Ranging with ultrawide bandwidth signals in multipath environments

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Ranging With Ultrawide Bandwidth Signals in Multipath Environments

Understanding the fundamental limits of UWB ranging accuracy may open the way to development of high-definition localization systems.

By Davide Dardari, Senior Member IEEE, Andrea Conti, Member IEEE, Ulric Ferner, Student Member IEEE, Andrea Giorgetti, Member IEEE, and Moe Z. Win, Fellow IEEE

ABSTRACT | Over the coming decades, high-definition situationally-aware networks have the potential to create revolutionary applications in the social, scientific, commercial, and military sectors. Ultrawide bandwidth (UWB) technology is a viable candidate for enabling accurate localization capabilities through time-of-arrival (TOA)-based ranging techniques. These techniques exploit the fine delay resolution property of UWB signals by estimating the TOA of the first signal path. Exploiting the full capabilities of UWB TOA estimation can be challenging, especially when operating in harsh propagation environments, since the direct path may not exist or it may not be the strongest. In this paper, we first give an overview of ranging techniques together with the primary sources of TOA error (including propagation effects, clock drift, and interference). We then describe fundamental TOA bounds (such as the Cramér–Rao bound and the tighter Ziv–Zakai bound) in both ideal and multipath environments. These bounds serve as useful benchmarks in assessing the performance of TOA estimation techniques. We also explore practical low-complexity TOA estimation techniques and analyze their performance in the presence of multipath and interference using IEEE 802.15.4a channel models as well as experimental data measured in indoor residential environments.

KEYWORDS | Interference; localization; multipath channel; ranging; time-of-arrival; ultrawide bandwidth

I. INTRODUCTION

Highly accurate position information is of great importance in many commercial, public safety, and military applications. To this end, the integration of Global Positioning System (GPS) into cellular phones, in conjunction with WiFi localization, is igniting a new era of ubiquitous location-awareness [1]. In the coming years, we will see the emergence of high-definition situation-aware (HDSA) applications with capability to operate in harsh propagation environments where GPS typically fails, such as inside buildings and in caves. Such applications require localization systems with submeter accuracy.

Reliable localization in such conditions is a key enabler for a diverse set of applications including logistics, security tracking (the localization of authorized persons in high-security areas), medical services (the monitoring of patients), search and rescue operations (communications with fire fighters or natural disaster victims), control of home appliances, automotive safety, military systems, and a large set of emerging wireless sensor network (WSN) applications [2]. Other applications include networking protocols taking advantage of position to improve the...
The performance of routing algorithms (georouting), as well as interference avoidance techniques in future cognitive radios [3], [4].

The purpose of localization algorithms is to find the unknown positions of agent nodes given a set of measurements. Localization occurs in two main steps: (i) selected measurements are performed between nodes; (ii) these measurements are processed to determine the position of agent nodes. Further, localization techniques can be classified based on measurements between nodes such as range-based, angle-based, and proximity-based localization. Among them, range-based systems (i.e., based on distance estimates) are more suitable for high-definition localization accuracy, especially when low-complexity devices are available. We refer the reader to [29]–[36] and references therein. It is expected that the advantages of UWB-based localization will be exploited in future HDSA systems that utilize coexisting networks of sensors, controllers, and peripheral devices. The IEEE 802.15.4a is the first UWB-based standard for low-rate wireless personal-area networks (WPANs) with localization capability [37], [38]; ranging accuracy is expected to be one meter or submeter at least 90% of the time. Ranging techniques based on TOA estimation of the first arriving signal path are mainly affected by noise, multipath components, obstacles, interference, and clock drift [19], [20]. In dense multipath environments the first path is often the strongest, making estimation of the TOA challenging [36], [39]–[42].

The aim of this paper is to provide an overview of the basic theory as well as some practical schemes for first path TOA estimation with UWB received signals, and to highlight the main sources of estimation error. In particular, this paper is organized as follows. Section II provides an overview of ranging techniques. Section III describes challenges in accurate ranging based on first-path TOA estimation. Section IV describes the theoretical limits of UWB ranging systems operating in both ideal and multipath environments. Section V describes the maximum likelihood (ML) TOA estimator together with some practical low-complexity schemes to mitigate multipath and interference. TOA estimation performance of these practical schemes in realistic environments is investigated in Section VI. Finally, selected future research directions are discussed in Section VII and some conclusions are given in Section VIII.

### II. RANGING TECHNIQUES

Ranging techniques have significant effects on localization accuracy, system complexity, and system cost [8]. This section outlines the two most widely used ranging techniques: time-based and received signal strength (RSS)-based systems.

#### A. Time-Based Ranging

The information related to the separation distance between a pair of nodes A and B can be obtained using measurements of the signal propagation delay, or time-of-flight (TOF) \( \tau_1 = d/c \), where \( d \) is the actual distance between the two nodes and \( c \) is the speed of electromagnetic waves \( (c \approx 3 \times 10^8 \text{ m/s}) \). This can be accomplished using one-way TOA, two-way TOA, or time difference-of-arrival (TDOA) ranging techniques, each of which is briefly described below.

1) One-Way TOA Ranging: At time \( t_1 \), node A transmits a packet that contains the timestamp \( t_1 \) at which A's packet was sent. Node B receives the packet at time \( t_2 \). Under ideal conditions, that is, when node clocks are perfectly synchronized to a common time reference, it is clear that \( \tau_1 \) can be determined at node B as \( \tau_1 = t_2 - t_1 \), from which the distance can be estimated. Synchronization error can significantly affect ranging error (this aspect is discussed in Section III).

2) Two-Way TOA Ranging: In two-way ranging, the system estimates the signal round-trip time (RTT) without a common time reference. Node A transmits a packet to node B, which replies by transmitting an acknowledgment packet to A after a response delay \( \tau_4 \) [43], [44]. The RTT at A is determined by \( \tau_{RTT} = 2\tau_1 + \tau_4 \), from which the distance can be estimated assuming \( \tau_4 \) is known.

While two-way ranging eliminates the error due to imperfect synchronization between nodes, relative clock drift still affects ranging accuracy. In fact, propagation delay in indoor applications is typically on the order of nanoseconds, but the response delay \( \tau_4 \) can be of a few microseconds due to bit synchronization and channel estimation delays [44]. Even a small clock offset between nodes will correspond to a large error in \( \tau_1 \) estimation due to error accumulation over \( \tau_4 \) (this aspect is discussed further in Section III).

3) Time Difference-of-Arrival: TDOA-based localization systems do not rely on absolute distance estimates between pairs of nodes. Such systems typically employ one of two schemes as follows. In the first, multiple signals are broadcast from synchronized fixed nodes (named anchors) located at known positions and agents measure the TDOA (similar technique is used by GPS). In the second scheme, a reference signal is broadcast by an agent and is received by several anchors. The anchors share their estimated TOA and compute the TDOA. Typically anchors are
synchronized through a wired network connection. To calculate the position of the agent, at least three anchors with known position and two TDOA measurements are required. Each TDOA measurement can be geometrically interpreted as a hyperbola formed by a set of points with constant range differences (time differences) from two anchors [45], [46].

**B. Received Signal Strength-Based Ranging**

RSS ranging is based on the principle that the greater the distance between two nodes, the weaker their relative received signals. This technique is commonly used in low-cost systems such as WSNs because hardware requirements and costs can be more favorable compared to time-based techniques. In RSS-based systems, a receiving node B estimates the distance to a transmitting node A by measuring the RSS from A and then using theoretical and/or empirical path-loss models to translate the RSS into a distance estimate. These models strongly affect ranging accuracy [47].

A widely used model to characterize the RSS at node B from node A’s transmission is given by [48]

$$P_r(d) = P_0 - 10\gamma \log_{10} d + S$$

where $P_r(d)$ (dBm) is the received signal power, $P_0$ is the received power (dBm) at a reference distance of 1 m (which depends on the radio characteristics as well as the signal wavelength), $d$ (meters) is the separation between A and B, and $S$ (dB) represents the large-scale fading variations (i.e., shadowing). It is common to model $S$ (dB) as a Gaussian random variable (RV) with zero mean and standard deviation $\sigma_S$ [49]. The parameter $\gamma$ is known as the path-loss exponent, which typically assumes values between 2 and 6 [48].

The primary disadvantage of RSS ranging is that in cluttered environments the propagation phenomena cause the attenuation of the signal to poorly correlate with distance, resulting in inaccurate distance estimates [47], [50].

**III. ERROR SOURCES IN TIME-BASED RANGING**

Localization performance is highly dependent on the quality of range measurements. For this reason, it is important to understand the diverse set of factors that can degrade ranging accuracy. This section focuses on time-based ranging with UWB signals because it can provide high ranging accuracy relative to other techniques.

Let us first define a few terms with reference to Fig. 1. We refer to a range measurement between two nodes as a DP measurement if the measured signal traveled in a straight line between the two nodes through a medium with constant and known permittivity (such as in air, as per Tx-Rx1). A measurement can be non-DP if, for example, the DP signal is obstructed (Tx-Rx2) or if the first arriving path has been reflected by obstacles (Tx-Rx3). A line-of-sight (LOS) measurement occurs when the first arriving path is along the DP, while a non-LOS (NLOS) measurement occurs either when DP is completely blocked (in which case the first arriving path comes from the reflected signal) or from DP excess delay (in which the signal traverses through different materials in a straight line resulting in additional TOF delays).
Typical sources of error that affect time-based ranging are described in the following sections.

A. Propagation

Sources of error from wireless signal propagation include multipath, DP excess delay, and DP blockage.

1) Multipath: Multipath fading is caused by the destructive and constructive interference of signals arriving at the receiver via different propagation paths. When narrowband systems are used in cluttered environments, the signals arriving via different propagation paths usually cannot be resolved. This results in the destructive and constructive interference of signals and makes the detection of the DP component, if it exists, difficult. Although the use of UWB signals can resolve these multipath components [23], [27], the large number of multipath components in harsh environments makes DP detection nontrivial.

We consider a scenario in which a unitary energy pulse \( p(t) \) is transmitted (with duration \( T_p \)) through a channel with multipath and thermal noise.\(^1\) The received signal can be written as

\[
    r(t) = s(t) + n(t)
\]

where \( s(t) \) is the channel response (CR) to the transmitted pulse \( p(t) \) and \( n(t) \) is additive white Gaussian noise (AWGN) with zero mean and two-sided power spectral density \( N_0/2 \). We consider frequency-selective fading channel, in which case [27]

\[
    s(t) = \sqrt{E_p} \sum_{l=1}^{L} \alpha_l p(t - \tau_l) + n(t)
\]

where \( L \) is the number of received multipath components with \( \alpha_l \) and \( \tau_l \) denoting the amplitude and delay of the \( l \)th path, respectively, with the normalization \( \sum_{l=1}^{L} E\{\alpha_l^2\} = 1 \) such that \( E_p \) represents the average received energy.\(^2\) The goal is to estimate the TOA \( \tau = \tau_1 \) of the first path by observing the received signal \( r(t) \). This task can be challenging due to the presence of thermal noise and multipath components.

2) DP Excess Delay: Another source of error is the DP excess delay caused by the propagation of a partially obstructed DP component that travels through different obstacles like walls in buildings. The propagation time of these signals depends not only upon the traveled distance but also on the encountered materials. Since the propagation of electromagnetic waves is slower in some materials compared to air the signal arrives with excess delay, thereby introducing a positive bias in the range estimate.

The ranging error due to DP excess delay can be characterized as follows: the speed of electromagnetic waves traveling in a homogeneous material is reduced by a factor of \( \sqrt{\varepsilon_r} \) with respect to the speed of light \( c \), where \( \varepsilon_r \) is the relative electrical permittivity of the material. The DP excess delay \( \Delta \tau \) introduced by the material with thickness \( d_W \) is given by [51], [52]

\[
    \Delta \tau = (\sqrt{\varepsilon_r} - 1) \frac{d_W}{c}.
\]

Experiments in common office environments have shown that this delay significantly affects ranging accuracy of UWB systems. Recent measurement campaigns show that the mean of the ranging error (the bias in range estimates) due to DP excess delay is on the order of the thickness of the obstacles, such as walls, between two transmitting nodes [53].

3) DP Blockage: When the DP to a node is completely obstructed the node’s receiver can only observe NLOS components, resulting in estimated distances larger than the true distance. An important observation is that the effect of DP blockage and DP excess delay is the same: they both add a positive bias to the range estimate.

Both DP excess delay and DP blockage conditions occur in NLOS conditions. Identification of the NLOS condition is a precursor for any ranging/localization algorithm [54]–[61]. Once the NLOS condition has been identified, a simple way to mitigate the NLOS effects is to ignore NLOS measurements [62]. However, in [21] and [63]–[70], it is shown that a NLOS measurement provides information that can improve the localization accuracy. Quadratic programming and linear programming approaches have been used to mitigate NLOS effects in [54] and [71], respectively. Another method to account for NLOS propagation is to weight the range measurements in the localization algorithm according to the range estimate accuracy [61], [72]. Prior knowledge of the environment or cooperation between agents are other possible ways to improve localization performance in NLOS conditions [21], [53], [73]. The effect of NLOS propagation can also be mitigated by introducing memory into the system through the adoption of tracking techniques based on, for example, Bayesian filtering [70], [74], [75]. Additional NLOS mitigation techniques can be found in [60], [76], and [77].

B. Clock Drift

Time-based ranging requires precise time interval measurements (with errors on the order of 1 ns or less
when centimeter accuracy is required). To this end, nodes are equipped with an oscillator from which an internal clock reference is derived to measure the true time $t$. Numerous physical effects cause oscillators to experience independent frequency drifts, which result in large timing errors. These errors can be particularly significant in systems with poor oscillators (such as in low-cost WSNs).

The local time of a clock in a device can be expressed as a function $C(t)$ of the true time $t$, where $C(t) = t$ for a perfect clock. As a consequence, only an estimate $\hat{t} = C(t)$ of the true time $t$ is available. For short time intervals, $C(t)$ can be modeled as [78]

$$C(t) = (1 + \delta)t + \mu$$

(5)

where $\delta$ is the clock drift relative to the correct rate and $\mu$ is the clock offset. Note that the rate $dC(t)/dt$ of a perfect clock is one (i.e., $\delta = 0$).

Clock drift affects the time interval measurement. Specifically, if a single node wants to measure a time interval of true duration $\tau = t_2 - t_1$ seconds, then the corresponding estimated value $\hat{\tau}$ is

$$\hat{\tau} = C(t_2) - C(t_1) = \tau(1 + \delta).$$

(6)

On the other hand, a node effectively generates delay

$$\hat{\tau}_d = \frac{\tau_d}{1 + \delta}$$

(7)

whenever it attempts to generate a delay of $\tau_d$ seconds. The following sections describe the effect of clock drift and offset in one-way and two-way ranging protocols when node A and B have clock drifts $\delta_A$, $\delta_B$ and offsets $\mu_A$, $\mu_B$, respectively.

1) Clock Drift in One-Way Ranging Protocols: In one-way ranging, nodes are synchronized to a common time-base by using, for example, a network synchronization protocol [78]. In this case, $\mu_A$ and $\mu_B$ are the residual offsets caused by imperfect synchronization with respect to the common time-base, assuming, without loss of generality, that the last network synchronization phase occurred at time $t = 0$.

A packet is transmitted by node A at its local time $C_A(t_1)$ and is received by node B at its local time $C_B(t_2)$.

Node B calculates the estimated propagation delay as

$$\hat{\tau}_d = C_B(t_2) - C_A(t_1) = \tau_1 + \delta_{B}t_2 - \delta_{A}t_1 + \mu_B - \mu_A.$$  

(8)

Equation (8) shows that time estimation is affected both by clock drift and clock offset. However, clock offset could be on the order of several microseconds, and hence ranging accuracy depends highly on network synchronization performance [78]. For this reason, one-way ranging requires stringent network synchronization that may not be feasible in some systems.

2) Clock Drift in Two-Way Ranging Protocols: In two-way ranging, both accurate delay generation and time measurement are important for accurate ranging. According to (7), the effective response delay $\hat{\tau}_d$ generated by node B in the presence of a clock drift $\delta_B$ is given by

$$\hat{\tau}_d = \tau_1/(1 + \delta_B),$$

whereas the estimated RTT denoted by $\hat{\tau}_{RT}$, according to node A's time-scale, is

$$\hat{\tau}_{RT} = 2\tau_1(1 + \delta_A) + \tau_0(1 + \delta_A)/(1 + \delta_B).$$

(9)

In the absence of other information, node A can estimate the propagation time $\hat{\tau}_1$ by equating (9) with the supposed round-trip time $2\hat{\tau}_1 + \hat{\tau}_d$, leading to

$$\hat{\tau}_1 = \tau_1(1 + \delta_A) + \frac{\varepsilon \tau_d}{2(1 + \delta_A - \varepsilon)}$$

(10)

where $\varepsilon \triangleq \delta_A - \delta_B$ is the relative clock offset.

To demonstrate typical ranging errors due to the relative clock offset, Fig. 2 shows error in TOF estimation $\hat{\tau}_1 - \tau_1$ as a function of $\varepsilon$ using a typical value $\delta_A = 10^{-5}$ (10 ppm) for different response delays. Fig. 2 shows that both the relative clock offsets as well as response delay strongly affect ranging accuracy. For example, a target TOF estimation error of 33 ps (about 1 cm) can be satisfied for $\varepsilon$ up to $10^{-5}$ if the response delay $\tau_d$ is below 10 $\mu$s. For a less pessimistic value of $\varepsilon = 10^{-6}$, the requirement on the maximum response delay relaxes to $\tau_d \approx 100 \mu$s.

Another measure of clock precision is part-per-million (ppm), which is defined as the maximum number of extra (or missed) clock counts over a total of $10^6$ counts, that is, $\delta \cdot 10^6$.  

4The timestamp $C_A(t_1)$ is included in the transmitted packet and hence is also known by node B.

5Network synchronization requirements are reduced when systems use ranging signals, such as ultrasound, with propagation speeds significantly slower than electromagnetic waves. The propagation speed of acoustic waves (2-5 m/s) corresponds to propagation delay values that are several orders of magnitude larger than the respective network synchronization errors. This makes one-way ranging technique attractive for acoustic-based localization systems [8].
RTT measurement accuracy can be improved by adopting high-precision oscillators (not feasible in low-cost WSNs) or by adopting rapid synchronization techniques at the physical layer [2], [44], [79]. In addition, response packet generation can be implemented at the medium access control (MAC) layer to avoid large delays produced by upper protocol stack layers.

### C. Interference

UWB is intended to operate as an underlay technology, hence it has to coexist with others communication systems.\(^6\) Therefore, location-aware systems based on UWB technology are expected to operate in the presence of both narrowband interference (NBI) and UWB multi-user interference (MUI), which can degrade TOA estimation performance. In such scenarios, the received signal can be expressed as

\[
r(t) = s(t) + n(t) + i(t)
\]

where \(i(t)\) accounts for the interference.

It has been shown in [81] that a single NBI can be well approximated by a tone as

\[
i(t) = \sqrt{2}I\alpha_1 \cos(2\pi f_1 t + \phi_1)
\]

where \(I\) is the average received power, \(f_1\) is the center frequency of the NBI spectrum, and \(\alpha_1\) and \(\phi_1\) are the amplitude and phase, respectively, of the fading associated with the NBI. On the other hand, the expression for the MUI depends on the structure of transmitted signals as well as the channel characteristics (an example is described in Section VI-B).

The effect of NBI and MUI on TOA estimation as well as the explanation of some mitigation techniques will be discussed more in detail in Section VI-B.

### IV. THEORETICAL PERFORMANCE LIMITS OF TIME-BASED RANGING

We have seen that time-based ranging requires the estimation of the TOA of the first arriving path. Thus, understanding the theoretical performance limits of TOA estimation plays an important role in the design of TOA estimators. This general problem is summarized as follows.

- We are interested in the estimation of the TOA \(\tau = \tau_1\) of the direct path by observing the received signal \(r(t)\) within the observation interval \([0,T_{ob}]\).
- We consider \(\tau\) to be uniformly distributed in the interval \([0,T_a]\), which represents the initial uncertainty about \(\tau\), with \(T_a < T_{ob}\) so that all multipath components fall within the observation interval.
- According to (2) and (3), the received signal depends on the set of nuisance parameters \(\mathcal{U} = \{\tau_2, \tau_3, \ldots, \tau_1, \alpha_1, \alpha_2, \ldots, \alpha_L\}\) that, due to noise and fading, can strongly affect the TOA estimation. The complete set of unknown channel parameters is \(\mathcal{V} = \{\tau_1, \tau_2, \ldots, \tau_L, \alpha_1, \alpha_2, \ldots, \alpha_L\}\).

We begin by describing the theoretical performance limits of time-based ranging in ideal propagation conditions and then relate the theoretical performance with the primary sources of error, as outlined in Section III. For a more general description of the theoretical limits of range-based localization, we refer the reader to [19], [21], [63]–[67], [69], [77], [82], [83] and the references therein.

### A. Theoretical Limits in Ideal Propagation Conditions

This section presents a detailed analysis of the performance limits of TOA estimation in AWGN channels to understand which fundamental system parameters dominate ranging accuracy. In an AWGN channel (in the absence of other error sources), the received signal can be written as

\[
r(t) = \sqrt{E}_n p(t - \tau) + n(t).
\]

Under this model, TOA estimation is a classical nonlinear parameter estimation problem, with a solution based on a matched filter (MF) receiver (the block diagram of which is shown in Fig. 3). The received signal is first processed by a filter matched to the pulse \(p(t)\) (or, equivalently, by a correlator with template \(p(t)\)). TOA estimate is given by the time instant corresponding to the maximum absolute peak at the output of the MF over the observation interval.

![Fig. 2. TOF estimation error in two-way ranging due to relative clock drifts with \(\xi_a = 10^{-5}\) and \(\tau_1 = 100\) ns.](image-url)
This scheme yields a ML estimate, which is known to be asymptotically efficient, that is, the performance of the estimator achieves the Cramér–Rao bound (CRB) for large signal-to-noise ratios (SNRs) [84].

1) Cramér-Rao Bound: The mean square error (MSE) of any unbiased estimate \( \hat{\tau} \) of \( \tau \) can be lower bounded by

\[
\mathbb{V}(\hat{\tau}) = \mathbb{E}(\xi^2) \geq \text{CRB}
\]  

(14)

where \( \xi = \hat{\tau} - \tau \) is the estimation error and

\[
\text{CRB} = \frac{N_0/2}{(2\pi)^2 E_p \beta^2} = \frac{1}{8\pi^2 \beta^2 \text{SNR}}
\]  

(15)

is known as the CRB [84]. Here \( \text{SNR} \triangleq E_p/N_0 \) and the parameter \( \beta^2 \) represents the second moment of the spectrum \( P(f) \) of \( p(t) \),

\[
\beta^2 \triangleq \frac{\int_{-\infty}^{\infty} f^2 |P(f)|^2 df}{\int_{-\infty}^{\infty} |P(f)|^2 df}.
\]  

(16)

Thus the best achievable accuracy of a range estimate \( \hat{d} \) derived from TOA estimation satisfies the following inequality:

\[
\mathbb{V}(\hat{d}) \geq \frac{c^2}{8\pi^2 \beta^2 \text{SNR}}.
\]  

(17)

Notice that the lower bound in (17) decreases with both SNR and the constant \( \beta^2 \), which depends on the shape of the pulse. This reveals that signals with high power and wide transmission bandwidth are beneficial for ranging. As a comparison, the CRB for a distance estimate \( \hat{d} \) based on RSS measurements and path-loss model (1) is given by [19]

\[
\mathbb{V}(\hat{d}) \geq \left( \frac{\ln 10 \sigma S}{10 \gamma d} \right)^2.
\]  

(18)

In contrast to time-based methods, ranging capability using RSS measurements does not depend on the shape of the transmitted signal.

For time-based ranging methods, the shape, and hence the bandwidth, of the signal plays an important role in ranging accuracy. Typically the \( n \)th derivative of the basic Gaussian pulse \( p_0(t) = \exp(-2\pi^2 / \tau_p^2) \) is adopted in UWB systems [85]. In this case, we have

\[
p(t) = p_0^{(n)}(t) \sqrt{\frac{(n-1)!}{(2n-1)! \pi^n \tau_p^{1-2n}}} \quad \text{for } n > 0
\]  

(19)

where \( \tau_p \) is a parameter affecting the pulse width and \( p_0^{(n)}(t) \) denotes the \( n \)th order derivative of \( p_0(t) \) with respect \( t \). It is easy to show that \( p(t) \) has unitary energy and that the effective bandwidth \( \beta^{(n)} \) is given by

\[
\beta^{(n)} = \sqrt{\frac{2n + 1}{2\pi \tau_p^2}}
\]  

(20)

From (20), we can observe that a lower CRB for TOA estimation can be achieved by increasing \( n \) or decreasing \( \tau_p \).

Alternatively, the IEEE 802.15.4a standard suggests the following bandpass pulse with center frequency \( f_0 \) and root

\[
\text{The square-law device in Fig. 3 removes the sign of signal peaks and, for example, can be replaced equivalently by a full-wave rectifying device.}
\]

\[
\text{Parameter } \beta \text{ is often called effective bandwidth.}
\]
raised cosine (RRC) envelope [37]

\[
p(t) = \frac{4\nu\sqrt{2}}{\pi\sqrt{\tau_p}} \cos\left(\frac{(1 + \nu)\pi t}{\tau_p} + \frac{\sin((1-\nu)\pi t/\tau_p)}{4\nu/\tau_p}\right) - \cos(2\pi f_0 t)
\]

(21)

where parameter \( \tau_p \) and roll-off factor \( \nu \) determine the bandwidth \( W = (1 + \nu)/\tau_p \).\(^{10}\)

2) Ziv–Zakai Bound: In nonlinear estimation problems, the CRB is typically used as a performance benchmark. However, it is well known that the CRB is not accurate at low and moderate SNR. In fact, the performance of the TOA estimator, like all nonlinear estimators, is characterized by the presence of distinct SNR regions (low, medium, and high SNRs) corresponding to different modes of operation. This behavior is referred to as the threshold effect and has been studied in a variety of contexts (e.g., [86]–[91]). In a low SNR region (also known as the a priori region), signal observation does not provide significant additional information, and the MSE is close to that obtained solely from the a priori information about the TOA. In a high SNR region (also known as the asymptotic region), the MSE is accurately described by the CRB. Between these two extremes, there may be an additional region (also known as the transition region or ambiguity region) where observations are subject to ambiguities that are not accounted for by the CRB [84]. Therefore other bounds, which are more complicated but tighter than the CRB, have been proposed in the literature. In particular, the Barakian bound identifies the SNR values (thresholds) that distinguish the ambiguity region [92], [93]. The Ziv–Zakai bound (ZZB) [94], with its improved versions such as the Bellini–Tartara bound [95] and the Chazan–Zakai–Ziv bound [96], as well as the Weiss–Weinstein bound (WWB) [89] are more accurate than the Barakin bound. They can be applied to a wider range of SNRs and account for both ambiguity effects and a priori information of the parameter to be estimated. However, they may not be analytically tractable in many cases or require more complicated evaluations compared to the CRB [35], [97]–[99].

We now briefly review the ZZB. It can be derived starting from the following general identity for MSE estimation\(^{11}\):

\[
\mathbb{V}(\hat{\tau}) = \mathbb{E}\{\xi^2\} = \frac{1}{2} \int_0^\infty z \cdot P\{\xi \geq z\} dz
\]

(22)

and then by finding a lower bound on \( P\{\xi \geq z/2\} \) [95]. In particular, \( P\{\xi \geq z/2\} \) is related to the error probability of a classical binary detection scheme with equally probable hypothesis\(^{12}\)

\[
\mathcal{H}_1: r(t) \sim p(r(t)|\tau) \mathcal{H}_2: r(t) \sim p(r(t)|\tau + z)
\]

(23)

when using a suboptimum decision rule as described in [96]. It can be shown that (22) can be lower bounded using the error probability corresponding to the optimum decision rule based on the likelihood ratio test (LRT)

\[
\Lambda(r(t)) = \frac{p(r(t)|\tau)}{p(r(t)|\tau + z)}
\]

(24)

When \( \tau \) is uniformly distributed in \([0, T_a]\), the ZZB is given by [95], [96]

\[
\text{ZZB} = \frac{1}{T_a} \int_0^{T_a} z(T_a - z) P_{\min}(z) dz
\]

(25)

where \( P_{\min}(z) \) is the error probability corresponding to the optimum decision rule. In an AWGN channel, this error probability is

\[
P_{\min}(z) = Q\left(\sqrt{\text{SNR}(1 - \rho_p(z))}\right)
\]

(26)

where \( Q(\cdot) \) is the Gaussian Q-function and \( \rho_p(z) \) is the autocorrelation function of \( p(t) \) [100]. Note that the ZZB is obtained by recognizing that the performance evaluation of an estimation problem can be transformed to a binary detection problem. We will exploit this observation in Section IV-B to obtain the ZZB for multipath environments.

Fig. 4 shows the root mean square error (RMSE) for CRB and ZZB using the second- and sixth-order Gaussian monocyte pulses with \( \tau_p = 0.192 \) ns, as well as bandpass RRC pulses with center frequency \( f_0 = 4 \) GHz, roll-off \( \nu = 0.6 \), \( \tau_p = 1 \) ns, and \( \tau_p = 3.2 \) ns. The presence of the threshold effect is evident from the ZZB. In fact, the a priori region can be observed by ZZB which approaches to \( T_a/\sqrt{12} \) for low SNR values. On the other hand, this behavior cannot be observed by the CRB. For high SNRs, that is, in the asymptotic region, the estimation error approaches that predicted by the CRB. Note that as

\(^{10}\)Two different values of \( \tau_p \) are recommended [37]: \( \tau_p = 1 \) ns and \( \tau_p = 3.2 \) ns with \( \nu = 0.6 \), corresponding to two different bandwidths \( W = 1.6 \) GHz and \( W = 500 \) MHz, respectively.

\(^{11}\)Here the expectation is performed with respect to \( \tau \) and \( r(t) \).

\(^{12}\)The notation \( \sim p(\cdot) \) stands for “with distribution \( p(\cdot) \).”
Global errors: In low SNR conditions, the output of the MF designed only for the AWGN channel (which is no longer the optimum estimator in multipath environments) exhibits adjacent peaks with comparable amplitude to the correct peak. This engenders ambiguity in the selection of the correct peak (see \( v(t) \) in Fig. 3) and hence estimation performance is dominated by large errors with magnitude greater than the width of the transmitted pulse. As a consequence, the TOA estimate tends to be biased, and the corresponding MSE increases drastically for SNR below a certain value (i.e., the ambiguity effect is observed) [101].

Local errors: When a system operates in high SNR conditions, the effect of large errors due to multipath can be made negligible. In such cases, TOA estimation performance is dominated by small errors with magnitude on the order of the transmitted pulse width and is well accounted for by the CRB. However, high SNR conditions often cannot be met in UWB systems since they are primarily intended to operate in harsh multipath conditions with low SNR values.

1) Cramér–Rao Bound: In the presence of multipath, the log-likelihood function for the set of channel parameters \( \mathcal{V} \) takes the form [102], [103]

\[
\mathcal{L}(\mathcal{V}) = -\frac{1}{N_0} \int_0^{\tau_{\text{obs}}} \left[ r(t) - \sqrt{E_p} \sum_{l=1}^L a_l p(t - \tau_l) \right]^2 \, dt
\]  

(27)

from which the CRB on \( \tau \) can be derived by calculating the corresponding Fisher information matrix. As shown in [104], the Fisher information matrix can be assembled in the form

\[
J = \begin{bmatrix} f_{1,1} & B^T \\ B & C \end{bmatrix}
\]  

(28)

where \( f_{1,1} \) is the Fisher information corresponding to the single path case, whereas the submatrices \( B \) and \( C \) depend on the multipath parameters. The CRB can be evaluated using (28) as

\[
\text{CRB} = (f_{1,1} - B C^{-1} B^T)^{-1}.
\]  

(29)

When the resolvable multipath condition holds (i.e., \( |\tau_i - \tau_j| \geq T_p \quad \forall i \neq j \)), it can be shown that \( B = 0 \); thus (29) reduces to

\[
\text{CRB} = \frac{N_0}{8\pi^2 E_p \sigma_f^2 T_p^2}.
\]  

(30)

This has the same form as the AWGN counterpart given by (15), which could give credence to the belief that the estimation of the first path is decoupled from other multipath components. However, practical estimation of the first path may be dependent on the presence of others paths, even in resolvable channels, in the medium or low SNR regimes where the CRB is loose.
2) Ziv–Zakai Bound: The evaluation of the ZZB for joint estimation of $\tau$ and $U$ requires the optimization of a constrained 2L-dimensional cost function involving 2L-fold integrals [97]. Since we are only interested in the estimation of $\tau$, we use the single parameter bound given by (25), considering $U$ as a nuisance set of parameters with a different interpretation of $P_{\text{min}}(z)$. In particular, $P_{\text{min}}(z)$ corresponds to the error probability of the binary composite hypothesis test.\(^{13}\) The derivation of $P_{\text{min}}(z)$ depends on the a priori information about the multipath conditions at the receiver. In this context, we consider the following two cases: i) $U$ is a set of known deterministic parameters and ii) $U$ is a set of RVs with known distributions. Note that the former case assumes the receiver has perfect knowledge of the channel realization. This approach was first proposed in [35] and later extended to a wideband Gaussian random channel in [105]. In the following, we explain the latter case when $U$ is a set of RVs with known distributions, which typically gives an improved bound.

The LRT in (24) requires the average of the conditional probability distribution function (PDF) $p\{r(t)|\tau, U\}$ over all nuisance parameters $U$ as

$$p\{r(t)|\tau\} = \int_{\mathcal{U}} f_U(U)p\{r(t)|\tau, U\}dU$$  \hspace{1cm} (31)

where $f_U(U)$ is the joint PDF of $U$.

In general, the performance of the estimator resulting from (31) cannot be evaluated analytically except in a few special cases [106]. Alternatively, the ZZB can be evaluated when the CR belongs to a random process with a (typically large) finite ensemble of realizations $\{s^{(k)}(t)\}_{k=1}^{N_{ch}}$, where $N_{ch}$ is the number of realizations [35]. These CRs can be measured or generated using a statistical model in a Monte Carlo approach. In particular, $P_{\text{min}}(z)$ can be evaluated by averaging the conditional error probabilities, conditioned on a given CR, as follows

$$P_{\text{min}}(z) = P\{H_1\}P\{\Lambda(r(t)) < 1|H_1\}$$

$$+ P\{H_2\}P\{\Lambda(r(t)) > 1|H_2\}$$

$$= P\{\Lambda(r(t)) < 1|H_1\}$$

$$+ \sum_{k=1}^{N_{ch}} f_k(k)P\{\Lambda\left(\frac{r^{(k)}(t)}{\nu}\right) < 1\}$$  \hspace{1cm} (32)

where $r^{(k)}(t) = s^{(k)}(t - \tau) + n(t)$ and $f_k(k)$ is the probability of the $k$th CR (we assume CRs are equiprobable, thus $f_k(k) = 1/N_{ch}$ for all $k = 1, 2, \ldots, N_{ch}$). The LRT in (32) is

$$\Lambda(r(t)) = \frac{\sum_{i=1}^{N_{ch}} f_k(i)p\{r(t)|\tau, s^{(i)}(t)\}}{\sum_{i=1}^{N_{ch}} f_k(i)p\{r(t)|\tau + z, s^{(i)}(t)\}}$$

$$= \sum_{i=1}^{N_{ch}} \exp\left\{-\frac{1}{2N_{ch}} \int_{T_{ch}} [r(t) - s^{(i)}(t - \tau)]^2 dt\right\}$$

$$\sum_{i=1}^{N_{ch}} \exp\left\{-\frac{1}{2N_{ch}} \int_{T_{ch}} [r(t) - s^{(i)}(t - \tau - z)]^2 dt\right\} z.$$  \hspace{1cm} (33)

To simplify this problem, [35] presents some simple approximated expressions for (32). In particular, considering only the dominant terms, the LRT in (33) is plotted as a reference. These results clearly show the a priori region where the RMSE approaches CRs $s^{(k)}(t - \tau)$ and $s^{(i)}(t - \tau - z)$. In particular, $P_{\text{min}}(z)$ can be approximated by

$$P_{\text{min}}(z) \approx \frac{1}{N_{ch}} \sum_{k=1}^{N_{ch}} Q\left(\sqrt{\frac{\text{SNR}}{2} d_{k,\text{ch}}^2(z)}\right).$$  \hspace{1cm} (35)

Substituting (35) into (25), we obtain the desired ZZB for TOA estimation in multipath fading channels.

We now demonstrate applications of ZZB using the IEEE 802.15.4a channel models as well as experimentally measured data. ZZB is plotted in Fig. 5 as a function of the SNR for the IEEE 802.15.4a channel models CM1, CM4, and CM5, which correspond to, respectively, the indoor residential LOS, indoor office LOS, and outdoor NLOS environments [28], [37]. In each environment $N_{ch} = 100$ CRs were generated. An RRC bandpass pulse with center frequency $f_0 = 4$ GHz, rolloff factor $\nu = 0.6$, and $T_p = 1$ ns is considered. Note that the CRB in AWGN using (15) is plotted as a reference. These results clearly show the a priori region where the RMSE approaches $T_d/\sqrt{12}$, the ambiguity region, and the asymptotic region where the RMSE approaches the CRB. Significant deficiency of the CRB compared to the ZZB in predicting MSE of TOA estimation is observed for a wide range of practical SNR values.

Fig. 6 shows the ZZB as a function of the SNR using measured data obtained in a variety of indoor residential

Vol. 97, No. 2, February 2009 | PROCEEDINGS OF THE IEEE 413
environments.\textsuperscript{14} TOA estimation degrades when the agent node moves from LOS to NLOS conditions. It is shown that the required SNR to achieve a target RMSE strongly depends on the multipath conditions.

V. TOA ESTIMATORS

The performance of any TOA estimator can be bounded using theoretical bounds described in the previous section. We now describe the ML TOA estimator as well as some practical low-complexity TOA estimators.

A. Maximum Likelihood Estimator

ML estimators are known to be asymptotically efficient, that is, their performance achieves the CRB in high SNR region. As discussed previously, a TOA estimate can be obtained by using a MF (or, equivalently, a correlator) matched to the received signal; the TOA estimate is equal to the time delay that maximizes the MF output. Note that in the presence of multipath, the template in the receiver (also known as the locally generated reference) should be proportional to the CR\(\mathrm{s}(t)\) instead of \(\rho\mathrm{s}(t)\) in the case of an AWGN channel. However, it is difficult to implement this estimator since the received waveform must be estimated. In addition, each received pulse (echo) may have a different shape than the transmitted pulse due to varying antenna characteristics and materials for different propagation paths.

When channel parameters are unknown, TOA estimation in multipath environments is closely related to channel estimation. In this case, path amplitudes and TOAs are jointly estimated using, for example, the ML approach [24], [88], [90], [91], [102], [107]. In the presence of AWGN the ML criterion is equivalent to the minimum-mean-square-error (MMSE) criterion. Given an observation \(r(t)\), the ML estimate of the set \(\mathcal{V}\) of unknown channel parameters corresponds to the set of values that maximizes (27).\textsuperscript{15}

We now outline the derivation for the ML estimate of the path arrival times and their respective amplitudes. Let

\[
\begin{align*}
\tau & = [\tau_1, \tau_2, \ldots, \tau_L]^T, \\
\alpha & = [\alpha_1, \alpha_2, \ldots, \alpha_L]^T, \quad \text{and} \\
R(\tau) & \text{ be the autocorrelation matrix of } p(t) \text{ with elements } R_{ij} = \rho_p(\tau_i - \tau_j). \\
\end{align*}
\]

It can be shown that the ML estimates of \(\tau\) and \(\alpha\) are, respectively, given by [24]\textsuperscript{16}

\[
\hat{\tau} = \arg \max_{\tau} \left\{ \mathbf{\chi}^\dagger(\hat{\tau}) R^{-1}(\hat{\tau}) \mathbf{\chi}(\hat{\tau}) \right\} 
\]

and

\[
\hat{\alpha} = R^{-1}(\hat{\tau}) \mathbf{\chi}(\hat{\tau}) 
\]

where

\[
\mathbf{\chi}(\tau) \triangleq \int_{0}^{\tau_{obs}} \begin{bmatrix} p(t - \tau_1) \\ p(t - \tau_2) \\ \vdots \\ p(t - \tau_L) \end{bmatrix} dt 
\]

\textsuperscript{14}The interested reader is referred to [25], [26] for a detailed explanation of the different environments under test.

\textsuperscript{15}The ML estimation can also be performed by adopting super-resolution techniques that operate in the frequency domain [108]–[111].

\textsuperscript{16}\(\mathbf{X}^\dagger\) denotes the conjugate transpose of the matrix \(\mathbf{X}\).
is the correlation between the received signal and differently delayed replicas of the transmitted pulse. From (36) and (37), we note in general that, when the multipath is not resolvable, the estimate of the first path can depend strongly on other channel parameters.

On the other hand, when the channel is resolvable, (36) and (37) simplify to [24], [112]

\[ \hat{\tau} = \arg \max_{\tau} \mathbf{X}^T(\tau) \mathbf{X}(\tau) \]  

(39)

\[ \hat{\tau} = \arg \max_{\tau} \left\{ \sum_{l=1}^{L} \int_{0}^{T_{ch}} r(t) p(t - \hat{\tau}_l) dt \right\}^2 \]  

(40)

and \( \mathbf{X} = \mathbf{X}(\hat{\tau}) \). In this case, the estimation of the TOA of the direct path is decoupled from the estimation of the other channel parameters and reduces to a simple process involving a single correlator or a MF matched to the pulse \( p(t) \), as in the AWGN case.

As already mentioned, the computational complexity of ML estimators limits their implementation. To alleviate this problem, several practical suboptimal TOA estimators have been proposed in the literature. For instance, [43] presents a generalized ML-based TOA estimation technique that estimates the relative delay of the first path by assuming that the strongest path is perfectly locked. Different TOA estimation techniques with different levels of complexity, based on channel estimation algorithms [24], are compared using experimental measurements in [112]. Another suboptimum technique can be found in [113].

B. Practical TOA Estimators

A primary barrier to the implementation of ML estimators is that they usually require implementation at the Nyquist sampling rate or higher, and these sampling rates can be impractical due to the large bandwidth of UWB signals. As an alternative, subsampling TOA estimation schemes based on energy detectors (EDs) have recently received significant attention [39], [40], [114]–[121]. These TOA estimation techniques rely on the energy collected at sub-Nyquist rates over several time slots.

We now describe some practical ED-based TOA estimators that have been proposed in the literature by considering a typical signal structure for multiple access in which the generic user \( u \) transmits a packet with a preamble [122]. The preamble (for acquisition, synchronization, and ranging) consists of \( N_{sym} \) symbols (see Fig. 7). Each symbol of duration \( T_s \) is an unmodulated TH signal that is divided into \( N_s \) frames. Each frame of duration \( T_f \) is further decomposed into smaller time slots, called chips, having duration \( T_c \). A single unitary energy pulse \( p(t) \) with duration \( T_p < T_c \) is transmitted in each frame in a position specified by a user-specific pseudorandom TH sequence \( \{ c_{k}^{(u)} \} \) having period \( N_{s} \) [30]. Hence the preamble is composed of \( N_{p} = N_{sym} \cdot N_{s} \) pulses. Without loss of generality, \( u = 1 \) denotes the desired user.

A typical ED-based TOA estimator scheme is shown in Fig. 8. The received signal is first passed through a bandpass zonal filter (BPZF) with center frequency \( f_0 \) and bandwidth \( W \) to eliminate the out-of-band noise. The output of the BPZF can be written as

\[ r_{ZF}(t) = s_{ZF}(t) + i(t) + n(t) \]  

(41)
where

$$s_{ZF}(t) = \sum_{n=0}^{N_t-1} w^{(u)}_n (t - c_n^{(1)} T_c - n T_f)$$  \hspace{1cm} (42)$$

and

$$w^{(u)}(t) = \sqrt{E^{(u)}_0} \sum_{l=1}^{L} \alpha_l^{(u)} p(t - \tau_l^{(u)}).$$  \hspace{1cm} (43)$$

The parameters \( \{\tau_1^{(u)}, \tau_2^{(u)}, \ldots, \tau_L^{(u)}, \alpha_1^{(u)}, \alpha_2^{(u)}, \ldots, \alpha_L^{(u)}\} \) in (43) represent the delays and path gains associated with the \( u \)th user. The TOA to be estimated is \( \tau = \tau_1^{(1)} \). We consider for all users \( \sum_{l=1}^{L} E\{\alpha_l^{(u)}\} = 1 \) and hence \( E^{(u)}_0 \) represents the average received symbol energy from user \( u \). The component \( i(t) \) represents the interfering term whose expression depends on the nature of the interference, and \( n(t) \) represents the thermal noise.

We assume that the receiver has acquired the sequence of the desired user for estimating the TOA \( \tau \) of the first path based on the observation of the received signal \( r_{ZF}(t) \).\textsuperscript{17} The estimator utilizes a portion of \( r_{ZF}(t) \) consisting of \( N_t \) sub-intervals \( I_n = [c_1^{(1)} T_c + n T_f, c_2^{(1)} T_c + n T_f + T_{ob}], \) each of duration \( T_{ob} < T_f \), with \( n = 0, 1, \ldots, N_t - 1 \).

The observed signal forms the input to the ED, whose output is sampled at every \( T_{int} \) seconds (the integration time of the ED); thus \( K = \lfloor T_{ob}/T_{int} \rfloor \) samples are collected in each subinterval (with indexes \( [0,1,\ldots,K-1] \) corresponding to \( K \) time slots). The true TOA \( \tau \) is contained in the time slot \( n_{TOA} = \lfloor \tau/T_{int} \rfloor \). In the absence of other information, the system can assume that \( \tau \) is uniformly distributed in the interval \([0,T_n]\), with \( T_a < T_{ob} \); as a consequence the discrete RV \( n_{TOA} \) is uniformly distributed on the integers \( 0,1,\ldots,N_{TOA}-1 \), where \( N_{TOA} = \lfloor T_n/T_{int} \rfloor \). Note that the first \( n_{TOA} \) samples contain noise and possibly interference (called the noise region), followed by the \( (n_{TOA}+1) \)th sample containing the first path, and the remaining \( K - n_{TOA} - 1 \) samples possibly containing the echoes of the useful signal (the multipath region), in addition to noise and interference.

The integration time \( T_{int} \) of the ED determines the resolution in estimating the TOA, and thus the minimum achievable RMSE on TOA estimation is given by \( T_{int}/\sqrt{12} \).

Prior to the TOA estimation, the collected samples can be preprocessed or filtered in order to reduce the interference effects and to improve the detection of the first path \([114],[115]\). For notational convenience, the \( N_t \times K \) collected samples (\( N_t \) intervals with \( K \) samples each) at the output of the ED can be arranged in an \( N_t \times K \) matrix, denoted by \( V \), with elements

$$v_{n,k} = \int_{kT_{int}}^{(k+1)T_{int}} |r_n(t)|^2 dt.$$  \hspace{1cm} (44)$$

where \( n = 0,\ldots,N_t - 1, k = 0,\ldots,K - 1, \) and \( r_n(t) \) is the portion of the received signal after the dehopping process in the subinterval \( I_n \). We now introduce a generic transformation \( \mathbb{F}[] \) (see Fig. 8), whose output \( z_k = \{z_k\} \) is a vector to be used by the TOA estimator. In general

$$z_k = \mathbb{F}\{v_{n,k}\}, \hspace{1cm} k = 0,\ldots,K - 1.$$  \hspace{1cm} (45)$$

where \( \mathbb{F}[] \) can be either a linear or a nonlinear transformation. A variety of filtering schemes can be adopted. The conventional way to obtain the decision vector \( z \) is through simple column averaging (averaging filter), that is [115]

$$z_k = \sum_{n=0}^{N_t-1} v_{n,k}, \hspace{1cm} k = 0,\ldots,K - 1.$$  \hspace{1cm} (46)$$

In the following, we describe some recently proposed techniques to detect the first path from \( z \).\textsuperscript{18}

\textsuperscript{17}The IEEE 802.15.4a standard proposes a preamble consisting only of amplitude modulated pulses (i.e., \( c_0 = 0 \)) using a length-31 ternary sequence with an ideal periodic autocorrelation function [38], [123]. This preamble enables robust sequence acquisition and allows the use of both coherent and noncoherent TOA estimation.

\textsuperscript{18}For completeness, additional techniques can be found in [124]–[137].
1) Max: The Max criterion is based on the selection of the largest sample in $z$. Specifically, the TOA estimate of the received signal is given by

$$\hat{\tau} = T_{\text{int}} \cdot k_{\text{max}} + \frac{T_{\text{int}}}{2}$$

(47)

where $k_{\text{max}}$ is the time index corresponding to the largest sample (see Fig. 9).

This criterion has the advantage of not requiring extra parameters to be optimized accordingly to the received signal (channel, interference, and noise levels). However, as will be shown in Section VI, it suffers performance degradation when the first path is not the strongest. This typically occurs in NLOS propagation.

2) P-Max: The P-Max criterion is based on the selection of the earliest sample among the $P$ largest in $z$. Specifically, the TOA estimate of the received signal is given by [112]

$$\hat{\tau} = T_{\text{int}} \cdot \min_{i \in \{1,2,\ldots,P\}} \{k_i\} + \frac{T_{\text{int}}}{2}$$

(48)

where $k_1, k_2, \ldots, k_P$ are the time indexes corresponding to the $P$ largest samples (see Fig. 9). As will be shown in Section VI, TOA estimation performance depends on the parameter $P$, where $P$ can be optimized according to received signal characteristics.

3) Simple Thresholding: The simple thresholding (ST) criterion is based on an estimate of $N_{\text{TOA}}$, and hence $\tau$, by comparing each element of $z$ within the observation interval to a fixed threshold $\eta$ [39]. In particular, the TOA estimate is taken as the first threshold crossing event (see Fig. 9). The design of the threshold depends on the received signal characteristics, operating conditions, and channel statistics. The ST criteria can also be used instead of Max search criteria in the MF-based estimator shown in Fig. 3. With the MF ST scheme, the resolution of TOA estimation is not limited by the integration time associated with the ED ST. The main advantage of ST is that it can be implemented completely in analog hardware, and this is particularly attractive for low-cost battery-powered devices such as WSNs [2], [36], [39], [41], [138], [139].

The choice of the threshold $\eta$ strongly influences the performance of ST TOA estimation. When $\eta$ is small, we expect a high probability of early detection prior to the first path due to noise and interference. On the other hand, if $\eta$ is large, we expect a low probability of detecting the first path and a high probability of detecting an erroneous path due to fading. In [36], a simple criterion to determine the suitable threshold based on the evaluation of the probability of early detection $P_{\text{ed}}$ is proposed, which is given by

$$P_{\text{ed}} = 1 + \left(1 - q_o\right)^{N_{\text{TOA}}} - 1$$

(49)

where

$$q_o \triangleq \exp\left(-\frac{T_{\text{NR}}}{T_0} \sum_{i=0}^{M/2-1} \frac{T_{\text{NR}}}{i!}\right)$$

(50)

with $T_{\text{NR}} \triangleq \eta/N_0$ and $M = 2N_0T_{\text{int}}W$. This expression can be used to determine the threshold $\eta$ through $T_{\text{NR}}$, corresponding to a target $P_{\text{ed}}$. While this choice of $\eta$ does not require knowledge of the received signal characteristics, it was shown in [36] that this methodology does not lead to a significant performance degradation with respect to the optimum choice of $\eta$.

4) Jump Back and Search Forward: The jump back and search forward (JBSF) criterion is based on the detection of the strongest sample and a forward search procedure [140].
It assumes that the receiver is synchronized to the strongest path and the leading edge of the signal is searched element-by-element in a window of length \( W_{sb}+1 \) containing \( W_{sb} \) samples preceding the strongest one. The search begins from the sample in \( z \), with index \( k_{\text{max}} - W_{sb} \), and the search proceeds forward until the sample-under-test crosses the threshold \( \eta \) (see Fig. 9). In particular, the TOA estimate is given by

\[
\hat{\tau} = T_{\text{int}} \cdot \min \{k \in \{k_{\text{max}} - W_{sb}, k_{\text{max}} - W_{sb} + 1, \ldots, k_{\text{max}}\} | z_k > \eta \} + \frac{T_{\text{int}}}{2}. \tag{51}
\]

Note that the optimal selection of \( W_{sb} \) and \( \eta \) depends on the received signal characteristics.

5) Serial Backward Search: The serial backward search (SBS) criterion is based on the detection of the largest sample and a search-back procedure. The search begins from the largest sample in \( z \), with index \( k_{\text{max}} \), and the search proceeds element by element backward in a window of length \( W_{sb} \) until the sample under test goes below the threshold \( \eta \) (see Fig. 9). In particular, the TOA estimate of the received signal is given by

\[
\hat{\tau} = T_{\text{int}} \cdot \max \{k \in \{k_{\text{max}}, k_{\text{max}} - 1, \ldots, k_{\text{max}} - W_{sb}\} | z_k > \eta \text{ and } z_{k-1} < \eta \} + \frac{T_{\text{int}}}{2}. \tag{52}
\]

Some criteria to calculate suitable values of \( \eta \) and \( W_{sb} \) are presented in [140].

6) Serial Backward Search for Multiple Clusters: In typical UWB channels, multipath components arrive at the receiver in multiple clusters that are separated by noise-only samples. In this case, the SBS algorithm may choose a sample that arrives later than the leading edge. This clustering problem can be alleviated by continuing the backward search until more than \( D \) consecutive noise samples are encountered. In particular, this TOA estimate is given by

\[
\hat{\tau} = T_{\text{int}} \cdot \max \{k \in \{k_{\text{max}}, \ldots, k_{\text{max}} - W_{sb}\} | z_k > \eta \text{ and } \max \{z_{k-h} \cdot \ldots \cdot z_{k_{\text{max}}-D, k_{\text{max}}-W_{sb}}\} < \eta \} + \frac{T_{\text{int}}}{2}. \tag{53}
\]

This is referred to as the serial backward search for multiple clusters (SBSMC). Some criteria to calculate suitable values of \( \eta, W_{sb}, \) and \( D \) are presented in [141].

VI. PERFORMANCE OF PRACTICAL TOA ESTIMATORS

We now investigate the performance of practical TOA estimators in the presence of multipath and interference, and with bandwidth limitations.

A. The Effect of Multipath

To compare the performance of TOA estimators described in Section V, we consider a TH signal with RRC bandpass pulse (21) in an IEEE 802.15.4a CM4 channel [28]. We begin with the ST TOA estimator in the absence of interference.\(^{19}\) In Fig. 10, the RMSE as a function of SNR is shown for MF ST based and ED ST based estimators with optimal values of the threshold \( \eta \) (i.e., for each SNR, the value of \( \eta \) is chosen such that the RMSE is minimized). The SNR is defined as \( \text{SNR} \triangleq E_s / N_0 \) and the transmission of a single pulse in the absence of interference (i.e., \( N_{\text{sym}} = 1, N_s = 1 \)) as well as multiple pulses (i.e., \( N_{\text{sym}} = 400, N_s = 4 \)) are considered. The MF ST estimator achieves better accuracy than the ED ST for high SNRs. This becomes more evident when the preamble length \( N_s \cdot N_{\text{sym}} \) increases due to the fact that MF ST based estimator performs coherent summation of the energy of each pulse, at the contrary of the ED ST estimator. One can also observe that for low SNRs, the TOA estimation error

\(^{19}\)The RMSE of TOA estimation is evaluated through Monte Carlo simulation. The corresponding RMSE in distance estimation can be simply obtained by multiplying that of TOA by the speed of light \( c \).
approaches $T_n/\sqrt{12}$. As expected, for high SNRs the ED ST estimator performance has a floor equal to $T_{\text{int}}/\sqrt{12}$, which is due to the time resolution that is dependent on the integration interval. On the other hand, the floor exhibited by the MF ST estimator is due to template mismatching, that is, due to the partial overlap of unresolved pulses in the IEEE 802.15.4a channel model.

The RMSE as a function of the SNR for MF and ED ST estimators with optimum threshold is shown in Fig. 11 for LOS and NLOS conditions using data measured in indoor residential environments. Note that under LOS conditions, the MF ST estimator can achieve RMSE of 0.1 ns for SNR > 15 dB, while ED ST estimator can achieve RMSE of 1 ns for SNR > 20 dB. On the other hand, in NLOS conditions, the overall performance of both estimators degrades and the MF ST estimator achieves RMSE of 0.1 ns only when SNR > 36 dB.

We now compare the performance of the ED TOA estimators described in Section V-B when using an ideal BPZF with bandwidth $W = 1.6$ GHz. Fig. 12 shows the RMSE as a function of the SNR for Max, P-Max, ST, JBSF, SBS, and SBSMC. Recall that all these schemes, except Max, require an optimal selection of parameters (specifically $\eta$, $W_{\text{th}}$, $D$, or $P$) depending on received signal characteristics. In the ST scheme, this problem can be avoided by adopting the suboptimal criteria given in (49) and setting $P_{\text{ed}} = 10^{-4}$ [36]. The best scheme for TOA estimation depends on operating SNRs. However, one can compare them at high SNRs since all schemes exhibit error floors. Note that the ST and JBSF schemes achieve the minimum attainable error floor of ED-based TOA estimators, that is, $T_{\text{int}}/\sqrt{12} = 0.57$ ns. The SBS and SBSMC schemes exhibit an increased error floor due to the search back procedure that can stop after the first path in a clustered channel. The Max and P-Max schemes result in higher error floor due to the fact that, even in the presence of negligible noise, they identify the strongest paths instead of the first path, which often is not the strongest one.

B. The Effect of Interference

The performance of TOA estimators is affected by both NBI and MUI [142]. Despite this, relatively few publications address such effects on TOA estimation. A technique based on nonlinear filtering of received samples is proposed in [114] to mitigate the effect of MUI on ranging. The joint effect of NBI and MUI was considered in [115], where a technique to mitigate both of these effects is proposed.

As mentioned in Section III-C, a single NBI can be well approximated using (12). In this case, we define the interference-to-noise ratio (INR) as $\text{INR} \equiv IT_0/N_0$.

When $U$ UWB interferers are present, the MUI contribution to $i(t)$ can be written as

$$i(t) = \sum_{u=2}^{U} \sum_{n=0}^{N_u-1} w^{(u)} \left( t - c^{(u)}_n T_c - nT_i \right).$$

Typical energy matrix $V$ in (44) is plotted in Fig. 13(a), which shows the energy collected at the output.
of the ED in the presence of both NBI and MUI (the reader is reminded that \( N_t \) is the number of pulses in the preamble and \( K \) is the number of time slots in each observation subinterval). It can be seen that the energy samples corresponding to the desired signal are time aligned and partially buried under a floor caused by NBI and noise. This floor may increase the probability of early detection and thus degrades the TOA estimator performance. Note that, unlike the desired signal, the energy samples corresponding to MUI are not time aligned, resulting in impulsive behavior, due to the use of TH technique.

Fig. 13(a) suggests the use of filtering techniques to reduce the effects of interference on TOA estimation. To reduce the impulse behavior of MUI, nonlinear processing can be applied to each column of \( V \). In particular, the use of a \textit{min filter} gives

\[
Z_k = \sum_{n=0}^{N_t-H} \min \{ v_{n,k}, v_{n+1,k}, \ldots, v_{n+H-1,k} \} \quad (55)
\]

where \( k = 0, \ldots, K-1 \), and \( H \) is the length of the filter \[114\]. In scenarios where both NBI and MUI are present, \[115\] proposed to adopt a double-filtering scheme. Specifically, \textit{min filter} is applied to each matrix column as in (55), whose outputs \( \{ Z_k \} \) are then further processed by a \textit{differential filter} giving

\[
Z_k = \tilde{Z}_k - \tilde{Z}_{k+1}, \quad k = 0, \ldots, K-1 \quad (56)
\]

with \( \tilde{Z}_k = 0 \). The purpose of (56) is to reduce the presence of floors, typically caused by NBI, and to emphasize the beginning of the multipath cluster.\footnote{Differential filtering is typically used in image processing to emphasize the presence of edges.} The energy matrix at the output of the differential filter is shown in Fig. 13(b). The ability of this scheme to mitigate both NBI and MUI is demonstrated by comparing Fig. 13(a) and (b).

Various filtering techniques are compared in Fig. 14 in terms of RMSE of an ED ST estimator in the absence and in the presence of NBI and MUI using the IEEE 802.15.4a CM4 channel \[28\]. In particular, the NBI is modeled as a tone with frequency \( f_I = 3.5 \text{ GHz} \) subject to Rayleigh fading. To model MUI, we consider an additional user and define the signal-to-interference ratio (SIR) as

\[
\text{SIR} = \frac{E(1)}{E(2)} \quad \text{s} = \frac{E(3)}{E(4)}
\]

It can be seen that the performance of single averaging, \textit{min}, and \textit{differential filters} drastically degrades in the presence of NBI and MUI. On the contrary, the cascaded \textit{min} and \textit{differential filter} provides substantial performance improvements with respect to other schemes due to its capability to mitigate the combination of both types of interference.

C. The Effect of Bandwidth

It has been discussed in Section IV that theoretical performance limits on TOA estimation can be increased by adopting signals with larger bandwidths. However, a bandwidth increase beyond certain values may cease to be advantageous depending on the specific system parameters and operating environments. The effect of transmission bandwidth on the accuracy of UWB ranging systems is investigated in \[143\]–\[145\]. In particular, in scenarios with undetected first paths, an optimum transmission bandwidth exists that minimizes the ranging error \[143\]. In \[144\], the relationship between transmission bandwidth and ranging accuracy is investigated using experimental data collected in various apartments. This work showed
that while ranging accuracy increases significantly with bandwidths up to 2.5 GHz, these improvements become less significant above 5 GHz. Similar experimental results are presented in [145].

TOA estimator structure and system parameters may affect ranging accuracy regardless of the transmission bandwidth. For example, the RMSE of ED-based TOA estimators experiences a theoretical floor equal to $\frac{T_{\text{int}}}{\sqrt{12}}$. Therefore, even with the use of larger transmission bandwidths, improvements in ranging accuracy may not be significant if the estimator is poorly designed.

VII. FURTHER RESEARCH DIRECTIONS

Ranging in UWB channels has experienced a flurry of research in recent years. However there still remain multiple areas of open research that will help systems to meet the requirements of HDSA applications. The rapid deployment of location-based networks, for which accurate ranging is likely pivotal, will be aided by the development of the following techniques.

- **Interference mitigation:** NBI and MUI are inevitably present in most systems. To date, however, the majority of research effort ignores the effects of interference on TOA estimation accuracy, and few papers propose robust interference mitigation techniques [115]. There are opportunities for further research in robust interference-resistant ranging, such as the design of jam-resistant schemes.

- **Secure ranging:** In certain scenarios ranging may be subject to hostile attacks. While some works have presented secure localization algorithms (see, e.g., [146][149]), less attention has been paid to secure ranging. In order to make impostor and snooper attacks more difficult, the IEEE 802.15.4a standard includes an optional private ranging mode: using a secure communication protocol, nodes exchange information, after a preliminary authentication step, on both the sequences to be used in the next ranging cycle as well as their timestamp reports.

- **Cognitive ranging and localization:** Cognitive radio is a promising technology for efficient utilization of the spectrum due to its capability to sense environmental conditions and adapt its communication and localization techniques [3], [150], [151]. There are opportunities for further research in cognitive positioning systems, where the localization accuracy can be varied according to bandwidth availability [152], [153]. In general, UWB signals are well suited for cognitive radio that aims to improve the coexistence between systems operating in common frequency bands [4], [154].

VIII. CONCLUSION

UWB signals have the potential for a wide range of new HDSA systems. Numerous applications of UWB ranging includes logistics, security, healthcare, search and rescue, control of home appliances, automotive safety, and military applications. Ubiquitous deployment of these systems will present new research directions and challenges in terms of spectral usage, radio propagation, low-complexity architectures, and energy-efficient design. One of the underpinning requirements for the deployment of the HDSA systems is accurate, robust ranging, and this requires understanding the fundamental limits of ranging so that low-complexity techniques can be developed to approach these limits.

This paper presented a survey of time-base ranging UWB signals in multipath environments. The basic concepts of TOA estimation in ideal and realistic conditions are explained. An investigation into the theoretical performance limits is outlined and the main sources of error in TOA estimation, such as multipath, interference, and clocks drift, are discussed. Furthermore, practical TOA estimation schemes are illustrated and their performance compared using IEEE 802.15.4a channel models as well as measured data.

Acknowledgment

The authors would like to thank O. Andrisano, M. Chiani, C.-C. Chong, D. Jourdan, M. Mazzotti, and Y. Shen for helpful discussions.
REFERENCES


[125] I. Guvenc, Z. Sahinoglu, F. M. Gharsa, and P. Orlik, “Non-coherent TOA estimation in IR-UWB systems with different signal...
Dardari et al.: Ranging With Ultrawide Bandwidth Signals in Multipath Environments

ABOUT THE AUTHORS

Davide Dardari (Senior Member, IEEE) is an Associate Professor at the University of Bologna at Cesena, Cesena (FC), Italy. He is a Research Affiliate with WiLAB and the Wireless Communications Group, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology (MIT), Cambridge. His current research interests are in ultrawideband communication and localization, wireless sensor networks, and OFDM systems. He has published more than 100 technical papers and played several important roles in various national and European projects. He was Lead Editor for the EURASIP Journal on Advances in Signal Processing (Special Issue on Cooperative Localization in Wireless Ad Hoc and Sensor Networks). He was a Guest Editor for the Physical Communication Journal (Special Issue on Advances in Ultra-Wideband Wireless Communications). He is a coauthor of Wireless Sensor and Actuator Networks: Enabling Technologies, Information Processing and Protocol Design (Amsterdam, The Netherlands: Elsevier, 2008). He is a Reviewer for TRANSACTIONS/Journals and conferences, and a Technical Program Committee member for numerous international conferences.

Prof. Dardari is the current Vice-Chair for the Radio Communications Committee of the IEEE Communication Society. He was Cochair of the Wireless Communications Symposium of IEEE ICC 2007 and of the IEEE International Conference on UWB (ICUWB) 2006. He is a Guest Editor for the PROCEEDINGS OF THE IEEE (Special Issue on UWB Technology and Emerging Applications). Currently, he is an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Andrea Conti (Member, IEEE) received the Dr.ing. degree (with honors) in telecommunications engineering and the Ph.D. degree in electronic engineering and computer science from the University of Bologna, Bologna, Italy, in 1997 and 2001, respectively.

From 1999 to 2002, he was with the Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), University of Bologna. In November 2002, he joined the Istituto di Elettronica e di Ingegneria dell’Informazione e delle Telecomunicazioni, National Research Council, Research Unit of Bologna (IEIIT-BO/CNR). In 2005, he joined the University of Ferrara, Ferrara, Italy, where he is currently an Assistant Professor. In summer 2001, he joined the Wireless Section of AT&T Research Laboratories, Middletown, NJ, working on the performance of digital telecommunication systems with diversity reception. From February 2003, he visited the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, working on the performance of adaptive and nonadaptive multilevel quadrature amplitude modulation (M-QAM) in faded multi-channel reception and localization techniques. He is a Research Affiliate with WiLAB, IEIIT-BO/CNR, CNIT, and the Wireless Communications Group at LIDS. His current research interests are in the area of wireless communications, including localization, adaptive transmission and multichannel reception, coding in faded multiple-input multiple-output channels, wireless cooperative networks, wireless sensor networks, and cooperative distributed measurement laboratories. He is Lead Editor for the EURASIP Journal on Advances in Signal Processing (Special Issue on Wireless Cooperative Networks). He is a coauthor of Wireless Sensor and Actuator Networks: Enabling Technologies, Information Processing and Protocol Design (Amsterdam, The Netherlands: Elsevier, 2008).

He is an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Ulric Ferner (Student Member, IEEE) received the B.Eng. degree (with highest honors) in electrical engineering from the University of Auckland, New Zealand, in 2006. He is currently pursuing the Ph.D. degree at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge.

His main research interests are in wireless communications, specifically localization, ultra-wide bandwidth systems, and cognitive radio.

In 2007, Mr. Ferner received a Fulbright Scholarship and the James Gordon Goodfellow Award from the University of Auckland.
Andrea Giorgetti (Member, IEEE) received the Dr.Ing. degree (magna cum laude) in electronic engineering and the Ph.D. degree in electronic engineering and computer science from the University of Bologna, Bologna, Italy, in 1999 and 2003, respectively.

Since 2003, he has been with the Istituto di Elettronica e di Ingegneria dell’Informazione e delle Telecomunicazioni (IEIIT-BO) research unit at Bologna, National Research Council (CNR), Bologna. In 2005, he was a Researcher with the National Research Council. Since 2006, he has been an Assistant Professor at the II Engineering Faculty, University of Bologna, where he joined the Department of Electronics, Computer Sciences and Systems. During the spring of 2006, he was a Research Affiliate with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, working on coexistence issues between ultra-wideband (UWB) and narrowband wireless systems. His research interests include ultrawide bandwidth communication systems, wireless sensor networks, and multiple-antenna systems.

Dr. Giorgetti was Cochair of the Wireless Networking Symposium, IEEE International Conference on Communications (ICC 2008), Beijing, China, May 2008; and is Cochair of the MAC track of the IEEE Wireless Communications and Networking Conference (WCNC 2009), Budapest, Hungary, April 2009.

Moe Z. Win (Fellow, IEEE) received both the Ph.D. in electrical engineering and M.S. in applied mathematics as a Presidential Fellow at the University of Southern California (USC) in 1998. He received an M.S. in electrical engineering from USC in 1989, and a B.S. (magna cum laude) in Electrical Engineering from Texas A&M University in 1987.

Dr. Win is an Associate Professor at the Massachusetts Institute of Technology (MIT). Prior to joining MIT, he was at AT&T Research Laboratories for five years and at the Jet Propulsion Laboratory for seven years. His research encompasses developing fundamental theory, designing algorithms, and conducting experimentation for a broad range of real-world problems. His current research topics include location-aware networks, time-varying channels, multiple antenna systems, ultra-wide bandwidth systems, optical transmission systems, and space communications systems.

Professor Win is an IEEE Distinguished Lecturer and an elected Fellow of the IEEE, cited for “contributions to wideband wireless transmission.” He was honored with the IEEE Eric E. Sumner Award (2006), an IEEE Technical Field Award for “pioneering contributions to ultra-wide band communications science and technology.” Together with students and colleagues, his papers have received several awards including the IEEE Communications Society’s Guglielmo Marconi Best Paper Award (2008) and the IEEE Antennas and Propagation Society’s Sergei A. Schelkunoff Transactions Prize Paper Award (2003). His other recognitions include the Laurea Honoris Causa from the University of Ferrara, Italy (2008), the Technical Recognition Award of the IEEE ComSoc Radio Communications Committee (2008), Wireless Educator of the Year Award (2007), the Fulbright Foundation Senior Scholar Lecturing and Research Fellowship (2004), the U.S. Presidential Early Career Award for Scientists and Engineers (2004), the AIAA Young Aerospace Engineer of the Year (2004), and the Office of Naval Research Young Investigator Award (2003).

Professor Win has been actively involved in organizing and chairing a number of international conferences. He served as the Technical Program Chair for the IEEE Wireless Communications and Networking Conference in 2009, the IEEE Conference on Ultra Wideband in 2006, the IEEE Communication Theory Symposia of ICC-2004 and Globecom-2000, and the IEEE Conference on Ultra Wideband Systems and Technologies in 2002; Technical Program Vice-Chair for the IEEE International Conference on Communications in 2002; and the Tutorial Chair for ICC-2009 and the IEEE Semiannual International Vehicular Technology Conference in Fall 2001. He was the chair (2004–2006) and secretary (2002–2004) for the Radio Communications Committee of the IEEE Communications Society. Dr. Win is currently an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He served as Area Editor for Modulation and Signal Design (2003–2006), Editor for Wideband Wireless and Diversity (2003–2006), and Editor for Equalization and Diversity (1998–2003), all for the IEEE TRANSACTIONS ON COMMUNICATIONS. He was Guest-Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Special Issue on Ultra-Wideband Radio in Multiaccess Wireless Communications) in 2002.