Wi-Fi butterfly effect in indoor localization: The impact of imprecise ground truth and small-scale fading

Andrei Popleteev
University of Luxembourg, Luxembourg
E-mail: contact@popleteev.com

***Abstract*—The increasing accuracy of indoor positioning systems makes their evaluation an increasingly challenging task. A number of factors are already known to affect performance of fingerprint-based systems: hardware diversity, device orientation, environment dynamics.**

This paper presents a new butterfly-like effect in localization experiments. The effect is caused by minor ground truth (GT) errors — that is, small deviations between calibration and test positions. While such deviations are widely considered as purely additive and thus negligible, we demonstrate that even centimeter-scale GT errors are amplified by small-scale radio fading and lead to severe multi-meter Wi-Fi positioning errors.

The results show that fingerprint-based localization accuracy quickly deteriorates as GT errors increase towards 0.4 wavelength (5 cm for 2.4 GHz). Beyond that threshold, system's accuracy saturates to about one-third of its original level achievable with precise GT. This effect challenges the impact of the already known accuracy-limiting factors (such as cross-user tests, receiver diversity, device orientation and temporal variations), as they can be partially explained by minor GT errors. Moreover, for smartphone-in-a-hand experiments, this effect directly associates the evaluation outcomes with experimenters' diligence.

***Index Terms*—Indoor localization, small-scale fading, ground truth, fingerprinting, radio propagation, Wi-Fi butterfly effect, performance evaluation, experiment design.**

I. Introduction

Due to the wide availability of Wi-Fi infrastructure and client devices, Wi-Fi fingerprinting represents one of the most used indoor positioning technology. A number of factors can affect signal distributions and thus performance of such systems: user body orientation, environment changes, human presence and hardware diversity [1]–[3]. These factors are well-known and their impact can be minimized by performing the evaluation on the same day, by the same person, holding the same device in the same way, in the same test points.

However, same-point evaluation is almost unachievable in practice. Due to the limited accuracy of current ground truth (GT) methodologies, actual test points can deviate from the calibration points by centimeters or even decimeters [4]. Given that typical Wi-Fi positioning errors are orders of magnitude larger [5], GT errors seem negligible and are generally ignored.

In this paper, we analyze the impact of small GT errors on the measured accuracy of a fingerprint-based indoor localization system. By testing the system in slightly different points than it was calibrated in, we demonstrate that minor GT errors are amplified by small-scale fading effects [6], and can substantially degrade system's accuracy. The contribution of this study is more methodological than technological, as we focus on accurate *evaluation* of localization performance rather than on methods for improving the latter.

978-1-5386-3089-1/17/$31.00

II. Background

A. Fingerprint-based indoor localization

Location fingerprinting based on Wi-Fi signals is one of the most widely used indoor localization approaches. It consists of two phases. In the calibration (training) phase, the environment is surveyed to build an empirical spatial model of signal fingerprint distribution. Here, a fingerprint is a vector of received signal strength (RSS) or channel state information (CSI) values received from stationary transmitters, such as Wi-Fi access points. In the localization (testing) phase, the system employs the created model and machine learning methods to identify locations by the fingerprints.

Over the last decades, multiple authors have comprehensively investigated various aspects of fingerprint-based localization. While the efforts have primarily focused on readily available Wi-Fi infrastructure [7]–[9], fingerprinting has also been successfully used with broadcast signals of opportunity, such as GSM [10] and FM signals [11]. A number of methods have been proposed for reducing the calibration efforts at the deployment phase, with a particular focus on crowdsourcing-based techniques [12]–[15]. Additionally, the accuracy of fingerprinting-based systems have been found to depend on a variety of factors, including testbed layout, its dimensions, orientation-dependent signal shadowing by user body, diverse characteristics of the client devices, and human activity nearby. A detailed review of current Wi-Fi fingerprinting systems falls beyond the scope of this paper; a recent survey can be found in [16].

B. Localization performance evaluation

The wide diversity of experimental conditions and evaluation methodologies represents a major challenge for fair comparison between different localization algorithms. This issue is currently addressed by formal indoor localization competitions, where participating systems are evaluated in the same conditions or on the same datasets [17]–[20]. The

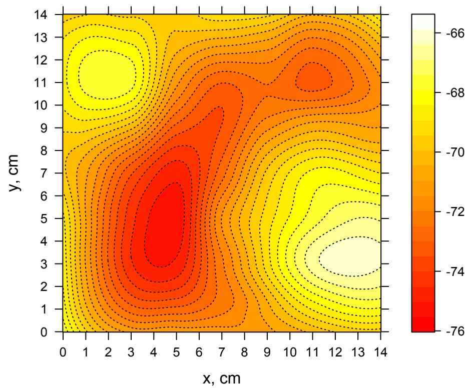

Fig. 1. Small-scale fading of Wi-Fi signals. Empirical measurements show 10 dB RSS variations in a 14×14 cm area sampled with 1 cm resolution.

same objective of fair comparison is pursued by automatic benchmarking platforms such as EVARILOS [21].

Performance evaluation of a localization system typically involves comparison of system's outputs with known ground truth (GT). However, the latter is also subject to measurement errors. Such GT errors become particularly important in experiments where the fingerprinting-based system is calibrated and tested in the same points. In reality, however, these 'same' points can deviate by up to several decimeters from their original positions, depending on the chosen GT methodology [4].

While such GT deviations are widely considered negligible in comparison to meter-scale Wi-Fi localization errors, in this paper we argue otherwise. Due to the specifics of radio wave propagation, GT errors may result in 'the same' point having very different calibration and testing fingerprints, with adverse impact on system evaluation results [22]. The next section provides the physical background of this phenomenon.

C. Spatial variations of RSS

Spatial variations of RSS fingerprints are defined by three main factors: free-space propagation loss, large-scale fading and small-scale fading [6]. *Free-space propagation loss (FSPL)* describes radio wave propagation as a function of the transmitter-to-receiver distance, without any obstacles. *Large-scale fading*, in turn, relates to shadowing by larger-than-wavelength obstacles, be it buildings or furniture. Finally, *small-scale fading* is caused by multipath propagation, where radio wave interferes with its own delayed and attenuated reflections from surrounding objects. The areas of constructive and destructive interference interleave at approximately $\lambda/2$ intervals (where λ is the wavelength), and thus even small sub-λ movements can significantly change signal reception conditions (see Fig. 1).

The impact of small-scale fading on indoor positioning accuracy has been addressed in several Wi-Fi based location tracking systems ($\lambda = 12.5$ cm). One of the first such systems, Horus, detected the influence of small-scale RSS variations by rapid changes of the estimated position during mobile client's movement through the environment [23]. The detected variations were mitigated by 'perturbing' the fingerprints; this technique improved the localization accuracy by 8% [23]. Another system, PinLoc [24], mitigated the impact of small-scale fading on Wi-Fi channel state information (CSI) fingerprints by collecting rich calibration data inside 1×1 m patches ('spots'), rather then in separate points. Lee et al. [25] proposed an extended Kalman filtering approach for simultaneous mitigation of RSS variations and position tracking. Other systems address small-scale fading indirectly, by filtering out too fast or out-the-map position changes [12].

Notably, all the mitigation methods are designed for *position tracking* in dynamic evaluation scenarios, where the user moves along a predefined GT path. However, multiple studies employ *single-shot positioning*, where the mobile device is evaluated in predefined test points and does not track transitions between them [26]. As a result, motion-based detection or mitigation of small-scale fading is not possible, and performance of the tested localization system theoretically becomes sensitive to GT precision.

To the best of our knowledge, previous work on fingerprint-based indoor positioning considered GT errors as purely additive and thus negligible. In contrast, this paper is the first detailed investigation of the impact of small GT errors on Wi-Fi localization performance.

III. Small-scale fading and RSS fingerprinting

Statistical theory of small-scale fading, developed by Clarke [27], provides analytical expressions for spatial correlation of received signals in environments with isotropic scattering (with later generalizations to non-isotropic scenarios [28], [29]). According to Clarke, when a receiver with two antennas moves through a multipath environment, signal amplitudes received by the antennas are correlated. For omnidirectional vertical whip antennas, their normalized covariance is [27, Eq. 16]:

$$\rho(d) \cong \left[J_0\left(\frac{2\pi}{\lambda}d\right)\right]^2$$

where J_0 is the zero-order Bessel function of the first kind, and d is the distance between antennas. The covariance function, shown in Fig. 2, has its maximum at $d = 0$ and gradually decreases until RSS values become completely uncorrelated at a distance of about 0.4λ (≈ 5 cm for 2.4 GHz Wi-Fi).

In order to investigate the impact of this spatial correlation on RSS fingerprints, we experimentally measured Wi-Fi RSS in a 14×14 cm area with a 1-cm grid step. Then we calculated pairwise Euclidean distances between all the test points, both in space and fingerprint domains. The resulting 25200 combinations were grouped by spatial proximity d and averaged. Groups with less than 100 measurements were discarded in order to reduce statistical noise.

The result, shown in Fig. 3, is well in line with Clarke's theory. Indeed, for nearby points ($d \leq 5\ cm$), even small movements quickly increase the distance between fingerprints; a linear fit has a 1.8 dB/cm increase rate. However, beyond the 0.4λ threshold, the impact of small-scale fading decreases and fingerprint changes are dominated by much slower large-scale effects (0.3 dB/cm).

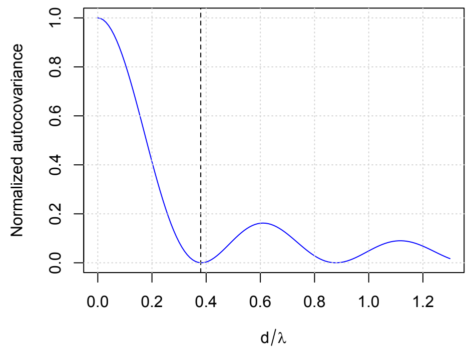

Fig. 2. Theoretical covariance of signal amplitudes received in two points distance d apart, in a multipath environment [27, Eq. 16].

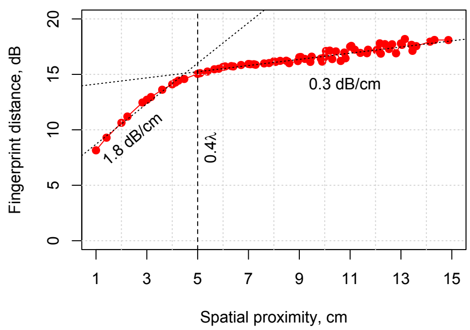

Fig. 3. Experimentally measured similarity between Wi-Fi RSS fingerprints in nearby points (indoor measurements).

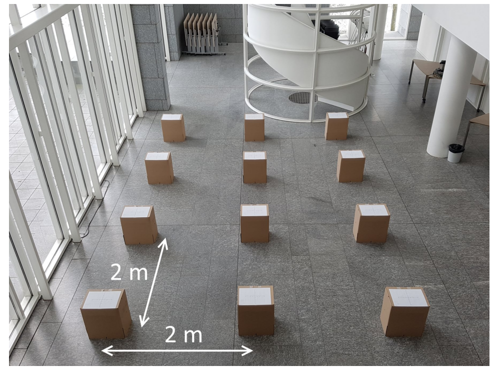

Fig. 4. Experimental testbed.

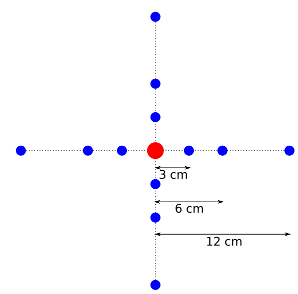

Fig. 5. Layout of a test location: central test point (red) surrounded with auxiliary test points (blue).

As shown above, both theoretical and experimental results suggest that RSS fingerprints are particularly sensitive to small sub-wavelength receiver dislocations. The following sections investigate how such dislocations (in form of GT errors) affect fingerprint-based Wi-Fi positioning performance.

IV. Experimental setup

The experiment is a modification of the classical Wi-Fi fingerprinting setup with static point-based evaluation. Data acquisition has been performed in a spacious foyer of an office building (Fig. 4). We defined 12 test locations, in the nodes of a 3×4 grid with 2 m step. To simulate the effect of inconsistent GT, each test location featured one central point and 12 auxiliary points that were 3, 6 and 12 cm away, as shown in Fig. 5. (The distances roughly correspond to $\lambda/4$, $\lambda/2$ and the full wavelength λ of 2.4 GHz Wi-Fi signals.)

In order to distinguish the impact of small-scale fading from the other interfering factors — such as environment dynamics (people movement nearby, opening and closing doors) and shadowing by the experimenter's body — the experiment was specifically designed to minimize their influence. Firstly, data acquisition took place during a weekend (when both the building and surrounding streets were virtually empty) and the training dataset was collected immediately after the testing one. Secondly, the test smartphone was placed on stationary tables rather than in operator's hand; to avoid possible interference from metallic parts, the tables were improvised by empty cardboard boxes. Finally, during the sampling the operator stepped at least 1.5 m away from the device.

In order to accurately establish centimeter-scale GT displacements, GT positioning was performed in two steps. Firstly, in each test location we installed a coordinate table (cardboard box) and positioned it using a laser rangefinder and dedicated targets placed on the walls. Secondly, the smartphone was positioned on the table, using one of the 13 test points marked on the latter's surface. Given the ± 2 mm rangefinder error, the ± 2 mm table-in-testbed alignment error, and the ± 1 mm smartphone-on-table alignment error, the absolute GT error was ± 5 mm per axis, while GT deviations between the two datasets were within 2.8 mm (± 2 mm per axis).

RSS fingerprints were collected by a Motorola Moto G smartphone running Android 4.4 OS. At each test point the device collected 20 Wi-Fi RSS fingerprints. Wi-Fi access points that were not available in some locations, as well as those with highly correlated RSS values (that is, collocated), were excluded from further processing.

Finally, the actual localization — identification of test point by signal fingerprints — was performed by three machine learning algorithms, including k-nearest neighbor (kNN), support vector machine (SVM) and random forest (RF) classifiers. By default, we report RF-based results, as it demonstrated the highest average accuracy. However, for completeness we also include the results of the other classifiers.

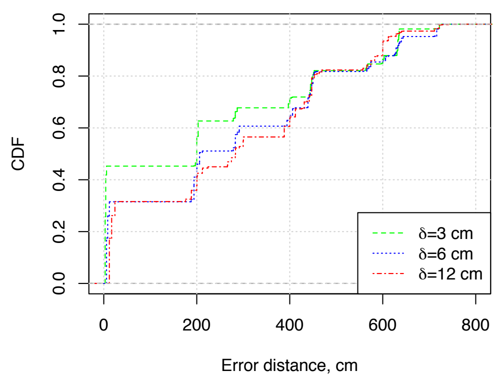

Fig. 6. Leave-one-out positioning performance with different GT errors.

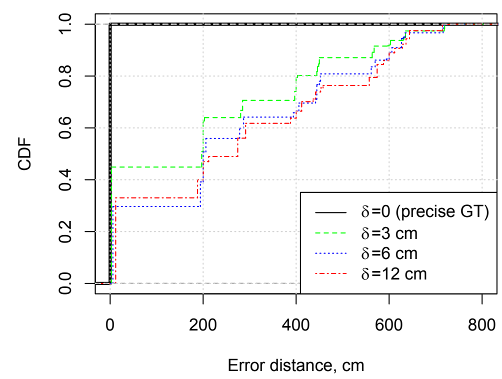

Fig. 7. Two-dataset positioning performance with different GT errors.

TABLE I
LOCALIZATION ACCURACY (LEAVE-ONE-OUT EVALUATION).

GT error	kNN	SVM	RF
3 cm ($\lambda/4$)	42%	41%	46%
6 cm ($\lambda/2$)	33%	32%	31%
12 cm (λ)	33%	28%	32%

TABLE II
LOCALIZATION ACCURACY (INDEPENDENT DATASETS).

GT error	kNN	SVM	RF
Precise GT	93%	98%	100%
3 cm ($\lambda/4$)	46%	44%	45%
6 cm ($\lambda/2$)	32%	30%	29%
12 cm (λ)	30%	29%	33%

V. RESULTS

A. Leave-one-out evaluation

Leave-one-out evaluation is widely employed for estimating positioning performance using a single dataset. There, points are taken one-by-one and used for testing, while the rest of the dataset is used for training the system. According to the main assumption of the fingerprinting approach, fingerprints of nearby points are more similar than those of distant points. Therefore, when presented with a test fingerprint, the classifier should return one of the points nearest to the testing one.

In order to verify this, we divided one dataset into four subsets with increasing GT error δ, so that in each subset each test location was represented by the central point and the four auxiliary points δ cm away from the center. Ideally, when one of the points is taken for testing, its fingerprints should be most similar to the nearest points, and the localization error would thus be limited to δ. The results, however, prove different.

As Fig. 6 shows, with the smallest tested GT error ($\delta = 3\ cm$) only 46% of the fingerprints were correctly recognized as similar to those of nearest points. When the points were moved further apart, the accuracy further decreased and saturated at 31–32% level. This behavior was consistent across all the classifiers (see Table I).

The results are consistent with Section III: with larger GT errors ($\delta \geq \lambda/2$) the localization accuracy is relatively low, but does not depend on δ. When GT error improves to $\delta = \lambda/4$, system's accuracy also improves, confirming that fingerprint similarity becomes related to the spatial proximity of the test points.

B. Two-dataset evaluation

While leave-one-out approach is useful for single-dataset evaluation, it provides a somewhat artificial estimate of the system performance. More practical evaluation scenarios involve dedicated training and testing datasets. In this experiment, we simulate different levels of GT mismatch between the calibration and the testing datasets, when the system is tested on slightly different points than those used for training.

Firstly, we trained the classifier on the 12 central points of one dataset. The other dataset, in turn, was divided into four non-overlapping subsets: an 'ideal' one without any GT error (including only 12 central points), and three imperfect subsets with increasing GT error (each containing only 12×4 auxiliary points; the central points were excluded). The analysis was based on the 19 access points present in both training and testing datasets.

The results are presented in Fig. 7. With exact GT, the system correctly recognized 100% of fingerprints, confirming that our efforts on creating an idealized experimental setup with minimal interference (see Section IV) were successful. However, even the minimal 3 cm ($\lambda/4$) GT deviation reduced the recognition rate to mere 45%. With GT errors of $\lambda/2$ and higher, only a third of fingerprints were recognized to their correct test locations (regardless of the GT error value); the rest of fingerprints were attributed to other test locations meters away. As shown in Table II, the results were consistent across different classifiers.

VI. CONCLUSION AND DISCUSSION

In this paper we investigated how minor GT errors (deviations between calibration and testing points) affect the performance evaluation of indoor positioning systems based on RSS fingerprinting. In contrast to the widespread assumption that GT errors are purely additive and thus negligible, our findings demonstrate that small GT deviations may severely affect the evaluation outcomes.

In particular, our results — supported by the Clarke's theory of small-scale fading [27] and direct experimental verification — show that with smaller GT errors ($\delta < 0.4\lambda$), localization performance directly depends on GT quality and degrades with the increase of GT errors. Beyond the 0.4λ threshold, in turn, the localization accuracy becomes insensitive to the increasing GT error and saturates at a relatively low level (around one-third of the precise-GT baseline).

For Wi-Fi RSS fingerprinting, this means that even centimeter-scale GT imperfections may introduce large and unpredictable localization errors. This phenomenon is surprisingly reminiscent of the butterfly effect from the chaos theory [30], where small initial disturbances lead to major differences in final results. Given the context, the main finding of this study may thus be described as a '*Wi-Fi butterfly effect*'.

While this effect is mainly inherent to single-shot positioning systems (which cannot mitigate small-scale fading), it has wide implications on indoor localization research. In particular, it questions certain findings of previous Wi-Fi localization studies. Due to the Wi-Fi butterfly effect, the experimental results of such studies may have underestimated the actual performance of the evaluated systems — but to a random degree, depending on the used GT methodology and even on the personal performance of the experimenters. Indeed, was system A more accurate than system B, or it was simply evaluated with a more precise GT method? Was cross-device fingerprinting performance more affected by different receiver characteristics or by different antenna placement? Did localization accuracy degrade over time, or the experimenter was tired and held the device a few centimeters lower?

Future studies can avoid such ambiguities by reducing the impact of the butterfly effect. A straightforward solution would be to employ a highly precise GT methodology (for example, laser rangefinders and tripods); this would minimize GT errors and potentially enable the system to demonstrate its upper-bound accuracy. As an alternative and arguably more robust approach, one could intentionally define the test points at least 0.4λ away from the calibration ones. In this case, the system would provide only a conservative lower-bound performance estimate, but it would be immune to minor GT errors.

ACKNOWLEDGMENTS

This work was supported by the National Research Fund of Luxembourg (INDOORS project, grant number C14/IS/8311593).

REFERENCES

[1] Y. Kim, H. Shin, Y. Chon, and H. Cha, "Smartphone-based wi-fi tracking system exploiting the rss peak to overcome the rss variance problem," *Pervasive and Mobile Computing*, vol. 9, no. 3, pp. 406–420, 2013.

[2] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for WLAN location fingerprinting," in *Proceedings of MobiQuitous-2004*. Los Alamitos, CA, USA: IEEE Computer Society, 2004, pp. 14–23.

[3] M. B. Kjærgaard and C. V. Munk, "Hyperbolic location fingerprinting: A calibration-free solution for handling differences in signal strength," in *Proceedings of PerCom-2008*, 2008, pp. 110–116.

[4] A. Popleteev, "HIPS: Human-based indoor positioning system," in *Proceedings of IPIN-2016*. IEEE, 4–7 Oct. 2016.

[5] G. Deak, K. Curran, and J. Condell, "A survey of active and passive indoor localisation systems," *Computer Communications*, vol. 35, no. 16, pp. 1939–1954, 2012.

[6] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, 2002.

[7] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proceedings of INFOCOM-2000*, vol. 2, 26-30 March 2000, pp. 775–784.

[8] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *Proceedings of MobiSys-2005*, 2005, pp. 205–218.

[9] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless LANs," *Computer Networks*, vol. 47, no. 6, pp. 825–845, 2005.

[10] A. Varshavsky, E. de Lara, J. Hightower, A. LaMarca, and V. Otsason, "GSM indoor localization," *Pervasive and Mobile Computing*, vol. 3, no. 6, pp. 698–720, December 2007.

[11] A. Popleteev, "Indoor positioning using FM radio signals," Ph.D. dissertation, University of Trento, 2011.

[12] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proceedings of Mobicom-2012*. ACM, 2012, pp. 293–304.

[13] M. Alzantot and M. Youssef, "Crowdinside: automatic construction of indoor floorplans," in *Proceedings of ACM SIGSPATIAL GIS-2012*. ACM, 2012, pp. 99–108.

[14] R. Faragher and R. Harle, "SmartSLAM – an efficient smartphone indoor positioning system exploiting machine learning and opportunistic sensing," in *Proceedings of ION GNSS-2013*, 2013.

[15] C. Wu, Z. Yang, and Y. Liu, "Smartphones based crowdsourcing for indoor localization," *IEEE Transactions on Mobile Computing*, no. 2, pp. 444–457, 2015.

[16] S. He and S.-H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[17] "Microsoft indoor localization competition," Online: http://research.microsoft.com/en-us/events/msindoorloccompetition2016/default.aspx, 2016, (Collocated with IPSN 2016).

[18] F. Potortì, P. Barsocchi, M. Girolami, J. Torres-Sospedra, and R. Montoliu, "Evaluating indoor localization solutions in large environments through competitive benchmarking: The EvAAL-ETRI competition," in *Proceedings of IPIN-2015*, Oct. 2015, pp. 1–10.

[19] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *Proceedings of IPIN-2014*. IEEE, 2014, pp. 261–270.

[20] N. Moayeri, M. O. Ergin, F. Lemic, V. Handziski, and A. Wolisz, "PerfLoc (Part 1): An extensive data repository for development of smartphone indoor localization apps," in *Proceedings of PIMRC-2016*. IEEE, 2016, pp. 1–7.

[21] T. Van Haute, E. De Poorter, F. Lemic, V. Handziski, N. Wirström, T. Voigt, A. Wolisz, and I. Moerman, "Platform for benchmarking of RF-based indoor localization solutions," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 126–133, 2015.

[22] A. Popleteev, "Poster: Impact of ground truth errors on Wi-Fi localization accuracy," in *Proceedings of MobiSys-2017*, 2017.

[23] M. Youssef and A. Agrawala, "Small-scale compensation for WLAN location determination systems," in *Proceedings of WCNC-2003*, vol. 3. IEEE, 2003, pp. 1974–1978.

[24] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the Mona Lisa: spot localization using PHY layer information," in *Proceedings of MobiSys-2012*. ACM, 2012, pp. 183–196.

[25] S. Lee, S. W. Choi, C. Laoudias, and S. Kim, "Heterogeneous device tracking with RSS variation mitigation over a radio map," *IEEE Wireless Communications Letters*, vol. 5, no. 5, pp. 552–555, 2016.

[26] S. Adler, S. Schmitt, K. Wolter, and M. Kyas, "A survey of experimental evaluation in indoor localization research," in *Proceedings of IPIN-2015*, 2015.

[27] R. Clarke, "A statistical theory of mobile-radio reception," *Bell Labs Technical Journal*, vol. 47, no. 6, pp. 957–1000, 1968.

[28] M. J. Gans, "A power-spectral theory of propagation in the mobile-radio environment," *IEEE Transactions on Vehicular Technology*, vol. 21, no. 1, pp. 27–38, 1972.

[29] G. D. Durgin and T. S. Rappaport, "Theory of multipath shape factors for small-scale fading wireless channels," *IEEE Transactions on Antennas and Propagation*, vol. 48, no. 5, pp. 682–693, 2000.

[30] E. Lorenz, *The Chaos Avant-garde: Memories of the Early Days of Chaos Theory*. World Scientific, 2000, ch. 7. The Butterfly Effect.