HAL open science

CSI-Based Probabilistic Indoor Position Determination: An Entropy Solution

Luan Chen, Iness Ahriz, Didier Le Ruyet

To cite this version:

Luan Chen, Iness Ahriz, Didier Le Ruyet. CSI-Based Probabilistic Indoor Position Determination: An Entropy Solution. IEEE Access, 2019, 7, pp.170048-170061. 10.1109/ACCESS.2019.2955747 . hal-02444914

HAL Id: hal-02444914

https://hal.science/hal-02444914v1

Submitted on 20 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 23, 2019, accepted November 20, 2019, date of publication November 25, 2019, date of current version December 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2955747

CSI-Based Probabilistic Indoor Position Determination: An Entropy Solution

LUAN CHEN[1,2], INESS AHRIZ[1], AND DIDIER LE RUYET[1], (Senior Member, IEEE)
[1]CEDRIC/LAETITIA Laboratory, Conservatoire National des Arts et Métiers, 75003 Paris, France
[2]School of Electronic Information, Wuhan University, Wuhan 430072, China

Corresponding author: Luan Chen (luan.chen@whu.edu.cn)

This work was supported in part by the China Scholarship Council, Beijing, China, under Grant CSC 201606270203.

ABSTRACT Location Fingerprinting (LF) is a promising localization technique that enables enormous commercial and industrial Location-Based Services (LBS). Existing approaches either appeal to the simple Received Signal Strength (RSS), which suffers from dramatic performance degradation due to sophisticated environmental dynamics, or rely on the fine-grained physical layer Channel State Information (CSI), whose intricate structure leads to an increased computational complexity. In this paper, we adopt Autoregressive (AR) modeling based entropy of CSI amplitude as location fingerprint, which shares the structural simplicity of RSS while exploiting the most location-specific statistical channel information. On this basis, we design *EntLoc*, a CSI-based probabilistic indoor localization system using commercial off-the-shelf Wi-Fi devices. *EntLoc* is deployed in an office building covering over 200 m^2. Extensive indoor scenario experiments corroborate that our proposed system yields superior localization accuracy over previous approaches even with only one signal transmitter.

INDEX TERMS Indoor localization, location fingerprinting, channel state information, entropy, autoregressive modeling, Kernel regression.

I. INTRODUCTION

Indoor Location-based Services (ILBS) have drawn tremendous attractions in recent years due to its huge potential values for wide-scale commercial and industrial applications [1]–[3], such as tracking products through manufacturing lines, shop advertising for target customers, security surveillance in banking system, first-responder navigation at medical center, etc.. Despite the up-to-date Global Navigation Satellite System (GNSS) can already provide precise location-aware information in the outdoor space, it is functionally ineligible indoors due to the signal blocking of architectures. Accurate, reliable and ubiquitous indoor localization solutions have been extensively studied in recent years. Examples include Wi-Fi [4]–[6], Bluetooth [7], Radio Frequency Identification (RFID) [8], Ultra Wideband (UWB) [9], infrared [10], visible light [11], sound [12], geomagnetic field [13] and so forth. Among these techniques, Wi-Fi based positioning is probably of the greatest popularity, mainly owing to the pervasive availability of the high-throughput and low-cost Wi-Fi technology. Accordingly, indoor position determination can be then operated in Wi-Fi based communication systems through firmware upgrades and software implementations.

In general, traditional indoor localization is commonly conducted via the geometric mapping or the Location Fingerprinting (LF)-based methods [14]. For geometric mapping, intermediate spatial parameters like distance or direction with regard to the Reference Points (RPs) are first derived from certain physical measurements. Typical parameters include Time of Flight (ToF) [15] and Angle of Arrival (AoA) [16]. Target's physical location can be further inferred by using geometric algorithms (e.g., trilateration or triangulation). Nevertheless, the performance of geometric mapping heavily relies on the Line-of-Sight (LoS) condition. The rich indoor multipath and shadowing effects, to a great extent, blur the monotonous relation between physical measurements and distances, complicate RF propagation modeling, and thus degenerate positioning accuracy. As an alternative to analyze the sophisticated signal propagation, location fingerprinting adopts a pattern-matching approach. The main idea is to collect signal features from predefined RP locations in the area of interest to construct a fingerprint radio map (a.k.a. site survey) in the offline phase. Subsequently, in the online

The associate editor coordinating the review of this manuscript and approving it for publication was Gongbo Zhou.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

phase, localization can be simply accomplished by matching the measured fingerprint at an unknown location with those in the offline database to return the best-fitted location estimation.

In IEEE 802.11n standard, Wi-Fi networks use Multiple Input Multiple Output Orthogonal Frequency Division Multiplexing (MIMO-OFDM) technique to modulate data on different orthogonal sub-channels and transmit them over multiple transmit-receive (TX-RX) antenna pairs simultaneously. Therefore, it can reflect the fine-grained channel feature known as channel response, which can be partially extracted from many commercial off-the-shelf Wi-Fi Network Interface Cards (NICs) [17], [18] in the format of Channel State Information (CSI). Specifically, CSI is aggregated by a set of channel estimations depicting the amplitude and phase information of each OFDM subcarrier. Different from the Received Signal Strength (RSS) as the indicator of the Medium Access Control (MAC) layer's link quality, the Physical (PHY) layer CSI measurement can serve as a preferable location signature characterized by the small-scale multipath fading, which significantly deteriorates the quality of its RSS counterpart. Furthermore, CSI indicates channel qualities on the level of multiple subcarriers and thus provides richer location-specific information than RSS-based localization schemes.

Generally, indoor location fingerprinting algorithms consist of deterministic and probabilistic methods [5]. Deterministic algorithms can be easily implemented but fail to fully exploit information about channel fluctuations in the environment and consequently lead to error-prone location estimate. On the contrary, probabilistic algorithms embrace the channel variation by inferring a probabilistic model reflected by the signal distribution, thus achieving better localization performance than deterministic ones [19]. The most adopted hypothesis in fingerprint localization problem is to model the signal signatures (RSS or CSI power summation) as well-known Gaussian random variables [19], [20]. In practice, however, we experimentally observe that certain devices record noisy signals with intricate non-Gaussian distributions, which complicate the fingerprinting process and incur ambiguity for location estimation. Moreover, the complex nature of indoor scenario makes probabilistic approaches more vulnerable when the signal statistics appear in the format of multivariate structure (e.g. multi-subcarrier CSI). Therefore, a fingerprint which shares the simplicity of RSS (scalar) but also preserves rich statistical location-dependent information (like CSI) would be highly desirable.

In this paper, to address the above underlying challenges for Wi-Fi fingerprinting based localization, we resort to the Shannon entropy metric which derives from the direct probabilistic transformation of the channel state information. This entropy-based channel interpretation enables us to perceive the indoor statistical diversity with a new insight. Since we avoid the massive storage of measurements driven by accurate Probability Density Function (PDF) estimation, this entropy transformation process largely reduces the offline fingerprint storage as well as the online computational complexity of pattern-matching. Through extensive experiments conducted in a typical indoor environment, we will show that the PDF of the amplitude for each OFDM subcarrier can be converted into entropy metric which outperforms its original channel response or RSS fingerprints. The reason why we ignore the phase information in this case is that CFR phases of one subcarrier are normally uniformly distributed [21], which incurs equally maximal entropy values for all locations and thus makes the phase-based entropy more trivial in our localization problem.

Particularly, we design *EntLoc*, a CSI entropy based probabilistic indoor localization system using commodity Wi-Fi devices. To reduce the measurement noise, we first convert the raw CSI measurements into the time domain Channel Impulse Response (CIR) via Inverse Fast Fourier Transform (IFFT). A power-based tap filtering scheme is then introduced to preserve the location-specific CIR taps, which contains the most relevant information related to multipath effects. After applying Fast Fourier Transform (FFT), a smoothed version of Channel Frequency Response (CFR) can be generated for the upcoming fingerprinting process. Subsequently, for the radio map construction, an Autoregressive (AR) modeling technique is adopted to accurately estimate the entropy of filtered CFR amplitude as fingerprint. Meanwhile, we experimentally observe that the entropy values belonging to two endpoint subcarriers of each RX antenna cause the most obvious ambiguity with other subcarrier values. This motivate us to eliminate these endpoint entropies, which further simplifies the fingerprint structure through this dimension reduction process. Accordingly, due to the structural simplicity of estimated CFR amplitude entropy, the succinct and convenient Manhattan distance [22] can be fully competent as the similarity metric in our online positioning phase. Afterwards, we leverage the optimal kernel regression scheme to accurately infer the target's location. The whole set of experiments were carried out on the lightweight HummingBoard platform [23], which tremendously facilitates the labor-intensive and time-consuming fingerprinting implementation. Furthermore, the experimental results also demonstrate the superior performance of our proposed system over other channel response based localization schemes.

In summary, our main contributions of this paper are set out below.

- As far as we are aware of, this is the first work to statistically study AR modeling based entropy signature in CSI fingerprint localization system. This simple fingerprint structure helps decrease the pattern-matching complexity and its informative statistical embodiment also facilitates the location estimation accuracy.
- We propose a power based pre-processing filtering scheme to mitigate the irrelevant noisy component in CSI measurements, thus further improving the location fingerprinting performance.

- We implement extensive positioning experiments on the lightweight HummingBoard Pro device, which remarkably enhances the experimental efficiency.

The rest of this paper is organized as follows. In Section II, we review the state-of-the-art related works. The CSI preliminaries are presented in Section III. We elaborate the overall architecture design of our proposed localization system along with detailed methodology in Section IV, and the experimental results are provided in Section V. Conclusions are drawn in Section VI.

II. RELATED WORK

Prevalent Wi-Fi fingerprint localization approaches mainly exploit two types of wireless signal properties: the received signal strength and the channel response. We present related works in accordance with these two categories.

A. FINGERPRINTING VIA RSS

Due to the easy acquisition of wireless signal power measurements, RSS-based fingerprinting plans have been widely adopted in various mainstream indoor positioning systems. *RADAR* [24] performed comprehensive site surveys for the first time and utilized RSS measurements recorded by wireless infrastructure to generate position fingerprints. Subsequently, the authors of *RADAR* used the deterministic k-Nearest Neighbors (kNN) technique to estimate the user's location with an average precision of 3 meters. In contrast, Youssef et al. employed the probabilistic approach in *Horus* [20] by using joint clustering algorithm and provided an accuracy improvement of 2.1 m, which outperformed *RADAR* even with less computational complexity. However, the instability of RSS still remains challenging. More recently, researchers of *LiFS* [25] brought up a novel fingerprint space by utilizing the spatial relations of RSS measurements, yielding low human cost for site survey and competitive accuracy over *RADAR*. Khatab et al. [26] used auto-encoder based deep extreme learning machine to extract high level data features from RSS fingerprint, which further improved the localization performance. Moreover, Wu et al. designed *DorFin* [27], a RSS-based location fingerprinting system which successfully tackled error mitigation problem by quantifying APs' distinction, alleviating RSS outliers and amending transitional RSS recordings. It reduced the mean and 95th percentile errors to respective 2.5 m and 6.2 m, outperforming both *RADAR* and *Horus* by nearly 50% accuracy improvement.

B. FINGERPRINTING VIA CHANNEL RESPONSE

In recent years, channel response based fingerprinting approaches have attracted massive attention due to their capability of harnessing the rich multipath information indoors. Sen et al. proposed the *PinLoc* [28], in which the localization process was conducted on a set of 1 $m \times 1$ m spots. The main observation in *PinLoc* was that the probability density function of CFR on a single subcarrier illustrated clustered distributions on the complex plane. Thus, the measurements of channel frequency response on each subcarrier were modeled to be Gaussian mixture distributed. The experimental result of *PinLoc* showed an 89% mean accuracy for 100 spots. In *FIFS* [19], the authors leveraged the spatial and frequency diversity of the channel response for Wi-Fi fingerprinting localization. In addition, *FIFS* used the power summation over all independent subcarriers as location fingerprint and applied Maximum A Posteriori (MAP) algorithm to achieve an improved accuracy over RSS based *Horus* system. Wang et al. presented *DeepFi* [29], a deep learning based indoor fingerprint positioning system using CFR information. *DeepFi* managed to train all the weights of a deep network as fingerprints in the offline stage, and used Radial Basis Function (RBF) based probabilistic method to acquire estimated location in the online stage. It achieved 20% accuracy improvement over *FIFS*. Furthermore, when it comes to the time domain CIR, authors in [30] proposed to exploit the amplitude of CIR (ACIR) vector to accomplish location estimation through nonparametric kernel regression scheme. Simulation results showed a distinguished performance superiority over the traditional RSS based fingerprinting methods.

III. CSI PRELIMINARIES

In wireless communication systems, the receiver conducts channel estimation through a mechanism named *channel sounding*. For the packet-based MIMO-OFDM IEEE 802.11n system, training sequences, namely High Throughput Long Training Fields (HT-LTF), are sent in the preamble, which is instantly used by the receiver to derive channel state information. Technically, CSI represents the PHY layer channel properties and reveals the combined effects of signal multipath propagation including amplitude attenuation and phase shift. Each CSI entry represents the Channel Frequency Response (CFR), which can be denoted as [14]

$$H(f_k) = \|H(f_k)\| e^{j\angle H(f_k)} \tag{1}$$

where $H(f_k)$ is the complex CFR at the subcarrier with central frequency of f_k. $\|H(f_k)\|$ and $\angle H(f_k)$ denotes its amplitude and phase, respectively.

In the time domain, to fully characterize the indoor multipaths, the wireless propagation channel is modeled as a temporal linear filter, known as Channel Impulse Response (CIR). Mathematically, it is expressed by [14]

$$h(\tau) = \sum_{i=1}^{T} \alpha_i e^{-j\varphi_i} \delta(\tau - \tau_i) \tag{2}$$

where α_i, φ_i and τ_i are the amplitude, phase and time delay spread of the ith path, respectively. T is the total number of multipaths and $\delta(\cdot)$ is the Dirac delta function.

However, practically, it is worth mentioning that, all the experiments in our work are based on Linux CSI tool [17], whose Intel 5300 NIC reports 30 out of 56 OFDM subcarriers for 20MHz bandwidth CFR packet. After applying IFFT on

measured CFR, the time domain CIR will be approximated with an equivalent number of 30 channel filter taps.

IV. SYSTEM DESIGN

In this section, we lay out the detailed architecture of the proposed indoor fingerprint positioning system.

A. OVERVIEW

The overall architecture of the proposed system is illustrated in Fig. 1. In general, it consists of two major functionality components: the offline fingerprint radio map construction and the online target's position estimation. For CSI fingerprint database construction, once getting the received raw CFR packets through signal war-driving, we first employ a tap filtering based pre-processing scheme to extract the most informative and location-dependent components in the multipath-rich indoor scenario. Subsequently, we model the statistical features of filtered CFR amplitude by calculating an AR modeling based entropy metric and then build a representative fingerprint radio map after removing the ambiguous endpoint subcarriers. Afterwards, for online location estimation process, when a mobile target arrives into the area of interest, it executes the same procedures to acquire the entropy vector and matches against the learned offline attributes. Finally, the simple Manhattan distance based kernel regression approach can be fully leveraged to accomplish the physical position estimation of the mobile target.

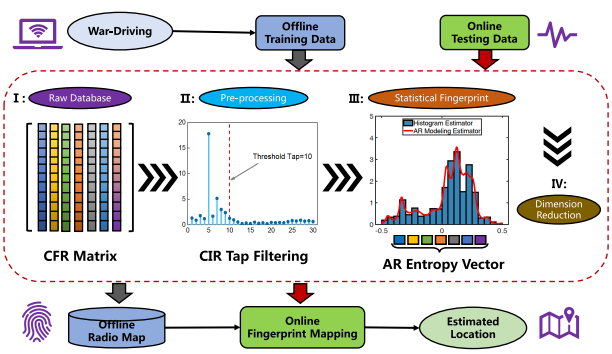


FIGURE 1. The proposed system architecture.

In what follows, we will dissect each component of the proposed system in a divide-and-conquer manner.

B. OFFLINE FINGERPRINT SITE SURVEY

1) PROBLEM FORMULATION

First of all, we begin the fingerprinting methodology with the presentation of problem formulation. In the offline stage, the area of interest is properly marked into M Reference Point (RP) locations. Each RP location is represented by a coordinate $\boldsymbol{\ell}_m = (x_m, y_m)$, where x_m and y_m are the X- and Y- coordinates of the mth RP location and $m \in [1, M]$. Considering that there are S Access Points (APs) as signal transmitters (TX) and one mobile user as the receiver (RX). Each AP has N_t TX antennas and the mobile user has N_r RX antennas. So each pair of TX-RX contains $N_t \cdot N_r$ antenna links. In addition, every CSI link has the same number of K OFDM subcarriers.

In the offline stage, the radio signature at each RP location is composed of concatenated CFR packets (samples) measured from S available APs. Thus each CFR signature has the total dimensionality of $S \cdot R$, where R is the signature's dimensionality from a single AP and $R = N_t \cdot N_r \cdot K$. To be more specific, the offline radio signature measured at m^{th} RP location from all S APs is given by the set $\mathcal{H}_m = \{\mathbf{H}_m^1, \ldots, \mathbf{H}_m^s, \ldots, \mathbf{H}_m^S\}$ and $s \in [1, S]$. Here $\mathbf{H}_m^s \in \mathbb{C}^{N \times R}$ contains N consecutive $1 \times R$ dimensional CFR samples measured at RP location $\boldsymbol{\ell}_m$ from the sth AP. This CFR matrix can be expressed by the following equation

$$\mathbf{H}_m^s = \begin{bmatrix} H_m^s(1,1) & \cdots & H_m^s(1,r) & \cdots & H_m^s(1,R) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ H_m^s(n,1) & \cdots & H_m^s(n,r) & \cdots & H_m^s(n,R) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ H_m^s(N,1) & \cdots & H_m^s(N,r) & \cdots & H_m^s(N,R) \end{bmatrix} \quad (3)$$

where $n \in [1, N]$ and $r \in [1, R]$.

In the online stage, a mobile device (target) at an unknown location $\boldsymbol{\ell}_\theta = (x_\theta, y_\theta)$ measures the CFR matrix from the sth AP which is given by $\mathbf{G}_\theta^s$ with the same dimension of $\mathbf{H}_m^s$. Likewise, the online measured CFR signature at the location $\boldsymbol{\ell}_\theta$ can be denoted by the set $\mathcal{G}_\theta = \{\mathbf{G}_\theta^1, \ldots, \mathbf{G}_\theta^s, \ldots, \mathbf{G}_\theta^S\}$. Accordingly, the mobile target's position can then be estimated as $\hat{\boldsymbol{\ell}}_\theta = (\hat{x}_\theta, \hat{y}_\theta)$ by exploiting these online measurements and the stored offline signal database.

2) CSI PRE-PROCESSING SCHEME

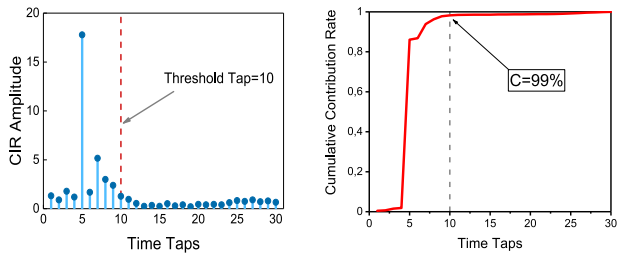
Recall that channel state information completely characterizes the multipath channel and preserves the fine location dependency, which makes it a good choice for location fingerprint. However, it would be fair to state that for the existing Wi-Fi networks, bandwidth limitation introduces severe location ambiguity which leads to limited localization accuracy [31]. By using the commodity Wi-Fi with center frequency of 2.4 GHz, the bandwidth of the system is therefore 20 MHz in this case. Since CFR can be converted into CIR via inverse fast Fourier transform, an estimation of CIR with time resolution of 1/20MHz = 50 ns is exposed. Since typical indoor maximum excess delay τ_{max} is smaller than 500 ns [32], given a time resolution of 50 ns, approximately only the first 10 out of the 30 accessible CIR time taps are relevant to multipath propagation. In other words, the remaining 20 taps are irrelevant for localization purpose. Moreover, when the Signal-to-Noise Ratio (SNR) is not high enough, the receiver's Additive White Gaussian Noise (AWGN) at these time taps will only make the accuracy worse.

Hence, based on the system bandwidth, a reasonable number of relevant time samples should be chosen for the sake of computation efficiency and accuracy. In this research, we design a power-based tap filtering method to preserve the most informative channel features for fingerprinting.

Specifically, for the conciseness of expression, we define the individual raw CFR signature as $\mathbf{H} \in \mathbb{C}^{1\times K}$. Through IFFT, we first convert $\mathbf{H}$ into the same dimensional CIR vector $\mathbf{h}$. For each $1 \times K$ CIR packet, we calculate the average channel power for each time tap, denoted by $U = (u_1, \ldots, u_k, \ldots, u_K), k \in [1, K]$, where $u_k = |\mathrm{h}[k]|^2$ and $\mathrm{h}[k]$ denotes the kth complex tap value of one CIR packet. Then, we define a cumulative contribution rate of the first k taps as

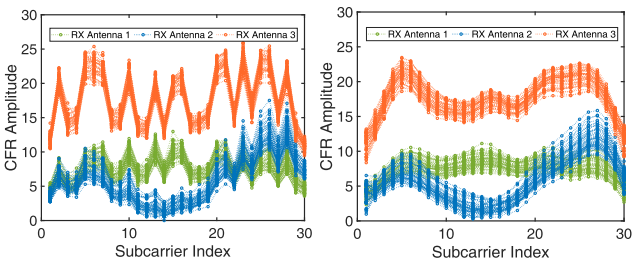
$$C_k = \sum_{i=1}^{k} u_i \Big/ \sum_{i=1}^{K} u_i. \tag{4}$$

If the cumulative contribution rate of the first L taps, i.e., C_L, is greater than the predefined threshold C, we then apply a simple rectangular window with length L to truncate the rest $(K - L)$ taps. Next, FFT is further utilized on filtered CIR to yield the smoothed version of CFR. As displayed in Fig. 2, we define the threshold of the cumulative contribution rate as 99%, the first 10 taps are thereby selected to preserve the most relevant multipath information for localization. Moreover, the differences between the raw CFR measurements and their smoothed versions can also be observed in Fig. 3.



(a) Channel Impulse Response (b) Power Threshold

FIGURE 2. Power threshold based tap filtering scheme: (a) channel impulse response; (b) power threshold.. (a) Channel Impulse Response. (b) Power Threshold.



(a) Raw CFR (b) CFR After Tap Filtering

FIGURE 3. Tap filtering based CFR smoothing: (a) raw CFR; (b) CFR after tap filtering.. (a) Raw CFR. (b) CFR After Tap Filtering.

3) AR-MODELING BASED ENTROPY ESTIMATION

As aforementioned in Section I, probabilistic positioning algorithms analytically outperform their deterministic counterparts. Additionally, numerous literatures [19], [20], [33] further reveal such superior localization performance of the probabilistic algorithms over their deterministic rivals in the complex indoor environment. Generic probabilistic methods include the Bayesian network [34], Kullback-Leibler Divergence (KLD) [35], [36], Gaussian process [37], etc.. The essential cause resides in the fact that PDF contains the complete statistical characterizations of the complex random variables, which are capable of providing better location-specific RF signatures.

The simplest probabilistic model for CFR is based on the assumption that there are a large number of statistically independent reflected and scattered paths with random amplitudes corresponding to a single subcarrier. By the central limit theorem, it can be reasonably modeled as circularly-symmetric Gaussian (complex Gaussian) random variables [38]. Thus, the amplitudes of the complex Gaussian process are essentially Rayleigh distributed. If the channel has a fixed LoS component, the received signal then equals the superposition of a complex Gaussian component and this LoS component. In this case, the CFR amplitude follows the Rician distribution. However, due to the sophisticated indoor environment and the imperfection of wireless devices, most measured CFR values are non-Gaussian distributed or even do not fit any known distribution [28], [39], [40]. Meanwhile, for multivariate fingerprint structure (e.g., multi-subcarrier CFR in our case), existing statistical tools only work under the condition of identifiably distributed measurements [36]. Besides, most probabilistic approaches require sufficient number of measurements stored in the fingerprint database, which guarantees an accurate PDF estimation but suffers huge system burden.

Therefore, in this paper, we resort to the well-known Shannon entropy [41] as the fingerprint alternative in our localization system. Given the offline and online CFR amplitude PDF estimates $\hat{p}_{\mathbf{H}_m^s}(\beta; r)$ and $\hat{p}_{\mathbf{G}_\theta^s}(\beta; r)$, both of which are from the rth subcarrier and the sth AP. For the simplicity of presentation, here we define β as a general expression of CFR amplitude from the same subcarrier. Thus, the offline entropy definition can be expressed by

$$\hat{\phi}^r_{\mathbf{H}_m^s} = -\int_{-\infty}^{\infty} \hat{p}_{\mathbf{H}_m^s}(\beta; r) \log \hat{p}_{\mathbf{H}_m^s}(\beta; r) d\beta \tag{5}$$

Similarly, the online CFR entropy $\hat{\phi}^r_{\mathbf{G}_\theta^s}$ can also be calculated as the fingerprint for the subsequent stage of target's location determination.

In practice, it is a challenging task to implement direct evaluation of the Shannon entropy from real data [42], [43]. The reason behind this dilemma is twofold: (i) Entropy has to be approximated from the mere sample data due to the fact that probability density function is generally unknown. (ii) Equation (5) requires numerical integration since a closed-form solution of the entropy does not exist. Typical data-adaptive PDF estimation methods comprise histogram estimator [44], order statistics [45] or kernel method (a.k.a. Parzen method) [46]. However, all of them share the major drawback of slow convergence rate.

In this paper, we address the entropy estimation problem by leveraging the more accurate and consistent Autoregressive (AR) modeling approach [47]. The basic principle of this approach is to estimate the unknown PDF in the form of Power Spectral Density (PSD) of an unit variance AR process. This unit variance condition ensures that PSD shares the basic requirements of PDF (i.e., positive function that integrates to one).

Given the general expression of amplitude β, we define the input CFR amplitudes from one certain subcarrier as $\boldsymbol{\beta}_I = [\beta(1), \cdots, \beta(n), \cdots, \beta(N)]^\top$, where N is the number of CFR packets and $(\cdot)^\top$ is the transpose operator. Since the law is modeled as the spectrum restriction on the interval of $[-0.5, +0.5]$, the amplitude data have to be first rescaled on this interval. Meanwhile, an order p AR process $W(n)$ is defined as the output of an all-poles filter driven by a white noise $\epsilon(n)$ with variance δ_ϵ^2. It can be mathematically denoted as [47]

$$W(n) = \sum_{i=1}^{p} a_i W(n-i) + \epsilon(n). \tag{6}$$

where $\mathbf{a} = \{a_i\}_{1 \le i \le p}$ are the AR model parameters.

Since the CFR amplitude PDF $p(\beta)$ can be equivalently depicted by the PSD $S_W(\beta)$ of this AR process which is parameterized by a set of AR parameters, the entire relations can be then presented as [48]

$$p(\beta) = S_W(\beta) = \frac{\sigma_\epsilon^2}{\left|1 + \sum_{i=1}^{p} a_i e^{-j2\pi i\beta}\right|^2}, \quad \beta \in [-0.5, 0.5] \tag{7}$$

where σ_ϵ^2 is the model prediction error which is chosen so that $\int_{-0.5}^{0.5} S_W(\beta) d\beta = 1$. It is notable that AR model order needs to be chosen appropriately at first since a low order leads to inadequate resolution (estimator bias) while a high order incurs spurious peaks (excessive variance). Through extensive experiments, a well-run model order selection technique known as the Exponentially Embedded Family (EEF) [49] is adopted to select a proper p which maximizes the following criterion.

$$\mathcal{F}(p) = \begin{cases} \xi_p - p(\log(\frac{\xi_p}{p}) + 1), & \text{if } \xi_p \ge p \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Here ξ_p is the Generalized Likelihood Ratio Test (GLRT) statistic which can be asymptotically computed as

$$\xi_p = (N-p)\log\left(\frac{\boldsymbol{\lambda}_p^\top \boldsymbol{\lambda}_p}{\boldsymbol{\lambda}_p^\top (\mathbf{I} - \Lambda_p \left(\Lambda_p^\top \Lambda_p\right)^{-1} \Lambda_p^\top) \boldsymbol{\lambda}_p}\right) \tag{9}$$

where $\boldsymbol{\lambda}_p = [\beta(p+1), \beta(p+2), \cdots, \beta(N)]^\top$ and $\Lambda_p = \left[\boldsymbol{\lambda}_{p-1}, \boldsymbol{\lambda}_{p-2}, \cdots, \boldsymbol{\lambda}_0\right]$. The detailed procedures are explicitly described in Algorithm 1.

Thereby, the succeeding task of estimating the AR parameters consists of two major steps [48]:

Algorithm 1 Model Order Selection Using EEF

Require:
N-dimensional CFR amplitude sample vector $\boldsymbol{\beta}_I$ of one certain subcarrier;
Predefined maximum AR model order p_{max}

Ensure:
Selected AR model order p

```
Rescale the input vector β_I into [−0.5, +0.5];
for each i ∈ [1, p_max] do
    Calculate the GLRT statistic ξ_i by (9);
    if ξ_i ≥ i then
        Obtain F(i) = ξ_i − i(log(ξ_i/i) + 1) by using EEF criterion in (8);
    else
        Make F(i) be zero;
    end if
end for
Execute p = argmax_{i∈[1,p_max]} F(i);
Return p;
```

(i) We first estimate the Autocorrelation Function (ACF) of the CFR amplitude data sequence $\boldsymbol{\beta}_I$ by applying the sample moment estimator, which is the statistical average correlation estimate:

$$R_W(i) = \frac{1}{N}\sum_{n=1}^{N} e^{j2\pi i\beta(n)}, \quad i \in [0, p] \tag{10}$$

(ii) AR coefficient estimation is then achieved by solving the Yule-Walker equations using the Levinson-Durbin recursion:

$$\mathbf{R}_W \mathbf{a} = -\mathbf{r}_W(0) \tag{11}$$

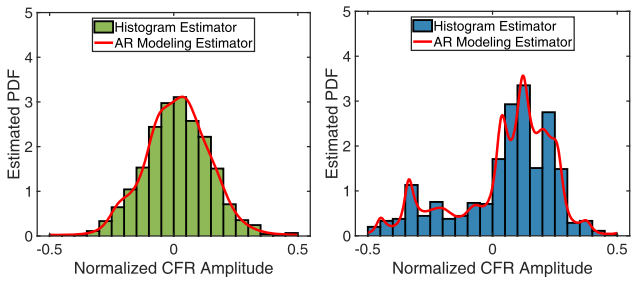
where $\mathbf{R}_W = [\mathbf{r}_W(1), \mathbf{r}_W(2), \cdots, \mathbf{r}_W(p)]$ and $\mathbf{r}_W(i) = [R_W(1-i), R_W(2-i), \cdots, R_W(p-i)]^\top$. Once the AR parameters have been estimated, say $\hat{\mathbf{a}} = [\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_p]^\top$, the AR model prediction error can be then computed by

$$\hat{\sigma}_\epsilon^2 = R_W(0) + \sum_{i=1}^{p} \hat{a}_i R_W(-i) \tag{12}$$

When AR PSD is determined, according to (7), the entropy estimation can be then converted to the following form:

$$\begin{aligned} \hat{\phi}_\beta &= -\int_{-0.5}^{0.5} \hat{p}(\beta)\log\hat{p}(\beta)d\beta \\ &= -\int_{-0.5}^{0.5} \hat{S}_W(\beta)\log\hat{S}_W(\beta)d\beta \end{aligned} \tag{13}$$

Additionally, a more feasible closed-form expression without any numerical integration can be obtained by applying Plancherel-Parseval formula to the right hand side of (13) [43]



(a) Location #1 (b) Location #2

FIGURE 4. AR modeling based PDF estimation at two sample locations.. (a) Location #1. (b) Location #2.

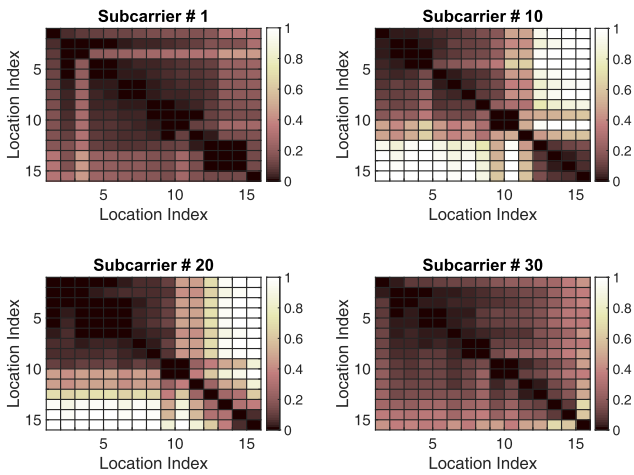


FIGURE 5. Ambiguity test for 4 subcarriers, namely #1, #10, #20 and #30.

Algorithm 2 AR Modeling Based Entropy Estimation

Require:
N-dimensional CFR amplitude sample vector $\boldsymbol{\beta}_I$ of one certain subcarrier;
Selected AR model order p from Algorithm 1

Ensure:
Estimated CFR amplitude entropy $\hat{\phi}_\beta$

1: Rescale the input vector $\boldsymbol{\beta}_I$ into $[-0.5, +0.5]$;
2: Compute the autocorrelation function $\{R_W(i)\}_{0\le i\le p}$ of the rescaled data via (10);
3: Calculate AR model parameters $\hat{\mathbf{a}}$ and prediction error $\hat{\sigma}_\epsilon^2$ by solving Yule-Walker equations in (11) and (12);
4: Estimate AR model PSD $\hat{S}_W(\beta)$ by applying (7);
5: Calculate AR entropy $\hat{\phi}_\beta$ using (14);
6: **Return** $\hat{\phi}_\beta$;

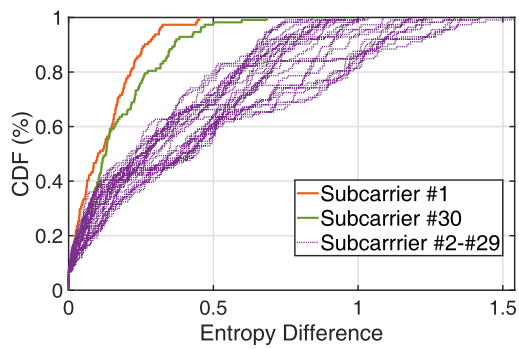


FIGURE 6. Ambiguity test for all 30 available subcarriers.

and yielding

$$\hat{\phi}_\beta = -\sum_{i=-\infty}^{\infty} R_W(i) Z_W^*(i) \tag{14}$$

where $(\cdot)^*$ is the conjugate operator and $Z_W(i)$ denotes the ith component of the AR process's cepstrum, which can be calculated by proceeding the inverse Fourier transform of $\log \hat{S}_W(\beta)$. The whole entropy estimation process is presented in Algorithm 2.

Fig. 4 illustrates the AR modeling based PDF estimates of CFR amplitude samples at two sample locations. Both of the estimated PDFs (Gaussian-like and non-Gaussian distributions) show good fit to the histograms which further justifies the AR-modeling scheme in practical work. Moreover, some shining points of AR modeling based entropy as location fingerprint will also be experimentally discussed in Section V-B.

4) ENDPOINT SUBCARRIER REMOVAL

In this part, we continue to exploit the AR modeling based entropy and shed light on some interesting observations in the sequel.

For location fingerprinting, the spatial resolvability is a key performance indicator for the proposed location fingerprint. In order to validate such property of our AR entropy-based fingerprint, a simple test was taken in our lab corridor. Concretely, we linearly selected 15 sample locations with 1 m spacing. A RF transmitter was placed at one end of the corridor, sending wireless packets continuously. In the meantime, we moved a mobile receiver in sequence at these sample locations. Around 500 CFR measurements were collected at each location. After calculating the AR entropies of all subcarriers at each position, we applied confusion matrix to portray the entropy differences among these 15 locations for each CFR subcarrier. In Fig. 5, for the visual clarity, we only exhibit subcarrier index 1, 10, 20, 30 and experimentally observe that the endpoint subcarriers #1 and #30 show a clear ambiguity in terms of location differentiation.

Given that this is only a visual indication, we then utilize the statistical Cumulative Distribution Function (CDF) to carefully study the behavior of these entropy differences. As depicted in Fig. 6, most subcarriers display an obvious entropy differences for different locations while the endpoint subcarrier #1 and #30 still show the opposite, inducing potential location differentiation errors in the next online pattern-matching stage.

Therefore, we propose in this paper to remove the two null endpoint subcarriers from the estimated AR entropies, which also serves as a dimension reduction strategy to

further improve the execution efficiency of our AR entropy approach. Recall that we combine the CFR measurements of the total R subcarriers from all N_r receiving antennas, the estimated AR entropy fingerprint in the offline stage can be hereby represented as

$$\hat{\mathbf{\Phi}}_{\mathbf{H}_m^s} = [\hat{\phi}^1_{\mathbf{H}_m^s}, \ldots, \hat{\phi}^r_{\mathbf{H}_m^s}, \ldots, \hat{\phi}^{R'}_{\mathbf{H}_m^s}], r \in [1, R'] \tag{15}$$

where $R' = R - 2 \cdot N_r$ is the reduced number of subcarriers in this case. Likewise, the online estimated AR entropy is denoted by $\hat{\mathbf{\Phi}}_{\mathbf{G}_\theta^s} = [\hat{\phi}^1_{\mathbf{G}_\theta^s}, \ldots, \hat{\phi}^r_{\mathbf{G}_\theta^s}, \ldots, \hat{\phi}^{R'}_{\mathbf{G}_\theta^s}], r \in [1, R']$.

C. ONLINE POSITION ESTIMATION

For the online location determination, the mobile target is required to be accurately mapped to the pre-designed radio map. To quantitatively measure the similarity between the stored entropy fingerprints and the estimated online CFR entropies, we employ Manhattan distance [22] which is also known as taxicab metric, capable of measuring the gap between two points through the summation of the absolute differences of their corresponding components.

Given the offline and online entropy fingerprints $\hat{\mathbf{\Phi}}_{\mathbf{H}_m^s}$ and $\hat{\mathbf{\Phi}}_{\mathbf{G}_\theta^s}$, we define the Manhattan distance between them as

$$\mathscr{D}_m^s = \|\hat{\mathbf{\Phi}}_{\mathbf{H}_m^s} - \hat{\mathbf{\Phi}}_{\mathbf{G}_\theta^s}\|_1 = \sum_{i=1}^{R'} \left|\hat{\phi}^i_{\mathbf{H}_m^s} - \hat{\phi}^i_{\mathbf{G}_\theta^s}\right| \tag{16}$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm. It concisely reveals the physical similarity between the online fingerprints at an unknown position and the offline dataset at the mth RP location, both of which are measured from the sth AP. Moreover, by using the chain rule for Shannon entropy [41], it can be proved that the Manhattan distance of a joint entropy of independent variables is equal to the sum of the distance for each variable's entropy. Under S independent AP assumption, we therefore have the Manhattan distance for all available APs as follows.

$$\mathscr{D}_m = \sum_{s=1}^{S} \mathscr{D}_m^s \tag{17}$$

In order to properly obtain the location estimation of the target, the weighted kernel regression is further adopted by employing the distance based kernel function $\mathcal{K}$ and the whole set of known reference points [36]. The estimated location can be derived from the following equation.

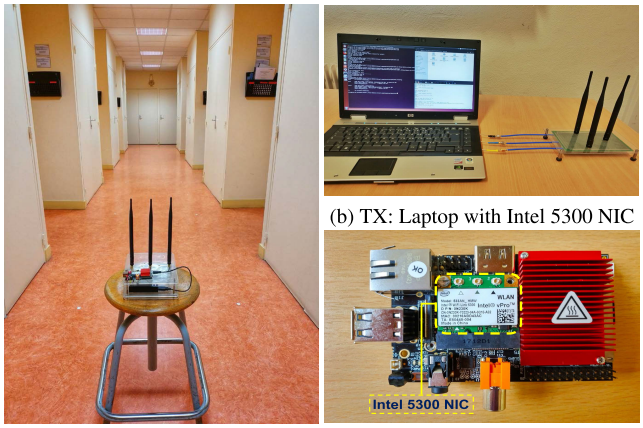
$$\hat{\boldsymbol{\ell}}_\theta = \frac{\sum_{m=1}^{M} \mathcal{K}_m \boldsymbol{\ell}_m}{\sum_{m=1}^{M} \mathcal{K}_m} \tag{18}$$

Here $\mathcal{K}_m$ is defined as the probability kernel of the mth RP position by exponentiating its corresponding Manhattan distance, which is presented as follows:

$$\mathcal{K}_m = \exp(-\rho \mathscr{D}_m) \tag{19}$$

where ρ is the kernel coefficient which is determined to optimally minimize the fingerprinting error by leave-one-out cross-validation in the offline phase [35]. It is notable that the kernel $\mathcal{K}_m$ is equal to one if the distributions of the given two fingerprints are identical (i.e., $\mathscr{D}_m = 0$) and decays to zero as the dissimilarity of the two fingerprints increases. In other words, this probability kernel provides a flexible way to naturally handle the CFR data and hence takes full advantage of our probabilistic AR entropy model, thus leading to an improved localization performance.

The performance of aforementioned fingerprinting approaches will be evaluated in the following section.



(a) CNAM lab scenario (b) TX: Laptop with Intel 5300 NIC (c) RX: HummingBoard Pro

FIGURE 7. CNAM lab scenario with HummingBoard Pro as receiver: (a) CNAM lab scenario; (b) TX: Laptop with Intel 5300 NIC; (c) RX: Humming Board Pro.. (a) CNAM lab scenario. (b) TX: Laptop with Intel 5300 NIC. (c) RX: HummingBoard Pro.

V. PERFORMANCE EVALUATION

In this section, we present the experimental evaluation of our proposed localization system. First of all, we start by introducing the experimental setup and the detailed implementation methodology. Then, the results of localization performance will be discussed in Section V-B.

A. EXPERIMENTAL SETUP

1) EXPERIMENTAL PRESENTATION

(a) **Environment**: The entire experiments are conducted in the CEDRIC laboratory of CNAM (a typical office environment in a multistorey building as shown in Fig. 7a). This lab office is a large room with an area of over 200 m^2. The indoor space is partitioned into several office and meeting rooms with many desks, chairs, computers, shelves furnished inside, which forms a complex radio propagation environment. The whole CSI database was collected during the working time in February, 2019.

(b) **Configuration**: We conduct our real experiments on commodity-ready off-the-shelf Wi-Fi devices [17]. Specifically, by working in the 5GHz band of IEEE 802.11n monitor mode, we use an HP Elitebook 8530w laptop as the signal transmitter (TX) and an HummingBoard Pro (HMB) as the mobile receiver (RX), which are exhibited in Fig. 7b and Fig. 7C. Both devices are equipped with Intel Wi-Fi Link (IWL) 5300 NIC and

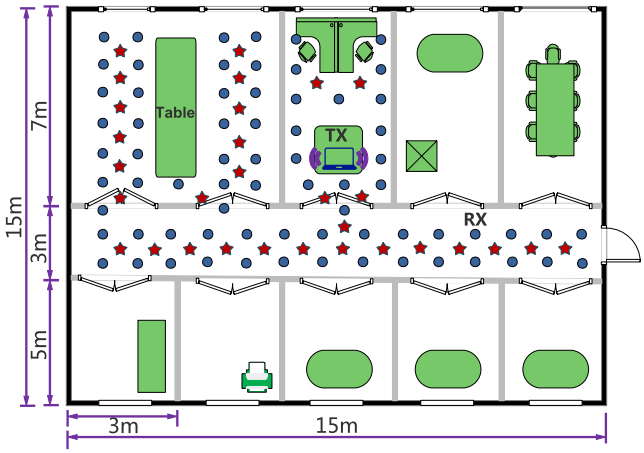


FIGURE 8. Floor plan of the lab scenario.

run 64-bit Ubuntu 14.04 OS and Debian 8.0 OS, respectively. Additionally, for our antenna settings, each Wi-Fi device is equipped with three omni-directional antennas to support 3×3 MIMO configuration.

(c) **Implementation**: As mentioned above, we implement the CSI data collection in our lab scenario. Fig. 8 shows the floor plan of this $15\,m \times 15\,m$ laboratory with a main corridor alongside several office and meeting rooms. The HP laptop serving as signal transmitter is fixed on the table of the central office room. Under injection mode, it is designed to intermittently transmit at the rate of 100 packets per second using only one transmitting antenna. It is worth mentioning that one transmitter setting is highly sufficient and well-performed in this lab scenario. If necessary, we may resort to multiple transmitters for the future larger testbed. The blue dots shown in Fig. 8 denote the 70 training reference points with one meter spacing and the 30 testing locations are marked as red stars. In the offline training phase, the CSI measurements are collected by the lightweight HMB at these reference points to build up the raw radio map. At each point, around 5000 CSI packets are stored as RF signatures in the firmware. In the online phase, we then move the HMB receiver among 30 testing locations to obtain the same size of CSI packets. In addition, all receiver ends are placed at the same height, constructing a simple 2-D platform for the precise indoor position estimation.

2) BENCHMARKS AND PERFORMANCE METRICS

In this part, we evaluate three existing probabilistic fingerprint positioning systems for the comparison purpose. As discussed in Section II, these include *Horus* [20], *FIFS* [19] and *PinLoc* [28]. Considering that the original *PinLoc* system conducted war-driving procedure in a set of predefined $1\,m \times 1\,m$ grids, known as spots, in order to provide a fair comparison, we modify *PinLoc* to use the same training set that we use in the proposed *EntLoc* system.

As for performance metrics, we define the localization error as Euclidean distance between the estimated location and the mobile user's actual position, which is presented as $\|\hat{\boldsymbol{\ell}}_\theta - \boldsymbol{\ell}_\theta\| = \sqrt{(\hat{x}_\theta - x_\theta)^2 + (\hat{y}_\theta - y_\theta)^2}$. When there are N_a testing locations, we evaluate the localization performance by using Mean Error (ME) metric which can be calculated as

$$ME = \frac{1}{N_a}\sum_{i=1}^{N_a}\sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2} \tag{20}$$

where (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$ are the actual and estimated coordinates at the ith testing location, respectively.

B. EXPERIMENTAL RESULTS

In this section, we evaluate the experimental performance and provide numerical results with relevant discussions.

1) AR ENTROPY PROPERTY STUDY

Since AR modeling based CFR amplitude entropy is the cornerstone of our fingerprint localization system, prior to accuracy analysis, we first evaluate the following two key characteristics of our proposed AR entropy fingerprint in location fingerprinting.

Temporal Stability: Practically, the channel response fluctuate frequently as the indoor environment varies over time. To investigate the robustness of our AR entropy based fingerprinting system, we design and implement a daytime measurement test in our lab. Specifically, the HummingBoard Pro

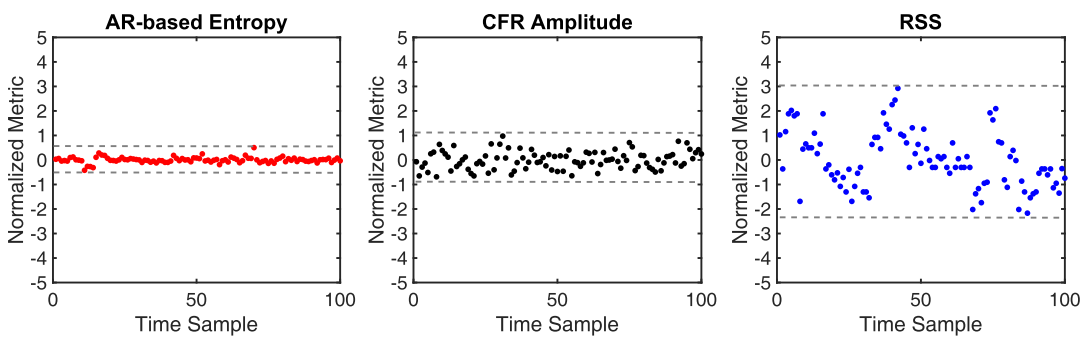


FIGURE 9. Temporal stability for three fingerprint signatures during the entire working times of one day (8 hours).

was configured to periodically record CFR measurements at a fixed position from a transmitter placed in the next-door room from 9 a.m. to 5 p.m. during a busy working day. About 500 CFR packets were collected every 10 minutes. Indoor furniture remained static with several personnel in the vicinity moving around. Next, we divide the whole measurements into 100 groups and compute the AR entropy, averaged CFR amplitude and the corresponding RSS mean value, respectively. For the purpose of fair comparison, we normalize the three metrics in the same range. As shown in Fig. 9, our AR entropy based fingerprint displays the lowest variance while the coarse-grained MAC layer RSS suffers the most severe fluctuations. It is reasonable that the environment changes do impact the time-varying channel response but cause less influence over its statistical entropy derivate.

Spatial Proximity: For indoor fingerprinting localization system, a good online signature is deemed to be qualified when it is capable of presenting similar trait with the offline signatures from the neighboring reference points. Based on the realistic testbed shown in Fig. 8, we experimentally chose two testing locations which are under LoS and NLoS condition, respectively. At each location, the multi-dimensional estimated AR entropy from all three RX antennas is compared with the entropy vectors of the corresponding four neighboring RP positions. Results illustrated in Fig. 10a demonstrate that for LoS test location #1, our AR modeling based entropy shows good spatial proximity with the fingerprints in the vicinity. For most subcarrier indices, the entropy value fits well in the center of its four neighbors. Especially for NLoS test location #2 in Fig. 10b, even though the neighboring entropies are relatively inconsistent (differ from different RX antennas), the overall multi-dimensional entropies at the center location can still capture the local minimal differences (Manhattan distance in our case) from its neighbors. This robust spatial property enables our AR modeling based

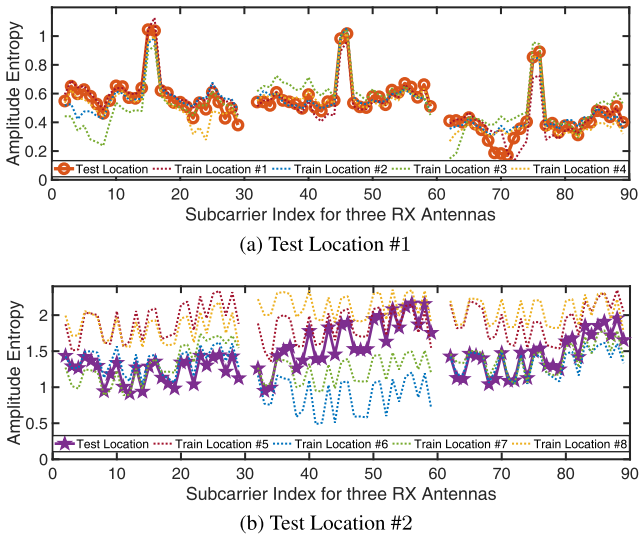


FIGURE 10. **Proximity test under (a) LoS condition; (b) NLoS condition.. (a) Test Location #1. (b) Test Location #2.**

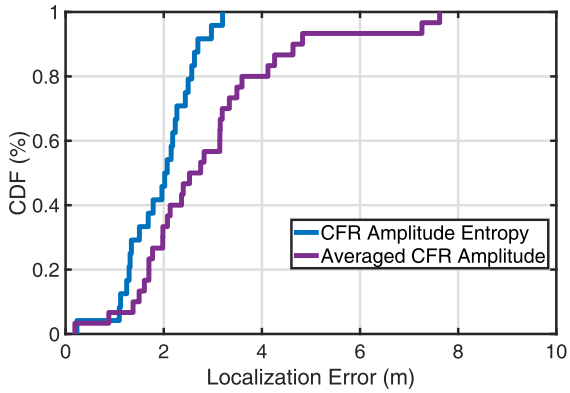


FIGURE 11. **Accuracy of AR entropy against its original CFR amplitude.**

entropy to be a strong candidate as fingerprint for most existing indoor positioning systems.

2) LOCALIZATION ACCURACY

This section provides a variety of numerical results in respect of localization accuracy, which firmly validates the superiority of our proposed localization system over other indoor geolocation schemes.

Comparison with CFR: We begin the localization accuracy evaluation by comparing our proposed AR entropy fingerprint with its original CFR amplitude. It is worth mentioning that these two fingerprint schemes follow the same online protocol (i.e., using Manhattan distance as similarity metric and kernel regression to figure out user location). Fig. 11 shows the CDF of localization errors for AR based entropy fingerprint and its original CFR amplitude. Specifically, our AR entropy approach shows a better performance with 90% positioning errors less than 2.69 m while CFR amplitude signature can only reach the level of 50th percentile. The 1.84 m ME of our proposed entropy scheme also precedes CFR amplitude based method whose mean sum error rises to 2.92 m. Since AR modeling based entropy accurately reflects the statistical distribution of the given CFR amplitudes, which unfortunately endure much more channel fluctuations, it can thus achieve better localization performance.

Comparison with CIR: Given that CIR is the inverse Fourier version of CFR, both of them should convey equivalent physical information. One may anticipate similar localization performance for these two channel response signatures. However, as shown in Fig. 12, our AR entropy based scheme maintains less than 2.69 meters localization error with the probability of 0.9, which outperforms CIR amplitude based entropy with only 63% percentage of the same positioning error. Meanwhile, the mean error of the proposed CFR amplitude entropy scheme is 1.84 m, which is also superior over CIR entropy approach with the mean error of 2.64 m. A possible explanation would be that most variations of CIR distribute within only a few time indices (i.e., first 10 taps), while the frequency diversity spans the entire range of CFR subcarrier indices, making the structures of CFR more

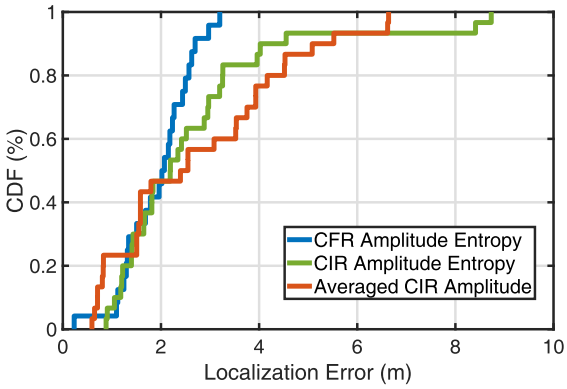


FIGURE 12. Accuracy of AR entropy against its time domain CIR.

TABLE 1. Detailed localization accuracy of all different methods.

Methods	Max. err.	Min. err.	Mean err.	Acc. at 90%
EntLoc	**3.20m**	0.23m	**1.84m**	**2.69m**
PinLoc-like	5.85m	0.46m	2.53m	4.15m
FIFS	7.70m	**0.15m**	2.83m	5.56m
Horus	9.77m	0.55m	3.50m	5.64m

distinguishable with each other [14]. Moreover, as expected, CIR entropy based fingerprint has better location estimation precision than its original CIR amplitude signature. The former precedes around 1.2 *m* localization error of the 90th percentile accuracy.

Comparison with the state-of-the-art: After comparing with the two most potential competitors, namely CFR and CIR amplitude schemes, our *EntLoc* system is then readily set to challenge other existing location fingerprinting systems. More specifically, as mentioned in Section V-A.2, we design a fair framework to compare our proposed AR entropy based localization approach with *PinLoc*-like, *FIFS* and *Horus* systems, respectively. As can be observed in Fig. 13, our proposed system achieves the 90th percentile error of 2.69 *m*, which outperforms *PinLoc*-like approach, *FIFS* and *Horus*

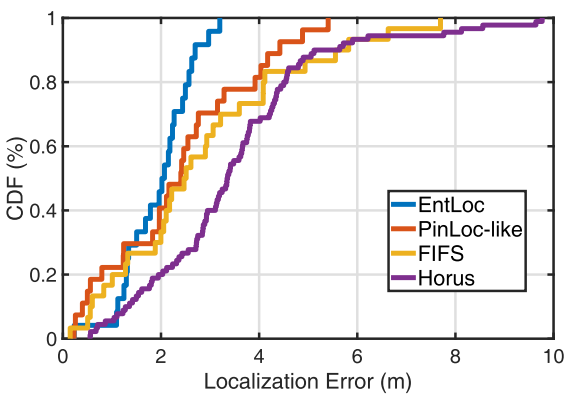


FIGURE 13. Accuracy of proposed *EntLoc* against state-of-the-arts.

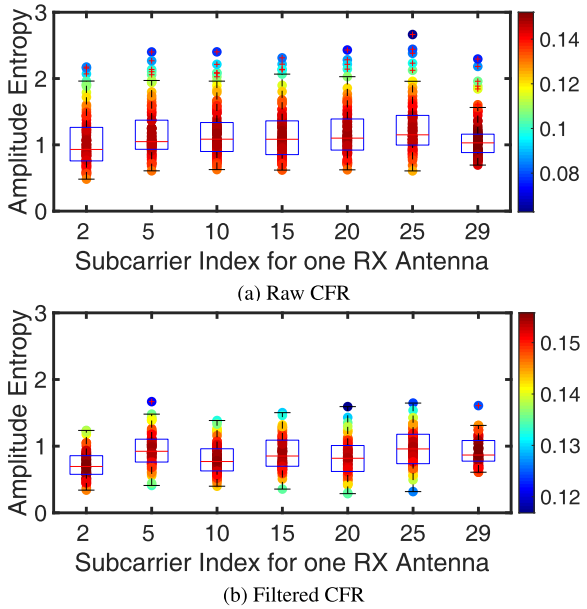


FIGURE 14. AR entropy box plot for (a) Raw CFR and (b) Filtered CFR.. (a) Raw CFR. (b) Filtered CFR.

with the same error level of 63%, 57% and 28%, respectively. Additionally, in order to provide an in-depth and comprehensive comparison for these localization systems, we enumerate the respective maximum error (Max. err.), minimum error (Min. err.), mean error (Mean err.) and the 90th percentile accuracy (Acc. at 90%) in Table 1. Apart from the 90th percentile accuracy, our *EntLoc* system is able to achieve the lowest mean error of 1.84 *m* compared with *PinLoc*-like, *FIFS* and *Horus* systems, improving the localization precision by 27.3%, 34.9% and 47.4%, respectively. As for maximum and minimum errors, *EntLoc* can still dominate the general accuracy evaluation. It only falls behind *FIFS* with 0.08 *m* in terms of minimum error, which can be neglected in realistic indoor environment.

3) IMPACT OF PRE-PROCESSING TECHNIQUE

Recall that we present a tap filtering based pre-processing technique in Section IV-B.2. Firstly the raw CFR measurements are converted into its time domain CIR by IFFT. Once removing irrelevant noise component in CIR, we can subsequently obtain a smoothed and finer version of CFR by applying FFT. It is interesting to study the impact of this approach to see how it can improve our localization performance. To this end, we design a fingerprint robustness based evaluation scheme. In particular, we manually record 10000 raw CFR measurements at one predefined location. By taking into account three RX antennas, we divide these CFRs into 100 subgroups and calculate the AR entropy of their amplitudes for each subgroup. Afterwards, we conduct the same procedures on the filtered CFR measurements and lay out the differences. As displayed in Fig. 14, we show the AR entropy box plot of selected subcarriers from one

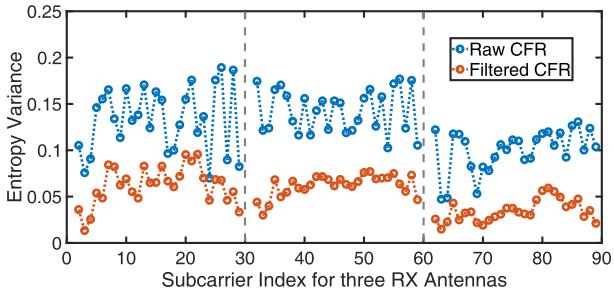


FIGURE 15. AR entropy variances for raw CFR and filtered CFR.

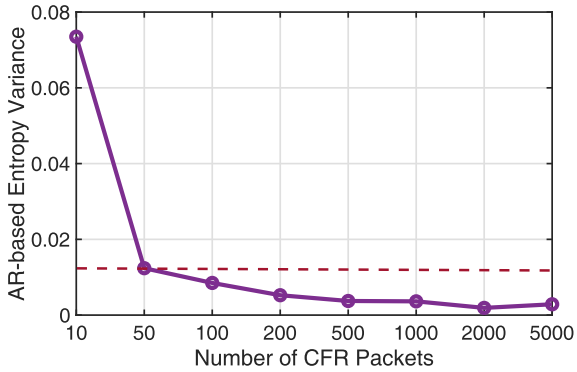


FIGURE 16. AR entropy variance changes with different CFR packet number selections.

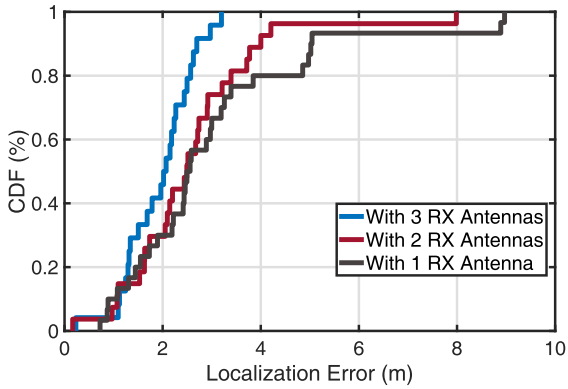


FIGURE 17. Localization accuracy under three different RX antenna configurations.

RX antenna out of visual clarity. Specifically, the filtered CFR entropy achieves less variation and reduces the statistical outliers to a great extent. Furthermore, we can also observe from Fig. 15 that for considering all three RX antennas, the filtered CFR entropy has an overall lower variance than its original raw CFRs. The above observations reveal that our pre-processing technique makes AR entropy based fingerprint more robust and can thus guarantee a preferable localization performance in the online location estimation phase.

4) IMPACT OF PACKET NUMBER FOR ENTROPY ESTIMATION

Since AR entropy estimation process requires sufficient CFR samples, larger number of samples can provide more accurate entropy estimation while increasing computational complexity. How to determinate the CFR packet number for entropy calculation becomes a trade-off problem which needs to be balanced in our localization system. Here we devise an AR entropy variance based scheme to select the optimal number of CFR packets. The motive lies in the fact that if the entropy variance is small enough, which can already guarantee a good accuracy, there is no need to import more CFR samples to increase computational burden. To be more specific, by testing the packet number ranging from 10 to 5000, we observe in Fig. 16 that 50 CFR packets can provide stable enough AR entropy estimates, which can further promote robust fingerprinting performance. So we choose and fix this packet number for all entropy estimation processes in our indoor positioning implementations.

5) IMPACT OF RX ANTENNA NUMBERS

In this part, we study the impact of RX antenna number on the localization performance. Intuitively, using more antennas at receiver end brings about more diverse channel response measurements, thus containing more location-specific information. We then study the localization accuracy differences for three RX antennas to deepen the understanding of our proposed localization system. As exhibited in Fig. 17, our AR entropy based localization system with three RX antennas is able to obtain superior estimation error precision over the same platforms with less antennas. Numerically, the three-antenna configuration can achieve less than 2.69 m localization error within the probability of 0.9, while the two and single antenna structures can only reach the same percentage level with the larger error of 4.1 m and 5.2 m, respectively. It validates the aforementioned assumption and encourages us to make full use of all three RX antennas in our indoor location fingerprinting system.

VI. CONCLUSION

In this paper, we presented *EntLoc*, an AR entropy based indoor location fingerprinting system using CSI amplitude information. In *EntLoc*, a tap filtering scheme was first utilized to remove the noisy component in raw CFR measurements. To capture the most informative statistical information of CFR while maintaining a simple structure, we adopted AR modeling based entropy as the fingerprint to construct a robust offline radio map. In the online phase, we proposed to use Manhattan distance as similarity metric and resorted to kernel regression scheme to infer the target's location. Experimental results from the lightweight HummingBoard device showed a superior localization performance of our proposed *EntLoc* system with an average accuracy improvement of 27.3%, 34.9% and 47.4%, in comparison with prominent *PinLoc*, *FIFS* and *Horus* system, respectively. In addition, we also examined the impacts of several different parameters on *EntLoc*'s performance, which enables us with deepening

insights to efficiently and productively implement our proposed localization system.

For our future work, since the phase information is abandoned in this work due to the entropy's limited capability of differentiating uniformly distributed signatures, one possible solution may turn to incorporating the CFR phase-based AoA information as the additional fingerprint to further improve localization accuracy.

ACKNOWLEDGMENT

The authors would like to thank Mr. Christophe Alexandre for his valuable assistance in the domain of Linux development as well as hardware implementation.

REFERENCES

[1] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568–2599, 3rd Quart., 2019.

[2] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A survey on wireless indoor localization from the device perspective," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 25:1–25:31, Oct. 2016.

[3] D. Dardari, P. Closas, and P. M. Djurić, "Indoor tracking: Theory, methods, and technologies," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1263–1278, Feb. 2015.

[4] C. Yang and H.-R. Shao, "WiFi-based indoor positioning," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 150–157, Mar. 2015.

[5] S. He and S.-H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 466–490, Jan. 2015.

[6] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, p. 46, 2019.

[7] P. Kriz, F. Maly, and T. Kozel, "Improving indoor localization using Bluetooth low energy beacons," *Mobile Inf. Syst.*, vol. 2016, Mar. 2016, Art. no. 2083094.

[8] Z. Zhou, L. Shangguan, X. Zheng, L. Yang, and Y. Liu, "Design and implementation of an RFID-based customer shopping behavior mining system," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2405–2418, Apr. 2017.

[9] B. Kempke, P. Pannuto, B. Campbell, and P. Dutta, "SurePoint: Exploiting ultra wideband flooding and diversity to provide robust, scalable, high-fidelity indoor localization," in *Proc. 14th ACM Conf. Embedded Netw. Sensor Syst. (CD-ROM)*, 2016, pp. 137–149.

[10] D. Hauschildt and N. Kirchhof, "Advances in thermal infrared localization: Challenges and solutions," in *Proc. IEEE Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2010, pp. 1–8.

[11] P. H. Pathak, X. Feng, P. Hu, and P. Mohapatra, "Visible light communication, networking, and sensing: A survey, potential and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2047–2077, 4th Quart., 2015.

[12] J. N. Moutinho, R. E. Araújo, and D. Freitas, "Indoor localization with audible sound—Towards practical implementation," *Pervasive Mobile Comput.*, vol. 29, pp. 1–16, Jul. 2016.

[13] Y. Shu, C. Bo, G. Shen, C. Zhao, L. Li, and F. Zhao, "Magicol: Indoor localization using pervasive magnetic field and opportunistic WiFi sensing," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1443–1457, Jul. 2015.

[14] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 25:1–25:32, Dec. 2013.

[15] N. Tadayon, M. T. Rahman, S. Han, S. Valaee, and W. Yu, "Decimeter ranging with channel state information," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3453–3468, Jul. 2019.

[16] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 269–282, 2015.

[17] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.

[18] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity Wi-Fi," *IEEE Trans. Mobile Comput.*, vol. 18, no. 6, pp. 1342–1355, Jul. 2018.

[19] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, "FIFS: Fine-grained indoor fingerprinting system," in *Proc. IEEE Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2012, pp. 1–7.

[20] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *Proc. 3rd Int. Conf. Mobile Syst., Appl., Services*, 2005, pp. 205–218.

[21] A. F. Molisch, *Wireless Communications*, vol. 34. Hoboken, NJ, USA: Wiley, 2012.

[22] E. F. Krause, *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*. New York, NY, USA: Dover, 1986.

[23] *HummingBoard Pro*. Accessed: Oct. 1, 2018. [Online]. Available: https://www.solid-run.com/product-tag/hummingboard-pro

[24] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. 19th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 2, Mar. 2000, pp. 775–784.

[25] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 269–280.

[26] Z. E. Khatab, A. Hajihoseini, and S. A. Ghorashi, "A fingerprint method for indoor localization using autoencoder based deep extreme learning machine," *IEEE Sensors Lett.*, vol. 2, no. 1, Mar. 2017, Art. no. 6000204.

[27] C. Wu, Z. Yang, Z. Zhou, Y. Liu, and M. Liu, "Mitigating large errors in WiFi-based indoor localization for smartphones," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6246–6257, Jul. 2017.

[28] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the Mona Lisa: Spot localization using PHY layer information," in *Proc. 10th Int. Conf. Mobile Syst., Appl., Services*, 2012, pp. 183–196.

[29] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.

[30] Y. Jin, W.-S. Soh, and W.-C. Wong, "Indoor localization with channel impulse response based fingerprint and nonparametric regression," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1120–1127, Mar. 2010.

[31] C. Chen, Y. Chen, Y. Han, H.-Q. Lai, F. Zhang, and K. R. Liu, "Achieving centimeter-accuracy indoor localization on WiFi platforms: A multi-antenna approach," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 122–134, Feb. 2017.

[32] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, H. Cai, Y. Liu, and L. M. Ni, "WiFi-based indoor line-of-sight identification," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6125–6136, Nov. 2015.

[33] S. Shi, S. Sigg, L. Chen, and Y. Ji, "Accurate location tracking from CSI-based passive device-free probabilistic fingerprinting," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5217–5230, Jun. 2018.

[34] R. Nandakumar, K. K. Chintalapudi, and V. N. Padmanabhan, "Centaur: Locating devices in an office environment," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 281–292.

[35] P. Mirowski, P. Whiting, H. Steck, R. Palaniappan, M. MacDonald, D. Hartmann, and T. K. Ho, "Probability kernel regression for WiFi localisation," *J. Location Based Services*, vol. 6, no. 2, pp. 81–100, 2012.

[36] L. Chen, I. Ahriz, D. Le Ruyet, and H. Sun, "Probabilistic indoor position determination via channel impulse response," in *Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2018, pp. 829–834.

[37] E. Homayounvala, M. Nabati, R. Shahbazian, S. A. Ghorashi, and V. Moghtadaiee, "A novel smartphone application for indoor positioning of users based on machine learning," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. ACM Int. Symp. Wearable Comput.*, 2019, pp. 430–437.

[38] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[39] N. Alsindi, Z. Chaloupka, N. AlKhanbashi, and J. Aweya, "An empirical evaluation of a probabilistic RF signature for WLAN location fingerprinting," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3257–3268, Apr. 2014.

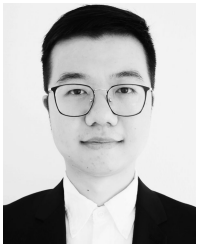
[40] X. Wang, L. Gao, and S. Mao, "BiLoc: Bi-modal deep learning for indoor localization with commodity 5GHz WiFi," *IEEE Access*, vol. 5, pp. 4209–4220, 2017.

[41] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.

[42] S. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[43] J. F. Bercher and C. Vignat, "Estimating the entropy of a signal with applications," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1687–1694, Jun. 2000.

[44] R. Moddemeijer, "On estimation of entropy and mutual information of continuous distributions," *Signal Process.*, vol. 16, no. 3, pp. 233–248, 1989.
[45] J. C. Correa, "A new estimator of entropy," *Commun. Statist.-Theory Methods*, vol. 24, no. 10, pp. 2439–2449, 1995.
[46] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.
[47] S. Kay, *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*, vol. 3. London, U.K.: Pearson, 2013.
[48] S. Kay, "Model-based probability density function estimation," *IEEE Signal Process. Lett.*, vol. 5, no. 12, pp. 318–320, Dec. 1998.
[49] S. Kay, "Exponentially embedded families-new approaches to model order estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 1, pp. 333–345, Apr. 2005.



LUAN CHEN received the B.E. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2013, and the M.S. degree from Wuhan University, Wuhan, China, in 2016. He is currently pursuing the Ph.D. degree with the Conservatoire National des Arts et Métiers (CNAM), Paris, France. His main research interests include wireless indoor localization and statistical signal processing.

INESS AHRIZ received the Ph.D. degree from the Pierre and Marie Curie University, Paris, France. Since 2011, she has been an Associate Professor with the CEDRIC Research Laboratory of Conservatoire National des Arts et Métiers (CNAM), Paris. Her research interests are indoor localization and machine learning.

DIDIER LE RUYET received the Eng. and Ph.D. degrees from the Conservatoire National des Arts et Métiers (CNAM), in 1994 and 2001, respectively. In 2009, he received the "Habilitation à diriger des recherches" from Paris XIII University.

From 1988 to 1996, he was a Senior Member of the Technical Staff at the SAGEM Defense and Telecommunication, France. He joined the Signal and Systems Laboratory, CNAM-Paris, as a Research Assistant, in 1996. From 2002 to 2010, he was an Assistant Professor with the Electronic and Communication Laboratory, CNAM-Paris. Since 2010, he has been a Full Professor at CNAM in the CEDRIC Research Laboratory. His main research interests lie in the areas of digital communications and signal processing, including channel coding, detection and estimation algorithms, filter bank based multicarrier communication, and multiantenna transmission. He has published about 200 articles in refereed journals and conference proceedings and published five books in the area of digital communication. He has been involved in several National and European projects dealing with multicarrier transmission techniques and multiantenna transmission. He has served as a Technical Program Committee member in major IEEE ComSoc and VTS conferences, and as the General Chair of the ninth edition of ISWCS'2012 Conference and the Co-Chair of the ISWCS 2013 and 2019 edition.