Timing and Synchronization of Low Data Rate Ultra-wideband Systems using Data-aided Auto-correlation Method

by

Rongrong Zhang

B. Eng., Shanghai Jiao Tong University, Shanghai, China, 2004

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Applied Science

in the Department of Electrical and Computer Engineering

Timing and Synchronization of Low Data Rate Ultra-wideband Systems using Data-aided Auto-correlation Method

by

Rongrong Zhang

B. Eng., Shanghai Jiao Tong University, Shanghai, China, 2004

Supervisory Committee

Dr. Xiaodai Dong, Supervisor (Department of Electrical and Computer Engineering)

Dr. Aaron Gulliver, Member (Department of Electrical and Computer Engineering)

Dr. Jianping Pan, Outside Member (Department of Computer Science)

Dr. Kui Wu, External Examiner (Department of Computer Science)

Supervisory Committee

Dr. Xiaodai Dong, Supervisor (Department of Electrical and Computer Engineering)

Dr. Aaron Gulliver, Member (Department of Electrical and Computer Engineering)

Dr. Jianping Pan, Outside Member (Department of Computer Science)

Dr. Kui Wu, External Examiner (Department of Computer Science)

Abstract

For low data rate ultra-wideband (UWB) communication systems employing non-coherent detection and autocorrelation detection schemes, timing of integration region significantly affects their error rate performance. Time-of-arrival (TOA) estimation of the first channel tap is also the foundation of the UWB based ranging applications. In this thesis, a data-aided, autocorrelation based timing and synchronization method is developed. First, estimation of the optimal integration region, i.e., the initial point and the length of the integration, using the new timing method is presented. It is shown that the proposed method enhances the error rate performance compared to non-optimal integration region-determining methods. After that, TOA estimation using the proposed timing method is studied for the dual pulse (DP) signal structure. The performance improvement of this approach over the conventional energy detection based method is demonstrated via simulation.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ADC	analog-to-digital converter
AWGN	additive white Gaussian noise
BER	bit error rate
BPPM	binary pulse position modulation
CDMA	code division multiple access
CIR	channel impulse response
CM	channel model
DoD	Department of Defense
DP	dual pulse
DS	direct sequence
ED	energy detection
FCC	Federal Communications Committee
GML	generalized maximum likelihood
GPS	global positioning system
IEEE	Institute of Electrical and Electronics Engineers

IEEE-SA	IEEE Standards Association
IFI	inter-frame-interference
IPI	inter-pulse-interference
IR	impulse radio
ISI	inter-symbol-interference
MAE	mean absolute error
MB-OFDM	multiband orthogonal frequency division multiplexing
MED	maximum energy detection
(N)LOS	(non)-line-of-sight
NR	noise region
PAPR	peak-to-average power ratio
PAR	project authorization request
PDP	power delay profile
PN	pseudorandom noise
PPAM	pulse position amplitude
PSD	power spectral density
RF	radio frequency
RRC	root raised cosine
SNR	signal-to-noise ratio
SR	signal region
TC	threshold crossing

TG	task group
TH-PPM	time-hopping pulse position modulation modulation
TOA	time of arrival
TR	transmitted reference
UWB	ultra-wideband
WPAN	wireless personal area network

Acknowledgements

First and foremost I would like to thank my supervisor, Dr. Xiaodai Dong, for her valuable guidance, continuous encouragement and insightful technical advice throughout my study. This thesis work could not have been completed without the support and help from Dr. Dong.

I would also like to thank Dr. Aaron Gulliver and Dr. Jianping Pan for the valuable suggestions on revising my thesis.

Thanks to many of my colleagues and friends at University of Victoria for being so nice and helpful, which makes my stay in this foreign country a great pleasure. Especially, I would like to thank Dr. Yue Wang, Dr. Wei Li, Li, Fengdan, Massoud, Shiva, Ruonan and Omar for their priceless help.

Special thanks to Steve, Erik, Duncan, Vicky, Sarah, Moneca, Lynne and Mary-Anne for the many patient and constant help from them.

Last but not least, I would like to thank my wife and parents for being so supportive all through these years. It is hard to put into words how much I appreciate their love and how grateful I feel to have them with me.

Dedication

To my dear wife, Amy Shen

Chapter 1

Introduction

Ultra-wideband (UWB) technology is a promising candidate for low power, low complexity, short distance communications. Ever since the Federal Communications Committee (FCC) granted the unlicensed 3.1-10.1 GHz spectrum for UWB systems [2], a large number of UWB techniques, algorithms and product prototypes have been developed each year. In August 2007, IEEE standard board approved UWB as an alternative physical layer technology for low-rate wireless personal area network (WPAN) applications (IEEE 802.15.4a [3]). It is reasonable to expect that the debut of UWB based personal communication products on the wireless market is not far away.

In this chapter, a brief overview of ultra-wideband technology is presented from a historical perspective, with focus on the recently standardized low data-rate UWB techniques. Strengths and challenges of low data-rate UWB applications are introduced, which lay as the foundation and start point of the research that forms this thesis.

1.1 Brief Overview of Ultra-wideband Communication

Ultra-wideband communication is not a new technology at all. When Guglielmo Marconi started his pioneer works of wireless telegraph transmission, the information was conveyed on a series of electrical sparks, which is nothing but carrierless, impulse based, ultra-wideband communication. During 1960s and 1970s, UWB has been developed for applications such as radar systems and ground penetrating geological survey systems [4]. Ross, Robbins and other researchers made some of the earliest contributions to the UWB communication systems [4]. However, the terminology "Ultra-Wideband" did not come into being until 1989 when the U.S. Department of Defense (DoD) decided to use this term for devices occupying a bandwidth no less than 1.5 GHz or a -20 dB fractional bandwidth exceeding 25% [1].

In April 2002, FCC released its first report and order *(R&O)* about UWB communications, in which it granted the bandwidth below 960 MHz and from 3.1 GHz to 10.6 GHz for UWB use [2]. The FCC spectral mask for indoor UWB systems is shown in Fig. 1. Specifically, the FCC spectrum mask requires the power spectral density (PSD) of indoor UWB transmitted signal to be smaller than -41.3 dBm/MHz, which is the Part 15 limit for spurious radio frequency (RF) emission, for the frequency bands below 960 MHz and between 3.1 GHz – 10.6 GHz. This means the UWB signal is below the noise floor of other systems that share the spectrum with it. Moreover, the UWB signal shall be below -75 dBm/MHz between 0.96 GHz to 1.61 GHz, which gives way to near-noise-floor communications such as global positioning system (GPS).

The first *R&O* also redefines UWB as a system with a -10 dB bandwidth no less

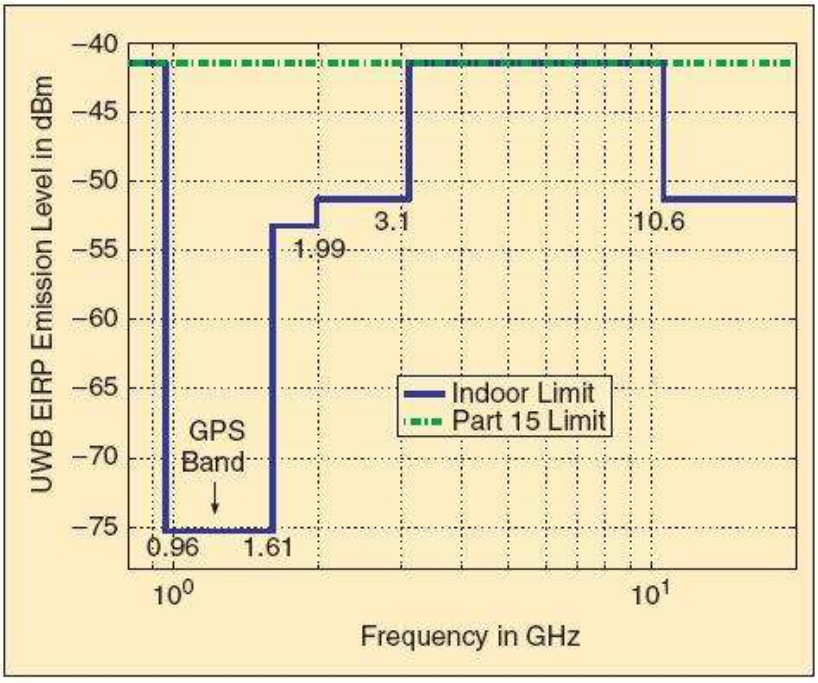

Figure 1.1: FCC spectral mask for indoor commercial UWB systems [1].

than 500 MHz or a fractional bandwidth exceeding 20%. The fractional bandwidth is defined as B/f_c, where $B = f_H - f_L$ is the -10 dB bandwidth with f_H being the upper frequency of the -10 dB emission point and f_L being the lower frequency of the -10 dB emission point, and $f_c \triangleq (f_H + f_L)/2$.

According to Shannon's theorem on channel capacity, the maximum information rate of a band-limited additive white Gaussian noise (AWGN) channel follows [5]

$$C = W \log(1 + \frac{P_{av}}{W N_0}) \tag{1.1}$$

where W is the signal bandwidth, P_{av} is the average power of the transmitted signal, and N_0 is the noise variance. Eq. (1.1) shows that given a certain amount of transmit power and noise variance, a larger bandwidth yields a greater channel capacity. For

example, supposing the signal to noise ratio P_{av}/WN_0 is as low as 0 dB, if the whole 7.5 GHz regulated bandwidth is efficiently utilized, the data rate could be 7.5 Gbps. Therefore, even if the PSD of UWB signal is limited to a sub-noise-floor level, at least in theory the UWB systems are able to achieve much higher data-rate than the currently used narrowband and wideband systems. From another perspective, given the ultra-wide bandwidth, a much lower transmitted power is sufficient for communication at a comparable speed to existing wireless systems. High data-rate, low power, along with other features such as high time resolution, ability of penetrating objects, great security ability, etc., provide strong motivations for people to pursue advances in the UWB technology.

Two task groups (TGs) were formed by IEEE to study the possibility of adopting UWB as an alternative physical layer option for wireless personal area network (WPAN) (range from 1 m to 10 m). TG 802.15.3a is aiming at providing a higher speed UWB physical layer enhancement amendment to IEEE 802.15.3 for applications which involves imaging and multimedia transmission; TG 802.15.4a is targeting at communications and high precision ranging/localization systems (1 meter accuracy and better) with high aggregate throughput, ultra low power and low cost.

Two technical proposals have been filed to TG3a. One is a single carrier UWB using direct sequence (DS) spread spectrum technology introduced by XtremeSpectrum [6], and the other is a multiband orthogonal frequency division multiplexing (MB-OFDM) based technology suppported by the Multiband OFDM Alliance (MBOA) [7]. DS-UWB transmits ultra-short pulses at baseband using the whole available spectrum, thus frees the requirements on (de)modulator. But otherwise it is similar to the conventional CDMA technology, and suffers from complicate design

of rake receivers to collect enough multipath diversity for reasonable system performance. On the other hand, MB-OFDM transmits signals using a 128-subcarrier OFDM scheme occupying a 500 MHz band. The signal band is frequency hopped in the available spectrum on a block-by-block basis. Although OFDM systems have simple receiver architecture, it also has drawbacks, such as OFDM's inherent high peak-to-average power ratio (PAPR) [8].

Since neither of the technologies has dominant advantage over the other, the standardization went into a deadlock. In January 2006, TG3a decided to withdraw the December 2002 project authorization request (PAR) that initiated the development of high data rate UWB standards and leave the decision to the market. If there is a surviving approach in a year or two and the technology has proven itself to be commercially viable, then IEEE will come back and revisit whether it makes sense to create an IEEE standard for it.

Similar to high data rate UWB, a number of techniques have been developed as candidates for low data rate UWB applications. These include time-hopping pulse position modulation (TH-PPM) [9], pulse position amplitude modulation (PPAM) [10], transmitted reference (TR) [11], dual pulse (DP) [12], etc. But unlike TG3a, TG4a was able to reach an accordance on a draft standard, which was officially approved by IEEE-SA Standard Board on March 22nd, 2007 as an amendment to IEEE Std 802.15.4-2006 [3]. The standard specifies a clustered pulse train signal pattern, a regulation on the transimtted pulse shape and operating frequency bands, but leaves large freedom in signal structure and receiver design. According to the standard, each transmitted information symbol may contain either a single or multiple of clustered pulse trains. The number of pulses per clustered pulse train varies from 1

to 128. This clustered pulse train pattern is compatible with all of the aforementioned low data rate UWB signal schemes. This provides the opportunity and possibility for researchers to develop novel receiver algorithms for various applications.

1.2 Strengths and Possible Applications of Low Data Rate UWB

Typically a low data rate UWB signal consists of a series of sub-nanosecond pulses. The bandwidth of the signal is roughly the reciprocal of the pulse duration, which ranges from 500 MHz up to a few GHz. Besides the benefit of low transmit power, as mentioned in the previous section, the UWB signal enjoys the following strengths as well.

1. High multipath diversity

   Because the UWB pulse has ultra-short duration, the possibility that pulses from two different multipaths overlap and cancel each other is much less than in narrowband systems. Intuitively, rich multipath diversity enhances the reliability of the wireless link since it increases the possibility that at least some of the multipath signal can go around obstacles. In other words, the multipath fading is less significant in a UWB system. A discrete equivalent UWB channel is usually comprised of hundreds of densely located multipath taps. A large amount of multipath diversity can be obtained if the signal energy on most of the multipath taps is collected by the receiver.

2. High timing resolution

   Timing resolution is inversely proportional to the bandwidth of the signal [13].

Because of the giga-Hertz bandwidth of the UWB signal, the timing resolution is in the order of nano-seconds, which leads to a centimeter-level ranging capability.

3. High security

   The sub-noise floor transmitted power of the UWB signal, along with the great variety of signal pattern possibilities, makes UWB signal virtually impossible to be detected by a third party. UWB may be the most secure means of wireless transmission technology available.

4. Co-existence with narrowband and wideband systems

   Since the power spectral density of UWB signal is below the FCC Part 15 limit, it introduces only negligible interference into the existing narrowband and wideband communication systems. This allows UWB systems to be adopted in various indoor applications without worrying about affecting co-existing systems.

Attributed by these merits, UWB is well suited to a lot of applications. First of all, UWB could be the communication backbone of the future "digital home" [14]. UWB wireless link can be used for sharing audio, video and data files between TV, computers, printers and consumer electronics. Connecting to a sensor network, it can also be used to wirelessly control the house heating and lighting systems to achieve best energy efficiency. Besides, UWB can be used in medical imaging systems, where small devices could be injected into human bodies to actually "see" the illness area and transmit the image back via a low power UWB connection.

Moreover, with its accurate ranging capability, UWB can be used to enhance location based services [15]. UWB localization devices can be attached to infants at

home, inmates at prison or packages in a logistic warehouse for better surveillance; accurate localization systems can also help fire fighters and earthquake or mine accident rescuers to get better knowledge about the location of themselves.

However, the ultra-wide bandwidth and low power also impose quite a few challenges on UWB system implementation.

- With current analog-to-digital converters (ADCs), it is impractical to sample the UWB signal at Nyquist rate which may be several GHz. Analog correlation and frame rate sampling are suggested in most of the proposed system designs, e.g., [16–19].

- The great number of multipath taps brings difficulties on channel estimation and full-rank rake receiver implementation [1,20]. Alternatives such as selective rake receivers with partial channel estimation or transmitted reference autocorrelation receivers are used in the literature (see [21] and references therein).

- Due to the wide bandwidth distinct frequency components may experience different channel environments, causing distortion on the UWB pulse shape [22], which adds extra difficulty in coherent detection.

- Although UWB signals have fine timing resolution, achieving this resolution is not easy [15]. On one hand, the nanosecond pulse duration requires the UWB receiver to have nanosecond clock accuracy. Even the slightest timing jitter compromises the accuracy of the ranging result. On the other hand, ultra-short pulse duration and relatively long multipath delay spread extend the uncertainty region of timing and synchronization, which increases the possible delay positions to be searched.

1.3 Thesis Outline

The research that forms this thesis mainly tackles the last challenge in the previous section. A data-aided timing and synchronization method is proposed for UWB systems. The proposed method bypasses the first three challenges mentioned in the last section. That is, the proposed method is based on non-coherent and autocorrelation receivers, thus requires neither channel estimation nor rake receiver. By using a training sequence, the proposed method is able to work at frame rate sampling. Furthermore, the proposed method is also applied on a dual pulse (DP) system to achieve a more accurate time-of-arrival (TOA) estimation which can be directly used for localization.

Because the proposed timing and synchronization algorithm focuses on low data rate UWB systems, from the next chapter on the term UWB denotes low data rate UWB. High data rate UWB systems are generally not in discussion, unless otherwise specifically mentioned.

The rest of the thesis is organized as follows.

Chapter 2 formulates the UWB system in discussion. Particularly, the UWB systems with pulse position modulation, transmitted reference systems and dual pulse systems are discussed. Different receiver schemes, such as coherent receiver, non-coherent receiver with square-law detection and autocorrelation receiver are described. On top of that, a brief introduction to the UWB channel model is also included in this chapter.

In Chapter 3, a method of timing and synchronization is presented. The method is able to estimate the optimal integration region, i.e., the initial point and the

length of the integration, for the non-coherent and autocorrelation receivers. Following a theoretical BER analysis, a data-aided estimation method using the idea of inter-symbol correlation is proposed. It is shown that using noise corrupted received signals, the proposed method is not only practically applicable, but also enhances the performance compared to non-optimal timing methods.

In Chapter 4, the timing method is modified for TOA estimation using the DP signal structure. Inter-symbol signal autocorrelation and threshold crossing are used to detect the direct path in a line-of-sight (LOS) channel. The effects of different threshold values and training sequence lengths on the estimation accuracy of the proposed method are studied. The performance improvement of this approach over the conventional energy detection based method is also demonstrated via simulation.

Finally, Chapter 5 concludes the thesis and suggests future research topics.

Chapter 2

UWB Communication System

2.1 UWB Impulse Radio

One prevailing way to realize ultra-wideband communication is through impulse radio (IR), which is characterized by the usage of unmodulated, nanosecond-width pulses with a very low duty cycle. The ultra-narrow pulse duration makes it possible to spread the energy over a large bandwidth, which may range from near DC to a few gigahertz. To avoid inter-symbol-interference (ISI) and inter-frame-interference (IFI), adjacent pulses or pulse clusters are separated further than the longest channel delay spread, making the transmitter to work at only a small fraction of time, and remain silent at most time. This allows the IR transmitter to operate with low power consumption.

The shape of the pulse may vary from system to system. In some cases even multiple types of pulses may co-exist in one system [23]. As per IEEE 802.15.4a standard [3], a pulse $p(t)$ is compliant to the standard if its normalized cross-correlation with the reference pulse $r(t)$ has a main lobe greater than 0.8 for at least a duration of T_w, and no side lobes above 0.3, regardless of the actual shape of the pulse. The

correlation is defined as $\frac{1}{\sqrt{E_r E_p}}\Re\Big\{\int_{-\infty}^{\infty} r(t)p^*(t+\tau)\mathrm{d}t\Big\}$ where E_r and E_p are the energies of $r(t)$ and $p(t)$, respectively. The reference pulse is a root raised cosine (RRC) pulse defined as

$$r(t) = \frac{4\beta}{\pi\sqrt{T_p}} \frac{\cos\left[\frac{(1+\beta)\pi t}{T_p} + \frac{\sin[(1-\beta)\pi t/T_p]}{4\beta t/T_p}\right]}{(4\beta t/T_p)^2 - 1} \tag{2.1}$$

where $\beta = 0.6$ is the roll-off factor and T_p is the width of the reference pulse. The minimum main lobe width T_w is 0.5 ns for pulses of 500 MHz bandwidth and 0.2 ns for pulses with larger bandwidth.

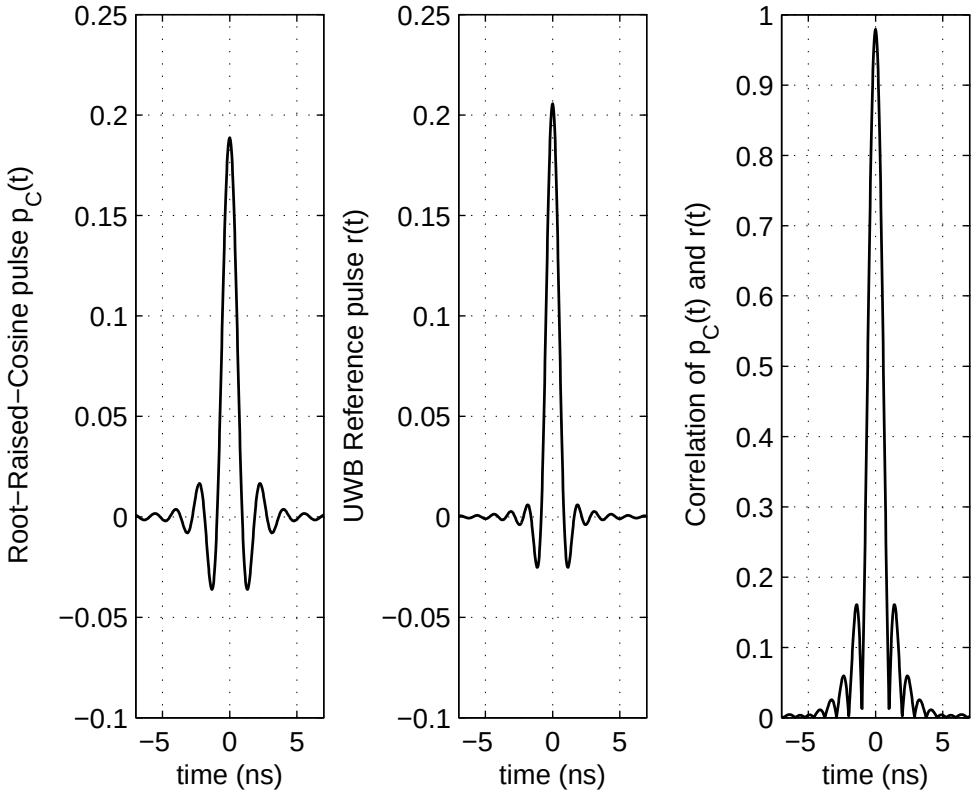

Figure 2.1: The RRC pulse used in this thesis is compliant with the standard.

In this thesis, an RRC pulse $p_C(t)$ with 500 MHz bandwidth is used as transmitted

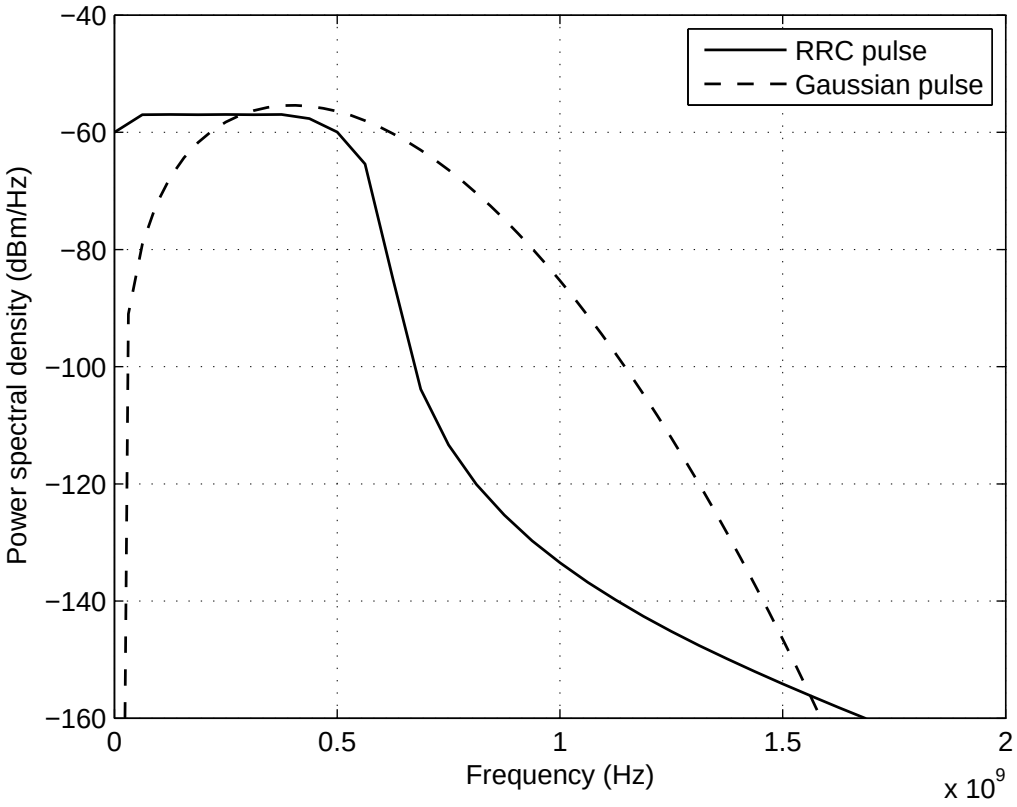

Figure 2.2: The spectrums of the RRC pulse and second order Gaussian pulse.

pulse in most of the simulations. It follows the same mathematical form as (2.1), but the roll-off factor of $p_C(t)$ is 0.25. Fig. 2.1 shows the shape of $p_C(t)$, $r(t)$ and their cross-correlation. It is shown that this pulse complies with the 802.15.4a standard as the main lobe of the normalized cross-correlation is greater than 0.8 for about 0.65 ns and no side lobe is greater than 0.3.

Besides the RRC pulse, second order derivative of Gaussian pulse is used in the simulation in Chapter 4 to make a performance comparison between different pulse shapes. The second order Gaussian pulse is widely used in UWB literatures [1, 17,

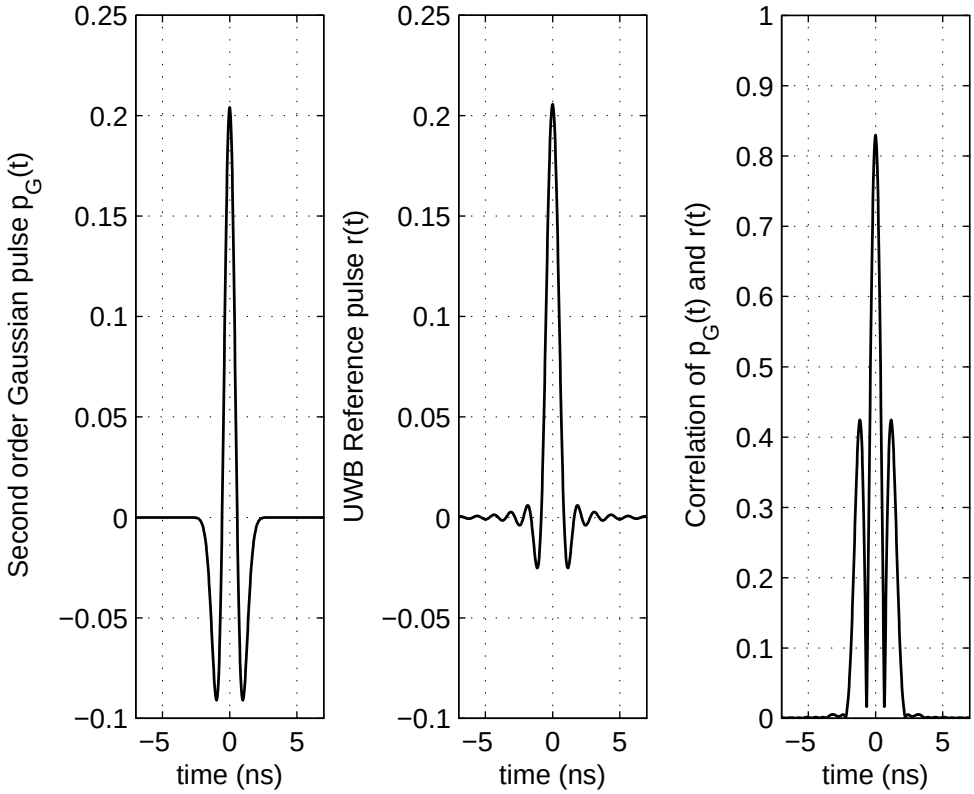

Figure 2.3: The second order Gaussian pulse used for comparison.

24, 25]. Its mathematical representation is

$$p_G(t) = \left[1 - 4\pi(t/T_p)^2\right] e^{-2\pi(t/T_p)^2} \tag{2.2}$$

where T_p also denotes the pulse duration. With the same pulse duration, the second order Gaussian pulse has very similar -10 dB bandwidth as the RRC pulse, as shown in Fig. 2.2. The pulse shape of $p_G(t)$ and its correlation with $r(t)$ is shown in Fig. 2.3. Unfortunately, the second order Gaussian pulse does not comply with the IEEE standard because the side lobe of the cross-correlation is higher than 0.3, and the

main lobe is above 0.8 for only around 0.2 ns.

In impulse radio UWB, data modulation is usually on the position or the polarization of the pulse, rather than on its magnitude or phase. This is because in a rich multipath environment it is more difficult to make a correct detection on the magnitude or phase of the received signal. Typically, one data symbol is comprised of several frames. Each frame may contain a single pulse, a pair of pulses, or a cluster of pulses. The information bit is modulated onto all these frames identically, which may be seen as a kind of repetition coding, which makes the i^{th} transmitted symbol be

$$s^{(i)}(t) = \sqrt{\frac{E_b}{N_f}} \sum_{n=0}^{N_f-1} s_f^{(i)} \left[t - nT_f - T_c^{(i)}(n) \right] \tag{2.3}$$

where E_b is the transmitted energy per bit, N_f is the number of frames per symbol, $s_f(t)$ denotes the signal within a frame where various modulation schemes can be applied and $T_c^{(i)}(n)$ is a sequence of time delays for the frames in the i^{th} symbol that is used to introduce time hopping to the system. Time hopping not only enables the system to accommodate multiple users, but also scrambles the transmitted signal so as to suppress the spectral spikes caused by the repetition of the signal. However, multiple access and scrambling are out of the scope of this thesis, thus are not included in the system model for timing and synchronization. As a result, $T_c^{(i)}(n)$ is omitted in the following discussion. For systems with the proposed timing methods, time hopping can be adopted in the data transmission period once timing and synchronization are done.

2.2 UWB Channel Model

As mentioned in Chapter 1, due to its ultra short pulse width, the UWB signal is able to resolve more multipath taps than conventional narrow band signals. This determines that the channel models used for narrow-band systems are no longer applicable for UWB studies. Based on various measurement results and modeling recommendations filed to the channel modeling subgroup of IEEE 802.15.4a, the task group proposed a generic UWB channel model in November 2004 [26]. It assumes the channel bins arrive in the form of clusters following the S-V model [27] and the channel fading is slow so that the channel stays constant during one block of data burst.

According to the S-V model, the channel impulse response can be represented as

$$h(t) = \sum_{l=1}^{L} \sum_{k=1}^{K} \alpha_{k,l} \exp(j\phi_{k,l}) \delta(t - T_l - \tau_{k,l}) \tag{2.4}$$

where T_l denotes the delay of the l^{th} cluster, $\tau_{k,l}$ denotes the delay of the k^{th} channel tap of the l^{th} cluster relative to T_l, $\alpha_{k,l}$ and $\phi_{k,l}$ denote the magnitude and phase of the k^{th} channel tap in the l^{th} cluster, respectively. The total number of clusters in the channel power delay profile (PDP) L is a random variable following Poisson distribution. That is

$$p(L) = \frac{\bar{L}^L \exp(-\bar{L})}{L!} \tag{2.5}$$

where $\bar{L}$ is the expectation of L. The cluster arrival time T_l is defined as a Poisson process, i.e., the time difference of adjacent clusters follows the exponential distribu-

tion, which can be written as

$$p(T_l|T_{l-1}) = \Lambda \exp[-\Lambda(T_l - T_{l-1})], \quad l > 0, \tag{2.6}$$

where Λ is the cluster arrival rate. On the other hand, the channel bin arrival time $\tau_{k,l}$ is defined as the mixture of two Poisson processes, which is

$$p(\tau_{k,l}|\tau_{(k-l),l}) = \beta\lambda_1 \exp[-\lambda_1(\tau_{k,l} - \tau_{k-1,l})] + (1-\beta)\lambda_2 \exp[-\lambda_2(\tau_{k,l} - \tau_{k-1,l})], \quad k > 0 \tag{2.7}$$

where β is called the mixture probability whose value varies over different channel environments, and λ_1 and λ_2 are the ray arrival rates.

The tap gain $\alpha_{k,l}$ follows Nakagami distribution. Its probability density function can be written as

$$p_{\alpha_{k,l}}(\alpha) = \frac{2}{\Gamma(m)}\Big(\frac{m}{\Omega_{k,l}}\Big)^m \alpha^{2m-1} \exp\Big(-\frac{m}{\Omega_{k,l}}\alpha^2\Big) \tag{2.8}$$

where $m \leq 1/2$ is the Nakagami factor, $\Gamma(m)$ is Gamma function and Ω is the mean square value of α. The Nakagami factor is modeled as a log-normal distributed random variable whose logarithm has mean m_0 and standard deviation $\hat{m}_0$. The mean power of the channel taps $\Omega_{k,l}$ follows exponential distribution in each cluster. That is

$$\Omega_{k,l} = \frac{\Omega_l \exp(-\tau_{k,l}/\gamma_l)}{\gamma_l[(1-\beta)\lambda_1 + \beta\lambda_2 + 1]} \tag{2.9}$$

where γ_l is the intra-cluster decay time constant which is linearly dependant on the cluster arrival time, and Ω_l denotes the total channel tap power within the l^{th} cluster.

In [26], the above mentioned parameters, e.g. $\bar{L}$, Λ, m_0, etc., are specified for nine different scenarios, including indoor residential environments, indoor office environments, outdoor environments, indoor industrial environments and open outdoor environments, and the former four environments are divided into line-of-sight (LOS) and non-line-of-sight (NLOS) cases. Fig. 2.4 shows two examples of UWB channel realizations. The upper plot is the channel impulse response (CIR) of a realization of IEEE 802.15.4a channel model (CM) 1, which is an indoor residential LOS channel. It has relatively short root-mean-square delay and fewer channel taps. The lower plot is the CIR of a CM8 realization, which is an industrial NLOS channel. Its tap delay is much longer than the CM1 case and there is virtually no visible tap clustering. This is because in such an environment dense arrival multipath components are observed, which means that each resolvable channel tap contains significant energy. The channel model is thus virtually degraded to a conventional tapped delay line model with regular tap spacing.

In the analysis of this thesis, the more general tapped delay line model is used for all environments instead of the more complicated model (2.4) just for representational clarity. It can be written as

$$h(t) = \sum_{l=1}^{L} \alpha_l \delta(t - \tau_l) \tag{2.10}$$

where L is the total number of resolvable multipath taps, α_l and τ_l are the complex magnitude and delay of the l^{th} tap, respectively. However, in all simulations and numerical calculations, the channel model (2.4) as specified in [26] is used.

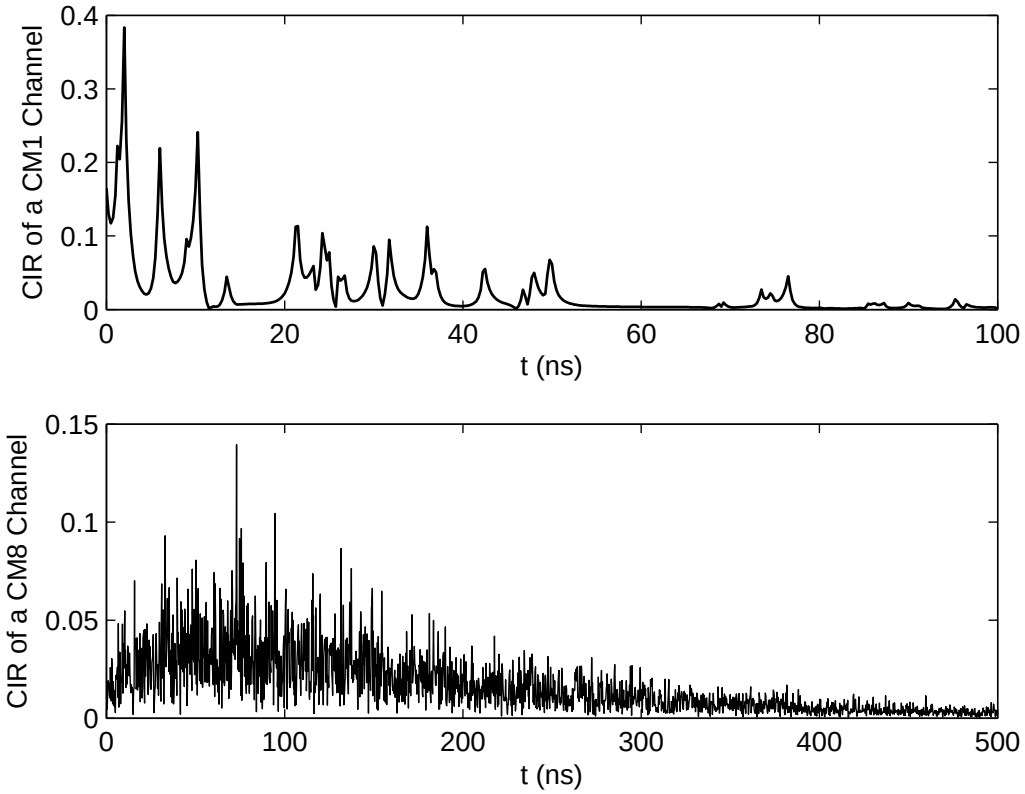

Figure 2.4: CIR of two UWB channel model realizations [26]. The upper plot is a CM1 realization and the lower plot is a CM8 realization.

2.3 UWB Modulation and Detection Schemes

Three modulation schemes are studied in this thesis, including binary pulse position modulation (BPPM), transmitted reference (TR) and dual pulse (DP) schemes.

2.3.1 Binary Pulse Position Modulation

The binary pulse position modulation is an orthogonal modulation. It equally divides one frame duration into two halves. Each frame contains a single pulse which locates in either the first half or the second half of the frame, depending on the data being

"0" or "1". The BPPM signal of the i^{th} symbol can be written as

$$s_{BPPM}^{(i)} = \sqrt{\frac{E_b}{N_f}} \sum_{n=0}^{N_f-1} \Big[(1-d^{(i)})p(t-nT_f) + d^{(i)}p(t-nT_f-T_f/2)\Big] \tag{2.11}$$

where $p(t)$ is the transmitted shaping pulse; $d^{(i)} \in \{0,1\}$ is a binary input data. When the input data d is "0", the transmitted pulse locates at the first half of the frame; when d is "1", the pulse is at the second half of the frame. The frame duration T_f is chosen to satisfy $T_f > 2(\tau_L - \tau_1 + T_p)$ so as to avoid IFI at the receiver, where T_p is the duration of $p(t) * f(t)$ where $f(t)$ is the receiver filter matched to $p(t)$ and $*$ is operator of convolution.

Denote $g(t) \triangleq p(t) * h(t) * f(t)$ as the impulse response of the equivalent channel, the received waveform of the i^{th} symbol after passing through filter $f(t)$ is given by

$$r_{BPPM}^{(i)}(t) = \sqrt{\frac{E_b}{N_f}} \sum_{n=0}^{N_f-1} \Big[(1-d^{(i)})g(t-nT_f) + d^{(i)}g(t-nT_f-T_f/2)\Big] + n(t) \tag{2.12}$$

where the additive band limited complex Gaussian noise $n(t)$ has a variance of N_0. Assuming an ideal low-pass filter (LPF) is used for $f(t)$, the autocorrelation function of $n(t)$ is $R_n(\tau) = 2BN_0\text{sinc}(2B\tau)$, where B is the bandwidth of $n(t)$, or equivalently the bandwidth of $f(t)$. An LPF matched to $p(t)$ can also be used in the receiver, bringing only slight modification on $R_n(\tau)$. Due to the orthogonal nature of BPPM signaling, both coherent and non-coherent detection methods can be applied to demodulate the received signal in (2.12).

For coherent detection, knowledge of the entire channel impulse response is required. In other words, a noiseless template of $g(t)$ should be stored in the receiver.

Since the non-zero support of $g(t)$ is $[\tau_1, \tau_L + T_p]$, the coherent detector performs cross-correlation in the following two time regions

$$y_1^{(i)} = \sum_{n=0}^{N_f-1} \Re\left\{ \int_{\tau_1}^{\tau_L+T_p} r_{BPPM}^{(i)}(t + nT_f) \cdot g^*(t)\mathrm{d}t \right\} \tag{2.13}$$

$$y_2^{(i)} = \sum_{n=0}^{N_f-1} \Re\left\{ \int_{T_f/2+\tau_1}^{T_f/2+\tau_L+T_p} r_{BPPM}^{(i)}(t + nT_f) \cdot g^*(t)\mathrm{d}t \right\} \tag{2.14}$$

where the operator * denotes complex conjugate and $\Re\{x\}$ takes the real part of x. The decision is then made by picking the greater one between y_1 and y_2, which can be written as $y_1^{(i)} \underset{\text{“0”}}{\overset{\text{“1”}}{\lessgtr}} y_2^{(i)}$.

For non-coherent detection, the detector simply calculates the received signal energy in the two possible time slots in one frame, i.e.,

$$y_1^{(i)} = \sum_{n=0}^{N_f-1} \int_{T_0}^{T_0+T} \left| r_{BPPM}^{(i)}(t + nT_f) \right|^2 \mathrm{d}t \tag{2.15}$$

$$y_2^{(i)} = \sum_{n=0}^{N_f-1} \int_{T_f/2+T_0}^{T_f/2+T_0+T} \left| r_{BPPM}^{(i)}(t + nT_f) \right|^2 \mathrm{d}t \tag{2.16}$$

where T_0 and T are the start point and the length of the integration region, respectively. As in the non-coherent detection the channel state information is assumed to be unknown, the integration start point and length need to be estimated by the receiver. Because the choices of T_0 and T determine the integrated signal energy and noise energy, which are the decisive factors of the bit error rate performance of the system, both parameters need to be optimized according to the current channel condition. This is the target of the algorithm described in Chapter 3. Following the

integrations in (2.15)-(2.16), the decision is then made in the same manner as the coherent detection, as $y_1^{(i)} \underset{\text{“0”}}{\overset{\text{“1”}}{\lessgtr}} y_2^{(i)}$.

In practice, the channel information is neither known by the transmitter nor by the receiver beforehand, and channel estimation from the noise corrupted received signal is shown to be overwhelmingly costly. Therefore, although non-coherent detection has poorer performance than the coherent detection, it remains a good compromise between complexity and performance.

2.3.2 Transmitted Reference

Originally, the transmitted reference system was proposed in the 1960s for communications in the situation where the channel is unknown and difficult to estimate [28,29]. Since proposed by Hoctor and Tomlinson [11] as an alternative modulation scheme for UWB, transmitted reference scheme has drawn much attention among UWB researchers (see [17, 25, 30, 31], etc.). The advantage of the TR system is that the detector is simply an autocorrelator. No channel estimation is needed, as is the BPPM non-coherent detection. The main drawback of the TR system is its requirement on a wideband analog delay line for the receiver to realize the long time delay. This is not easy to implement.

In a TR system, each frame contains two pulses. The first one that sits at the beginning of the frame is the reference pulse, and the second one which is delayed from the first pulse by T_d is the data pulse. When the transmitting data is bit “0”, the data pulse is a replica of the reference pulse; when a bit “1” is transmitted, the data pulse is the latter's antipode. The i^{th} symbol of a TR signal can thus be represented

as

$$s_{TR}^{(i)} = \sqrt{\frac{E_b}{2N_f}} \sum_{n=0}^{N_f-1} \Big[p(t - nT_f) + (1 - 2d^{(i)}) p(t - nT_f - T_d) \Big]. \tag{2.17}$$

The frame duration T_f in the TR system should also be at least $2(\tau_L - \tau_1 + T_p)$ so as to avoid IFI. Moreover, the time delay T_d and the time difference between the data pulse and the reference pulse of the next frame should both be at least $(\tau_L - \tau_1 + T_p)$ so as to avoid the inter-pulse-interference (IPI). Therefore, one convenient way to configure the TR signal structure is making T_f exactly twice as T_d. This is the case adopted in the discussion in Chapter 3, and thus T_d will be replaced by $T_f/2$ within the rest of this subsection and the next chapter.

Similar to the BPPM case, the received TR signal after passing the receiver filter $f(t)$ can be written as

$$r_{TR}^{(i)}(t) = \sqrt{\frac{E_b}{2N_f}} \sum_{n=0}^{N_f-1} \Big[g(t - nT_f) + (1 - 2d^{(i)}) g(t - nT_f - T_f/2) \Big] + n(t). \tag{2.18}$$

Autocorrelation detector is used to detect TR signals. It correlates the reference pulse with the data pulse, which is

$$y_{TR}^{(i)} = \Re \Bigg\{ \sum_{n=0}^{N_f-1} \int_{T_0}^{T_0+T} r_{TR}^{(i)}(t + nT_f) r_{TR}^{(i)*}(t + nT_f + T_f/2) \mathrm{d}t \Bigg\}, \tag{2.19}$$

and then the decision is made as $y_{TR}^{(i)} \underset{\text{“0”}}{\overset{\text{“1”}}{\gtrless}} 0$. In (2.19), T_0 and T represent the integration start point and integration length, respectively. Similar to BPPM modulation with non-coherent detection, both parameters are not known a priori and thus need to be estimated and optimized at the receiver.

2.3.3 Dual Pulse Scheme

Realizing the difficulty of implementing the $T_f/2$ long wideband analog delay line in the TR system, Dong, et. al. proposed a dual pulse (DP) system that uses only pulse duration delay [12]. Similar to the TR signal, a dual pulse signal contains a reference pulse immediately followed by a data pulse in each symbol, as shown in Fig. 2.5. The transmitted DP signal of the i^{th} symbol can be represented as

$$s_{DP}^{(i)}(t) = \sqrt{\frac{E_b}{2N_f}} \sum_{n=0}^{N_f-1} \Big[p(t - nT_f) + (1 - 2d^{(i)}) p(t - nT_f - T_p) \Big]. \tag{2.20}$$

Comparing (2.20) with (2.17), the delay between the reference pulse and the data pulse T_d in conventional TR systems is longer than the channel maximum excess delay, but in the DP scheme T_d equals to the pulse duration T_p. Though certain amount of IPI is present in the DP system, it is shown that the system performance is just slightly degraded [12]. It is also found that the DP signal is an appropriate choice to realize the TOA estimation method proposed in Chapter 4.

The received signal at the output of the receiver filter is given by

$$r^{(i)}(t) = \sqrt{\frac{E_b}{2N_f}} \sum_{n=0}^{N_f-1} \Big[g(t - nT_f) + (1 - 2d^{(i)}) g(t - nT_f - T_p) \Big] + n(t). \tag{2.21}$$

Autocorrelation is also used for data detection of the DP signal. The decision variable of the i^{th} symbol is the following correlation

$$y_{DP}^{(i)} = \Re \Bigg\{ \sum_{n=0}^{N_f-1} \int_{T_0}^{T_0+T} r_{DP}^{(i)}(t + nT_f) r_{DP}^{(i)*}(t + nT_f + T_p) \mathrm{d}t \Bigg\}. \tag{2.22}$$

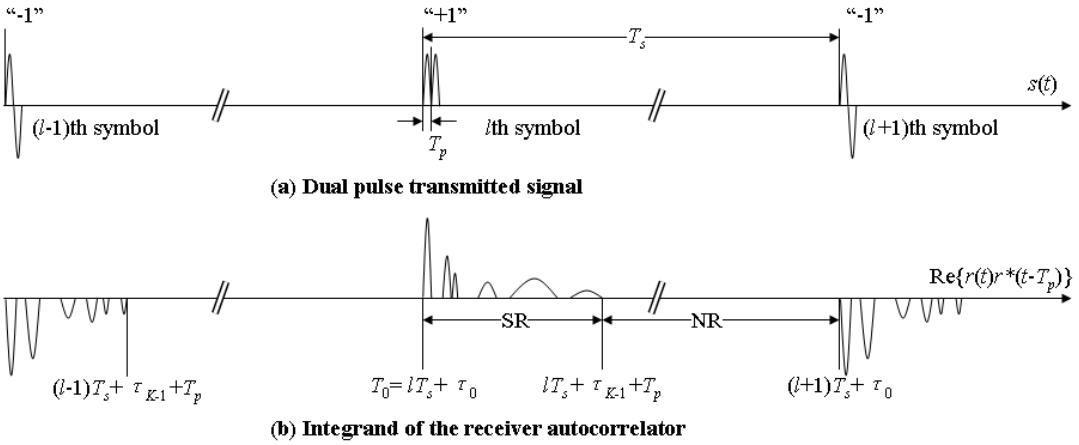

Figure 2.5: Illustration of (a) the transmitted signal and (b) $\Re\{r(t)r^*(t-T_p)\}$ of the DP system. The labeled time intervals in (b) correspond to the flat regions (noise region (NR) of "-1", NR of "+1") and the monotonically decreasing region (signal region (SR) of "+1") that the noiseless part of $x(\tau)$ in eq. (4.2) experiences.

To further spread out the energy in the DP scheme into multiple DPs for lower peak-to-average power ratio, we can modify the transmitted signal to be a cluster of dual pulses spread by a pseudo random sequence, as given by

$$s^{(i)}(t) = \sqrt{\frac{E_b}{2N_c}} \sum_{n=0}^{N_f-1} \sum_{j=0}^{N_c-1} c_j \Big[p(t-nT_f-2jT_p) + (1-2d^{(i)}) p(t-nT_f-2jT_p-T_p) \Big] \quad (2.23)$$

where c_j is the pseudo random code sequence with length N_c. At the receiver, the received signal is matched to the following pseudo random pulse sequence

$$c(t) = \sum_{j=0}^{N_c-1} c_j p(t-2jT_p) \quad (2.24)$$

and the resultant signal is close to the single DP case for use in the autocorrelation based TOA estimation. In the subsequent discussion we will not distinguish the multi-DP scheme and the single DP scheme, while in the simulation single DP signal

is used.

Chapter 3

Integration Region Optimization for Non-Coherent Detection and Autocorrelation Detection UWB Systems

3.1 Background

As mentioned in the previous chapters, coherent detection of UWB signals requires formidably complex channel estimation. Hence suboptimal receivers that do not require channel estimation were proposed for low complexity and low data rate applications, using either non-coherent detection of pulse position modulation signals (e.g., [32, 33]) or autocorrelation detection of transmitted reference signals (e.g., [11, 17]). These two schemes simply perform "integration-and-dump" operation at the receiver, which requires no channel estimation and only frame or symbol rate sampling. They yield reasonable performance with a sufficiently low complexity. For these schemes, the position and length of the integration region greatly affect their performance, because the start and end points of the integration determine how much signal energy

and noise energy will be captured, which in turn determines the bit error rate (BER). This issue was studied in some literature, e.g., [30, 31, 34]. In [30] and [31], the integration region was divided into a large number of smaller intervals, and a weighted summation was performed on the integration results so that the signal-to-noise ratio (SNR) at the detector is maximized. Both [30] and [31] focused on the methods of finding the optimal combining weights, while assuming the synchronization was done beforehand and the integration intervals were fixed. It is a different approach from the one in this paper, which does a synchronization first and then determines the position of the integration interval. In [34], synchronization for noncoherent schemes was performed iteratively until the integration interval is approximately the same as the signal region (SR), i.e., the time interval during which the transmitted pulses and their multipath components are received. This led to some performance improvement since the noise only region is excluded in the integration. However, covering the entire signal region is not necessarily an optimal solution.

In this paper, determining the optimal integration region includes the estimation of two parameters: the start point of the integration and the length of the integration. The start point estimation is a similar problem to the fine synchronization that estimates the time of arrival (TOA) of the first significant tap. Previous literature, e.g. [35, 36], proposed methods such as energy detection or maximum likelihood detection on the TOA estimation problem. Performance sensitivity to the timing inaccuracy of a non-coherent UWB system was derived in [24]. Theoretical discussion on the relationship between BER performance and the integration length of PPM non-coherent receivers and TR receivers can be found in [32] and [37], respectively. Except for the SNR maximization criterion derived, no practical method was given

on how to carry out the optimal integration length estimation in both papers. In this chapter, we develop an optimum integration region estimation method that is able to perform timing acquisition from the noise corrupted received signal. The method we propose is based on the idea of maximizing the captured receiving SNR at the integrator. In particular, we propose a data-aided approach that first performs a frame-level coarse timing, followed by accurate timing acquisition that determines the start point of the detection integration and then estimates the optimal integration length.

The idea of data-aided timing using inter-symbol correlation can be traced back to previous research on the synchronization for UWB signals in [38] and [39]. These two papers studied frame-level synchronization, which is similar to the coarse timing step in our method. Since both [38] and [39] focused on bipolar modulation with coherent detection, they did not deal with the integration length problem. In this chapter, we further apply the idea of inter-symbol correlation with training sequence to non-coherent and autocorrelation schemes to perform fine timing and determine the optimal integration interval. Due to the transmitted reference signal structure, we are able to devise a relatively simple training sequence.

The rest of this chapter is organized as follows. A theoretical BER performance analysis of the optimal integration region is given in Section 3.2. Section 3.3 proposes a practical estimation approach using training sequences. Section 3.4 presents some simulation results using the developed method and Section 3.5 gives concluding remarks.

3.2 BER Performance Analysis

In this section, the performance analysis will first focus on BPPM modulation, and the results can be easily extended to the transmitted reference scheme thereafter. BPPM and TR signals follow the signal models in Chapter 2. Particularly, (2.15) and (2.16) are used to detect the BPPM signal. In the absence of IFI and IPI, demodulation and detection can be carried out symbol by symbol. Thus in the subsequent discussion in this section the index i is omitted for notation brevity. Intuitively, since more noise will be counted into the decision statistics with a longer integration length, the integration region should be chosen within the non-zero support of $g(t)$, i.e., $T_0 \geqslant \tau_1$ and $T_0 + T \leqslant \tau_L + T_p$. Suppose in the i^{th} symbol interval $d = 0$ is transmitted, the detector outputs are then given by

$$\begin{aligned} y_1 &= E_b \int_{T_0}^{T_0+T} |g(t)|^2 \mathrm{d}t + \sqrt{\frac{E_b}{N_f}} \sum_{n=1}^{N_f} \int_{T_0}^{T_0+T} g(t) n^*(t + nT_f) \mathrm{d}t \\ &\quad + \sqrt{\frac{E_b}{N_f}} \sum_{n=1}^{N_f} \int_{T_0}^{T_0+T} g^*(t) n(t + nT_f) \mathrm{d}t + \sum_{n=1}^{N_f} \int_{T_0}^{T_0+T} |n(t + nT_f)|^2 \mathrm{d}t \\ &\triangleq E_{cap}(T_0, T) + \zeta_1 + \zeta_2 + \zeta_3 \end{aligned} \tag{3.1}$$

$$y_2 = \sum_{n=1}^{N_f} \int_{T_f/2+T_0}^{T_f/2+T_0+T} |n(t + nT_f)|^2 \mathrm{d}t \triangleq \zeta_4 \tag{3.2}$$

where $E_{cap}(T_o, T) = E_b \int_{T_0}^{T_0+T} |g(t)|^2 \mathrm{d}t$ is the signal energy captured in the integration, ζ_1 and ζ_2 are the signal-noise cross terms, and ζ_3 and ζ_4 are the noise-noise cross terms. As shown in Appendix A, when $T \gg T_p$, they can be approximated as independent Gaussian random variables of which the distributions are respectively

$\zeta_1, \zeta_2 \sim \mathcal{N}\big(0, N_0 E_{cap}(T_0, T)\big)$ and $\zeta_3, \zeta_4 \sim \mathcal{N}\big(2N_f N_0 BT, 2N_f N_0^2 BT\big)$.

Similar to the expression given in [32], the bit error rate of BPPM non-coherent detection is given by

$$\begin{aligned} P_{e,PPM} &= P(y_1 < y_2) = P\big(E_{cap}(T_0, T) + \zeta_1 + \zeta_2 + \zeta_3 - \zeta_4 < 0\big) \\ &= Q\left(\sqrt{\frac{E_{cap}^2(T_0, T)}{4N_f N_0^2 BT + 2N_0 E_{cap}(T_0, T)}}\right) \end{aligned} \tag{3.3}$$

where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} \mathrm{d}t$. Eq. (3.3) shows that the BER of BPPM signal largely depends on the choice of T_0 and T. Therefore the optimization of T_0 and T is equivalent to the maximization

$$(\hat{T}_0, \hat{T}) = \arg\max_{T_0, T} \frac{E_{cap}^2(T_0, T)}{4N_f N_0^2 BT + 2N_0 E_{cap}(T_0, T)}. \tag{3.4}$$

Noticing that most of the energy in $h(t)$ is in the front part while the latter part contains relatively small scattering and low-energy pulses traveling from far. By including the latter part of the signal region into the integration interval the additional collected signal energy may not compensate the additional noise energy. Therefore, the optimal integration length T is usually smaller than the length of signal region.

Although the two-dimensional maximization in (3.4) can be solved by trying all the possible values within $T_0 \in [\tau_1, \tau_L + T_p]$ and $T \in [0, \tau_L + T_p - T_0]$ for both variables, this approach has a computation complexity of $O\big((\frac{\tau_L + T_p - \tau_1}{\Delta t})^2\big)$, which grows rapidly as $\tau_L + T_p - \tau_1$ gets larger or the trying step size Δt gets smaller. In order to alleviate the computation task, we hope to fix one degree of freedom first and then perform the maximization over the other variable. It is found through simulation that deeming

T_0 as the arrival time of the first "significant" tap, i.e., the first tap whose captured energy exceeds a small threshold ξ, involves only negligible performance degradation compared to the optimum. Thus a threshold crossing (TC) scheme can be used in the estimation of T_0. That is,

$$\begin{cases} \hat{T}_0 = \max\left(\tau | E_{cap}(\tau_1, \tau - \tau_1) < \xi\right) \\ \hat{T} = \arg\max\limits_{T} \dfrac{E_{cap}^2(\hat{T}_0, T)}{4N_f N_0^2 BT + 2N_0 E_{cap}(\hat{T}_0, T)}. \end{cases} \tag{3.5}$$

Similar derivation can lead to the optimal integration interval of the transmitted reference signals. Again, suppose $d = 0$ is transmitted, the autocorrelator output given by (2.19) can be expressed as

$$\begin{aligned} y_{TR} = & -\frac{E_b}{2}\int_{T_0}^{T_0+T} |g(t)|^2 \mathrm{d}t + \sqrt{\frac{E_b}{2N_f}} \sum_{n=1}^{N_f} \int_{T_0}^{T_0+T} \Re\big\{g(t) n^*(t + nT_f + \frac{T_f}{2})\big\}\mathrm{d}t \\ & + \sqrt{\frac{E_b}{2N_f}} \sum_{n=1}^{N_f} \int_{T_0}^{T_0+T} \Re\big\{g^*(t) n(t + nT_f)\big\}\mathrm{d}t \\ & + \sum_{n=1}^{N_f} \int_{T_0}^{T_0+T} \Re\big\{n(t + nT_f) n^*(t + nT_f + \frac{T_f}{2})\big\}\mathrm{d}t \\ \triangleq & -E_{cap}(T_0, T)/2 + \zeta_1 + \zeta_2 + \zeta_3. \end{aligned} \tag{3.6}$$

Similar to the PPM case, the noise terms ζ_1, ζ_2, ζ_3 can be approximated as independent Gaussian random variables when $T \gg T_p$, with distributions as $\zeta_1, \zeta_2 \sim \mathcal{N}\big(0, \frac{N_0 E_{cap}(T_0,T)}{4}\big)$, $\zeta_3 \sim \mathcal{N}\big(0, N_f N_0^2 BT\big)$. Thus the BER of a TR receiver can be

written as

$$\begin{aligned} P_{e,TR} &= P(y_{TR} > 0) = P(-E_{cap}(T_0, T)/2 + \zeta_1 + \zeta_2 + \zeta_3 > 0) \\ &= Q\left(\sqrt{\frac{E_{cap}^2(T_0, T)}{4N_f N_0^2 BT + 2N_0 E_{cap}(T_0, T)}}\right). \end{aligned} \tag{3.7}$$

A similar expression for TR can be found in [17]. From (3.7), the optimal integration interval for TR system is given by

$$(\hat{T}_0, \hat{T}) = \arg\max_{T_0, T} \frac{E_{cap}^2(T_0, T)}{4N_f N_0^2 BT + 2N_0 E_{cap}(T_0, T)}. \tag{3.8}$$

3.3 Optimization Using Training Sequence

Comparing (3.3) and (3.7), the BER performance of BPPM and TR are exactly the same, provided that all the parameters in (3.3) and (3.7) are equal to each other. This implies the optimum integration region of these two schemes should be exactly the same for the same channel condition. In this section, a data-aided $\{T_0, T\}$-estimation method for both schemes is presented. Because the designed training sequence is based on the TR signal structure, we introduce our method for the TR scheme. However, exactly the same training pulses and the estimation method apply to the BPPM scheme, since both schemes have the equivalent optimal integration region. Suppose at the i^{th} bit interval, $d^{(i)} = 1$, and at the $(i+1)^{th}$ bit interval, $d^{(i+1)} = 0$. Fig. 3.1 depicts the noiseless version of the two received TR signals $r^{(i)}(t)$ and $r^{(i+1)}(t)$ with $N_f = 4$. The integration region optimization procedure consists of three steps:

1) Coarse timing

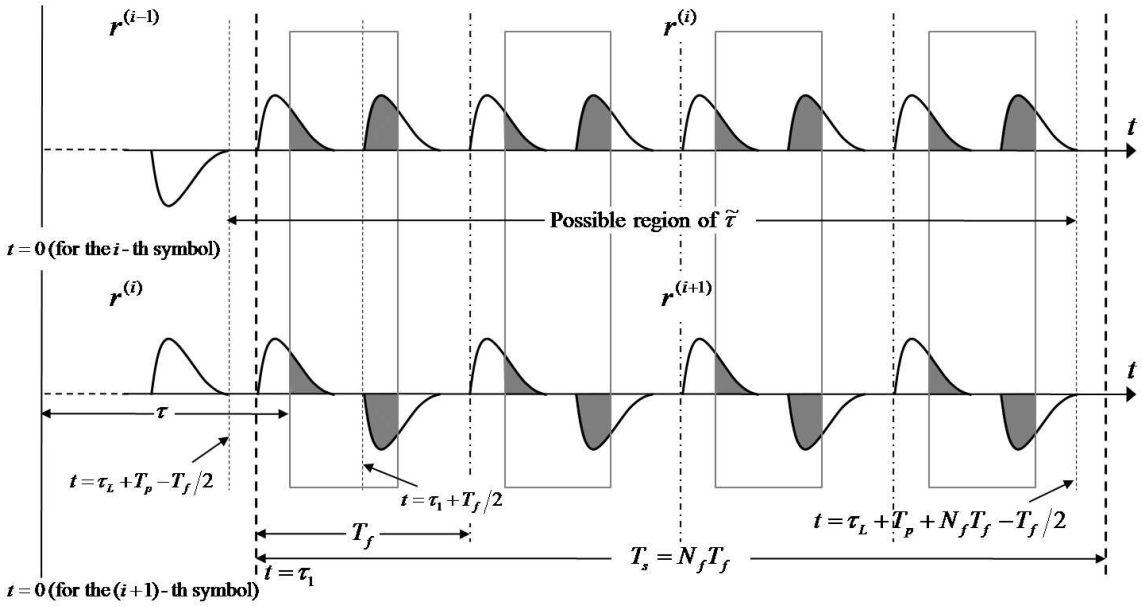

Figure 3.1: Received TR signal (noiseless) $r^{(i)}(t)$ and $r^{(i+1)}(t)$, $N_f = 4$.

Let the receiver catch τ at an arbitrary point $\tau_0 = iN_fT_f + \tilde{\tau}$ at the very beginning, where i is an integer and $\tilde{\tau}$ is in the region $[\tau_L + T_p - T_f/2, \tau_L + T_p + (N_f - 1/2)T_f]$ as shown in Fig. 3.1. Perform the following integrations

$$I_j = \left| \sum_{n=0}^{N_f-1} \int_{\tilde{\tau}}^{\tilde{\tau}+\frac{T_f}{2}} r^{(i)}(t+nT_f+j\frac{T_f}{2}) r^{(i)*}(t+nT_f+\frac{T_f}{2}+j\frac{T_f}{2}) \mathrm{d}t \right|, \quad j = 0 \cdots 2N_f - 1. \tag{3.9}$$

If for a certain j, $\tilde{\tau} + j\frac{T_f}{2}$ falls in the first half of the first frame in a symbol plus a small interval from the end of the last received pulse of the previous symbol to the beginning of the current symbol, i.e., $\tilde{\tau} + j\frac{T_f}{2} \in [\tau_L + T_p - T_f/2, \tau_1 + T_f/2]$, the noiseless part of I_j will be $\frac{E_b}{2}$. Then for another j, if $\tilde{\tau} + j\frac{T_f}{2}$ falls in the second half of the first frame, the noiseless part of I_j becomes $\frac{E_b}{2}\big(1 - 2E(\dot{\tilde{\tau}})/N_f\big) < \frac{E_b}{2}$, where $\dot{\tilde{\tau}} \triangleq (\tilde{\tau} - \tau_1 \mathrm{mod}\ \frac{T_f}{2}) + \tau_1$. For other j's, $\tilde{\tau} + j\frac{T_f}{2}$ fall into other frames of the current

symbol, and the noiseless part of I_j will be even smaller. Therefore simply shifting τ to

$$\hat{\tau} = \tau_0 + \frac{T_f}{2}\arg\max_j I_j \tag{3.10}$$

will ensure $\hat{\tau}$ to be in the first half of the first frame in a symbol plus a small interval before it, i.e., $[iN_fT_f + \tau_L + T_p - T_f/2, iN_fT_f + \tau_1 + T_f/2]$ or $[(i+1)N_fT_f + \tau_L + T_p - T_f/2, (i+1)N_fT_f + \tau_1 + T_f/2]$, which is the prerequisites of the following fine timing steps.

2) Estimating T_0, or equivalently τ_1

After the coarse timing step we have $\hat{\tau} \in [\tau_L + T_p - T_f/2, \tau_1 + T_f/2]$, which is indicated in Fig. 3.1. Note that since whether $\hat{\tau}$ is at the beginning of the i^{th} symbol or the $(i+1)^{\text{th}}$ symbol does not affect the fine timing step, parameter i is omitted here for brevity. Now do the following integration

$$x(\tau) = \Re\left\{ \sum_{n=0}^{N_f-1} \int_{\tau}^{\tau+\frac{T_f}{2}} r^{(i)}(t+nT_f) r^{(i+1)*}(t+nT_f)\mathrm{d}t \right\}, \;\; \hat{\tau} - \frac{T_f}{2} \leqslant \tau \leqslant \hat{\tau} + \frac{T_f}{2} \tag{3.11}$$

for all τ's in the range $[\hat{\tau} - \frac{T_f}{2}, \hat{\tau} + \frac{T_f}{2}]$, which is just the correlation between the shadowed parts of $r^{(i)}(t)$ and $r^{(i+1)}(t)$ in Fig. 3.1. Define $E(\tau) \triangleq \int_0^{\tau} |g(t)|^2 \mathrm{d}t$, then

$$x(\tau) = \frac{E_b}{2}\big[1 - 2E(\tau)\big] + \zeta(\tau) \tag{3.12}$$

where $\zeta(\tau)$ is approximately a zero-mean Gaussian noise process with variance $\sigma_\zeta^2 = (N_fN_0^2BT_f + N_0E_b)/2$. A plot of $x(\tau)$ for $\tau \geqslant \tau_L + T_p - T_f/2$ including the effect of noise ζ is illustrated in Fig. 3.2.

When $\tau \in [\tau_L + T_p - T_f/2, \tau_1]$, which is the blank space from the end of the last

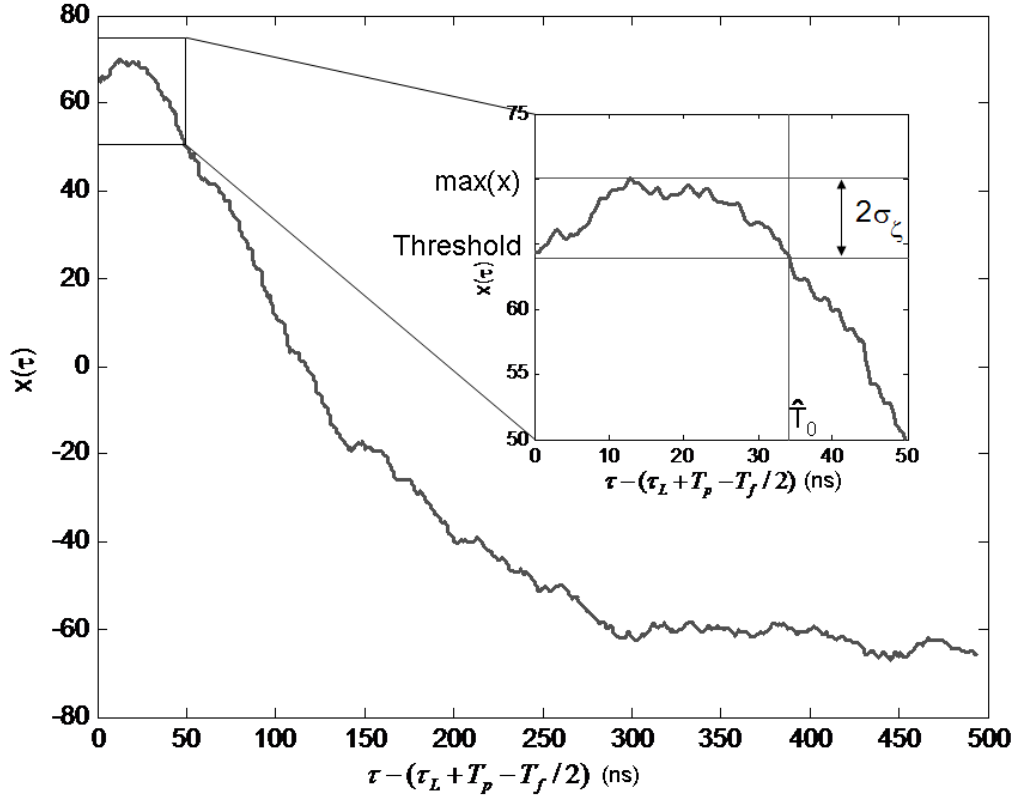

Figure 3.2: Typical curve of $x(\tau)$ with detailed display of the threshold and estimated T_0.

received pulse to the begining of the current pulse, $E(\tau)$ achieves its minimum value 0, which leads $x(\tau)$ to its maximum $E_b/2$ if neglecting $\zeta(\tau)$. When $\tau < \tau_L + T_p - T_f/2$, $x(\tau)$ monotonously increases, whereas when $\tau > \tau_1$, $x(\tau)$ monotonously decreases. In other words, a noiseless $x(\tau)$ has a plateau shape with flat top in the interval $\tau \in [\tau_L + T_p - \frac{T_f}{2}, \tau_1]$. Thus we can estimate T_0 as

$$\hat{T}_0 = \max\big(\tau | x(\tau) > x_0\big),\ \hat{\tau} - \frac{T_f}{2} \leqslant \tau \leqslant \hat{\tau} + \frac{T_f}{2} \tag{3.13}$$

where the threshold $x_0 \approx E_b\big(\frac{1}{2} - \xi\big)$. The rule used to decide the value of x_0 is given later.

3) Estimating T

The candidates of T are confined in the region $[0, T_f/2]$. Comparing the definition of $E(\tau)$ and $E_{cap}(\hat{T}_0, T)$ and using (3.5) and (3.2), T can be estimated as

$$\hat{T} = \arg\max_{\tau} \frac{\big[x(\hat{T}_0) - x(\tau + \hat{T}_0)\big]^2}{4N_f N_0^2 B\tau + 2N_0\big[x(\hat{T}_0) - x(\tau + \hat{T}_0)\big]}. \tag{3.14}$$

Defining the SNR $\gamma \triangleq E_b/N_0$, and normalizing $x(\tau)$ to $\tilde{x}(\tau) \triangleq x(\tau)/E_b$, (3.14) can be rewritten as

$$\hat{T} = \arg\max_{\tau} \frac{\gamma^2\big[\tilde{x}(\hat{T}_0) - \tilde{x}(\tau + \hat{T}_0)\big]^2}{4N_f B\tau + 2\gamma\big[\tilde{x}(\hat{T}_0) - \tilde{x}(\tau + \hat{T}_0)\big]}. \tag{3.15}$$

To use (3.14) or (3.15) the knowledge of the SNR γ is required at the receiver. When γ is not too large, e.g. when $\gamma < 15$ dB for CM1 and when $\gamma < 20$ dB for CM8, which are quite reasonable in practical environments, the second term in the denominator is much smaller than the first one. For example, when $N_f = 4$, $B = 500$

MHz, $T_f = 1$ μs, $\gamma = 17$ dB, τ will be more than 125 ns for a typical CM8 channel, thus the first term $4N_f B\tau > 1000$ while the second term $2\gamma(\tilde{x}(\hat{T}_0) - \tilde{x}(\tau + \hat{T}_0)) < 100$. Therefore we can simply ignore the second term in the denominator of (3.14) or (3.15) so that they can be further simplified to

$$\hat{T} = \arg\max_{\tau} \frac{\left[x(\hat{T}_0) - x(\tau + \hat{T}_0)\right]^2}{\tau}. \tag{3.16}$$

Eq. (3.16) is the formula adopted in Section V to estimate T. For a system operating in a high SNR environment we can use (3.15) instead of (3.16) and substituting γ by a fixed SNR value less than the true SNR. It is found in our simulation that if the true SNR is in the 10–20 dB region, γ fixed at 10 dB will work well.

The discussion so far has not considered the nuisance noise term $\zeta(\tau)$. When the noise term is taken into account, the useful signal waveform will be distorted, sometimes even inundated, making accurate estimation almost impossible without necessary denoising treatment. Since $\zeta(\tau)$ is approximately a Gaussian random process, an effective way of denoising is by averaging. To better fulfill this task, a training sequence is implemented. Observing (3.11), we can see that $x(\tau)$ is essentially the correlation of two consecutive TR symbols, one conveying data "1" and the other conveying "0". Thus a training sequence can be designed such that "0"s and "1"s appear alternately, i.e.,

$$d_t^{(i)} = (-1)^{i-1}, \quad i = 1 \cdots N \tag{3.17}$$

where an even integer N denotes the length of the sequence. With the N-symbol training sequence, there are now $N - 1$ pairs of consecutive symbols that can be used to generate $x(\tau)$. We can average the $N - 1$ correlation outcomes to arrive at

a revised $x(\tau)$ as

$$\bar{x}(\tau) = \frac{1}{N-1}\Re\left\{ \sum_{i=1}^{N-1}\sum_{n=0}^{N_f-1} \int_{\tau}^{\tau+\frac{T_f}{2}} r^{(i)}(t+nT_f)\cdot r^{*(i+1)}(t+nT_f)\mathrm{d}t \right\}. \tag{3.18}$$

It is easy to find that $\bar{x}(\tau)$ has a smaller noise variance of $\sigma_{\bar{\zeta}}^2 = (N_f N_0^2 BT_f + N_0 E_b)/2(N-1)$.

Furthermore, if we are able to implement analog averaging over several symbols, the variance of the noise term can be further reduced by first averaging over the analog symbol waveforms and then correlating the averaged signals like

$$\bar{\bar{x}}(\tau) = \Re\left\{ \sum_{n=0}^{N_f-1} \int_{\tau}^{\tau+T_f/2} \Big(\frac{2}{N}\sum_{i=1}^{N/2} r^{(2i-1)}(t+nT_f)\Big)\cdot\Big(\frac{2}{N}\sum_{i=1}^{N/2} r^{*(2i)}(t+nT_f)\Big)\mathrm{d}t \right\}. \tag{3.19}$$

The variance of noise term of $\bar{\bar{x}}(\tau)$ is $\sigma_{\bar{\bar{\zeta}}}^2 = (2N_f N_0^2 BT_f + NN_0 E_b)/N^2$. Obviously, a longer training sequence will result in smaller $\sigma_{\bar{\zeta}}^2$ and $\sigma_{\bar{\bar{\zeta}}}^2$, which makes the estimation more accurate. With a large N, $\sigma_{\bar{\zeta}}^2$ will be much greater than $\sigma_{\bar{\bar{\zeta}}}^2$, thus the improvement of $\bar{\bar{x}}(\tau)$ becomes more prominent than $\bar{x}(\tau)$.

The value of the threshold x_0 can be decided according to the chosen averaging scheme. In simulation, we choose x_0 as $\max\big(x(\tau)\big) - 2\sigma_\zeta$ if no denoising process is employed at all. If (3.18) or (3.19) is used for averaging, the threshold should be adjusted to $\bar{x}_0 = \max\big(\bar{x}(\tau)\big) - 2\sigma_{\bar{\zeta}}$ or $\bar{\bar{x}}_0 = \max\big(\bar{\bar{x}}(\tau)\big) - 2\sigma_{\bar{\bar{\zeta}}}$, respectively. An example of the threshold x_0 and the estimated $\hat{T}_0$ is shown in Fig. 3.2.

Similar to (3.18) and (3.19), averaging process can also be performed on the coarse

timing step, using the following equations instead of (3.9),

$$\bar{I}_j = \frac{1}{M}\sum_{i=0}^{M-1}\left|\sum_{n=0}^{N_f-1}\int_{\tilde{\tau}}^{\tilde{\tau}+\frac{T_f}{2}} r^{(m+i)}(t+nT_f+j\frac{T_f}{2})\cdot r^{*(m+i)}(t+nT_f+\frac{T_f}{2}+j\frac{T_f}{2})\mathrm{d}t\right| \qquad j=0\cdots 2N_f-1 \tag{3.20}$$

or

$$\begin{aligned}\bar{\bar{I}}_j =& \frac{2}{M^2}\left|\sum_{n=0}^{N_f-1}\int_{\tilde{\tau}}^{\tilde{\tau}+\frac{T_f}{2}}\Big(\sum_{i=0}^{\frac{M}{2}-1} r^{(m+2i)}(t+nT_f+j\frac{T_f}{2})\Big)\right.\\ &\left.\cdot\Big(\sum_{i=0}^{\frac{M}{2}-1} r^{*(m+2i)}(t+nT_f+\frac{T_f}{2}+j\frac{T_f}{2})\Big)\mathrm{d}t\right|\\ &+\frac{2}{M^2}\left|\sum_{n=0}^{N_f-1}\int_{\tilde{\tau}}^{\tilde{\tau}+\frac{T_f}{2}}\Big(\sum_{i=0}^{\frac{M}{2}-1} r^{(m+2i+1)}(t+nT_f+j\frac{T_f}{2})\Big)\right.\\ &\left.\cdot\Big(\sum_{i=0}^{\frac{M}{2}-1} r^{*(m+2i+1)}(t+nT_f+\frac{T_f}{2}+j\frac{T_f}{2})\Big)\mathrm{d}t\right|\\ &j=0\cdots 2N_f-1\end{aligned} \tag{3.21}$$

where m is the index of the first symbol in the received signal and M is the length of the training sequence that is used to do the coarse timing. One thing worth noting is that with coarse timing and fine synchronization together, the length of the training sequence should be at least $M+N$ so that there are M training symbols used for coarse timing and there are another N symbols used for fine synchronization and estimating the optimal integration length T.

Furthermore, some implementation considerations are presented in the following. First, the above discussion of the timing method is based on the continuous signal.

To search for the estimation of T_0 and T exhaustively in the continuous region $[\hat{\tau} - \frac{T_f}{2}, \hat{\tau} + \frac{T_f}{2}]$ and $[\hat{T}_0, \hat{T}_0 + T_f/2]$, respectively, is a formidable task, if not impossible at all. To reduce the complexity, the search can be performed on a discrete series of time instants instead. The distance between adjacent time instants, denoted by T_b, depends on the required synchronization resolution.

Since the searching process requires the whole training sequence to be sent once for every calculation of $x(\tau)$, a long preamble would be needed which affects the achievable data rate. To avoid this problem we can substitute the integration in (3.11) with a series of integrations over smaller intervals. That is,

$$x_k = \Re\left\{ \sum_{n=0}^{N_f-1} \int_{\hat{\tau}+kT_b}^{\hat{\tau}+(k+1)T_b} r^{(i)}(t+nT_f) r^{(i+1)*}(t+nT_f) \mathrm{d}t \right\}, \;\; k = -Q, \cdots, 2Q-1 \quad (3.22)$$

where $Q = \lfloor \frac{T_f}{2T_b} \rfloor$ and $\lfloor x \rfloor$ takes the integral part of x. The x_k's are stored in the registers and then (3.11) can be obtained by digital processing as

$$x(\tau)|_{\tau=\hat{\tau}+kT_b} = \sum_{j=k}^{k+Q-1} x_j, \quad k = -Q, \cdots, Q-1. \quad (3.23)$$

Consequently, (3.13) and (3.16) can be expressed as

$$\hat{T}_0 = T_b \max\left(\{k | x(\hat{\tau} + kT_b) > x_0, \; k = -Q, \cdots, Q-1\} \right) \quad (3.24)$$

and

$$\hat{T} = T_b \cdot \arg\max_k \frac{\left[x(\hat{T}_0) - x(kT_b + \hat{T}_0) \right]^2}{kT_b}, \quad k = 1, \cdots, Q \quad (3.25)$$

and (3.18) and (3.19) can be modified accordingly.

Note that the algorithm still requires an analog delay line as long as a symbol duration, which is out of the capability of a normal wideband analog delay line. This is a well known open problem in the TR literature. Future research will focus on modifying the algorithm to avoid the long delay requirement.

3.4 Simulation Results

In this section, simulation results of PPM non-coherent detection and transmitted reference scheme are presented for IEEE 802.15.4a CM1 (indoor residential line-of-sight environment) and CM8 (indoor industrial non-line-of-sight environment) channel models, respectively. The default parameters for the simulations are as follows: number of frames in a symbol $N_f = 1$, frame duration $T_f = 1$ μs, the bandwidth $B = 494$ MHz, the sampling rate for the simulation $f_s = 3.952$ GHz, the shaping pulse $p(t)$ is a root raised cosine pulse with roll-off factor $\beta = 0.25$, the resolution T_b is the simulation sampling rate. All the simulation results are the average of the error rate in 100 different channel realizations. Synchronization is performed on each of the channel realizations.

First, the impact of integration region on the BPPM and TR system performance are studied to show the importance of conducting an integration region optimization on each of the channel realizations. In Fig. 3.3, the effect of T_0 or T mistiming is evaluated for TR systems in CM1 and CM8 channels, respectively. We can see that the BER performance in CM1 channels is quite sensitive to the accuracy of T_0 and T estimation. The BER performance in CM8 channels is less sensitive to the estimation accuracy, but estimation errors can still cause visible BER degradation. Fig. 3.4 plots the distributions of the optimal T_0, T and the corresponding BER of

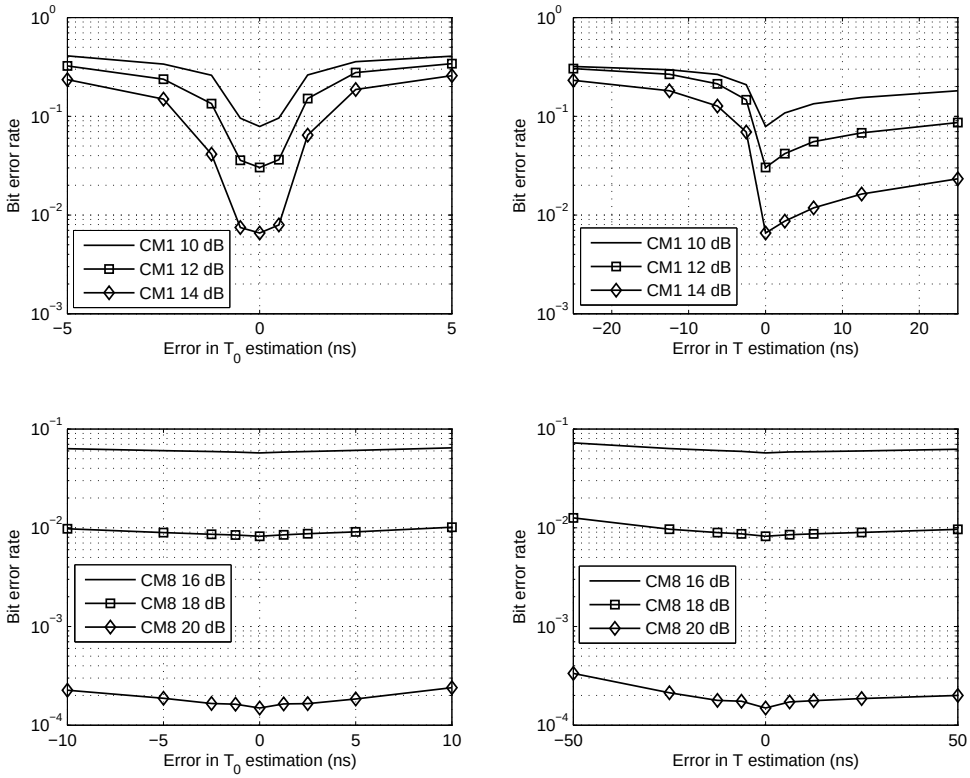

Figure 3.3: The effect of a timing error on the BER performance of the non-coherent and TR schemes. The top two plots represent error in T_0 and T for CM1 channels; the bottom two plots represent error in T_0 and T for CM8 channels.

50 channel realizations for CM1 and CM8. It is shown that both T_0 and T may vary in a large range between different channel realizations, and the corresponding BER also changes greatly. Therefore, for every single channel realization an integration region optimization is necessary for the non-coherent detection and autocorrelation detection systems.

Figs. 3.5-3.6 show the BER performance of four different scenarios, including PPM modulation in the CM1 channel model, TR scheme in the CM1 channel model, PPM

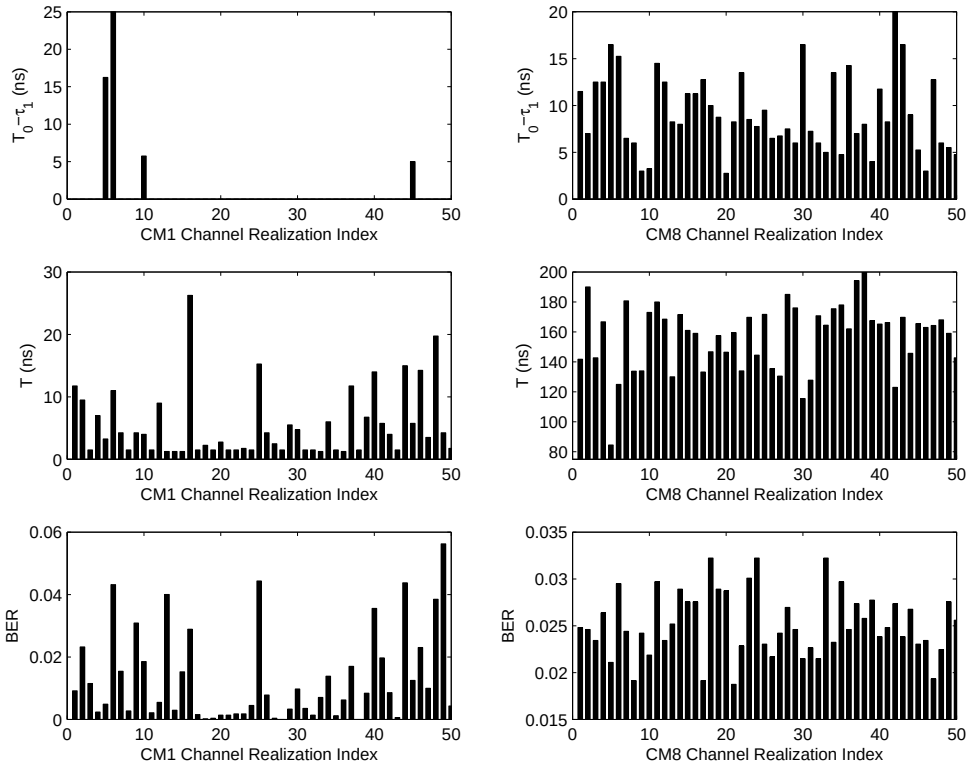

Figure 3.4: The distribution of the optimal integration region and the corresponding BER of 50 CM1 and CM8 channel realizations.

modulation in the CM8 channel model, and TR scheme in the CM8 channel model. In all simulations, we assume that the coarse timing is perfectly done beforehand and we only focus on the fine timing steps. Each figure contains five BER vs. SNR curves. Among them, the theoretically optimal curve stands for the method given by (3.4) and (3.8). With the assumption that full channel state information is known at the receiver, the theoretically optimal integration region is obtained from numerical calculation of (3.4) and (3.8), with the calculation step size be the reciprocal of sampling rate, which is one eighth of the pulse duration. The "Estimated with $\bar{x}$" curve represents the proposed estimation method using (3.13) and (3.16)-(3.18). Eq. (3.17) is used to construct a training sequence, (3.13) and (3.16) are used to estimate T_0 and T respectively, and eq. (3.18) is used for averaging. The "Estimated with $\bar{\bar{x}}$" curve represents the estimation method that utilizes almost the same equations as the "Estimated with $\bar{x}$" method, except this time eqs. (3.19) is employed to do the averaging. It is clearly shown that this estimation method has a very close performance to the theoretically optimal method. The performance of the method using (3.18) for denoising is slightly worse than the method using (3.19), because the noise variances of (3.18) is larger than that of (3.19).

The rest two non-optimal methods, one determines the integration region from the amplitude of the channel gain and the other fixes the integration length as constants, are included in Figs. 3.5-3.6 for comparison to show the benefit of performing integration region optimization. The "Scale=0.5" method defines the integration start and end points to be the first and last taps that have magnitude greater than or equal to 50% of the strongest tap. Since it uses the channel amplitude instead of the channel energy to determine the integration region, this approach has worse

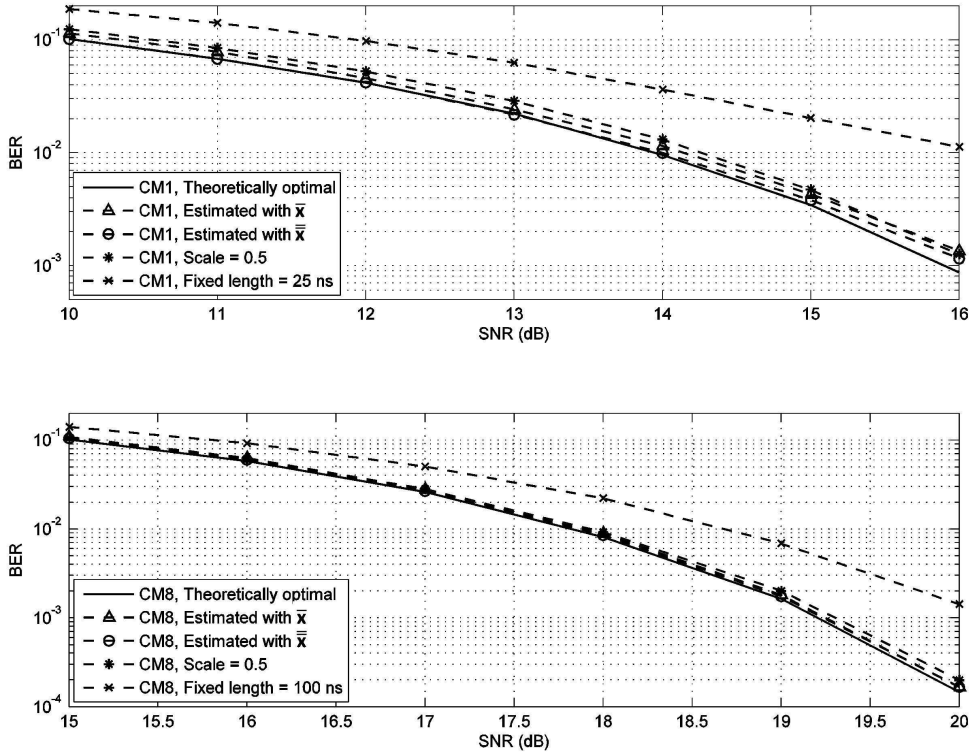

Figure 3.5: BER obtained with 5 different integration regions for PPM non-coherent detection in CM1 and CM8 channel realizations.

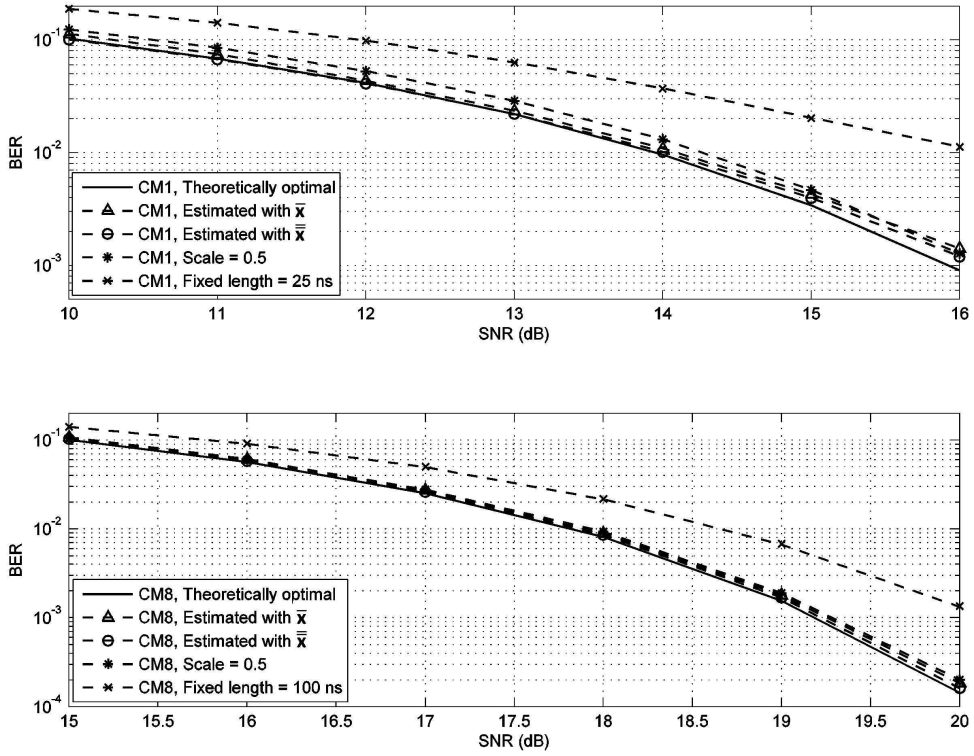

Figure 3.6: BER obtained with 5 different integration regions for transmitted reference in CM1 and CM8 channel realizations.

performance than the theoretical and estimated optimal ones. Note that this method also requires channel state information, thus is difficult to implement. The "Fixed length" curve integrates the signal from the beginning of a frame (assuming known) through a fixed integration length which is set to be 25 ns for the CM1 channel and 100 ns for the CM8 channel. Apparently, this non-adaptive method gives the worst performance among the five. CM1 is found to be more sensitive to the choice of integration region than CM8. The parameters in the scale method and fixed length method are carefully chosen through simulation trials such that the average BER over different channel realizations can be reduced as much as possible.

We have also performed simulations using (3.15) instead of (3.16), by fixing γ to a value smaller than the actual SNR. The results are shown in Figs. 3.7-3.8. It can be seen that different fixed γ has a negligible impact on a system in CM8 channels. Moreover, fixing γ at a value that is smaller than the true value provides very small performance degradation for CM1 channels. Therefore as an extreme case we can fix γ at 0, which means dropping the second term in the denominator in (3.15). For CM1 channels in a high SNR environment, we can fix γ at a close to but smaller than true SNR value. For example if the true SNR is in the 10–20 dB region, γ fixed at 10 dB will work well.

Figs. 3.9-3.10 display the effect of different training sequence lengths on the BER performance. Since the PPM non-coherent detection case and TR scheme are equivalent in terms of BER performance, we plotted the curves of using (3.18) with PPM non-coherent detection and the curves of using (3.19) with the TR scheme. Actually the results are interchangeable for PPM and TR. As discussed in the previous section, a longer training sequence results in a smaller noise variance and consequently

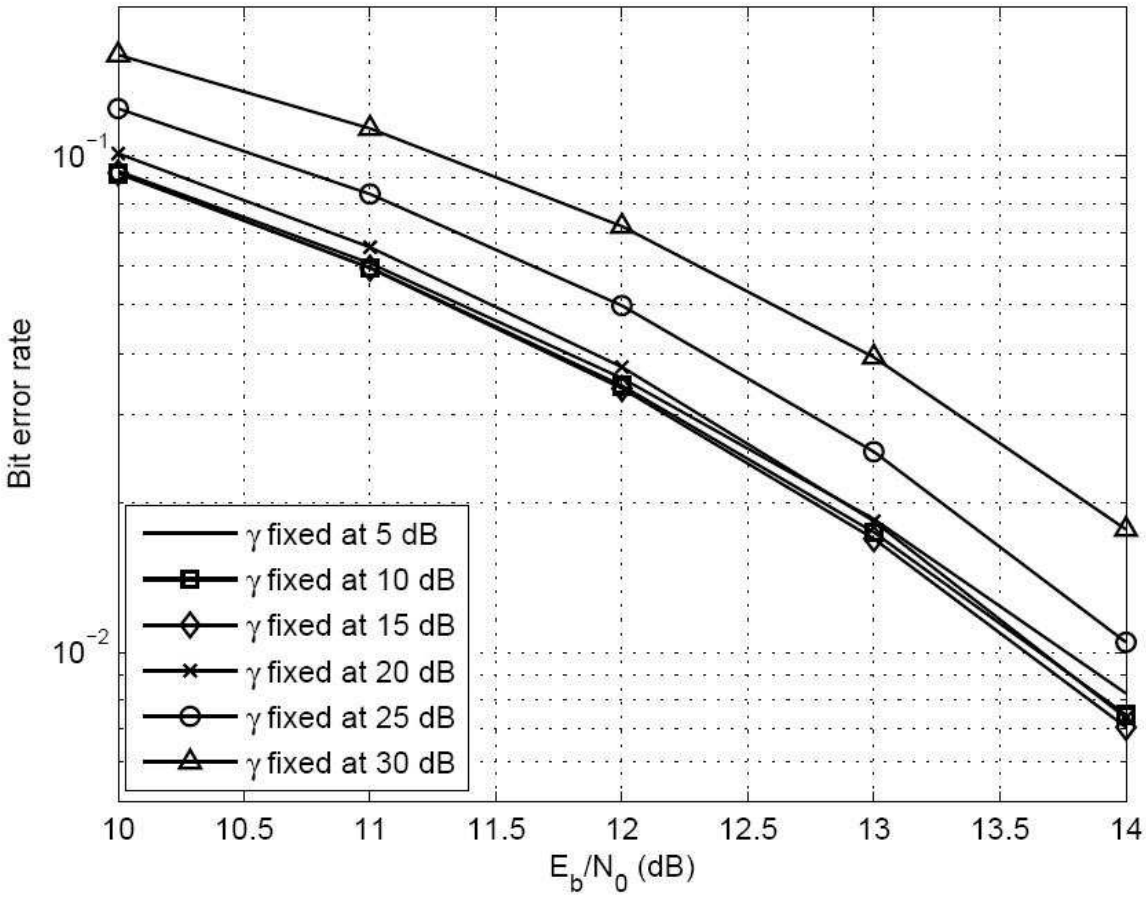

Figure 3.7: The effect of fixing γ in (3.15) for CM1 channels.

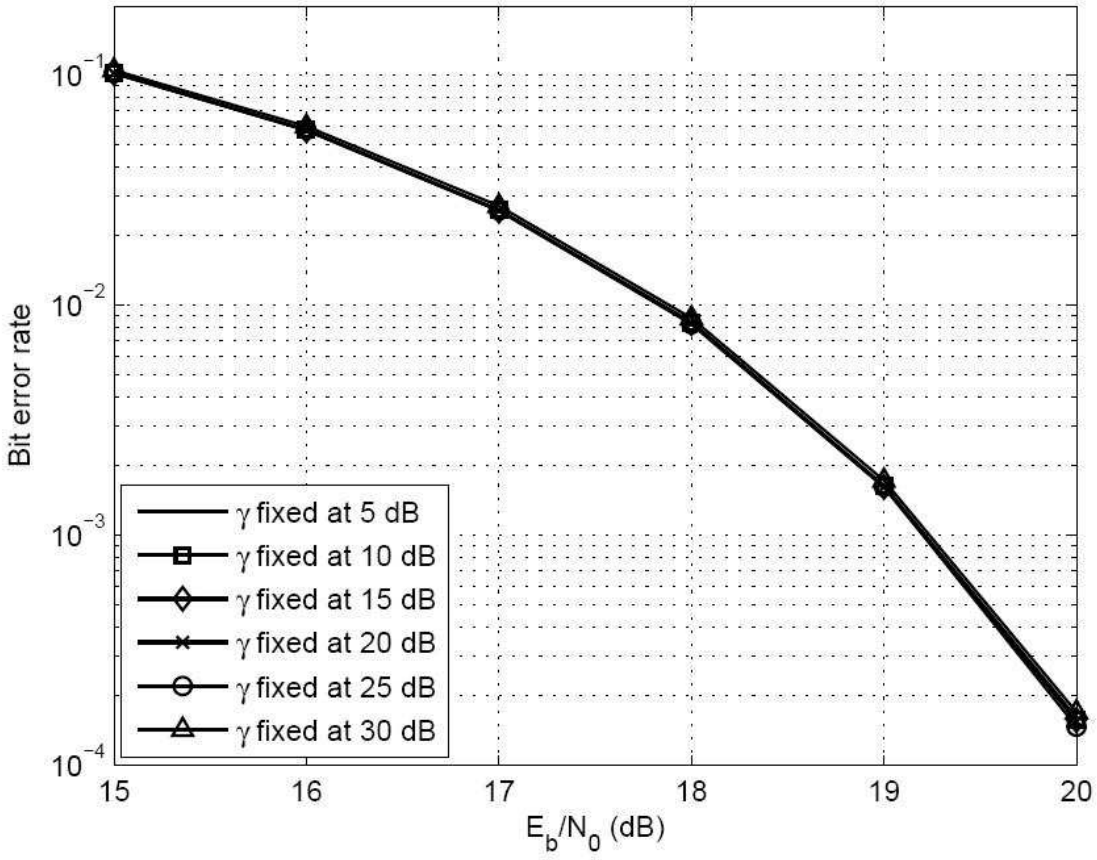

Figure 3.8: The effect of fixing γ in (3.15) for CM8 channels.

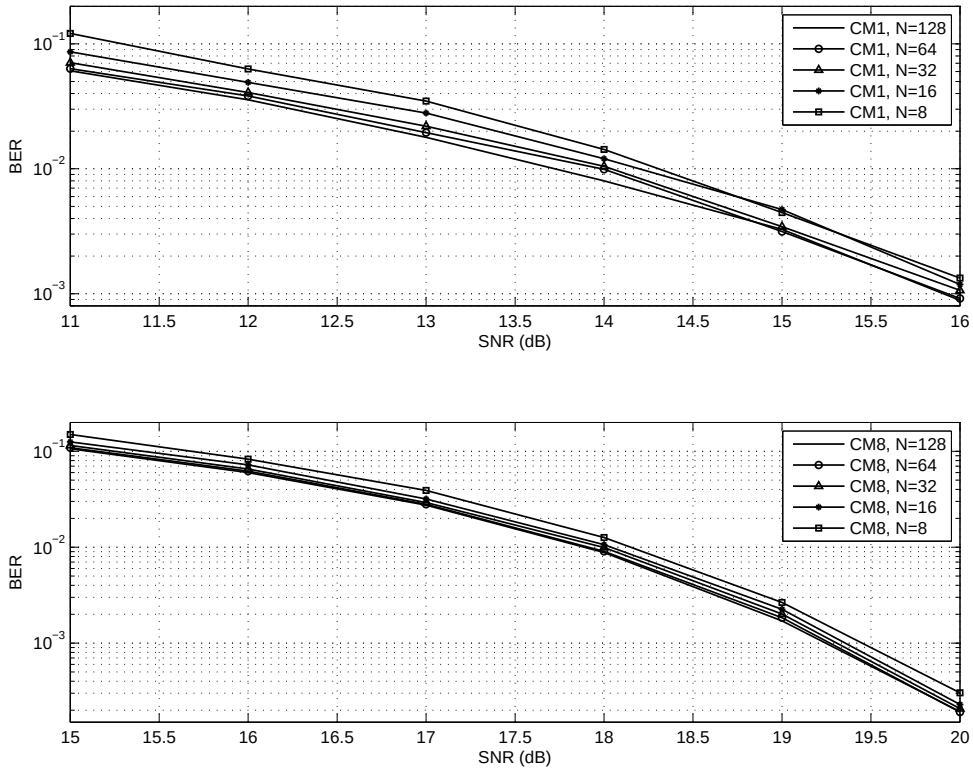

Figure 3.9: The effect of different training sequence length on the performance of PPM non-coherent detection in CM1 and CM8 channel realizations. Eq. (3.18) is used for averaging.

lower BER. Generally, if (3.18) is used, a training sequence with $N = 64$ is sufficient, while if (3.19) is used, $N = 32$ or more makes very little difference on the BER performance.

Fig. 3.11 shows the simulation results of the $N_f = 2$ and $N_f = 4$ cases in CM1 channels for the TR scheme. Compared to Figs. 3.5-3.6, we can find that the difference between a multi-frame signal structure and a single-frame structure is that the former performs integration on multiple intervals thus involves more noise and has a worse

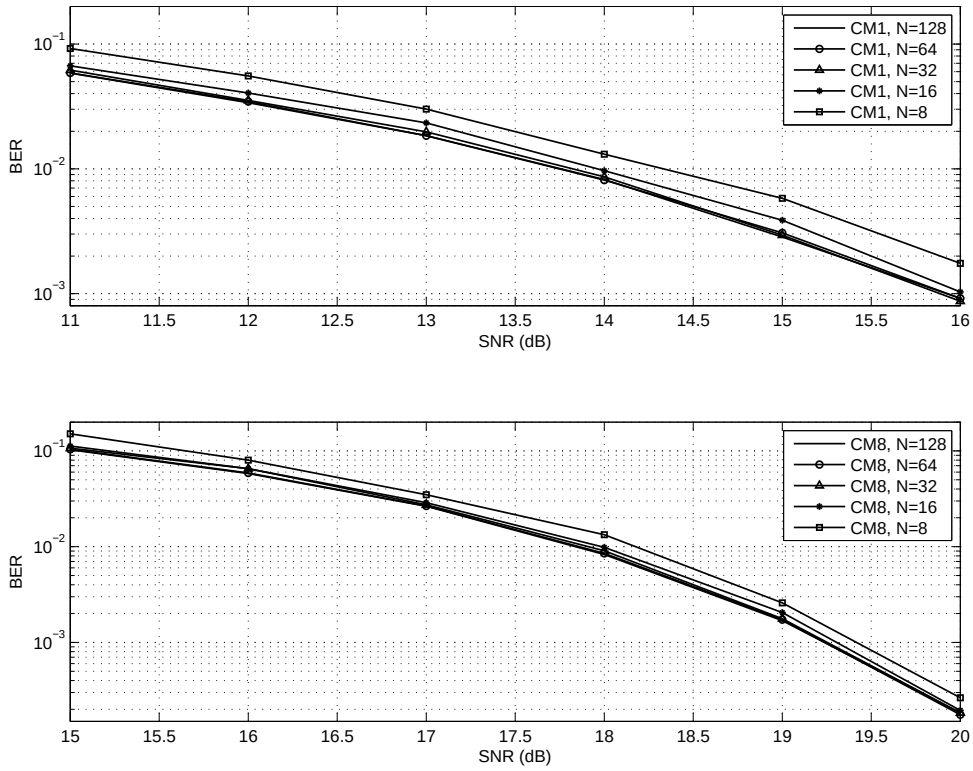

Figure 3.10: The effect of different training sequence length on the performance of transmitted reference in CM1 and CM8 channel realizations. Eq. (3.19) is used for averaging.

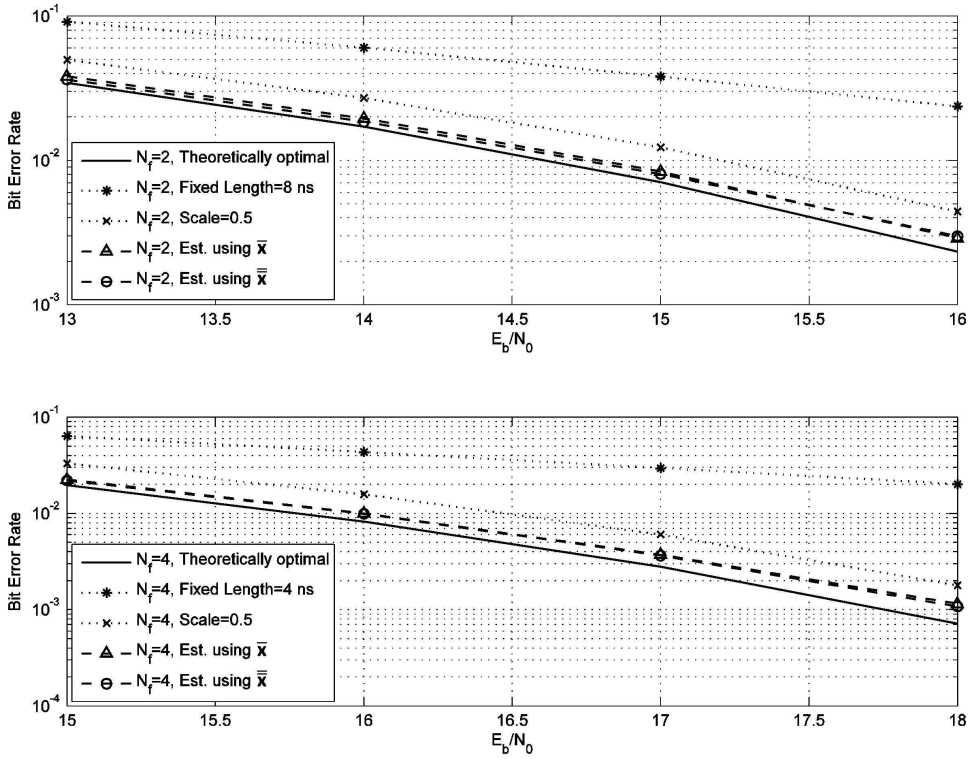

Figure 3.11: The BER performance of 5 different integration region determining methods. The top plot represents the $N_f = 2$ case, the bottom plot represents the $N_f = 4$ case.

performance. For example, performance of an $N_f = 2$ system is roughly 1.5 dB worse than an $N_f = 1$ system, and the same difference exists between an $N_f = 4$ system and an $N_f = 2$ system.

We have also tested the coarse timing step of our algorithm. We find when the SNR is greater than 11 dB in CM1 channels or greater than 15 dB in CM8 channels, an $M = 16$ is sufficient if applying (3.21) and $M = 64$ if employing (3.20) to ensure 90% probability of correct coarse timing. Our simulation also shows that the coarse timing step of the proposed method is more reliable than the methods presented

in [38] and [39]. This is mainly because the method of [38], in which a {+1, +1, -1, -1} pattern training sequence is employed, utilizes only $M/2$ correlations compared to M correlations in the proposed method, and the method of [39] requires an additional analog averaging process to construct a template beforehand.

3.5 Summary

In this chapter, a data-aided algorithm of synchronization and optimal integration region estimation has been proposed. The algorithm is based on the theoretical analysis of the optimal integration region that minimizes the probability of error. The proposed method employs a {0,1,0,1} pattern training sequence to first perform a frame-level coarse timing, then does a fine synchronization and finally estimates the optimal integration length by autocorrelation operations on consecutive symbol waveforms. To enhance the estimation accuracy, averaging over the training sequence has been applied for the purpose of denoising. Simulation results have shown that the proposed method can greatly reduce the bit error rate compared to the conventional fixed length method and achieves close-to-theoretical optimum performance.

Chapter 4

Time of Arrival Estimation Using Dual Pulse Signals

4.1 Background

Due to the ultra-wide bandwidth induced high timing resolution, UWB is considered as a promising candidate for accurate localization applications [15]. Among the localization methods, the time of arrival (TOA) approach is most suitable for use with UWB signals [15]. The TOA approach estimates the travel time of the first multipath component (or the direct path component) and thus acquires the distance between two nodes by multiplying the travel time by the speed of the light. The challenge to obtain a precise TOA estimation lies in detecting the first arriving pulse from a noise corrupted signal in a dense multipath environment, where the first arriving pulse may not come from the strongest path.

Various TOA estimation methods have been proposed in the literature [15, 35, 40–44]. In [15], the received signal is correlated with a previously constructed template on different intervals that have the same length as SR (assuming known) and

picks the maximum correlation output as the TOA. Another method that requires the knowledge of the SR length was proposed in [43], where a maximum energy detection (MED) method was proposed, which integrates the received signal energies within intervals that have the length of SR but start from different points, and picks the one with maximal energy as the SR. Both [15] and [43] have the drawback that their performances are closely related to the accuracy of the SR length estimation. Moreover, constructing the noisy template in [15] is also a challenge in practical implementation. In [35], a generalized maximum likelihood (GML) detection method was proposed. It assumes the position of the strongest tap is correctly acquired at first and then detects the taps before the strongest tap one by one until a threshold is reached and assigns the last detected tap to be the direct path component. Threshold crossing (TC) methods based on energy detection (ED) or correlation are proposed in [40–42]. These methods used a threshold on the received signal energy to distinguish between the SR and the noise-only region (NR) (the time interval during which only additive noise is received). In [40], a two-stage (namely the coarse timing phase and the fine timing phase) correlation based method was proposed for UWB timing acquisition, and the algorithm for finding the first arrival pulse in a received symbol is applicable to the TOA estimation problem. The coarse timing phase adopts a fixed threshold, and the fine timing phase uses an adaptive threshold based on the noise energy estimation obtained from the first phase. Guvenc *et. al.* proposed an adaptive threshold crossing method [41,42], whose threshold depends on the maximal and minimal values of the received signal energy. The approach given by [44] distinguishes the SR and the NR based on the difference in their statistical characteristics. It requires Nyquist rate sampling and a large number of samples to

get a reliable estimation of the statistical parameters of a random process.

In this chapter, a practical TOA estimation method based on the autocorrelation of a DP signal is proposed. In the proposed method, a training sequence of bipolar modulated DP symbols with alternating signs is transmitted and the received signal is correlated with its delayed version in the receiver. The autocorrelations are performed starting from different positions in a search region for a length equal to the symbol duration, and the outcomes drop rapidly when the autocorrelation start point encounters a multipath component. A threshold is then applied to detect the TOA by determining the position of the first significant descent of the autocorrelation results. The autocorrelations are averaged across several training symbols to reduce noise effect. This process is similar to the algorithm presented in Chapter 3 for estimating the integration start point of a TR autocorrelation receiver. However the objective of TOA estimation of the direct path is somewhat different from estimating the integration start point of a TR receiver which is determined from the perspective of integrated energy. Moreover, the method in Chapter 3 involved symbol duration long analog delay line. This drawback is avoided in the current chapter thanks to the adoption of the DP scheme.

The rest of this chapter is organized as follows. The proposed TOA estimation method is discussed in Section 4.2. Performance analysis of the developed TOA estimation method is given in Section 4.3. Section 4.4 presents simulation results and discussions, and Section 4.5 summarizes this chapter.

4.2 TOA Estimation Using Autocorrelation

The DP signal model used in this Chapter follows (2.20) with the number of frame $N_f = 1$. The received DP signal can be rewritten as

$$\begin{aligned} r_{DP}(t) &= \left(\sum_{l=1}^{L} \alpha_l s_{DP}(t - \tau_l) \right) * f(t) + n(t) \\ &= \sqrt{\frac{E_b}{2}} \sum_{i=-\infty}^{\infty} \sum_{l=1}^{L} \alpha_l \big[g_{DP}(t - iT_s - \tau_l) + g_{DP}(t - iT_s - \tau_l - T_p) \big] + n(t) \end{aligned} \tag{4.1}$$

where T_s denotes the symbol duration, $g_{DP}(t)$ is the composition of the transmitted pulse $p(t)$ and the receiver matching pulse $f(t)$, $n(t)$ is the additive white Gaussian noise with power spectral density of $N_0/2$ passed through the band-limited filter $f(t)$. The tapped delay line model described in (2.10) is used as channel model in (4.1). Particularly, τ_1, which is the time delay of the direct path, is the parameter of interest in the TOA estimation.

Let $\tilde{\tau}$ denote the initial time instant that the receiver locks to at the beginning of the communication. The arrival time of the direct path of the i^{th} symbol is $T_0 = \tau_1 + iT_s$, and the objective is to estimate T_0, given the arbitrary $\tilde{\tau}$ falls in the i^{th} received symbol or its vicinity. That is, $(i-1)T_s + \tau_L + T_p < \tilde{\tau} < (i+1)T_s + \tau_1$. Let the i^{th} transmitted data be "+1" while the $(i+1)^{\text{th}}$ transmitted data be "-1". The following integration

$$x(\tilde{\tau}) = \int_{\tilde{\tau}}^{\tilde{\tau}+T_s} \Re\big\{ r_{DP}(t) r_{DP}^*(t + T_d) \big\} \mathrm{d}t \tag{4.2}$$

can provide an estimation of where $\tilde{\tau}$ situates relative to the i^{th} symbol. Referring to

Fig. 2.5, when $\tilde{\tau}$ is within the noise region of the $(i-1)^{\text{th}}$ symbol, i.e., $\tilde{\tau} \in [(i-1)T_s + \tau_L + T_p, T_0]$, the value of $x(\tilde{\tau})$ is $E_b/2$ plus the noise and IPI. When $\tilde{\tau}$ is within the noise region of the i^{th} symbol, i.e., $\tilde{\tau} \in [iT_s + \tau_L + T_p, (i+1)T_s + \tau_1]$, the noiseless and IPI free part of $x(\tilde{\tau})$ is $-E_b/2$ since all of the signal region of the $(i+1)^{\text{th}}$ symbol is included in the integration (4.2). In between the above two intervals, i.e., when $\tilde{\tau} \in [T_0, iT_s + \tau_L + T_p]$, the noiseless and IPI free part of $x(\tilde{\tau})$ is in between $E_b/2$ and $-E_b/2$ due to the fact that integrals over the sub-intervals $\int_{\tilde{\tau}}^{iT_s+\tau_L+T_p} \Re\{r_{DP}(t) r_{DP}^*(t+T_d)\} \mathrm{d}t$ and $\int_{(i+1)T_s+\tau_1}^{\tilde{\tau}+T_s} \Re\{r_{DP}(t) r_{DP}^*(t+T_d)\} \mathrm{d}t$ have opposite signs and thus will partially cancel each other. Therefore, let τ be a variable that varies in the range $[\tilde{\tau}, \tilde{\tau} + T_s]$ and define

$$x(\tau) = \int_{\tau}^{\tau+T_s} \Re\{r_{DP}(t) r_{DP}^*(t+T_d)\} \mathrm{d}t, \quad \tau \in [\tilde{\tau}, \tilde{\tau} + T_s]. \tag{4.3}$$

If $\tilde{\tau} \in [(i-1)T_s + \tau_L + T_p, T_0]$, T_0 can be determined as the time when $x(\tau)$, ignoring the noise term, starts to decrease from its maximum, i.e.,

$$\hat{T}_0 = \max\{\tau | x(\tau) = \frac{E_b}{2} - E_{IPI}, \tilde{\tau} < \tau < \tilde{\tau} + T_s\} \tag{4.4}$$

where E_{IPI} is the signal energy loss due to IPI. If $\tilde{\tau} \in [T_0, (i+1)T_s + \tau_1]$, T_0 can be determined as the time when $x(\tau)$ starts to increase from its minimum, i.e.,

$$\hat{T}_0 = \max\{\tau | x(\tau) = -\frac{E_b}{2} + E_{IPI}, \tilde{\tau} < \tau < \tilde{\tau} + T_s\}. \tag{4.5}$$

Taking the noise and IPI into account, the curve of $x(\tau)$ contains fluctuation. Thus the exact first falling edge of $x(\tau)$ is obscured, as shown in Fig. 4.1. To combat

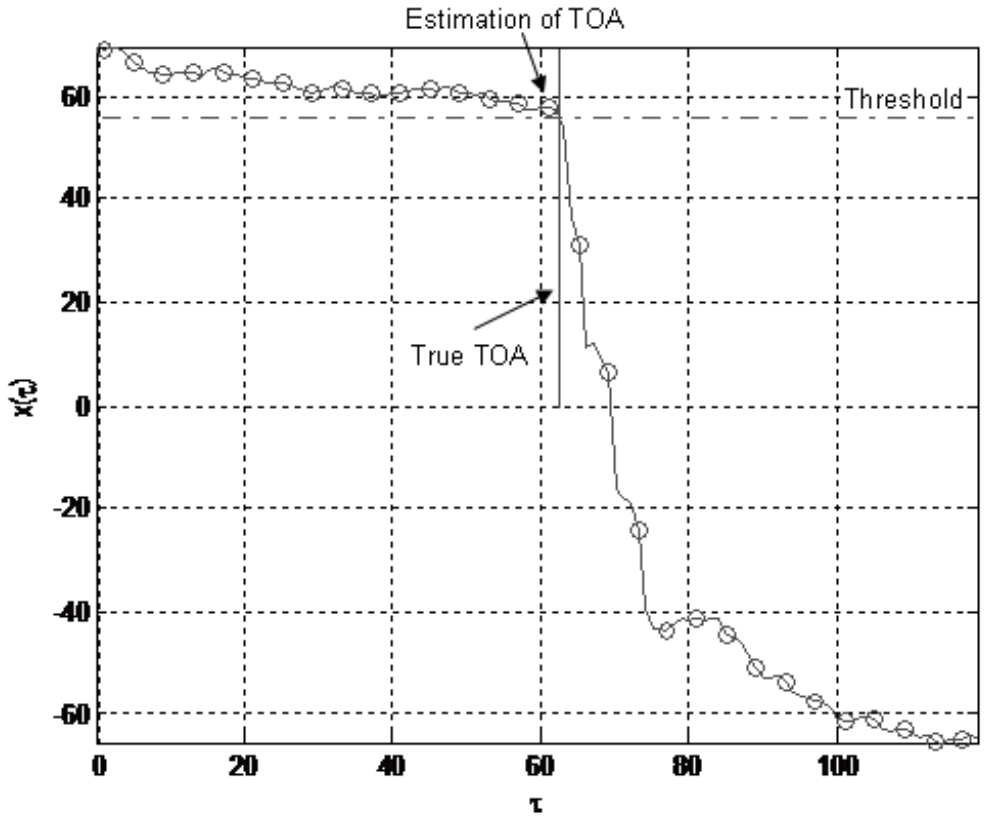

Figure 4.1: The noisy $x(\tau)$ curve with an example threshold crossing of $\eta = 0.9$ and $E_b/N_0 = 16$ dB.

noise and IPI, three steps are devised to form a complete algorithm based on the autocorrelation (4.3) and the edge detection (4.4) or (4.5).

First, the autocorrelation (4.3) requires the adjacent data symbols to have opposite signs. Thus a training sequence comprised of alternating "-1"s and "+1"s is transmitted, i.e.,

$$b_i = (-1)^{i-1}, \quad l = 0, \cdots, N-1 \tag{4.6}$$

where N is the length of the training sequence.

Second, the edge detection (4.4) requires an exhaustive search of τ's in the entire

T_s-long uncertainty region. For practical implementation, the integration region in (4.3) is divided into $Q = \lfloor \frac{T_s}{T_b} \rfloor$ sections where T_b is the desired timing resolution. Then the integration results over sections are averaged over $N-1$ training symbols through digital processing and then summed up with different starting points. The procedure can be written as

$$\bar{x}_k = \frac{1}{N-1} \sum_{m=k}^{k+Q-1} \sum_{i=0}^{N-2} (-1)^{i-1} \int_{\tilde{\tau}+mT_b+iT_s}^{\tilde{\tau}+(m+1)T_b+iT_s} \Re\Big\{ r_{DP}(t) r_{DP}^*(t+T_d) \Big\} \mathrm{d}t, \;\; k = 0, \cdots, Q-1 \tag{4.7}$$

where $\bar{x}_k$ can be viewed as the sampled sequence of $x(\tau)$ using sampling rate $1/T_b$, yet the noise power in $\bar{x}_k$'s is greatly reduced through averaging.

Third, a threshold is used to determine the TOA of the direct path from $x(\tau)$. This is done by choosing the time instant corresponding to the last sample of $\bar{x}_k$'s that is greater than a threshold ξ as the TOA. The TOA estimation can be written as

$$\hat{T}_0 = \tilde{\tau} + T_b \cdot \max\big(\{k | \bar{x}_k > \xi, \; k = 0, \cdots, Q-1\}\big). \tag{4.8}$$

The threshold crossing process is depicted in Fig. 4.1. Instead of choosing a fixed threshold ξ, we set the threshold to be proportional to the maximum value of $\bar{x}_k$. That is,

$$\xi = \eta \cdot \bar{x}_m \tag{4.9}$$

where $\bar{x}_m = \max\{\bar{x}_k, \; k = 0, \cdots, Q-1\}$ and $\eta \in (0, 1)$ is termed as the normalized threshold.

Note that here we assume the initial time $\tilde{\tau}$ is within one symbol duration ahead of the TOA of the first training symbol. If this assumption does not hold, i.e., $\tilde{\tau}$

is behind the first training symbol, more training symbols are needed so that the training sequence can cover the region $[\tilde{\tau}, \tilde{\tau} + NT_s]$. Moreover, if $\tilde{\tau}$ falls in the SR of a "+1" training symbol, $x(\tau)$ would start somewhere between max and min values, keep decreasing to the minimum, stay around minimum values and eventually start to increase to the maximum. If $\tilde{\tau}$ falls in the NR of a "+1" symbol, $x(\tau)$ would increase from its minimum to its maximum. In both cases, it is the rising edge that tells the TOA of the next "-1" training symbol as given by (4.5). The signs of all $\bar{x}_k$'s should be flipped so that (4.8) still applies. In our simulation, if $\bar{x}_1 < 0$ or $\frac{1}{i_{max}} \sum_{k=1}^{i_{max}} \bar{x}_k$ is smaller than $-\frac{\xi}{Q-i_{min}} \sum_{k=i_{min}}^{Q-1} \bar{x}_k$, where $\bar{x}_{i_{max}}$ and $\bar{x}_{i_{min}}$ are the maximum and minimum of $\bar{x}_k$'s with index i_{max} and i_{min}, respectively, the signs of all $\bar{x}_k$'s are flipped. Although this sign decision scheme does not guarantee 100% correctness, it only adds negligible extra error to the mean absolute error. When the SNR is greater than 16 dB, the error rate occurred at this flipping step is found to be less than 0.1% and when the SNR is lower, a flipping error introduces an equivalent early false alarm error while the error rate of the former is far less than the latter.

4.3 Analysis on P_{fa}, P_m and MAE of the Proposed TOA Method

Define the probability of early false alarm P_{fa} as the probability that the estimated TOA is in front of the true TOA [35], i.e., $\hat{T}_0 < T_0 - T_b$. Suppose the true TOA T_0 falls in the section $[jT_b + \tilde{\tau}, (j+1)T_b + \tilde{\tau}]$, i.e., j satisfies $jT_b + \tilde{\tau} < T_0 < (j+1)T_b + \tilde{\tau}$. There is

$$P_{fa} = \Pr\left\{\bar{x}_k < \xi, \text{ for } k = j, j+1, \cdots, Q-1\right\} \approx \Pr\{\bar{x}_j < \xi\}. \tag{4.10}$$

The approximation is justified based on the reason that for a given channel realization, all $\bar{x}_k$'s for $k = j, \cdots, Q-1$ are highly correlated Gaussian random variables with monotonically decreasing mean values and the same variance $\sigma_{\bar{x}}^2 \approx \frac{N_0 E_b + N_0^2 B T_s}{2(N-1)}$, where B is the bandwidth of the shaping pulse. Next, we consider a fixed threshold ξ to facilitate analysis tractability. Since $\bar{x}_j \sim \mathcal{N}(\frac{E_b}{2} - E_{IPI}, \sigma_{\bar{x}}^2)$, we have

$$P_{fa} \approx Q\left(\frac{E_b/2 - E_{IPI} - \xi}{\sigma_{\bar{x}}}\right) \tag{4.11}$$

where the Q-function is defined as $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathrm{d}t$.

Similarly, let P_m denote the probability of missed-direct-path error which is the probability that the estimated TOA is behind the true TOA [35], i.e., $\hat{T}_0 > T_0 + T_b$. We have

$$P_m = 1 - \Pr\left\{\bar{x}_k < \xi, \text{ for } k = j+1, \cdots, Q-1\right\} > \Pr\{\bar{x}_{j+1} > \xi\}. \tag{4.12}$$

It can be shown that $\bar{x}_{j+1}$ can be written as $\bar{x}_{j+1} = \frac{E_b}{2} - E_{IPI} - E_b|\alpha_0|^2 - n_{j+1}$ where the noise term n_{j+1} is a zero mean Gaussian random variable with variance $\sigma_{\bar{x}}^2$ and $|\alpha_0|^2$ is the energy of the first multipath tap which follows Gamma distribution (since α_0 follows Nakagami distribution [26]) given by $f_{\alpha^2}(x) = \frac{x^{m-1}}{\Gamma(m)} (\frac{m}{\Omega})^m \exp(-\frac{mx}{\Omega})$. Therefore,

$$\begin{aligned} P_m &> \int_{-\infty}^{\frac{E_b}{2} - E_{IPI} - \xi} \int_0^{\frac{1}{2} - \frac{E_{IPI} + \xi + t}{E_b}} f_{\alpha^2}(x) f_{n_{j+1}}(t) \mathrm{d}x \mathrm{d}t \\ &= \int_{-\infty}^{\frac{E_b}{2} - E_{IPI} - \xi} \frac{\gamma\big(m, \frac{m}{\Omega}(\frac{1}{2} - \frac{E_{IPI} + \xi + t}{E_b})\big)}{\Gamma(m)} \frac{e^{-t^2/2\sigma_{\bar{x}}^2}}{\sqrt{2\pi}\sigma_{\bar{x}}} \mathrm{d}t \end{aligned} \tag{4.13}$$

where $\gamma(a, x) = \int_0^x t^{a-1} e^{-t} \mathrm{d}t$ is the lower incomplete gamma function. In the high SNR environment, where $n_{j+1} \ll |\alpha_0|^2$, the pdf of n_{j+1} can be approximated as

$f_{n_{j+1}}(t) \approx \delta(t)$. Thus $P_m > \gamma\big(m, \frac{m}{\Omega}(\frac{1}{2} - \frac{E_{IPI}+\xi}{E_b})\big)/\Gamma(m)$. This shows that when the SNR is high, P_m depends only on the choice of threshold, but is independent of SNR.

Fig. 4.2 plots the analytical and simulation results of P_{fa} and P_m. In [26], the parameters m and Ω are both defined as log-normal random variables. In our analytical evaluation, m and Ω are determined as $m = 1.2$ and $\Omega = 0.4$ by curve fitting the distribution of the first taps of 1000 channel realizations in CM1 channels. Moreover, note that E_{IPI} varies from channel to channel, and for a same channel realization, the "+1" DP symbol leads to different IPI energy from the "-1" DP symbol due to their asymmetrical structures. Therefore, we calculate the "+1" symbol E_{IPI} and "-1" symbol E_{IPI} for each channel realization and insert the mean value into (4.11) and (4.13). The probabilities of error are then obtained by averaging over 1000 channel realizations. It is observed from Fig. 4.2 that the approximate analysis for P_{fa} and the lower bound of P_m agree with simulation. In addition, a larger threshold ξ results in a larger P_{fa} but a smaller P_m.

Furthermore, the mean absolute error (MAE), which is defined as the ensemble average of the absolute estimation errors over all channel realizations, can be written as

$$\text{MAE} = \sum_{i=0}^{Q-1} |i-j| T_b \Pr\left\{\bar{x}_i > \eta \cdot \bar{x}_m, \bar{x}_k < \eta \cdot \bar{x}_m, \text{ for } k = i+1, \cdots, Q-1\right\}. \quad (4.14)$$

Unfortunately, we find that for the proposed TOA estimation algorithm, the theoretical formula of MAE is very difficult to derive even for a simple AWGN channel

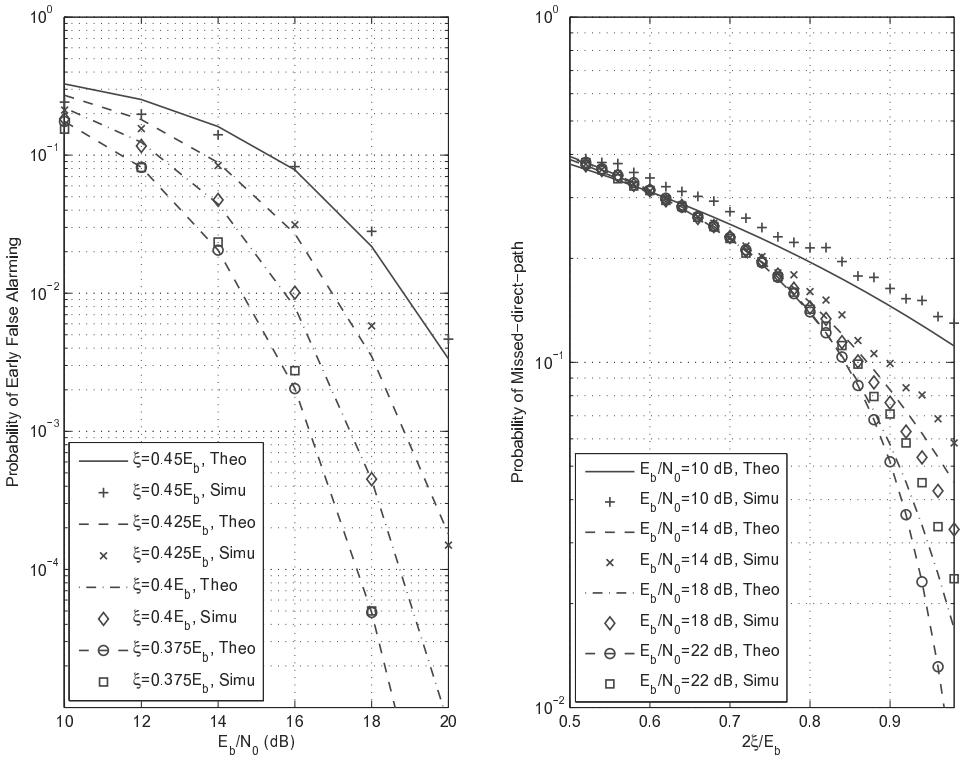

Figure 4.2: The probability of early false alarm and probability of missed-direct-path errors of the proposed TOA estimation method in CM1 channels with fixed thresholds, $N = 32$, and $T_b = 1$ ns.

and fixed threshold. Denote

$$y_j = \int_{\tilde{\tau}+jT_b}^{\tilde{\tau}+(j+1)T_b} \Re\{r_{DP}(t) r_{DP}^*(t+T_d)\} \mathrm{d}t \quad \text{for } j = 0, 1, \cdots, 2Q-2 \tag{4.15}$$

and

$$x_k = \sum_{j=k}^{k+Q-1} y_j \quad \text{for } k = 0, \cdots, Q-1 \tag{4.16}$$

where x_k is the sample of $x(\tau)$ at $\tau = \tilde{\tau} + kT_b$. In an AWGN channel, y_j's are approximately independent non-identical distribution (i.n.d.) random variables. However, x_k's are highly correlated random variables because x_k and x_{k+1} both have the common term $\sum_{j=k+1}^{k+Q-1} y_j$. Assume the covariance matrix of x_k's is calculable and denoted as $\mathbf{M}$, their joint pdf can be written as

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{Q/2}\sqrt{|\mathbf{M}|}} \exp\big[-\frac{1}{2}(\mathbf{x}-\mathbf{a})^{\mathcal{T}} \mathbf{M}^{-1} (\mathbf{x}-\mathbf{a})\big] \tag{4.17}$$

where $\mathbf{x} = [x_0, \cdots, x_{Q-1}]^{\mathcal{T}}$, $\mathbf{a}$ is the mean of $\mathbf{x}$ and $^{\mathcal{T}}$ is the matrix transpose. Thus MAE can be calculated as $\text{MAE} = \sum_{i=0}^{Q-1} |i-j| T_b \Pr\big\{\bar{x}_i > \xi, \bar{x}_k < \xi, \text{ for } k = i+1, \cdots, Q-1\big\}$, where

$$\Pr\big\{\bar{x}_i > \xi, \bar{x}_k < \xi, \text{ for } k = i+1, \cdots, Q-1\big\} = \int_{\xi}^{\infty} \int_{-\infty}^{\xi} \cdots \int_{-\infty}^{\xi} f(\mathbf{x}) \mathrm{d}x_{Q-1} \cdots \mathrm{d}x_i. \tag{4.18}$$

Eq. (4.18) involves $(Q-i)$-tuple integral and cannot be decomposed into multiplication of a series of single integrals, thus it is very difficult to derive the closed form of this probability. In the next section, we study the effect of normalized threshold η on MAE via simulations.

4.4 Simulation Results

In this section, we present the simulation results of the proposed algorithm. The performance is evaluated based on three criteria: MAE, P_{fa} and P_m. Simulations are carried out on each of the 1000 channel realizations of the IEEE 802.15.4a channel model 1 (CM1) for indoor residential line-of-sight environments and the channel model 4 (CM4) for indoor office NLOS environments as described in [26]. The symbol duration of the dual pulse signal is $T_s = 125$ ns. The signal bandwidth is $B = 494$ MHz and the shaping pulse $p(t)$ is a root raised cosine pulse with roll-off factor $\beta = 0.25$. The threshold (4.9) is used in all the simulation.

Figs. 4.3–4.4 depict the estimation accuracy as a function of the normalized threshold η. Similar to using a fixed threshold, a larger η results in a larger P_{fa} but a smaller P_m. This is because a larger η leads to larger $\Pr\left\{\bar{x}_k < \eta \cdot \bar{x}_m, \text{ for all } k\right\}$. Moreover, Fig. 4.3 shows that P_{fa} varies with the SNR while P_m is almost independent of the SNR. This is because when the noise power is high, $x(\tau)$ has larger fluctuation in the region $[\tilde{\tau}, T_0]$, which often leads to a larger $\bar{x}_m$. Consequently the threshold ξ is more likely to be higher than $\bar{x}_j$, where $[jT_b + \tilde{\tau}, (j+1)T_b + \tilde{\tau}]$ is the section where the true TOA lies, resulting in a larger P_{fa}. On the other hand, the probability of missed-direct-path error mainly depends on the impulse response of the channel realization, or more exactly, the magnitude of the first multipath tap relative to the threshold. A sufficiently strong first tap can provide enough descent from $\bar{x}_j$ to $\bar{x}_{j+1}$ so that all $\bar{x}_k$'s for $k > j$ are smaller than the threshold $\eta \cdot \bar{x}_m$. Comparing Fig. 4.4 with Fig. 4.3, it is also found that at lower SNRs, the MAE follows the trend of P_{fa}, whereas P_m determines the MAE floor at high SNRs.

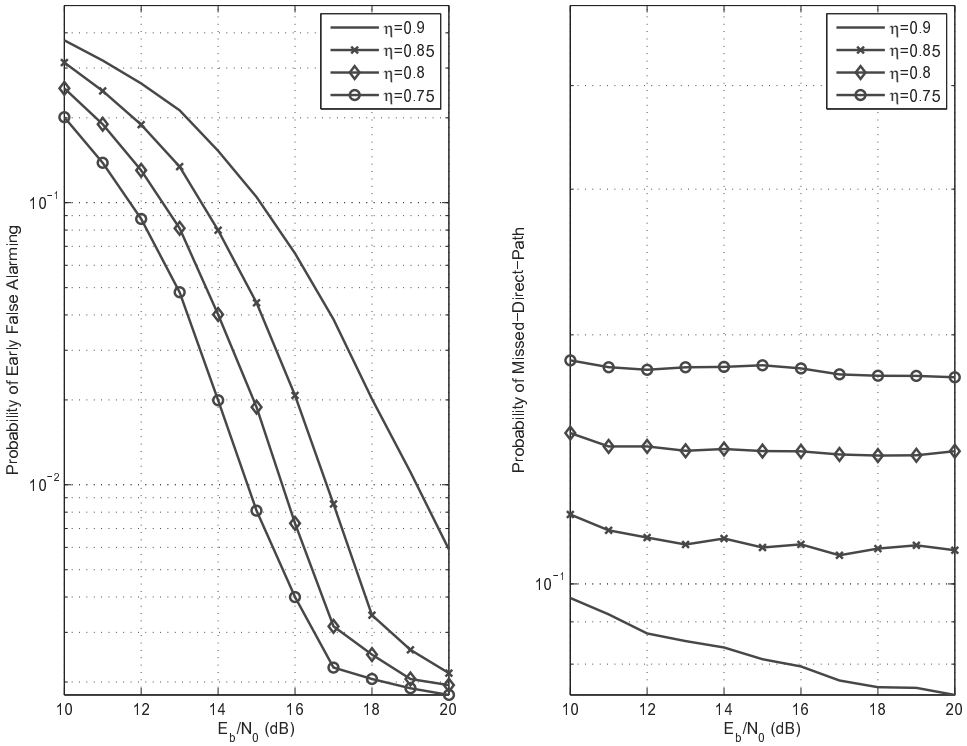

Figure 4.3: The probability of early false alarm and probability of missed-direct-path errors with normalized threshold η, $N = 32$ and $T_b = 1$ ns.

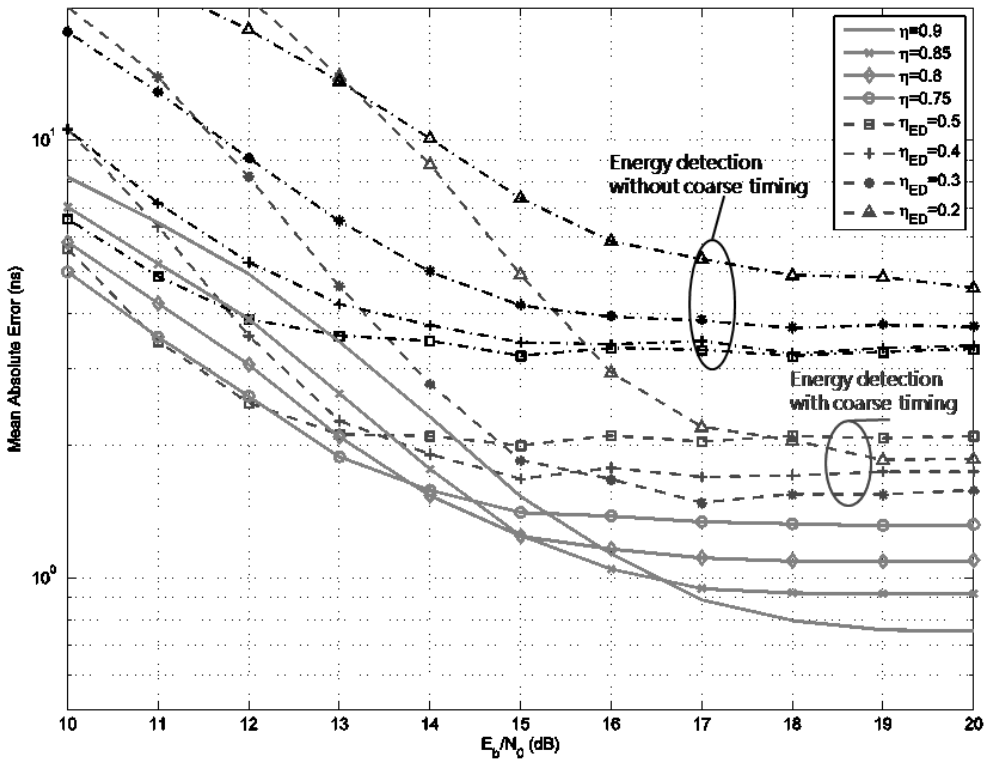

Figure 4.4: The effect of different values of normalized threshold η on the mean absolute error, with $N = 32$ and $T_b = 1$ ns. Solid lines represent the DP based estimation method, dashed lines represent the ED based estimation method with perfect coarse timing, and dashed-dotted lines represent the ED based method without coarse timing.

The DP autocorrelation based TOA estimation method is compared with the energy detection (ED) based TOA estimation [41] in Fig. 4.4. In the simulation, the transmitted signal for the ED method contains a *single* BPSK modulated pulse in each symbol duration, which has higher peak power than the DP structure. The ED receiver integrates the energy of the received signal over each T_b interval, and determines the TOA as the first interval whose energy integration $z_k = \int_{\tilde{\tau}+kT_b}^{\tilde{\tau}+(k+1)T_b} |r(t)|^2 \mathrm{d}t$ is greater than a threshold. The threshold ξ_{ED} for the energy detection method is chosen as [41]

$$\xi_{ED} = \eta_{ED}\big(\max\{z_k\} - \min\{z_k\}\big) + \min\{z_k\} \tag{4.19}$$

where $\eta_{ED} \in (0, 1)$ is the normalized threshold for the ED method. The same N bit training sequence as (4.6) is used and the z_k's are averaged over the training symbols to increase the SNR. For the ED method, a coarse timing step is required to make sure the initial point $\tilde{\tau}$ falls in the interval $[T_0 + \tau_L - T_s, T_0]$. Otherwise, averaging over multiple training symbols starting from the position $\tilde{\tau}$ will divide a continuous signal region into two parts. In this case, it is difficult to distinguish whether the first pulse crossing the threshold is the earliest channel tap in the current symbol interval or a delayed pulse of the previous symbol. The DP method does not have this problem and therefore does not require coarse timing. In Fig. 4.4, both perfect coarse timing and no coarse timing are tested for ED based TOA estimation. Simulation results show that the ED method in general has worse performance than the DP autocorrelation method. Especially, in the high SNR region, a proper η for the DP autocorrelation method leads to a significantly lower MAE floor (about 0.7 ns) whereas for the ED based method with perfect coarse timing the lowest MAE floor is around 1.5 ns. To explain this, it is noticed that the DP method chooses the

last threshold crossing point as the estimated TOA whereas the ED method picks the first one. For a threshold that leads to a low P_m, e.g., for $\eta = 0.9$ and $\eta_{ED} = 0.2$, the DP mechanism is less sensitive to the noise region fluctuation in $\bar{x}_k$'s, while for the ED method the first z_k greater than the threshold in the noise region tends to cause a large early false alarm error. Besides this, the descent of $\bar{x}_k$'s in the DP method caused by a multipath component is about twice as much as the energy collected in the corresponding interval of the ED method.

A threshold analysis can be drawn from the curves in Fig. 4.5. In the figure the MAE with respect to a wide range of normalized threshold values is shown for different SNR values in CM1 channels. Given a certain SNR, an optimal normalized threshold can be picked. Generally, the optimal threshold value goes from about 0.5 to about 0.9 as the SNR varies from 10 dB to 24 dB. The MAE is more sensitive to the choice of threshold in high SNR environments.

The MAE results of using a second order Gaussian pulse are shown in Fig. 4.6. The second order Gaussian pulse is generated following (2.2). Comparing it with Fig. 4.4, the results using second order Gaussian pulse are slightly worse than those using root raised cosine pulse, while they show similar features in terms of MAE floors and the advantage of the proposed method over the ED based method. This shows that the estimation performances change just slightly for different pulse shapes as long as the bandwidth of the pulses are similar.

Fig. 4.7 depicts the impact of different search resolution T_b on the MAE. As shown in the figure, the smaller T_b is, the better estimation performance will be. When $\eta = 0.9$, the mean absolute error approaches $T_b/2$ in the high SNR region, which is the theoretical optimal result.

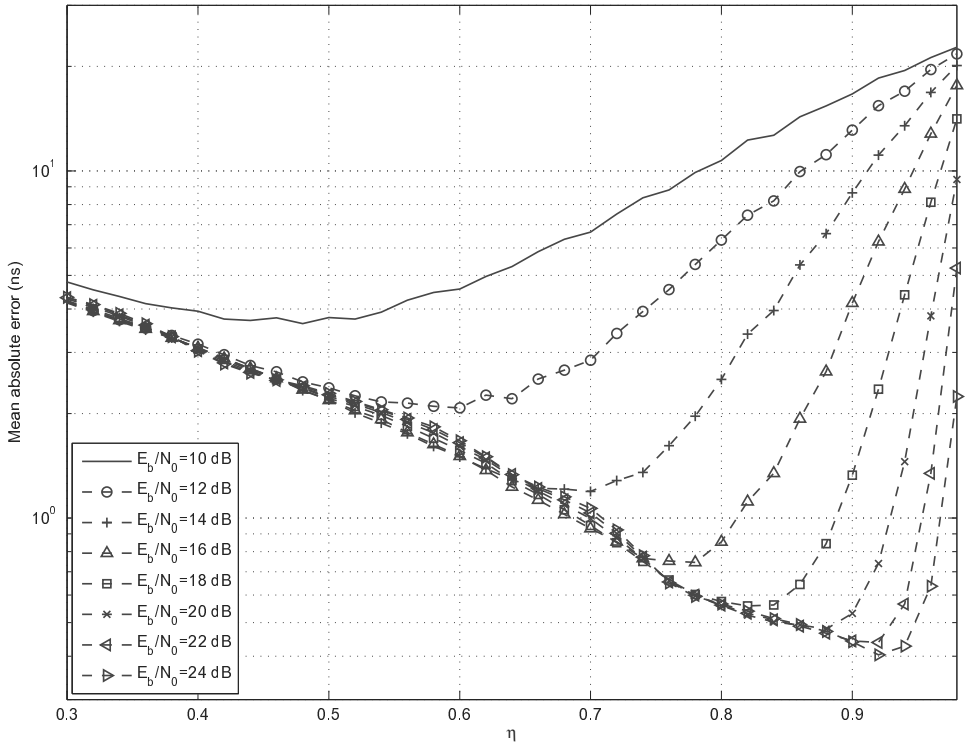

Figure 4.5: MAE versus normalized threshold for different SNR values in CM1 channels with $N = 32$ and $T_b = 1$ ns.

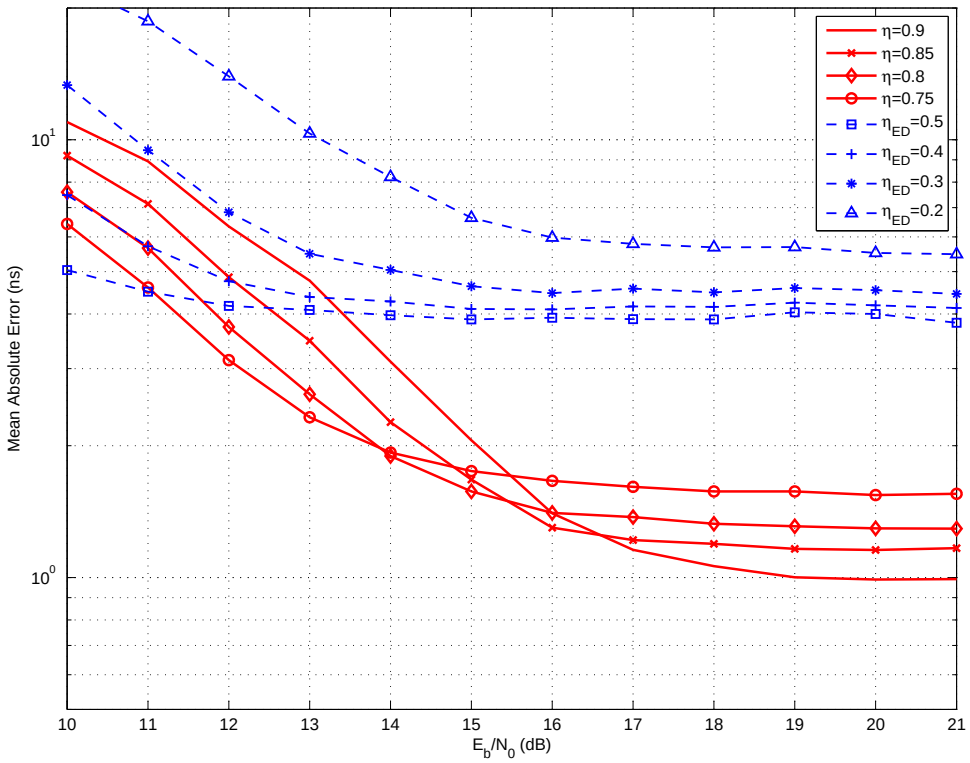

Figure 4.6: The mean absolute error of using a 2nd order Gaussian pulse with different values of the normalized threshold, $N = 32$ and $T_b = 1$ ns. The solid lines represent DP based estimation method, and the dashed line represent the ED based estimation method without coarse timing.

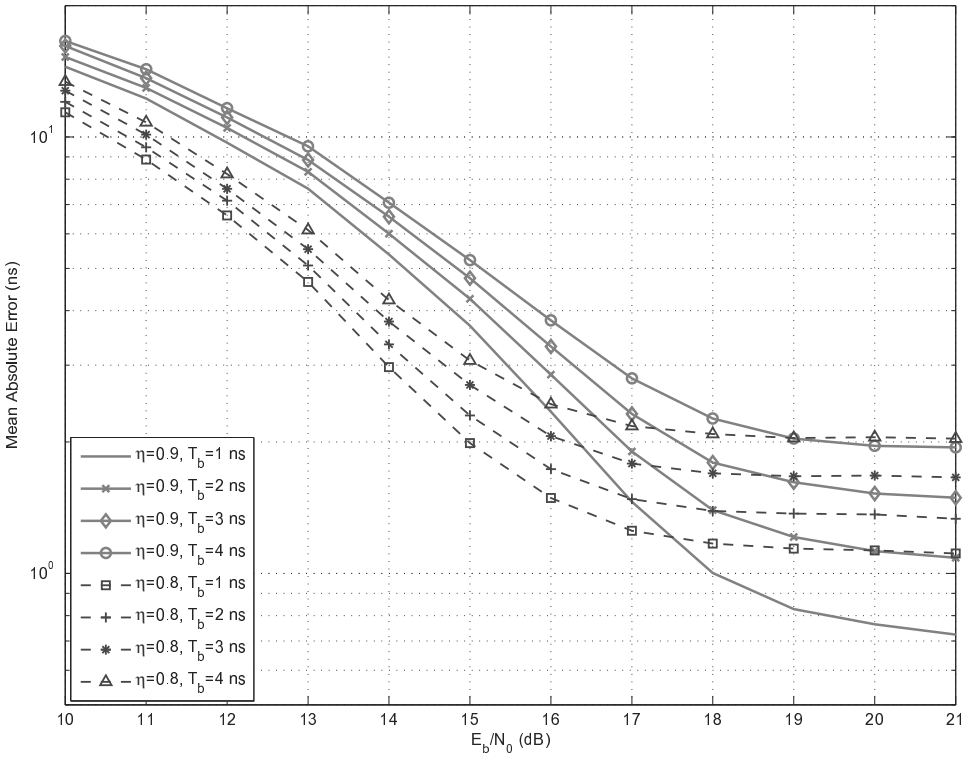

Figure 4.7: The mean absolute errors with different search stepsize T_b in CM1 channels, and $N = 16$.

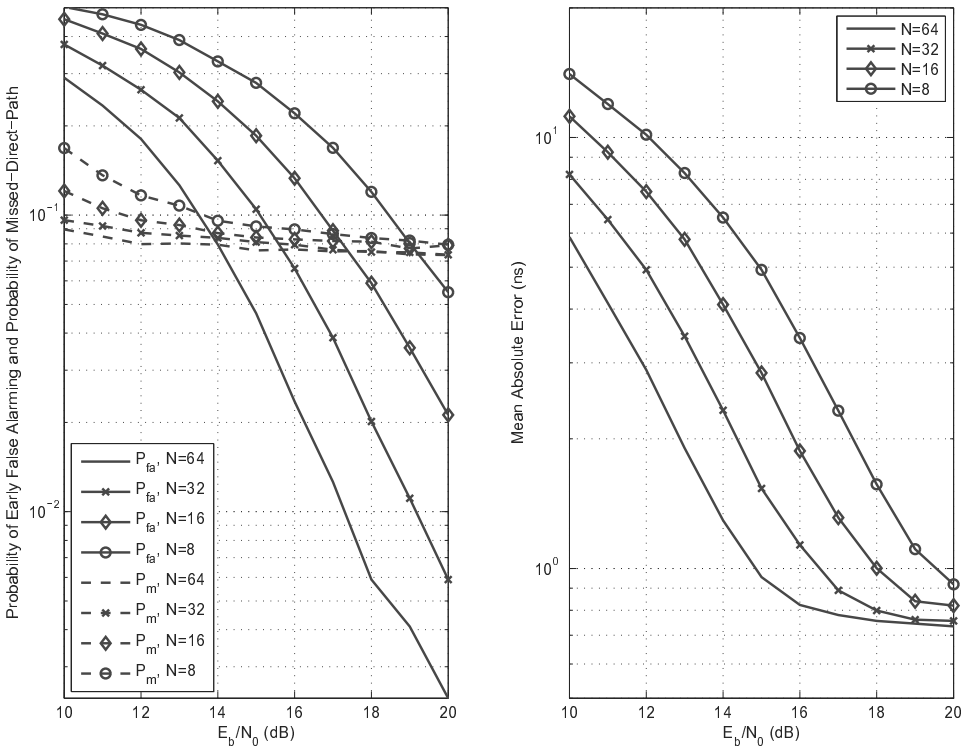

Figure 4.8: The effect of different training sequence length N on the DP based TOA estimation performance, with $\eta = 0.9$ and $T_b = 1$ ns.

The effect of training sequence length N on the estimation performance is shown in Fig. 4.8. A longer training sequence leads to lower noise levels in $\bar{x}_k$'s. Therefore the larger N is, the smaller P_{fa} will be. However, it is also shown in Fig. 4.8 that P_m just slightly changes when increasing N due to its insensitivity to SNR. Similar to Fig. 4.4, the MAE is closely related to P_{fa} at lower SNRs. But since all N's yield similar P_m, the MAE curves tend to converge to an approximately same error floor at higher SNRs.

In the non-line-of-sight channels where the first channel tap is severely attenuated,

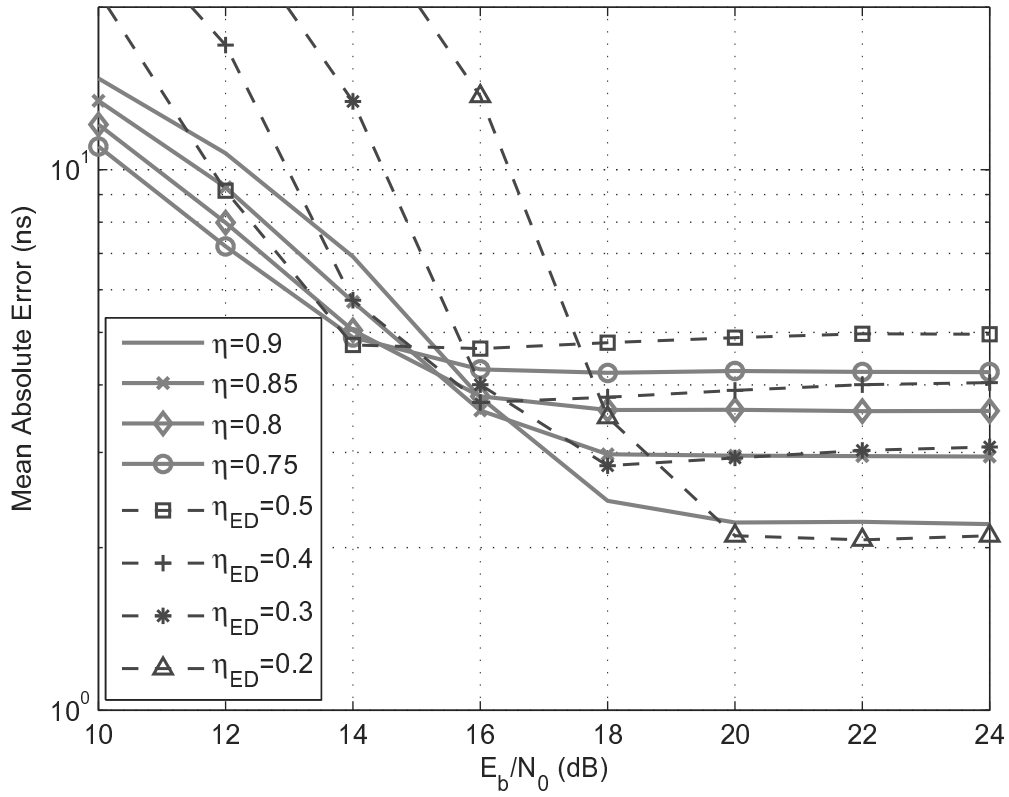

Figure 4.9: The mean absolute errors of the proposed TOA estimation method in CM4 channels with $N = 32$ and $T_b = 1$ ns.

often the arrival time of the first significant channel tap, instead of the very first tap, is located. The performance of the proposed TOA estimation algorithm and the ED method in CM4 non-line-of-sight channels are presented in Fig. 4.9. It is shown that the proposed DP based method outperforms the ED based method at low SNRs, but has a similar MAE floor at high SNRs in CM4 channels.

4.5 Summary

A dual pulse autocorrelation based TOA estimation method has been proposed in this chapter. This method results in more accurate TOA estimates (in terms of lower MAE) than the existing ED based method and does not require the additional coarse timing step in the ED approach. It has also been shown that for the proposed method, the probability of early false alarm is sensitive to the SNR, the probability of missed-direct-path error closely relates to the channel condition, and MAE depends on the combined effects of the two.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Timing and synchronization play a pivotal role in the digital communication world. Due to the ultra-short pulse width and rich multipath environment, timing and synchronization for ultra-wideband systems encounter particular difficulties. In this thesis an autocorrelation based timing and synchronization method is designed for low data rate UWB systems. The proposed timing method is able to provide an estimate of the optimal integration region for the BPPM non-coherent detection system and the TR system. When applied to the DP scheme, the proposed timing method can also provide an accurate time-of-arrival estimation.

In Chapter 3, the optimal integration region estimation algorithm is described. It is based on inter-symbol correlation on the received signal of a {0,1,0,1} pattern training sequence. The algorithm consists of three consecutive steps: coarse synchronization which locates the first half of a symbol's first frame; estimation of the integration start point; and estimation of the optimal integration length. The latter two steps are based on the theoretical derivation of the bit error rate of the BPPM

and TR systems. Significant performance improvement has been observed for the systems using the proposed method against the conventional fixed integration region method. The estimation accuracy is related to the length of the training sequence and the search stepsize. It has been found that with several tens of training symbols, the proposed method is able to closely approach the theoretical optimal error rate performance. Although this data-aided timing method uses a TR signal as training symbols, it is possible to apply the algorithm onto both TR and BPPM systems, because in a given channel condition the BPPM non-coherent detection system and the TR autocorrelation system have the same error rate expression.

The autocorrelation based time-of-arrival estimation method has been studied in Chapter 4. Again, the {0,1,0,1} pattern training sequence is applied with the dual pulse signal scheme. After the autocorrelation, a normalized threshold is used to find the arrival time of the first significant drop in the correlation result. The performance of the new estimation method has been analyzed theoretically and through simulations. It has smaller mean absolute error than the energy detection method. Moreover, when using single frame signal structure, the proposed method does not require any coarse timing, which is a critical step for the energy detection method.

Besides their advantage in performance, another important merit of the timing method developed in this thesis is its applicability. First, the signal structure used in the timing method fully complies with the IEEE 802.15.4a standard. Second, the method treats the channel response as a whole and avoids the complexity of searching for individual channel taps. Third, the method requires frame or symbol rate sampling rather than Nyquist rate sampling, thus alleviates the receiver complexity.

5.2 Future Work

Nonetheless, several critical issues and more in-depth research work should be further dealt with before the implementation of the proposed timing method. These issues, as listed below, will be the topics of future study.

1. For the integration region optimization algorithm, a $T_f/2$ long analog delay line is required. As widely known, manufacturing of the accurate wideband delay line is very difficult. It is necessary to find a way to remove this requirement from the algorithm. One method of removing the long delay line requirement is changing the transmitted signal pattern, as is in the DP scheme which uses a short T_d. On the other hand, if the DP scheme is used for data transmission, the integration region of its autocorrelator should also be estimated and optimized. It is worthwhile to investigate the relationship between the error rate and integration region for DP scheme, which is more complicated than the conventional TR system since it involves certain amount of IPI.

2. The repetition of the training symbols for both the integration region optimization algorithm and the TOA estimation algorithm may induce spectral spikes which violate the FCC mask. As mentioned in Chapter 2, one solution is to spread the energy into a cluster of PN sequence modulated pulses. Other pulse cluster structures may also do the work. The impact of using the pulse cluster signal structure needs to be studied in detail in the future.

3. As mentioned in Chapter 2, time hopping is a widely used method to accommodate multiple users. It can also alleviate the magnitude of spectral spikes.

In this thesis, the proposed timing method deals with single user only. Further study could focus on modifying the method to be compatible with the UWB system with time hopping.

4. The TOA estimation method presented in Chapter 4 provides accurate ranging capability. With a number of nodes that are able to measure the distance between each other, an accurate sensor network can be formed. Some algorithms could be developed to coordinate the ranging among these nodes so as to calculate the location of some desired node. Multi-user detection and ad-hoc network technologies are involved in this problem.

Bibliography

[1] L. Yang and G. Giannakis, "Ultra-wideband communications: an idea whose time has come," *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 26–54, 11.

[2] "First report and order: in the matter of revision of part 15 of the commission's rules regarding ultra-wideband transmission systems," FCC, Tech. Rep. 02-48, April 2002.

[3] "Part 15.4: wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (WPANs), amendment 1: add alternate phys," IEEE std, Tech. Rep. 802.15.4a-2007, Aug. 2007.

[4] T. W. Barett, "History of ultra wideband (UWB) radar and communications: pioneers and innovators," in *Proc. Progress in Electromagnetics Symposium*, 2000.

[5] J. G. Proakis, *Digital Communication*, 4th ed. New York: Prentice Hall, 2001.

[6] "DS-UWB physical layer submission to 802.15 task group 3a," IEEE P802.15.3a Working Group, Tech. Rep. P802.15-04/0137r3, July 2004.

[7] "Multi-band OFDM physical layer proposal for IEEE 802.15 task group 3a," IEEE P802.15.3a Working Group, Tech. Rep. P802.15-03/0268r2, Nov. 2003.

[8] Y. Wang, "Single carrier block transmission in ultra-wideband communications," Ph.D. dissertation, University of Victoria, 2006.

[9] M. Z. Win and R. A. Scholtz, "Impulse radio: how it works," *IEEE Trans. Communications*, vol. 53, no. 8, pp. 1269–1273, Aug. 2005.

[10] H. Zhang, W. Li, and T. A. Gulliver, "Pulse position amplitude modulation for time-hopping multiple-access UWB communications," *IEEE Communications Letters*, vol. 2, no. 2, pp. 36–38, Feb. 1998.

[11] R. Hoctor and H. Tomlinson, "Delay-hopped transmitted-reference RF communications," in *Proc. IEEE Conference on Ultra-wideband System and Technology (UWBST)*, pp. 265–269, May 2002.

[12] X. Dong, A. Lee, and L. Xiao, "A new UWB dual pulse transmission and detection technique," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 4, pp. 2835–2839, May 2005.

[13] J.-Y. Lee, "Ultra-wideband ranging in dense multipath environments," Ph.D. dissertation, University of Southen California, 2002.

[14] H. Zhang, T. Udagawa, T. Arita, and M. Nakagawa, "Home entertainment network: combination of IEEE 1394 and ultra-wideband solutions," in *Proc. IEEE Conference on Ultra-wideband System and Technology (UWBST)*, pp. 144–145, May 2002.

[15] S. Gezici and et. al., "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Processing Magazine*, vol. 11, no. 4, pp. 70–84, July 2005.

[16] M. Win and R. Scholtz, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications," *IEEE Trans. Communications*, vol. 48, no. 4, pp. 679–691, April 2000.

[17] Y.-L. Chao and R. A. Scholtz, "Ultra-wideband transmitted reference systems," *IEEE Trans. Vehicular Technology*, vol. 54, no. 5, pp. 1556–1569, Sep. 2005.

[18] L. Yang and G. Giannakis, "Optimal pilot waveform assisted modulation for ultrawideband communications," *IEEE Trans. Wireless Communications*, vol. 3, no. 4, pp. 1236–1249, July 2004.

[19] F. Tufvesson, S. Gezici, and A. Molisch, "Ultra-wideband communications using hybrid matched filter correlation receivers," *IEEE Trans. Wireless Communications*, vol. 5, no. 11, pp. 3119–3129, Nov. 2006.

[20] V. Lottici, A. D'Andrea, and U. Mengali, "Channel estimation for ultra-wideband communications," *IEEE J. Selected Areas in Communications*, vol. 20, no. 9, pp. 1638–1645, Dec. 2002.

[21] J. Choi and W. Stark, "Performance of ultra-wideband communications with suboptimal receivers in multipath channels," *IEEE J. Selected Areas in Communications*, vol. 20, no. 9, pp. 1754–1766, Dec. 2002.

[22] M. Win and R. Scholtz, "Characterization of ultra-wide bandwidth wireless indoor channels: a communication-theoretic view," *IEEE J. Selected Areas in Communications*, vol. 20, no. 9, pp. 1613–1627, Dec. 2002.

[23] S. Gezici, H. K. Z. Sahinoglu, and H. Poor, "Ultra-wideband impulse radio systems with multiple pulse types," *IEEE J. Selected Areas in Communications*, vol. 24, no. 4, pp. 892–898, Apr. 2006.

[24] N. He and C. Tepedelenlioglu, "Performance analysis of non-coherent UWB receivers at different synchronization levels," *IEEE Trans. Wireless Communications*, vol. 5, pp. 1266–1273, June 2006.

[25] Q. Dang, A. Trindade, A.-J. van der Veen, and G. Leus, "Signal model and receiver algorithms for a transmit-reference ultra-wideband communication system," *IEEE J. Selected Areas in Communications*, vol. 24, no. 4, pp. 773–779, Apr 2006.

[26] A. F. Molisch and et al., "IEEE 802.15.4a channel model - final report," IEEE, technical report IEEE 802.15-04-0662-02-004a, Nov. 2004.

[27] A. Salch and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Selected Areas in Communications*, vol. 5, no. 2, pp. 128–137, Feb. 1987.

[28] C. Rushforth, "Transmitted-reference techniques for random or unknown channels," *IEEE Trans. Info. Theory*, vol. 10, no. 1, pp. 39–42, Jan. 1964.

[29] G. Hingorani and J. Hancock, "A transmitted reference system for communication in random or unknown channels," *IEEE Trans. Communications*, vol. 13, no. 3, pp. 293–301, Sep. 1965.

[30] J. Romme and K. Witrisal, "Transmitted-reference UWB systems using weighted autocorrelation receivers," *IEEE Trans. Microwave Theory and Technology*, vol. 54, no. 4, pp. 1754–1761, Apr. 2006.

[31] G. Leus and A. van der Veen, "A weighted autocorrrelation receiver for transmitted reference ultra wideband communications," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 965–969, June 2005.

[32] M. Weisenhorn and W. Hirt, "Robust noncoherent receiver exploiting UWB channel properties," in *Proc. IEEE Conference on Ultra-wideband System and Technology (UWBST)*, pp. 156–160, May 2004.

[33] S. Jayaweera, "Signal design for noncoherent ppm modulation with applications to UWB communications," *IEEE Communications Letters*, vol. 9, no. 5, pp. 411–413, May 2005.

[34] N. He and C. Tepedelenlioglu, "Adaptive synchronization for non-coherent UWB receivers," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 517–520, May 2004.

[35] J. Y. Lee and R. A. Scholtz, "Ranging in a dense multipath environment using an UWB radio link," *IEEE J. Selected Areas in Communications*, vol. 20, no. 9, pp. 1677–1683, Dec. 2002.

[36] A. Rabbachin and I. Oppermann, "Synchronization analysis for UWB systems with a low-complexity energy collection receiver," in *Proc. IEEE Conference on Ultra-wideband System and Technology (UWBST)*, pp. 288–292, May 2004.

[37] S. Franze and U. Mitra, "Integration interval optimization and performance analysis for UWB transmitted reference systems," in *Proc. IEEE Conference on Ultra-wideband System and Technology (UWBST)*, pp. 26–30, May 2004.

[38] L. Yang and G. Giannakis, "Timing ultra-wideband signals with dirty templates," *IEEE Trans. Communications*, vol. 53, no. 11, pp. 1952–1963, Nov. 2005.

[39] Z. Tian and G. Giannakis, "A GLRT approach to data-aided timing acquisition in UWB radios–part I: Algorithms," *IEEE Trans. Wireless Communications*, vol. 4, no. 6, pp. 2956–2967, Nov. 2005.

[40] J. Ibrahim and R. Buehrer, "Two-stage acquisition for UWB in dense multipath," *IEEE J. Selected Areas in Communications*, vol. 24, no. 4, pp. 801–807, Apr. 2006.

[41] I. Guvenc and Z. Sahinoglu, "Threshold-based TOA estimation for impulse radio UWB systems," in *Proc. IEEE Conference on Ultra-wideband (ICUWB)*, pp. 420–425, Sep. 2005.

[42] I. Guvenc, Z. Sahinoglu, and P. Orlik, "TOA estimation for IR-UWB systems with different transceiver types," *IEEE Trans. Microwave Theory and Technology*, vol. 54, no. 4, pp. 1876–1886, June 2006.

[43] C. Carbonelli and U. Mengali, "A ranging technique for UWB indoor channel based on power delay profile analysis," *IEEE Trans. Communications*, vol. 54, no. 2, pp. 329–338, Feb. 2006.

[44] C. Mazzucco, U. Spagnolini, and G. Mulas, "Threshold-based TOA estimation for impulse radio UWB systems," in *Proc. IEEE Vehicular Technology Conference (VTC)*, vol. 5, pp. 2595–2599, May 2004.

Appendix A

The distributions of the noise terms, i.e. ζ_i $i = 1, \cdots, 4$, in (3.1) and (3.2) for PPM are derived here in detail. The derivation is based on the fact that $\{\zeta_i\}$'s are approximately Gaussian random variables and $T \gg T_p$ along with the following two equations,

$$\mathrm{E}[n(t)n^*(u)] = R_n(t-u) = 2BN_0\mathrm{sinc}(2B(t-u)) \tag{1}$$

$$\mathrm{E}[n(t)n(u)] = \mathrm{E}[n^*(t)n^*(u)] = 0. \tag{2}$$

It is trivial that $\mathrm{E}[\zeta_1] = \mathrm{E}[\zeta_2] = 0$. The autocorrelation and cross-correlation of ζ_1 and ζ_2 are given as

$$\begin{aligned}
\mathrm{E}[\zeta_1\zeta_1^*] = \mathrm{E}[\zeta_2\zeta_2^*] &= E_b \int_{T_0}^{T_0+T}\int_{T_0}^{T_0+T} g(t)g^*(u)\mathrm{E}\big[n(t)n^*(u)\big]\mathrm{d}t\mathrm{d}u \\
&= 2BN_0E_b \int_{T_0}^{T_0+T}\int_{T_0}^{T_0+T} g(t)g^*(u)\mathrm{sinc}(2B(t-u))\mathrm{d}t\mathrm{d}u \\
&= N_0E_b \int_{T_0}^{T_0+T}\int_{T_0}^{T_0+T} g(t)g^*(u)\int_{-B}^{B} e^{j2\pi f(t-u)}\mathrm{d}f\mathrm{d}t\mathrm{d}u \\
&\approx N_0E_b \int_{T_0}^{T_0+T} g(t)\int_{-B}^{B} e^{j2\pi ft}\int_{-\infty}^{\infty} g^*(u)e^{-j2\pi fu}\mathrm{d}u\mathrm{d}f\mathrm{d}t \\
&= N_0E_b \int_{T_0}^{T_0+T} g(t)\int_{-B}^{B} G^*(-f)e^{j2\pi ft}\mathrm{d}f\mathrm{d}t \\
&= N_0E_b \int_{T_0}^{T_0+T} g(t)g^*(t)\mathrm{d}t \\
&= N_0E_{cap}(T_0, T)
\end{aligned} \tag{3}$$

$$\begin{aligned}
\mathrm{E}[\zeta_1\zeta_2^*] = \mathrm{E}[\zeta_2\zeta_1^*] &= \mathrm{E}\Bigg[\sqrt{\frac{E_b}{N_f}}\sum_{n=1}^{N_f}\int_{T_0}^{T_0+T} g(t)n^*(t+nT_f)\mathrm{d}t \\
&\qquad \cdot \sqrt{\frac{E_b}{N_f}}\sum_{n=1}^{N_f}\int_{T_0}^{T_0+T} g(u)n^*(u+nT_f)\mathrm{d}u\Bigg] \\
&= E_b\int_{T_0}^{T_0+T}\int_{T_0}^{T_0+T} g(t)g(u)\mathrm{E}\big[n^*(t)n^*(u)\big]\mathrm{d}t\mathrm{d}u \\
&= 0,
\end{aligned} \tag{4}$$

thus ζ_1 and ζ_2 are independent zero-mean Gaussian random variables with variance $N_0E_{cap}(T_0,T)$. For ζ_3 and ζ_4 we have

$$\mathrm{E}[\zeta_3] = \mathrm{E}[\zeta_4] = \sum_{n=1}^{N_f}\int_{T_0}^{T_0+T} R_n(0)\mathrm{d}t = 2N_fN_0BT, \tag{5}$$

$$\begin{aligned}
\mathrm{E}[\zeta_3\zeta_3^*] = \mathrm{E}[\zeta_4\zeta_4^*] &= \mathrm{E}\Bigg[\sum_{n=1}^{N_f}\int_{T_0}^{T_0+T} n(t+nT_f)n^*(t+nT_f)\mathrm{d}t \\
&\qquad \cdot \sum_{n=1}^{N_f}\int_{T_0}^{T_0+T} n(u+nT_f)n^*(u+nT_f)\mathrm{d}u\Bigg] \\
&= \int_0^T\int_0^T \big(N_f^2R_n^2(0) + N_fR_n^2(t-u)\big)\mathrm{d}t\mathrm{d}u \\
&= 4N_f^2N_0^2B^2T^2 + 2BN_fN_0^2\int_0^T\int_0^T \mathrm{sinc}(2B(u-t))\int_{-B}^{B} e^{j2\pi f(t-u)}\mathrm{d}f\mathrm{d}t\mathrm{d}u \\
&\approx 4N_f^2N_0^2B^2T^2 + N_fN_0^2\int_0^T\int_{-B}^{B}\int_{-\infty}^{\infty} 2B\mathrm{sinc}(2Bu)e^{-j2\pi fu}\mathrm{d}u\mathrm{d}f\mathrm{d}t \\
&= 4N_f^2N_0^2B^2T^2 + N_fN_0^2\int_0^T\int_{-B}^{B}\mathrm{d}f\mathrm{d}t \\
&= 4N_f^2N_0^2B^2T^2 + 2N_fN_0^2BT.
\end{aligned} \tag{6}$$

So $\mathrm{var}[\zeta_3] = \mathrm{var}[\zeta_4] = 2N_fN_0^2BT$.