LiFi: Line-Of-Sight Identification with WiFi

Zimu Zhou*, Zheng Yang†, Chenshu Wu†, Wei Sun* and Yunhao Liu†

*CSE, Hong Kong University of Science and Technology

†School of Software and TNList, Tsinghua University

{zhouzimu, yang, wu, sunwei, yunhao}@greenorbs.com

Abstract—**Wireless LANs, especially WiFi, have been pervasively deployed and have fostered myriad wireless communication services and ubiquitous computing applications. A primary concern in designing each scenario-tailored application is to combat harsh indoor propagation environments, particularly Non-Line-Of-Sight (NLOS) propagation. The ability to distinguish Line-Of-Sight (LOS) path from NLOS paths acts as a key enabler for adaptive communication, cognitive radios, robust localization, etc. Enabling such capability on commodity WiFi infrastructure, however, is prohibitive due to the coarse multipath resolution with mere MAC layer RSSI. In this work, we dive into the PHY layer and strive to eliminate irrelevant noise and NLOS paths with long delays from the multipath channel responses. To further break away from the intrinsic bandwidth limit of WiFi, we extend to the spatial domain and harness natural mobility to magnify the randomness of NLOS paths while retaining the deterministic nature of the LOS component. We prototype LiFi, a statistical LOS identification scheme for commodity WiFi infrastructure and evaluate it in typical indoor environments covering an area of 1500m^2. Experimental results demonstrate an overall LOS identification rate of 90.4% with a false alarm rate of 9.3%.**

I. INTRODUCTION

The ubiquitously deployed WiFi networks indoors serve as more than a vehicle for communication. Fast emerging applications, e.g. indoor localization [1], through-wall tomography [2], human gesture recognition [3] etc., are continuously revolutionizing the horizon. For each innovative design to excel in multipath-dense indoor scenarios, Non-Line-Of-Sight (NLOS) propagation lurks as a primary threat that cannot be snapped shut outside. The severe attenuation of NLOS propagation deteriorates communication link quality and degrades theoretical propagation models. The past decade has witnessed extensive research to combat such a phenomenon [4]–[7], where the ability to distinguish Line-Of-Sight (LOS) and NLOS propagation acts as a fundamental primitive.

According to the NLOS/LOS conditions, PHY layer settings can be tuned for high throughput and reliable communication. For instance, in case of LOS dominant propagation, transmitters can switch to denser modulation and thus higher data rates [8]. Under severe NLOS conditions, on the other hand, particular receiver parameters (e.g. finger number of Raker receiver [9]) can be configured to remain effective with slightly higher complexity.

Besides adaptive wireless communication, numerous research domains also rely heavily on or even build upon the presence of the LOS path. For instance, NLOS propagation induces positive bias in time and power based ranging [10] [6], and generates spurious angular peaks for angle estimation [11] [12]. Even for fingerprinting-based localization, the fierce

signal strength fluctuations due to multipath superposition still pose substantial challenges in producing recurring radio fingerprints [13] [14]. The availability of a clear and short-range LOS path also benefits other novel applications such as wireless energy harvesting by ensuring tight electromagnetic coupling and thus high charging efficiency [15]. In a nutshell, the awareness of LOS and NLOS conditions, and further disentangling the LOS component, paves the way for and enhances all these frameworks.

Achieving such capability with commodity WiFi infrastructure, however, entails a range of challenges. Although vast theoretical channel models have been proposed for LOS and NLOS propagation [16], a practical LOS identification scheme either requires precise channel profiles, which involves dedicated channel sounders, or assumes abundant randomness to bring the statistical models in effect. Towards more pervasive solutions, most existing approaches either employ extremely wideband signals like UWB [17], or resort to relatively long-range communications like cellular networks [18], and often halt at simulations. Unfortunately, current WiFi operates with a bandwidth of only 20MHz, thus unable to resolve paths with distance difference shorter than 15m, yet often targets at inbuilding services of meter-level accuracy. Such scale mismatch of operating bandwidth and geographic space hampers direct adoption of either category of existing approaches to WiFi due to the coarse-grained channel measurements and short-range indoor propagation environments. Pioneer works [11] [12] extend to the spatial dimension leveraging MIMO techniques, but still require hardware modification, impeding immediate viability.

In this work, we aim to design a pervasive primitive to identify the availability of the LOS path under multipath propagation with only commodity WiFi devices. Since the presence and obstruction of the LOS path are mutually exclusive, we harness the hypothesis test framework for statistical LOS identification [18]. To capture the distinctions between LOS and NLOS conditions with merely off-the-shelf WiFi infrastructure, we exploit two key observations. 1) The recently exposed PHY layer information on commercial WiFi devices reveals multipath channel characteristics at the granularity of OFDM subcarriers [19], which is much finer-grained than the traditional MAC layer RSSI. 2) The spatial disturbance induced by natural mobility tends to magnify the randomness of NLOS paths, while retaining the deterministic nature of the LOS path, thus facilitating LOS identification via the statistical characteristics of the received signals.

We prototype LiFi, a LOS identification scheme for commodity WiFi infrastructure. Leveraging the PHY layer channel state information reported by the off-the-shelf Intel 5300 Network Interface Card (NIC), we eliminate irrelevant noise

and NLOS paths with large delays. On observing that mobility magnifies the discrepancies between LOS and NLOS paths due to their intrinsic difference in spatial degree of freedom, we involve natural receiver movement (e.g. walking with an ultrabook at hand) to enhance LOS identification. Combined with mobility, we extract representative features from the envelope distribution of the filtered channel state information to quantify its distinctions under LOS and NLOS conditions. Through extensive evaluation, LiFi achieves an overall LOS detection rate of 90.42% with a false alarm rate of 9.34%, and proves to be robust to different propagation distances, channel attenuation and blockage diversity.

The main contributions of this work are summarized as follows:

- We exploit PHY layer channel state information to identify the availability of the LOS component in multipath-dense indoor scenarios. As far as we are aware of, this is the first LOS identification scheme built upon merely commodity WiFi infrastructure without hardware modification leveraging PHY layer information, which allows pervasive adoption.

- We harness natural mobility to magnify the distinctions between LOS and NLOS conditions, and put LOS identification into mobile context, indicating viability with truly mobile devices.

- We prototype LiFi, a pervasive LOS identification scheme and validate its performance in various indoor office environments covering a total area of 1500m^2. Experimental results demonstrate that LiFi outperforms RSSI based approaches, achieving both LOS and NLOS detection rates of above 90%.

In summary, the existence of the LOS path can be regarded as a primary characteristic of wireless channels. We envision the primitive to identify LOS and NLOS dominant conditions as an enhancement for current 802.11 standards and future communication protocols, and a synergy for myriad applications including AP association, network routing, topology maintenance, human-computer interaction, etc.

In the rest of this paper, we first provide preliminary background in Section II, introduce feature extraction in Section III, and detail our design in Section IV. Section V presents the performance evaluation. We review the related work in Section VI. Section VII concludes this work.

II. PRELIMINARIES

In essence, LOS identification is tasked to infer the channel state via certain feature metrics of the received signals. In this section, we review the statistical decision framework of LOS identification, followed by candidate features from both theoretic analysis and commodity WiFi infrastructure.

A. The LOS Identification Problem

Twisty corridors, capsuled rooms and scattering furniture indoors often create a labyrinth for radio signals, where they have to propagate via multiple intricate NLOS paths. As shown in Fig. 1, it is common for the LOS path to be mixed with multiple aliased NLOS paths (Case 1), or too
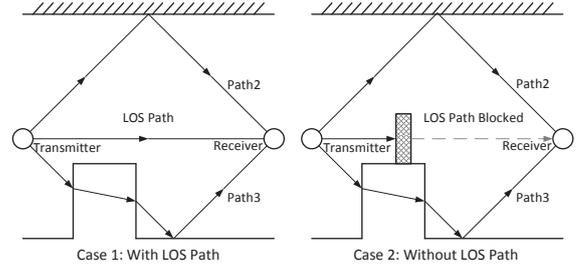

Fig. 1. An illustration of multipath propagation and LOS/NLOS conditions.

harshly attenuated to be perceivable against the noise floor (Case 2). Hence the LOS identification problem is to discern the availability of the LOS path under multipath propagation. Mathematically, the identification procedure is formulated as a binary hypothesis test with LOS hypothesis (H_0) and NLOS hypothesis (H_1) [20]. Given a generic feature metric ξ, the conditional Probability Density Function (PDF) under the two hypotheses $p(\xi|LOS)$ and $p(\xi|NLOS)$ are then applied to the classical decision theory with a likelihood ratio test:

$$\frac{p(\xi|LOS)}{p(\xi|NLOS)} \underset{H_1}{\overset{H_0}{\gtrless}} \frac{P(NLOS)}{P(LOS)} \tag{1}$$

where $P(LOS)$ and $P(NLOS)$ denote the prior probabilities of LOS and NLOS propagation, respectively.

A distinctive feature ξ, therefore, lies in the core of effective LOS identification schemes. Intuitively, since the LOS path, if present, always arrive ahead of NLOS paths, the delay characteristics of received signals differ under LOS and NLOS conditions. For instance, received signals under LOS condition normally experience smaller average delay, and this forms the basis for most UWB based LOS identification schemes, where high resolution channel information is available [21] [9]. In the time domain, the multipath channel is modeled as a temporal linear filter, known as Channel Impulse Response (CIR) [16] $h(\tau)$:

$$h(\tau) = \sum_{i=1}^{N} a_i e^{-j\theta_i} \delta(\tau - \tau_i) \tag{2}$$

where a_i, θ_i and τ_i are the amplitude, phase and time delay of the ith path, respectively. N is the total number of paths and $\delta(\tau)$ is the Dirac delta function. Various statistics depicting the average delay (e.g. mean excess delay [21]) are then utilized as indicators for LOS/NLOS conditions. Nevertheless, commodity wireless infrastructure often fails to support precise CIR estimation [6] [7]. Theoretical analysis hence resorts to modeling the distributions of the received signal envelope. For instance, received signal envelope exhibits Rician distribution under LOS propagation yet Rayleigh distribution in case of NLOS condition [20]. In practice, due to the noisy readings of conventional signal strength indicators (e.g., MAC layer RSSI), such frameworks generally halt at simulations or only partially applicable to long-range outdoor communications like cellular networks.

B. Channel State Information

Towards a practical LOS identification scheme with commodity WiFi infrastructure, we explore the recently available
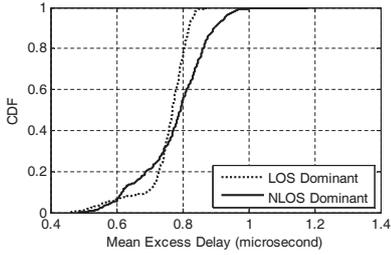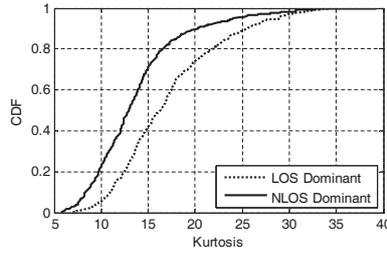
Fig. 2. Distributions of Mean Excess Delay

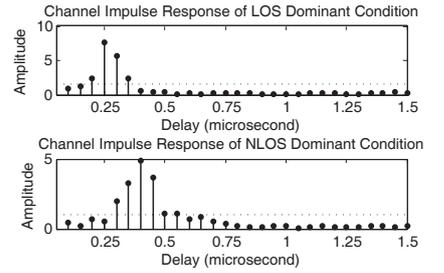Fig. 3. Distributions of Kurtosis of CIR

Fig. 4. CIR under LOS/NLOS Conditions

PHY layer information. Leveraging the off-the-shelf Intel 5300 NIC and a modified driver, a sampled version of Channel Frequency Response (CFR) within WiFi bandwidth is revealed to upper layers in the format of Channel State Information (CSI) [19]. Each CSI depicts the amplitude and phase of a subcarrier:

$$H(f_k) = \|H(f_k)\| e^{j \sin(\angle H(f_k))} \qquad (3)$$

where $H(f_k)$ is the CSI at the subcarrier with central frequency of f_k, and $\angle H(f_k)$ denotes its phase.

Since CFR can be converted into CIR via Inverse Fourier Transform (IFT), an estimation of CIR with time resolution of $1/20MHz = 50ns$ is exposed. Although its time resolution remains capable of discriminating only clusters of multipath components [6], CSI holds potential to eliminate the impact of irrelevant noise and NLOS paths with large delays when identifying the LOS path.

Compared with the conventional MAC layer RSSI, CSI portrays a finer-grained temporal and spectral structure of wireless links. However, the insufficient bandwidth and time resolution of WiFi with respect to UWB pose a series of challenges to codify the new information into effective features. We strive to extract proper features from CSI for LOS identification in the following section.

III. FEATURE EXTRACTION

Despite elegant theories underpinning delay characteristics based LOS identification, we demonstrate its limitations on bandwidth-constrained WiFi infrastructure even with PHY layer CSI. Combined with natural mobility, though, we observe the potential feasibility of envelope distribution based features. This section presents primary measurements from a typical office building, and focuses on mobility-enhanced envelope features for LOS identification.

A. Crude Delay Characteristics of CSI

The rationale for delay characteristics based LOS identification is twofold. (1) For a particular wireless link, signals transmitted via the LOS path always arrive first. (2) If the LOS path is not obstructed, it usually experiences weaker attenuation compared with NLOS paths. Prevalent feature metrics include mean excess delay [21] and kurtosis [9], which approximate the weighted average and peakedness of the received signal power delay profile, respectively. The limited bandwidth of current WiFi, however, yields insufficient multipath resolution, thus impeding direct adoption of this thread of features even with PHY layer CSI.

We extracted CSIs from 5000 packets measured under typical LOS and NLOS dominant conditions, and calculated the corresponding CIRs via IFFT. Fig. 2 and Fig. 3 illustrate the CDFs of the mean excess delay and kurtosis metrics, respectively. In general, LOS dominant conditions have shorter mean excess delay (i.e., shorter average delay) and larger kurtosis (i.e., a more sharply distributed power delay profile). However, a threshold to discriminate LOS and NLOS conditions would always lead to high false identification rate. The primary hurdle lies in the crude CIR samples. Given an operating bandwidth of 20MHz, commodity WiFi yields a time resolution of 50ns. Therefore paths with length difference smaller than 15m might be mixed in one CIR sample. Moreover, as shown in Fig. 4, there is an uncertain time lag at the start of measured CIR samples. In case of low time resolution and lack of synchronization, it is rather error-prone to align the CIR samples with respect to the first arriving path.

Although the CSI on current WiFi fails to estimate precise CIR, the emerging trend towards wider bandwidth WiFi (e.g. 802.11ac up to 160MHz) holds potential for practical LOS identification via delay characteristics. On the other hand, it suffices to extract signals of the dominant paths and eliminate irrelevant noise [6]. The following subsection hence explore the envelope distribution features of the dominant paths.

B. Envelope Distribution of Dominant Paths with Mobility

Unlike delay characteristics based LOS identification which models the time domain distinctions between LOS and NLOS conditions, envelope distribution based schemes explore the spatial domain. The intuition is that NLOS paths often involve large numbers of obstacles that reflect, refract and diffract wireless signals. Consequently, signals travelling along NLOS paths tend to behave more randomly compared with those along a clear LOS path. Therefore, from a statistical perspective, the distributions of received signal envelope differ under LOS and NLOS conditions due to varied extent of spacial randomness. In the communications communities, the envelope distribution is often modeled as Rayleigh fading for NLOS dominant conditions and Rician fading for LOS dominant conditions [16], and this has been numerically verified via simulation for long-range communication (e.g. cellular) [22]. For indoor WiFi networks, though, two challenges remain.

- Conventional MAC layer RSSI usually contains enormous noise, thus inducing irrelevant interference to the LOS path, making it less deterministic.

- Constrained by particular indoor floor plans and the relatively short transmission distances, the NLOS
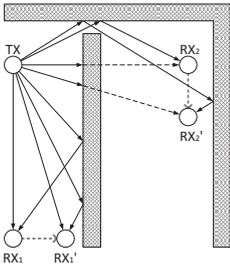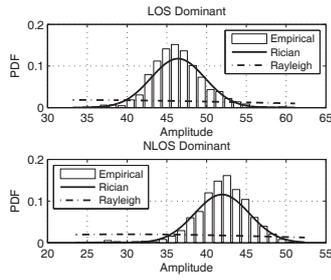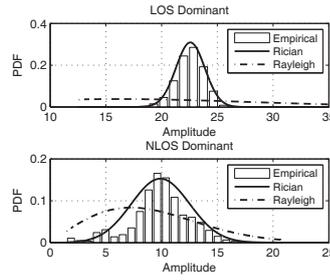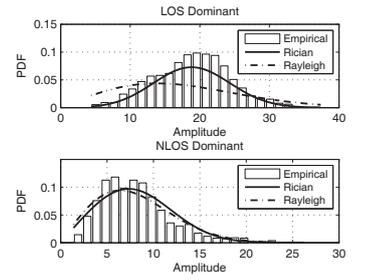
Fig. 5.   Impact of Mobility on LOS/NLOS Propagation

(a) RSS under Static Links

(b) CSI under Static Links

(c) CSI under Mobile Links

Fig. 6.   Received Envelope Distribution under LOS/NLOS Conditions

paths may not be adequately random, which potentially degrades the viability of theoretical models.

To wipe out unwanted interference when identifying the LOS path, we explore the PHY layer CSI instead of traditional MAC layer RSSI. As pointed out in [6], CSI provides a time resolution of multipath-clusters. We therefore exploit CSI to disentangle the dominant paths to mitigate the impact of NLOS paths with long delays as well as irrelevant noise to enhance the deterministic nature of the LOS path.

More specifically, we filter the CIR samples obtained from CSI as follows:

- Although the CSI tool [19] provides an estimation of the noise floor, which facilitates a threshold based signal arrival detection scheme, we find it more reliable to detect signal arrival by finding the maximum slope in a CIR sequence. The reason is that the slope based detection better captures the energy switch from noise to signals, even if the signal power is not strong enough.

- Since typical indoor maximum excess delay is smaller than 500ns [23], given a time resolution of 50ns, only the first 10 out of the 30 accessible CIR samples are relevant to multipath propagation. Accounting for the alignment errors due to uncertain time lag, the CIR sample next to the first detected path may contain the LOS path as well. We thus summate over the CIR sample with the maximum slope along with the next CIR sample for envelope distribution feature extraction.

Fig. 6a and Fig. 6b plot the distributions of received envelopes as well as Rician and Rayleigh fitting of overall RSS and filtered CSI for a static link. As is shown, the envelope of filtered CSI distributes more narrowly than that of overall RSS, indicating that weeding out irrelevant paths enhances the deterministic nature of the LOS path. However, although the empirical distributions are well fitted by Rician fading, Rayleigh fitting almost fails even for NLOS conditions. This seemingly counter-theoretical result indicates that it is possible for NLOS paths to behave deterministically in static environments with relatively short propagation distances and insufficient scatters. Therefore, it is still infeasible to simply employ filtered CSI for LOS identification, since it also reduces the randomness of NLOS paths.

To induce more randomness on NLOS paths, we explore to involve mobility into our scheme. The intuition is that, mobility may induce notable spatial disturbance on NLOS paths. As illustrated in Fig. 5, when Receiver1 moves from RX_1 to RX_1', the LOS path experiences only slight variation, while NLOS paths suffer notable changes in transmission distances, arriving angles, channel attenuation, etc. Such mixture of a clear LOS path with randomly attenuated NLOS paths is well captured by Rician fading. On the other hand, in case of undeceivable LOS path, almost all paths would fluctuate considerably during Receiver2's movement from RX_2 to RX_2', which would better fit Rayleigh fading due to the abundant randomness. Fig. 6c plot the envelope distributions of filtered CSI for a mobile link (by moving the receiver back and forth). As is shown, with natural mobility, the received envelope under LOS condition distributes almost symmetrically, while the distribution exhibits a notable skew under NLOS conditions and better fits Rayleigh fading.

C. Candidate Envelope Distribution Features

To codify the above observations into quantitative features, an intuitive method is to measure the envelope distribution of filtered CSI for a mobile link, and compare it with theoretical Rician and Rayleigh distribution. However, it often involves large amounts of measurement to obtain an accurate probability density estimation [20], which limits its realtime applicability. Towards a light-weight yet effective feature, we propose two candidates, Rician-K factor and skewness, for our LOS identification scheme.

1) Rician-K Factor: Rician-K factor [8] is defined as the ratio of the power in the LOS component to the power in the scattered NLOS paths. In theory, if the PDF of received signal envelope r obeys Rician distribution, it is associated with Rician-K factor K as follows:

$$p(r) = \frac{2(K+1)r}{\Omega} e^{(-K - \frac{(K+1)r^2}{\Omega})} I_0(2r\sqrt{\frac{K(K+1)}{\Omega}}) \quad (4)$$

where $I_0(\cdot)$ is the zero order modified Bessel function of the first kind, and Ω denotes the total received power.

Although the concept of Rician-K factor is built upon Rician distribution, its calculation does not involve probability density estimation and fitting. However, as demonstrated in Eq. 4, it involves a function inverse operation when obtaining an accurate Rician-K factor estimation from the sampled received

envelopes. In this paper we utilize a practical estimator for the Rician-K factor [8] leveraging only the empirical moments.

$$\hat{K} = \frac{-2\hat{\mu}_2^2 + \hat{\mu}_4 - \hat{\mu}_2\sqrt{2\hat{\mu}_2^2 - \hat{\mu}_4}}{\hat{\mu}_2^2 - \hat{\mu}_4} \quad (5)$$

where $\hat{\mu}_2$ and $\hat{\mu}_4$ are the empirical second and fourth order moments of the measured data, respectively. A large K indicates strong LOS power and thus, a high probability of LOS dominant conditions.

2) Skewness: Skewness is a general metric depicting the the skewed shape of a distribution. Mathematically, skewness s is defined as:

$$s = \frac{E\{x - \mu\}^3}{\sigma^3} \quad (6)$$

where x, μ and σ denote the measurement, mean, and standard deviation, respectively. A positive (negative) skewness indicates that the measured data spread out more to the right (left) of the sample mean.

As the received envelope of filtered CSI distributes more asymmetrically in NLOS conditions (Fig. 6c), we employ skewness as a candidate feature, which is agnostic to specific distributions and is more computation-effective. The skewed distribution of received signal envelope for NLOS paths has also been observed in [24], and modeled as a skewed Laplace distribution with respect to fade level, but it involves the prerequisite of transmitter-receiver distance between a static link. Conversely, we do not assume specific distribution and only calculate the skewness from filtered CSI for mobile links, which is irrespective of propagation distances, transmission power, and channel attenuation.

In summary, both Rician-K factor and skewness quantify the differences of the skewed envelope distribution under LOS and NLOS dominant conditions. In the next section we first present a generic LOS identification framework and postpone the detailed performance comparison between the two features to Section V.

IV. LOS IDENTIFICATION

In this section, we present our LiFi LOS identification scheme. The CSI samples reported from the receiver are first preprocessed to mitigate random phase noise and are normalized to eliminate the impact of transmitting power. The reassembled CFRs are then converted into CIR via IFFT. The two candidate features are extracted from a set of filtered CIR samples from N packets. The identification procedure is formulated as a statistical hypothesis test with a pre-calibrated threshold for each of the feature metrics. The following subsections elaborate on the detailed operations for each processing stage.

A. Preprocessing

The lack of time and frequency synchronization induces random phase noise when measuring the complex channel response at each subcarrier [25] [26]. Given the carrier frequency f, initial phase of $\phi_t(f)$ and propagation time t, the ideal received phase $\phi_r(f)$ is equal to $\phi_t + 2\pi ft$. However, the clock offset Δt and frequency difference Δf result in unknown phase shifts $2\pi f\Delta t$ and $2\pi\Delta ft$, respectively [25]. Since phase

shifts in the frequency domain is equivalent to delays in the time domain, the random phase noise leads to unknown time lags when calculating CIR samples from raw CSI samples as shown in Fig. 4. We hence utilize the linear revision as in [26] to mitigate the CIR aligning errors incurred by random phase noise, where the revised phase $\phi_r'(f)$ is equal to $\phi_r(f) - \alpha f - \beta$ and α and β denote the slope of the phase change and average phase change over all the subcarriers, respectively.

The revised phase information is thereafter reassembled with the amplitude measurement as the complex CFR samples. The corresponding CIR samples are obtained via a 32-point IFFT on the CFR samples. Envelope distribution features are thereafter extracted from the dominant paths detected by the maximum slope scheme as discussed in Section III-B.

B. Normalization

One advantage of the PHY layer CSI over the traditional MAC layer RSSI is that CSI eliminates the unknown transmission power and only characterizes the attenuation of the propagation channel. However, it is well-known that the radiation power decreases with propagation distance in free space. To make the LOS identification scheme independent of the power attenuation of the channel, we normalize the CIR samples measured at one time by dividing them by the average amplitude, i.e., setting the mean signal amplitude to 1, before extracting envelope distribution features from the CIR samples. The rationale for normalizing the mean signal amplitude to 1 instead of 0 is that the received signal amplitude is always greater than 0 (not in dB).

C. Identification

Given a set of normalized and filtered CIR samples from N packets, the Rician-K factor K and skewness s are calculated as introduced in Section III-B. Then LOS identification is formulated as a classical binary hypothesis test with LOS condition H_0 and NLOS condition H_1.

For the Rician-K factor, the hypothesis test is:

$$\begin{cases} H_0: & K > K_{th} \\ H_1: & K < K_{th} \end{cases} \quad (7)$$

and for skewness based LOS identification,

$$\begin{cases} H_0: & s < s_{th} \\ H_1: & s > s_{th} \end{cases} \quad (8)$$

where K_{th} and s_{th} represent the corresponding identification threshold for Rician-K factor and skewness, respectively. The thresholds are pre-calibrated and according to our measurements, a unified threshold for each feature metric would fit various scenarios including different propagation distances, channel attenuation, and blockage diversity.

V. PERFORMANCE

In this section, we first interpret the experiment setup and the methodology, followed by detailed performance evaluation of LiFi in various indoor scenarios, as well as comparative studies against RSS based approaches.
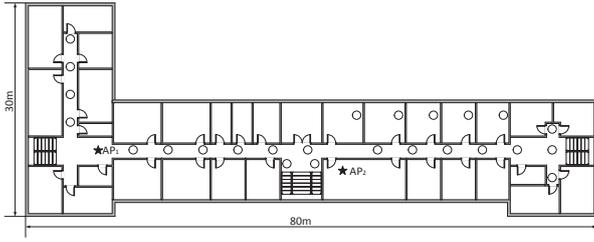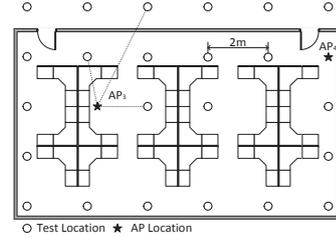
Fig. 7. Floorplan of the Testing Building

Fig. 8. Floorplan of the Testing Office

A. Methodology

We conduct the measurement campaign over one week in typical office environments including corridors and rooms, covering an area of approximately 1500m². The corridors are enclosed with concrete bearing walls and hollow non-bearing walls. The rooms are furnished with cubicle desks partitioned by glass and metal boards, computers, and other plastic, wooden and metallic furniture. The doors are kept open during the measurements and occasionally there are people passing by.

The floor plan of the testing building is illustrated in Fig. 7. For the corridors, we collect CSIs for LOS, through-wall and around-corner propagation with a maximum transmitter-receiver distance of 30m. For the rooms, we select a grid of 23 testing locations separated by 2m and 2 AP locations (Fig. 8). The direct link between one transmitter and one receiver could be a clear LOS path, partially blocked by furniture and humans, or through-wall propagation, as shown in Fig. 8. Different AP heights are also tested including 0m, 0.8m and 2m above the floor. To simulate natural human mobility, the receiver is placed on a wheeled desk of 0.8m in height, and is pushed by 2 different volunteers. For each measurement, the receiver moves randomly for 1m to 2m. We collect 2000 packets for each measurement, and in total we conduct 1000 measurements. For fair comparison, we also collect 1000 measurements for static links, with each measurement collected roughly along the user moving traces. Each category of measurements include 500 LOS dominant conditions and 500 NLOS dominant conditions.

During the measurements, a TP-LINK TLWR741N wireless router is employed as the transmitter operating in IEEE 802.11n AP mode at 2.4GHz. A LENOVO laptop equipped with Intel 5300 NIC and modified firmware as in [19] is used as the receiver pinging packets from the AP. A group of 30 CSIs are extracted from each packet and processed as in Section IV-A.

We mainly focus on the following metrics to evaluate our scheme. (1) LOS Detection Rate P_D: The fraction of cases where the receiver correctly identifies a LOS condition for all LOS cases. (2) False Alarm Rate P_{FA}: The fraction of cases where the receiver mistakes a NLOS condition for LOS condition for all NLOS cases. (Note that the NLOS detection rate is thus simply $1 - P_{FA}$.) The LOS detection rate and false alarm rate for Rician-K factor are defined as:

$$P_{D,K} = \int_{K_{th}}^{+\infty} f_K(\xi|H_0)d\xi$$
$$P_{FA,K} = \int_{K_{th}}^{+\infty} f_K(\xi|H_1)d\xi$$

where K_{th}, $f_K(\xi|H_0)$ and $f_K(\xi|H_1)$ denote the identification threshold and conditional probability densities under the two hypothesesfor Rician-K factor, respectively, while for skewness:

$$P_{D,s} = \int_{-\infty}^{s_{th}} f_s(\xi|H_0)d\xi$$
$$P_{FA,s} = \int_{-\infty}^{s_{th}} f_s(\xi|H_1)d\xi$$

where s_{th}, $f_s(\xi|H_0)$ and $f_s(\xi|H_1)$ represent the corresponding identification threshold and conditional probability densities under the two hypotheses for skewness, respectively.

B. Overall Identification Performance

Fig. 9 and Fig. 10 demonstrate the distributions of Rician-K factor and skewness under LOS and NLOS dominant conditions for (a) overall RSS[1] of static links (b) filtered CSI of static links (c) overall RSS of mobile links (d) filtered CSI of mobile links, respectively.

Static Links vs. Mobile Links: In general, the histograms of both Rician-K factor and skewness of static links span a larger range compared with mobile links, indicating mobility decreases the variations of the feature estimation. The rationale is that static links occasionally suffer from environmental dynamics (e.g. pedestrians), whereas the locomotion of mobile links tend to mask the impact of background instability. Therefore, mobile links are more robust to accidental disturbance since receiver motion dominates the changes of propagation paths. The disadvantage of mobile links, though, is that the LOS component also experiences fluctuations, thus leading to potential ambiguity between LOS and NLOS conditions. As shown in Fig. 9 and Fig. 10, the distributions of Rician-K factor under LOS conditions demonstrate a notable negative shift while those of skewness do a slight positive shift. The shifts under NLOS conditions, on the other hand, vary for the two features. According to our measurements, the distributions of Rician-K factor under NLOS conditions seem to experience less shift compared with those under LOS conditions, leading to more ambiguity for LOS identification (Fig. 9a and Fig. 9c). In contrast, the distributions of skewness under NLOS conditions are more sensitive to mobility induced path changes compared with those under LOS conditions, and hence, result in more distinguishable feature distributions.

---

[1]We summate over all the CIR samples to approximate the overall RSS. Note that such approximation in decibel is proportional to the MAC layer RSSI. The difference is that the summation of CIR samples characterizes only the overall channel attenuation, whereas RSS also contains unknown transition power. Therefore we believe that its relative performance would be at least comparable with RSS.
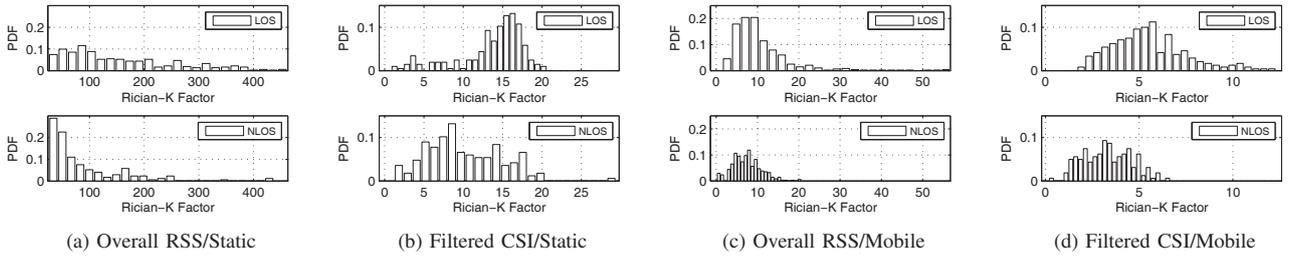
(a) Overall RSS/Static   (b) Filtered CSI/Static   (c) Overall RSS/Mobile   (d) Filtered CSI/Mobile

Fig. 9.   Distribution of Rician-K Factor under LOS/NLOS Conditions

(a) Overall RSS/Static   (b) Filtered CSI/Static   (c) Overall RSS/Mobile   (d) Filtered CSI/Mobile
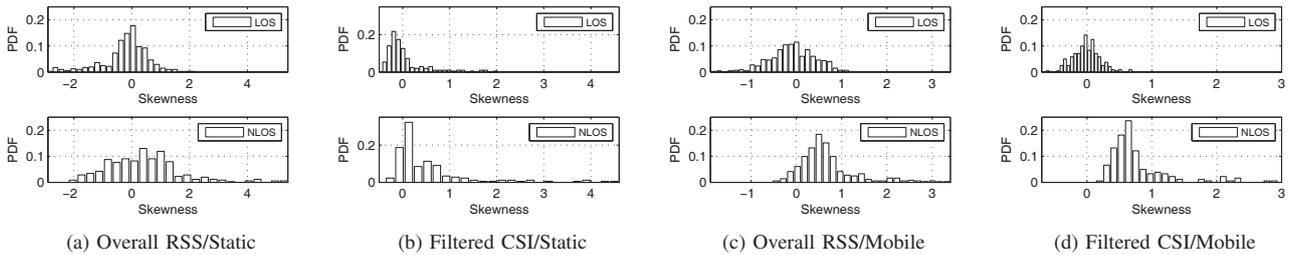
Fig. 10.   Distribution of Skewness under LOS/NLOS Conditions

Overall RSS vs. Filtered CSI: As shown in Fig. 9 and Fig. 10, the feature distributions of filtered CSI are more consistent than those of overall RSS. This is because the irrelevant noise and NLOS paths with long propagation distances are eliminated. Consequently, only the changes of the LOS path (if present) and NLOS paths with short delays are preserved. One exception, is that the distribution of Rician-K factor with overall RSS seems to span more concentratedly than that with filtered CSI under NLOS dominant conditions as shown in Fig. 9a and Fig. 9b. A partial explanation might be that Rician-K factor is more representative and effective for NLOS conditions in case of sufficient scatters. For static and relatively short distance indoor propagation scenarios, the noise and other NLOS paths in the overall RSS potentially compensate for the insufficient randomness, which possibly contributes to more consistent estimate of Rician-K factor.

Rician-K Factor vs. Skewness: To quantitatively evaluate the overall LOS identification performances of the two features, we plot the Receiver Operating Characteristic (ROC) curves of the two features in Fig. 11 and Fig. 12 for the above 4 combinations. ROC curves plot the LOS detection probability P_D against the probability of false alarms P_{FA}. It is a classical graphical view of the tradeoff between false positives and false negatives of a detection algorithm by evaluating a wide range of thresholds. In general, the ROC curve closer to the upper left corner indicates better overall detection performance.

As shown in Fig. 11 and Fig. 12, given a constant false alarm rate of 10%, the LOS detection rates using Rician-K factor are all below 60%, with the highest detection rate of 58.75% for mobile links with filtered CSI. With skewness features, in contrast, the highest detection rate is 90.83%. The reasons for such performance gap are as follows. The effectiveness of Rician-K factor implicitly relies on how the data fits Rician (or Rayleigh) fading. Despite human mobility compensated spatial disturbance, the abundance of scatters

between an indoor AP and a laptop is still incomparable with that within a base station and a mobile client at the center of the densely-built Manhattan, where these theoretical models cater for [27]. Conversely, skewness is applicable for any distributions, depicting the extent of leans of a distribution to one side of its mean. Therefore, skewness is a lightweight and general metric, and when combined with natural human mobility, outperforms other schemes and achieves an overall LOS identification accuracy of 90.42% and NLOS identification accuracy of 90.66%, respectively.

In the following subsections, we only utilize skewness and evaluate the impact of distances, number of packets and different obstacles on the identification performance using the optimal threshold of -0.205.

C. Impact of Propagation Distance

As LiFi aims to provide a generic LOS identification scheme, it is envisioned that a single threshold would fit a wide range of propagation distances. We collect data in the corridor with transmitter-receiver distances ranging from 5m to 30m. The corresponding LOS detection rates and false alarm rates are shown in Fig. 13. There is no direct correlation between the LOS identification performance and propagation distances, indicating a single pre-calibrated threshold tends to be independent of propagation distances and is applicable to most of the tested locations. Modest performance degradation, however, is observed for both short distance (5m) and relatively long distance (25m). The degeneration in short distance cases is partially due to the lack of randomness as discussed before, since short distance between the transmitter and the receiver would constrain the potential paths that wireless signals propagate, especially in the relatively narrow corridors. With distant receivers, on the other hand, the low SNR would incur enormous noise when estimating CSI, and consequently, degenerated skewness features.
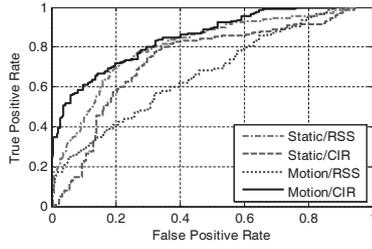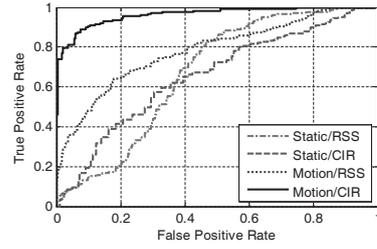
Fig. 11.   ROC Curve of Rician-K Factor
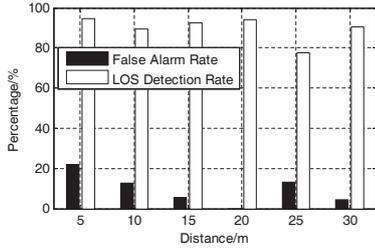

Fig. 12.   ROC Curve of Skewness

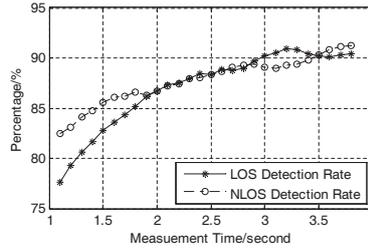Fig. 13.   Impact of Propagation Distances
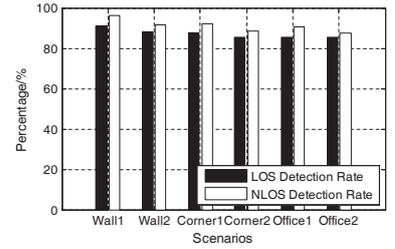

Fig. 14.   Impact of Packet Number

Fig. 15.   Impact of Obstacle Diversity

D. Impact of Packet Quantity

To evaluate the realtime performance of LiFi, we calculate the LOS and NLOS detection rate with different number of packets, ranging from 500 packets to 2000 packets per measurement. Since the receiver is downloading packets from the AP at the rate of 500 packets per second, this corresponds to a time range of 1s to 4s. As shown in Fig. 14, both the LOS and NLOS detection rates retain around 90% with 3 to 4 seconds of measurements. Even with measurements of only about 1s, LiFi demonstrates reasonable LOS and NLOS detection rates of 77.65% and 82.5%, respectively. The reason for the moderate performance degradation is that, although an effective skewness does not depend on a particular distribution, a stable estimation of skewness relies on adequate received envelope samples, especially with mobile links, unpredictable human behaviors and uncertain background dynamics.

E. Impact of Obstacle Diversity

To evaluate the robustness of LiFi under different obstruction scenarios, we investigate the identification performance with different blockage: (1) through wall (concrete bearing walls and hollow non-bearing walls), (2) around the corners (could partially be trough wall propagation) and (3) office (obstructed by metallic and wooden furniture). For each scenario, we test two AP locations as shown in Fig. 7 and Fig. 8, and plot the overall detection performances of the 3 scenarios as in Fig. 15. As is shown, there is no clear performance gap among the 3 scenarios with lowest LOS/NLOS detection rate of 85.42% for the 6 tested cases. The two wall cases slightly outperform the others partially because through wall propagation magnifies the difference between LOS and NLOS conditions. And although furniture blockage induces weaker attenuation and disturbance on the LOS component, the relatively open space in offices compared with narrow corridors allows more freedom of the propagation paths. Hence the detection performance is comparable with the other two

scenarios even with short propagation distances.

VI.   Related Work

The design of LiFi is closely related to the following three categories of research areas.

Channel Statistics based LOS Identification: To distinguish NLOS and LOS propagation via channel characteristics, various features have been proposed [20]. Since signals travelling along LOS paths always arrive earlier than those travelling along NLOS paths, delay characteristics such as mean excess delay [21] and kurtosis [9] are proposed. However they assume high resolution channel response and are thus primarily used in UWB systems. Narrowband and wideband systems often resort to analyzing the distributions of the more accessible signal envelopes from multiple packets at the cost of longer latency [22]. However, most of them mainly focused on analysis and simulation, with few practical solutions available, especially for WiFi. The aim of our work is therefore to implement a pervasive LOS identification scheme on commercial bandwidth-limited WiFi devices.

Leveraging MIMO and Mobility: Besides boosting wireless capacity [19], [28], the ever-increasing numbers of antennas on WiFi APs have also extended the context of LOS identification to the spatial dimension. The key insight is that a slight movement of the transmitting link induces little change for the LOS path, yet usually significant deviations for the NLOS paths. Xiong et al. [11] proposed a multipath suppression algorithm by removing peaks in the angle spectra that experience significant angular changes measured by two adjacent antenna arrays. Joshi et al. [12] determined the existence of LOS component via clustering on the angle and delay measurements calculated by cyclic correlation techniques. Sen et al. [7] weeded out NLOS components by verifying the angular changes at two distinct locations with the one extracted by distance measurements. Our work also grounds upon the stableness of LOS path and the random variations of NLOS

paths in case of slight link movements. However, instead of deriving angle information, our scheme analyzes the channel characteristics extracted from multiple packet and can be combined with MIMO based frameworks.

PHY assisted Indoor Localization: The availability of finer-grained channel state information on commercial WiFi devices has fostered thriving enthusiasm on utilizing PHY layer features for indoor localization with meter-level accuracy [29]. Sen et al. [26] and Xiao et al. [30] employed channel statistics as fingerprints for spot localization and fine-grained device-free motion detection. In this work, we exploit the channel state information available to capture the distinctive characteristics of LOS and NLOS conditions. Wu et al. [6] and Sen et al. [7] extracted the power and the arriving angle of the direct path for accurate ranging and direction estimation, respectively. In the absence of the LOS path, though, these frameworks might potentially mistake the first arriving signals along a NLOS path for those along the LOS path. Our work thus can serve as a prerequisite to first validate the existence of the LOS path.

VII. Conclusion

In this study, we explore PHY layer information to identify LOS dominant conditions with commodity WiFi infrastructure. On observing that natural mobility magnifies the randomness of NLOS paths while retaining the deterministic nature of the LOS component, we leverage the skewed distribution features of the received envelopes in mobile links, and prototype LiFi, a statistical LOS identification scheme with off-the-shelf 802.11 NIC. Extensive experimental evaluation considering various propagation distances, channel attenuation and obstruction diversity have validated the feasibility of LiFi, with an overall LOS detection rate of 90.42% and a false alarm rate of 9.34%. We envision this work as an early step towards a generic, pervasive, and fine-grained channel profiling framework, which paves the way for WLAN based communication, sensing and control services in complex indoor environments.

Acknowledgment

References

[1] M. Youssef and A. Agrawala, "The Horus WLAN Location Determination System," in *Proc. of ACM MobiSys*, 2005.

[2] F. Adib and D. Katabi, "See Through Walls with Wi-Fi!" in *Proc. of ACM SIGCOMM*, 2013.

[3] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-Home Gesture Recognition Using Wireless Signals," in *Proc. of ACM MobiCom*, 2013.

[4] A. Bhartia, Y.-C. Chen, S. Rallapalli, and L. Qiu, "Harnessing Frequency Diversity in Wi-Fi Networks," in *Proc. of ACM MobiCom*, 2011.

[5] Z. Yang and Y. Liu, "Quality of Trilateration: Confidence-based Iterative Localization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 5, pp. 631–640, 2010.

[6] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "CSI-Based Indoor Localization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 7, pp. 1300–1309, 2013.

[7] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Back to the Basics: Avoiding Multipath to Revive Inbuilding WiFi Localization," in *Proc. of ACM MobiSys*, 2013.

[8] C. Tepedelenlioglu, A. Abdi, and G. Giannakis, "The Ricean K factor: Estimation and Performance Analysis," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 799–810, 2003.

[9] L. Mucchi and P. Marcocci, "A New Parameter for UWB Indoor Channel Profile Dentification," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1597–1602, 2009.

[10] D. Giustiniano and S. Mangold, "CAESAR: Carrier Sense-based Ranging in off-the-shelf 802.11 Wireless LAN," in *Proc. of ACM CoNEXT*, 2011.

[11] J. Xiong and K. Jamieson, "ArrayTrack: A Fine-Grained Indoor Location System," in *Proc. of USENIX NSDI*, 2013.

[12] K. Joshi, S. Hong, and S. Katti, "PinPoint: Localizing Interfering Radios," in *Proc. of USENIX NSDI*, 2013.

[13] D. Turner, S. Savage, and A. C. Snoeren, "On the Empirical Performance of Self-Calibrating WiFi Location Systems," in *Proc. of IEEE LCN*, 2011.

[14] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Push the Limit of WiFi based Localization for Smartphones," in *Proc. of ACM MobiCom*, 2012.

[15] J. Lin, "Wireless Power Transfer for Mobile Applications, and Health Effects," *IEEE Antennas and Propagation Magazine*, vol. 55, no. 2, pp. 250–253, 2013.

[16] T. Rappaport, *Wireless Communications: Principles and Practice (2nd)*. Prentice Hall PTR, 2002.

[17] I. Guvenc, C.-C. Chong, and F. Watanabe, "NLOS Identification and Mitigation for UWB Localization Systems," in *Proc. of IEEE WCNC*, 2007.

[18] J. Borras, P. Hatrack, and N. Mandayam, "Decision Theoretic Framework for NLOS Identification," in *Proc. of IEEE VTC*, 1998.

[19] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 Packet Delivery from Wireless Channel Measurements," in *Proc. of ACM SIGCOMM*, 2010.

[20] C. Gentile, N. Alsindi, R. Raulefs, and C. Teolis, "Multipath and NLOS Mitigation Algorithms," in *Geolocation Techniques*. Springer, 2013, pp. 59–97.

[21] I. Guvenc, C.-C. Chong, F. Watanabe, and H. Inamura, "NLOS Identification and Weighted Least-Squares Localization for UWB Systems using Multipath Channel Statistics," *EURASIP Journal on Advances in Signal Processing*, no. 36, 2008.

[22] F. Benedetto, G. Giunta, A. Toscano, and L. Vegni, "Dynamic LOS/NLOS Statistical Discrimination of Wireless Mobile Channels," in *Proc. of IEEE VTC*, 2007.

[23] Y. Jin, W.-S. Soh, and W.-C. Wong, "Indoor Localization with Channel Impulse Response based Fingerprint and Nonparametric Regression," *IEEE Transactions on Wireless Communications*, vol. 9, no. 3, pp. 1120–1127, Mar 2010.

[24] J. Wilson and N. Patwari, "A Fade-Level Skew-Laplace Signal Strength Model for Device-Free Localization with Wireless Networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 947–958, 2012.

[25] J. Zhang, M. H. Firooz, N. Patwari, and S. K. Kasera, "Advancing Wireless Link Signatures for Location Distinction," in *Proc. of ACM MobiCom*, 2008.

[26] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are Facing the Mona Lisa: Spot Localization using PHY Layer Information," in *Proc. of ACM MobiSys*, 2012.

[27] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, and K. Huber, "Multiple-Input-Multiple-Output Measurements and Modeling in Manhattan," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 321–331, 2003.

[28] L. Fu, Y. Qin, X. Wang, and X. Liu, "Throughput and Delay Analysis for Convergecast with MIMO in Wireless Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 4, pp. 768–775, 2012.

[29] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor Localization via Channel Response," *ACM Computing Surveys*, vol. 46, 2014.

[30] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "FIMD: Fine-grained Device-free Motion Detection," in *Proc. of IEEE ICPADS*, 2012.