx^k$$

$$+C_2\sum_{p=1}^{Q}\sum_{q=1}^{Q}\sum_{n=1,m=n+k}^{Q-k}\gamma(p,q)\Gamma(p,m)\Gamma(n,q)kx^k$$

$$+C_2\sum_{p=1}^{Q}\sum_{m=1}^{Q}\sum_{q=1,n=q+k}^{Q-k}\gamma(p,q)\Gamma(p,m)\Gamma(m,n)kx^k$$

$$+C_2\sum_{n=1}^{Q}\sum_{q=1}^{Q}\sum_{m=1,p=m+k}^{Q-k}\gamma(p,q)\Gamma(m,n)\Gamma(n,q)kx^k\}$$

$$= \sum_{k=1}^{Q-1}\alpha_p(k)x^k, \tag{5.45}$$

Similarly, we can collect all the items with negative $k$ into another expression and have

$$\{\frac{\partial}{\partial \varepsilon_{e,i}}\Lambda(\boldsymbol{\varepsilon})\}_- = \sum_{k=-1}^{1-Q}\{\sum_{p=1-k,q=p+k}^{Q}F_1(p,q)x^k + \sum_{q=1-k,p=q+k}^{Q}F_2(p,q)x^k$$

$$+C_1\sum_{p=1}^{Q}\sum_{q=1-k,n=q+k}^{Q}\gamma(p,q)\Gamma(p,n)kx^k$$

$$+C_1\sum_{q=1}^{Q}\sum_{n=1-k,p=n+k}^{Q}\gamma(p,q)\Gamma(n,q)kx^k$$

$$+C_2\sum_{p=1}^{Q}\sum_{q=1}^{Q}\sum_{n=1-k,m=n+k}^{Q}\gamma(p,q)\Gamma(p,m)\Gamma(n,q)kx^k$$

$$+C_2\sum_{p=1}^{Q}\sum_{m=1}^{Q}\sum_{q=1-k,n=q+k}^{Q}\gamma(p,q)\Gamma(p,m)\Gamma(m,n)kx^k$$

$$+C_2\sum_{n=1}^{Q}\sum_{q=1}^{Q}\sum_{m=1-k,p=m+k}^{Q}\gamma(p,q)\Gamma(m,n)\Gamma(n,q)kx^k\}.$$

Using some variable transformations, we can obtain a similar result with (5.45) :

$$
\begin{aligned}
\{\frac{\partial}{\partial \varepsilon_{e,i}} \Lambda(\varepsilon)\}_{-} &= \sum_{k=1}^{Q-1} \{ \sum_{q=1,p=q+k}^{Q-k} F_1(p,q)x^{-k} + \sum_{p=1,q=p+k}^{Q-k} F_2(p,q)x^{-k} \\
&\quad -C_1 \sum_{p=1}^{Q} \sum_{n=1,q=n+k}^{Q-k} \gamma(p,q)\Gamma(p,n)kx^{-k} \\
&\quad -C_1 \sum_{q=1}^{Q} \sum_{p=1,n=p+k}^{Q-k} \gamma(p,q)\Gamma(n,q)kx^{-k} \\
&\quad -C_2 \sum_{p=1}^{Q} \sum_{q=1}^{Q} \sum_{m=1,n=m+k}^{Q-k} \gamma(p,q)\Gamma(p,m)\Gamma(n,q)kx^{-k} \\
&\quad -C_2 \sum_{p=1}^{Q} \sum_{m=1}^{Q} \sum_{n=1,q=n+k}^{Q-k} \gamma(p,q)\Gamma(p,m)\Gamma(m,n)kx^{-k} \\
&\quad -C_2 \sum_{n=1}^{Q} \sum_{q=1}^{Q} \sum_{p=1,m=p+k}^{Q-k} \gamma(p,q)\Gamma(m,n)\Gamma(n,q)kx^{-k}\} \\
&= \sum_{k=1}^{Q-1} \alpha_n(k)x^{-k}.
\end{aligned}
\tag{5.46}
$$

So, the derivative of the logarithm likelihood function (3.22) can then be re-expressed as

$$
\frac{\partial}{\partial \varepsilon_{e,i}} \Lambda(\varepsilon) = \sum_{k=1}^{Q-1} \alpha_p(k)x^k + \sum_{k=1}^{Q-1} \alpha_n(k)x^{-k}.
\tag{5.47}
$$

# Appendix C

## § C.1   Detailed Derivation of (4.37)

The conditional Fisher information in (4.37) is a product of four $N_s \times N_s$ matrices, i.e. $\mathbf{V} = \mathbf{R}_{y|h}^{-1} \frac{\partial \mathbf{R}_{y|h}}{\partial \varepsilon_{SD}} \mathbf{R}_{y|h}^{-1} \frac{\partial \mathbf{R}_{y|h}}{\partial \varepsilon_{SD}}$. From (4.36), $\gamma'(1) \approx 0$ when $N_s$ is large. Using the property, we can see that the off-diagonal terms of $\mathbf{R}'_{y|h}$ in (4.34) are zeroes, simplifying the derivation of $\mathbf{V}$. Note that $\mathbf{V}$ contains $N^2$ $Q \times Q$ submatrices. Denoting the $(p,q)$th submatrix of $\mathbf{V}$ as $\mathbf{V}_{p,q}$, we can have

$$\mathbf{V}_{p,q} = \sum_{k=1}^{N} \mathbf{T}_{p,k} \mathbf{R}'_1 \mathbf{T}_{k,q} \mathbf{R}'_1, \tag{5.48}$$

where $\mathbf{T}_{p,q}$ is the $(p,q)$th submatrix of $\mathbf{R}_{y|h}^{-1}$ as defined in (4.31). The trace of $\mathbf{V}$ can then be expressed as

$$
\begin{aligned}
\mathrm{tr}(\mathbf{V}) &= \mathrm{tr}\left( \sum_{p=1}^{N} \sum_{q=1}^{N} \mathbf{T}_{p,q} \mathbf{R}'_1 \mathbf{T}_{q,p} \mathbf{R}'_1 \right) \\
&= \mathrm{tr}\left( \sum_{p=q=1}^{N} \mathbf{T}_{p,p} \mathbf{R}'_1 \mathbf{T}_{p,p} \mathbf{R}'_1 + \sum_{p=1}^{N} \sum_{q>p}^{N} \mathbf{T}_{p,q} \mathbf{R}'_1 \mathbf{T}_{q,p} \mathbf{R}'_1 + \sum_{q=1}^{N} \sum_{p>q}^{N} \mathbf{T}_{p,q} \mathbf{R}'_1 \mathbf{T}_{q,p} \mathbf{R}'_1 \right) (5.49)
\end{aligned}
$$

Substituting (4.31) into (5.49), we can have

$$\mathrm{tr}(\mathbf{V}) = \mathrm{tr}\left( 2 \sum_{p=1}^{N-1} \sum_{q>p}^{N} \mathbf{W}_{p,q} + \sum_{p=1}^{N} \mathbf{W}_{p,p} \right) \tag{5.50}$$

where $\mathbf{W}_{p,q} = (|\gamma(1)|^2)^{q-p} (\mathbf{P}_{p-1} \mathbf{Q}_{q+1} \mathbf{P}_N^{-1} \mathbf{R}'_1)^2$ and (4.37) results.

# Bibliography

[1] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*, Boston, MA:Artech House, 2000.

[2] J. Heiskala and J. Terry, *OFDM wireless LANs: A Theoretical and Practical Guide*, Indianapolis, IN:Sams, 2002.

[3] Amitava Ghosh, Rapeepat Ratasuk, Bishwarup Mondal, Nitin Mangalvedhe, and Tim Thomas, Motorola Inc, "LTE-Advanced: Next-Generation Wireless Broadband Technology," *IEEE Trans. Wireless Commun.*, Vol. 17, Issue 3, June 2010.

[4] W. D. Warner and C. Leung, "OFDM/FM fram synchroniztion for mobile radio data communication," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 302-313, May 1993.

[5] S. Y. Park and C. G. Kang, "Performance of pilot-assisted channel estimation for OFDM system under time-varying multi-path Rayleigh fading with frequency offset compensation," in Proc. *IEEE Veh. Technol. Conf.*, vol. 2, pp. 1245-1249, May 2000.

[6] Seung Young Park , Bo Seok Seo , Chung Gu Kang, "Effects of frequency offset compensation error on channel estimation for OFDM system under mobile radio channels," *Signal Process.*, vol. 83 no.12, pp.2621-2630, December 2003.

[7] Koivisto T. and Koivunen V., "Impact of time and frequency offsets on cooperative multiuser MIMO-OFDM systems," *IEEE PIMRC 2009*, Tokyo, Japan, pp. 3119 - 3123, 2009.

[8] Luca Rugini and , Polo Banelli. "BER of OFDM systems imparied by carrier frequency offset in multipath fading channels," *IEEE Trans. Wireless commun.*, vol. 4, no. 5, pp. 2279-2288, Sep. 2008.

[9] P.C. Weeraddana, N. Rajatheva, and H. Minn, "Probability of error analysis of BPSK OFDM systems with random residual frequency offset," *IEEE Trans. Commun.*, vol. 57, No. 1, pp. 106-116, Jan. 2009.

[10] J.J. van de Beek, M. Sandell, and P.O.Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Process.*, vol. 45, pp.1800-1805, July 1997.

[11] Navid Lashkarian,"Class of cyclic-based estimators for frequency-offset estimation of OFDM Systems," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2139-2149, Dec. 2000.

[12] Yik-Chung Wu, E. Serpedin, "Comments on "Class of cyclic-based estimators for frequency-offset estimation of OFDM systems"," *IEEE Trans. Commun.* vol. 53, no. 3, pp.413-414, March 2005.

[13] M. Tanda, "Blind symbol-timing and frequency-offset estimation in OFDM systems with real data symbols," *IEEE Trans. Commun.* vol. 52, no. 10, pp. 1609-1612, Oct. 2004.

[14] H. Meyr, M. Moeneclaey and S. A. Fechtel, *Digital Communication Receivers-Synchronization , Channel Estimation, and Signal Processing*, New York, NY:Wiley, 1998.

[15] E. Chiavaccini and G. M. Vitetta, "Maximum-likelihood frequency recovery for OFDM signals transmitted over multipath fading channels," *IEEE Trans. Commun.*, vol.52, no. 2, pp.244-251, Feb. 2004.

[16] D. Lee and K. Cheun, "Coarse symbol synchronization algorithms for OFDM systems in multipath channels," *IEEE Commun. Lett.*, vol. 6, no. 1, pp. 446-448, Oct. 2002.

[17] Simon Haykin, *Adaptive Filter Theory 4E*, Prentice-Hall, 2002

[18] M. Morelli, A. N. D'Andrea, and U. Mengali, "Frequency ambiguity resolution in OFDM systems," *IEEE Commun. Lett.*, vol. 4, pp. 134-136, April 2000.

[19] Kyung-Taek Lee and Jong-Soo Seo, "Pilot-aided iterative frequency offset estimation for digital video broadcasting (DVB) systems," *IEEE Trans. Consum. Electron.*, vol. 53, no. 1, pp. 11-16, Feb. 2007.

[20] Paul H. Moose, "A technique of orthogonal frequency division multiplexing frequency offset correlation," *IEEE Trans. Commun.* vol. 42, No. 10, pp. 2908-2914, October 1994.

[21] Defeng (David) Huang, and Khaled Ben Letaief, "Carrier frequency offset estimation for OFDM systems using null subcarriers," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 813-823, May 2006.

[22] T.M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613-1621, Dec. 1997.

[23] M. Morelli and U. Mengali, "An improved frequency offset estimator for OFDM applications," *IEEE Commun. Lett.*, vol. 3, pp. 75-77, Mar. 1999.

[24] H. Minn, P. Tarasak, and V. K. Bhargava, "OFDM frequency offset estimation methods based on BLUE principle," in *Proc. IEEE Veh. Technol. Conf.*, Vancouver, BC, Canada, pp. 1230-1234, Sept. 2002.

[25] J. Li, G. Liu and G. B. Giannakis, "Carrier frequency offset estimation for OFDM-based WLANs," *IEEE Signal Processing Lett.*, vol. 8, no. 3, pp.80-82, Mar. 2001.

[26] Jiun H. Yu, Yu T. Su, "Pilot-assisted maximum likelihood frequency-Offset estimation for OFDM systems," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1997-2008, Nov. 2004.

[27] M. H. Cheng, C. C. Chou, "Maximum-likelihood estimation of frequency and time offsets in OFDM systems with multiple sets of identical data," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2848-2852, July 2006.

[28] J. Chen, Y. C. Wu, S. C. Chan and T. S. Ng, "Joint maximum-likelihood CFO and channel estimation for OFDMA uplink using importance sampling, " *IEEE Trans. Veh. technol.*, vol. 57, No. 6, pp. 3462-3470, Nov. 2008.

[29] M. O. Pun, M. Morelli, and C. C. Jay Kuo, "Maximum-likelihood synchronization and channel estimation for OFDMA uplink transmissions, " *IEEE Trans. Comm.*, vol. 54, No. 4, pp. 726-736, April 2006.

[30] Z. Wang, Y. Xin, and G. Mathew., "Iterative Carrier-Frequency Offset Estimation for Generalized OFDMA Uplink Transmission," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, Mar. 2009.

[31] S. Sezginer and P. Bianchi, "Asymptotically efficient reduced complexity frequency offset and channel estimators for uplink MIMO-OFDMA systems," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp.964-979, Mar. 2008.

[32] J. J. van de Beek, P. O. Borjesson, M. L. Boucheret, D. Landstrom, J. M. Arenas, O. Odling, M. Wahlqvist, and S. K. Wilson, "A time and frequency synchronization scheme for multiuser OFDM," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1900-1914, Nov. 1999.

[33] S. Barbarossa, M. Pompili, and G. B. Giannakis, "Channel-independent synchronization of orthogonal frequency division multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 2, pp. 474-486, Feb. 2002.

[34] Z. Cao, U. Tureli, and Y. D. Yao, "Deterministic multiuser carrierfrequency offset estimation for interleaved OFDMA uplink," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1585-1594, Sep. 2004.

[35] J. Lee, S. Lee, K. J. Bang, S. Cha, and D. Hong, "Carrier frequency offset estimation using ESPRIT for interved OFDMA uplink systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 3227-3231, Sep. 2007.

[36] Z. Li, and Xiang-Gen Xia, "An Alamouti coded OFDM transmission for cooperative systems robust to both timing errors and frequency offsets," *IEEE Trans. Wireless Commun.*, vol. 7, No. 5, pp. 1839-1844, May, 2008.

[37] Q. Huang, J. Wei, and P. Ciblat, "Practical Timing and Frequency Synchronization for OFDM-Based Cooperative Systems ," *IEEE Trans. Signal Process.*, vol. 58, No. 7, pp. 3706-3716, July, 2010.

[38] Z. Zhang, W. Zhang, and C. Tellambura, "OFDMA uplink frequency offset estimation via cooperative relaying ," *IEEE Trans. Wireless Commun.*, vol. 8, No. 9, pp. 4450-4456, May, 2009.

[39] Y. S. Choi, Voltz, P.J., Cassara, F.A., "ML estimation of carrier frequency offset for multicarrier signals in Rayleigh fading channels, " *IEEE Trans. Veh. Technol.*, vol. 50, No. 2, pp. 644-655, March 2001.

[40] Steven M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory*, Prentice Hall, c1993.

[41] H.Vincent Poor, *An Introduction to Signal Detection and Estimation 2E*, Springer, 1994.

[42] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720-741, May 1989.

[43] Y. Jian, H. Minn, X. Gao, X. You, and Y. Li, "Frequency offset estimation and training sequence design for MIMO OFDM," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1244-1254, Apr. 2008.

[44] S. Tretter, "Estimating the frequency of a noisy sinusoid by linear regression," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 832-835, Nov. 1985.

[45] William S. McCormick and , James L. L. "Efficient parallel rooting of complex polynomials on the unit circle," *IEEE Trans. Signal Processing*, vol. 39, No. 10, pp.2347-2351, Oct. 1991.

[46] T. J. Lv and J. Chen,"ML estimation of timing and frequency offset using multiple OFDM symbols in OFDM systems," in *IEEE Global Telecommunications Conference*, vol. 4, Dec. 2003, pp. 2280-2284.

[47] ETSI, BRAN; HIPERLAN Type 2; Physical (PHY) Layer Specification, 2001. Tech. Spec. 101 475. [Online]. Available: http://www.etsi.org

[48] S. Kay and S. Saha, "Mean likelihood frequency estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 1937-1946, Jul. 2000.

[49] Carl D. Meyer, *Matrix Analysis and Applied Linear Algebra* . Society for Industrial and Applied Mathematics (SIAM), 2000.

[50] K. Berberidis, S. Rantos, and J. Palicot, "A step-by-step Quasi-Newton algorithm in the frequency doamin and its application to adaptive channel equalization," *IEEE Trans. Comm.*, vol. 52, No. 9, pp. 1585-1594, Sep. 2004.

[51] Harry L. Van Trees, *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing*, John Wiley, 2002.

[52] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory,* Princeton, 2005.

[53] D. Kincaid and W. Cheney, *Mathematics of Scientific Computing*, 2nd ed. Pacific Grove, CA: Brooks/Cole, 1996.

[54] Ali H. Sayed, *Fundamentals of Adaptive Filtering*, John Wiley, 2003.

[55] Man-On Pun, Michele Morelli, and C.-C. Jay Kuo, "Iterative Detection and Frequency Synchronization for OFDMA Uplink Transmissions," *IEEE Trans. Wireless Commun.*, vol. 6, No. 2, pp. 629-639, Feb. 2007

[56] M. Guenach, F. Simoens, H. Wymeersch, and M. Moeneclaey, "Uplink acquisition of a new user accessing a fixed wireless DS-CDMA system," in *Proc. of 6th Workshop on Signal Processing Advances in Wireless Communications (SPAWC2005), IEEE*, New York City, NY, June 2005, pp. 353-357.

[57] A. Nosratinia, E. Hunter, A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, pp. 74-80, October 2004.

[58] Huawei, R1-090822 "System performance evaluation of downlink CoMP".

[59] Y. Hu, Y. Jiang, and X. You, "SNR Degradation due to Carrier Frequency Offset in Amplify-and-Forward Relay System for Fading Channels," *IEEE Veh. Technol. Conf.*, VTC-Fall, 2009.

[60] J. Chen, Y.-C. Wu, S. Ma, and T.-S. Ng, "Joint CFO and channel estimation for multiuser MIMO-OFDM systems with optimal training sequences," *IEEE Trans. Signal Process.*, vol. 56, pp. 4008-4019, Aug. 2008.

[61] C.M. Da Fonseca, "On the eigenvalues of some tridiagonal matrices", J. Comput. Appl. Math., Vol. 200, pp. 283-286, 2007.

[62] R. Usmani, "Inversion of a tridiagonal Jacobi matrix," Linear Algebra Appl. 212/213 (1994) pp.413-414.

[63] D.P. Palomar, and S. Verdú, "Gradient of mutual information in linear vector gaussian channels," *IEEE Trans. Inf. Theory*, Vol. 52, No. 1, pp. 141-154 , Jan. 2006.

[64] J. Medbo and P. Schramm, "Channel Models for HIPERLAN/2," ETSI/BRAN doc. No. 3ERI085B, 1998

[65] A. Hjørungnes, and D, Gesbert "Complex-Valued Matrix Differentiation: Techniques and Key Results," *IEEE Trans. Signal Processing*, vol. 55, No. 6, pp.2740-2746, June 2007.

[66] R. Mo, and Y. H. Chew, "MMSE-based joint source and relay precoding design for amplify-and-forward MIMO relay networks," *IEEE Trans. Wireless Comm.*, vol. 8, No. 9, pp.4668-4676, Sep. 2009.

[67] B. Wang, Z. Han, and K. J. Ray Liu, "Distributed Relay Selection and Power Control for Multiuser Cooperative Communication Networks Using Buyer/Seller Game," *Proc. IEEE INFOCOM '07*, pp.544-552, May 2007.

[68] I. Ziskind and M.Wax,"Maximum likelohood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 10, pp. 1553-1560, Oct. 1988.