Machine Learning for Time-of-Arrival Estimation With 5G Signals in Indoor Positioning

Zhaoliang Liu, Liang Chen, Xin Zhou, Zhenhang Jiao, Guangyi Guo, and Ruizhi Chen

***Abstract*—Location-based service in the indoor environment is playing a crucial role in different application scenarios. The introduction of technologies, such as ultradense network and massive multiple-input multiple-output enables fifth-generation (5G) cellular signals, as a new generation of cellular network signals, to show unique advantages in indoor positioning. This article describes 5G reference signal structures that can be used for navigation. A high-precision time-of-arrival estimation method based on 5G downlink signal is proposed that can be realized by edge computing. A software-defined receiver (SDR) based on machine learning to extract navigation observations from 5G signals is then developed. In simulation, the error sources of SDR in additive white gaussian noise channel and multipath channel were analyzed, and the possible ranging accuracy achieved by 5G signals in the developed SDR was evaluated. In field experiments, commercial 5G signals deployed by operators were collected, and the performance of SDR in practical applications was evaluated. The feasibility in practical applications of the proposed SDR is demonstrated, and high pseudorange measurement accuracy can be achieved.**

***Index Terms*—Discriminator function, fifth-generation, Kalman filter (KF), machine learning (ML), time-of-arrival (TOA).**

I. Introduction

With the advancement of technology and economic development, location-based service (LBS) is playing an increasingly important role in people's daily life, social public services, emergency prevention, and other fields [1], [2]. At the same time, LBS is also a major component of the Internet of Things (IoT) [3]. Accurate location information is the basis for various IoT application systems to achieve service functions. Traditional LBS mostly provides users with location information in the outdoor environment via the global navigation satellite system (GNSS) [4], [5]. However, emerging IoT services, such as indoor emergent rescue, precision marketing in shopping malls, indoor crowd monitoring in epidemics, virtual reality games, and LBS social media have a more urgent demand for location information in the indoor environments. Moreover, GNSS is limited by the application scenarios, so providing ubiquitous and high-precision LBS to users in the indoor environment is difficult. The indoor positioning technology is a technical means designed to provide users with high-precision LBS in GNSS-restricted areas [6]. Based on the characteristics of the Signal of Opportunity (SoP) acquired by a user in the indoor environment, the accurate location information of the user can be obtained.

Presently, the widely used SoP in indoor positioning mainly includes WiFi [7], [8], [9], [10], [11], Bluetooth [12], [13], [14], and cellular wireless network signals [15], [16], [17], [18]. Among them, WiFi and Bluetooth have received more attention because of their lower cost and power consumption. For example, fingerprint positioning based on the received signal strength indication (RSSI) [7], [8] feature of WiFi signals has been used commercially in numerous fields because of the simplicity of the method and ease of implementation. In [9] and [10], the channel state information of WiFi signals has also been used in indoor positioning research because it contains more fingerprint information. In [12] and [13], Bluetooth signals have achieved satisfactory positioning results in the research of indoor positioning because of low cost and power consumption. Although these indoor positioning methods based on signal fingerprint features can achieve high-precision positioning in a certain area, the applicability of the positioning method based on signal fingerprint characteristics is poor. In addition, substantial manpower is required to complete the establishment and update of the fingerprint database of the positioning system. In [11] and [14], the ranging positioning method based on WiFi or Bluetooth signal path loss model is investigated, but this method still requires preliminary data acquisition for model training to improve the ranging accuracy of the system in different scenarios.

In contrast with WiFi or Bluetooth signals, cellular wireless network signals were first used in the long-term evolution (LTE) era for research into the indoor positioning technology. Although the fingerprint positioning methods based on LTE signals are investigated in [19] and [20], high-precision tracking and ranging methods based on LTE signals are more interesting to researchers because LTE signals incorporate a stricter time synchronization on the base stations (BSs) than other SoP. In [15] and [16], researchers developed several high-precision software-defined receivers (SDRs) that can

Manuscript received 2 April 2022; revised 15 December 2022; accepted 1 January 2023. Date of publication 4 January 2023; date of current version 23 May 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42171417; in part by the Special Fund of Hubei Luojia Laboratory under Grant 220100008; in part by the Central Guided Local Science and Technology Development Special Project under Grant 20222ZDH04090; in part by the Medium- and Long-Term Scientific and Technological Planning Projects for Radio, Television and Audio-Visual Networks; and in part by the Key Research and Development Program of Hubei Province under Grant 2021BAA166. *(Corresponding author: Liang Chen.)*

The authors are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, Hubei, China (e-mail: zhaoliang.liu@whu.edu.cn; l.chen@whu.edu.cn; x.zhou.whu@whu.edu.cn; zhenhang.jiao@whu.edu.cn; guangyi.guo@whu.edu.cn; ruizhi.chen@whu.edu.cn).

Digital Object Identifier 10.1109/JIOT.2023.3234123

be used in the complex urban environment based on the time-of-arrival (TOA) estimation of LTE signals. In particular, the multipath delay estimation algorithm for LTE signals in [17] and [18] reduces the impact of multipath effects on TOA estimation accuracy in the indoor environment to improve the accuracy and stability of the delay estimation.

With the emergence and commercial application of fifth-generation (5G) signals, the latest generation of cellular network signals, the number of 5G network users worldwide has reached 220 million by the end of 2020. The introduction of the massive multiple-input multiple-output (MIMO) and ultradense network (UDN) technology enables 5G signals to show unique advantages in the indoor positioning technology compared with other SoP. Massive MIMO can provide high-precision pseudorange measurement information for trilateral positioning in the indoor environment to estimate user's location. In addition, the introduction of the UDN technology ensures that enough line-of-sight (LOS) signals can be received in the indoor environment to complete trilateral positioning. In [21] and [22], the researchers constructed offline databases based on the feature of 5G signals and implemented fingerprint positioning in the indoor environments based on deep learning methods, which achieved stable positioning results in their field tests. Shamaei and Kassas [23] constructed an SDR for TOA estimation by combining delay-locked loop (DLL) and phase-locked loop (PLL) to track 5G signals in an outdoor urban environment, which achieved the standard deviation (SD) of the range error in 1.19 m. In [24], a joint angle and delay estimation scheme is designed based on the antenna arrays and large bandwidth technology for 5G signals, which achieves a triangulation positioning error of 0.44 m for 90% cases. Chen et al. [25] developed an SDR platform and proposed a method of carrier phase ranging with real 5G downlink reference signal for indoor positioning. The field tests verified the method with high-precision performance even with multipath interference, which reflects 5G signals are promising for indoor navigation.

In this article, machine learning (ML) algorithms are introduced into the TOA estimation of 5G signals to achieve the delay and phase tracking of the signal through regression classification of the received pilot signal discriminator function. Compared with the fingerprint positioning method, this trilateral positioning without pretraining can allows for a more extensive spread of the indoor positioning technology. Complex ML algorithms are also not required in the developed SDR. As shown in Fig. 1, the use of typical lightweight ML algorithms allows SDR to not only meet the accuracy requirements of LBSs but also has the potential to enable edge computing in the majority of mobile terminals [26]. At the same time, because the developed SDR uses 5G downlink signals to extract navigation observations, personal information is not exchanged between the user and the BSs, which can better ensure user information security [27], [28]. In the continuous tracking of 5G signals, we use the Kalman filtering (KF) to transform discrete ML classification results into continuous tracking results. The addition of the KF not only improves the accuracy of SDR for signal tracking and the stability of pseudorange measurements but also weakens the interference of multipath. The specific contributions of this article are summarized as follows.

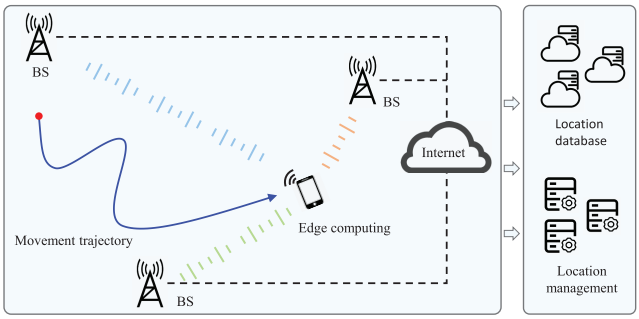


Fig. 1. Schematic block diagram of 5G positioning based on ML that can be realized through edge computing.

1) This article uses the *S*-curve of 5G signal as feature to develop a method for TOA estimation by introducing ML algorithms combined with KF, which achieves high-precision and stable signal tracking in the form of an SDR without changing the structure of the hardware equipment.
2) The SDR developed in this article extracts navigation observations for mobile device terminals via 5G downlink signals, which can also reduce the computational pressure on the BSs while ensuring user privacy and data security.
3) The accuracy of the developed SDR under the additive white Gaussian noise (AWGN) channel and the multipath channel is tested through simulation experiments to verify the effectiveness of SDR. The field experiments are carried out to use commercial 5G downlink signals in the indoor environment, and they achieve relatively well pseudorange measurement results.

The remainder of this article is organized as follows. Section II introduces the 5G standard and signal model. Section III puts forward the TOA estimation of 5G signals based on ML algorithms in this article. Section IV is the analysis of system errors. Section V presents the field experiments and the related result analysis. Section VI is the conclusion of this article.

II. 5G Standard and Signal Model

As a new generation of cellular network signals, 5G signals consist of downlink and uplink signals to enable communication between user equipment (UE) and BSs. The acquisition of accurate location information, protection of user location security, and rational use of computing resources are all important components to consider in the area of wireless positioning. In this article, we aim to develop a method for TOA estimation using 5G downlink signals, which not only secure the location information of the user but can also be implemented by means of edge computing to reduce computational resources of the BS [26], [29]. This section mainly introduces the commercial 5G downlink signal protocol and signal transmission model that can be used for wireless positioning on the FR1 frequency band [30], [31], [32].

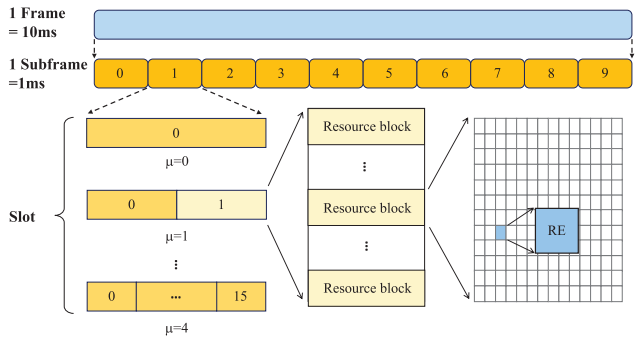


Fig. 2. Schematic of the frame structure of 5G signals.

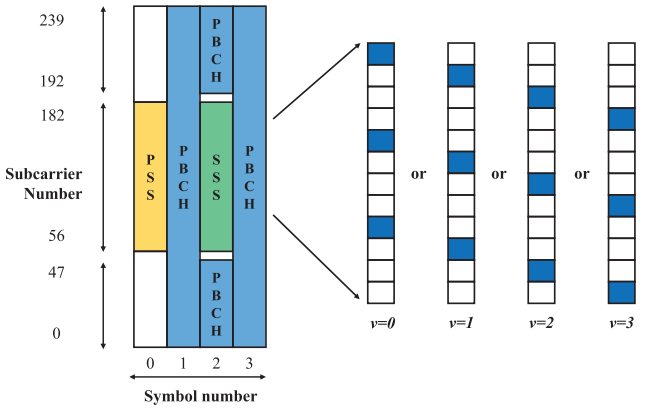


Fig. 3. Structure of the SSB.

A. Overview of 5G Standard From the Perspective of Wireless Positioning

The 5G downlink transmission is based on orthogonal frequency division multiplexing (OFDM) modulation with cyclic prefixes (CPs). This modulation not only reduces mutual interference between subcarriers in the signal and improves spectrum utilization, but also weakens the inter-symbol interference caused by multipath delay spreading. A schematic of the frame structure of 5G signals is shown in Fig. 2. The duration of a 5G frame is 10 ms, and each frame is composed of ten subframes with a duration of 1 ms. In the time domain, although each time slot of 5G signal still consists of 14 OFDM symbols, the number of time slots contained in each subframe can be flexibly defined depending on the different numerologies $\mu \in \{0, \ldots, 4\}$. In this article, the 5G signals in the simulation experiments and commercial 5G field experiments involved are for the case of $\mu = 1$, i.e., each subframe consists of two-time slots with a subcarrier spacing of 30 kHz. In the frequency domain, each subframe can be divided into numerous resource grids, each of which has multiple resource blocks (RBs) with 12 subcarriers. An RB can be divided into smaller elements, known as the resource element (RE), which is the smallest element in the 5G frame structure.

The synchronization signal and PBCH block (SSB) is an important pilot channel in the 5G downlink signals, and the signals carried by it can be used for wireless positioning. The SSB is composed of primary synchronization signals (PSSs), secondary synchronization signals (SSSs), and physical broadcast channel (PBCH). As shown in Fig. 3, the SSB occupies four OFDM symbols in the time domain and N ($N = 240$) subcarriers in the frequency domain which correspond to the numbers $\{0, 1, \ldots, 239\}$. The PSS is located on 127 subcarriers in the middle of the 0th symbol of SSB, and the SSS is located on 127 subcarriers in the middle of the second symbol of SSB. When 5G signals are received, the UE must first perform cell search, to obtain synchronization information, and complete OFDM demodulation before the effective information transmitted in the signal can be extracted. The PSS and SSS are orthogonal m-sequences of length 127. For the signals transmitted by all the 5G BSs, there are three known possible PSSs, corresponding to $N_{\mathrm{ID}}^{(2)} = (0, 1, 2)$. In addition, there are 336 known possible SSSs, corresponding to $N_{\mathrm{ID}}^{(1)} = \{0, 1, \ldots, 335\}$. The $N_{\mathrm{ID}}^{(1)}$, $N_{\mathrm{ID}}^{(2)}$, and the frame start position of the received signal can be obtained through two steps: the PSS detection and the SSS detection. The cell ID of the received signal is calculated according to

$$N_{\mathrm{ID}}^{cell} = 3N_{\mathrm{ID}}^{(1)} + N_{\mathrm{ID}}^{(2)}. \tag{1}$$

The demodulation reference signal (DMRS) is in the first and third symbols of the SSB, on the subcarrier with the sequence number $d = [g\Delta p + \nu]$. Where $g = \{0, 1, \ldots, G-1\}$, G ($G = 60$) is the total number of all DMRS on one symbol of the SSB, and Δp ($\Delta p = 4$) is the subcarrier distance of two DMRS. The DMRS has four different starting positions $\nu \in \{0, 1, 2, 3\}$ on one OFDM symbol, which can be determined based on the cell ID and the transmit antenna serial number. As shown in Fig. 3, the state changes of the signal during transmission can be better reflected owing to the uniform arrangement of DMRS in the frequency domain. Therefore, DMRS can be used not only for conventional channel estimation but also for high-precision continuous tracking in positioning systems.

B. Signal Model

The OFDM symbol on the 5G signal SSB composes of N subcarriers. Let $\{t(n)|n = 0, \ldots, N-1\}$ denote the subcarrier symbol, where n represents the subcarrier number. After the inverse fast Fourier transform (FFT) operation for every OFDM symbol, the samples of the transmitted baseband signal can be expressed as follows:

$$s(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} t(n) e^{j2\pi kn/N}, -N_{cp} \le k \le (N-1) \tag{2}$$

where $j = \sqrt{-1}$, N_{cp} is the number of guard samples and k is the number of subcarriers.

During the transmission of a 5G signal, $s(k)$ changes in amplitude, delay, and phase owing to the continuous increase of the transmission distance. The 5G signal received at the receiver is also affected by noise. Here, we consider that the signal is transmitted on a frequency-selective fading channel of length M, where $\{m = 0, \ldots, M-1\}$. Hence, the received

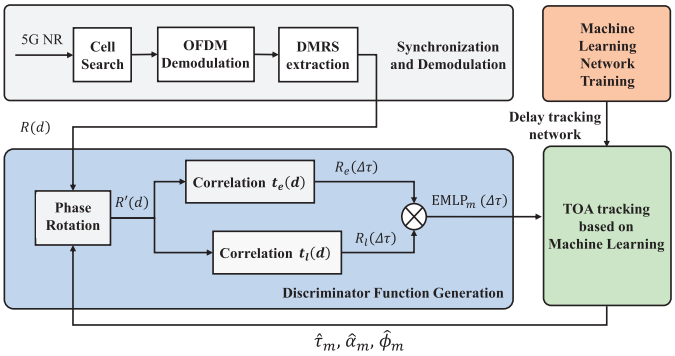


Fig. 4. Diagram of the system flow.

baseband signal can be written as follows:

$$r(k) = \sum_{m=0}^{M-1} r_m(k) = \sum_{m=0}^{M-1} \alpha_m(k) \cdot s(k-\tau_m) e^{j\phi_m(k)} + n(k) \tag{3}$$

where $\alpha_m(k)$, τ_m, and $\phi_m(k)$ stand for the amplitude, delay, and phase of the mth path received baseband signal. $n(k)$ is the sample of zero-mean complex Gaussian noise process with variance σ^2. The phase $\phi_m(k)$ can be written as follows:

$$\phi_m(k) = 2\pi k f_m / N + \phi_0 \tag{4}$$

where f_m is the Doppler frequency normalized by the subcarrier spacing of mth path signal, and ϕ_0 is the initial phase of the carrier.

III. TOA Estimation of 5G Signals Based on ML Algorithms

This section describes the system model for implementing continuous tracking of 5G signals using ML algorithms. The main steps in the system are described in detail. As shown in Fig. 4, the system includes the main steps of synchronization and demodulation, discriminator function generation, ML network training, and TOA tracking based on ML.

A. Synchronization and Demodulation

The purpose of this step is to obtain the correct position of the FFT window of the received signal SSB through coarse synchronization and extract the DMRS. The developed SDR needs to perform cell search, OFDM demodulation, and DMRS extraction. The details are as follows.

1) In the 5G protocol, the coarse synchronization of the signal can be accomplished utilizing cell search to obtain information, such as the correct position of the FFT window of the received signal SSB and the cell ID. In cell search, the $N_{\mathrm{ID}}^{(2)}$ is first obtained using three known possible PSSs correlated with the received signal. The $N_{\mathrm{ID}}^{(1)}$ is then obtained using 336 known possible SSSs correlated with the received signal. The correct position $\hat{\epsilon}_{\max}$ of the FFT window of the received signal SSB can be determined through the above two steps of detection. The cell ID of the received signal can be obtained by substituting $N_{\mathrm{ID}}^{(2)}$ and $N_{\mathrm{ID}}^{(1)}$ into (1).
2) Based on the results of the coarse synchronization of the received signal, the time offset can be corrected directly for the received time-domain signal. After removing the guard samples with the length of N_{cp}, the time-domain signal can be converted into the frequency-domain signal through the FFT, and the extraction of DMRS can be completed. To this end, an FFT is performed on the samples belonging to the same OFDM symbol within SSB, which can be expressed as follows:

$$R(n) = \mathbf{FFT}\left\{r\left(k + \hat{\epsilon}_{\max}\right)\right\} \tag{5}$$

where $\mathbf{FFT}\{\cdot\}$ is the discrete-time transform operator. At the same time, the sequence number d of the DMRS subcarrier on the different OFDM symbols of the SSB can be obtained from the cell ID. After cell search, the cell ID of the received signal is known. Therefore, the DMRS on different symbols in the SSB can be extracted separately from the received signal.

B. Discriminator Function Generation

After completing the cell search of the 5G signal, the starting position of the integer symbol of the received signal can be obtained, that is, the integer delay $\mathrm{Int}\{\cdot\}$. In the signal TOA estimation, the early minus-late power (EMLP) discriminator function is used to obtain the fractional delay $\mathrm{Frac}\{\cdot\}$ of the received signal to improve the tracking accuracy of the received signal. Owing to the characteristics of the DMRS decentralized pilot, we can use the DMRS as the pilot signal to generate the discriminator function for the 5G signal. According to the 5G protocol, the DMRS can be mapped to an all-zero sequence of length N according to the position to obtain the received pilot signal $R(d)$. The local reference pilot on each symbol of SSB can also be generated and mapped to the all-zero sequence of length N according to the same position to obtain the reference pilot signal $t(d)$. As the delay in the time domain is equivalent to a phase rotation in the frequency domain, the early and late code signals of $t(d)$ can be obtained respectively as follows:

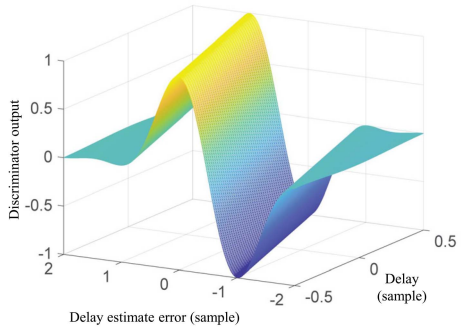
$$\begin{aligned} t_e(d) &= e^{-j2\pi d\xi/N} t(d) \\ t_l(d) &= e^{+j2\pi d\xi/N} t(d) \end{aligned} \tag{6}$$

where $\xi (0 < \xi < 1/2)$ is the advanced (and retarded) interval, which is normalized to the OFDM sample interval. When tracking the received signal continuously, we need to perform a phase rotation of the received pilot signal using the normalized symbol delay $\hat{\tau}$. Therefore, the received pilot signal after phase rotation can be expressed as follows:

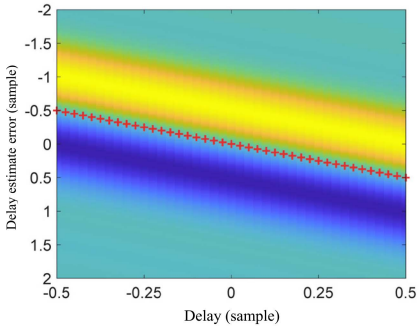
$$R'\left(\hat{\tau}\right) = e^{-j2\pi d\hat{\tau}/N} R(d). \tag{7}$$

The early and late correlations branch output in the frequency domain can be written respectively as follows:

$$\begin{aligned} \Re_e(\Delta\tau) &= \frac{1}{G}\sum_{g=0}^{G-1} R'\left(\hat{\tau}\right) t_e^*(d) \\ \Re_l(\Delta\tau) &= \frac{1}{G}\sum_{g=0}^{G-1} R'\left(\hat{\tau}\right) t_l^*(d) \end{aligned} \tag{8}$$



(a)



(b)

Fig. 5. Standard normalized discriminator function with different fractional delays. (a) 3-D view, (b) 2-D view.

where $\Delta\tau$ is the normalized symbol delay variation of the received signal pilot with respect to $\hat{\tau}$. In a channel without multipath and noise, the discriminator function is defined as follows:

$$\text{EMLP}(\Delta\tau) = |\Re_e(\Delta\tau)|^2 - |\Re_l(\Delta\tau)|^2 = G^2 A S(\Delta\tau, \xi) \quad (9)$$

where A is the received signal power and $S(\Delta\tau, \xi)$ is the normalized S-curve of the received signal pilot, which can be expressed as follows:

$$S(\Delta\tau, \xi) = \left(\frac{\sin(\pi(\Delta\tau + \xi))}{G\sin(\pi(\Delta\tau + \xi)/G)}\right) - \left(\frac{\sin(\pi(\Delta\tau - \xi))}{G\sin(\pi(\Delta\tau - \xi)/G)}\right). \quad (10)$$

When $\hat{\tau} = 0$ and $\Delta\tau = \tau$, the standard normalized S-curve with different fractional delays τ is shown in Fig. 5. Among them, Fig. 5(a) and (b) show the overall 3-D and 2-D views of the standard normalized S-curve with normalized symbol delay $\tau = (-0.5, 0.5)$, respectively. The red cross in Fig. 5(b) is the fractional delay corresponding to the standard normalized S-curve.

C. ML Network Training

In the developed SDR based on the ML algorithm, the training of the delay tracking network must be completed in the off-line phase. The SDR for TOA estimation not only needs the input data to the ML algorithms to contain sufficient variation delay information but also more requirements on the size of the input data. The discriminator function defined according to the received signal pilot is sufficiently lightweight on the basis of including the delay variation information, so it is used as the feature data for ML network training. The features of the standard normalized discriminator function with the same fractional delay are identical under noiseless conditions. Therefore, the ML network with the discriminator function feature data is not affected by changes in the 5G BS and has excellent applicability. Fig. 6 shows an overview of the algorithm for generating a delay-tracking network. The specific process is as follows.

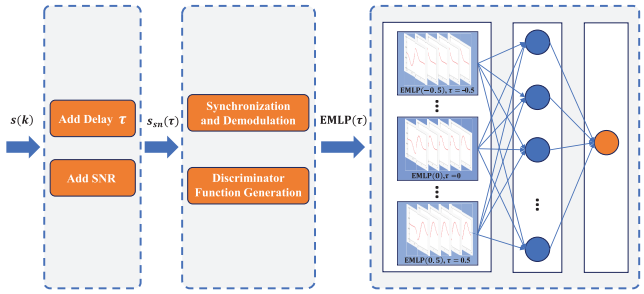


Fig. 6. Overview of the algorithm for generating a delay tracking network.

1) In the first step, the original signal data with different fractional delays need to be generated. The transmit signal $s(k)$ is first generated according to the 5G protocol. The signal $s(\tau)$ with different delays is obtained by adding different normalized symbol delays τ ($\tau = [0 \pm N_b \cdot \Upsilon]$) to $s(k)$, respectively. Where $N_b \cdot \Upsilon \geq 0.5$, which is used to ensure that the trained delay tracking networks can identify fractional delays within one sample point. To obtain high-precision TOA estimation results in SDR, the trained delay tracking networks should have a high resolution for delay variation. Therefore, we set the minimum normalized symbol delay variation resolution $\Upsilon = 0.025$, and set $N_b = \{0, 1, 2, \ldots, 20\}$. This means that the trained delay tracking networks can identify fractional delays within ± 0.5 sample points at a resolution of 0.025 sample points. Notice, the parameter settings in this section are empirical values that take into account the efficiency and accuracy of the operation, and adjustments to the parameter settings may affect the final experimental results. To increase the stability and anti-interference ability of the ML training network, a different SNR is added on $s(\tau)$ separately to obtain the signal $s_{sn}(\tau)$ with different SNR. Considering the SNR requirements for the available 5G signals in the communication network, $sn = \{0, 5, 10, \ldots, 50\}$ in dB is set.
2) In the next step, the feature data for input into the ML algorithms need to be extracted from the original signal data with different fractional delays. First, according to Section III-A, the DMRS must be extracted from generated $s_{sn}(\tau)$ with different delays through the cell search, OFDM demodulation, and DMRS extraction to obtain $R_{sn}(\tau)$. Following, the discriminator function corresponding to each $R_{sn}(\tau)$ is computed according to the method in Section III-B. In calculating the discriminator function, $\hat{\tau}$ is set to 0, which means $\tau = \Delta\tau$. Therefore,

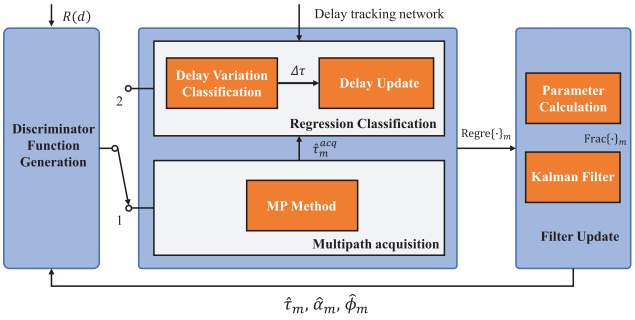


Fig. 7. Flow of TOA tracking based on ML.

the discriminator function EMLP(τ) with different *sn* can be obtained.

3) Finally, the feature data must be input into the ML algorithms to obtain the delay tracking networks. With the above two steps, the feature data for ML network training is obtained. In order to ensure the developed SDR based on the ML algorithm can obtain TOA estimation on mobile terminals through edge computing, the ML algorithm applied to SDR should have lightweight characteristics. In this article, we select four typical lightweight ML algorithms, namely, classification tree (CT) [33], *k*-nearest neighbor (KNN) [34], [35], classification ensembles (CEs) [36], [37], and support vector machines (SVMs) [38], for the developed SDR and use the same test data for the comparative analysis of the results. In the delay tracking network training, τ is used as the data label and EMLP(τ) is used as the data set, which are respectively input into different ML algorithms to obtain the corresponding delay tracing network.

D. TOA Tracking Based on Machine Learning

In conventional delay tracking algorithms, the output at the zero value of the discriminator delay estimation error is usually used as the result of the signal delay estimation. This method can usually achieve finer accuracy results in open areas outdoors. However, in an indoor environment, serious multipath effects and noise interference will distort the discriminator, and the output at the discriminator delay estimation error value of zero is no longer representative of the current delay estimation result. The TOA tracking based on ML first obtains the initial value of the multipath delay of the received signal through the matching pursuit (MP) method [39], [40]. In the follow-up tracking process, the tracking loop is used to refine the TOA estimation and track any changes in the multipath signal. In this tracking loop, for each SSB of the 5G signal, the discriminator function $\text{EMLP}_m(\Delta\tau)$ of the *m*th path signal should be obtained separately. The delay tracking network uses the whole of $\text{EMLP}_m(\Delta\tau)$ as the criterion for the signal delay to perform ML regression classification and integrate the KF to obtain the delay variation $\Delta\tau_m$ of adjacent SSB. The flow of TOA tracking based on ML is shown in Fig. 7. When the received signal is tracked for the first time, the multipath signal delay is obtained by 1, and each subsequent time the delay variation is obtained by 2.

1) Multipath Acquisition: In the signal propagation, the influence of complex environments and human activities produces serious multipath effects. For the received 5G signal, it is a combination of multipath signals with different delays and energies. Based on the multipath signal model described before, the received signal with locally generated reference code to get the autocorrelation function $\Re(\tau)$ can be expressed as follows:

$$\Re(\tau) = \sum_{m=0}^{M-1} \Re_m(\tau_m) \tag{11}$$

$$\Re_m(\tau_m) = \frac{1}{G}\sum_{g=0}^{G-1} R(\tau_m)t^*(d). \tag{12}$$

Therefore, based on each demodulated SSB, the delay estimates for different paths can be given as follows:

$$\hat{\tau}_m = \text{Int}\{\cdot\}_m + \text{Frac}\{\cdot\}_m \tag{13}$$

where $\text{Int}\{\cdot\}_m$ and $\text{Frac}\{\cdot\}_m$ are the integer delay and fractional delay of the *m*th path signal, respectively. In this case, based on the minimum mean square error principle and signal model, the autocorrelation function $\Re(\tau_m)$ of *m*th signal can be expressed as follows:

$$\Re(\tau_m) = \Re(\tau) - \sum_{i=0}^{m-1} \alpha_i \cdot s(\tau - \tau_i)e^{j\phi_i}. \tag{14}$$

In the cell search, the coarse synchronization of the received signal is already completed, so the integer delay $\text{Int}\{\cdot\}_m$ is determined to be 0. We only need the MP method to derive finer accuracy $\text{Frac}\{\cdot\}_m$ to obtain the delay estimation results $\hat{\tau}_m$ for different paths. As the number *M* of a multipath signal is unknown, the order-recursive least-square MP (LS-MP) algorithm [41] is used to obtain the delayed initial value $\hat{\tau}_m^{\text{acq}}$ of the multipath signal.

2) Regression Classification: In the continuous tracking loop of the received signal, the estimate of the autocorrelation function $\Re(\hat{\tau}_m)$ of *m*th path signal is first obtained at the current moment according to (14). The discriminator function $\text{EMLP}_m(\Delta\tau_m)$ of *m*th path signal in discriminator function generation is then obtained. The normalized discriminator of the *m*th path received signal must be first brought into the pretrained delay tracking networks to obtain the delay variation $\Delta\tau_m$. Therefore, the regression classification delay at the current moment can be expressed as follows:

$$\text{Regre}\{\cdot\}_m = \hat{\tau}_m + \Delta\tau_m \tag{15}$$

where the normalized symbol delay $\hat{\tau}_m$ is set to delay acquisition results $\hat{\tau}_m^{\text{acq}}$ during the first fractional delay regression classification of the received signal. For the delay regression classification estimated with each subsequent cycle iterative, the normalized symbol delay $\hat{\tau}_m$ is set to be the result of the previous cycle iterative estimation $\text{Frac}\{\cdot\}_m$. The logical structure of the cycle iterative is expressed in Algorithm 1.

Algorithm 1 Iteration Process of TOA Tracking Based on ML

```
set loop = 1.
while signal tracking do
    if loop = 1 then
        set τ̂_m = τ̂_m^acq
    else
        set τ̂_m = Frac{·}_m
    end if
    Extract the autocorrelation function ℜ(τ̂_m) of mth path signal using (14).
    Calculate the discriminator function EMLP_m(Δτ_m) of mth path signal using (9).
    Calculate ML regression classification delay Regre{·}_m based on Δτ_m using (15).
    Calculate fractional delay estimates Frac{·}_m based on KF using (16).
    Calculate the parameter α̂_m and φ̂_m using (17), (18).
    set loop = loop + 1.
end while
```

3) Filter Update: The resolution of the ML regression classification delay variation $\Delta\tau_m$ is equal to the minimum normalized symbol delay variation resolution Υ when training the ML network. Therefore, it is impossible to obtain a continuous delay estimation result through ML regression classification. By introducing a KF [42], [43], the discrete delay estimation results are converted into continuous delay estimation results, and the estimation error caused by misclassification in the regression classification process is reduced. As shown in (16), the regression classification delay is brought into the KF, and then the fractional delay estimate $\hat{\tau}_m$ of mth path signal for the current moment is obtained

$$\hat{\tau}_m = \text{Frac}\{\cdot\}_m = \textbf{Kalman}\{\text{Regre}\{\cdot\}_m\} \tag{16}$$

where $\textbf{Kalman}\{\cdot\}$ is the KF operator. Based on the $\hat{\tau}_m$, the estimates of the amplitude $\hat{\alpha}_m$ and phase $\hat{\phi}_m$ of mth path signal at the current moment can be expressed as follows:

$$\hat{\alpha}_m = \Re\left\{\left[\Re\left(\hat{\tau}_m\right) - \sum_{i=0}^{m-1}\hat{\alpha}_i \cdot s\left(\tau - \hat{\tau}_i\right)e^{j\hat{\phi}_i}\right]e^{-j\hat{\phi}_m}\right\} \tag{17}$$

$$\hat{\phi}_m = \arg\left[\Re\left(\hat{\tau}_m\right) - \sum_{i=0}^{m-1}\hat{\alpha}_i \cdot s\left(\tau - \hat{\tau}_i\right)e^{j\hat{\phi}_i}\right]. \tag{18}$$

IV. Analysis of System Errors

In the developed SDR, noise and multipath signals in the received signal can seriously affect the TOA estimation accuracy of the system. In this section, the system error analysis is carried out separately for the AWGN channel and multipath channel [44]. The simulation signals are all generated according to the 5G protocol, with different white noise and multipath signals added according to the experimental requirements.

A. Error Analysis in AWGN Channel

In the AWGN channel, the white noise of the received signal will introduce an error V_E in the discriminator function [45]. The discriminator function is defined as follows:

$$\text{EMLP}(\Delta\tau) = G^2AS(\Delta\tau, \xi) + V_E \tag{19}$$

where

$$\begin{aligned} V_E &= \text{Var}\left[\left|\Re_e(\Delta\tau)\right|^2\right] + \text{Var}\left[\left|\Re_l(\Delta\tau)\right|^2\right] \\ &\quad - 2\text{Cov}\left[\left|\Re_e(\Delta\tau)\right|^2, \left|\Re_l(\Delta\tau)\right|^2\right] \\ &\le \text{Var}\left[\left|\Re_e(\Delta\tau)\right|^2\right] + \text{Var}\left[\left|\Re_l(\Delta\tau)\right|^2\right] \\ &= 2G^2\sigma^4\left\{1 + \frac{A}{G\sigma^2}\left[\frac{\sin(\pi(\Delta\tau - \xi))}{\sin(\pi(\Delta\tau - \xi)/G)}\right]^2\right. \\ &\quad \left. + \frac{A}{G\sigma^2}\left[\frac{\sin(\pi(\Delta\tau + \xi))}{\sin(\pi(\Delta\tau + \xi)/G)}\right]^2\right\} \end{aligned} \tag{20}$$

where covariance $2\text{Cov}[|\Re_e(\Delta\tau)|^2, |\Re_l(\Delta\tau)|^2]$ should be nonnegative to guarantee that the inequality in (20) holds, so that $\Delta\tau$ is 0 and $0 < \xi \le 1/2$. ξ should be set to 0.5 to ensure that $|\Re_e(\Delta\tau)|^2$ and $|\Re_l(\Delta\tau)|^2$ are uncorrelated [44]. Thus, (20) can be written as follows:

$$V_E = 2G^2\sigma^4\left\{1 + \frac{2A}{G\sigma^2}\left[\frac{\sin(\pi\xi)}{\sin(\pi\xi/G)}\right]^2\right\}. \tag{21}$$

In the AWGN channel, the error of the discriminator function is closely related to the number of pilots G and the SNR ($sn = A/\sigma^2$) of received signals. As the number of pilots G is constant in the SSB structure of 5G signals, simulations of 5G signals with different SNRs were carried out using the SDR based on ML algorithms. The complete tracking process was performed for the same simulated 5G signals using four different ML algorithms. The root mean square error (RMSE) at different SNR is shown in Fig. 8.

According to the results in Fig. 8, the accuracy of the tracking algorithm increases as the SNR of the simulated 5G signal increases when based on the same ML algorithm. Fig. 9 shows the 95% cumulative distribution function (CDF) statistical results for the RMSE of the four ML algorithms at different SNRs. The performance of the ML tracking algorithm is closely related to the SNR of the signal. The tracking accuracy stabilizes when the SNR is more than 15 dB. For different ML algorithms, the KNN is more sensitive to the SNR, and the tracking accuracy is poor under the condition of low SNR. The remaining three ML algorithms performed similarly under different SNR conditions and achieved smaller errors.

B. Error Analysis in Multipath Channel

In the multipath channel, the accumulation of multipath signals will introduce errors χ_1 and χ_2 in the discriminator function [44]. The discriminator function is defined as follows:

$$\text{EMLP}(\Delta\tau) = G^2AS(\Delta\tau, \xi) + \chi_1 + \chi_2 \tag{22}$$

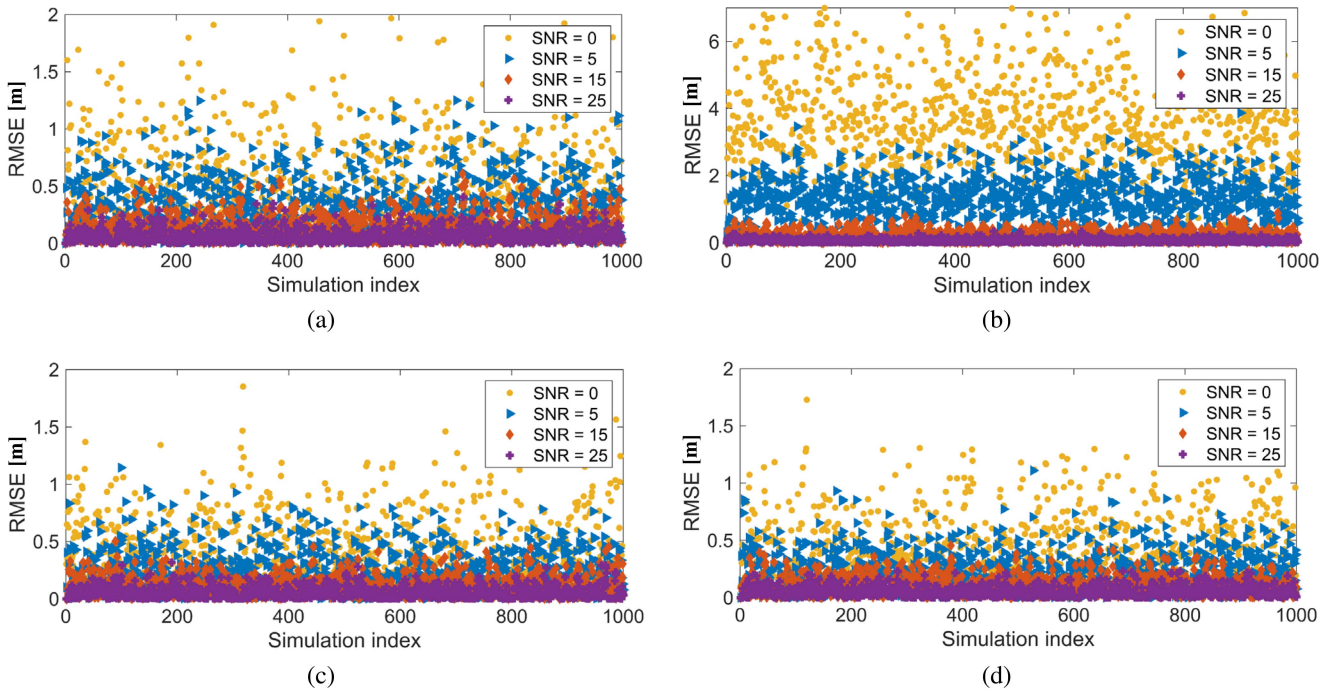


Fig. 8. Tracking results of simulation experiments with different SNR. (a) CT. (b) KNN. (c) CE. (d) SVM.

where

$$\chi_1 = \left| \sum_{g=0}^{G-1} \sum_{m=0}^{M-1} \alpha_m e^{-j2\pi(d/N)((\tau_m/t_s)+\Delta\tau+\xi)} \right|^2 - \left| \sum_{g=0}^{G-1} \sum_{m=0}^{M-1} \alpha_m e^{-j2\pi(d/N)((\tau_m/t_s)+\Delta\tau-\xi)} \right|^2 \quad (23)$$

$$\chi_2 = 2Re\left\{ \left[\sum_{g=0}^{G-1} e^{-j2\pi(d/N)(\Delta\tau+\xi)} \right] \cdot \left[\sum_{g'=0}^{G-1} \sum_{m=0}^{M-1} \alpha_m^* e^{j2\pi(d'/N)((\tau_m/t_s)+\Delta\tau+\xi)} \right] \right\} - 2Re\left\{ \left[\sum_{g=0}^{G-1} e^{-j2\pi(d/N)(\Delta\tau-\xi)} \right] \cdot \left[\sum_{g'=0}^{G-1} \sum_{m=0}^{M-1} \alpha_m^* e^{j2\pi(d'/N)((\tau_m/t_s)+\Delta\tau-\xi)} \right] \right\} \quad (24)$$

where ξ is also 0.5. The errors χ_1 and χ_2 introduced in discriminator function EMLP($\Delta\tau$) are only related to multipath signals, and they will distort the discriminator function [23], [44]. The error of the discriminator function in multipath channels is not only related to the multipath delay τ_m but also to the multipath amplitude α_m and the sampling time t_s. Under the premise that numerologies $\mu = 1$, the sampling time t_s is a constant value, so we next analyze the effect of different multipath amplitudes and delays on the output of the discriminator function.

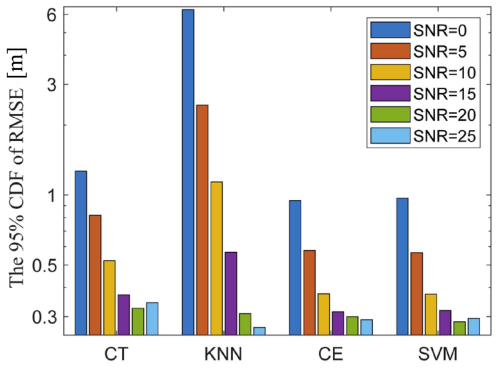


Fig. 9. 95% CDF of simulation experiments with different SNR.

A simplified channel impulse response (CIR) model was used to analyze the effect of multipath signals on the discriminator function output. In this simplified CIR model, the amplitude α_0 of the first-path signal of the first path is set to a constant value of 1 and the delay τ_0 is set to a constant value of zero. A multipath signal is introduced, a constant multipath amplitude is used, and the multipath delay is continuously increased to obtain the statistics of the output error caused by the multipath signal. To analyze the effect of multipath signals with different amplitudes on the discriminator function output, four sets of multipath signals with different amplitudes were tested. The multipath signal amplitude $|\alpha_1|$ was 6, 9, 12, and 15 dB lower than the first-path signal amplitude $|\alpha_0|$, respectively (i.e. $|\alpha_1|$ was set to 0.2512, 0.1259, 0.0631, and 0.0316, respectively).

As shown in Fig. 10, the solid and dashed lines indicate the statistics of the pseudorange measurement error at value zero of the discriminator function for constructive and

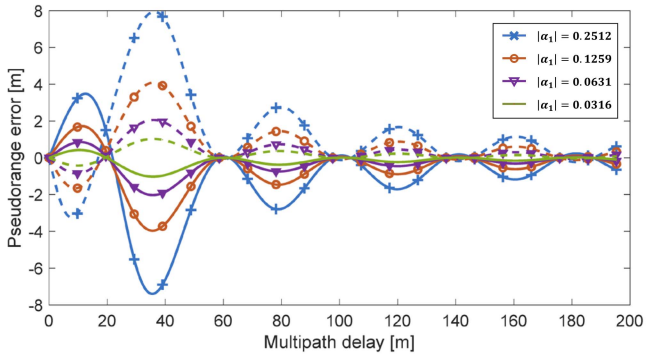


Fig. 10. Statistics of the pseudorange measurement error caused by multipath signals in the discriminator function. The solid and dashed lines represent constructive and destructive interferences, respectively. The pseudorange measurement error is the discriminator output at the zero value of the discriminator function.

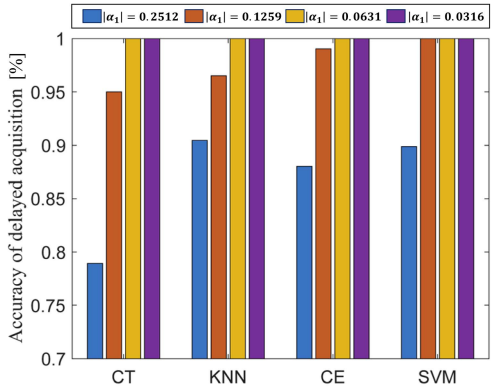


Fig. 11. First-path signal delay tracking accuracy statistics.

destructive multipath signals, respectively. When the multipath signal delay is constant, the discriminator function output error introduced by the multipath signal decreases as the multipath amplitude decreases. When the multipath signal amplitude is constant, the effect of the multipath signal delay on the discriminator function output error varies periodically. As the multipath delay tends to zero, the first-path signal coincides with the multipath signal, so errors are no longer introduced. As the multipath delay increases, the first-path signal is separated from the multipath signal and the error introduced by it decreases.

In the TOA estimation of 5G signal based on ML algorithms, instead of using the output at the zero value of the discriminator delay estimation error as the result of the pseudorange measurement, the overall characteristics of the discriminator function are used as the output basis of the pseudorange measurement. The purpose of the ML algorithm is to accurately identify the delay of first-path signals with the resolution of the delay tracking network according to the characteristics of the discriminator function. Therefore, in the subsequent part of this section, the accuracy of different ML algorithms for first-path signal delay tracking is analyzed.

The same simplified CIR model was used to analyze the accuracy of the first-path signal delay tracking and the same parameters were set for first-path and multipath signals as in the previous part of this section. As the resolution $\Upsilon = 0.025$, we consider the tracking of the first-path signal delay to be unbiased when the deviation of the tracking result for the first-path signal delay from the true first-path signal delay is less than 0.025 sample points. The accuracy of the four ML algorithms for the tracking of the first-path signal delay is shown in Fig. 11. In the signal tracking, the reduction of the multipath signal amplitude will improve the accuracy of the first-path signal tracking. When $|\alpha_1| \leq 0.0631$, the accuracy of all four ML algorithms for the tracking of the first-path signal delay is close to 100%.

V. Field Experiment

To evaluate the proposed method, a field experiment was performed with commercial 5G signals in a typical conference room at Wuhan University, Hubei, China. In this section, the hardware and software setup of the field experiment are first presented. Then, the experimental results are also presented. Finally, comparative experiments have been added from two perspectives.

A. Experimental Hardware and Software Setup

The construction of commercial BSs based on the 5G protocol is advancing globally as the preliminary determination of the 5G protocol. The Chinese operators China Mobile, China Unicom, and China Telecom have already completed basic 5G network coverage in some major cities across the country. At the location and time of the experiment in this article, there are commercial 5G micro BSs deployed by China Unicom in every room of the building. However, due to the limitations of the experimental scenario, there is only one 5G BS available in each room. Deriving the position of the receiver in one pseudorange measurement is infeasible. The results from the experiments were only used to evaluate the accuracy of the proposed method in pseudorange measurement.

The experimental scenario and a schematic diagram of the true trajectory of the movement of the experimenter is shown in Fig. 12. During the experiment, the experimenter held a cellular antenna on a bracket to receive the signal at the center frequency of the 5G signal SSB. The center frequency was obtained through preliminary investigation and searching for all possible frequencies [25]. Based on the information provided by the service provider and actual measurements, the subcarrier spacing for receiving 5G signal SSB is 30 kHz, and the SSB in this band has 7.2-MHz bandwidth. The cellular antenna was connected to a universal software radio peripheral USRP B210 through cable [46], driven by a GPS disciplined oscillator (GPSDO) to down-mix and sample 5G signals at 10 Mb/s. The experimenter held the bracket fixed at a height of 1.3 m and walked back and forth between A and B according to the planned true value trajectory at a constant speed. Notice that since there is a 2.0-m height difference between the 5G BS and the receiving antenna, the true value of the motion trajectory of the experimenter is 9.0 m and the actual distance variation between the BS and the receiving antenna is 7.2 m. A laptop was connected to the USRP B210 to record

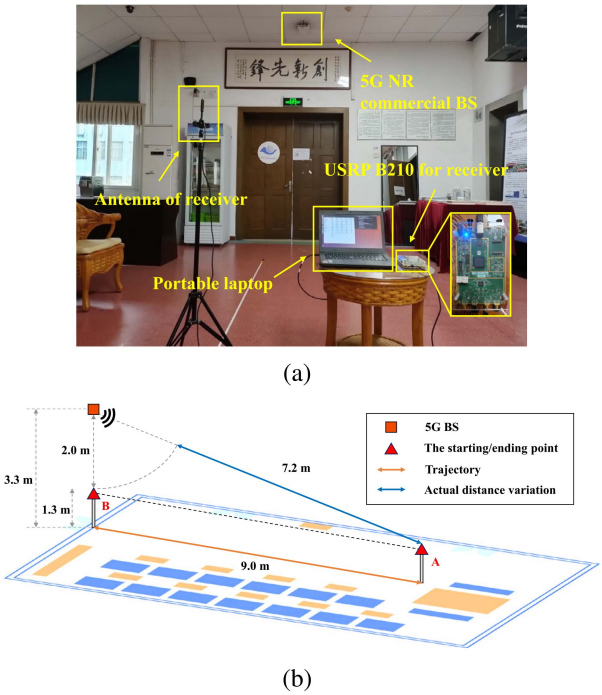


Fig. 12. (a) Experimental scenario and (b) schematic of the true value trajectory of the movement of the experimenter.

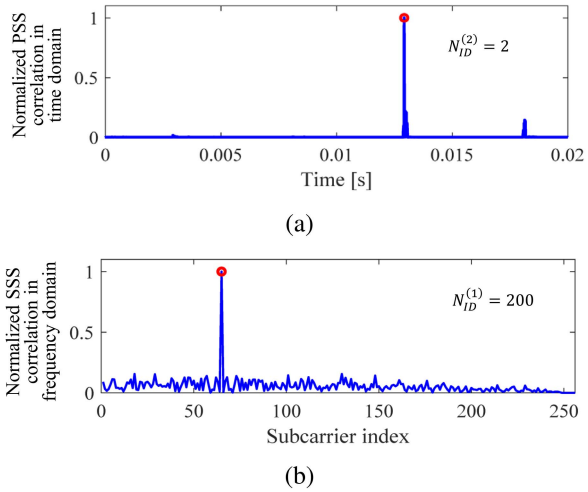


Fig. 13. (a) Normalized PSS correlation of two consecutive frames in the time domain for received 5G signals. (b) Normalized SSS correlation in the frequency domain for received 5G signal SSB (SSS detection is performed using all subcarrier data on the second OFDM symbol in the SSB).

data using GNU Radio software [47]. The recorded data was processed offline by an SDR developed on MATLAB [48].

B. Experimental Results

Pseudorange measurements of the 5G signal can be obtained by substituting the received signals into the developed SDR. In this process, the received signal was first correlated with all the possible PSS and the one with the highest peak value was selected. Fig. 13(a) shows the maximum of the PSS correlations values in one 5G signal frame, which is normalized by the highest value. Depending on the position of the PSS maximum, the signal can be demodulated into a frame structure to extract the SSB. At this time, the received SSS was correlated with all the possible SSS, and the one with the highest peak value was selected. Fig. 13(b) shows the maximum of the SSS correlation values in one 5G signal frame, which is normalized by the highest value. According to the results of the two correlations, the cell ID of the received signal can be calculated, and the DMRS on the SSB can be extracted.

Next, the DMRS is extracted as the pilot to calculate the discriminator function. In this process, in order to ensure the same interval time between adjacent discriminator functions, we only use the DMRS on the first symbol of each adjacent SSB as the pilot to calculate the discriminator function. The discriminator function curve of the actual received signal is distorted to varying degrees due to noise and multipath effects, and the discriminator output no longer fully represents the fractional delay of the current 5G signal. In this article, the ML algorithms are used to incorporate the whole discriminator function curve as a feature in the fractional delay calculation. It is worth noting that the ML network only needs to be trained once and it can be applied to any 5G signal without being affected by different cell IDs. Therefore, ML regression classification calculations with fractional delay can be performed directly in the SDR using the pretrained delay tracking networks.

Although separate ML algorithms in SDR can also achieve TOA estimation of 5G signals, KF can better prevent the accumulation of pseudorange measurement errors caused by serious multipath signal variations and noise during the actual use of SDR. As shown in Fig. 14, to better compare the role of KF in SDR, pseudorange measurements of the SDR output were calculated separately for the w/o and w/ KF cases using the same test data. In the existing results, the pseudorange measurement bias is significantly reduced with the addition of KF compared with the results of the w/o KF case. During the movement, the phenomenon of missing tracking targets caused by the accumulation of pseudorange measurement errors is effectively alleviated.

To compare more intuitively the TOA estimation accuracy of the four ML algorithms w/o and w/ KF, the pseudorange measurements were fitted to the true trajectories to obtain the pseudorange measurement errors for the four ML algorithm tracking results. Fig. 15(a) shows the pseudorange error statistical results of the four ML algorithms w/o KF. Fig. 15(b) shows those of the four ML algorithms w/ KF. Although four ML algorithms enable tracking measurements of 5G signals, the tracking accuracy varies. From the current results, the improvement of the KF on the accuracy of the KNN algorithm in practical applications is more obvious. The 95% CDF of the pseudorange measurement error is reduced from 2.91 to 0.75 m when KF is incorporated into the SDR. The remaining three algorithms improve the pseudorange measurement accuracy after KF. Among them, the pseudorange measurement error based on the SVM algorithm is the smallest. After KF, the 95% CDF of the pseudorange measurement error is 0.50 m. The specific pseudorange measurement error statistics from the experiment are given in Table I. The SD, maximum

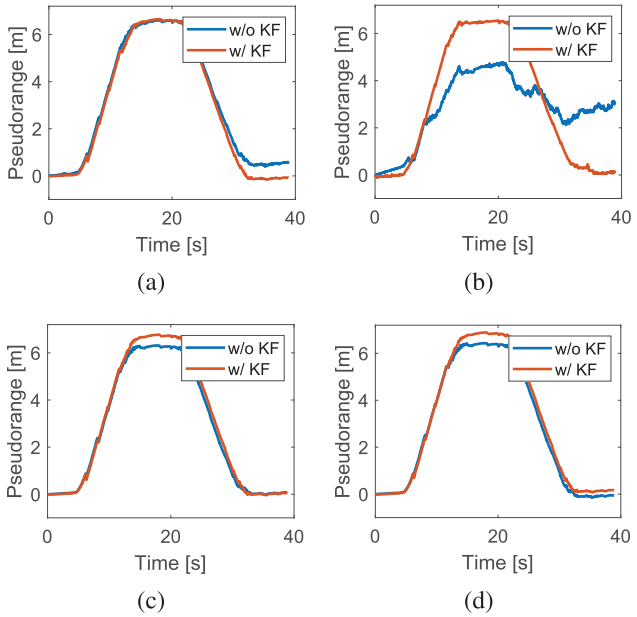


Fig. 14. Pseudorange measurement results of four ML algorithms. (a) CT, (b) KNN, (c) CEN, and (d) SVM.

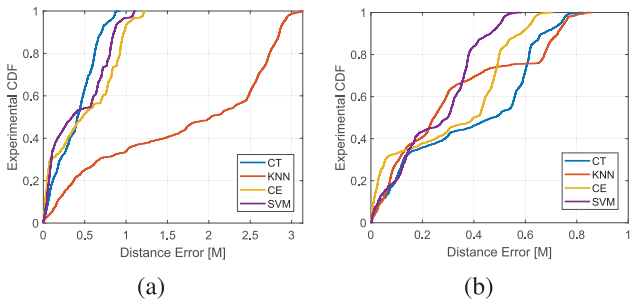


Fig. 15. Comparison of distance errors in calculation results of four ML algorithms. (a) Without KF. (b) With KF.

TABLE I
STATISTICAL RESULTS OF PSEUDORANGE MEASUREMENT ERRORS FOR 5G SIGNALS

Type	Method	SD (m)	ME (m)	RMSE (m)	95% CDF (m)
w/o KF	CT	0.42	0.92	0.39	0.78
	KNN	1.95	3.14	1.69	2.91
	CE	0.42	1.24	0.50	1.04
	SVM	0.36	1.12	0.42	0.92
w/ KF	CT	0.26	0.83	0.38	0.74
	KNN	0.33	0.86	0.31	0.75
	CE	0.24	0.71	0.31	0.62
	SVM	0.22	0.71	0.25	0.50

error (ME), RMSE, and 95% CDF of the pseudorange measurement error after KF are all improved. Therefore, in the developed SDR, the tracking based on ML algorithms is the foundation, and KF is an important component in achieving high accuracy and stable TOA estimation.

C. Experimental Comparison

To evaluate the performance of the developed SDR in relation to other tracking methods and its applicability to other

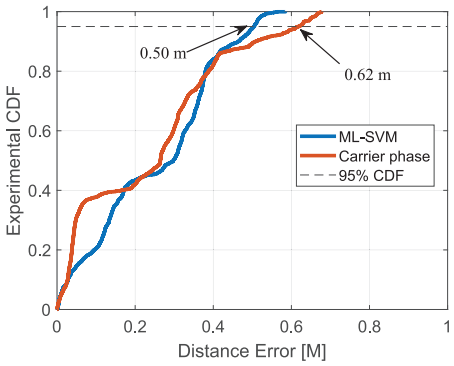


Fig. 16. Comparison of distance errors in calculation results of two methods.

signals. First, we compare the results based on a 5G signal carrier phase ranging method with the results of the method proposed in this article. Next, we introduce the LTE signal and obtain the TOA estimation result based on the signal structure of the LTE signal using the proposed method. The details are as follows.

1) Comparison Analysis of Different Methods: For the research on indoor positioning tracking methods with 5G signals, ideal indoor ranging results were achieved in [25] by estimating the phase of the 5G signal carrier. To analyze the ranging performance of the developed SDR in this article, the method in [25] is used for comparative analysis. In Section V-B, although the four ML algorithms can achieve TOA estimation in the indoor environment, the pseudorange measurement accuracy varies. The comparative analysis shows that the developed SDR based on the SVM algorithm has better performance in terms of the accuracy and stability of pseudorange measurement. Therefore, the developed SDR based on the SVM algorithm is used as a typical SDR to compare the TOA estimation accuracy. To control the interference in the measurement results caused by changes in the experimenter and experimental scenario, pseudorange measurement errors were obtained using the same test data.

Fig. 16 shows the pseudorange measurement error statistical results of the two methods based on the same test data.The 95% CDF of the pseudorange measurement error based on the 5G signal carrier phase ranging method is 0.62 m, and the 95% CDF of the pseudorange measurement error based on the 5G signal SVM method is 0.50 m. The overall error distributions of the two methods are essentially the same, but the developed SDR in this article has higher accuracy. At the same time, considering that ML algorithms may have better stability in serious multipath and noisy environments, and the ranging accuracy of the method can be further improved by adjusting the minimum resolution of the delay tracking network. As the performance of smart terminal equipment continues to improve, the TOA estimation methods based on lightweight ML algorithms show stronger advantages in practical use.

2) Comparison Analysis of Different Signals: Owing to the initial deployment and development of 5G networks, the joint networking of LTE and 5G signals may be a trend for a long time in the future. In response to this reality, this article constructs the application of SDR in LTE signals, which will

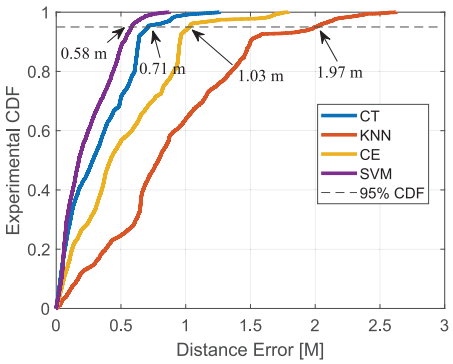


Fig. 17. Comparison of distance errors in calculation results of two signals.

TABLE II
STATISTICAL RESULTS OF PSEUDORANGE MEASUREMENT ERRORS FOR LTE SIGNALS

Signal	Method	SD (m)	ME (m)	RMSE (m)	95% CDF (m)
LTE	CT	0.32	1.26	0.32	0.71
	KNN	0.89	2.63	0.87	1.97
	CE	0.54	1.80	0.51	1.03
	SVM	0.27	0.87	0.24	0.58

greatly extend the application scenario of the method. In the LTE signal protocol [15], [16], [17], [18], [49], the downlink channel is usually estimated via cell reference signal (CRS) for channel estimation. As an important reference signal in LTE, CRS is evenly distributed throughout the downlink channel frequency domain and is transmitted periodically at high frequencies. In [16], [17], and [18], researchers used CRS to construct SDRs to achieve TOA estimation of LTE signals in outdoor environments.

Based on the method in this article, the CRS of commercial LTE signals will be used to construct SDR for delay variation tracking. Owing to the joint networking of LTE and 5G signals, the same experiment is performed in the experimental scenario of Section V-A. Here, each of the four typical lightweight ML algorithms was used for the developed SDR, and specific pseudorange measurement errors were obtained. As shown in Fig. 17, four typical lightweight ML algorithms achieve effective TOA estimation in the developed SDR. Among them, the accuracy of TOA estimation based on the KNN algorithm is low, and the 95% CDF of the pseudorange measurement error is 1.97 m. The accuracy of TOA estimation based on the SVM algorithm is the highest, and the 95% CDF of the pseudorange measurement error is 0.58 m. The specific pseudorange measurement error statistics from the experiment are provided in Table II.

VI. CONCLUSION

In this article, an SDR based on ML is developed for the TOA estimation of 5G signal in the indoor environment. The results of simulation and commercial 5G signal experiments demonstrate the feasibility of using 5G signals for pseudorange measurements in the indoor environment. For this purpose, first, the 5G standard and signal model that can be exploited for navigation is presented. Next, the complete process using ML algorithms for SDR to extract navigation observations from 5G signals is described.

The work proposed in this article has been tested through commercial 5G signals, and four typical lightweight ML algorithms have been tested in the developed SDR to obtain pseudorange measurement results. From the current results, the ML algorithm based on SVM achieved the best pseudorange measurement results in the field experiments. The results show that the 95% CDF of the pseudorange measurement error using commercial 5G signals in the indoor environment is 0.50 m. The results of the comparison experiments indicate that the tracking method based on ML algorithms can achieve an equal or higher level of accuracy than the traditional carrier phase ranging tracking method. Moreover, ML methods may have better stability than the traditional carrier phase ranging method in serious multipath and noisy environments. In LTE application experiments, the developed SDR based on ML algorithms can be simultaneously applied to the signal-tracking process with good accuracy. At the same time, we notice the influence of multipath effects on wireless positioning accuracy in the indoor environment. In the future, we will improve the accuracy of TOA estimation of SDR under the influence of obstacles and develop a dedicated ML algorithm for 5G signal characteristics to improve the accuracy and stability of pseudorange measurements.

REFERENCES

[1] D. Zou, W. Meng, S. Han, K. He, and Z. Zhang, "Toward ubiquitous LBS: Multi-radio localization and seamless positioning," *IEEE Wireless Commun.*, vol. 23, no. 6, pp. 107–113, Dec. 2016.
[2] L. Chen et al., "Robustness, security and privacy in location-based services for future IoT: A survey," *IEEE Access*, vol. 5, pp. 8956–8977, 2017.
[3] Y. Li et al., "Toward location-enabled IoT (LE-IoT): IoT positioning techniques, error sources, and error mitigation," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4035–4062, Mar. 2021.
[4] N. Shen et al., "Interactive multiplemodel vertical vibration detection of structures based on high-frequency GNSS observations," *GPS Solut.*, vol. 26, no. 2, pp. 1521–1886, 2022.
[5] E. S. Lohan, D. Alonso de Diego, J. A. Lopez-Salcedo, G. Seco-Granados, P. Boto, and P. Fernandes, "Unambiguous techniques modernized GNSS signals: Surveying the solutions," *IEEE Signal Process. Mag.*, vol. 34, no. 5, pp. 38–52, Sep. 2017.
[6] A. Yassin et al., "Recent advances in indoor Localization: A survey on theoretical approaches and applications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1327–1346, 2nd Quart., 2017.
[7] Y. Shu et al., "Gradient-based fingerprinting for indoor Localization and tracking," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2424–2433, Apr. 2016.
[8] J. Yan, L. Zhao, J. Tang, Y. Chen, R. Chen, and L. Chen, "Hybrid kernel based machine learning using received signal strength measurements for indoor localization," *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2824–2829, Mar. 2018.
[9] Z. Gao, Y. Gao, S. Wang, D. Li, and Y. Xu, "CRISLoc: Reconstructable CSI fingerprinting for indoor Smartphone localization," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3422–3437, Mar. 2021.
[10] R. Zhou, X. Lu, P. Zhao, and J. Chen, "Device-free presence detection and localization with SVM and CSI fingerprinting," *IEEE Sensors J.*, vol. 17, no. 23, pp. 7990–7999, Dec. 2017.
[11] S. Mazuelas et al., "Robust indoor positioning provided by real-time RSSI values in unmodified WLAN networks," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 5, pp. 821–831, Oct. 2009.
[12] L. Chen, L. Pei, H. Kuusniemi, Y. Chen, T. KröGer, and R. Chen, "Bayesian fusion for indoor positioning using Bluetooth fingerprints," *Wireless Pers. Commun.*, vol. 70, no. 4, pp. 1735–1745, 2013.

[13] R. Faragher and R. Harle, "Location fingerprinting with Bluetooth low energy beacons," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2418–2428, Nov. 2015.

[14] Y. Zhuang, J. Yang, L. Qi, Y. Li, Y. Cao, and N. El-Sheimy, "A pervasive integration platform of low-cost MEMS sensors and wireless signals for indoor localization," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4616–4631, Dec. 2018.

[15] M. Driusso, C. Marshall, M. Sabathy, F. Knutti, H. Mathis, and F. Babich, "Vehicular position tracking using LTE signals," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3376–3391, Apr. 2017.

[16] K. Shamaei and Z. M. Kassas, "LTE receiver design and multipath analysis for navigation in urban environments," *Navigation*, vol. 65, no. 4, pp. 655–675, 2018.

[17] P. Wang and Y. J. Morton, "Multipath estimating delay lock loop for LTE signal TOA estimation in indoor and urban environments," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5518–5530, Aug. 2020.

[18] Z. Liu, L. Chen, X. Zhou, N. Shen, and R. Chen, "Multipath tracking with LTE signals for accurate TOA estimation in the application of indoor positioning," *Geo-Spat. Inf. Sci.*, Oct. 2022, doi: 10.1080/10095020.2022.2108344.

[19] L. Ma, N. Jin, Y. Cui, and Y. Xu, "LTE user equipment RSRP difference elimination method using multidimensional scaling for LTE fingerprint-based positioning system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2017, pp. 1–6.

[20] Y. Wang, S. Han, Y. Tian, C. Xiu, and D. Yang, "Is centimeter accuracy achievable for LTE-CSI fingerprint-based indoor positioning?" *IEEE Access*, vol. 8, pp. 75249–75255, 2020.

[21] K. Gao, H. Wang, H. Lv, and W. Liu, "Toward 5G NR high-precision indoor positioning via channel frequency response: A new paradigm and dataset generation method," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2233–2247, Jul. 2022.

[22] Y. Ruan, L. Chen, X. Zhou, G. Guo, and R. Chen, "Hi-Loc: Hybrid indoor localization via enhanced 5G NR CSI," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–15, 2022.

[23] K. Shamaei and Z. M. Kassas, "Receiver design and time of arrival estimation for opportunistic localization with 5G signals," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4716–4731, Jul. 2021.

[24] M. Pan et al., "Efficient joint DOA and TOA estimation for indoor positioning with 5G Picocell base stations," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–19, 2022.

[25] L. Chen, X. Zhou, F. Chen, L.-L. Yang, and R. Chen, "Carrier phase ranging for indoor positioning with 5G NR signals," *IEEE Internet Things J.*, vol. 9, no. 13, pp. 10908–10919, Jul. 2022.

[26] Z. Tian, Y. Wang, Y. Sun, and J. Qiu, "Location privacy challenges in mobile edge computing: Classification and exploration," *IEEE Netw.*, vol. 34, no. 2, pp. 52–56, Mar./Apr. 2020.

[27] S. J. Vaughan-Nichols, "Will mobile computing's future be location, location, location?" *Computer*, vol. 42, no. 2, pp. 14–17, Feb. 2009.

[28] S. Wang, Q. Hu, Y. Sun, and J. Huang, "Privacy preservation in location-based services," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 134–140, Mar. 2018.

[29] S. Wang et al., "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1205–1221, Jun. 2019.

[30] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5G evolution: A view on 5G cellular technology beyond 3GPP release 15," *IEEE Access*, vol. 7, pp. 127639–127651, 2019.

[31] "Physical channels and modulation." Accessed: Mar. 16, 2022. [Online]. Available: https://www.etsi.org/deliver/etsi-ts/138200--138299/138211/15.02.00--60/ts-138211v150200p.pdf

[32] M. Shafi et al., "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[33] S. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 3, pp. 660–674, May/Jun.1991.

[34] S. Kulkarni, G. Lugosi, and S. Venkatesh, "Learning pattern classification–A survey," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2178–2206, Oct. 1998.

[35] I. Triguero, J. Derrac, S. Garcia, and F. Herrera, "A taxonomy and experimental study on prototype generation for nearest neighbor classification," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 1, pp. 86–100, Jan. 2012.

[36] Y. Ren, L. Zhang, and P. Suganthan, "Ensemble classification and regression-recent developments, applications and future directions [review article]," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 41–53, Feb. 2016.

[37] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022.

[38] H. Ahmed and A. K. Nandi, *Support Vector Machines (SVMs)*. Hoboken, NJ, USA: 2019, pp. 259–277.

[39] S. Cotter and B. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.

[40] L. Chen, L.-L. Yang, J. Yan, and R. Chen, "Joint wireless positioning and emitter identification in DVB-T single frequency networks," *IEEE Trans. Broadcast.*, vol. 63, no. 3, pp. 577–582, Sep. 2017.

[41] W. Li and J. C. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE J. Ocean. Eng.*, vol. 32, no. 4, pp. 927–939, Oct. 2007.

[42] P. Hargrave, "A tutorial introduction to Kalman filtering," in *Proc. IEE Colloq. Kalman Filt. Introduct. Appl. Future Develop.*, 1989, pp. 1–6.

[43] W. Gong, J. Li, and J. Yang, "A novel positioning algorithm with pseudo range and carrier phase based on Kalman filter," in *Proc. Int. Conf. Netw. Secur. Wireless Commun. Trusted Comput.*, vol. 1, 2009, pp. 417–420.

[44] B. Yang, K. Letaief, R. Cheng, and Z. Cao, "Timing recovery for OFDM transmission," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2278–2291, Nov. 2000.

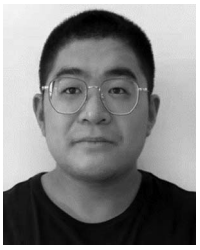
[45] L. Chen, P. Thevenon, G. Seco-Granados, O. Julien, and H. Kuusniemi, "Analysis on the TOA tracking with DVB-T signals for positioning," *IEEE Trans. Broadcast.*, vol. 62, no. 4, pp. 957–961, Dec. 2016.

[46] "USRP B210." Accessed: Mar. 16, 2022. [Online]. Available: https://www.ettus.com/all-products/ub210-kit/

[47] "GNU radio." Accessed: Mar. 16, 2022. [Online]. Available: https://www.gnuradio.org/

[48] "MATLAB." Accessed: Mar. 16, 2022. [Online]. Available: https://www.mathworks.com

[49] K. Shamaei, J. Khalife, and Z. M. Kassas, "Exploiting LTE signals for navigation: Theory to implementation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2173–2189, Apr. 2018.



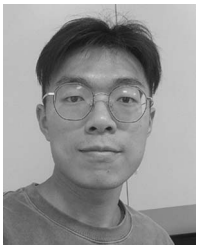
Zhaoliang Liu received the M.S. degree from Wuhan University, Wuhan, China, in 2020, where he is currently pursuing the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing.

His research interests include indoor positioning and navigation technology based on signal of opportunity, wireless communications, and the Internet of Things.

Liang Chen received the Ph.D. degree from Southeast University, Nanjing, China, in 2009.

In March 2017, he has been with Wuhan University, Wuhan, China, where he is the Professor with the State Key Laboratory of Surveying, Mapping and Remote Sensing Information Engineering, as well as the Deputy Head of the Institute of Artificial Intelligence in Geomatics. He has published over 70 scientific research papers.

Dr. Chen served as an Associate Editor to *Navigation, the Journal of Institute of Navigation*, and *The Journal of Navigation*. His research interests include signal processing for navigation and positioning, and indoor positioning.

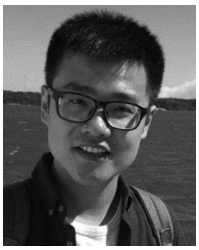


Xin Zhou received the B.S. and M.S. degrees from Wuhan University, Wuhan, China, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree in geodesy and survey engineering with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing.

His research interests include indoor positioning and navigation technology based on signal of opportunity, wireless communications, and the Internet of Things.

Zhenhang Jiao is currently pursuing the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

His research interests focus on signals-of-opportunity-based positioning and data fusion.



Guangyi Guo received the B.S. and M.S. degrees in geographic information science from Hubei University, Wuhan, Hubei, China, in 2012 and 2015, respectively.

He is currently a Postdoctoral Fellow with Wuhan University, Wuhan. His research interests include indoor localization and navigation, data processing, and machine learning.

Ruizhi Chen is a Professor and the Director of the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, and an Academician of the Finnish Academy of Science and Letters. He is an International Scholar working in the field of navigation and positioning. He is committed to the theoretical research and core technology development for seamless indoor/outdoor positioning using smartphones.