2006

MIMO-OFDM communication systems: channel estimation and wireless location

Zhongshan Wu
Louisiana State University and Agricultural and Mechanical College, zwu1@lsu.edu

MIMO-OFDM COMMUNICATION SYSTEMS: CHANNEL ESTIMATION AND WIRELESS LOCATION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Electrical and Computer Engineering

by
Zhongshan Wu
B.S., Northeastern University, China, 1996
M.S., Louisiana State University, US, 2001
May 2006

To my parents.

Acknowledgments

Throughout my six years at LSU, I have many people to thank for helping to make my experience here both enriching and rewarding.

First and foremost, I wish to thank my advisor and committee chair, Dr. Guoxiang Gu. I am grateful to Dr. Gu for his offering me such an invaluable chance to study here, for his being a constant source of research ideas, insightful discussions and inspiring words in times of needs and for his unique attitude of being strict with academic research which will shape my career forever.

My heartful appreciation also goes to Dr. Kemin Zhou whose breadth of knowledge and perspectiveness have instilled in me great interest in bridging theoretical research and practical implementation. I would like to thank Dr. Shuangqing Wei for his fresh talks in his seminar and his generous sharing research resource with us.

I am deeply indebted to Dr. John M. Tyler for his taking his time to serve as my graduate committee member and his sincere encouragement. For providing me with the mathematical knowledge and skills imperative to the work in this dissertation, I would like to thank my minor professor, Dr. Peter Wolenski for his precious time.

For all my EE friends, Jianqiang He, Bin Fu, Nike Liu, Xiaobo Li, Rachinayani

Kumar Phalguna and Shuguang Hao, I cherish all the wonderful time we have together.

Through it all, I owe the greatest debt to my parents and my sisters. Especially my father, he will be living in my memory for endless time.

Zhongshan Wu

October, 2005

Contents

List of Figures

Notation and Symbols

$\mathbf{A_{M\times N}}$: M-row N-column matrix

$\mathbf{A^{-1}}$: Inverse of $\mathbf{A}$

$\mathrm{Tr}(\mathbf{A})$: Trace of $\mathbf{A}$, $\mathrm{Tr}(\mathbf{A}) = \sum_{\mathbf{i}} \mathbf{A_{ii}}$

$\mathbf{A^T}$: Transpose of $\mathbf{A}$

$\mathbf{A}^*$: Complex conjugate transpose of $\mathbf{A}$

$\mathbf{I_N}$: Identity matrix of size $N \times N$

List of Acronyms

MIMO	multiple input and multiple outut
OFDM	orthogonal frequency division multiplexing
LS	least square
MS	mobile station
TDOA	time difference of arrival
AOA	angle of arrival
WiMax	worldwide interoperability for microwave access
ML	maximum-likelihood
AWGN	additive white Gaussian noise
WMAN	wireless metropolitan area network
ICI	inter-carrier interference
ISI	inter-symbol interference
FFT	fast Fourier transform
WLAN	wireless local area network
CP	cyclic prefix
BER	bit error rate
MMSE	minimum mean squared error
GPS	global positioning system
WiFi	wireless fidelity

Abstract

In this new information age, high data rate and strong reliability features our wireless communication systems and is becoming the dominant factor for a successful deployment of commercial networks. MIMO-OFDM (multiple input multiple output-orthogonal frequency division multiplexing), a new wireless broadband technology, has gained great popularity for its capability of high rate transmission and its robustness against multi-path fading and other channel impairments.

A major challenge to MIMO-OFDM systems is how to obtain the channel state information accurately and promptly for coherent detection of information symbols and channel synchronization. In the first part, this dissertation formulates the channel estimation problem for MIMO-OFDM systems and proposes a pilot-tone based estimation algorithm. A complex equivalent baseband MIMO-OFDM signal model is presented by matrix representation. By choosing L equally-spaced and equally-powered pilot tones from N sub-carriers in one OFDM symbol, a down-sampled version of the original signal model is obtained. Furthermore, this signal model is transformed into a linear form solvable for the LS (least-square) estimation algorithm. Based on the resultant model, a simple pilot-tone design is proposed in the form of a unitary

matrix, whose rows stand for different pilot-tone sets in the frequency domain and whose columns represent distinct transmit antennas in the spatial domain. From the analysis and synthesis of the pilot-tone design in this dissertation, our estimation algorithm can reduce the computational complexity inherited in MIMO systems by the fact that the pilot-tone matrix is essentially a unitary matrix, and is proven an optimal channel estimator in the sense of achieving the minimum MSE (mean squared error) of channel estimation for a fixed power of pilot tones.

In the second part, this dissertation addresses the wireless location problem in WiMax (worldwide interoperability for microwave access) networks, which is mainly based on the MIMO-OFDM technology. From the measurement data of TDOA (time difference of arrival), AOA (angle of arrival) or a combination of those two, a quasi-linear form is formulated for an LS-type solution. It is assumed that the observation data is corrupted by a zero-mean AWGN (additive white Gaussian noise) with a very small variance. Under this assumption, the noise term in the quasi-liner form is proved to hold a normal distribution approximately. Hence the ML (maximum-likelihood) estimation and the LS-type solution are equivalent. But the ML estimation technique is not feasible here due to its computational complexity and the possible nonexistence of the optimal solution. Our proposed method is capable of estimating the MS location very accurately with a much less amount of computations. A final result of the MS (mobile station) location estimation, however, cannot be obtained directly from the LS-type solution without bringing in another independent constraint. To solve

this problem, the Lagrange multiplier is explored to find the optimal solution to the constrained LS-type optimization problem.

Chapter 1

Introduction

Wireless technologies have evolved remarkably since Guglielmo Marconi first demonstrated radio's ability to provide continuous contact with ships sailing in the English channel in 1897. New theories and applications of wireless technologies have been developed by hundreds and thousands of scientists and engineers through the world ever since. Wireless communications can be regarded as the most important development that has an extremely wide range of applications from TV remote control and cordless phones to cellular phones and satellite-based TV systems. It changed people's life style in every aspect. Especially during the last decade, the mobile radio communications industry has grown by an exponentially increasing rate, fueled by the digital and RF (radio frequency) circuits design, fabrication and integration techniques and more computing power in chips. This trend will continue with an even greater pace in the near future.

The advances and developments in the technique field have partially helped to realize our dreams on fast and reliable communicating "any time any where". But we

are expecting to have more experience in this wireless world such as wireless Internet surfing and interactive multimedia messaging so on. One natural question is: how can we put high-rate data streams over radio links to satisfy our needs? New wireless broadband access techniques are anticipated to answer this question. For example, the coming 3G (third generation) cellular technology can provide us with up to 2Mbps (bits per second) data service. But that still does not meet the data rate required by multimedia media communications like HDTV (high-definition television) and video conference. Recently MIMO-OFDM systems have gained considerable attentions from the leading industry companies and the active academic community [28, 30, 42, 50]. A collection of problems including channel measurements and modeling, channel estimation, synchronization, IQ (in phase-quadrature)imbalance and PAPR (peak-to-average power ratio) have been widely studied by researchers [48, 11, 14, 15, 13]. Clearly all the performance improvement and capacity increase are based on accurate channel state information. Channel estimation plays a significant role for MIMO-OFDM systems. For this reason, it is the first part of my dissertation to work on channel estimation of MIMO-OFDM systems.

The maturing of MIMO-OFDM technology will lead it to a much wider variety of applications. WMAN (wireless metropolitan area network) has adopted this technology. Similar to current network-based wireless location technique [53], we consider the wireless location problem on the WiMax network, which is based on MIMO-OFDM technology. The work in this area contributes to the second part of my dissertation.

1.1 Overview

OFDM [5] is becoming a very popular multi-carrier modulation technique for transmission of signals over wireless channels. It converts a frequency-selective fading channel into a collection of parallel flat fading subchannels, which greatly simplifies the structure of the receiver. The time domain waveform of the subcarriers are orthogonal (*subchannel* and *subcarrier* will be used interchangeably hereinafter), yet the signal spectral corresponding to different subcarriers overlap in frequency domain. Hence, the available bandwidth is utilized very efficiently in OFDM systems without causing the ICI (inter-carrier interference). By combining multiple low-data-rate subcarriers, OFDM systems can provide a composite high-data-rate with a long symbol duration. That helps to eliminate the ISI (inter-symbol interference), which often occurs along with signals of a short symbol duration in a multipath channel. Simply speaking, we can list its pros and cons as follows [31].

Advantage of OFDM systems are:

- High spectral efficiency;
- Simple implementation by FFT (fast Fourier transform);
- Low receiver complexity;
- Robustability for high-data-rate transmission over multipath fading channel
- High flexibility in terms of link adaptation;

- Low complexity multiple access schemes such as orthogonal frequency division multiple access.

Disadvantages of OFDM systems are:

- Sensitive to frequency offsets, timing errors and phase noise;
- Relatively higher peak-to-average power ratio compared to single carrier system, which tends to reduce the power efficiency of the RF amplifier.

1.1.1 OFDM System Model

The OFDM technology is widely used in two types of working environments, i.e., a wired environment and a wireless environment. When used to transmit signals through wires like twisted wire pairs and coaxial cables, it is usually called as DMT (digital multi-tone). For instance, DMT is the core technology for all the xDSL (digital subscriber lines) systems which provide high-speed data service via existing telephone networks. However, in a wireless environment such as radio broadcasting system and WLAN (wireless local area network), it is referred to as OFDM. Since we aim at performance enhancement for wireless communication systems, we use the term OFDM throughout this thesis. Furthermore, we only use the term MIMO-OFDM while explicitly addressing the OFDM systems combined with multiple antennas at both ends of a wireless link.

The history of OFDM can all the way date back to the mid 1960s, when Chang [2] published a paper on the synthesis of bandlimited orthogonal signals for multichannel

data transmission. He presented a new principle of transmitting signals simultaneously over a bandlimited channel without the ICI and the ISI. Right after Chang's publication of his paper, Saltzburg [3] demonstrated the performance of the efficient parallel data transmission systems in 1967, where he concluded that "the strategy of designing an efficient parallel system should concentrate on reducing crosstalk between adjacent channels than on perfecting the individual channels themselves". His conclusion has been proven far-sighted today in the digital baseband signal processing to battle the ICI.

Through the developments of OFDM technology, there are two remarkable contributions to OFDM which transform the original "analog" multicarrier system to today's digitally implemented OFDM. The use of DFT (discrete Fourier transform) to perform baseband modulation and demodulation was the first milestone when Weinstein and Ebert [4] published their paper in 1971. Their method eliminated the banks of subcarrier oscillators and coherent demodulators required by frequency-division multiplexing and hence reduced the cost of OFDM systems. Moreover, DFT-based frequency-division multiplexing can be completely implemented in digital baseband, not by bandpass filtering, for highly efficient processing. FFT, a fast algorithm for computing DFT, can further reduce the number of arithmetic operations from N^2 to NlogN (N is FFT size). Recent advances in VLSI (very large scale integration) technology has made high-speed, large-size FFT chips commercially available. In Weinstein's paper [4], they used a guard interval between consecutive symbols and the

raised-cosine windowing in the time-domain to combat the ISI and the ICI. But their system could not keep perfect orthogonality between subcarriers over a time dispersive channel. This problem was first tackled by Peled and Ruiz [6] in 1980 with the introduction of CP (cyclic prefix) or cyclic extension. They creatively filled the empty guard interval with a cyclic extension of the OFDM symbol. If the length of CP is longer than the impulse response of the channel, the ISI can be eliminated completely. Furthermore, this effectively simulates a channel performing cyclic convolution which implies orthogonality between subcarriers over a time dispersive channel. Though this introduces an energy loss proportional to the length of CP when the CP part in the received signal is removed, the zero ICI generally pays the loss. And it is the second major contribution to OFDM systems.

With OFDM systems getting more popular applications, the requirements for a better performance is becoming higher. Hence more research efforts are poured into the investigation of OFDM systems. Pulse shaping [7, 8], at an interference point view, is beneficial for OFDM systems since the spectrum of an OFDM signal can be shaped to be more well-localized in frequency; Synchronization [9, 10, 11] in time domain and in frequency domain renders OFDM systems robust against timing errors, phase noise, sampling frequency errors and carrier frequency offsets; For coherent detection, channel estimation [46, 49, 48] provides accurate channel state information to enhance performance of OFDM systems; Various effective techniques are exploited to reduce the relatively high PAPR [12, 13] such as clipping and peak windowing.

The principle of OFDM is to divide a single high-data-rate stream into a number of lower rate streams that are transmitted simultaneously over some narrower subchannels. Hence it is not only a modulation (frequency modulation) technique, but also a multiplexing (frequency-division multiplexing) technique. Before we mathematically describe the transmitter-channel-receiver structure of OFDM systems, a couple of graphical intuitions will make it much easier to understand how OFDM works. OFDM starts with the "O", i.e., orthogonal. That orthogonality differs OFDM from conventional FDM (frequency-division multiplexing) and is the source where all the advantages of OFDM come from. The difference between OFDM and conventional FDM is illustrated in Figure 1.1.

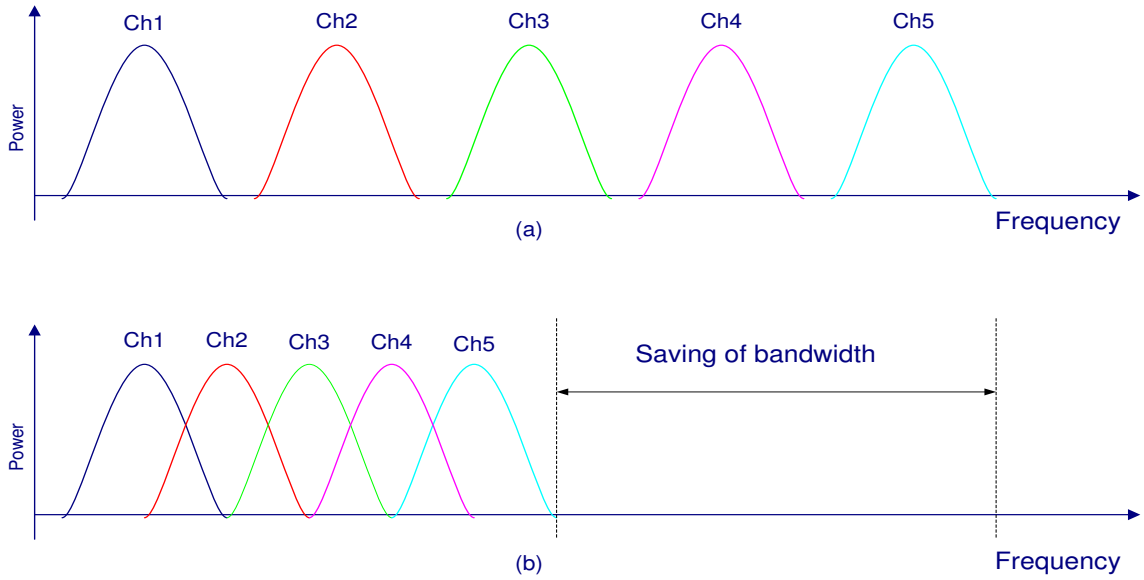

Figure 1.1: Comparison between conventional FDM and OFDM

It can be seen from Figure 1.1, in order to implement the conventional parallel data transmission by FDM, a guard band must be introduced between the different

carriers to eliminate the interchannel interference. This leads to an inefficient use of the rare and expensive spectrum resource. Hence it stimulated the searching for an FDM scheme with overlapping multicarrier modulation in the mid of 1960s. To realize the overlapping multicarrier technique, however we need to get rid of the ICI, which means that we need perfect orthogonality between the different modulated carriers. The word "orthogonality" implies that there is a precise mathematical relationship between the frequencies of the individual subcarriers in the system. In OFDM systems, assume that the OFDM symbol period is T_{sym}, then the minimum subcarrier spacing is $1/T_{\text{sym}}$. By this strict mathematical constraint, the integration of the product of the received signal and any one of the subcarriers f_{sub} over one symbol period T_{sym} will extract that subcarrier f_{sub} only, because the integration of the product of f_{sub} and any other subcarriers over T_{sym} results zero. That indicates no ICI in the OFDM system while achieving almost 50% bandwidth savings. In the sense of multiplexing, we refer to Figure 1.2 to illustrate the concept of OFDM. Every T_{sym} seconds, a total of N complex-valued numbers S_k from different QAM/PSK (quadrature and amplitude modulation/phase shift keying) constellation points are used to modulate N different complex carriers centered at frequency $f_k, 1 \leq k \leq N$. The composite signal is obtained by summing up all the N modulated carriers.

It is worth noting that OFDM achieves frequency-division multiplexing by baseband processing rather than by bandpass filtering. Indeed, as shown in Figure 1.3, the individual spectra has sinc shape. Even though they are not bandlimited, each

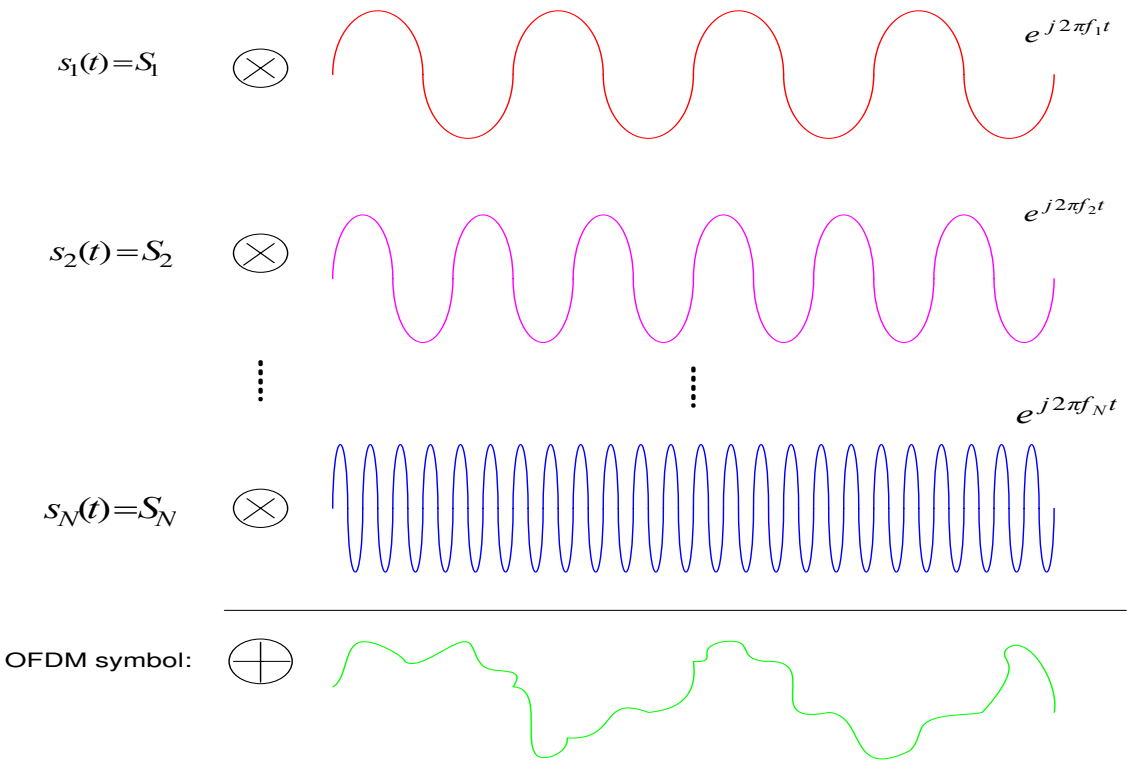

Figure 1.2: Graphical interpretation of OFDM concept

subcarrier can still be separated from the others since orthogonality guarantees that the interfering sincs have nulls at the frequency where the sinc of interest has a peak.

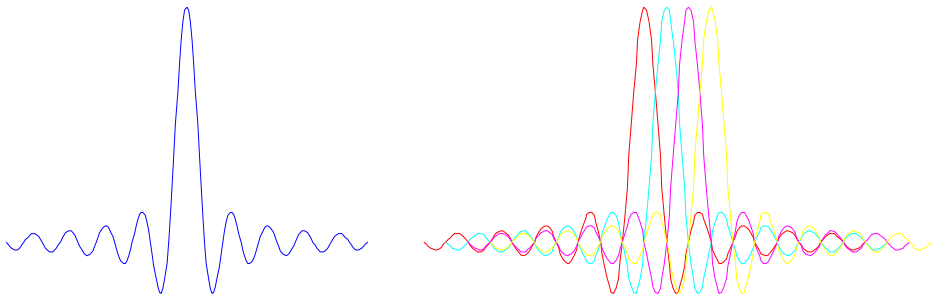

Figure 1.3: Spectra of (a) an OFDM subchannel (b) an OFDM symbol

The use of IDFT (inverse discrete Fourier transform), instead of local oscillators, was an important breakthrough in the history of OFDM. It is an imperative part for OFDM system today. It transforms the data from frequency domain to time domain. Figure 1.4 shows the preliminary concept of DFT used in an OFDM system. When the DFT of a time domain signal is computed, the frequency domain results are a function of the sampling period T and the number of sample points N. The fundamental frequency of the DFT is equal to $\frac{1}{NT}$ (1/total sample time). Each frequency represented in the DFT is an integer multiple of the fundamental frequency. The maximum frequency that can be represented by a time domain signal sampled at rate $\frac{1}{T}$ is $f_{max} = \frac{1}{2T}$ as given by the Nyquist sampling theorem. This frequency is located in the center of the DFT points. The IDFT performs exactly the opposite operation to the DFT. It takes a signal defined by frequency components and converts them to a time domain signal. The time duration of the IDFT time signal is equal to NT. In

essence, IDFT and DFT is a reversable pair. It is not necessary to require that IDFT be used in the transmitter side. It is perfectly valid to use DFT at transmitter and then to use IDFT at receiver side.

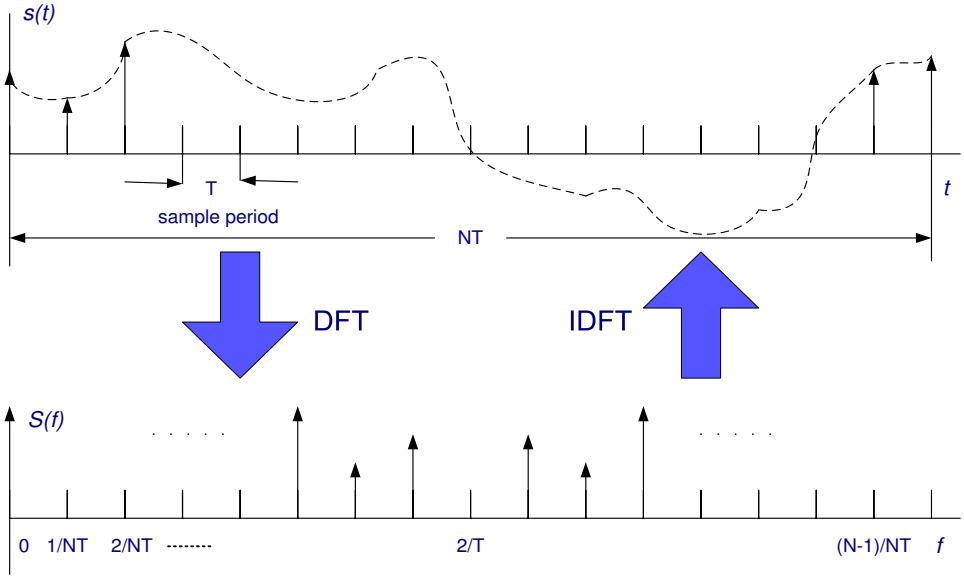

Figure 1.4: Preliminary concept of DFT

After the graphical description of the basic principles of OFDM such as orthogonality, frequency modulation and multiplexing and use of DFT in baseband processing, it is a time to look in more details at the signals flowing between the blocks of an OFDM system and their mathematical relations. At this point, we employ the following assumptions for the OFDM system we consider.

- a CP is used;
- the channel impulse response is shorter than the CP, in terms of their respective length;

- there is perfect synchronization between the transmitter and the receiver;
- channel nosise is additive, white and complex Gaussian;
- the fading is slowing enough for the channel to be considered constant during the transmission of one OFDM symbol.

For a tractable analysis of OFDM systems, we take a common practice to use the simplified mathematical model. Though the first OFDM system was implemented by analogue technology, here we choose to investigate a discrete-time model of OFDM step by step since digital baseband synthesis is widely exploited for today's OFDM systems. Figure 1.5 shows a block diagram of a baseband OFDM modem which is based on PHY (physical layer) of IEEE standard 802.11a [37].

Before describing the mathematical model, we define the symbols and notations used in this dissertation. Capital and lower-case letters denote signals in frequency domain and in time domain respectively. Arrow bar indicates a vector and boldface letter without an arrow bar represents a matrix. It is packed into a table as follows.

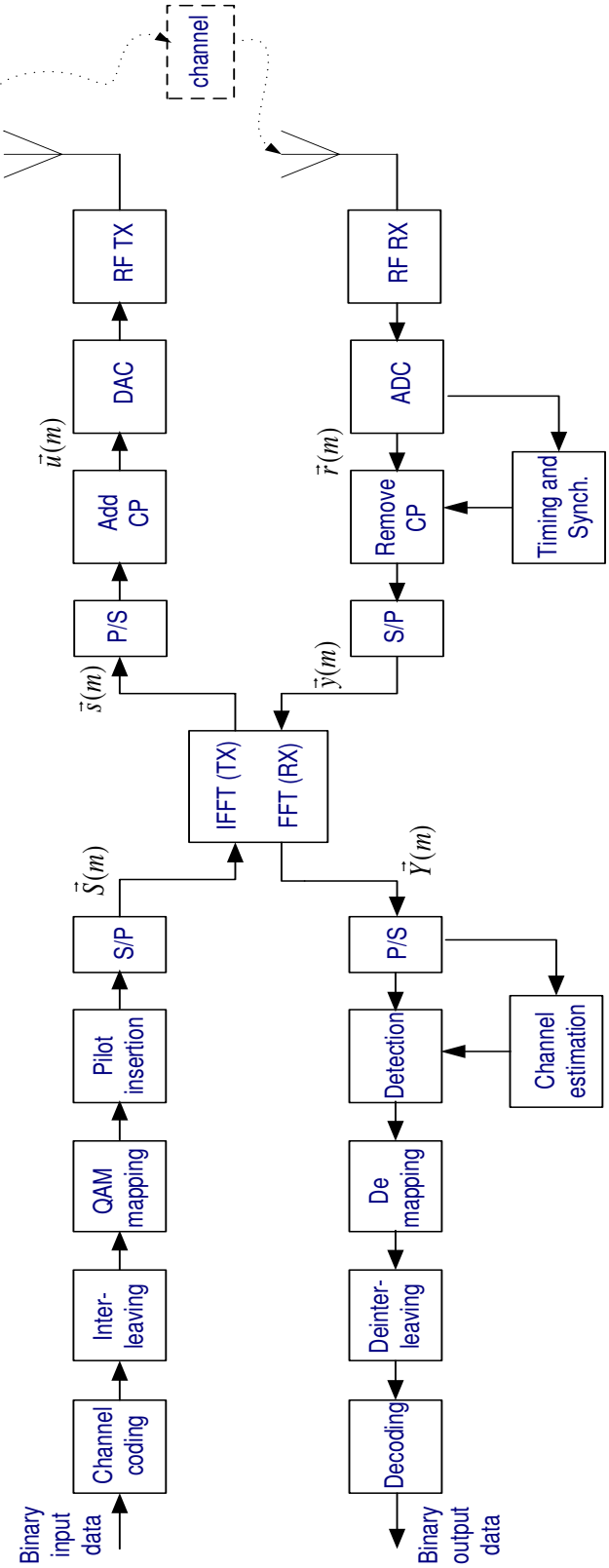

Figure 1.5: Block diagram of a baseband OFDM transceiver

$\mathbf{A}_{p\times q}$	$p \times q$ matrix
$\vec{a}$	column vector
$\mathbf{I}_p$	$p \times p$ identity matrix
$\mathbf{0}$	zero matrix
diag($\vec{\mathrm{a}}$)	diagonal matrix with $\bar{a}$'s elements on the diagonal
$\mathbf{A}^T$	transpose of $\mathbf{A}$
$\mathbf{A}^*$	complex conjugate of $\mathbf{A}$
$\mathbf{A}^H$	Hermitian of $\mathbf{A}$
tr($\mathbf{A}$)	trace of $\mathbf{A}$
rank($\mathbf{A}$)	rank of $\mathbf{A}$
det($\mathbf{A}$)	determinant of $\mathbf{A}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of $\mathbf{A}$ and $\mathbf{B}$

As shown in Figure 1.5, the input serial binary data will be processed by a data scrambler first and then channel coding is applied to the input data to improve the BER (bit error rate) performance of the system. The encoded data stream is further interleaved to reduce the burst symbol error rate. Dependent on the channel condition like fading, different base modulation modes such as BPSK (binary phase shift keying), QPSK (quadrature phase shift keying) and QAM are adaptively used to boost the data rate. The modulation mode can be changed even during the transmission of data frames. The resulting complex numbers are grouped into column vectors which have the same number of elements as the FFT size, N. For simplicity of presentation and ease of understanding, we choose to use matrix and vector to describe the mathematical model. Let $\vec{S}(m)$ represent the m-th OFDM symbol in

the frequency domain, i.e.,

$$\vec{S}(m) = \begin{bmatrix} S(mN) \\ \vdots \\ S(mN+N-1) \end{bmatrix}_{N\times 1},$$

where m is the index of OFDM symbols. We assume that the complex-valued elements $\{S(mN), S(mN+1), \ldots, S(mN+N-1)\}$ of $\vec{S}(m)$ are zero mean and uncorrelated random variables whose sample space is the signal constellation of the base modulation (BPSK, QPSK and QAM). To achieve the same average power for all different mappings, a normalization factor K_{MOD} [37] is multiplied to each elements of $\vec{S}(m)$ such that the average power of the different mappings is normalized to unity. To obtain the time domain samples, as shown by the IDFT block in Figure 1.5, an IFFT (inverse fast Fourier transform) operation is represented by a matrix multiplication. Let $\mathbf{F}_N$ be the N-point DFT matrix whose (p,q)-th elements is $e^{-j\frac{2\pi}{N}(p-1)(q-1)}$. The resulting time domain samples $\vec{s}(m)$ can be described by

$$\begin{aligned} \vec{s}(m) &= \begin{bmatrix} s(mN) \\ \vdots \\ s(mN+N-1) \end{bmatrix}_{N\times 1} \\ &= (\tfrac{1}{N})\mathbf{F}_N^H \vec{S}(m). \end{aligned} \tag{1.1}$$

Compared to the costly and complicated modulation and multiplexing of conventional FDM systems, OFDM systems easily implement them by using FFT in baseband processing. To combat the multipath delay spread in wireless channels, the time-domain samples $\vec{s}(m)$ is cyclically extended by copying the last N_g samples and pasting them to the front, as shown in Figure 1.6(a) [6].

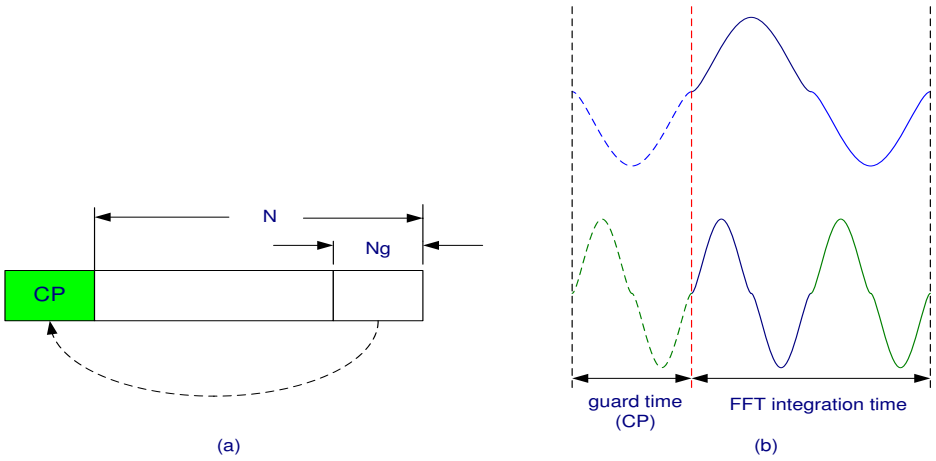

Figure 1.6: (a) Concept of CP; (b) OFDM symbol with cyclic extension

Let $\vec{u}(m)$ denote the cyclically extended OFDM symbol as

$$\vec{u}(m) = \begin{bmatrix} u(mN_{tot}) \\ \vdots \\ u(mN_{tot} + N_{tot} - 1) \end{bmatrix} = \begin{bmatrix} CP \\ \vec{s}(m) \end{bmatrix}_{N_{tot} \times 1},$$

where $N_{tot} = N + N_g$ is the length of $\vec{u}(m)$. In the form of matrix, the CP insertion can be readily expressed as a matrix product of $\vec{s}(m)$ and an $N_{tot} \times N$ matrix $\mathbf{A}_{CP}$. By straight computation, it holds that

$$\vec{u}(m) = \mathbf{A}_{CP}\vec{s}(m), \tag{1.2}$$

where

$$\mathbf{A}_{CP} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{N_g} \\ \mathbf{I}_{N-N_g} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_g} \end{bmatrix}_{(N+N_g)\times N}.$$

One of the challenges from the harsh wireless channels is the multipath delay spread. If the delay spread is relatively large compared to the symbol duration, then a delayed copy of a previous symbol will overlap the current one which implies severe ISI. To

eliminate the ISI almost completely, a CP is introduced for each OFDM symbol and the length of CP, N_g must be chosen longer than the experienced delay spread, L, i.e., $N_g \geq L$. In addition, CP is capable of maintaining the orthogonality among subcarriers which implies zero ICI. It is because the OFDM symbol is cyclically extended and this ensures that the delayed replicas of the OFDM symbol always have an integer number of cycles within the FFT interval, as long as the delay is smaller than the CP. It is clearly illustrated in Figure 1.6(b). No matter where the FFT window starts, provided that it is within the CP, there will be always one or two complete cycles within FFT integration time for the symbol on top and at below respectively. In IEEE 802.11a standard [37], N_g is at least 16. The obtained OFDM symbol (including the CP) $\vec{u}(m)$, as shown in Figure 1.5, must be converted to the analogue domain by an DAC (digital-to-analog converter) and then up-converted for RF transmission since it is currently not practical to generate the OFDM symbol directly at RF rates. To remain in the discrete-time domain, the OFDM symbol could be up-sampled and added to a discrete carrier frequency. This carrier could be an IF (intermediate frequency) whose sample rate is handled by current technology. It could then be converted to analog and increased to the final transmit frequency using analog frequency conversion methods. Alternatively, the OFDM modulation could be immediately converted to analog and directly increased to the desired RF transmit frequency. Either way has its advantages and disadvantages. Cost, power consumption and complexity must be taken into consideration for the selected technique.

The RF signal is transmitted over the air. For the wireless channel, it is assumed in this thesis as a quasi-static frequency-selective Rayleigh fading channel [71]. It indicates that the channel remains constant during the transmission of one OFDM symbol. Suppose that the multipath channel can be modeled by a discrete-time baseband equivalent $(L-1)$th-order FIR (finite impulse response) filter with filter taps $\{h_0, h_1, \ldots, h_l, \ldots, h_{L-1}\}$. It is further assumed that the channel impulse response,i.e., the equivalent FIR filter taps, are independent zero mean complex Gaussian random variables with variance of $\frac{1}{2}\mathcal{P}_l$ per dimension. The ensemble of $\{\mathcal{P}_0, \ldots, \mathcal{P}_l, \ldots, \mathcal{P}_{L-1}\}$ is the PDP (power delay profile) of the channel and usually the total power of the PDP is normalized to be 1 as the unit average channel attenuation. Denote the CIR (channel impulse response) vector $\vec{h}_m$ as

$$\vec{h}_m = \begin{bmatrix} h_{0,m} \\ \vdots \\ h_{L-1,m} \end{bmatrix}_{L\times 1},$$

where the subscript m is kept to imply that the channel may vary from one OFDM symbol to the next one. Then the complex baseband equivalent received signal can be represented by a discrete-time convolution as

$$r(mN_{tot}+n) = \sum_{l=0}^{L-1} h_{l,m}u(mN_{tot}+n-l) + v(mN_{tot}+n), \tag{1.3}$$

where $mN_{tot}+n$ means the n-th received sample during the m-th OFDM symbol and $0 \le n \le N_{tot}-1$. The term $v(mN_{tot}+n)$ represents the complex AWGN at the $(mN_{tot}+n)$-th time sample with zero mean and variance of $\frac{1}{2}\sigma_v^2$ per dimension. Hence, the expected SNR (signal-to-noise ratio) per received signal is $\rho = \frac{1}{\sigma_v^2}$. In

order for the parallel processing by the DFT block in Figure 1.5, we will rewrite the equation (1.3) into a matrix form. First we define

$$\vec{r}(m) = \begin{bmatrix} r(mN_{tot}) \\ \vdots \\ r(mN_{tot} + N_{tot} - 1) \end{bmatrix}; \quad \vec{v}(m) = \begin{bmatrix} v(mN_{tot}) \\ \vdots \\ v(mN_{tot} + N_{tot} - 1) \end{bmatrix}, \tag{1.4}$$

and

$$\mathbf{h}_{m,Toep} = \begin{bmatrix} h_{0,m} & & & & \\ \vdots & \ddots & & & \\ h_{L-1,m} & \cdots & h_{0,m} & & \\ & \ddots & \vdots & \ddots & \\ & & h_{L-1,m} & \cdots & h_{0,m} \end{bmatrix}; \quad \mathbf{h}^{(c)}_{m,Toep} = \begin{bmatrix} h_{L-1,m} & \cdots & h_{1,m} \\ & \ddots & \vdots \\ & & h_{L-1,m} \\ & & \\ & & \end{bmatrix}. \tag{1.5}$$

Then it is straight forward to have the following input-output relationship with regard to the channel

$$\vec{r}(m) = \mathbf{h}_{m,Toep}\vec{u}(m) + \mathbf{h}^{(c)}_{m,Toep}\vec{u}(m-1) + \vec{v}(m). \tag{1.6}$$

It is easy to see in (1.6) that the first $L-1$ terms of $\vec{r}(m)$, i.e., $\{r(mN_{tot}), \ldots, r(mN_{tot}+L-2)\}$, will be affected by the ISI term $\mathbf{h}^{(c)}_{m,Toep}\vec{u}(m-1)$ since the Toeplitz and upper triangular matrix $\mathbf{h}^{(c)}_{m,Toep}$ has non-zero entries in the first $L-1$ rows. In order to remove the ISI term, we transform the $N_{tot} \times 1$ vector $\vec{r}(m)$ into an $N \times 1$ vector $\vec{y}(m)$ by simply cutting off the first N_g possibly ISI affected elements. For complete elimination of ISI, $N_g \geq L$ must be satisfied. It is a reverse operation of the cyclic extension as implemented in the transmitter side. Consistently this transformation

can also be expresses as matrix-vector product

$$\vec{y}(m) = \begin{bmatrix} y(mN) \\ \vdots \\ y(mN+N-1) \end{bmatrix} = \mathbf{A}_{DeCP}\vec{r}(m) \ , \tag{1.7}$$

where

$$\mathbf{A}_{DeCP} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_N \end{bmatrix}_{N \times N_{tot}} .$$

As shown in Figure 1.5, the ISI-free received signal $\vec{y}(m)$ is demodulated by FFT and hence it is converted back to the frequency domain received signal $\vec{Y}(m)$. It is described by

$$\vec{Y}(m) = \begin{bmatrix} Y(mN) \\ \vdots \\ Y(mN+N-1) \end{bmatrix} = \mathbf{F}_N\vec{y}(m) \ . \tag{1.8}$$

After obtaining the received signal $\vec{Y}(m)$, symbol detection can be implemented if the channel state information is known or it can be estimated by some channel estimation algorithms. The detected symbol will pass through a series of reverse operations to retrieve the input binary information, corresponding to the encoding, interleaving and mapping in the transmitter side. Following the signal flow from the transmitted signal $\vec{S}(m)$ to the receive signal $\vec{Y}(m)$, a simple relationship between them can be expressed as

$$\vec{Y}(m) = \mathbf{H}_{m,diag}\vec{S}(m) + \vec{V}(m), \tag{1.9}$$

where the diagonal matrix $\mathbf{H}_{m,diag}$ is

$$\mathbf{H}_{m,diag} = \begin{bmatrix} H_{0,m} & & \\ & \ddots & \\ & & H_{N-1,m} \end{bmatrix}; \quad H_{k,m} = \sum_{l=0}^{L-1} h_l e^{-j\frac{2\pi}{N}kl}, 0 \le k \le N.$$

and $\vec{V}(m)$ is the complex AWGN in frequency domain. This simple transmitter-and-receiver structure is well known in all the literatures [42, 46, 48, 49] and it is an important reason for the wide application of OFDM systems. The transmitted signal can be easily extracted by simply dividing the channel frequency response for the specific subcarrier. Hence it eliminates the needs of a complicated equalizer at the receive side. In this thesis, we do not directly jump on this known conclusion for two reasons. First, following through the baseband block diagram in Figure 1.5, we use a matrix form of presentation to describe all the input-output relationship with respect to each block. This gives us a clear and thorough understanding of all the signal processing within the OFDM system. It is a different view from those in literatures which can be summarized by the fact that the discrete Fourier transform of a cyclic convolution (IDFT($\vec{S}(m)$) and $\vec{h}_m$) in time domain leads to a product of the frequency responses ($\vec{S}(m)$ and DFT($\vec{h}_m$)) of the two convoluted terms. Second, this provides a base for our channel estimator design in the following chapter. Next, the simple relation in (1.9) is shown by going through the signal flow backwards from

$\vec{Y}(m)$ to $\vec{S}(m)$ that

$$\begin{aligned}
\vec{Y}(m) &= \mathbf{F}_N \vec{y}(m) \\
&= \mathbf{F}_N(\mathbf{A}_{DeCP}\vec{r}(m)) \\
&= \mathbf{F}_N\{\mathbf{A}_{DeCP}[\mathbf{h}_{m,Toep}\vec{u}(m) + \mathbf{h}^{(c)}_{m,Toep}\vec{u}(m-1) + \vec{v}(m)]\} \\
&= \mathbf{F}_N[\mathbf{A}_{DeCP}\mathbf{h}_{m,Toep}\vec{u}(m) + \mathbf{A}_{DeCP}\vec{v}(m)] \\
&= \mathbf{F}_N[\mathbf{A}_{DeCP}\mathbf{h}_{m,Toep}\mathbf{A}_{CP}\vec{s}(m) + \mathbf{A}_{DeCP}\vec{v}(m)] \\
&= \mathbf{F}_N[\mathbf{A}_{DeCP}\mathbf{h}_{m,Toep}\mathbf{A}_{CP}(\tfrac{1}{N})\mathbf{F}^H_N\vec{S}(m) + \mathbf{A}_{DeCP}\vec{v}(m)] \\
&= \tfrac{1}{N}\mathbf{F}_N[\mathbf{A}_{DeCP}\mathbf{h}_{m,Toep}\mathbf{A}_{CP}]\mathbf{F}^H_N\vec{S}(m) + \mathbf{F}_N(\mathbf{A}_{DeCP}\vec{v}(m)) \\
&= \tfrac{1}{N}[\mathbf{F}_N\mathbf{h}_{Cir}\mathbf{F}^H_N]\vec{S}(m) + \vec{V}(m)
\end{aligned}, \tag{1.10}$$

where $\vec{V}(m) = \mathbf{F}_N(\mathbf{A}_{DeCP}\vec{v}(m))$ and $\mathbf{h}_{Cir} = \mathbf{A}_{DeCP}\mathbf{h}_{m,Toep}\mathbf{A}_{CP}$ is an $N \times N$ circulant matrix with some special properties. It is parameterized as

$$\mathbf{h}_{m,Cir} = \begin{bmatrix}
h_{0,m} & 0 & \cdots & \cdots & 0 & h_{L-1,m} & h_{L-2,m} & \cdots & h_{1,m} \\
h_{1,m} & h_{0,m} & 0 & \cdots & 0 & 0 & h_{L-1,m} & \cdots & h_{2,m} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
h_{L-2,m} & \cdots & \cdots & h_{0,m} & 0 & \cdots & \cdots & 0 & h_{L-1,m} \\
h_{L-1,m} & \cdots & \cdots & \cdots & h_{0,m} & 0 & \cdots & \cdots & 0 \\
0 & h_{L-1,m} & \cdots & \cdots & \cdots & h_{0,m} & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & h_{L-1,m} & h_{L-2,m} & \cdots & \cdots & h_{0,m}
\end{bmatrix}_{N\times N}. \tag{1.11}$$

As stated in [38], an $N \times N$ circulant matrix has some important properties:

- All the $N \times N$ circulant matrices have the same eigenvectors and they are the columns of $F_N^{\mathcal{H}}$, where F_N is the N-point FFT matrix;
- The corresponding eigenvalues $\{\lambda_1, \cdots, \lambda_N\}$ are the FFT of the first column of the circulant matrix;

The first column of the circulant matrix $\mathbf{h}_{m,Cir}$ is $[h_{0,m}^T, \ldots, h_{L-1,m}^T, 0, \ldots, 0]^T$. Hence, the eigenvalues of $\mathbf{h}_{m,Cir}$ is

$$\begin{bmatrix} H_{0,m} \\ H_{1,m} \\ \vdots \\ H_{N-1,m} \end{bmatrix} = \mathbf{F}_N \begin{bmatrix} h_{0,m} \\ \vdots \\ h_{L-1,m} \\ 0_{(N-L)\times 1} \end{bmatrix}.$$

Taking eigenvalue decomposition of $\mathbf{h}_{m,Cir}$, we have

$$\mathbf{h}_{m,Cir} = \frac{1}{N}\mathbf{F}_N^H \begin{bmatrix} H_{0,m} & & \\ & \ddots & \\ & & H_{N-1,m} \end{bmatrix} \mathbf{F}_N. \tag{1.12}$$

Simply substituting (1.12) into (1.10) shows that (1.9) is true.

The simple model in (1.9) is widely exploited for theoretical research. It is, however, based on all of the assumptions we make at the beginning of this section. In the practical OFDM systems, a lot of efforts were made in research to keep the OFDM systems as close to this model as possible. Perfect synchronization in time domain and frequency domain is the most challenging subject. The orthogonality could be easily destroyed by a few factors such as the Doppler shift resulting from the relative movement between the transmitter and the receiver, the frequency mismatch between the oscillators at two ends, large timing errors and phase noise. Meanwhile, accurate channel state information is critical for reducing the BER and improving the system performance. Hence, joint channel estimation and synchronization with low complexity is an active research area for current OFDM systems. As long as the orthogonality is obtained, OFDM is a simple and efficient multicarrier data transmission technique.

1.2 Dissertation Contributions

In the first part, this dissertation addresses one of the most fundamental problems in MIMO-OFDM communication system design, i.e., the fast and reliable channel estimation. By using the pilot symbols, a MIMO-OFDM channel estimator is proposed in this dissertation which is capable of estimating the time-dispersive and frequency-selective fading channel. Our contribution to this dissertation are as follows.

- Great Simplicity:

  For an $N_t \times N_r$ MIMO (N_t : number of transmit antennas,N_r : number of receive antennas) system, the complexity of any kinds of signal processing algorithms at the physical layer is increased usually by a factor of $N_t N_r$. Hence, simplicity plays an important role in the system design. We propose a pilot tone design for MIMO-OFDM channel estimation that N_t disjoint set of pilot tones are placed on one OFDM block at each transmit antenna. For each pilot tone set, it has L (L: channel length) pilot tones which are equally-spaced and equally-powered. The pilot tones from different transmit antennas comprise a unitary matrix and then a simple least square estimation of the MIMO channel is easily implemented by taking advantage of the unitarity of the pilot tone matrix. There is no need to compute the inverse of large-size matrix which is usually required by LS algorithm. Contrast to some other simplified channel estimation methods by assuming that there are only a few dominant paths among L of them

and then neglecting the rest weaker paths in the channel, our method estimates the full channel information with a reduced complexity.

- Estimation of Fast Time-varying Channel:

  In a highly mobile environment, like a mobile user in a vehicle riding at more than 100km/hr, the wireless channel may change within one or a small number of symbols. But the information packet could contain hundreds of data symbols or even more. In the literature [50] there are some preamble designs that the wireless channel is only estimated at the preamble part of a whole data packet and is assumed to be constant during the transmission of the rest data part. Different from the preamble design, our scheme is proposed that we distribute the pilot symbols in the preamble to each OFDM block for channel estimation. Since the pilot tones are placed on each OFDM block, the channel state information can be estimated accurately and quickly, no matter how fast the channel condition is varying.

- Link to SFC (Space-frequency code):

  Usually channel estimation and space-frequency code design of MIMO-OFDM systems are taken as two independent subject, especially for those algorithms generalized from their counterparts in the SISO (single-input single-output) case. Some researchers [48, 50] propose some orthogonal structures for pilot tone design and try to reduce the complexity of computing. However, each

individual structure is isolated and it is not easy to generalize their structures to the MIMO system with any number of transmit antennas and receive antennas. In this dissertation, the orthogonal pilot tone matrix we propose is indeed a space-frequency code. The row direction of the matrix stands for different pilot tone sets in the frequency domain, and the column direction represents the individual transmit antennas in spatial domain. And it can be readily extended to an $N_t \times N_r$ MIMO system by constructing an $N_t \times N_t$ orthogonal matrix. With this explicit relation to space-frequency code, the design of pilot-tone matrix for MIMO-OFDM channel estimation can be conducted in a more broad perspective. This link will shed light on each other.

In the second part of this dissertation, we contribute to the formulation of the location estimation into a constrained LS-type optimization problem. As surveyed in [53], there are different methods for location estimation based on measurements of TOA, TDOA, AOA and amplitude. There are two problems which are not given full attention and may increase the complexity of the algorithm. One problem is that only an intermediate solution can be first obtained by solving the LS estimation problem. It means that the intermediate solution is still a function of the unknown target location. Extra constraints are needed to get the final target estimation. Though such a constraint exists, solving the quadratic equation may end up with nonexistence of a real positive root. Another problem is that it is unclear how the measurement noise variance affect the estimation accuracy. Intuitively, a small variance is always pre-

ferred. In our proposed algorithm, the constrained LS-type optimization problem is solved by using Lagrange multiplier. And it is pointed out that the noise variance is closely related to the equivalent SNR. For example, in the case of TDOA, the equivalent SNR is the ratio of the time for a signal traveling from the target to the k-th base station over the noise variance. A smaller noise variance then indicates a higher SNR which leads to more accurate location estimation. The formulation of a constrained LS-type optimization has its advantages. First it holds a performance which is close to the ML algorithm, provided that the assumption about the measurement noise variance is satisfied. Second it inherits the simplicity from the LS algorithm.

1.3 Organization of the Dissertation

This dissertation is organized as follows. In Chapter 1, the principle of OFDM is illustrated through instructive figures and the signal mode of OFDM systems is described by matrix representation in details. Also, a review of research on channel estimation for OFDM systems is covered in Chapter 1. In Chapter 2, it is mainly focused on the pilot tone based channel estimation of MIMO-OFDM systems. It ends up with intensive computer simulations of different estimation algorithms and effects of some key OFDM parameters on estimator performance. Chapter 3 devotes to wireless location on WiMax network. A constrained LS-type optimization problem is formulated under a mild assumption and it is solved by using Lagrange multiplier method. Finally this dissertation is summarized in Chapter 5 by suggesting some open research subjects on the way.

Chapter 2

MIMO-OFDM Channel Estimation

2.1 Introduction

With the ever increasing number of wireless subscribers and their seemingly "greedy" demands for high-data-rate services, radio spectrum becomes an extremely rare and invaluable resource for all the countries in the world. Efficient use of radio spectrum requires that modulated carriers be placed as close as possible without causing any ICI and be capable of carrying as many bits as possible. Optimally, the bandwidth of each carrier would be adjacent to its neighbors, so there would be no wasted bands. In practice, a guard band must be placed between neighboring carriers to provide a guard space where a shaping filter can attenuate a neighboring carrier's signal. These guard bands are waste of spectrum. In order to transmit high-rate data, short symbol periods must be used. The symbol period T_{sym} is the inverse of the baseband data rate R ($R = 1/T_{sym}$), so as R increases, T_{sym} must decrease. In a multipath environment, however, a shorter symbol period leads to an increased degree of ISI, and thus performance loss. OFDM addresses both of the two problems with its

unique modulation and multiplexing technique. OFDM divides the high-rate stream into parallel lower rate data and hence prolongs the symbol duration, thus helping to eliminate ISI. It also allows the bandwidth of subcarriers to overlap without ICI as long as the modulated carriers are orthogonal. OFDM therefore is considered as a good candidate modulation technique for broadband access in a very dispersive environments [42, 43].

However, relying solely on OFDM technology to improve the spectral efficiency gives us only a partial solution. At the end of 1990s, seminal work by Foshini and Gans [21] and, independently, by Teltar [22] showed that there is another alternative to accomplish high-data-rate over wireless channels: the use of multiple antennas at the both ends of the wireless link, often referred to as MA (multiple antenna) or MIMO in the literature [21, 22, 17, 16, 25, 26]. The MIMO technique does not require any bandwidth expansions or any extra transmission power. Therefore, it provides a promising means to increase the spectral efficiency of a system. In his paper about the capacity of multi-antenna Gaussian channels [22], Telatar showed that given a wireless system employing N_t TX (transmit) antennas and N_r RX (receive) antennas, the maximum data rate at which error-free transmission over a fading channel is theoretically possible is proportional to the minimum of N_t and N_r (provided that the $N_t N_r$ transmission paths between the TX and RX antennas are statistically independent). Hence huge throughput gains may be achieved by adopting $N_t \times N_r$ MIMO systems compared to conventional 1×1 systems that use single antenna at

both ends of the link with the same requirement of power and bandwidth. With multiple antennas, a new domain,namely, the spatial domain is explored, as opposed to the existing systems in which the time and frequency domain are utilized.

Now let's come back to the previous question: what can be done in order to enhance the data rate of a wireless communication systems? The combination of MIMO systems with OFDM technology provides a promising candidate for next generation fixed and mobile wireless systems [42]. In practice for coherent detection, however, accurate channel state information in terms of channel impulse response (CIR) or channel frequency response (CFR) is critical to guarantee the diversity gains and the projected increase in data rate.

The channel state information can be obtained through two types of methods. One is called blind channel estimation [44, 45, 46], which explores the statistical information of the channel and certain properties of the transmitted signals. The other is called training-based channel estimation, which is based on the training data sent at the transmitter and known *a priori* at the receiver. Though the former has its advantage in that it has no overhead loss, it is only applicable to slowly time-varying channels due to its need for a long data record. Our work in this thesis focuses on the training-based channel estimation method, since we aim at mobile wireless applications where the channels are fast time-varying. The conventional training-based method [47, 48, 50] is used to estimate the channel by sending first a sequence of OFDM symbols, so-called preamble which is composed of known training symbols.

Then the channel state information is estimated based on the received signals corresponding to the known training OFDM symbols prior to any data transmission in a packet. The channel is hence assumed to be constant before the next sequence of training OFDM symbols. A drastic performance degradation then arises if applied to fast time-varying channels. In [49], optimal pilot-tone selection and placement were presented to aid channel estimation of single-input/single-output (SISO) systems. To use a set of pilot-tones within each OFDM block, not a sequence of training blocks ahead of a data packet to estimate the time-varying channel is the idea behind our work. However direct generalization of the channel estimation algorithm in [49] to MIMO-OFDM systems involves the inversion of a high-dimension matrix [47] due to the increased number of transmit and receive antennas, and thus entails high complexity and makes it infeasible for wireless communications over highly mobile channels. This becomes a bottleneck for applications to broadband wireless communications. To design a low-complexity channel estimator with comparable accuracy is the goal of this chapter.

The bottleneck problem of complexity for channel estimation in MIMO-OFDM systems has been studied by two different approaches. The first one shortens the sequence of training symbols to the length of the MIMO channel, as described in [50], leading to orthogonal structure for preamble design. Its drawback lies in the increase of the overhead due to the extra training OFDM blocks. The second one is the simplified channel estimation algorithm, as proposed in [48], that achieves optimum channel

estimation and also avoids the matrix inversion. However its construction of the pilot-tones is not explicit in terms of space-time codes (STC). We are motivated by both approaches in searching for new pilot-tone design. Our contribution in this chapter is the unification of the known results of [48, 50] in that the simplified channel estimation algorithm is generalized to explicit orthogonal space-frequency codes (SFC) that inherit the same computational advantage as in [48, 50], while eliminating their respective drawbacks. In addition, the drastic performance degradation occurred in [48, 50] is avoided by our pilot-tone design since the channel is estimated at each block. In fact we have formulated the channel estimation problem in frequency domain, and the CFR is parameterized by the pilot-tones in a convenient form for design of SFC. As a result a unitary matrix, composed of pilot-tones from each transmit antenna, can be readily constructed. It is interesting to observe that the LS algorithm based on SFC in this paper is parallel to that for conventional OFDM systems with single transmit/receive antenna. The use of multiple transmit/receive antennas offers more design freedom that provides further improvements on estimation performance.

2.2 System Description

The block diagram of a MIMO-OFDM system [27, 28] is shown in Figure 2.1. Basically, the MIMO-OFDM transmitter has N_t parallel transmission paths which are very similar to the single antenna OFDM system, each branch performing serial-to-parallel conversion, pilot insertion, N-point IFFT and cyclic extension before the final TX signals are up-converted to RF and transmitted. It is worth noting that

the channel encoder and the digital modulation, in some spatial multiplexing systems [28, 29], can also be done per branch, not necessarily implemented jointly over all the N_t branches. The receiver first must estimate and correct the possible symbol timing error and frequency offsets, e.g., by using some training symbols in the preamble as standardized in [37]. Subsequently, the CP is removed and N-point FFT is performed per receiver branch. In this thesis, the channel estimation algorithm we proposed is based on single carrier processing that implies MIMO detection has to be done per OFDM subcarrier. Therefore, the received signals of subcarrier k are routed to the k-th MIMO detector to recover all the N_t data signals transmitted on that subcarrier. Next, the transmitted symbol per TX antenna is combined and outputted for the subsequent operations like digital demodulation and decoding. Finally all the input binary data are recovered with certain BER.

As a MIMO signalling technique, N_t different signals are transmitted simultaneously over $N_t \times N_r$ transmission paths and each of those N_r received signals is a combination of all the N_t transmitted signals and the distorting noise. It brings in the diversity gain for enhanced system capacity as we desire. Meanwhile compared to the SISO system, it complicates the system design regarding to channel estimation and symbol detection due to the hugely increased number of channel coefficients.

2.2.1 Signal Model

To find the signal model of MIMO-OFDM system, we can follow the same approach as utilized in the SISO case. Because of the increased number of antennas, the signal

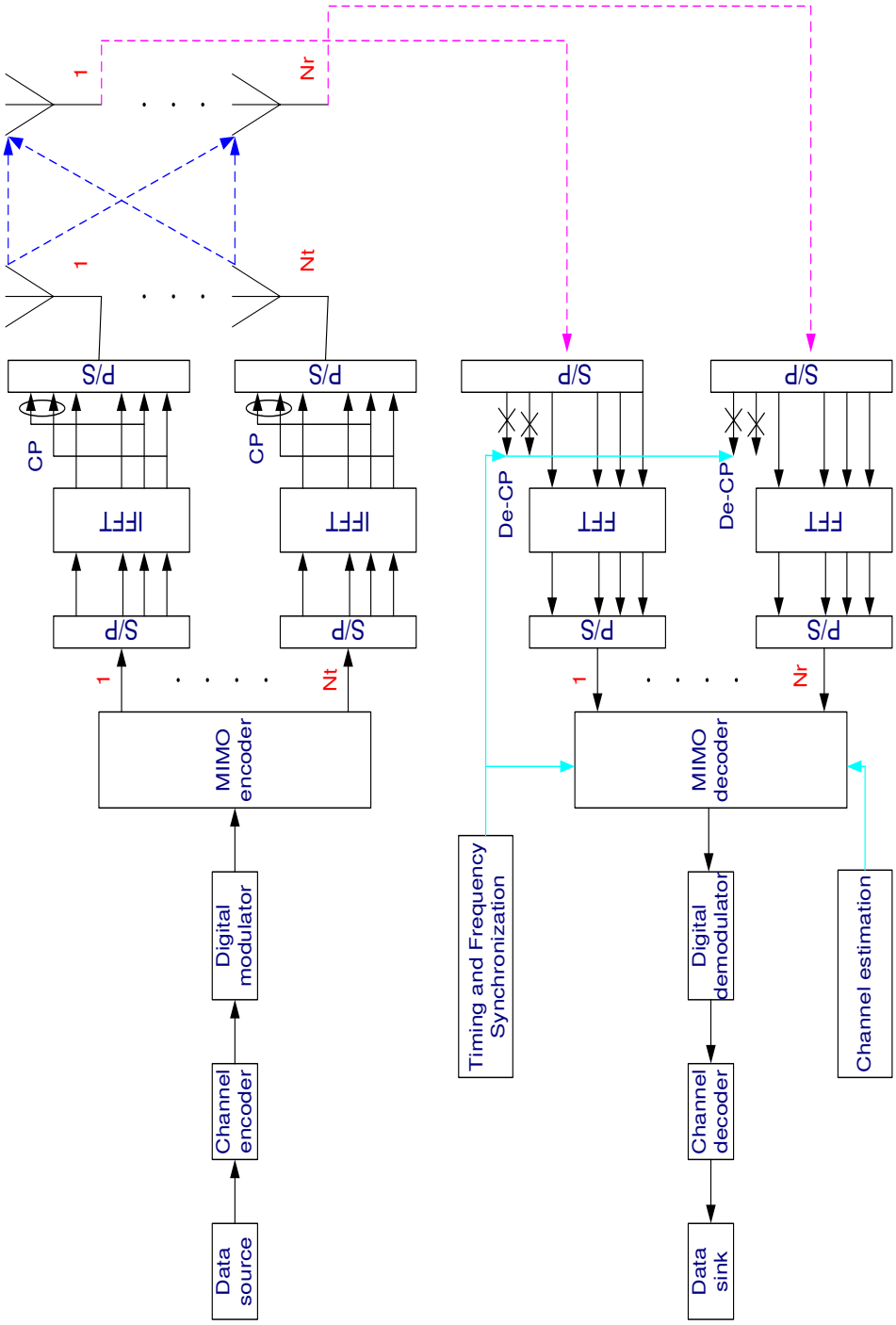

Figure 2.1: $N_t \times N_r$ MIMO-OFDM System model

dimension is changed. For instance, the transmitted signal on the k-th subcarrier in a MIMO system is an $N_t \times 1$ vector, instead of a scalar in the SISO case. For brevity of presentation, the same notations are used for both the SISO and MIMO cases. But they are explicitly defined in each case. There are N_t transmit antennas and hence on each of the N subcarriers, N_t modulated signals are transmitted simultaneously. Denote $\vec{S}(m)$ and $\vec{S}(mN+k)$ as the m-th modulated OFDM symbol in frequency domain and the k-th modulated subcarrier respectively as

$$\vec{S}(m) = \begin{bmatrix} \vec{S}(mN) \\ \vdots \\ \vec{S}(mN+N-1) \end{bmatrix} \qquad \vec{S}(mN+k) = \begin{bmatrix} S_1(mN+k) \\ \vdots \\ S_{N_t}(mN+k) \end{bmatrix}, \tag{2.1}$$

where $S_j(mN+k)$ represents the k-th modulated subcarrier for the m-th OFDM symbol transmitted by the j-th antenna. And it is normalized by a normalization factor K_{MOD} so that there is a unit normalized average power for all the mappings. Taking IFFT of $\vec{S}(m)$ as a baseband modulation, the resulting time-domain samples can be expressed as

$$\begin{aligned} \vec{s}(m) &= \begin{bmatrix} \vec{s}(mN) \\ \vdots \\ \vec{s}(mN+N-1) \end{bmatrix} \qquad \vec{s}(mN+n) = \begin{bmatrix} s_1(mN+n) \\ \vdots \\ s_{N_t}(mN+n) \end{bmatrix} \\ &= \tfrac{1}{N}(\mathbf{F}_N^H \otimes \mathbf{I}_{N_t})\vec{S}(m) \qquad . \end{aligned} \tag{2.2}$$

Here IFFT is a block-wise operation since each modulated subcarrier is a column vector and the generalized NN_t-point IFFT matrix is a Kronecker product of $\mathbf{F}_N$ and $\mathbf{I}_{N_t}$. This is just a mathematical expression. In the real OFDM systems, however, the generalized IFFT operation is still performed by N_t parallel N-point IFFT. To

eliminate the ISI and the ICI, a length-N_g ($N_g \geq L$) CP is prepended to the time-domain samples per branch. The resulting OFDM symbol $\vec{u}(m)$ is denoted as

$$\vec{u}(m) = \begin{bmatrix} \vec{u}(mN_{tot}) \\ \vdots \\ \vec{u}(mN_{tot} + N_{tot} - 1) \end{bmatrix} \quad \vec{u}(mN_{tot} + n) = \begin{bmatrix} u_1(mN_{tot} + n) \\ \vdots \\ u_{N_t}(mN_{tot} + n) \end{bmatrix}. \tag{2.3}$$

In a matrix form, there holds

$$\vec{u}(m) = \mathbf{A}_{CP}\vec{s}(m), \tag{2.4}$$

where

$$\mathbf{A}_{CP} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{N_g} \\ \mathbf{I}_{N-N_g} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_g} \end{bmatrix} \otimes \mathbf{I}_{N_t}.$$

The time-domain samples denoted by $\vec{u}(m)$ may be directly converted to RF for transmission or be up-converted to IF first and then transmitted over the wireless MIMO channel. For the MIMO channel, we assume in this thesis that the MIMO-OFDM system is operating in a frequency-selective Rayleigh fading environment and that the communication channel remains constant during a frame transmission, i.e., quasi-static fading. Suppose that the channel impulse response can be recorded with L time instances, i.e., time samples, then the multipath fading channel between the j-th TX and i-th RX antenna can be modeled by a discrete-time complex base-band equivalent $(L-1)$-th order FIR filter with filter coefficients $h_{ij}(l, m)$, with $l \subseteq \{0, \ldots, L-1\}$ and integer $m > 0$. As assumed in SISO case, these CIR coefficients $\{h_{ij}(0, m), \ldots, h_{ij}(L-1, m)\}$ are independent complex zero-mean Gaussian RV's with variance $\frac{1}{2}\mathcal{P}_l$ per dimension. The total power of the channel power delay

profile $\{\mathcal{P}_0, \ldots, \mathcal{P}_{L-1}\}$ is normalized to be $\sigma_c^2 = 1$. Let $\mathbf{h}_m$ be the CIR matrix and denote $\mathbf{h}_{l,m}$ as the l-th matrix-valued CIR coefficient.

$$\mathbf{h}_m = \begin{bmatrix} \mathbf{h}_{0,m} \\ \vdots \\ \mathbf{h}_{L-1,m} \end{bmatrix}; \qquad \mathbf{h}_{l,m} = \begin{bmatrix} h_{11}(l,m) & \cdots & h_{1N_t}(l,m) \\ \vdots & \ddots & \vdots \\ h_{N_r1}(l,m) & \cdots & h_{N_rN_t}(l,m) \end{bmatrix}. \tag{2.5}$$

In addition, we assume that those N_tN_r geographically co-located multipath channels are independent in an environments full of scattering. In information-theoretic point of view [21, 22], it guarantees the capacity gain of MIMO systems. For the practical MIMO-OFDM systems, it enforces a lower limit on the shortest distance between multiple antennas at a portable receiver unit. If the correlation between those channels exists, the diversity gain from MIMO system will be reduced and hence system performance is degraded.

At the receive side, an N_r-dimensional complex baseband equivalent receive signal can be obtained by a matrix-based discrete-time convolution as

$$\vec{r}(mN_{tot}+n) = \sum_{l=0}^{L-1} \mathbf{h}_{l,m}\vec{u}(mN_{tot}+n-l) + \vec{v}(mN_{tot}+n), \tag{2.6}$$

where

$$\vec{r}(mN_{tot}+n) = \begin{bmatrix} r_1(mN_{tot}+n) \\ \vdots \\ r_{N_r}(mN_{tot}+n) \end{bmatrix} \quad \vec{v}(mN_{tot}+n) = \begin{bmatrix} v_1(mN_{tot}+n) \\ \vdots \\ v_{N_r}(mN_{tot}+n) \end{bmatrix}.$$

Note that $v_i(mN_{tot}+n)$ is assumed to be complex AWGN with zero mean and variance of $\frac{1}{2}\sigma_v^2$ per dimension. Therefore, the expected signal-to-noise ratio (SNR) per receive antenna is $\frac{N_t}{\sigma_v^2}$. In order to have a fair comparison with SISO systems, the power

per TX antenna should be scaled down by a factor of N_t. By stacking the received samples at discrete time instances, $\vec{r}(m)$ can be described by

$$\vec{r}(m) = \begin{bmatrix} \vec{r}(mN_{tot}) \\ \vdots \\ \vec{r}(mN_{tot} + N_{tot} - 1) \end{bmatrix}. \tag{2.7}$$

To combat the ISI, the first $N_g N_r$ elements of $\vec{r}(m)$ must be removed completely. The resulting ISI-free OFDM symbol $\vec{y}(m)$ is

$$\vec{y}(m) = \begin{bmatrix} \vec{y}(mN) \\ \vdots \\ \vec{y}(mN + N - 1) \end{bmatrix} = \mathbf{A}_{DeCP}\vec{r}(m), \tag{2.8}$$

where

$$\mathbf{A}_{DeCP} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_N \end{bmatrix} \otimes \mathbf{I}_{N_r}.$$

By exploiting the property that $\vec{u}(m)$ is a cyclic extension of $\vec{s}(m)$ so that cyclic discrete-time convolution is valid, the relation between $\vec{s}(m)$ and $\vec{y}(m)$ can be expressed as

$$\vec{y}(m) = \mathbf{h}_{m,Cir}\vec{s}(m) + \mathbf{A}_{DeCP}\vec{v}(m), \tag{2.9}$$

where $\mathbf{h}_{m,Cir}$ is an $NN_r \times NN_t$ block circulant matrix. In general, an $NN_r \times NN_t$ block circulant matrix is fully defined by its first $NN_r \times N_t$ block matrices. In our case, $\mathbf{h}_{m,Cir}$ is determined by

$$\begin{bmatrix} \mathbf{h}_{0,m} \\ \vdots \\ \mathbf{h}_{L-1,m} \\ \mathbf{0}_{(N-L)N_r \times N_t} \end{bmatrix}.$$

Finally taking FFT on the $\vec{y}(m)$ at the receiver, we obtain the frequency domain MIMO-OFDM baseband signal model

$$\begin{aligned}\vec{Y}(m) &= (\mathbf{F}_N \otimes \mathbf{I}_{N_r})\vec{y}(m) \\ &= (\mathbf{F}_N \otimes \mathbf{I}_{N_r})(\mathbf{h}_{m,Cir}\vec{s}(m) + \mathbf{A}_{DeCP}\vec{v}(m)) \\ &= (\tfrac{1}{N})(\mathbf{F}_N \otimes \mathbf{I}_{N_r})\mathbf{h}_{m,Cir}(\mathbf{F}_N^H \otimes \mathbf{I}_{N_t})\vec{S}(m) + (\mathbf{F}_N \otimes \mathbf{I}_{N_r})\mathbf{A}_{DeCP}\vec{v}(m) \\ &= \mathbf{H}_{m,diag}\vec{S}(m) + \vec{V}(m).\end{aligned} \tag{2.10}$$

In the above expression, $\vec{V}(m)$ represents the frequency domain noise, which is i.i.d. (independent and identically distributed) zero-mean and complex Gaussian random variable with variance $\frac{1}{2}\sigma_v^2$ per dimension, and $\mathbf{H}_{m,diag}$ is a block diagonal matrix which is given by

$$\mathbf{H}_{m,diag} = \begin{bmatrix} \mathbf{H}_{0,m} & & \\ & \ddots & \\ & & \mathbf{H}_{N-1,m} \end{bmatrix}.$$

The k-th block diagonal element is the frequency response of the MIMO channel at the k-th subcarrier and can be shown to be $\mathbf{H}_{k,m} = \sum_{l=0}^{L-1} \mathbf{h}_{l,m} e^{-j\frac{2\pi}{N}kl}$. So for that subcarrier, we may write it in a simpler form

$$\vec{Y}(mN+k) = \mathbf{H}_{k,m}\vec{S}(mN+k) + \vec{V}(mN+k), \tag{2.11}$$

where

$$\mathbf{H}_{k,m} = \begin{bmatrix} H_{11}(k,m) & \cdots & H_{1N_t}(k,m) \\ \vdots & \ddots & \vdots \\ H_{N_r1}(k,m) & \cdots & H_{N_rN_t}(k,m) \end{bmatrix}.$$

This leads to a flat-fading signal model per subcarrier and it is similar to the SISO signal model, except that $\mathbf{H}_{k,m}$ is an $N_r \times N_t$ matrix.

2.2.2 Preliminary Analysis

Based on those assumptions such as perfect synchronization and block fading, we end up with a compact and simple signal model for both the single antenna OFDM and MIMO-OFDM systems. Surely it is an ideal model that says, considering first a noise free scenario, the received signal on the k-th subcarrier is just a product (or matrix product for MIMO case) of the transmitted signal on the k-th subcarrier and the discrete-time channel frequency response at the k-th subcarrier. Noise in frequency domain can also be modeled as an additive term. When it comes to channel estimation for OFDM systems, this model is still valid since there is no ICI as we assume.

For channel estimation of MIMO-OFDM systems, it is appropriate to estimate the channel in time domain rather than in frequency domain because there are few parameters in the impulse response ($N_t N_r L$ coefficients) than in the frequency response ($N_t N_r N$ coefficients). Given the limited number of training data that can be sent to estimate the fast time-varying channel, limiting the number of parameters to be estimated would increase the accuracy of the estimation. This is the thrust of the estimation technique in this thesis. The estimation algorithm we propose is based on pilot tones, namely known data in the frequency domain. Since the signal model of OFDM in (2.11) is in the frequency domain too, it is necessary to find the relations between the CFR and the CIR. Discrete-time Fourier transform is a perfect tool we

can use to describe the relation. It is shown as

$$\mathbf{H}_m = \mathcal{F}_{NN_r} \begin{bmatrix} \mathbf{h}_m \\ \mathbf{0}_{(N-L)N_r \times N_t} \end{bmatrix},$$

where

$$\mathbf{H}_m = \begin{bmatrix} \mathbf{H}_{0,m} \\ \vdots \\ \mathbf{H}_{N-1,m} \end{bmatrix}; \qquad \mathcal{F}_{NN_r} = \mathbf{F}_N \otimes \mathbf{I}_{N_r}.$$

Since the channel length L is less than the FFT size N, only the first LN_r columns of FFT matrix $\mathcal{F}_{NN_r}$ are involved in calculation. It gives us another form to describe the relation as

$$\mathbf{H}_m = \mathcal{F}_{NN_r}(1 : N_r L)\mathbf{h}_m, \tag{2.12}$$

where $\mathcal{F}_{NN_r}(:, 1 : N_r L)$ is an $NN_r \times N_r L$ submatrix of $\mathbf{F}_N$, consisting of its first $N_r L$ columns. $\mathcal{F}_{NN_r}(:, 1 : N_r L)$ is a 'tall' matrix and its left inverse exists. That implies the equation in (2.12) is an overdetermined system. To determine $\mathbf{h}_m$, we can easily multiply the left inverse of $\mathcal{F}_{NN_r}(:, 1 : N_r L)$ in the two sides of the equation. This requires full information for the channel frequency response matrix $\mathbf{H}_m$. That is not necessarily to be true. Actually if we know L of the N matrices $\{\mathbf{H}_{0,m}, \ldots, \mathbf{H}_{N-1,m}\}$, then $\mathbf{h}_m$ can be calculated. For example, in the SISO case, if we know the channel frequency response at any L subcarriers $\{H_{k_1,m}, \ldots, H_{k_L,m}\}$, then the channel impulse response $\vec{h}(m)$ can be uniquely determined. This is the base for pilot-tone based channel estimation of OFDM systems. Pilot-tones are the selected subcarriers over which the training data are sent. The question then arises as to which tones should be used as pilot-tones and the impact of pilot-tones selection on the quality of estimation.

Cioffi's paper [49] addressed this issue first that one should choose the sets of equally-spaced tones as pilot tones, to avoid the noise enhancement effect in interpolating the channel impulse response from the frequency response. Assume that $N = mL$ and the integer $m > 1$. This is a realistic assumption since the OFDM block size N is often chosen to be 128, 256 or even a larger value and the channel length of MIMO-OFDM channel is usually not greater than 30. For the typical urban (TU) model [47] of delay profile with RMS delay $\tau_{rms} = 1.06\mu s$, the channel length is $L = \tau_{rms} \times 20\text{MHz}+1 \approx 23$ in an 802.11a system with a bandwidth of 20MHz. In systems like DVB-T and WiMax [40, 41], N is even a much bigger integer. Since $N = ML$, there could be m equally-sized pilot tones sets. Define

$$\mathbf{H}_m^{(p)} = \begin{bmatrix} \mathbf{H}_{p,m} \\ \vdots \\ \mathbf{H}_{p+(L-1)M,m} \end{bmatrix} \qquad \mathbf{W}_N^{(p)} = \begin{bmatrix} 1 & & \\ & W_N^p & \\ & & W_N^{p(L-1)} \end{bmatrix} \otimes I_{N_r}, \tag{2.13}$$

where p is any integer such that $0 \leq p \leq m-1$ and $W_N = e^{-j\frac{2\pi}{N}}$. Clearly $\mathbf{H}_m^{(p)}$ is the p-th down-sampled version of $\mathbf{H}_m^{(p)}$, and $\mathbf{W}_N^{(p)}$ simply acts as a shift operator of order p. The CFR matrix $\mathbf{H}_m$ can be decomposed into M disjoint down-sampled submatrices $\{\mathbf{H}_m^{(p)}\}_{p=0}^{M-1}$, each composed of L equally-spaced CFR sample matrices. It can be verified via straightforward calculation that

$$\mathbf{H}_m^{(p)} = \mathcal{F}_{LN_r} \mathbf{W}_N^{(p)} \mathbf{h}_m \qquad p = 0, 1, \cdots, M-1, \tag{2.14}$$

where $\mathcal{F}_{LN_r}$ is a $LN_r \times LN_r$ DFT matrix. It indicates that the channel state information represented by $\mathbf{h}_m$ can be obtained from a down-sampled version of $\mathbf{H}_m$, i.e.,

$\mathbf{H}_m^{(p)}$, which only requires us to probe the unknown channel frequency response with some training data on the selected p-th pilot-tones set. The procedure of pilot-tone based channel estimation is illustrated in Figure 2.2.

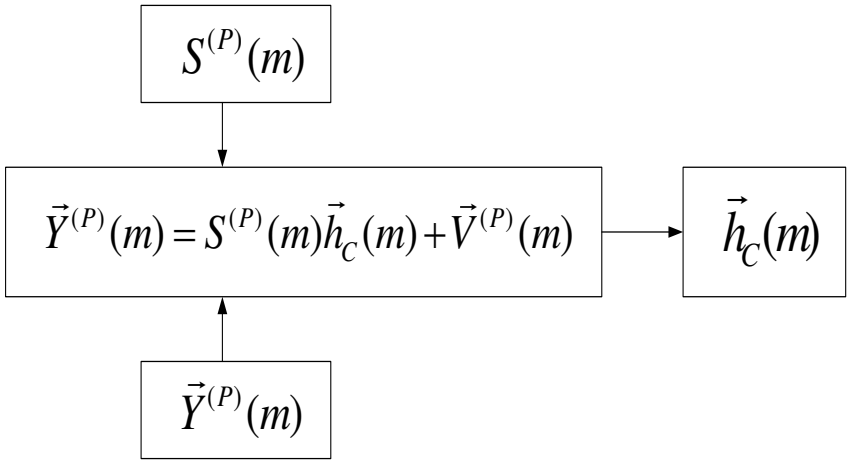

Figure 2.2: The concept of pilot-based channel estimation

And it is also true that

$$\mathbf{H}_m^{(p)}(:,i) = \mathcal{F}_{LN_r}\mathbf{W}_N^{(p)}\mathbf{h}_m(:,i), \tag{2.15}$$

where $\mathbf{H}_m^{(p)}(:,i)$ and $\mathbf{h}_m(:,i)$ are the i-th column of $\mathbf{H}_m^{(p)}$ and $\mathbf{h}_m$ respectively and $1 \leq i \leq N_t$. After discussing the relation between the CIR $\mathbf{h}_m$ and the p-th down-sampled CFR $\mathbf{H}_m$, we return to the input-output relationship of MIMO-OFDM system

$$\vec{Y}(mN+k) = \mathbf{H}_{k,m}\vec{S}(mN+k) + \vec{V}(mN+k), \tag{2.16}$$

where

$$\vec{Y}(mN+k) = \begin{bmatrix} Y_1(mN+k) \\ \vdots \\ Y_{N_r}(mN+k) \end{bmatrix}; \quad \vec{S}(mN+k) = \begin{bmatrix} S_1(mN+k) \\ \vdots \\ S_{N_t}(mN+k) \end{bmatrix}; \quad \vec{V}(mN+k) = \begin{bmatrix} V_1(mN+k) \\ \vdots \\ V_{N_r}(mN+k) \end{bmatrix}.$$

are the received signal, the transmitted signal and the noise term respectively as defined in the previous section. They are repeated here for convenience. In order to get a useful form for channel estimation based on pilot-tones, we have to manipulate the expression in (2.16) so that the transmitted signal and the CFR terms exchange their position in the product. (2.16) can be equivalently rewritten as

$$\vec{Y}(mN+k) = S_1(mN+k)\mathbf{H}_{k,m}(:,1) + \cdots + S_{N_t}(mN+k)\mathbf{H}_{k,m}(:,N_t) + \vec{V}(mN+k). \tag{2.17}$$

Basically we transform the product of a matrix and a vector into a summation of products of a scalar and a vector. The noise term remains unchanged. This transformation is specified to the k-th subcarrier. If we consider all the N subcarriers, we need stack $\{\vec{Y}(mN+k)\}$'s and $\{\mathbf{H}_m(:,i)\}$'s together and construct a block diagonal matrix for the $\{\vec{S}(mN+k)\}$'s. It can be shown that

$$\vec{Y}(m) = \mathbf{S}_{diag,1}(m)\mathbf{H}_m(:,1) + \cdots + \mathbf{S}_{diag,N_t}(m)\mathbf{H}_m(:,N_t) + \vec{V}(m), \tag{2.18}$$

where $\vec{Y}(m)$ and $\vec{V}(m)$ are the received signal and the noise term respectively given by

$$\vec{Y}(m) = \begin{bmatrix} \vec{Y}(mN) \\ \vdots \\ \vec{Y}(mN+N-1) \end{bmatrix}; \quad \vec{V}(m) = \begin{bmatrix} \vec{V}(mN) \\ \vdots \\ \vec{V}(mN+N-1) \end{bmatrix}; \quad \mathbf{H}_m(:,i) = \begin{bmatrix} \mathbf{H}_{0,m}(:,i) \\ \vdots \\ \mathbf{H}_{N-1m}(:,i) \end{bmatrix},$$

and

$$\mathbf{S}_{diag,i}(m) = \begin{bmatrix} S_{diag,i}(mN) & & \\ & \ddots & \\ & & S_{diag,i}(mN+N-1) \end{bmatrix} \otimes \mathbf{I}_{N_r}; \quad 1 \le i \le N_t.$$

Here the dimensions of the above column vectors and matrices are very large, for instance, $\vec{Y}(m)$ is an $NN_r \times 1$ column vector. The computational load, however, is

not changed since $\mathbf{S}_{diag,i}(m)$ is a block diagonal matrix, compared to the expression in (2.10).

As proved in [49], pilot-tones should be equally-powered and equally-spaced to achieve the MMSE (minimum mean squared error) of channel estimation. Let $\{S_i(mN+p), S_i(mN+M+p), \cdots, S_i(mN+(L-1)M+p)\}$ represent a set of L pilot-tones with index p which are transmitted simultaneously along with the other $N-L$ data signals at the m-th block from the i-th antenna. Obviously one pilot-tone is placed every M subcarriers in one OFDM block. Hence we can also have a down-sampled version of equation (2.18) by selecting a sampled element every M subcarriers. Since we assume that there is no ICI, we can neglect the data symbol which are transmitted together with pilot symbol. We only consider the p-th set of pilot-tones on the p-th, the $(p+M)$-th,... and the $(p+(L-1)M)$-th subcarriers, and so are the received signals. It turns out to be

$$\vec{Y}^{(p)}(m) = \mathbf{S}^{(p)}_{diag,1}(m)\mathbf{H}^{(p)}_m(:,1) + \cdots + \mathbf{S}^{(p)}_{diag,N_t}(m)\mathbf{H}^{(p)}_m(:,N_t) + \vec{V}^{(p)}(m), \qquad (2.19)$$

where

$$\vec{Y}^{(p)}(m) = \begin{bmatrix} \vec{Y}(mN+p) \\ \vdots \\ \vec{Y}(mN+(L-1)M+p) \end{bmatrix}; \quad \vec{V}^{(p)}(m) = \begin{bmatrix} \vec{V}(mN+p) \\ \vdots \\ \vec{V}(mN+(L-1)M+p) \end{bmatrix}; \quad \mathbf{H}^{(p)}_m(:,i) = \begin{bmatrix} \mathbf{H}_{p,m}(:,i) \\ \vdots \\ \mathbf{H}_{(L-1)M+p,m}(:,i) \end{bmatrix},$$

and

$$\mathbf{S}^{(p)}_{diag,i}(m) = \begin{bmatrix} S_{diag,i}(mN+p) & & \\ & \ddots & \\ & & S_{diag,i}(mN+(L-1)M+p) \end{bmatrix} \otimes \mathbf{I}_{N_r}; \quad 1 \le i \le N_t$$

are all the p-th down-sampled versions. In the equation (2.19), we obtain the relation between $\vec{Y}^{(p)}(m)$ and $\mathbf{H}^{(p)}_m(:,i)$. To estimate the channel in time domain, we need

explicitly relate $\vec{Y}^{(p)}(m)$ with $\mathbf{h}_m$. Plugging (2.15) into (2.19) yields

$$\vec{Y}^{(p)}(m) = \mathbf{S}^{(p)}_{diag,1}(m)\mathcal{F}_{LN_r}\mathbf{W}^{(p)}_N\mathbf{h}_m(:,1) + \cdots + \mathbf{S}^{(p)}_{diag,N_t}(m)\mathcal{F}_{LN_r}\mathbf{W}^{(p)}_N\mathbf{h}_m(:,N_t) + \vec{V}^{(p)}(m). \tag{2.20}$$

To estimate those unknown $\{\mathbf{h}_m(:,1), \cdots, \mathbf{h}_m(:,N_t)\}$, one set of pilot-tones is not adequate for estimation. That is different from the SISO case in which any one of the M pilot-tone sets can be utilized to estimate the channel. For MIMO-OFDM channel estimation, we need, at least, N_t disjoint sets of pilot-tones indexed by $\{p_1, p_2, \ldots, p_{N_t}\}$. It is assumed that $N = ML$ and hence there are totally $M = N/L$ different sets. It indicates a constraint imposed on the selection of FFT size N for MIMO systesm, i.e., $N \geq N_t L$. This observation tallies with the result in [48]. In practice, the selection of N determines the number of subcarriers utilized in the system. For systems like WLAN and WiMax [39, 40], N is not very large because a larger N means narrower subcarrier spacing which may cause severe ICI. Furthermore, those systems often operate in a low SNR environments.

2.3 Channel Estimation and Pilot-tone Design

2.3.1 LS Channel Estimation

Assume that we have N_t disjoint sets of pilot-tones. Then we have the following observation equations.

$$\begin{aligned}
&\vec{Y}^{(p_1)}(m) = \mathbf{S}^{(p_1)}_{diag,1}(m)\mathcal{F}_{LN_r}\mathbf{W}^{(p_1)}_N\mathbf{h}_m(:,1) + \cdots + \mathbf{S}^{(p_1)}_{diag,N_t}(m)\mathcal{F}_{LN_r}\mathbf{W}^{(p_1)}_N\mathbf{h}_m(:,N_t) + \vec{V}^{(p_1)}(m)\\
&\vdots\\
&\vec{Y}^{(p_{N_t})}(m) = \mathbf{S}^{(p_{N_t})}_{diag,1}(m)\mathcal{F}_{LN_r}\mathbf{W}^{(p_{N_t})}_N\mathbf{h}_m(:,1) + \cdots + \mathbf{S}^{(p_{N_t})}_{diag,N_t}(m)\mathcal{F}_{LN_r}\mathbf{W}^{(p_{N_t})}_N\mathbf{h}_m(:,N_t) + \vec{V}^{(p_{N_t})}(m)
\end{aligned} \tag{2.21}$$

To use LS (least square) method for channel estimation, we usually put those observation equations into a matrix form. LS is a well-known method and widely used for

estimation. We choose LS rather than other methods like MMSE channel estimation for the simplicity of implementation. In a matrix form, it is described by

$$\vec{Y}^{(P)}(m) = \mathbf{S}^{(P)}(m)\vec{h}_C(m) + \vec{V}^{(P)}(m), \tag{2.22}$$

where

$$\vec{Y}^{(P)}(m) = \begin{bmatrix} \vec{Y}^{(p_1)}(m) \\ \vdots \\ \vec{Y}^{(p_{N_t})}(m) \end{bmatrix}; \quad \vec{h}_C(m) = \begin{bmatrix} \mathbf{h}_m(:,1) \\ \vdots \\ \mathbf{h}_m(:,N_t) \end{bmatrix}; \quad \vec{V}^{(P)}(m) = \begin{bmatrix} \vec{V}^{(p_1)}(m) \\ \vdots \\ \vec{V}^{(p_{N_t})}(m) \end{bmatrix},$$

and

$$\mathbf{S}^{(P)}(m) = \begin{bmatrix} \mathbf{S}^{(p_1)}_{diag,1}(m)\mathcal{F}_{LN_r}\mathbf{W}^{(p_1)}_N & \cdots & \mathbf{S}^{(p_1)}_{diag,N_t}(m)\mathcal{F}_{LN_r}\mathbf{W}^{(p_1)}_N \\ \vdots & \ddots & \vdots \\ \mathbf{S}^{(p_{N_t})}_{diag,1}(m)\mathcal{F}_{LN_r}\mathbf{W}^{(p_{N_t})}_N & \cdots & \mathbf{S}^{(p_{N_t})}_{diag,N_t}(m)\mathcal{F}_{LN_r}\mathbf{W}^{(p_{N_t})}_N \end{bmatrix}.$$

In the above expression, $\mathbf{S}^{(P)}(m)$ is an $N_tN_rL \times N_tN_rL$ square matrix, composed of N_t^2 pilot-tone block matrices $\{\mathbf{S}^{(p_i)}_{diag,j}(m)\}_{i,j=1}^{N_t}$. At each transmit antenna N_t sets of pilot-tones are transmitted with the same index $\{p_1, p_2, \cdots, p_{N_t}\}$. Assume that $N_t \leq M = \frac{N}{L}$. It can also be seen that the total number of unknown CIR parameters N_tN_rL cannot be greater than the total number of received signals NN_r, i.e., $NtNrL \leq NNr \Leftrightarrow N_tL \leq N \Leftrightarrow N_t \leq \frac{N}{L}$.

The standard solution to the LS channel estimates [50] is known as

$$\hat{\vec{h}}_{C,LS}(m) = [(\mathbf{S}^{(P)}(m))^H\mathbf{S}^{(P)}(m)]^{-1}(\mathbf{S}^{(P)}(m))^H\vec{Y}^{(P)}(m). \tag{2.23}$$

Obviously the matrix $\mathbf{S}^{(P)}(m)$ is of huge size and it has $N_t^2N_r^2L^2$ elements. Computation of the inverse for such a large size matrix is undesirable. Therefore, an intuitive solution is to design the square matrix $\mathbf{S}^{(P)}(m)$ such that $(\mathbf{S}^{(P)}(m))^H\mathbf{S}^{(P)}(m) =$

$\mathbf{S}^{(P)}(m)(\mathbf{S}^{(P)}(m))^H = aI_{N_tN_rL}$, $a \in \mathcal{R}^+$, or equivalently $\frac{1}{\sqrt{a}}\mathbf{S}^{(P)}(m)$ is a unitary matrix. Then the LS channel estimates can be easily obtained as

$$\hat{\vec{h}}_{C,LS}(m) = \vec{h}_{C,LS}(m) + \frac{1}{a}(\mathbf{S}^{(P)}(m))^H \vec{V}^{(P)}(m). \tag{2.24}$$

2.3.2 Pilot-tone Design

In order to have a simple and efficient LS algorithm for channel estimation, we have to design the square matrix $\mathbf{S}^{(P)}(m)$ deliberately. In this section, the design will be illustrated by a theorem and an example.

The preamble design discussed in [50] adopted Tarokh's approach [18] to space-time block code construction. It could be related to orthogonal design to which our pilot-tone design also has a connection. In each of the first N_t training blocks in a frame, a group of at least L pilot-tones are equally-placed and all the other tones are set to zeros. LS channel estimation can then be obtained based on the known pilot-tones. The channel is assumed to be unchanged for the rest of the whole frame. In a mobile environment, however, we cannot guarantee that the channel state information estimated at the m-th block still holds true at the $(m + N_t)$-th block. Hence the preamble design in [50] is not suitable to be applied to the fast time-varying channels. In addition to this common disadvantage, the training sequences designed in [48] have to satisfy a condition called *local orthogonality*. It requires that, for the N_t different training sequences with length N, they are orthogonal over the minimum set of elements for any starting position. The pilot design proposed in this paper aims to remove the disadvantage and the constraint mentioned above. It actually has its

roots to Table I in [16], but it is not implemented in space and time domain. On the contrary, it is accomplished in space and frequency domain. We explicitly connect pilot-tone design with space-frequency coding so that we have more insights on its design. Denote E_P as the fixed total power for all the pilot-tones at each transmit antenna. Then the power allocated on each pilot-tone is $\frac{E_P}{N_t L}$ since pilot-tones are all equalspaced and equalpowered. In some systems, the power of those pilot-tones could be larger than the power of data symbols for a better estimation of the wireless channel. We assume in our work that the pilot-tones and other data are all equally normalized such that the average power for all different mappings is the same. Our pilot-tone design is illustrated in the following theorem.

Theorem 2.1 *Let* $\mathbf{S}_{diag,j}^{(p_i)}(m) = \alpha_{p_i,j} I_{LN_r}$, $|\alpha_{p_i,j}| = \sqrt{\frac{E_P}{N_t L}}$, $i, j = 1, 2, \cdots, N_t$, *then* $\frac{1}{\sqrt{E_P}}\mathbf{S}^{(P)}(m)$ *is a unitary matrix if*

$$\mathbf{S}_{SFC}^{(P)}(m) = \sqrt{\frac{L}{E_P}} \begin{bmatrix} \mathbf{S}_{diag,1}^{(p_1)}(m) & \cdots & \mathbf{S}_{diag,N_t}^{(p_1)}(m) \\ \vdots & \ddots & \vdots \\ \mathbf{S}_{diag,1}^{(p_{N_t})}(m) & \cdots & \mathbf{S}_{diag,N_t}^{(p_{N_t})}(m) \end{bmatrix}$$

is a unitary matrix.

Proof.

$$
\begin{aligned}
&\mathbf{S}^{(P)}(m) \\
&= \begin{bmatrix} \mathcal{F}_{LN_r}\mathbf{S}_{diag,1}^{(p_1)}(m)\mathbf{W}_N^{(p_1)} & \cdots & \mathcal{F}_{LN_r}\mathbf{S}_{diag,N_t}^{(p_1)}(m)\mathbf{W}_N^{(p_1)} \\ \vdots & \ddots & \vdots \\ \mathcal{F}_{LN_r}\mathbf{S}_{diag,1}^{(p_{N_t})}(m)\mathbf{W}_N^{(p_{N_t})} & \cdots & \mathcal{F}_{LN_r}\mathbf{S}_{diag,N_t}^{(p_{N_t})}(m)\mathbf{W}_N^{(p_{N_t})} \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{F}_{LN_r}\mathbf{W}_N^{(p_1)}\mathbf{S}_{diag,1}^{(p_1)}(m) & \cdots & \mathcal{F}_{LN_r}\mathbf{W}_N^{(p_1)}\mathbf{S}_{diag,N_t}^{(p_1)}(m) \\ \vdots & \ddots & \vdots \\ \mathcal{F}_{LN_r}\mathbf{W}_N^{(p_{N_t})}\mathbf{S}_{diag,1}^{(p_{N_t})}(m) & \cdots & \mathcal{F}_{LN_r}\mathbf{W}_N^{(p_{N_t})}\mathbf{S}_{diag,N_t}^{(p_{N_t})}(m) \end{bmatrix} \\
&= \mathbf{F}_{LN_r}\mathbf{W}_N^{(P)}(\sqrt{\tfrac{E_P}{L}})\mathbf{S}_{SFC}^{(P)}(m),
\end{aligned}
$$

where

$$
\mathbf{F}_{LN_r} = \begin{bmatrix} \mathcal{F}_{LN_r} & & \\ & \ddots & \\ & & \mathcal{F}_{LN_r} \end{bmatrix}, \mathbf{W}_N^{(P)} = \begin{bmatrix} \mathbf{W}_N^{(p_1)} & & \\ & \ddots & \\ & & \mathbf{W}_N^{(p_{N_t})} \end{bmatrix}.
$$

It is easy to see that $\mathcal{F}_{LN_r}\mathcal{F}_{LN_r}^H = \mathcal{F}_{LN_r}^H\mathcal{F}_{LN_r} = LI_{N_tN_rL}$ and $\mathbf{W}_N^{(P)}$ is a unitary matrix. Hence $\mathbf{S}^{(P)}(m)(\mathbf{S}^{(P)}(m))^H = (\mathbf{S}^{(P)}(m))^H\mathbf{S}^{(P)}(m) = E_PI_{N_tN_rL}$. This completes the proof. □

Clearly each of the N_t different pilot-tone sets has the same L elements. That is because, for example, $\mathbf{A}_{n\times n}\mathbf{B}_{n\times n} = \mathbf{B}_{n\times n}\mathbf{A}_{n\times n}$ if $\mathbf{B} = \mathbf{I}_n$. Or put it in another way, we can turn the product $\mathbf{AB}$ into $\mathbf{BA}$ by moving $\mathbf{B}$ to the front of $\mathbf{A}$. It is a simple manipulation of the mathematical derivation. In general, the product of two square matrices, AB is not equal to BA. But it turns out to be true if B is a square identity matrix. Then we can find that this assumption greatly simplifies the pilot-tone design for a MIMO-OFDM system with a large number of transmit antennas. It reduces to the design of a square orthogonal matrix. Hence we are more interested in the design

of $\mathbf{S}_{SFC}^{(P)}(m)$. First we consider a simple example with 2 transmit antennas and 2 receive antennas, i.e., $N_t = N_r = 2$ in the previous equations. Assume the channel length $L = 4$. By Theorem 2.1, we use Alamouti's structure [16]

$$\begin{bmatrix} x & y \\ -y^* & x^* \end{bmatrix}, \quad |x|^2 + |y|^2 = \tfrac{E_P}{4}, \quad x, y \in \mathcal{C}.$$

The above leads to the design

$$\mathbf{S}_{SFC}^{(P)}(m) = \sqrt{\frac{4}{E_P}} \begin{bmatrix} \mathbf{S}_{diag,1}^{(p_1)}(m) & \mathbf{S}_{diag,2}^{(p_1)}(m) \\ \mathbf{S}_{diag,1}^{(p_2)}(m) & \mathbf{S}_{diag,2}^{(p_2)}(m) \end{bmatrix}, \tag{2.25}$$

where

$$\begin{aligned} \mathbf{S}_{diag,1}^{(p_1)}(m)) &= xI_8, \qquad \mathbf{S}_{diag,2}^{(p_1)}(m)) = yI_8 \\ \mathbf{S}_{diag,1}^{(p_2)}(m) &= -y^*I_8, \quad \mathbf{S}_{diag,2}^{(p_2)}(m) = x^*I_8. \end{aligned}$$

The placement of pilot-tones in the example is shown in Figure 2.3. It can be seen in the figure that red and purple square boxes symbol the first and the second pilot-tone sets for TX antenna 1 respectively, and so are the green and light blue for TX antenna 2. They are all equally-spaced and the same color for each set implies that they are the same pilot symbols. For this example, there are total 16 pilot-tones and they are allocated to two TX antennas easily by our proposed method. The square matrix $\mathbf{S}_{SFC}^{(P)}(m)$ is actually a space-frequency code. In the column direction, it is signified by the TX antennas, namely the spatial domain; In the row direction, it is denoted by different pilot-tone sets, namely the frequency domain. Hence our design explicitly clarifies the connection between conventional pilot-tone design and the space-frequency code design [32, 33] aiming at performance enhancement.

When we have more than 2 transmit antennas, i.e., $N_t \geq 3$, it is also very easy

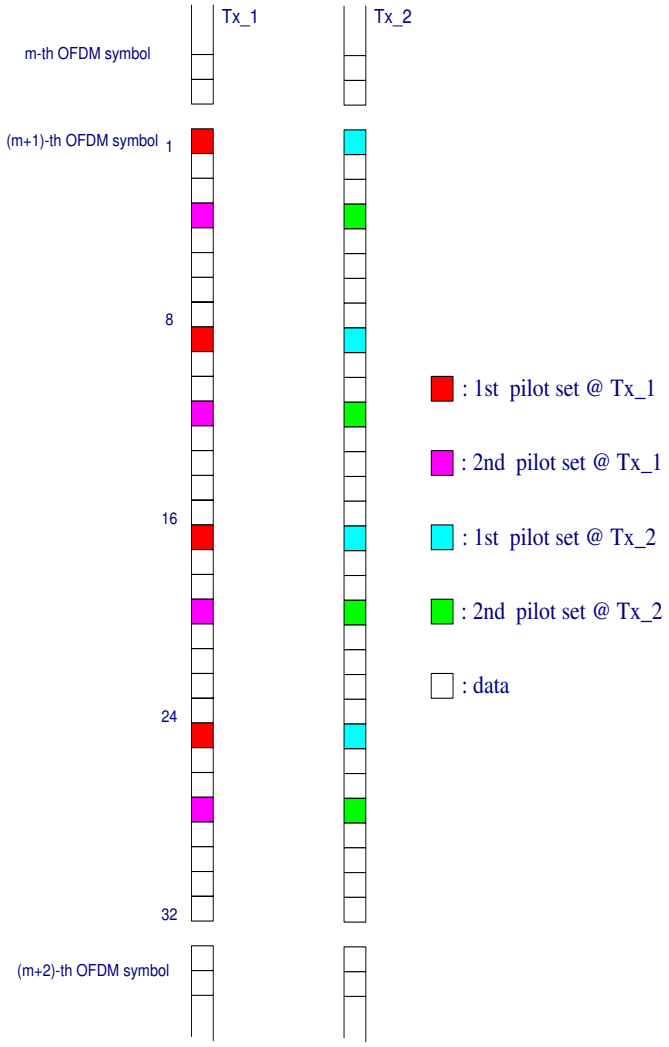

Figure 2.3: Pilot placement with $N_t = N_r = 2$

to design an $N_tN_rL \times N_tN_rL$ unitary matrix $\mathbf{S}_{SFC}^{(P)}(m)$. Based on the assumption in Theorem 2.1 that all the pilot-tones within one set are all the same, the design of $\mathbf{S}_{SFC}^{(P)}(m)$ can be simplified to the design of an $N_t \times N_t$ unitary matrix S and the complexity is reduced from N_tN_rL to N_t:

$$S = \sqrt{\frac{L}{E_P}} \begin{bmatrix} \alpha_{p_1,1} & \cdots & \alpha_{p_1,N_t} \\ \vdots & \ddots & \vdots \\ \alpha_{p_{N_t},1} & \cdots & \alpha_{p_{N_t},N_t} \end{bmatrix}_{N_t \times N_t}.$$

Choose $\alpha_{p_i,j} = \sqrt{\frac{E_P}{LN_t}} e^{-\tilde{j}\frac{2\pi}{N_t}ij}$, $\forall i,j \in \{1,2,\ldots,N_t\}$, $\tilde{j} = \sqrt{-1}$. Then S can be shown to be a unitary matrix. Basically it is very close to an N_t-point FFT matrix. After obtaining the $\{\alpha_{p_i,j}\}_{i,j=1}^{N_t}$, $\mathbf{S}_{SFC}^{(P)}(m)$ can be easily constructed from Theorem 2.1 by mapping a scalar to a diagonal matrix with its diagonal elements all equal to that scalar.

2.3.3 Performance Analysis

With the fixed total power E_P, the pilot-tones designed in the previous section can be shown to be optimal in the sense that it achieves the minimum mean squared error of the channel estimation. This is shown in the following. From (2.24), MSE of channel estimates $\hat{\vec{h}}_{C,LS}(m)$ is given by

$$\begin{aligned} \mathbf{MSE}_m &= \tfrac{1}{N_tN_rL}\mathcal{E}\{\|\hat{\vec{h}}_{C,LS}(m) - \vec{h}_{C,LS}(m)\|^2\} \\ &= \tfrac{1}{E_P^2N_tN_rL}\mathcal{E}\{\|(\mathbf{S}^{(P)}(m))^H\vec{V}^{(P)}(m)\|^2\} \\ &= \tfrac{1}{E_P^2N_tN_rL}tr\{(\mathbf{S}^{(P)}(m))^H\mathcal{E}[\vec{V}^{(P)}(m)\vec{V}^{(P)}(m)^H]\mathbf{S}^{(P)}(m)\} \\ &= \tfrac{\sigma_n^2}{E_P^2N_tN_rL}tr\{(\mathbf{S}^{(P)}(m))^H\mathbf{I}_{N_tN_rL}\mathbf{S}^{(P)}(m)\}. \end{aligned} \tag{2.26}$$

Since $\mathbf{S}^{(P)}(m)(\mathbf{S}^{(P)}(m))^H = (\mathbf{S}^{(P)}(m))^H\mathbf{S}^{(P)}(m) = E_PI_{N_tN_rL}$, then MSE achieves its minimum as $\mathbf{MSE}_{min} = \frac{\sigma_n^2}{E_P}$. At this point, we can find that the unitary matrix design

not only reduces the complexity of the channel estimator, but also ensures that it has the least estimation error, if the pilit-tones have fixed transmit power.

2.4 An Illustrative Example and Concluding Remarks

2.4.1 Comparison With Known Result

In this section, we demonstrate the performance of the proposed channel estimation based on our optimal pilot-tone design through computer simulations. In order to have a clear look at the performance improvement, other channel estimation technique [50] is also simulated. We consider a typical MIMO-OFDM system with 2 transmit antennas and 2 receive antennas. The OFDM block size is chosen as $N = 128$ and a CP with length of 16 is prepended to the beginning of each OFDM symbol. The four sub-channel paths denoted by $\{h_{11}, h_{12}, h_{21}, h_{22}\}$ are assumed to be independent to each other and have a CIR with length $L = 16$ individually. Those CIR coefficients in each sub-channel are simulated by the Jakes' model [51]. Our simulation is conducted in two ways:

- Method I: Place two sets of $L = 16$ pilot-tones into each OFDM block and the pilot-tones are equally-spaced and equally-powered as shown in Figure 2.3;
- Method II: Set the first two OFDM blocks of each data frame, which includes ten OFDM blocks, as preamble. Put $L = 16$ equally-spaced and equally-powered

pilot-tones into each of the first two preamble block and set all the other tones as zeros. (see [50] for detailes).

To illustrate the mobile environments, different Doppler shifts are simulated as $f_d = 5, 20, 40, 100$ and 200 Hz. The performance of the system is measured in terms of the MSE of the two different channel estimation schemes mentioned above and the symbol error rate (SER) versus SNR. For a reliable simulation, total 10,000 frames are transmitted for each test. Then the average values of MSE and SER are taken as the measurements. In Figure 2.4, the Doppler shift is 5 Hz and the two curves marked with "known channel" serve as the performance bound since we know the channel state information exactly. This is totally unrealistic and is just for the purpose of comparison. We can find the two curves corresponding to both RX antenna 1 and RX antenna 2 are nearly merged together. This matches our expectation since there is no difference between the two receive antennas statistically. It also can be found that the two curves generated by channel estimation based on our optimal pilot-tone design is close to the performance bound, just a narrow gap between them due to the ever-existing channel estimation error. On the contrast, the two curves generated by channel estimation based on the technique in [50] is far away from the performance limit, even with a large SNR. It justifies our point that the method based on preamble at the beginning of a frame is not applicable to a fast varying wireless channel. Through Figure 2.4 to Figure 2.6, the performance of the system based on the proposed pilot-tone design does not change a lot since it keeps tracking

the channel by the pilot-tones in each OFDM block. The difference between the two estimation schemes is illustrated in the MSE plots. In Figure 2.7, for a fixed SNR value, the curves for different Doppler spreads do not change that much and that implies that the method we proposed is able to track the fast time-varying channel. For a specific SNR value, the curves in Figure 2.8 do change along with the different Doppler shifts. It can be seen that the estimation error when $f_d = 200$ Hz is much larger than the one when $f_d = 5$ Hz in Figure 2.8. It indicates that the method based on preambles works poorly when Doppler spread is small, and does not work when the channel is changing quickly.

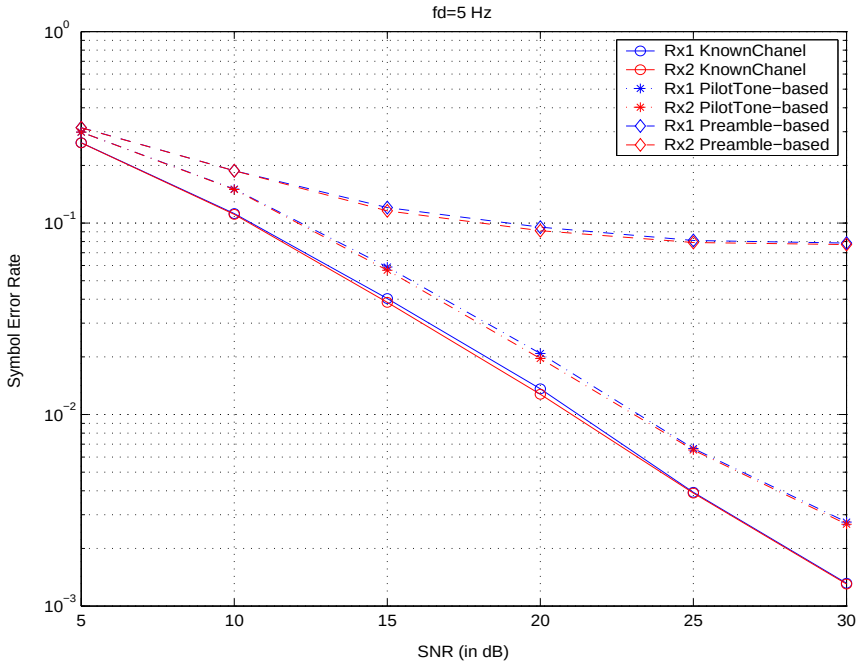

Figure 2.4: Symbol error rate versus SNR with Doppler shift=5 Hz

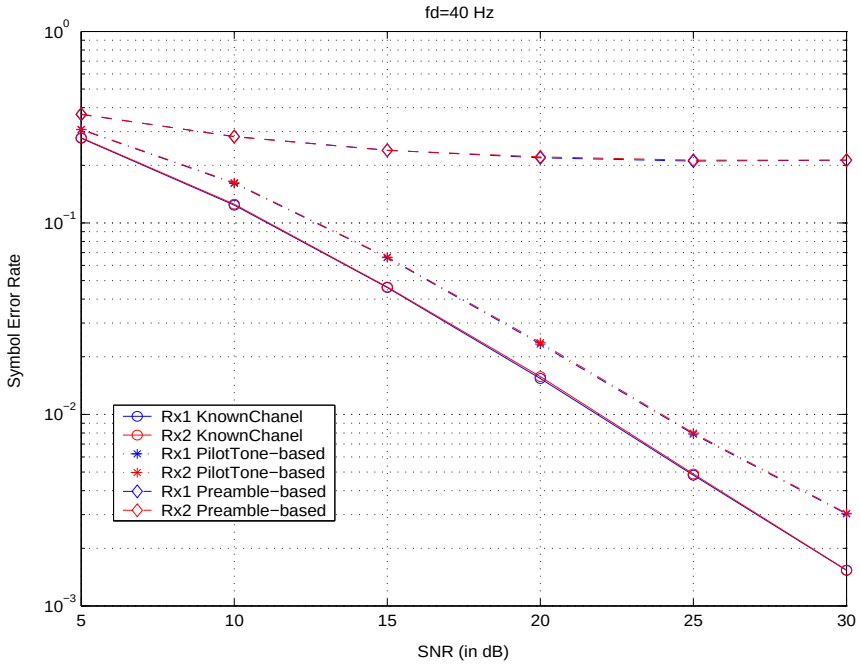

Figure 2.5: Symbol error rate versus SNR with Doppler shift=40 Hz

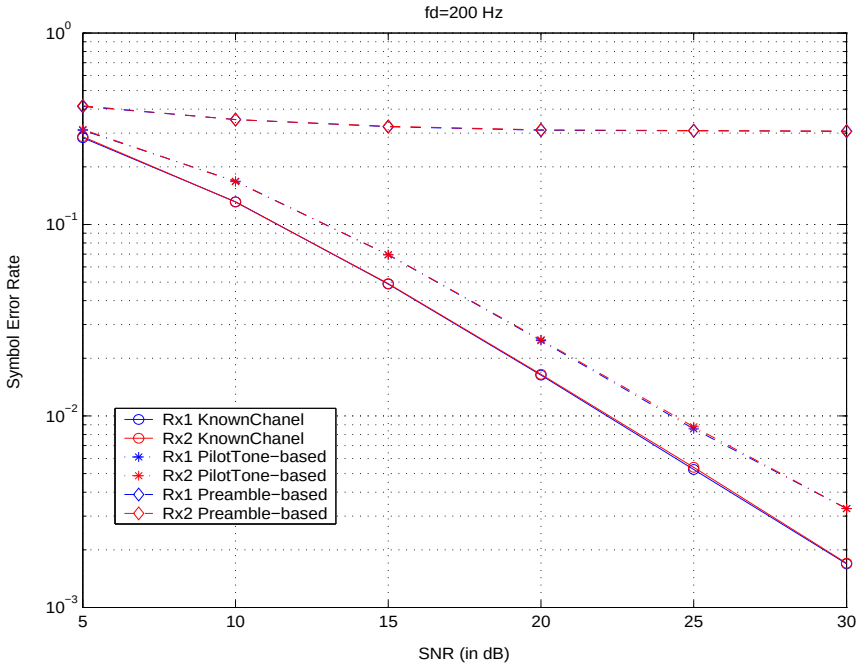

Figure 2.6: Symbol error rate versus SNR with Doppler shift=200 Hz

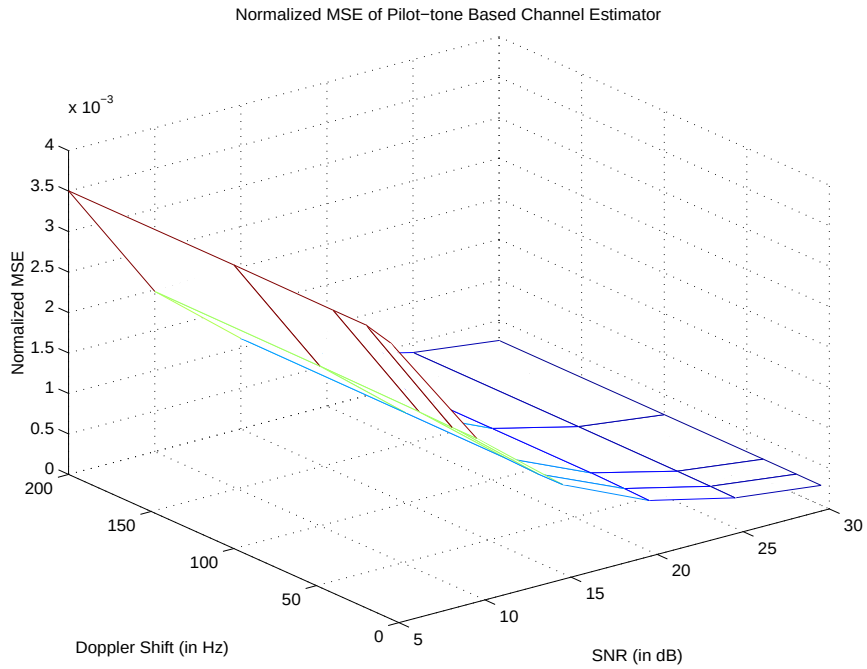

Figure 2.7: Normalized MSE of channel estimation based on optimal pilot-tone design

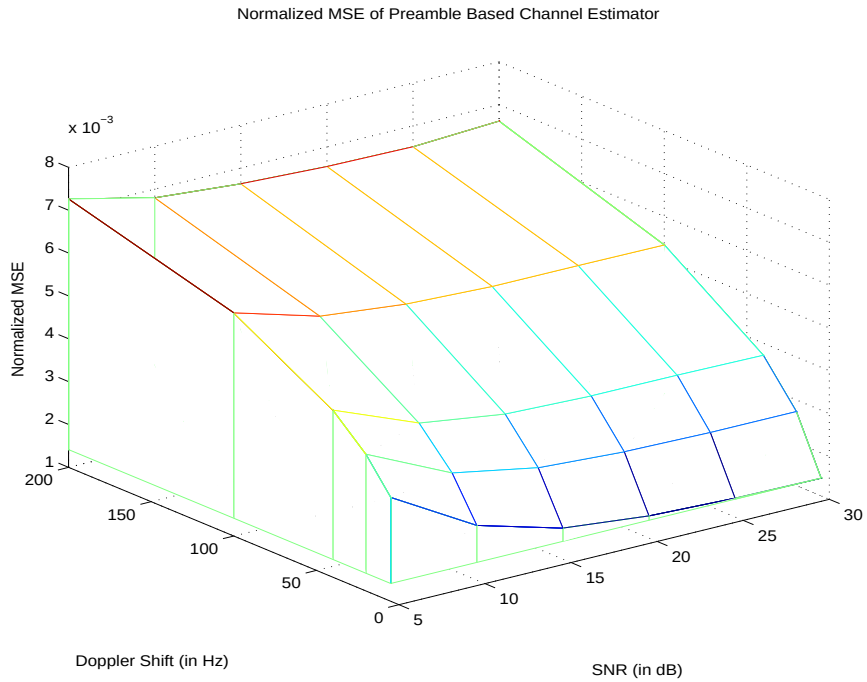

Figure 2.8: Normalized MSE of channel estimation based on preamble design

2.4.2 Chapter Summary

We presented a new optimal pilot-tone design for MIMO-OFDM channel estimation. N_t sets of L pilot-tones coded in $\mathbf{S}_{SFC}^{(P)}(m)$ are transmitted at each antenna simultaneously and the channel can be estimated optimally. The main advantage is rooted in its ability to handle fast time-varying system since channel can be estimated at each OFDM block and its simpleness since the orthogonal design makes the MIMO system be easily processed in a parallel way.

For an $N_t \times N_r$ MIMO system, the complexity of any kinds of signal processing algorithms at the physical layer is increased usually by a factor of $N_t N_r$. To name a few, channel estimation, carrier frequency offset estimation and correction and IQ imbalance compensation all become very challenging in MIMO case. In this chapter, we provide solutions to the following "how" questions. How many pilot tones are needed? How are they placed in one OFDM block? Most importantly, how fast can channel estimation be accomplished? We propose a pilot tone design for MIMO-OFDM channel estimation that N_t disjoint sets of pilot tones are placed on one OFDM block at each transmit antenna. For each pilot tone set, it has L pilot tones which are equally-spaced and equally-powered. The pilot tones from different transmit antennas comprise a unitary matrix and then a simple least square estimation of the MIMO channel is easily implemented by taking advantage of the unitarity of the pilot tone matrix. There is no need to compute the inverse of large-size matrix which is usually required by LS algorithm.

In a highly mobile environment, like a mobile user in a vehicle riding at more than 100km/hr, the wireless channel may change within one or a small number of symbols. For example as in [30], in IEEE 802.16-2004 Standard with $N = 256$, $G = 44$ (N: FFT size; G: guard interval) and 3.5MHz full bandwidth, the symbol duration is about 73 microseconds. For a user in a vehicle traveling with 100km/hr, the channel coherent time is about 1100 microseconds. That means the wireless channel varies after around 15 symbols. In a real-time communication scenario, the information packet could contain hundreds of data symbols or even more. Our scheme is proposed in this chapter that we distribute the pilot symbols in the preamble to each OFDM block for channel estimation. Since the pilot tones are placed on each OFDM block, the channel state information can be estimated accurately and quickly, no matter how fast the channel condition is varying. It is fair to point out that we may have a higher overhead rate compared to the methods in the literature. Therefore our pilot design can also be applied to a slow time-varying channel by placing pilot tones on every a few number of OFDM blocks. That can reduce the channel throughput loss.

The orthogonal pilot tone matrix is indeed a space-frequency code. The row direction of the matrix stands for different pilot tone sets in the frequency domain, and the column direction represents the individual transmit antennas in spatial domain. And it can be readily extended to an $N_t \times N_r$ MIMO system by constructing an $N_t \times N_t$ orthogonal matrix. With this explicit relation to space-frequency code, the

design of pilot tone matrix for MIMO-OFDM channel estimation can be conducted in a more broad perspective.

Chapter 3

Wireless Location for OFDM-based Systems

3.1 Introduction

Wireless networks are primarily designed and deployed for voice and data communications. The widespread availability of wireless nodes, however, makes it feasible to utilize these networks for wireless location purpose as an alternative to the GPS (global positioning system) location service. It is expected that location-based applications will play an important role in future wireless markets. The commercially available location technology is implemented on cellular networks and WLAN, such as E911 (Enhanced 911) and indoor positioning with WiFi (wireless fidelity). In this dissertation, we are investigating wireless location technology aimed at a different network, i.e., WiMax system.

3.1.1 Overview of WiMax

WiMax is an acronym for Worldwide Interoperatability for Microwave Access. It is not only a technical term indicating a new wireless broadband technology, but also is

referred to as a series of new products working on this network. The real WiMax-based wireless gears do not come to the market yet. But people are already very familiar with the WiFi-based products such as notebook wireless cards and wireless routers from Linksys, D-Link and Belkin, while they are checking their emails or surfing on Internet wirelessly on campus or at airports, hotels, bookstores and coffee shops. WiFi stands for Wireless Fidelity and it is the first available technology for WLAN and wireless home networking. However it is constrained by its limited coverage of about 50-100 meters and relatively low data rate. Different from WiFi, WiMax is another new broadband wireless access technology that provides very high data throughput over long distance in a point-to-multipoint and line of sight (LOS) or non-line of sight (NLOS) environments. In terms of the coverage, WiMax can provide seamless wireless services up to 20 or 30 miles away from the base station. It also has an IEEE name 802.16-2004. It is this IEEE standard that defines the specifics of air interface of WiMax.

WiMax Standards

Actually microwave access is not a new technology for broadband systems. Proprietary point-to-multipoint broadband access products from companies like Alcatel and Siemens have existed for decades. They did not get their popularity because they are extremely proprietary. Today's WiMax is attempting to standardize the technology to reduce the cost and to increase the range of applications. The current standard for WiMax is IEEE std 802.16-2004. It can be easily downloaded at IEEE website. With

its approval in June 2004, it renders the previous standard IEEE std 802.16-2001 and its two amendments 802.16a and 802.16c obsolete. Now IEEE 802.16-2004 can only address the fixed broadband systems. IEEE 802.16 Task Group e is working on an amendment to add mobility component to the standard. The new standard may be named as IEEE 802.16e.

WiMax Applications

We have seen a lot of marketing efforts on WiMax applications at conferences, exhibitions and other media. People are wondering if it is a must technology in the near future. Let's have a look at the fact that what kind of broadband services we can have today. We usually resort to a landline connection with T1, DSL and cable modems. WiMax or 802.16 is proposed to address the first mile/last mile wireless connection in a metropolitan area network. It can change the last-mile connection as much as 802.11 did for the change of the last hundred feet connection. It may change not only for the rural areas, but also for anyplace where the cost of laying or upgrading landline to broadband capabilities is prohibitively expensive. WiMax's primary use will most likely come in the form of metropolitan area network. In terms of services and applications, it is different from the traditional WiFi standards which include 802.11a, 802.11b and 802.11g. The WiFi technology with a maximum range of 800 feet outdoors mainly intend to be used in local area networks to provide services for residential homes, for public hot spots like airports, hotels and coffee shops, and for small business buildings. With its much longer range, in theory WiMax can reach a

maximum of 31 miles, and WiMax can provide broadband services to thousands of homes in a metropolitan area. Imagine that a broadband service provider can serve thousands of residential homes, small and large scale business buildings without the cost of laying out physically running lines and dispatching the technicians for installations and maintenance of the lines. The savings will push them to choose WiMax and to reduce the charge fees for their customers. Another driving force for WiMax is its speed. It can transfer the data with a rate up to 70 Mbps which is equivalent to almost 60 T1 lines. Combining its long range with the high-speed, it is why the application of WiMax is endless. All of these sound great enough though, the real WiMax products are not commercially available in the market yet. There are only some pre-WiMax products based on the standard coming up. But it will come soon. For example, Intel's PRO/Wirelss 5116 is a highly integrated IEEE std 802.16-2004 compliant system on chip for both licensed and license-exempt radio frequencies.

3.1.2 Overview to Wireless Location System

Wireless location refers to determination of the geographic coordinates, or even the velocity and the heading in a more general sense, of a mobile user/device in a cellular, WLAN or GPS environments. Usually wireless location technologies fall into two main categories: handset-based and network-based. In handset-based location systems [55], the mobile station equipped with extra electronics determines its location from signals received from the base stations or from the GPS satellites. In GPS-based estimations, the MS (mobile station) receives and measures the signal

parameters from at least four satellites of a currently existing constellation of 24 GPS satellites. The parameter of which the MS measures is the time for each satellite signal to reach the MS. GPS systems have a relatively higher degree of accuracy and they also provide global location information. However, embedding a GPS receiver into mobile devices leads to increased cost, size and battery consumption. It also requires the replacement of millions of mobile handsets that are already in the market with new GPS-featured handsets. In addition, the accuracy of GPS measurements degrades in urban and indoor environments. For these reasons, some wireless carriers may be unwilling to embrace GPS fully as the only location technology.

On the other hand, network-based location technology relies on the ever existing network infrastructures to determine the position of a mobile user by measuring its signal parameters when received at the network BSs (base stations). This may require some hardware upgrade or installation at the BSs, but the cost can be shared by a huge number of mobile subscribers and it does not affect the users in using their mobile devices. In this technology, the BSs measure the signal transmitted from an MS and relay them to a central site for further processing and data fusion to provide an estimate of the MS location. Network-based technologies have the significant advantage that the MS is not involved in the location-finding process; thus the technology does not require modifications to existing handsets. However, unlike GPS location systems, many aspects of network-based location are not yet fully studied. In Figure 3.1, network-based wireless location technology is illustrated.

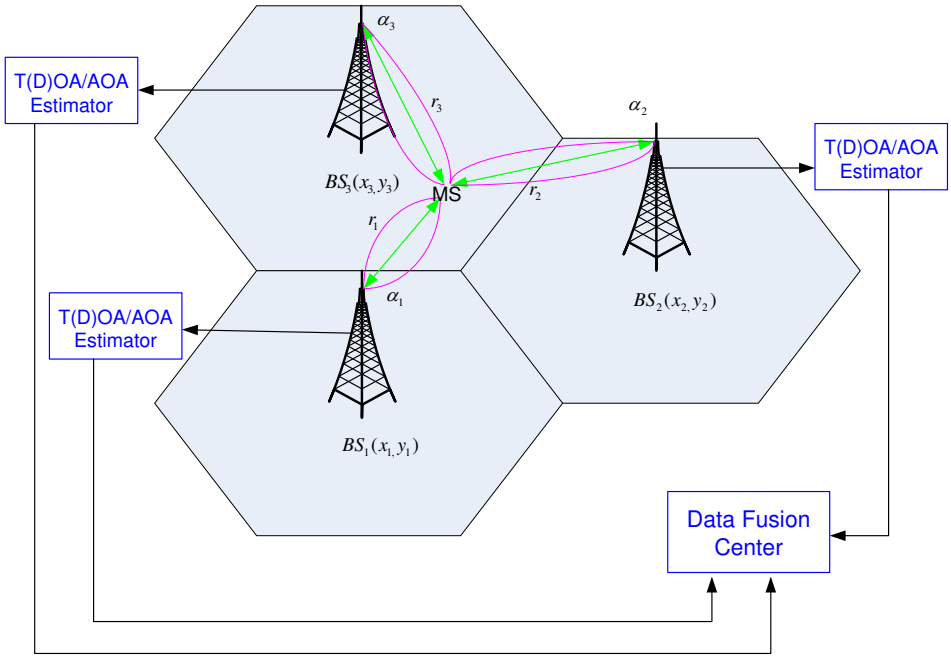

Figure 3.1: Network-based wireless location technology (outdoor environments)

Network-based wireless location technology gains more recognition with the increasing number of wireless subscribers and the demands for some location-oriented services such as E911. It is estimated [56, 57] that location based service will generate annual revenues in the order of $ 15B worldwide. In U.S. alone, about 170 million mobile subscribers are expected to become covered by the FCC mandated location accuracy for emergency services. The following is a partial list of applications that will be enhanced by using wireless location information [58].

- **E911**. Nowadays a high percentage of E911 calls are generated from mobile phones; the percentage is estimated [59, 60] to be at one third of all 911 calls (170,000 per day). These wireless 911 calls do no receive the same quality of emergency assistance as those fixed-network 911 calls enjoy. This is due to

the unknown position of the wireless 911 caller. To fix this problem, FCC issued an order on July 12, 1996 [59], requiring all wireless service providers to report accurate MS location to the E911 operator at the PSAP (public safety answering point). In the FCC order, it was mandated that within five years from the effective date of the order, October 1, 1996, wireless service providers must convey to the PSAP the location of the MS within 100 meters of its actual position for at least 67 percent of all wireless E911 calls. This FCC order has motivated considerable research efforts towards developing accurate wireless location algorithms for cellular networks and has led to significant enhancement to the wireless location technology.

- **Mobile advertising**. Location-specific advertising and marketing will benefit once the location information is available. For example, stores would be able to track customer locations and to attract them in by flashing customized coupons on their wireless devices [61]. In addition, a cellular phone or a PDA (personal digital assistant) could act as a smart handy mobile yellow pages on demand.

- **Asset tracking (indoor/outdoor)**. Wireless location technology can also assist in advanced public safety applications such as locating and retrieving lost children, patient, or even pets. In addition, it can be used to track personnel/assets in a hospital or a manufacturing site to provide a more efficient management of assets and personnel. One could also consider application such as smart and interactive tour guides, smart shopping guides that lead shoppers

based on their location in a store, smart traffic control in parking structures that guides cars to free parking slots. Department stores, enterprises, hospitals, manufacturing sites, malls, museums, and campuses are some of the potential end-users to benefit from the technology.

- **Fleet management**. Many fleet operators, such as police force, emergency vehicles, and other services including shuttle and taxi companies, can make use of the wireless location technology to track and operate their vehicles in an efficient way in order to minimize the response time. In addition, a large number of drivers on roads and highways carry cellular phones while driving. The wireless location technology can help track these phones, thus transforming them into sources of real-time traffic information that can be used to enhance transportation safety.

- **Location-based wireless security**. New location-based wireless security schemes can be developed to add a level of security to wireless networks against being intercepted or hacked into. By using location information, only people at certain specific areas could access certain files or databases through a WLAN.

- **Location sensitive billing**. Using the location information of wireless users, wireless service providers can offer variable-rate call plans or services that are based on the caller location.

3.1.3 Review of Data Fusion Methods

We assume that the location is specified by (x, y) for simplicity. As shown in Figure 3.1, data fusion center is to determine the mobile user location by exploring all the estimated signal parameters from BSs. The most common signal parameters are time, angle and amplitude of arrival of the MS signal. Therefore, different data fusion algorithms are proposed accordingly. The materials in this section are mainly based on the survey paper in [53].

- **Time**. By combining the estimates of the TOA (time of arrival) of the MS signal when received at the BSs, the MS location can be determined in a wireless network with three or more BSs. It is illustrated in Figure 3.3. Without loss of

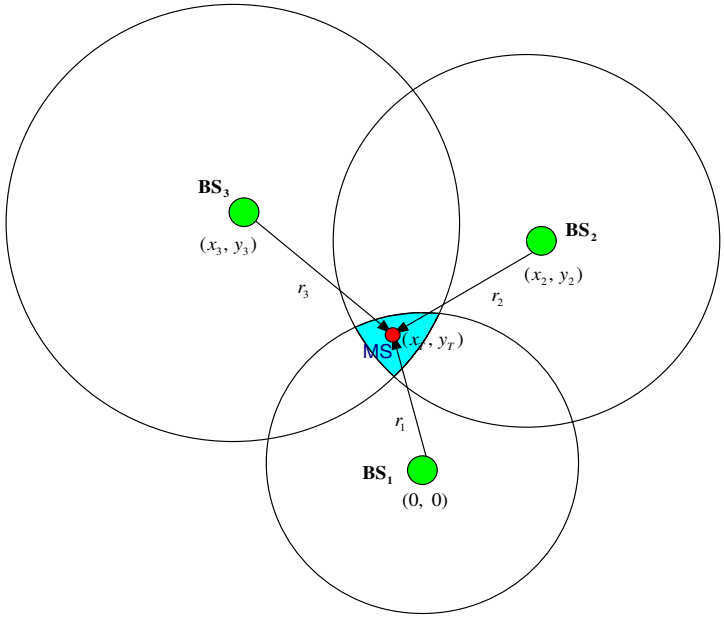

Figure 3.2: TOA/TDOA data fusion using three BSs

generality, the geometric coordinate of BS_1 is assumed to be $(0, 0)$. The location

of other BSs are denoted by $(x_k, y_k), k = 2, 3$. Obviously $x_1 = y_1 = 0$. Since the radio signal travels at the speed of light ($c = 3 \times 10^8 m/s$), the distance between the MS and BS_k is given by

$$R_{k,T} = (t_k - t_o)c, \tag{3.1}$$

where t_o is the time instant when the MS starts transmitting signal and t_k is the time of arrival of the MS signal at BS_k. The distances $\{R_{k,T}\}_{k=1}^{3}$ can be used to estimate the MS location (x_T, y_T) by solving the following set of equations

$$\begin{aligned} R_{1,T}^2 &= x_T^2 + y_T^2 \\ R_{2,T}^2 &= (x_2 - x_T)^2 + (y_2 - y_T)^2 \\ R_{3,T}^2 &= (x_3 - x_T)^2 + (y_3 - y_T)^2. \end{aligned} \tag{3.2}$$

To solve the above overdetermined nonlinear system of equations, we can reformulate (3.2) into an LS-type presentation by subtracting the first equation from the second and the third equations respectively. Hence the following equation is obtained

$$\begin{aligned} R_{2,T}^2 - R_{1,T}^2 &= x_2^2 + y_2^2 - 2(x_2 x_T + y_2 y_T) \\ R_{3,T}^2 - R_{1,T}^2 &= x_3^2 + y_3^2 - 2(x_3 x_T + y_3 y_T). \end{aligned} \tag{3.3}$$

In a matrix form, it can be rewritten as

$$\begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \end{bmatrix} \begin{bmatrix} x_T \\ y_T \end{bmatrix} = \frac{1}{2} \begin{bmatrix} R_2^2 - (R_{2,T}^2 - R_{1,T}^2) \\ R_3^2 - (R_{3,T}^2 - R_{1,T}^2) \end{bmatrix}, \tag{3.4}$$

where $R_k = \sqrt{x_k^2 + y_k^2}$ is the distance of the base station $\mathrm{BS_k}$ to the origin point in the coordinate, and clearly $R_1^2 = 0$. If we have more than three BSs, a compact form can be obtained in a similar way as

$$\vec{b} = \mathbf{A}\vec{\theta}, \tag{3.5}$$

where

$$\vec{b} = \frac{1}{2}\begin{bmatrix} R_2^2 - (R_{2,T}^2 - R_{1,T}^2) \\ R_3^2 - (R_{3,T}^2 - R_{1,T}^2) \\ R_4^2 - (R_{4,T}^2 - R_{1,T}^2) \\ \vdots \end{bmatrix}; \quad \mathbf{A} = \begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \\ \vdots & \end{bmatrix}; \quad \vec{\theta} = \begin{bmatrix} x_T \\ y_T \end{bmatrix}.$$

A standard LS estimation of $\vec{\theta}$ is given by

$$\hat{\vec{\theta}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{b}. \tag{3.6}$$

Note that $R_{1,T}^2$ is a function of x_T and y_T as defined in 3.2. Hence (3.6) only provides an intermediate solution and the estimates $\hat{x}_T$ and $\hat{y}_T$ can be obtained by solving the resultant quadratic equation. And clearly the TOA data fusion method requires perfect timing between the MS and the BSs since a small offset of a few microseconds between the MS clock and the BS clock will reflect into hundreds of meters of errors in location estimate. But the current wireless network standards only mandate tight timing synchronization among BSs [62]. The accuracy of TOA method is heavily dependent on the timing between BS and MS. There is another alternative of using the TDOA (time difference of arrivals) which help avoid the MS clock synchronization problem. Define the TDOA associated with the base station $\mathrm{BS_k}$ as $\Delta t_{k,1} = t_k - t_1$, i.e., the difference between the TOA of the MS signal at the BS $\mathrm{BS_k}$ and $\mathrm{BS_1}$. Then the difference between $R_{k,T}$ and $R_{1,T}$ can be related to $\Delta t_{k,1}$ as

$$\begin{aligned} \Delta R_{k,1} &= R_{k,T} - R_{1,T} \\ &= (t_k - t_o)c - (t_1 - t_o)c \\ &= \Delta t_{k,1} c. \end{aligned} \tag{3.7}$$

Clearly it is seen that the possible timing error on the MS clock t_o is canceled out. This insensitivity to t_o gives TDOA method the advantage over TOA. By substituting $R_{k,T}^2 = (R_{1,T} + \Delta R_{k,1})^2$ in (3.2) and rearranging some terms, we can obtain the following LS expression for any number of base stations as

$$R_{1,T}\vec{c} + \vec{d} = \mathbf{A}\vec{\theta}, \tag{3.8}$$

where

$$\vec{c} = \begin{bmatrix} -\Delta R_{2,1} \\ -\Delta R_{3,1} \\ -\Delta R_{4,1} \\ \vdots \end{bmatrix}; \quad \vec{d} = \frac{1}{2}\begin{bmatrix} R_2^2 - \Delta R_{2,1}^2 \\ R_3^2 - \Delta R_{3,1}^2 \\ R_4^2 - \Delta R_{4,1}^2 \\ \vdots \end{bmatrix}.$$

Notice that $R_{1,T} = x_T^2 + y_T^2$ is not known and hence only an intermediate solution can be obtained from the above LS formulation

$$\hat{\vec{\theta}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(R_{1,T}\vec{c} + \vec{d}). \tag{3.9}$$

Since

$$\|\hat{\vec{\theta}}\|^2 = R_{1,T}^2, \tag{3.10}$$

we can substitute (3.9) into (3.10) and solve $R_{1,T}$ from the resulting quadratic equation. A final solution for $\hat{\vec{\theta}}$ can be subsequently obtained by substitute the positive root of the quadratic equation into (3.9).

- **Angle**. The AOA (angle of arrival) can be obtained at a BS by using an antenna array. The direction of arrival of the MS signal can be calculated by measuring the phase difference between the antenna array elements or by measuring the power spectral density across the antenna array in what is known

as beamforming [64]. Intuitively, the MS location can be estimated by combining the AOA estimates from two BSs as shown in Figure 3.3. Compared

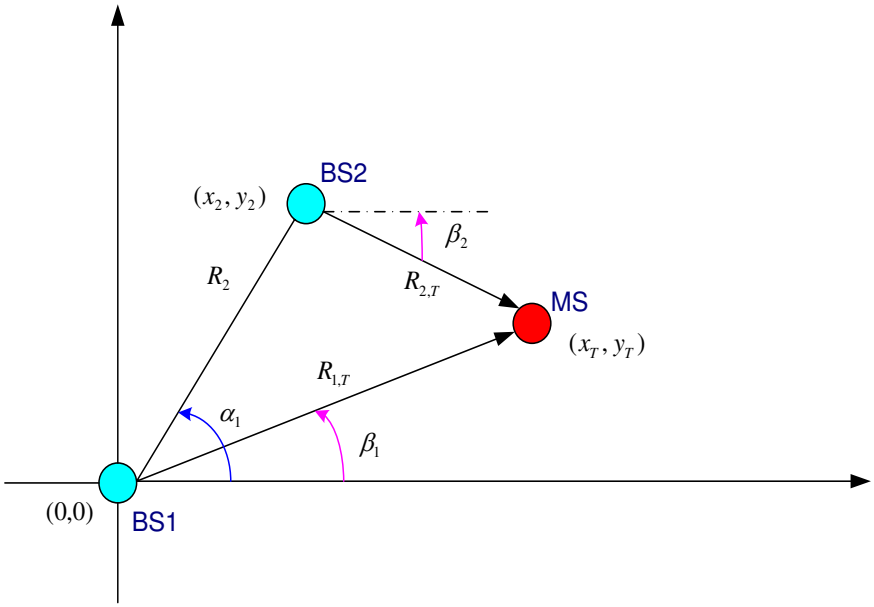

Figure 3.3: AOA data fusion with two BSs

to TOA/TDOA methods, the number of BSs needed for location is relatively smaller and there is no need for timing synchronization between base stations and MS clocks. However, one disadvantage is that antenna array used at the BS which is not available in 2G systems. It is planned for 3G cellular systems such as UMTS and CDMA2000 [65, 66]. As indicated in Figure 3.3, we have

$$\begin{bmatrix} x_T \\ y_T \end{bmatrix} = \begin{bmatrix} R_{1,T}\cos(\beta_1) \\ R_{1,T}\sin(\beta_1) \end{bmatrix}; \quad \begin{bmatrix} x_T \\ y_T \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + \begin{bmatrix} R_{2,T}\cos(\beta_2) \\ R_{2,T}\sin(\beta_2) \end{bmatrix}, \quad (3.11)$$

where

$$R_{2,T} = \sqrt{R_{1,T}^2 + R_2^2 - 2R_{1,T}R_2\cos(\alpha_1 - \beta_1)} = f(\alpha_1, \beta_1, R_{1,T}, R_2).$$

Since α_1, β_1 and R_2 is known, we simply denote $R_{2,T}$ as a function of $R_{1,T}$ as $R_{2,T} = f_2(R_{1,T})$. If there are more than two BSs, an LS formulation can be obtained by collecting the relations in (3.11) into a single equation as

$$\vec{b} = \mathbf{A}\vec{\theta}, \tag{3.12}$$

where

$$\vec{b} = \begin{bmatrix} R_{1,T}\cos(\beta_1) \\ R_{1,T}\sin(\beta_1) \\ \hline R_2 + f_2(R_{1,T})\cos(\beta_2) \\ R_2 + f_2(R_{1,T})\sin(\beta_2) \\ \hline \vdots \end{bmatrix}; \quad \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \hline 1 & 0 \\ 0 & 1 \\ \hline \vdots & \end{bmatrix}.$$

The LS solution for x is then

$$\hat{\vec{\theta}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{b}. \tag{3.13}$$

Since this intermediate solution involves the unknown $R_{1,T}$, we have to utilize the relation in (3.10) to get the positive root of the quadratic equation and then substitute $R_{1,T}$ back to (3.13) for a final solution of $\hat{\vec{\theta}}$.

- **Amplitude**. Amplitude-based wireless location technology is mainly used in indoor environments where WLAN standards such as 802.11a and 802.11g have been widely adopted. The WLAN connectivity has also become a standard feature for laptop computers and PDAs. As such, there is an increasing interest in utilizing these networks for location purposes to help provide a good coverage for indoor scenario. In 802.11b and 802.11g MAC layer, the information about

the signal strength and the signal-to-noise ratio is provided. Hence, a software-level location technique could be developed for WLAN networks based on the amplitude of arrival of the MS signal at different access points [67, 68, 69]. Specifically, when an IEEE 802.11 networks operate in the infrastructure mode, there are several APs (access point) and many end users within the network. RF-based systems that use the signal strength for location purposes can monitor the received signal strength from different APs and use the obtained statistics to build a conditional probability distribution network in order to estimate the location of the mobile client. These schemes usually work in two stages. The first stage is the offline training and data gathering phase and the second stage is the location determination phase using the online signal strength measurements. In the training phase, signal strength measurements are used to build an a priori probability distribution of the received signal strength at the mobile user from all APs. Assume there are N_a APs in the system and the radio map is created based on measurements from N_u user locations. It is illustrated in Figure 3.4.

The radio map model is described by [67]. Define $p(A_i \mid x_j, y_j)$ as the probability density function of the received signal strength from the i-th AP at the j-th measurement point (x_j, y_j). After constructing a Bayesian network, the online determination phase uses maximum likelyhood estimation to locate the mobile user. Thus assume that the mobile user measures the received signal strength from all APs, say $A_i, i = 1, 2, \ldots, N_a$. Then by Bayes' rule, the probability of

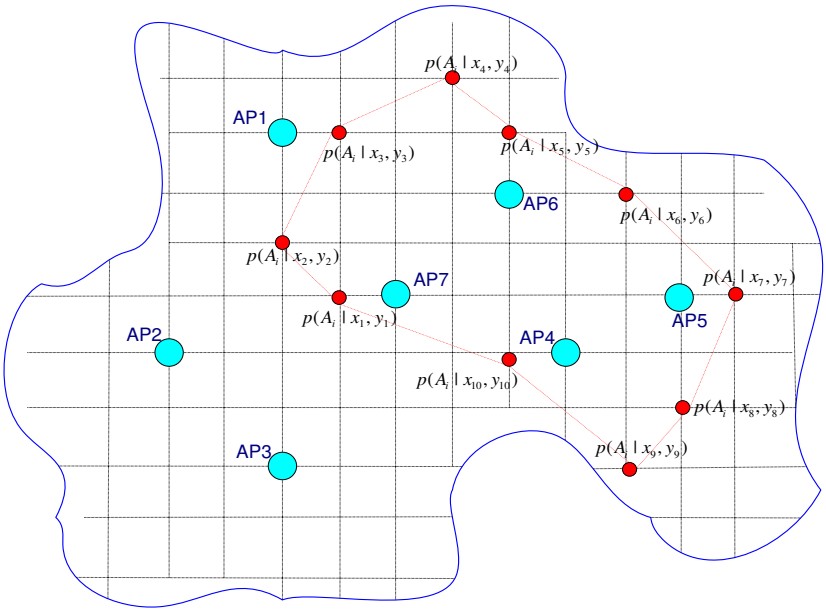

Figure 3.4: Magnitude-based data fusion in WLAN networks

having the mobile user at location (x_j, y_j) given the received signal strengths from all APs is given by

$$\begin{aligned} \mathbf{A} &= [A_1, \ldots, A_{N_a}]^T \\ p(x_j, y_j \mid \mathbf{A}) &= \frac{p(\mathbf{A}|x_j,y_j)p(x_j,y_j)}{p(\mathbf{A})} \\ &= \frac{p(x_j,y_j)\prod_{i=1}^{N_a} p(A_i|x_j,y_j)}{p(\mathbf{A})}, \end{aligned} \tag{3.14}$$

where $\prod_{i=1}^{N_a} p(A_i \mid x_j, y_j), 1 \leq j \leq N_u$ is the approximation for the conditional probability density function of the received signal strength when the location of the mobile is given. Thus the location of the mobile user can be estimated as

$$(\hat{x}_T, \hat{y}_T) = \arg\max_{\mathrm{x_j},\mathrm{y_j}} \mathrm{p}(\mathrm{x_j}, \mathrm{y_j} \mid \mathbf{A}) \qquad 1 \leq \mathrm{j} \leq \mathrm{N_u}. \tag{3.15}$$

We note that the location problem has been tackled by the LS approach as above. See also [53] for more details. However several problems exist. The first one is that

it is unclear for the physical meaning of these LS solutions, because of the lack of the statistical information on the measurements of the TOAs, TDOAs, AOAs and amplitudes, and the impact in transforming the nonlinear estimation for wireless location into quasi-linear estimation. This problem will be investigated in this thesis for location based on TDOAs and AOAs. The second one is the nuisance variables $R_{k,T}$, the distance from the k-th BS to the MS which is really unknown. Although we can use roots solving method, it works only if no noise is involved in measurement data and often no positive real roots exist. We will convert it into a constrained LS problem and provide a solution algorithm in this thesis. The final problem is location using more than one type of measurements. Because of the timing difficulty and lack of training, we will consider only measurements of TDOAs and AOAs for wireless location.

3.2 Least-square Location based on TDOA/AOA Estimates

3.2.1 Mathematical Preparations

Estimation problem, simply speaking, is to guess what you do not know base on what is given to you. In terms of its mathematical fundamentals, it is to estimate the unknown parameters based on some observation data by using some criteria which leads to an optimal estimator. The observation data usually is a function of the unknown parameters, either a linear function or a nonlinear one. For simplicity, let's

begin with a generic linear model as follows:

$$\vec{Z} = \mathbf{H}\vec{\theta} + \vec{V}. \tag{3.16}$$

In this model, $\vec{Z}$, of size $N \times 1$, is called the *measurement vector* ; $\vec{\theta}$, of size $n \times 1$, is called the *parameter vector* ; $\mathbf{H}$, of size $N \times n$, is called the *observation matrix* and $\vec{V}$, the same size as $\vec{Z}$, is called the *measurement noise vector*. Because $\vec{V}$ is random,$\vec{Z}$ is random too. Both $\mathbf{H}$ and $\vec{\theta}$ can be either deterministic or random. This is determined by the specific applications. Because of the simplicity, linear models are widely used in practice. Even in the case of nonlinear models, quasi-linear models that are close to nonlinear models are often pursued as in this thesis.

Here a question follows the linear model above: "How can we have the best estimate of $\vec{\theta}$ if we only know $\vec{Z}$?" This can be viewed as that we have made N times of independent experiments in order to estimate $\vec{\theta}$, which is composed of n unknown elements $\{\theta_1, \theta_2, \ldots, \theta_n\}$, where $n < N$. Inevitably, the experiment data is corrupted by some noise which is usually assumed to be additive Gaussian. To answer the question, there are generally three types of criteria to seek for the best estimate of $\vec{\theta}$ in the field of statistical signal processing. They are weighted least-square estimation (WLSE), minimum mean square estimation (MMSE) and maximum-likelihood estimation (MLE).

1. WLSE: It is the simplest method with the oldest history. The best estimate $\hat{\vec{\theta}}$ can be obtained by minimizing the cost function

$$J[\hat{\vec{\theta}}] = [\vec{Z} - \mathbf{H}\hat{\vec{\theta}}]^T \mathbf{W} [\vec{Z} - \mathbf{H}\hat{\vec{\theta}}], \tag{3.17}$$

where $\mathbf{W} = \mathbf{W}^T > 0$ is the weighting matrix.

2. MMSE: The optimal estimate minimizes the error variance. Given the measurements $\{Z(i)\}_{i=1}^N$, we shall determine an estimate of $\vec{\theta}$

$$\hat{\vec{\theta}} = f[z(1), z(2), \ldots, z(N)] \tag{3.18}$$

such that the mean squared error

$$J[\hat{\vec{\theta}}] = \mathcal{E}[\vec{\theta} - \hat{\vec{\theta}}]^T[\vec{\theta} - \hat{\vec{\theta}}] \tag{3.19}$$

is minimized.

3. MLE: It aims to maximize the likelihood function. Suppose that the measurement data $\{Z(i)\}_{i=1}^N$ are jointly distributed with a density function $p(\vec{Z}; \hat{\vec{\theta}})$. The optimal estimate is given by

$$\hat{\vec{\theta}}_{opt} = \arg\max_{\hat{\vec{\theta}}} \mathrm{p}(\vec{\mathrm{Z}}; \hat{\vec{\theta}}). \tag{3.20}$$

It is usually a nonlinear estimator since the likelihood function $p(\vec{Z}; \hat{\vec{\theta}})$ is nonlinear with respect to $\vec{\theta}(k)$. Hence the computational load could be high.

Then, how do we know whether or not the result obtained from one particular method is good? Or why is it better than other methods? We learn that, to answer this question, we must make use of the fact that all estimators represent transformations of random data and hence the estimate itself is random so that its properties must be studied from a statistical viewpoint. In this section, we introduce some fundamental

concepts such as unbiased estimator and efficient estimator, Cramer-Rao bound and Fisher information matrix [72].

Definition 3.1 *(Unbiasedness [72]) Suppose that the parameter vector $\vec{\theta}$ is deterministic. An estimator $\hat{\vec{\theta}}$ is unbiased if $\mathcal{E}\{\hat{\vec{\theta}}\} = \theta$.*

An unbiased estimate indicates that its mean value is the same as the true parameter vector. Hence as the number of observation increases, the estimate is assured to converge to the true parameter. However the unbiasedness itself is not adequate. We must study the dispersion about the mean, the variance of the estimator. Ideally, we would like our estimator to be unbiased and to have the smallest possible error variance.

Definition 3.2 *(Efficiency [72]) An unbiased estimate, $\hat{\vec{\theta}}$ of vector $\vec{\theta}$ is said to be more efficient than any other unbiased estimator, $\tilde{\vec{\theta}}$, of $\vec{\theta}$, if*

$$\mathcal{E}\{[\vec{\theta} - \hat{\vec{\theta}}][\vec{\theta} - \hat{\vec{\theta}}]^T\} \leq \mathcal{E}\{[\vec{\theta} - \tilde{\vec{\theta}}][\vec{\theta} - \tilde{\vec{\theta}}]^T\}. \tag{3.21}$$

A more efficient estimator has the smallest error covariance among all the unbiased estimators of $\vec{\theta}$, "smallest" in the sense that $\mathcal{E}\{[\vec{\theta} - \hat{\vec{\theta}}][\vec{\theta} - \hat{\vec{\theta}}]^T\} - \mathcal{E}\{[\vec{\theta} - \tilde{\vec{\theta}}][\vec{\theta} - \tilde{\vec{\theta}}]^T\}$ is negative semidefinite. Normally it does not make much sense to compare each pair of unbiased estimators. A lower bound, called CRB (Cramer-Rao Bound), about the minimum error variance achievable over all unbiased estimates exists and the efficiency of an unbiased estimator can be used to measure by how close it is to the CRB. The following theorem presents the CRB.

Theorem 3.1 *(Cramer-Rao Bound [72]) Let* $\vec{Z}$ *denote a set of* N *observation data, i.e.,* $\vec{Z} = [z(1), z(2), \ldots, z(N)]^T$ *which is characterized by the probability density function* $p(\vec{Z}; \theta) = p(\vec{Z})$*. If* $\hat{\vec{\theta}}$ *is an unbiased estimate of the deterministic* $\vec{\theta}$*, then the error convariance matrix,* $\mathcal{E}\{[\theta - \hat{\vec{\theta}}(k)][\theta - \hat{\vec{\theta}}(k)]^T\}$*, is bounded from below by*

$$\mathcal{E}\{[\theta - \hat{\vec{\theta}}(k)][\theta - \hat{\vec{\theta}}(k)]^T\} \geq \mathbf{J}^{-1}, \tag{3.22}$$

where $\mathbf{J}$ *is the Fisher information matrix, defined by*

$$\mathbf{J} = \mathcal{E}\left\{\left[\frac{\partial}{\partial\theta}\ln p(\vec{Z}(k))\right]\left[\frac{\partial}{\partial\theta}\ln p(\vec{Z}(k))\right]^T\right\}, \tag{3.23}$$

which can also be expressed equally as

$$\mathbf{J} = -\mathcal{E}\left\{\frac{\partial^{\mathbf{2}}}{\partial\theta^{\mathbf{2}}}\ln \mathbf{p}(\vec{\mathbf{Z}}(\mathbf{k}))\right\}. \tag{3.24}$$

Note that, for the theorem to be applicable, the vector derivatives in (3.23) must exist and the norm of $\partial p(\vec{Z})/\partial\theta$ must be absolutely integrable. Clearly, to compute the Cramer-Rao lower bound, we need to know the probability density function $p(\vec{Z})$. Often the exact information on $p(\vec{Z})$ is not available, for which we cannot evaluate this bound. However, in the case of normal distribution, i.e.,

$$p(\vec{Z}; \theta) = \frac{1}{(2\pi)^{N/2}|C|^{1/2}} e^{-\frac{[\vec{Z}-\mu]^T C^{-1}[\vec{Z}-\mu]}{2}}, \tag{3.25}$$

where μ and C are, respectively, the mean and the convariance matrix of $\vec{Z}$. Then we can compute the Cramer-Rao bound corresponding to the Gaussian data distribution by the Slepian-Bangs formula [74]

$$[\mathbf{J}^{-1}]_{ij} = \frac{1}{2}\mathrm{tr}\left[\mathrm{C}^{-1}\frac{\partial \mathrm{C}}{\partial\theta_\mathrm{i}}\mathrm{C}^{-1}\frac{\partial \mathrm{C}}{\partial\theta_\mathrm{j}}\right] + \left[\left(\frac{\partial\mu}{\partial\theta_\mathrm{i}}\right)^{\mathrm{T}}\mathrm{C}^{-1}\frac{\partial\mu}{\partial\theta_\mathrm{j}}\right]. \tag{3.26}$$

Because of the central limit theorem, Gaussian distribution holds approximately in applications such as location estimation.

3.2.2 Location based on TDOA

In this section, we investigate location estimation algorithms based on the measurements of TDOA and AOA. For simplicity, we assume that the mobile users travel at a low speed and can be taken as stationary targets approximately. Hence we do not consider the estimation of velocity of mobile users. Basically we explore all the available measurements $\{\Delta t_{k,1}\}_{k=2}^{N_b}$ (TDOA data) and $\{\beta_k\}_{k=1}^{N_b}$ (AOA data), where N_b is the total number of base stations to determine the location of the mobile user or the target, i.e., (x_T, y_T). It is seen that we consider only two-dimensional localization that is adequate, if the terrain elevation is known *a priori* or it could be neglected compared to the heights of the antenna towers.

We start with stationary target estimation based on the measurements of TDOA. As defined in section 3.1.3,

$$\begin{aligned} R_{k,T} &= \sqrt{(x_T - x_k)^2 + (y_T - y_k)^2} \\ \Delta t_{k,1} &= (R_{k,T} - R_{1,T})/c \\ &= (\sqrt{(x_T - x_k)^2 + (y_T - y_k)^2} - \sqrt{x_T^2 + y_T^2})/c. \end{aligned} \tag{3.27}$$

Besides the measurements $\{\Delta t_{k,1}\}_{k=2}^{N_b}$, the locations of all the base stations $\{(x_k, y_k)\}_{k=1}^{N_b}$ are also assumed to be known. Clearly $\Delta t_{k,1}$ is a nonlinear function of the unknown (x_T, y_T), i.e., $\Delta t_{k,1}(x_T, y_T)$. Here, for brevity of notation, (x_T, y_T) is omitted in TDOAs unless it is needed for clarification.

For all the TDOA measurements $\{\Delta\hat{t}_{2,1}, \Delta\hat{t}_{3,1}, \ldots, \Delta\hat{t}_{N_b,1}\}$, it is unavoidable that

there are measurement noises embedded within the data. Therefore, the measurement data are described by

$$\Delta\hat{t}_{k,1} = \Delta t_{k,1} + \delta t_k, \tag{3.28}$$

where $\{\delta t_k\}_{k=2}^{N_b}$ are assumed to be i.i.d. (independent and identical distributed) Gaussian random variables with zero mean and variance σ_t^2. It is an important but fair assumption given the fact that all the BSs are well synchronized and it is much less likely that a large deviation from the mean occurs. Since δt_k is a Gaussian random variable, so is $\Delta\hat{t}_{k,1}$. Based on the above assumption, we can define the $(N_b-1)\times 1$ multivariate Gaussian random variable vector $\vec{\Delta}\hat{t}$ and the associated mean $\vec{m}_{\Delta\hat{t}}$ and covariance matrix $\mathbf{M}_{\Delta\hat{t}}$ respectively as

$$\vec{\Delta}\hat{t} = \begin{bmatrix} \Delta\hat{t}_{2,1} \\ \vdots \\ \Delta\hat{t}_{N_b,1} \end{bmatrix}; \quad \vec{m}_{\Delta\hat{t}} = \begin{bmatrix} \Delta t_{2,1} \\ \vdots \\ \Delta t_{N_b,1} \end{bmatrix}; \quad \mathbf{M}_{\Delta\hat{t}} = \sigma_t^2\mathbf{I}_{(N_b-1)}. \tag{3.29}$$

As shown in [71], the joint PDF for $\vec{\Delta}\hat{t}$ is given by

$$\begin{aligned} p(\Delta\hat{t}) &= \frac{1}{(\sqrt{2\pi})^{N_b-1}\sqrt{\det \mathbf{M}_{\Delta\hat{t}}}} \exp[-\tfrac{1}{2}(\vec{\Delta}\hat{t}-\vec{m}_{\Delta\hat{t}})^T\mathbf{M}_{\Delta\hat{t}}^{-1}(\vec{\Delta}\hat{t}-\vec{m}_{\Delta\hat{t}})] \\ &= \frac{1}{(\sqrt{2\pi})^{N_b-1}\sigma_t^{N_b-1}} \exp[-\textstyle\sum_{k=2}^{N_b}\frac{(\Delta\hat{t}_{k,1}-\Delta t_{k,1}(x_T,y_T))^2}{2\sigma_t^2}]. \end{aligned} \tag{3.30}$$

This joint Gaussian PDF can completely describe the statistical characteristics of the measurement data and itself is affected by the two unknowns x_T and y_T. With a fixed data set of measurements, there must be only one pair of (x_T, y_T) such that the set of data is the most likely to occur. In light of the estimation theory, maximum-likelihood (ML) method can be explored to estimate the target location (x_T, y_T). Before providing the ML estimator, as shown in Theorem (3.1), we would like to

compute the Fisher information matrix and the Cramer-Rao bound such that we know how close the estimation can be. The Cramer-Rao bound is a benchmark for evaluating different types of unbiased estimators. Let $\mathbf{P}$ and $\mathbf{J}_{\text{FIM}}$ denote the estimation error convariance matrix and the Fish information matrix. It holds for any type of unbiased estimator [72] that

$$\mathbf{P} \geq \mathbf{J}_{\text{FIM}}^{-1}. \tag{3.31}$$

According to the *Slepian-Bangs* formula, the Fisher information matrix based on (3.30) can be calculated by

$$\mathbf{J}_{\text{tdoa}} = [\frac{1}{2}\text{tr}\{\mathbf{M}_{\Delta\hat{\text{t}}}^{-1}\frac{\partial \mathbf{M}_{\Delta\hat{\text{t}}}}{\partial \chi_{\text{i}}}\mathbf{M}_{\Delta\hat{\text{t}}}^{-1}\frac{\partial \mathbf{M}_{\Delta\hat{\text{t}}}}{\partial \chi_{\text{j}}}\} + (\frac{\partial \vec{\text{m}}_{\Delta\hat{\text{t}}}}{\partial \chi_{\text{i}}})^{\text{T}}\mathbf{M}_{\Delta\hat{\text{t}}}^{-1}(\frac{\partial \vec{\text{m}}_{\Delta\hat{\text{t}}}}{\partial \chi_{\text{j}}})]_{\text{i,j}=1,1}^{2,2}, \tag{3.32}$$

where $\chi_1 = x_T$ and $\chi_2 = y_T$. Since $\mathbf{M}_{\Delta\hat{t}} = \sigma_t^2\mathbf{I}_{N_b-1}$ is only related to σ_t^2, the first term in (3.32) is zero. By direct calculations, we have

$$\frac{\partial \vec{m}_{\Delta\hat{t}}}{\partial \chi_1} = \begin{bmatrix} \frac{1}{c}(\frac{x_T - x_2}{R_{2,T}} - \frac{x_T}{R_{1,T}}) \\ \frac{1}{c}(\frac{x_T - x_3}{R_{3,T}} - \frac{x_T}{R_{1,T}}) \\ \vdots \\ \frac{1}{c}(\frac{x_T - x_{N_b}}{R_{N_b,T}} - \frac{x_T}{R_{1,T}}) \end{bmatrix}; \quad \frac{\partial \vec{m}_{\Delta\hat{t}}}{\partial \chi_2} = \begin{bmatrix} \frac{1}{c}(\frac{y_T - y_2}{R_{2,T}} - \frac{y_T}{R_{1,T}}) \\ \frac{1}{c}(\frac{y_T - y_3}{R_{3,T}} - \frac{y_T}{R_{1,T}}) \\ \vdots \\ \frac{1}{c}(\frac{y_T - y_{N_b}}{R_{N_b,T}} - \frac{y_T}{R_{1,T}}) \end{bmatrix}; \quad \mathbf{M}_{\Delta\hat{t}}^{-1} = \frac{1}{\sigma_t^2}\mathbf{I}_{N_b-1}. \tag{3.33}$$

Then it is easy to obtain $\mathbf{J}_{\text{tdoa}}$ as the follow by substituting (3.33) into (3.32),

$$\begin{aligned} \mathbf{J}_{\text{tdoa}} &= [(\frac{\partial \vec{m}_{\Delta\hat{t}}}{\partial \chi_i})^T\mathbf{M}_{\Delta\hat{t}}^{-1}(\frac{\partial \vec{m}_{\Delta\hat{t}}}{\partial \chi_j})]_{i,j=1,1}^{2,2} \\ &= \sum_{k=2}^{N_b}\frac{1}{c^2\sigma_t^2}\begin{bmatrix} \frac{x_T - x_k}{R_{k,T}} - \frac{x_T}{R_{1,T}} \\ \frac{y_T - y_k}{R_{k,T}} - \frac{y_T}{R_{1,T}} \end{bmatrix}\begin{bmatrix} \frac{x_T - x_k}{R_{k,T}} - \frac{x_T}{R_{1,T}}, & \frac{y_T - y_k}{R_{k,T}} - \frac{y_T}{R_{1,T}} \end{bmatrix} \\ &= \sum_{k=2}^{N_b}\frac{1}{c^2\sigma_t^2}\begin{bmatrix} \cos(\beta_k) - \cos(\beta_1) \\ \sin(\beta_k) - \sin(\beta_1) \end{bmatrix}\begin{bmatrix} \cos(\beta_k) - \cos(\beta_1), & \sin(\beta_k) - \sin(\beta_1) \end{bmatrix}, \end{aligned} \tag{3.34}$$

where $\{\beta_k\}_{k=1}^{N_b}$ are shown in Figure 3.3 with $\tan(\beta_k) = (y_T - y_k)/(x_T - x_k)$. By taking an inverse of the Fisher information matrix $\mathbf{J}_{\text{tdoa}}$, the resultant matrix will be a lower bound of estimation error covariance for all the unbiased estimators.

In terms of the large-sample property, the ML estimate approaches the Cramer-Rao bound asymptotically, i.e, with an infinite number of data measurements. From (3.30), the ML location estimator seeks (x_T, y_T) to minimize the log-likelihood function of the form

$$L_{\Delta t}(x_T, y_T) = \sum_{k=2}^{N_b} \left| c\Delta\hat{t}_{k,1} - \sqrt{(x_T - x_k)^2 + (y_T - y_k)^2} + \sqrt{x_T^2 + y_T^2} \right|^2. \tag{3.35}$$

This is obtained by using the fact that e^{-x} is a monotonically decreasing function and scaling with a constant $c^2\sigma_t^2$ does not affect the likelihood function. There are two unknowns in $L_{\Delta t}(x_T, y_T)$, namely x_T and y_T. Differentiating $L_{\Delta t}(x_T, y_T)$ with respect to each and equating the resulting partial derivatives to zero gives the following necessary condition for the optimal solution (x_T^*, y_T^*)

$$\sum_{k=2}^{N_b} \begin{bmatrix} \frac{x_k - x_T^*}{R_{k,T}} + \frac{x_T^*}{R_{1,T}} \\ \frac{y_k - y_T^*}{R_{k,T}} + \frac{y_T^*}{R_{1,T}} \end{bmatrix} (c\Delta\hat{t}_{k,1} - R_{k,T} + R_{1,T}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{3.36}$$

The ML estimator is well studied and widely used in practice, especially in some applications which require high accuracy of estimation and computational complexity can be afforded via commercially available hardware and software. It has a variety of statical properties which is preferred in applications:

- It is unbiased: the expectation of the estimate is equal to the real value;
- It is the most efficient: it achieves the minimum error variance;

- It is consistent: it converges to the real value in probability.

Hence it is plausible to apply ML to our estimation problem for the highest possible accuracy of localization. However, solving the optimal solution (x_T^*, y_T^*) from (3.35) and (3.36) is not easy and involves nonlinear procedures such as Newton-type algorithms which are not discussed in this dissertation. The maximization of the likelihood function can be done by hands with some PDFs and even the commercial software does not guarantee to reach the ML solution because of the possible existence of the local minimum. In this thesis we take a quasi-linear approach as in [54] to convert the nonlinear optimization problem into a linear one that leads to an LS-type problem in order to simplify the solution algorithm. Or we can use the LS-type solution as an initial solution candidate in the Newton-type iterative algorithms to ensure the fast convergence to the true ML solution (x_T^*, y_T^*). For this purpose of bypassing the difficulty and complexity of the original ML estimator, we notice that the second equation in (3.27) leads to

$$(x_T - x_k)^2 + (y_T - y_k)^2 = \left(\sqrt{x_T^2 + y_T^2} + c\Delta t_{k,1}\right)^2. \tag{3.37}$$

By expanding and rearranging the terms, the above can be written as

$$\frac{1}{c^2}R_k^2 = \frac{2}{c^2}\begin{bmatrix} x_k & y_k \end{bmatrix}\begin{bmatrix} x_T \\ y_T \end{bmatrix} + \Delta t_{k,1}^2 + \frac{2}{c}R_{1,T}\Delta t_{k,1}. \tag{3.38}$$

Packing all the equations in (3.38) for $k = 2, 3, \ldots, N_b$ yields

$$\frac{1}{c^2}\begin{bmatrix} R_2^2 \\ \vdots \\ R_{N_b}^2 \end{bmatrix} = \frac{2}{c^2}\begin{bmatrix} x_2 & y_2 \\ \vdots & \vdots \\ x_{N_b} & y_{N_b} \end{bmatrix}\begin{bmatrix} x_T \\ y_T \end{bmatrix} + \begin{bmatrix} \Delta t_{2,1}^2 \\ \vdots \\ \Delta t_{N_b,1}^2 \end{bmatrix} + \frac{2}{c}\begin{bmatrix} \Delta t_{2,1} \\ \vdots \\ \Delta t_{N_b,1} \end{bmatrix} R_{1,T}. \tag{3.39}$$

If we have the perfect TDOA information, the target (x_T, y_T) is uniquely located with any 2 out of the $N_b - 1$ sets of data since it is an over-determined problem. To estimate the target location (x_T, y_T) in (3.39), however, we have to replace the perfect time difference $\Delta t_{k,1}$ with the available TDOA measurements $\Delta\hat{t}_{k,1}$. It then introduces a noise vector as follows, since $\Delta\hat{t}_{k,1} = \Delta t_{k,1} + \delta t_k$.

$$\begin{bmatrix} \eta_2 \\ \vdots \\ \eta_{N_b} \end{bmatrix} = -\frac{2}{c} \begin{bmatrix} \delta t_2 \\ \vdots \\ \delta t_{N_b} \end{bmatrix} R_{1,T} - 2 \begin{bmatrix} \Delta\hat{t}_{2,1}\delta t_2 \\ \vdots \\ \Delta\hat{t}_{N_b,1}\delta t_{N_b} \end{bmatrix} + \begin{bmatrix} \delta t_2^2 \\ \vdots \\ \delta t_{N_b}^2 \end{bmatrix}. \tag{3.40}$$

Each element of the noise vector is composed of the TDOA measurement noise δt_k and the corresponding squared term. Taking expectation at both sides of (3.40), we find that each element of the noise vector is with mean σ_t^2. In an effort to obtain an LS-type formulation, we define

$$a_k = \left(R_k^2/c^2 - \Delta\hat{t}_{k,1}^2 - \sigma_t^2\right), \qquad b_k = 2\Delta\hat{t}_{k,1}/c. \tag{3.41}$$

We can regard $\{a_k\}$ and $\{b_k\}$ as pseudo-measurements that leads to a constrained linear model:

$$\begin{bmatrix} a_2 \\ \vdots \\ a_{N_b} \end{bmatrix} - \begin{bmatrix} b_2 \\ \vdots \\ b_{N_b} \end{bmatrix} R_{1,T} = \frac{2}{c^2} \begin{bmatrix} x_2 & y_2 \\ \vdots & \vdots \\ x_{N_b} & y_{N_b} \end{bmatrix} \begin{bmatrix} x_T \\ y_T \end{bmatrix} + \begin{bmatrix} \eta_2 - \sigma_t^2 \\ \vdots \\ \eta_{N_b} - \sigma_t^2 \end{bmatrix} \tag{3.42}$$

where the constraint is $R_{1,T} = \sqrt{x_T^2 + y_T^2}$. It is worth noting that the composite-noise $\{\eta_k\}_{k=2}^{N_b}$ are not Gaussian random variables or to say, not in normal distribution. But if $\{\eta_k\}_{k=2}^{N_b}$ are Gaussian then the ML algorithm for location estimation is equivalent to a weighted LS problem involving a constraint. As stated in Corollary 11-1 of [72],

ML, LS and BLUE (Best Linear Unbiased Estimator) algorithms are all equivalent for a generic linear model with additive Gaussian noise term. By defining

$$\vec{a} = \begin{bmatrix} a_2 \\ \vdots \\ a_{N_b} \end{bmatrix}; \quad \vec{b} = \begin{bmatrix} b_2 \\ \vdots \\ b_{N_b} \end{bmatrix}; \quad \mathbf{H}_1 = \frac{2}{c^2} \begin{bmatrix} x_2 & y_2 \\ \vdots & \vdots \\ x_{N_b} & y_{N_b} \end{bmatrix}; \quad \vec{\eta}_1 = \begin{bmatrix} \eta_2 - \sigma_t^2 \\ \vdots \\ \eta_{N_b} - \sigma_t^2 \end{bmatrix}. \tag{3.43}$$

we can rewrite (3.42) into a more compact quasi-linear form:

$$\vec{a} - \vec{b}R_{1,T} = \mathbf{H}_1\vec{\theta} + \vec{\eta}_1. \tag{3.44}$$

The above expression is very similar to a generic linear model of the standard LS algorithm except that the pseudo-measurements vector $\vec{a} - \vec{b}R_{1,T}$ involves one unknown $R_{1,T} = \sqrt{x_T^2 + z_T^2}$. Fortunately we have an extra condition that helps to solve $R_{1,T}$. $\mathbf{H}_1$ is deterministic and $\vec{\eta}_1$ is a non-Gaussian vector but whose elements all have zero mean. Let $\mathbf{W}_1(R_{1,T})$ be a diagonal matrix with elements $\mathrm{E}\{|\eta_k - \sigma_t^2|^2\}$. Set

$$J_1 = \frac{1}{2}\left[\vec{a} - \vec{b}R_{1,T} - \mathbf{H}_1\vec{\theta}\right]^T \mathbf{W}_1^{-1}(R_{1,T})\left[\vec{a} - \vec{b}R_{1,T} - \mathbf{H}_1\vec{\theta}\right] \tag{3.45}$$

as the objective function to be minimized. Then it is well known that the minimizer is the ML solution provided that the noise vector is Gaussian with $\mathbf{W}_1(R_{1,T})$ as the covariance matrix. The weighted LS solution can be easily obtained as

$$\hat{\vec{\theta}} = (\mathbf{H}_1^T\mathbf{W}_1^{-1}(R_{1,T})\mathbf{H}_1)^{-1}\mathbf{H}_1^T\mathbf{W}_1^{-1}(R_{1,T})\left[\vec{a} - \vec{b}R_{1,T}\right] = \mathbf{\Phi}_1(R_{1,T})\left[\vec{a} - \vec{b}R_{1,T}\right], \tag{3.46}$$

where $\mathbf{\Phi}_1(R_{1,T}) = (\mathbf{H}_1^T\mathbf{W}_1^{-1}(R_{1,T})\mathbf{H}_1)^{-1}\mathbf{H}_1^T\mathbf{W}_1^{-1}(R_{1,T})$. Here $\hat{\vec{\theta}}$ is an intermediate solution since $R_{1,T}$ is unknown. By taking norm square on both sides, it yields

$$R_{1,T}^2 = \|\mathbf{\Phi}_1(R_{1,T})\left[\vec{a} - \vec{b}R_{1,T}\right]\|^2. \tag{3.47}$$

If one of the roots from such a nonlinear equation is real and positive of which the one yielding the smallest J_1 is the optimal solution to the constrained LS problem. It is commented that we convert the ML estimation problem into an LS-type estimation by replacing the perfect TDOA information with measurement data and the equivalence between the LS-type solution and ML estimator can be further established based on the assumption that the composite noise vector is Gaussian. If the noise vector in (3.40) is not exactly Gaussian, the constrained LS solution is not the ML solution either. It seems that we overemphasized the simplicity that LS-type algorithm may have and sacrificed the accuracy of estimation. However it is not too far away from the true ML solution under some mild conditions as shown below.

Let X be a Gaussian random variable with zero mean and variance σ^2. Then the high-order moments of X is given by [73]

$$\mathrm{E}\{X^{2n}\} = 1 \times 3 \times 5 \times \cdots \times (2n-1)\sigma^{2n}; \qquad \mathrm{E}\{\mathrm{X}^{2\mathrm{n}-1}\} = 0.$$

where $n > 0$ an integer. Let $Y = \alpha X + (X^2 - \sigma^2)$. Then $\mathrm{E}\{Y\} = 0$ and

$$\sigma_Y^2 = \mathrm{E}\{Y^2\} = \alpha^2\sigma^2 - \sigma^4 + \mathrm{E}\{X^4\} = \alpha^2\sigma^2 + 2\sigma^4 = \sigma^2(\alpha^2 + 2\sigma^2). \tag{3.48}$$

Gaussian random variables (GRV) admit some nice properties that the summation of any two GRV is still a GRV and the product of two independent GRV is a GRV [73]. But we cannot conclude that Y is a GRV since it includes the X^2 term. We are interested in under what condition Y is close to a GRV. By noting that $Y =$

$(X+\alpha/2)^2-(\sigma^2+\alpha^2/4)$, we have

$$X=-\alpha/2\pm\sqrt{Y+(\sigma^2+\alpha^2/4)},\qquad Y\geq-(\sigma^2+\alpha^2/4). \tag{3.49}$$

Since Y is a function of the GRV X, its PDF is thus given by

$$p_Y(y)=\frac{1}{\sqrt{2\pi\sigma^2}}\left\{\frac{e^{-\frac{1}{2\sigma^2}\left(\frac{\alpha}{2}-\sqrt{y+(\sigma^2+\alpha^2/4)}\right)^2}}{2\sqrt{y+(\sigma^2+\alpha^2/4)}}+\frac{e^{-\frac{1}{2\sigma^2}\left(\frac{\alpha}{2}+\sqrt{y+(\sigma^2+\alpha^2/4)}\right)^2}}{2\sqrt{y+(\sigma^2+\alpha^2/4)}}\right\},y\geq-(\sigma^2+\alpha^2/4). \tag{3.50}$$

From PDF's property, there holds $\int_{-(\sigma^2+\alpha^2/4)}^{\infty}p_Y(y)=1$. Interestingly, the integral of the first term in $p_Y(y)$ is

$$\begin{aligned}
I_Y &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-(\sigma^2+\alpha^2/4)}^{\infty}\frac{e^{-\frac{1}{2\sigma^2}\left(\frac{\alpha}{2}-\sqrt{y+(\sigma^2+\alpha^2/4)}\right)^2}}{2\sqrt{y+(\sigma^2+\alpha^2/4)}}\,dy\\
&= \frac{-1}{\sqrt{2\pi\sigma^2}}\int_{-(\sigma^2+\alpha^2/4)}^{\infty}e^{-\frac{1}{2\sigma^2}\left(\frac{\alpha}{2}-\sqrt{y+(\sigma^2+\alpha^2/4)}\right)^2}d[\frac{\alpha}{2}-\sqrt{y+(\sigma^2+\alpha^2/4)}]\\
&= \frac{-1}{\sqrt{2\pi\sigma^2}}\int_{\frac{\alpha}{2}}^{-\infty}e^{-\frac{z^2}{2\sigma^2}}dz\qquad\left(let:\quad z=\frac{\alpha}{2}-\sqrt{y+(\sigma^2+\alpha^2/4)}\right)\\
&= \frac{1}{\sqrt{2\pi}}\int_{\frac{-\alpha}{2\sigma}}^{\infty}e^{-\frac{\tilde{z}^2}{2\sigma^2}}d\tilde{z}\qquad\left(let:\quad\tilde{z}=-\frac{z}{\sigma}\right)\\
&= 1-Q\left(\frac{\alpha}{2\sigma}\right),
\end{aligned}$$

where $Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^{\infty}e^{-\frac{x^2}{2\sigma^2}}dx$ is the error function. Hence it is concluded that if α/σ is sufficiently large, then $I_Y\approx1$ and thus $p_Y(y)$ is dominated by the first term. Intuitively, it can be seen that the second term $(X^2-\sigma^2)$ in Y will fade out since its mean is zero and it has a small variance $\mathrm{E}\{(\mathrm{X}^2-\sigma^2)^2\}=2\sigma^4$ when α/σ is sufficiently large. It is also easy to see that σ_Y^2 is dominated by $\alpha^2\sigma^2$ based on the same assumption. Therefore, the random variable $Y=\alpha X+(X^2-\sigma^2)$ behaves like normal distributed, provided that α/σ is sufficiently large. Translating this result to

the random variables as in (3.40) with $\delta\tilde{t}_k = -\delta t_k$ leads to

$$Y_k = \alpha_k \delta\tilde{t}_k + (\delta\tilde{t}_k^2 - \sigma_t^2), \qquad \alpha_k = 2(R_{1,T} + c\Delta\hat{t}_{k,1})/c. \tag{3.51}$$

Then $\vec{\eta}_1 = [Y_2, Y_3, \ldots, Y_{N_b}]^T$ is a normally distributed vector, as δt_k is Gaussian with mean zero and variance σ_t^2. Thus Y_k is close to Gaussian provided that $\alpha_k/\sigma_t = 2(R_{1,T} + c\Delta\hat{t}_{k,1})/(c\sigma_t)$ is sufficiently large for all $k \geq 2$. It is worth noting that

$$(\frac{\alpha_k}{2\sigma_t})^2 = (R_{1,T}/c + \Delta\hat{t}_{k,1})^2/\sigma_t^2. \tag{3.52}$$

The right-hand side of the above equation indicates an approximation to the SNR, since its numerator represents the recorded signal of the traveling time from the target to the k-th BS and its denominator, σ_t^2, is the noise variance. If α_k/σ_t is sufficiently large, the variance of Y_k is, by (3.48),

$$\sigma_{Y_k}^2 = \mathrm{E}\{Y_k^2\} = \alpha_k^2\sigma_t^2 + 2\sigma_t^4 = \sigma_t^2(\alpha_k^2 + 2\sigma_t^2) \tag{3.53}$$

that is dominated by $\alpha_k^2\sigma_t^2$. It is emphasized that $\alpha_k = \alpha_k(R_{1,T})$ is a function of $R_{1,T}$. Recall that one question is raised in the previous part that how far is the LS-type solution obtained in (3.46) and (3.47) away from the true ML solution. Here a clear answer is that the LS-type algorithm approximates to the ML solution well as long as α_k/σ is very large for $2 \leq k \leq N_b$. Therefore the properties of the ML algorithm hold approximately.

Before ending this subsection, we would like to compute the Cramér-Rao bound associated with the weighted LS solution by assuming that $\{Y_k\}$ are normal dis-

tributed which holds true approximately under the condition discussed earlier. Recall that $\mathbf{W}_1(R_{1,T})$ in the weighted LS problem is the associated covariance matrix. Thus $\mathrm{E}\{Y_k^2\}$ is its element and the joint probability density function (PDF) for the pseudo-measurement data $\{a_k\}$ and $\{b_k\}$ in (3.42) is

$$\mathrm{PDF} = \frac{1}{\sqrt{(2\pi)^{n-1}\det[\mathbf{W}_1(R_{1,T})]}}\exp\left\{-\frac{1}{2}\left[\vec{a}-\vec{b}R_{1,T}-\mathbf{H}_1\vec{\theta}\right]^T\mathbf{W}_1^{-1}(R_{1,T})\left[\vec{a}-\vec{b}R_{1,T}-\mathbf{H}_1\vec{\theta}\right]\right\}. \tag{3.54}$$

Note that inside the exponent is exactly J_1 with a minus sign. The Fisher information matrix for the PDF in (3.42) can be computed by using the Slepian-Bangs formula in (3.32). Here we take the pseudo-measurement vector $\vec{a}$ as the data vector whose mean vector and convariance matrix are $\vec{m}_a = \vec{b}R_{1,T} + \mathbf{H}_1\vec{\theta}$ and $\mathbf{M}_a = \mathcal{E}\{\vec{\eta}_1\vec{\eta}_1^T\}$ respectively. Hence both mean and covariance are functions of (x_T, y_T). By some direct calculations, we have

$$\begin{aligned}
\frac{\partial\vec{m}_a}{\partial x_T} &= \begin{bmatrix} \frac{\partial}{\partial x_T}(b_2\sqrt{x_T^2+y_T^2}+\frac{2}{c^2}(x_2x_T+y_2y_T)) \\ \vdots \\ \frac{\partial}{\partial x_T}(b_{N_b}\sqrt{x_T^2+y_T^2}+\frac{2}{c^2}(x_{N_b}x_T+y_{N_b}y_T)) \end{bmatrix} \\
&= \begin{bmatrix} \frac{2}{c^2}x_2+b_2\cos(\beta_1) \\ \vdots \\ \frac{2}{c^2}x_{N_b}+b_{N_b}\cos(\beta_1) \end{bmatrix}; \\
\frac{\partial\vec{m}_a}{\partial y_T} &= \begin{bmatrix} \frac{\partial}{\partial y_T}(b_2\sqrt{x_T^2+y_T^2}+\frac{2}{c^2}(x_2x_T+y_2y_T)) \\ \vdots \\ \frac{\partial}{\partial y_T}(b_{N_b}\sqrt{x_T^2+y_T^2}+\frac{2}{c^2}(x_{N_b}x_T+y_{N_b}y_T)) \end{bmatrix} \\
&= \begin{bmatrix} \frac{2}{c^2}y_2+b_2\sin(\beta_1) \\ \vdots \\ \frac{2}{c^2}y_{N_b}+b_{N_b}\sin(\beta_1) \end{bmatrix}.
\end{aligned} \tag{3.55}$$

And since $\mathbf{M}_a$ is a diagonal matrix whose k-th diagonal element is $E\{Y_k^2\} = \sigma_2(\alpha_k^2 + 2\sigma^2)$ with $\alpha_k = \frac{2}{c}\sqrt{x_T^2+y_T^2}+2\Delta\hat{t}_{k,1}$, then taking the partial derivative of $E\{Y_k^2\}$ with respect to x_T and y_T gives

$$\frac{\partial}{\partial x_T}\mathrm{E}\{\mathrm{Y}_\mathrm{k}^2\} = \frac{4\sigma_\mathrm{t}^2\alpha_\mathrm{k}}{\mathrm{c}}\cos(\beta_1); \qquad \frac{\partial}{\partial \mathrm{y_T}}\mathrm{E}\{\mathrm{Y}_\mathrm{k}^2\} = \frac{4\sigma_\mathrm{t}^2\alpha_\mathrm{k}}{\mathrm{c}}\sin(\beta_1).$$

It is then straightforward to show that

$$\begin{aligned}
\frac{\partial}{x_T}\mathbf{M}_a &= \mathrm{diag}\{\tfrac{4\sigma_{\mathrm{t}}^2\cos(\beta_1)}{\mathrm{c}}[\alpha_2,\alpha_3,\ldots,\alpha_{\mathrm{N_b}}]\} \\
\frac{\partial}{y_T}\mathbf{M}_a &= \mathrm{diag}\{\tfrac{4\sigma_{\mathrm{t}}^2\sin(\beta_1)}{\mathrm{c}}[\alpha_2,\alpha_3,\ldots,\alpha_{\mathrm{N_b}}]\} \\
\mathbf{M}_a^{-1} &= \mathrm{diag}\{\tfrac{1}{\sigma_{\mathrm{t}}^2}[\tfrac{1}{\alpha_2^2+2\sigma_{\mathrm{t}}^2},\tfrac{1}{\alpha_3^2+2\sigma_{\mathrm{t}}^2},\ldots,\tfrac{1}{\alpha_{\mathrm{N_b}}^2+2\sigma_{\mathrm{t}}^2}]\}.
\end{aligned} \tag{3.56}$$

Now we can calculate the Fisher information matrix via *Slepian-Bangs* method in (3.32) as

$$\mathbf{J}_{tdoa,LS} = [\frac{1}{2}\mathrm{tr}\{\mathbf{M}_{\mathrm{a}}^{-1}\frac{\partial\mathbf{M}_{\mathrm{a}}}{\partial\chi_{\mathrm{i}}}\mathbf{M}_{\mathrm{a}}^{-1}\frac{\partial\mathbf{M}_{\mathrm{a}}}{\partial\chi_{\mathrm{j}}}\} + (\frac{\partial\vec{\mathrm{m}}_{\mathrm{a}}}{\partial\chi_{\mathrm{i}}})^{\mathrm{T}}\mathbf{M}_{\mathrm{a}}^{-1}(\frac{\partial\vec{\mathrm{m}}_{\mathrm{a}}}{\partial\chi_{\mathrm{j}}})]_{\mathrm{i,j}=1,1}^{2,2}, \tag{3.57}$$

where $\chi_1 = x_T$ and $\chi_2 = y_T$. By substituting (3.55) and (3.56) into (3.57), the Fisher information matrix is given by

$$\begin{aligned}
\mathbf{J}_{tdoa,LS} &= \sum_{k=2}^{n}\left(\frac{1}{\sigma_t^2(\alpha_k^2+2\sigma_t^2)}\right)\begin{bmatrix} 2x_k/c^2 + b_k\cos(\beta_1) \\ 2y_k/c^2 + b_k\sin(\beta_1)\end{bmatrix}\begin{bmatrix} 2x_k/c^2 + b_k\cos(\beta_1) \\ 2y_k/c^2 + b_k\sin(\beta_1)\end{bmatrix}^T \\
&\quad + \sum_{k=2}^{n}\left(\frac{8\alpha_k^2}{c^2(\alpha_k^2+2\sigma_t^2)^2}\right)\begin{bmatrix}\cos(\beta_1)\\ \sin(\beta_1)\end{bmatrix}\begin{bmatrix}\cos(\beta_1) & \sin(\beta_1)\end{bmatrix}
\end{aligned} \tag{3.58}$$

The above expression is different from $\mathbf{J}_{tdoa}$ in (3.34) no matter how large α_k/σ_t is and how small σ_t is. Such a discrepancy is caused by the omission of the second term in $p_Y(y)$ in computing the Fisher information matrix. The omitted term in $p_Y(y)$ may have negligible value in computing the probability but its derivative can be significant. Moreover no matter how small σ_t is, it can not be zero that contributes to this discrepancy.

3.2.3 Location based on AOA

The angle of arrival (AOA) of MS signals at a BS can be obtained by antenna arrays. Unlike TOA/TDOA based location methods, we do not need to consider timing syn-

chronization problems for an AOA based location algorithm. But there are something in common with TOA/TDOA that we have to fuse either TOA/TDOA or AOA measurements into the triangular relations between the BSs and the mobile users, i.e., the target. Suppose that the AOA measurement data are to be of the form $\hat{\beta}_k = \beta_k + \delta\beta_k$. Recall that $\tan(\beta_k) = (y_T - y_k)/(x_T - x_k)$. That is, $\beta_k = \beta_k(x_T, y_T)$. We again assume that $\{\delta\beta_k\}$ are uncorrelated with Gaussian distribution of mean zero and variance σ_β^2. Its joint PDF is given by

$$p_{\Delta\beta}(\delta\beta) = \frac{1}{\sqrt{(2\pi)^{N_b}}\sigma_\beta^{N_b}} \exp\left[-\sum_{k=1}^{N_b} \frac{1}{2\sigma_\beta^2}\left(\hat{\beta}_k - \beta_k(x_T, y_T)\right)^2\right]. \tag{3.59}$$

Since the AOA measurements are associated with additive Gaussian noise, it is easy to compute the Fisher information matrix whose inverse matrix is the Cramer-Rao bound for the covariance matrix of the estimation error. Simply speaking, the larger the Fisher information matrix, the smaller the estimation error variance. And that translates into a better estimator in terms of accuracy, provided that it is unbiased. The Fisher information matrix contains the relative rate (derivative) at which the probability density function changes with respect to the data. Note that the greater the expectation of a change is at a give value, say $(\hat{x}_T, \hat{y}_T)$, the easier it is to distinguish $(\hat{x}_T, \hat{y}_T)$ from neighboring values (locations), and hence the more precisely (x_T, y_T) can be estimated at (x_T, y_T)=$(\hat{x}_T, \hat{y}_T)$. To calcualte the Fisher information matrix, we still have to use the *Slepian-Bangs* formula as in (3.32). First, some primary

computations are carried out as

$$\begin{aligned} \frac{\partial \beta_k(x_T,y_T)}{\partial x_T} &= \frac{\partial}{\partial x_T}\left[\tan^{-1}\left(\frac{y_T-y_k}{x_T-x_k}\right)\right] & \frac{\partial \beta_k(x_T,y_T)}{\partial y_T} &= \frac{\partial}{\partial y_T}\left[\tan^{-1}\left(\frac{y_T-y_k}{x_T-x_k}\right)\right] \\ &= -\frac{y_T-y_k}{R_{k,T}^2}; & &= \frac{x_T-x_k}{R_{k,T}^2}. \end{aligned} \tag{3.60}$$

And we know that the mean vector is $\vec{m}_\beta = [\beta_1, \beta_2, \ldots, \beta_{N_b}]^T$ and the covariance matrix is $\mathbf{M}_\beta = \mathbf{I}_{N_b}$. With these primary calculation and results, the Fisher information matrix of AOA measurements is given by

$$\begin{aligned} \mathbf{J}_{aoa} &= \sum_{k=1}^{N_b} \frac{1}{\sigma_\beta^2 R_{k,T}(x_T,y_T)^2} \begin{bmatrix} -\dfrac{y_T-y_k}{R_{k,T}(x_T,y_T)} \\ \dfrac{x_T-x_k}{R_{k,T}(x_T,y_T)} \end{bmatrix} \begin{bmatrix} -\dfrac{y_T-y_k}{R_{k,T}(x_T,y_T)} & \dfrac{x_T-x_k}{R_{k,T}(x_T,y_T)} \end{bmatrix} \\ &= \sum_{k=1}^{N_b} \frac{1}{\sigma_\beta^2 R_{k,T}(x_T,y_T)^2} \begin{bmatrix} -\sin(\beta_k) \\ \cos(\beta_k) \end{bmatrix} \begin{bmatrix} -\sin(\beta_k) & \cos(\beta_k) \end{bmatrix}. \end{aligned} \tag{3.61}$$

With the information matrix above, we can calculate the Cramer-Rao bound (CRB)easily. In terms of CRB, ML estimator is the closest one among all the unbiased estimators. The ML algorithm is to minimize the likelihood function of the following form

$$L_{\Delta\beta}(x_T,y_T) = \sum_{k=1}^{N_b} \left(\hat{\beta}_k - \beta_k(x_T,y_T)\right)^2. \tag{3.62}$$

Then the necessary condition for (x_T^*, y_T^*) to be ML solution is

$$\sum_{k=1}^{N_b} \frac{1}{R_{k,T}} \begin{bmatrix} \sin(\beta_k) \\ -\cos(\beta_k) \end{bmatrix} \left(\hat{\beta}_k - \beta_k(x_T^*, y_T^*)\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{3.63}$$

No matter how many minimum points the nonlinear likelihood function may have, the true ML solution (x_T^*, y_T^*) must be one of them such that the partial derivative of $L_{\Delta\beta}(x_T, y_T)$ with respect to x_T and y_T at the location (x_T^*, y_T^*) are zeros. Again this is a difficult nonlinear optimization to solve and multiple solutions may exists. Thus we turn our attention to the LS-type algorithm before solving the ML solution.

Recall that the AOA measurements are given by $\hat{\beta}_k = \beta_k + \delta\beta_k$, or $\delta\beta_k = \hat{\beta}_k - \beta_k$. Hence $R_{k,T}\sin(\delta\beta_k) = R_{k,T}\sin(\hat{\beta}_k - \beta_k)$ and thus

$$\begin{aligned} R_{k,T}\sin(\delta\beta_k) &= R_{k,T}\sin(\hat{\beta}_k)\cos(\beta_k) - R_{k,T}\cos(\hat{\beta}_k)\sin(\beta_k) \\ &= \Delta x_k \sin(\hat{\beta}_k) - \Delta y_k \cos(\hat{\beta}_k). \end{aligned} \tag{3.64}$$

where $\Delta x_k = x_T - x_k$, $\Delta y_k = y_T - y_k$, and $R_{k,T} = \sqrt{\Delta x_k^2 + \Delta y_k^2}$. It follows that

$$\varphi_k = -x_k \sin(\hat{\beta}_k) + y_k \cos(\hat{\beta}_k) = -x_T \sin(\hat{\beta}_k) + y_T \cos(\hat{\beta}_k) + R_{k,T}\sin(\delta\beta_k). \tag{3.65}$$

We can regard φ_k as a pseudo-measurement constituting of the real measurements data $\hat{\beta}_k$ and the known BS location (x_k, y_k). For the term $R_{k,T}\sin(\delta\beta_k)$ at the right side of equation (3.65), we argue that even though $\{\sin(\delta\beta_k)\}$ are not Gaussian, they are close to Gaussian distributed provided that the variance σ_β^2 is adequately small by the fact that with $z = \sin(\delta\beta)$ [73],

$$p_Z(z) = \sum_{k=-\infty}^{\infty} \frac{\exp\left[-\frac{1}{2\sigma_\beta^2}\left(\sin^{-1}(z) + 2k\pi\right)^2\right] + \exp\left[-\frac{1}{2\sigma_\beta^2}\left(\sin^{-1}(z) + (2k+1)\pi\right)^2\right]}{\sqrt{2\pi\sigma_\beta^2(1-z^2)}} \tag{3.66}$$

for $|z| \leq 1$ and $p_Z(z) = 0$ for $|z| > 1$. Since σ_β^2 is sufficiently small, there holds

$$p_Z(z) \approx \frac{1}{\sqrt{2\pi\sigma_\beta^2}}\exp\left[-\frac{1}{2\sigma_\beta^2}\left(\sin^{-1}(z)\right)^2\right] \approx \frac{1}{\sqrt{2\pi\sigma_\beta^2}} e^{-\frac{1}{2\sigma_\beta^2}z^2} \tag{3.67}$$

for $z \approx 0$. The above implies $R_{k,T}\sin(\delta\beta_k)$ will behave like a GRV under the condition that $\delta\beta_k$ is very small. This can also be seen in an approximate way that $\sin(\delta\beta_k) \approx \delta\beta_k$, if $\delta\beta_k$ is very small. Hence $\sin(\delta\beta_k)$ and $\delta\beta_k$ will almost have the same PDF. We also would like to argue that the probability for $|\delta\beta| \geq \pi/2$ is zero generically. Otherwise it would imply the wrong direction of the angle of arrival completely. Hence

the PDF of $\delta\beta$ has a shape similar to the normal function but tends to zero for $|\delta\beta| = \pi/2$ and beyond that implies that $p_Z(z)$ behaves closely to Gaussian distributed.

Even if $\delta\beta$ is normal, the exact variance of $\sin(\delta\beta_k)$ can be computed as

$$E\{\sin^2(\delta\beta_k)\} = \frac{1}{2}\, E\{1-\cos(2\delta\beta_k)\} = \frac{1}{2}-\frac{1}{4}\, E\{e^{j2\delta\beta_k}+e^{-j2\delta\beta_k}\} = \frac{1}{2}\left(1-e^{-2\sigma_\beta^2}\right) \approx \sigma_\beta^2 \tag{3.68}$$

for the case when σ_β^2 is sufficiently small. Now the linear equations in (3.65) are of the form

$$\begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_{N_b} \end{bmatrix} = \begin{bmatrix} -\sin(\hat{\beta}_1) & \cos(\hat{\beta}_1) \\ -\sin(\hat{\beta}_2) & \cos(\hat{\beta}_2) \\ \vdots & \\ -\sin(\hat{\beta}_{N_b}) & \cos(\hat{\beta}_{N_b}) \end{bmatrix} \begin{bmatrix} x_T \\ y_T \end{bmatrix} + \begin{bmatrix} R_{1,T}\sin(\delta\beta_1) \\ R_{2,T}\sin(\delta\beta_2) \\ \vdots \\ R_{N_b,T}\sin(\delta\beta_{N_b}) \end{bmatrix}, \tag{3.69}$$

The noise vector on the right hand side is denoted by

$$\vec{\eta}_2 = \begin{bmatrix} R_{1,T}\sin(\delta\beta_1) & R_{2,T}\sin(\delta\beta_2) & \dots & R_{N_b,T}\sin(\delta\beta_{N_b}) \end{bmatrix}^T.$$

It has mean zero and covariance matrix $\mathbf{W}_2(R_{1,T})$ that is diagonal with the k-th element

$$R_{k,T}^2\left(1-e^{-2\sigma_\beta^2}\right)/2 \approx R_{k,T}^2\sigma_\beta^2 = [(x_T-x_k)^2+(y_T-y_k)^2]\sigma_\beta^2. \tag{3.70}$$

With the Gaussian assumption on the noise vector and $\{\varphi_k\}$ as pseudo-measurements, (3.69) has the form

$$\vec{\varphi} = \mathbf{H}_2\vec{\theta} + \vec{\eta}_2 \quad \Longrightarrow \quad J_2 = \frac{1}{2}\left(\vec{\varphi}-\mathbf{H}_2\vec{\theta}\right)^T \mathbf{W}_2^{-1}(R_{1,T})\left(\vec{\varphi}-\mathbf{H}_2\vec{\theta}\right) \tag{3.71}$$

is the objective function. Minimization of J_2 corresponds to the ML algorithm. The

ML solution is given by

$$\hat{\vec{\theta}} = \left(\mathbf{H}_2^T \mathbf{W}_2^{-1}(R_{1,T})\mathbf{H}_2\right)^{-1} \mathbf{H}_2^T \mathbf{W}_2^{-1}(R_{1,T})\vec{\varphi}. \tag{3.72}$$

However $\mathbf{W}_2^{-1}(R_{1,T})$ involves the unknown (x_T, y_T) and $R_{1,T} = \sqrt{x_T^2 + y_T^2}$, the above does not give the ML solution explicitly. It is interesting to notice that the weighted LS problem in this subsection is again a constrained LS-type problem. Indeed by noting that

$$R_{k,T}^2 = x_k^2 + y_k^2 + x_T^2 + y_T^2 - 2(x_k x_T + y_k y_T) = R_k^2 + R_T^2 - 2(x_k x_T + y_k y_T), \tag{3.73}$$

we can multiply (3.72) from left by $\begin{bmatrix} x_k & y_k \end{bmatrix}$ for $k = 2, \cdots, N_b$ to arrive at

$$\rho_{k,T} := x_k x_T + y_k y_T = \begin{bmatrix} x_k & y_k \end{bmatrix} \left(\mathbf{H}_2^T \mathbf{W}_2^{-1}(R_{1,T})\mathbf{H}_2\right)^{-1} \mathbf{H}_2^T \mathbf{W}_2^{-1}(R_{1,T}))\vec{\varphi}. \tag{3.74}$$

In addition $R_{k,T}^2 = R_k^2 + R_T^2 - 2\rho_{k,T}$. Thus taking norm square on both sides of (3.72) yields

$$R_{1,T}^2 = \|\mathbf{\Phi}_2(R_{1,T})\vec{\varphi}\|^2, \qquad \mathbf{\Phi}_2(R_{1,T}) = \left(\mathbf{H}_2^T \mathbf{W}_2^{-1}(R_{1,T})\mathbf{H}_2\right)^{-1} \mathbf{H}_2^T \mathbf{W}_2^{-1}(R_{1,T}). \tag{3.75}$$

Consequently we have N_b equations with N_b unknowns $\{R_{k,T}\}_{k=2}^{N_b}$ and $R_{1,T}$. Although these are nonlinear equations, they can be manipulated to solve at least one set of solutions for these N_b unknowns. These solutions can be substituted back to (3.72) to yield the ML solution (x_T, y_T). It is commented that for large N_b, the complexity for quasi-linear localization based on AOAs is much higher than the corresponding localization based on TDOAs. But if we have additional information on TDOAs,

then the complexity can be reduced tremendously that will be studied in the next subsection.

Before ending this subsection we present the Fisher information matrix associated with the LS-type problem as posed in (3.69). With the assumption on Gaussian distribution for the noise vector $\vec{\eta}_2$, the joint PDF has the expression

$$\text{PDF} = \frac{1}{\sqrt{(2\pi)^n \det[\mathbf{W}_2(R_{1,T})]}} \exp\left\{-\frac{1}{2}\left[\vec{\varphi} - \mathbf{H}_2\vec{\theta}\right]^T \mathbf{W}_2^{-1}(R_{1,T})\left[\vec{\varphi} - \mathbf{H}_2\vec{\theta}\right]\right\} \quad (3.76)$$

where $\vec{\varphi}$ can be regarded as pseudo-measurement vector. Hence $\mathbf{H}_2\vec{\theta}$ is the mean vector and $\mathbf{W}_2(R_{1,T})$ is the covariance matrix. An application of the Slepian and Bangs formula gives the corresponding Fisher information matrix:

$$\begin{aligned} \mathbf{J}_{aoa,LS} &= \sum_{k=1}^{n} \frac{2}{R_{k,T}^2} \begin{bmatrix} \cos(\beta_k) \\ \sin(\beta_k) \end{bmatrix} \begin{bmatrix} \cos(\beta_k) & \sin(\beta_k) \end{bmatrix} \quad (3.77) \\ &\quad + \sum_{k=1}^{n} \frac{1}{R_{k,T}^2 \sigma_\beta^2} \begin{bmatrix} -\sin(\hat{\beta}_k) \\ \cos(\hat{\beta}_k) \end{bmatrix} \begin{bmatrix} -\sin(\hat{\beta}_k) & \cos(\hat{\beta}_k) \end{bmatrix} \\ &\approx \sum_{k=1}^{n} \frac{1}{R_{k,T}^2 \sigma_\beta^2} \begin{bmatrix} -\sin(\hat{\beta}_k) \\ \cos(\hat{\beta}_k) \end{bmatrix} \begin{bmatrix} -\sin(\hat{\beta}_k) & \cos(\hat{\beta}_k) \end{bmatrix} \quad (3.78) \end{aligned}$$

where sufficiently small σ_β^2 is assumed. It is interesting to observe that the above approximate expression is the same as $\mathbf{J}_{aoa}$ in (3.61) except that $\{\beta_k\}$ are replaced by $\{\hat{\beta}_k\}$.

3.2.4 Location based on both TDOA and AOA

After discussing the location techniques based on either TDOA or AOA measurements in the previous two sections, we now assume that both AOAs and TDOAs are available

for target localization. Though it indicates more information and data are needed and consequently costs are increased for a location system, the improved accuracy may pay off all the expense. Hence it is meaningful to study the location method based on a combination of TDOA/AOA in the case of redundant information available and high location resolution mandated. Assuming the independence of the noises (δt_k and $\delta\beta_k$) in measuring the TDOAs and AOAs, the joint PDF is

$$\begin{aligned} p_{\Delta}(\delta t, \delta\beta) &= \frac{\exp\left[-\sum_{k=2}^{N_b}\frac{\left(\Delta\hat{t}_{k,1}-\Delta t_{k,1}(x_T,y_T)\right)^2}{2\sigma_t^2}-\sum_{k=1}^{N_b}\frac{\left(\hat{\beta}_k-\beta_k(x_T,y_T)\right)^2}{2\sigma_\beta^2}\right]}{\sqrt{(2\pi)^{N_b-1}}\sigma_t^{N_b-1}\sqrt{(2\pi)^{N_b}}\sigma_\beta^{N_b}} \\ &= p_{\Delta t}(\delta t)p_{\Delta\beta}(\delta\beta). \end{aligned} \tag{3.79}$$

Because of the independence between $\{\delta t_k\}_{k=2}^{N_b}$ and $\{\delta\beta_k\}_{k=1}^{N_b}$, the Fisher information matrix has the expression

$$\mathbf{J}_{tdoa/aoa} = \mathbf{J}_{tdoa} + \mathbf{J}_{aoa}, \tag{3.80}$$

where $\mathbf{J}_{tdoa}$ and $\mathbf{J}_{aoa}$ are the same as in (3.34) and (3.61), respectively. This can be easily shown [74] by

$$\begin{aligned} \mathbf{J}_{tdoa/aoa} &= \mathrm{E}\left\{\left[\frac{\partial[\ln(\mathrm{p}_{\Delta}(\delta\mathrm{t},\delta\beta))]}{\partial\vec{\mathrm{x}}}\right]\left[\frac{\partial[(\ln \mathrm{p}_{\Delta}(\delta\mathrm{t},\delta\beta))]}{\partial\vec{\mathrm{x}}}\right]^{\mathrm{T}}\right\} \\ &= \mathrm{E}\left\{\left[\frac{\partial[\ln(\mathrm{p}_{\Delta\mathrm{t}}(\delta\mathrm{t}))]}{\partial\vec{\mathrm{x}}}+\frac{\partial[\ln(\mathrm{p}_{\Delta\beta}(\delta\beta))]}{\partial\vec{\mathrm{x}}}\right]\left[\frac{\partial[\ln(\mathrm{p}_{\Delta\mathrm{t}}(\delta\mathrm{t}))]}{\partial\vec{\mathrm{x}}}+\frac{\partial[\ln(\mathrm{p}_{\Delta\beta}(\delta\beta))]}{\partial\vec{\mathrm{x}}}\right]^{\mathrm{T}}\right\} \\ &= \mathrm{E}\left\{\left[\frac{\partial[\ln(\mathrm{p}_{\Delta\mathrm{t}}(\delta\mathrm{t}))]}{\partial\vec{\mathrm{x}}}\right]\left[\frac{\partial[(\ln \mathrm{p}_{\Delta\mathrm{t}}(\delta\mathrm{t}))]}{\partial\vec{\mathrm{x}}}\right]^{\mathrm{T}}\right\}+\mathrm{E}\left\{\left[\frac{\partial[\ln(\mathrm{p}_{\Delta\beta}(\delta\beta))]}{\partial\vec{\mathrm{x}}}\right]\left[\frac{\partial[(\ln \mathrm{p}_{\Delta\beta}(\delta\beta))]}{\partial\vec{\mathrm{x}}}\right]^{\mathrm{T}}\right\} \\ &= \mathbf{J}_{tdoa}+\mathbf{J}_{aoa}. \end{aligned} \tag{3.81}$$

With respect to the joint PDF in (3.79), the corresponding likelihood-type function in this case has the form

$$L(x_T, y_T) = \sum_{k=2}^{N_b} \frac{1}{c^2\sigma_t^2} \left(c\Delta\hat{t}_{k,1} - R_{k,T}(x_T, y_T) + R_{1,T}\right)^2 + \sum_{k=1}^{N_b} \frac{1}{\sigma_\beta^2} \left(\hat{\beta}_k - \beta_k(x_T, y_T)\right)^2. \tag{3.82}$$

The ML algorithm seeks the maximum of the above likelihood function. The necessary condition for it to achieve maximum at (x_T^*, y_T^*) is:

$$\begin{aligned} \begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \textstyle\sum_{k=1}^{N_b} \frac{1}{R_{k,T}} \begin{bmatrix} \sin(\beta_k) \\ -\cos(\beta_k) \end{bmatrix} \left(\hat{\beta}_k - \beta_k(x_T^*, y_T^*)\right) \\ &+ \textstyle\sum_{k=2}^{N_b} \begin{bmatrix} \dfrac{x_k - x_T^*}{R_{k,T}} + \dfrac{x_T^*}{R_{1,T}} \\ \dfrac{y_k - y_T^*}{R_{k,T}} + \dfrac{y_T^*}{R_{1,T}} \end{bmatrix} \left(c\Delta\hat{t}_{k,1} - R_{k,T} + R_{1,T}\right). \end{aligned} \tag{3.83}$$

The Newton-type algorithms can be applied to solve the ML solution. Clearly the ML solution to the above nonlinear equations is hard to compute that may not be a global maximum for $L(x_T, y_T)$. An alternative method is the use of LS-type algorithm as in the previous two subsections. One possible way is to compute the constrained LS solutions $(\hat{x}_T^{(\mathrm{TDOA})}, \hat{y}_T^{(\mathrm{TDOA})})$ and $(\hat{x}_T^{(\mathrm{AOA})}, \hat{y}_T^{(\mathrm{AOA})})$ based on TDOAs and AOAs separately as in the previous subsections and then combine the two as [53]

$$\hat{x}_T = \gamma x_T^{(\mathrm{AOA})} + (1-\gamma)\hat{x}_T^{(\mathrm{TDOA})}, \qquad \hat{y}_T = \gamma y_T^{(\mathrm{AOA})} + (1-\gamma)\hat{y}_T^{(\mathrm{TDOA})} \tag{3.84}$$

where $0 < \gamma < 1$. Note that $R_{k,T} = R_{1,T} + c\Delta t_{k,T}$ can be used in (3.69) to avoid computing N_b unknowns with N_b equations. Indeed the noise terms in (3.69) have

zero mean and variance

$$E\{R_{k,T}^2 \sin^2(\beta_k)\} = \ E\{[R_{1,T} + c\Delta\hat{t}_{k,1} - c\delta t_k]^2 \sin^2(\beta_k)\} \approx [(R_{1,T} + c\Delta\hat{t}_{k,1})^2 + c^2\sigma_t^2]\sigma_\beta^2 \tag{3.85}$$

if σ_β^2 is sufficiently small. Hence only one unknown R_T is involved and $\rho_{k,T}$ are all eliminated which helps to simplify the computation of the LS-type solution to the target localization problem based on measurements of AOAs. However the determination of the optimal value of γ is not easy. Hence we opt to compute the LS-type solution directly.

Since both AOAs and TDOAs are available, we have the following linear equations:

$$\begin{bmatrix} \vec{a} - \vec{b}R_{1,T} \\ \vec{\varphi} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix} \begin{bmatrix} x_T \\ y_T \end{bmatrix} + \begin{bmatrix} \vec{\eta}_1 \\ \vec{\eta}_2 \end{bmatrix}. \tag{3.86}$$

Under the independence assumption for the noises $\vec{\eta}_1$ and $\vec{\eta}_2$, we have

$$\mathrm{E}\left\{ \begin{bmatrix} \vec{\eta}_1 \\ \vec{\eta}_2 \end{bmatrix} \right\} = 0, \qquad \mathbf{W} = \mathrm{E}\left\{ \begin{bmatrix} \vec{\eta}_1 \\ \vec{\eta}_2 \end{bmatrix} \begin{bmatrix} \vec{\eta}_1 & \vec{\eta}_2 \end{bmatrix} \right\} = \begin{bmatrix} \mathbf{W}_1(R_{1,T}) & 0 \\ 0 & \mathbf{W}_2(R_{1,T}) \end{bmatrix} \tag{3.87}$$

where the kth diagonal element of $\mathbf{W}_2(R_{1,T})$ is the same as in (3.85). By assuming uncorrelated Gaussian for $\vec{\eta}_1$ and $\vec{\eta}_2$, the ML solution to estimation of (x_T, y_T) can be computed through minimization the following objective function:

$$\begin{aligned} J_{1,2} \ &= \tfrac{1}{2}\left[\vec{a} - \vec{b}R_{1,T} - \mathbf{H}_1\vec{\theta}\right]^T \mathbf{W}_1^{-1}(R_{1,T})\left[\vec{a} - \vec{b}R_{1,T} - \mathbf{H}_1\vec{\theta}\right] \\ &\quad + \tfrac{1}{2}\left(\vec{\varphi} - \mathbf{H}_2\vec{\theta}\right)^T \mathbf{W}_2^{-1}(R_{1,T})\left(\vec{\varphi} - \mathbf{H}_2\vec{\theta}\right) = J_1 + J_2. \end{aligned} \tag{3.88}$$

Taking derivative of the cost function $J_{1,2}$ with respect to $\vec{\theta}$, we have

$$\frac{\partial J_{1,2}}{\partial \vec{\theta}} = -(\vec{a} - \vec{b}R_{1,T} - \mathbf{H}_1\vec{\theta})^T\mathbf{W}_1^{-1}\mathbf{H}_1 - (\vec{\varphi} - \mathbf{H}_2\vec{\theta})^T\mathbf{W}_2^{-1}\mathbf{H}_2. \tag{3.89}$$

By letting $\frac{\partial J_{1,2}}{\partial \vec{\theta}} = 0$, tt can be easily shown that the minimizer to the cost function $J_{1,2}$ is given by

$$\begin{bmatrix} \hat{x}_T \\ \hat{y}_T \end{bmatrix} = \left[\mathbf{H}_1^T \mathbf{W}_1^{-1}(R_{1,T})\mathbf{H}_1 + \mathbf{H}_2^T \mathbf{W}_2^{-1}(R_{1,T})\mathbf{H}_2\right]^{-1} \left[\mathbf{H}_1^T \mathbf{W}_1^{-1}(R_{1,T})\left(\vec{a} - \vec{b}R_{1,T}\right) + \mathbf{H}_2^T \mathbf{W}_2^{-1}(R_{1,T})\vec{\varphi}\right]. \tag{3.90}$$

Because the above solution involves an unknown $R_{1,T} = \sqrt{x_T^2 + y_T^2}$, we can again take norm square both sides to obtain an equation for $R_{1,T}$ first, and after computing its solution, the value of $R_{1,T}$ can be substituted into (3.90) to obtain the solution to the weighted LS problem. Note that $R_{1,T}$ is a positive real root to some nonlinear equation. One of the positive real roots corresponds to the constrained LS solution, which provides an initial guess for the true (nonlinear) ML solution.

It is commented that the optimal solution in (3.90) is not in the form of the convex combination of the two separate LS-type solutions as in (3.84). Rather it is in the form

$$\begin{bmatrix} \hat{x}_T \\ \hat{y}_T \end{bmatrix} = \Gamma \begin{bmatrix} \hat{x}_T^{(\text{TDOA})} \\ \hat{y}_T^{(\text{TDOA})} \end{bmatrix} + (I - \Gamma) \begin{bmatrix} \hat{x}_T^{(\text{AOA})} \\ \hat{y}_T^{(\text{AOA})} \end{bmatrix} \tag{3.91}$$

where Γ is a matrix. Specifically the solution in (3.90) can be written as

$$\begin{bmatrix} \hat{x}_T \\ \hat{y}_T \end{bmatrix} = [A_1 + A_2]^{-1}[B_1 + B_2] = \left[I + A_1^{-1}A_2\right]^{-1} \begin{bmatrix} \hat{x}_T^{(\text{TDOA})} \\ \hat{y}_T^{(\text{TDOA})} \end{bmatrix} + \left[I + A_2^{-1}A_1\right]^{-1} \begin{bmatrix} \hat{x}_T^{(\text{AOA})} \\ \hat{y}_T^{(\text{AOA})} \end{bmatrix} \tag{3.92}$$

where

$$\begin{aligned} A_1 &= \mathbf{H}_1^T \mathbf{W}_1^{-1}(R_{1,T})\mathbf{H}_1; \\ A_2 &= \mathbf{H}_2^T \mathbf{W}_2^{-1}(R_{1,T})1, T_2; \\ B_1 &= \mathbf{H}_1^T \mathbf{W}_1^{-1}(R_{1,T})\left(\vec{a} - \vec{b}R_{1,T}\right); \\ B_2 &= \mathbf{H}_2^T \mathbf{W}_2^{-1}(R_{1,T})\vec{\varphi}. \end{aligned}$$

Hence $A_1^{-1}B_1$ and $A_2^{-1}B_2$ are the LS-type solution based on TDOAs and AOAs, respectively. Now it is straightforward to show that

$$\left[I + A_1^{-1}A_2\right]^{-1} + \left[I + A_2^{-1}A_1\right]^{-1} = [A_1 + A_2]^{-1}A_1 + [A_1 + A_2]^{-1}A_2 = \Gamma + [I - \Gamma] = I. \tag{3.93}$$

Even though the LS solution in (3.90) is some kind of combination of the two separate LS solutions in (3.46) and (3.72), the unknown $R_{1,T}$ has to be computed based on (3.90).

Finally the Fisher information matrix associated with the linear model in (3.86) is

$$\begin{aligned}
\mathbf{P}_{tdoa/aoa-fim,LS} &= \sum_{k=2}^{N_b} \left(\frac{1}{\sigma_t^2(\alpha_k^2 + 2\sigma_t^2)} \right) \begin{bmatrix} 2x_k/c^2 + b_k\cos(\beta_1) \\ 2y_k/c^2 + b_k\sin(\beta_1) \end{bmatrix} \begin{bmatrix} 2x_k/c^2 + b_k\cos(\beta_1) \\ 2y_k/c^2 + b_k\sin(\beta_1) \end{bmatrix}^T \\
&\quad + \sum_{k=2}^{N_b} \left(\frac{8\alpha_k^2}{c^2(\alpha_k^2 + 2\sigma_t^2)^2} \right) \begin{bmatrix} \cos(\beta_1) \\ \sin(\beta_1) \end{bmatrix} \begin{bmatrix} \cos(\beta_1) & \sin(\beta_1) \end{bmatrix} \\
&\quad + \sum_{k=1}^{N_b} \frac{1}{R_{k,T}^2\sigma_\beta^2} \begin{bmatrix} -\sin(\hat{\beta}_k) \\ \cos(\hat{\beta}_k) \end{bmatrix} \begin{bmatrix} -\sin(\hat{\beta}_k) & \cos(\hat{\beta}_k) \end{bmatrix}
\end{aligned} \tag{3.94}$$

under the uncorrelated Gaussian assumption and sufficiently small σ_t^2 and σ_β^2.

3.3 Constrained Least-square Optimization

As shown in 3.46, the weighted LS solution $\hat{\vec{\theta}}$ is constrained by

$$R_{1,T}^2 = \|\mathbf{\Phi}_1[\vec{a} - \vec{b}R_{1,T}^2]\|^2, \tag{3.95}$$

from which some solutions $R_{1,T}$ can be solved. If there exist real solutions $R_{1,T}$, they can be substituted back into J_1 in 3.45 and obtain the optimal solution $R_{1,T}$ based on

which the optimal solution $\hat{\vec{\theta}}$ can be obtained. While this holds, (3.95) may not admit a real solution $R_{1,T}$ due to the existence of noises in observations. More specifically (3.95) is equivalent to the quadratic equation

$$(\vec{b}^T\mathbf{\Phi}^T\mathbf{\Phi}\vec{b}-1)R_{1,T}^2-2\vec{a}^T\mathbf{\Phi}^T\mathbf{\Phi}\vec{b}R_{1,T}+\vec{a}^T\mathbf{\Phi}^T\mathbf{\Phi}\vec{a}=0, \tag{3.96}$$

which admit real solution, if and only if

$$(\vec{a}^T\mathbf{\Phi}^T\mathbf{\Phi}\vec{b})^2+\vec{a}^T\mathbf{\Phi}^T\mathbf{\Phi}\vec{a}-(\vec{a}^T\mathbf{\Phi}^T\mathbf{\Phi}\vec{a})(\vec{b}^T\mathbf{\Phi}^T\mathbf{\Phi}\vec{b})\geq 0. \tag{3.97}$$

That is, (3.95) admits a real solution $R_{1,T}$ if and only if (3.97) holds. Simulation in [54] shows that the location estimate in (3.46) is very accurate if the condition (3.97) holds; Otherwise the location estimate is far away from the true location. The question is what if (3.97) does not hold which is generically true due to the existence of noise in the TDOA and AOA measurements.

Let us examine (3.45) again by rewriting J_1 into

$$J_1=\frac{1}{2}\left(\vec{a}-\begin{bmatrix}\mathbf{H}_1 & \vec{b}\end{bmatrix}\begin{bmatrix}p_T\\ R_{1,T}\end{bmatrix}\right)^T\mathbf{W}_1^{-1}\left(\vec{a}-\begin{bmatrix}\mathbf{H}_1 & \vec{b}\end{bmatrix}\begin{bmatrix}p_T\\ R_{1,T}\end{bmatrix}\right), \tag{3.98}$$

where $p_T=[\ \hat{x}_T\quad \hat{y}_T\ ]^T$. The nonlinear estimation problem aims to search p_T and $R_{1,T}$ such that J_1 is minimized, subject to the constraint $R_{1,T}=\|p_T\|$. Denote $\mathbf{\Sigma}=\mathbf{W}_1$ and

$$\mathbf{A}=\begin{bmatrix}\mathbf{H}_1 & \vec{b}\end{bmatrix},\quad \vec{\varphi}=\vec{a},\quad \vec{\theta}=\begin{bmatrix}p_T\\ R_{1,T}\end{bmatrix},\quad \mathbf{Q}=\begin{bmatrix}-\mathbf{I}_2 & 0\\ 0 & 1\end{bmatrix}.$$

Then we have the following more general constrained LS optimization problem:

$$\min_{\vec{\theta}^T\mathbf{Q}\vec{\theta}=0} J_1,\quad J_1=\tfrac{1}{2}(\mathbf{A}\vec{\theta}-\vec{\varphi})^T\mathbf{\Sigma}^{-1}(\mathbf{A}\vec{\theta}-\vec{\varphi}). \tag{3.99}$$

We will develop a solution algorithm to such a constrained LS optimization problem in the following. Assume that $\mathbf{\Sigma}$ is positive definite, $\mathbf{A}$ has full column rank and $\mathbf{Q}$ is nonsingular that has both positive and negative eigenvalues, i.e., $\mathbf{Q}$ is indefinite. We employ Lagrange multiplier to develop the solution algorithm. Let λ be real and consider

$$J = \frac{1}{2}\left[(\mathbf{A}\vec{\theta} - \vec{\varphi})^T\mathbf{\Sigma}^{-1}(\mathbf{A}\vec{\theta} - \vec{\varphi}) + \lambda\vec{\theta}^T\mathbf{Q}\vec{\theta}\right]. \tag{3.100}$$

Then the necessary condition for optimality yields the condition

$$\mathbf{A}^T\mathbf{\Sigma}^{-1}[\mathbf{A}\vec{\theta} - \vec{\varphi}] + \lambda\mathbf{Q}\vec{\theta} = \mathbf{0} \;\;\Leftrightarrow\;\; \vec{\theta} = [\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A} + \lambda\mathbf{Q}]^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1}\vec{\varphi}. \tag{3.101}$$

An optimal solution needs to satisfy the constraint $\vec{\theta}^T\mathbf{Q}\vec{\theta} = 0$ leading to

$$\vec{\varphi}^T\mathbf{\Sigma}^{-1}\mathbf{A}[\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A} + \lambda\mathbf{Q}]^{-1}\mathbf{Q}[\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A} + \lambda\mathbf{Q}]^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1}\vec{\varphi} = 0. \tag{3.102}$$

The solution algorithm hinges to the computation of the real root λ from the above equation and there can be more than one such real root. We employ the result of simultaneous diagonalization. Because $\mathbf{\Sigma} = \mathbf{\Sigma}^T > 0$ and $\mathbf{Q} = \mathbf{Q}^T > 0$, there exists a nonsingular matrix $\mathbf{S}$ such that $\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A} = \mathbf{S}\mathbf{D}_\Sigma\mathbf{S}^T$ and $\mathbf{Q} = \mathbf{S}\mathbf{D}_Q\mathbf{S}^T$ where $\mathbf{D}_\Sigma$ and $\mathbf{D}_Q$ are both diagonal. It is noted that $\mathbf{D}_\Sigma$ and $\mathbf{D}_Q$ have the same inertia as $\mathbf{\Sigma}$ and $\mathbf{Q}$, respectively. It follows that (3.102) is equivalent to

$$(\mathbf{S}^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1}\vec{\varphi})^T(\lambda\mathbf{I} + \mathbf{D}_\Sigma\mathbf{D}_Q^{-1})^{-1}\mathbf{D}_Q^{-1}(\lambda\mathbf{I} + \mathbf{D}_\Sigma\mathbf{D}_Q^{-1})^{-1}(\mathbf{S}^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1}\vec{\varphi}) = 0. \tag{3.103}$$

Let $\mathbf{D}_Q^{-1} = \mathrm{diag}(\mathrm{q}_1, \mathrm{q}_2, \ldots, \mathrm{q}_l)$ with $l \times l$ the size of $\mathbf{Q}$. Then it has the same number of negative and positive elements as $\mathbf{D} = \mathbf{D}_\Sigma\mathbf{D}_Q^{-1} = \mathrm{diag}(\mathrm{d}_1, \mathrm{d}_2, \ldots, \mathrm{d}_l)$ by the positivity

of $\mathbf{\Sigma}$ and $\mathbf{D}_\Sigma$. In fact, $q_i d_i > 0$. The matrices $\mathbf{S}$ and $\mathbf{D}$ can be obtained by eigenvalue decomposition of $\mathbf{A}^T\mathbf{S}^{-1}\mathbf{A}\mathbf{Q}^{-1} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}$. Let v_i be the i-th element of $\mathbf{S}^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1}\vec{\varphi}$. Then (3.103) is converted into the following:

$$(\mathbf{S}^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1}\vec{\varphi})^T(\lambda\mathbf{I}+\mathbf{D}_\Sigma\mathbf{D}_Q^{-1})^{-1}\mathbf{D}_Q^{-1}(\lambda\mathbf{I}+\mathbf{D}_\Sigma\mathbf{D}_Q^{-1})^{-1}(\mathbf{S}^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1}\vec{\varphi}) = \sum_{i=1}^{l}\frac{q_i v_i^2}{(\lambda+d_i)^2} = 0. \tag{3.104}$$

We comment that the above has real roots by examining the summation at $\lambda \approx -d_i$ and by the fact that $\{q_i\}$ have both positive and negative values but not zero. Recall the assumption on $\mathbf{Q}$. However there are only finitely many real λ values satisfying (3.104), which are denoted by $\{\lambda_k\}$. Now by (3.101),

$$\begin{aligned}
\mathbf{A}\vec{\theta} - \vec{\varphi} &= [\mathbf{A}(\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A} + \lambda_k\mathbf{Q})^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1} - \mathbf{I}]\vec{\varphi} \\
&= \left[(\mathbf{A}\mathbf{Q})^{-1}(\lambda_k\mathbf{I} + \mathbf{A}^T\mathbf{\Sigma}^{-1}(\mathbf{A}\mathbf{Q})^{-1})^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1} - \mathbf{I}\right]\vec{\varphi} \\
&= \left[(\lambda_k\mathbf{I} + \mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1})^{-1}\mathbf{A}\mathbf{Q})^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1} - \mathbf{I}\right]\vec{\varphi} \\
&= -\lambda_k(\lambda_k\mathbf{I} + \mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1})^{-1}\vec{\varphi} \\
&= -\lambda_k\mathbf{\Sigma}(\lambda_k\mathbf{\Sigma} + \mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T)^{-1}\vec{\varphi}.
\end{aligned} \tag{3.105}$$

Substituting the above into the performance index J in (3.100) leads to

$$2J = \lambda_k^2\vec{\varphi}^T(\lambda_k\mathbf{\Sigma} + \mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T)^{-1}\mathbf{\Sigma}(\lambda_k\mathbf{\Sigma} + \mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T)^{-1}\vec{\varphi}. \tag{3.106}$$

Let λ_k^{opt} be the value that minimizes J over $\{\lambda_k\}$. Then in light of (3.101), the optimal $\vec{\theta}$ is obtained as

$$\vec{\theta}^{\text{opt}} = [\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A} + \lambda_k^{\text{opt}}\mathbf{Q}]^{-1}\mathbf{A}^T\mathbf{\Sigma}^{-1}\vec{\varphi}. \tag{3.107}$$

To facilitate the MATLAB programming in simulation for roots computation we can convert (3.104) to

$$\sum_{i=1}^{l} q_i v_i^2 \prod_{k\neq i}(\lambda+d_k)^2 = 0. \tag{3.108}$$

Obviously the solution algorithm above is developed for location estimation with TDOA measurements available only. If both TDOA and AOA measurements are collected, as discussed in the previous section, the extra redundancy indicates an improved accuracy. According to (3.86), we can formulate it into a similar constrained LS optimization problem. Denote

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{W}_1 & \\ & \mathbf{W}_2 \end{bmatrix}; \quad \mathbf{A} = \begin{bmatrix} \mathbf{H}_1 & \vec{b} \\ \mathbf{H}_2 & 0 \end{bmatrix}; \vec{\theta} = \begin{bmatrix} p_T \\ R_{1,T} \end{bmatrix}; \vec{\varphi} = \begin{bmatrix} \vec{a} \\ \vec{\phi} \end{bmatrix}; \mathbf{Q} = \begin{bmatrix} -\mathbf{I}_2 & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.109}$$

Then we can use the same Lagrange multiplier method to give a solution.

3.4 Simulations

In this section, we present a set of simulation results that demonstrate the performance of our proposed estimation algorithm.

In the simulation, there are nine base stations which are equally spaced around a circle. In real WiMax system, the base stations may not exactly locate on a circle. This is simply for ease of presentation and it is not necessarily required in our algorithm which is applicable to any geographical distribution of any number of base stations. To test the accuracy of our location method, ten positions for the mobile user are chosen and they are distributed around a smaller circle too. For the same purpose of an easy demonstration, the above assumption about the MS route is made. The configuration is shown in Figure 3.5.

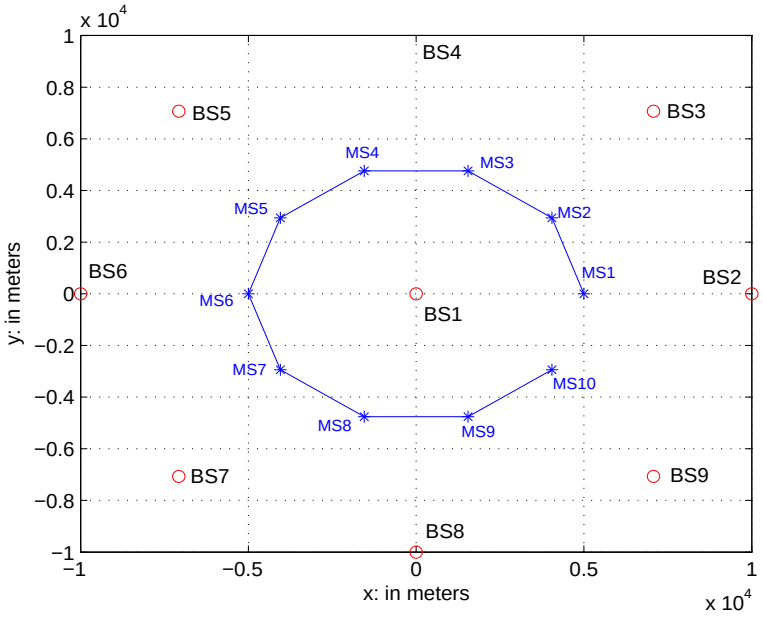

Figure 3.5: Base stations and mobile user locations

The base stations are at BS1 = $[0, 0]^{\mathrm{T}}$, BS2 = $[32000, 0]^{\mathrm{T}}$, BS3 = $[22627, 22627]^{\mathrm{T}}$, BS4 = $[0, 32000]^{\mathrm{T}}$, BS5 = $[-22627, 22627]^{\mathrm{T}}$, BS6 = $[-32000, 0]^{\mathrm{T}}$, BS7 = $[-22627, -22627]^{\mathrm{T}}$, BS8 = $[0, -32000]^{\mathrm{T}}$, BS9 = $[22627, -22627]^{\mathrm{T}}$. The unit is in meters. For each MS position, total number of 2000 different data sets are run and the MS location is obtained by averaging over all the 2000 estimates.

In the experiments, our location algorithm is simulated for TDOA data only and for a combination of AOA and TDOA data, respectively. In Figure 3.6, the green line is the result from a combination of AOA and TDOA data when the SNR's are $\mathrm{SNR}_{\mathrm{tdoa}}$ = 20dB and $\mathrm{SNR}_{\mathrm{aoa}}$ = 20dB respectively. It almost merges with the blue line which represents the real MS positions and is invisible in the figure. It shows the high accuracy of the estimation algorithm we propose in this thesis. With the same $\mathrm{SNR}_{\mathrm{tdoa}}$ = 20dB, the cyan line is the estimation result from the TDOA data only. It can be seen that there is small deviation from the real position. Intuitively, with the extra information from AOA measurement, the result in the green line is expected to be closer to the real positions. From the Fisher information matrices we calculated in the previous sections, the Cramer-Rao bound for the combination data of TDOA and AOA should be smaller than that of TDOA data only.

To have a closer look at the performance of the proposed algorithm, we calculate the approximate mean and standard deviation of the estimation error, i.e., the distance between the real MS position and the estimated position. It is obtained from a sample space of 2000 data points. In Figure 3.7, the average estimation error is less

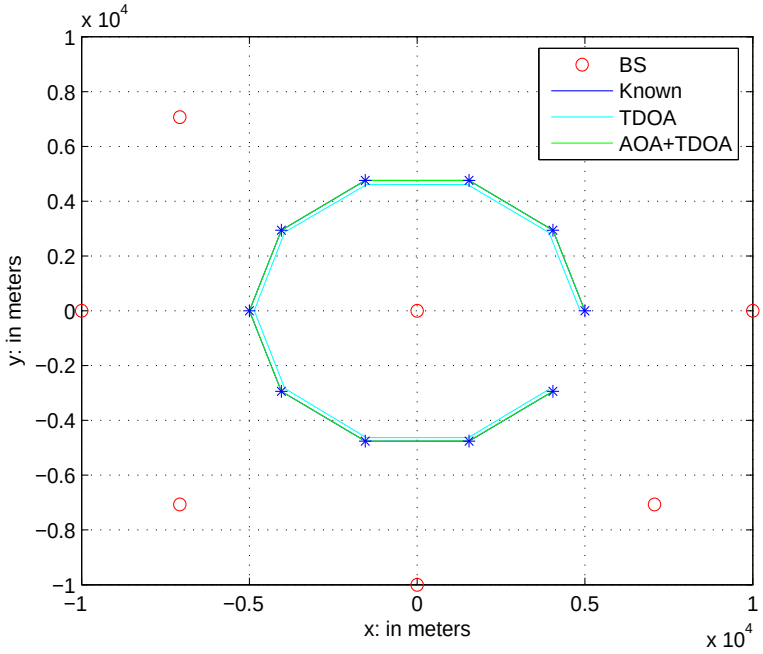

Figure 3.6: Location estimation with TDOA-only and AOA+TDOA data

than 4 meters for all the ten MS locations when the TDOA data is of high SNR. To study the effect of SNR on the performance of the proposed location algorithm, the MS position at MS2 is randomly selected and the mean and the standard deviation of the estimation error vary with SNR as shown in Figure 3.8. It is easily seen that at a low SNR, the estimation is not accurate enough and it is because our assumption about the measurement noise variance is not valid.

According to the FCC regulations, it requires that for 67% of the E911 calls, the wireless service providers must provide an estimated location with location error below 100m. As shown in Figure 3.9, the location error is below 100m for 98% of the time with $\mathrm{SNR}_{\mathrm{tdoa}} = 40\mathrm{dB}$. It is well above the requirement from FCC.

From the above figures, it is demonstrated that the proposed algorithm can provide

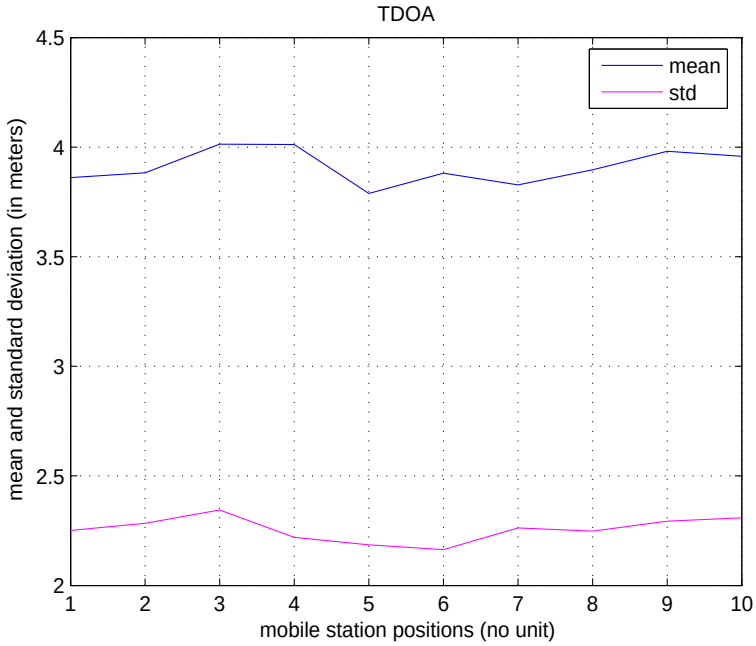

Figure 3.7: Location estimation performance

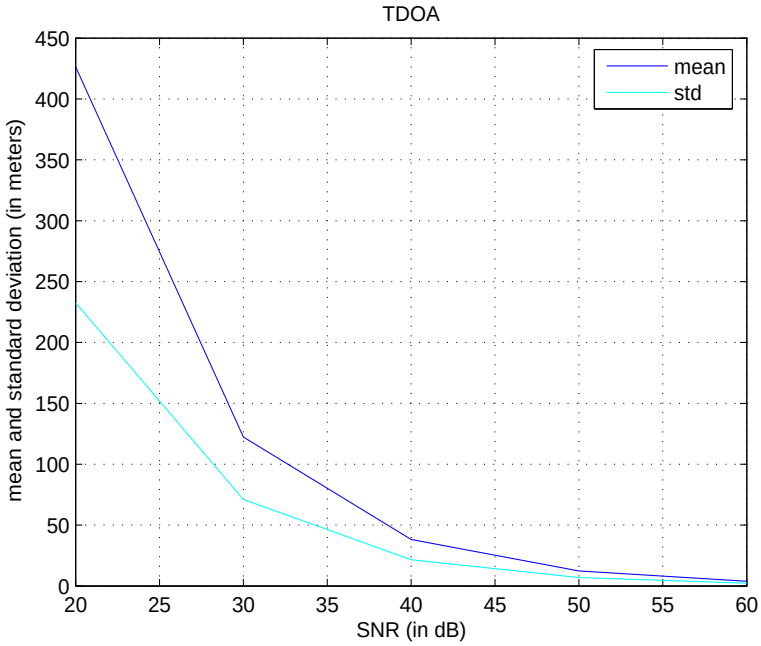

Figure 3.8: Effect of SNR on estimation accuracy

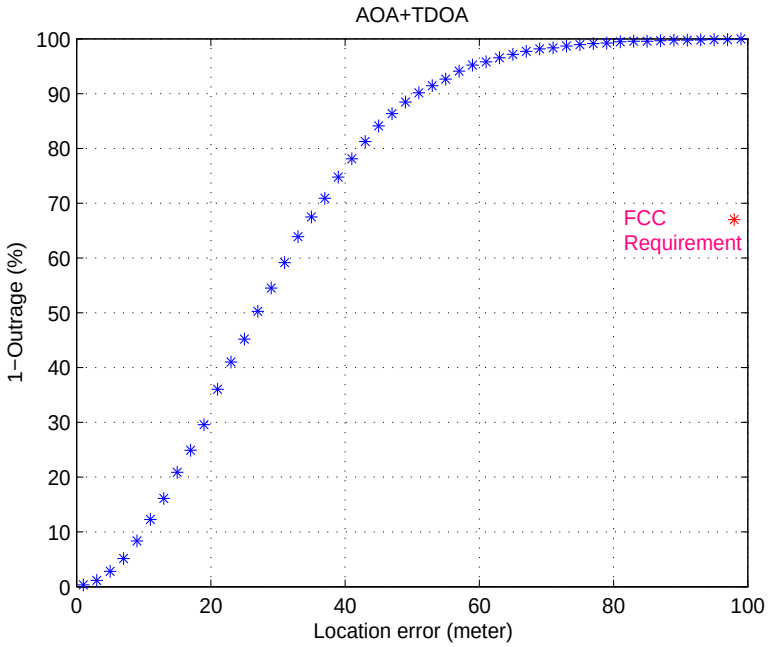

Figure 3.9: Outrage curve for location accuracy

accurate estimation for the MS location. It also meets the FCC requirement for outdoor network-based wireless location.

3.5 Chapter Summary

In this chapter, an introduction about WiMax networks and its IEEE standard evolution and applications in most aspects is given and the outdoor/indoor wireless location technologies based on measurements of TOA's, TDOA's, AOA's and amplitudes are reviewed.

With measurements of TDOA and AOA available, we present a constrained LS-type algorithm to estimate the target location. The proposed method is different from the commonly used ML algorithm, though the latter is heavily preferred in

some applications for its superior performance. Because of the large number of observation data and the additive measurement noise, maximizing the likelihood function involves a great amount of computational load. It even does not guarantee that the optimal estimation can be obtained due to the existence of local minimum. Under the assumption of zero-mean additive Gaussian noise with a very small variance, the location estimation problem is formulated into a quasi-linear form, which is solvable by the LS algorithm. The assumption is usually validated as in [54]. Therefore, our method holds the preferable properties of the ML algorithm in the sense that it approaches the Cramer-Rao bound with a large sample of observation data. More importantly, the computational complexity is reduced by the LS algorithm. As shown in this chapter, the LS algorithm also involves a constraint that $\|\hat{\vec{\theta}}\| = \|R_{1,T}\|$. The target location can only be obtained by substituting the intermediate LS solution into the constraint and solving the resultant quadratic equation. It brings complexity back to the solution. Hence the Lagrange multiplier is explored to solve the above constrained LS optimization problem. The simulation results show that our scheme is effective in location estimation.

Chapter 4

Conclusions

This dissertation, in the first part, addresses the problem of channel estimation of MIMO-OFDM systems. It starts from the matrix representation of the signal model of MIMO-OFDM systems, which clearly describes the relation of signals in frequency domain and time domain and expressing operations like adding CP and removing CP as matrix product. From the resulting MIMO-OFDM signal model, a pilot tone based channel estimation is proposed to estimate the fast time-varying and frequency-selective fading channel via the least-squares method. The least-squares is selected for the purpose of low complexity, though some other methods such as MMSE and ML may produce better estimation performance. To further reduce the computational complexity, the pilot tone matrix is designed as a unitary matrix to save the computation of the matrix inversion in the standard LS solution. The pilot tone matrix is designed in a simple way that N_t disjoint pilot tone sets are placed at one OFDM block on each transmit antenna. Each pilot tone set has L pilot tones which are equally-spaced and equally-powered. By choosing the pilot tones based on our de-

sign, those pilot tones comprise a unitary matrix. For a simple 2×2 case, Alamouti's orthogonal structure is exploited. And the design can be readily extended to a configurable MIMO-OFDM system with any number of transmit and receive antennas. For a fixed power of pilot tones, our design can be proved to be also optimal in the sense of achieving the minimum MSE of channel estimation. Compared with some relative pilot tone designs in the literature, our channel estimation method differs in its ability to estimate fast time-varying wireless channel since pilot tones are inserted into each OFDM block, and in its explicit relation with space-frequency code design which can benefit the channel estimation in return. Seeking for a robust channel estimator with lower complexity for MIMO-OFDM systems, we are looking at the following aspects in the future.

- Less overhead loss: It is worth noting that the use of pilot symbols for channel estimation decrease the spectrum efficiency of the wireless communication systems. It is a trade off between data throughput and estimation accuracy. It is of interest to investigate a scheme with even fewer number of pilot tones in each OFDM block by exploiting some statistical properties of the wireless channel itself. Intuitively, it is the best balance between overhead loss and estimation reliability if we can adaptively change the number of pilot tones depending on the channel condition through some feedback information.

- Joint channel estimation and CFO correction: Usually when we design the channel estimator, we assume that the OFDM system is perfectly synchronized and

there is carrier frequency offset at all. And some CFO compensation algorithms are also based on the assumption that channel is known at the receiver. It would be beneficial to combine the channel estimation and CFO compensation into an integrate algorithm since the performance of either one of the two individual algorithms can be degrade by the invalidity of their assumptions in the real world OFDM systems. There are already some research work in this area [34, 35], but more intensive investigation is still needed.

But we still have to consider the data rate loss caused by the pilot-tone overhead within each OFDM block. We are currently working on this issue with a goal that we can use a sequence of pilot-tones with length less than the channel length by exploring its diversity in the time domain.

In the second part of this dissertation, the wireless location on WiMax network is studied. Similar to the location technology applied to the cellular networks, the application scenario of locating the mobile user by using some signal parameters received at the antenna towers is considered. Location estimation methods based on TDOA, AOA and a combination of TDOA and AOA are presented, respectively. With the assumption that the measurement noise is zero-mean additive Gaussian noise with very small variance, the location estimation problem is formulated into a quasi-linear form. Then the simple LS algorithm can be used to solve the estimation problem, provided that the noise term in the quasi-linear form is Gaussian. In theory, the ML algorithm can be directly utilized to estimate the target location since the probability

density function of the observation data is known with our assumption. However, direct use of ML algorithm proves infeasible because of the difficulty of finding the real roots of a quadratic equation. An alternative to the ML algorithm is required, which should drastically reduce the complexity of the ML algorithm and provide a close performance. Our proposed method is such an alternative that it is essentially a constrained LS-type optimization technique. The approximation of the noise term in the quasi-linear form to a Gaussian random is also proved in this thesis under the assumption above. Hence it is concluded that the proposed method can estimate the target location very accurately, provided that the size of the observation data is large enough and the equivalent SNR is high. To solve the constrained LS-type optimization problem, the Lagrange multiplier method is used. It is because that the direct use of the constraint condition may lead to the same level of complexity for the algorithm and even positive real roots may not exist in the quadratic equation obtained by substituting the intermediate LS solution into the constraint. Finally,the extensive simulation studies has demonstrated the effectiveness of our proposed algorithm.

For future work on wireless location problem, the following aspects are open for research.

- Large variance: The approximation of the constrained LS-type optimization to the ML algorithm is dependent on the assumption that the measurement noise variance is very small, which is usually true. Further research on the case of

measurement noise with relatively large variance will improve the robustness of the proposed algorithm.

- Velocity Estimation: In the thesis, the target is considered stationary by assuming it is moving at a low speed. If the FDOA (frequency difference of arrivals) of the received signal is available, then the velocity of the target can be estimated too. This will extend the range of applications of the proposed algorithm.

Bibliography

[1] Richard Van Nee and Ramjee Prasad, *OFDM For Wireless Multimedia Communications*, Artech House Publishers, Norwood MA, 2000.

[2] R. W. Chang, "Synthesis of band-limited orthogonal signals for multichannel data," *BSTJ.*, pp. 1775-1797, Dec. 1996.

[3] B. R. Saltzburg, "Performance of an efficient parallel data transmission systems," *IEEE Trans. on Comm. Tech.*, pp. 805-811, Dec. 1967.

[4] S. B. Weinstein and P. M. Ebet, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Trans. on Commun.*, COM-19(5), pp. 628-634, Oct. 1971.

[5] L.J. Cimini, Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. on Communications.*, vol. 33, pp. 665-675, July 1985.

[6] A. Peled and A. Ruiz, "Frequency domain data transmission usng reduced computational complexity algorithms," In *Proc. IEEE ICASSP*, pp. 964-967, Denver, CO, 1980.

[7] A. Vahlin and N. Holte, "Optimal finite duration pulses for OFDM," *IEEE Trans. Commun.*, 44(1), pp. 10-14, Jan. 1996.

[8] B. Le Floch, M. Alard and C. Berrou, "Coded orthogonal frequency-division multiplexing," *Proc. IEEE*, 83(6), pp. 982-996, Jun. 1995.

[9] T. Pollet, M. Van Bladel and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *IEEE Trans. on Comm.*, Vol. 43, No. 2/3/4, pp. 191-193, Feb.-Apr., 1995.

[10] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. on Comm.*, Vol. 42, No. 10, pp. 2908-2914, Oct., 1994.

[11] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. on Comm.*, Vol. 45, No. 12, pp. 1613-1621, Dec., 1997.

[12] Van Nee and R. D. J., "OFDM codes for peak-to-average power reduction and error correction," *IEEE Global Telecommunications Conference*, London, pp. 740-744, Nov., 1996.

[13] J. A. Davis and J. Jedwab, "Peak-to-average power control in OFDM, Golay complementary sequences and Reed-Muller codes," HP Laboratories Technical Report, HPL-97-158, Dec., 1997.

[14] A. Tarighat and A. H. Sayed, "MIMO OFDM receivers for systems with IQ imbalances," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3583-3596, Sep. 2005.

[15] A. Tarighat, R. Bagheri, and A. H. Sayed, "Compensation schemes and performance analysis of IQ imbalances in OFDM receivers," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3257-3268, Aug. 2005.

[16] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Communication*, vol. 16, pp. 1451-1458, Oct., 1998.

[17] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, pp. 41-59, Autumn, 1996.

[18] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456-1467, July 1999.

[19] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744-765, March 1998.

[20] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139-157, Jan. 1999.

[21] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading evvironments when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311-335, Mar. 1998.

[22] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. Comun.*, vol. 10, no. 6, pp. 585-595, Nov.-Dec. 1999.

[23] A. Wittneben, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," *Proc. ICC*, pp. 1630-1634, 1993.

[24] Jan Mietzner and Peter A. Hoeher, "Boosting the performance of wireless communication systems: theory and practice of multiple-antenna techniques," *IEEE Communicatin Magazine*, no. 10, pp. 40-47, Oct. 2004.

[25] T. M. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading ," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139-157, 1999.

[26] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel ," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359-383, Feb. 2002.

[27] I. Barhumi, G. Leus and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Processing*, vol. 51, No. 6, pp. 1615-1624, Jun. 2003.

[28] Allert van Zelst and Tim C. W. Schenk, "Implementation of a MIMO OFDM-based Wireless LAN system," *IEEE Trans. Signal Processing*, vol. 52, No. 2, pp. 483-494, Feb. 2004.

[29] X. Li, H. Huang G. J. Foschini and R. A. Valenzuela, "Effects of iterative detection and decoding on the performance of BLAST," *IEEE Proc. Global Telecommun. Conf.*, vol. 2, No. 2, pp. 1061-1066, 2000.

[30] A. Salvekar, S. Sandhu, Q. Li, M. Vuong and X. Qian, "Multiple-Antenna Technology in WiMax Systems," *Intel Technology Journal*, vol. 8, No. 3, [online]: http://www.intel.com/technology/itj/2004/volume08issue03, Aug. 2004.

[31] Hongwei Yang, "A road to future broadband wireless access: MIMO-OFDM-Based air interface," *IEEE Communications Magazine*, Vol. 43, No. 1, pp. 53 - 60, Jan. 2005.

[32] H. Bölcskei, M. Borgmann and A. J. Paulraj, "Impact of the propagation environments on the performance of space-frequency coded MIMO-OFDM," *IEEE J. Select. Areas Commun.*, vol. 21, No. 3, pp. 427-439, Apr. 2003.

[33] H. Bölcskei, and A. J. Paulraj, "Space-frequency coded broadband OFDM systems," *Proc. IEEE WCNC*, pp. 1-6, Chicago, IL, Sep. 2000.

[34] X. Ma, H. Kobayashi and S. C. Schwartz, "Joint frequency offset and chanel estimation for OFDM," *Proc. of Global Telecommun. Conf.*, pp. 15-19, Dec. 2003.

[35] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. on Commun.*, vol. 51, No. 11, pp. 1910-1917, Nov. 2003.

[36] Nima Khajehnouri and Ali H. Sayed, "Adaptive angle of arrival estimation for multiuser wireless location systems," *Fifth IEEE Workshop on Signal Processing Advances in Wireless Communications*, Lisboa, Portugal, July 11-14, 2004.

[37] Part 11: Wireless LAN Medium Access Control (MAC) and Pyhsical Layer (PHY) Specifications—Amendment 1: High-speed Phyisical Layer in the 5 GHz Band, IEEE Standard 802.11a-1999.

[38] M. Brookers, "Matrix Reference Manual [online]", available: http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/.

[39] Part 11: Wireless LAN Medium Access Control (MAC) and Pyhsical Layer (PHY) Specifications—Amendment 1: High-speed Phyisical Layer in the 5 GHz Band, IEEE Standard 802.11a-1999.

[40] Part 16: Air Interface for Fixed Broadband Wireless Access Systems—Amendment 2: Medium Access Control Modifications and Additional Pyhsical Layer Specifications for 2-11 Ghz, IEEE Standard 802.16a-2003.

[41] Digital broadcasting systems for television, sound and data services. European Telcommunications Standard, prETS 300 744 (Draft, version 0.0.3), Apr. 1996.

[42] H. Sampath, S. Talwar, J. Tellado, V. Erceg and A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: design, performance and field trial results," *IEEE Communications Magazine*, No. 9, pp. 143-149, Sep., 2002.

[43] Justin Chuang and Nelson Sollenberger, "Beyond 3G: Wideband wireless data access based on OFDM and dynamic packet assignment," *IEEE Communications Magazine*, No. 7, pp. 78-87, Jul., 2000.

[44] Z. Liu, G. Giannakis, S. Barbarosa, and A. Scaglione, "Transmit-antennae space-time block coding for generalized OFDM in the presence of unknown multipath," *IEEE J. Select. Areas Communication*, vol. 19, no. 7, pp. 1352-1364, Jul. 2001.

[45] S. Yatawatta and A. P. Petropulu, "Blind channel estimation in MIMO OFDM systems," *IEEE Trans. Signal Processing*, submitted, http://www.ece.drexel.edu/CSPL/publications/ssp03sa
-rod.pdf

[46] H. Bölcskei, R. W. Heath Jr. and A. Paulraj, "Blind channel identification and equalization in OFDM-based multiantenna systems," *IEEE Trans. Signal Processing*, vol. 50, No. 1, pp. 96-109, Jan. 2002.

[47] Y. Li, N. Seshadri and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Select. Areas Communication*, vol. 17, pp. 461-471, March 1999.

[48] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Communications*, vol. 1, No. 1, pp. 67-75, Jan. 2002.

[49] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Cosumer Electronics*, vol. 44, No. 3, pp. 1122-1128, August 1998.

[50] G. L. Stüber, J. R. Barry, S. W. Mclaughlin, Y. Li, M. A. Ingram and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proceedings of the IEEE*, vol. 92, No. 2, pp. 271-294, Feb. 2004.

[51] W. C. Jakes, *Microwave Mobile Communications*, John Wiley and Sons, New York, 1974.

[52] R. O. Schmidt, "Multiple emitter location and signal parameter estimation", in *Proc. RADC, Spectral Estimation Workshop*, Rome, NY, pp. 243-258.

[53] A. H. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24-40, July 2005.

[54] K. C. Ho and Wenwei Xu, "An accurate algebraic solution for moving source location using TDOA and FDOA measurements," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2453-2463, Sep. 2004.

[55] "Wireless location technologies and service [online]," available: http://www.3gamericas.org/English/

[56] PELORUS Group. Report on wireless location-based markets. Technical Report, 2001

[57] In-Stat/MDR. Location-based services: Finding their place in the market . Technical Report, Feb. 2003

[58] A. H. Sayed and N. R. Yousef, Wireless location. *Wiley Encyclopedia of Telecommunications,* J. Proakis, editor, John Wiley & Sons, NY, 2003

[59] FCC Docket No. 94-102. Revision of the commissions rules to issue compatability with enhanced 911 emergency calling systems. Technical Report RM-8143, July 1996.

[60] State of New Jersey. Report on the New Jersey wireless enhanced 911 terms: The first 100 days. Technical Report, Jun. 1997

[61] M. Yunos, J. Zeyu Gao and S. Shim, Wireless advertising's challenges and opportunities. *IEEE Computer Magazine,* vol. 36, No. 5, pp. 30-37, May, 2003

[62] Telecommunications Industry Association. The CDMA2000 ITU-R RTT Candidate Submission V0.18, Jul. 1998.

[63] J. J. Caffery and G. L. Stuber, "Overview of radiolocation in CDMA cellular systems," *IEEE Communications Magazine*, vol. 36, No. 4, pp. 38-45, Apr. 98.

[64] H. Krim and M. Viberg, "Two decades of array signal processing research: Te parametric approach," *IEEE Signal Processing Magazine*, vol. 13, No. 4, pp. 67-94, Jul. 1996.

[65] T. Ojanpera and R. Rrasad, *Wideband CDMA for third generation mobile communications*. Arech House, Boston, MA 1998.

[66] R. Rrasad, W. Mohr and W. Konhauser, *Third generation mobile communications*. Arech House, Boston, MA 2000.

[67] P. Bahl and V. N. Padmanabhan, "Radar: an in-building RF-based user location and tracking system," *Proc. IEEE Conference INFOCOMM*, Vol. 2, pp. 775-784, Tel Aviv, March 2000.

[68] T. Ross, P. Myllymaki and H. Tirri, "A statistical modeling approach to location estimation," *IEEE Trans. On Mobile Computing*, Vol. 1, No. 1, pp. 59-69, Jan. 2002.

[69] M. Youssef, A. Agrawala and A. U. Shankar, "WLAN location determination via clustering and probability distributions," *Proc. IEEE Conference PerCom*, pp. 143-150, March 2003.

[70] G. H. Golub and C. F. Van Loan, "Matrix Computations", 2nd Edition, Baltimore: The Johns Hopkins University Press, 1989.

[71] John G. Proakis, "Digital Communications", 4th Edition, Prentice Hall, New Jersey, 2000

[72] Jerry M. Mendel, "Lessons in estimation theory for signal processing, communications and control," 2nd Edition, Prentice Hall PTR, Englewood Cliffs, New Jersey, March 1995.

[73] Athanasios Papoulis and S. Unnikrishna Pillai, "Probability , Random Variables and Stochastic Processes," 4h Edition, McGraw-Hill, Dec. 2001.

[74] P. Stoica, and R. Moses, "Introduction to Spectral Analysis." Upper Saddle River, NJ: Prentice Hall, 1997.

Vita

Zhongshan Wu was born in Anhui, China, on December 4, 1974. He received his bachelor of science degree in electrical engineering from Northeastern University in July 1996. In spring 2000, he entered the graduate program in the Department of Electrical and Computer Engineering at Louisiana State University. He got his master of science degree in electrical engineering in December 2001. Now he is a candidate for the degree of doctor of philosophy in electrical engineering.