A Nation-Wide Wi-Fi RSSI Dataset: Statistical Analysis and Resulting Insights

Germán Capdehourat*†, Federico Larroca† and Gastón Morales†
*Plan Ceibal, Montevideo, Uruguay
†Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay
{gcapde,flarroca,gmorales}@fing.edu.uy

***Abstract*—We present a dataset collected during ten months from a network comprising approximately 9500 double-band Access Points (APs), corresponding to Uruguay's nation-wide one-to-one computing program's internet provider. The dataset includes the transmission power, used channel and measured RSSI (Radio Signal Strength Indicator) that each AP senses every other AP in sight, with a granularity of an hour. This results in a total of more than 750 million measurements, one of the largest Wi-Fi datasets to date.**

In the study of this dataset we have first focused on a link-level analysis. Our contributions are fourfold. We verify that approximately only half of the RSSI time-series are actually stationary, and that in that case, they present strong time correlations. Moreover, the typical assumption that the channel is symmetrical is not true, even in the long-term, and we show that interference plays an important role on this asymmetry. Finally, we study attenuation in the 5 GHZ band and show that its upper section is prone to larger attenuation than what is predicted by classic models. The practical consequences of these observations are discussed throughout the article. We also present network-level indicators of the system (such as number of neighbors per AP and interference level). These are particularly useful for simulating a planned network such as the one discussed here.

***Index Terms*—attenuation, indoor propagation, hypothesis testing**

I. Introduction

Understanding and characterizing the received power of Wi-Fi devices is a crucial input to several problems. Indoor positioning [1], MAC access mechanisms [2], [3], power control [4], radio planning [5] and performance evaluation [6], [7] are only examples where certain assumptions are made on the attenuation (and thus received power) suffered by the Wi-Fi signal. However, a true understanding of the propagation mechanisms and the resulting received power is still difficult to attain, particularly in indoor scenarios. This results in sometimes contradictory assumptions being made on the literature, for instance regarding the symmetry of the channel, the distribution of fading, or the dependence of path-loss on distance or frequency (see for instance [8]).

We present a step towards overcoming this difficulty: a dataset including almost 9500 operative Wi-Fi Access Points (APs) in more than 1400 educational centers distributed across an entire country. During the data collection phase, we have gathered RSSI (Radio Signal Strength Indicator) measurements every hour between all the APs (independently of the channel in use) during ten months. This means that we have 24 measurements per day for each pair of APs visible to each other on each band (as all APs are double-band), during the complete school year, resulting in over 750 million measurements. In addition to RSSI measurements, the dataset includes the associated AP model, configured frequency channel and transmission power on both bands.

We are still working on the analysis of this huge dataset, and this is an ongoing effort, but we share here four insights that we believe will be highly interesting to the community. Firstly, even when APs use a fixed channel and transmission power, the measured RSSI is not stationary in approximately half of the studied RSSI time-series. Moreover, even if stationary, most of these time-series present strong time correlations for several hours, and very few of them are actually independent random variables. This is an important observation with practical consequences (in particular) for positioning systems, as we discuss later.

Thirdly, we study the symmetry of the links and show that it does not hold. That is to say, the mean RSSI (considering only stationary sequences) measured at certain AP of the transmission of another one is not necessarily equal on the reverse direction (even when both APs are configured to use the same frequency and transmission power). This observation is particularly important for power control algorithms. We further study the stationarity of both directions and use it to show that this asymmetry is mostly due to interference.

The fourth observation is that the upper part of the 5 GHz band presents a stronger attenuation than the rest of the band (stronger than what the Friis model predicts). We show evidence that this is mostly due to the walls between the APs. This is a relevant conclusion that has significant importance for radio planning. Even though the larger attenuation can cause problems, cell isolation is also relevant, in particular for scenarios with a high density of APs. Thus, in some situations the wall attenuation could become helpful, in order to avoid a high co-channel interference (CCI).

Finally, we present important network-level indicators of the complete system. For instance, the number of neighbors per AP and the level of CCI they suffer. These are basic parameters useful for simulating a planned network such as the one discussed here.

This work was partially supported by ANII (grant FMV_3_2018_1_148149) and was approved by Plan Ceibal's ethical and data privacy committee. ISBN 978-3-903176-28-7

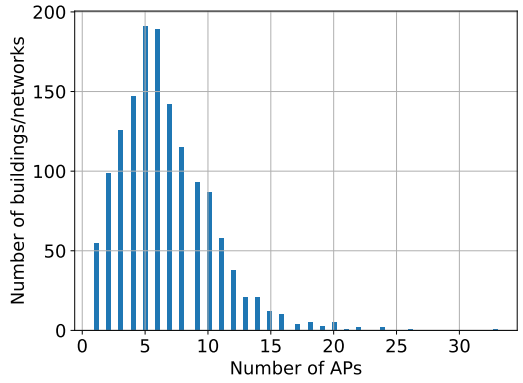

Fig. 1. Number of buildings/networks with a given number of APs

After presenting the details of the dataset on the following section, we discuss the related work on Sec. III. Section IV presents our link-level insights, whereas in Sec. V we present the network-level analysis. Conclusions and future work are discussed in Sec. VI.

II. Details of the Dataset

The dataset was obtained from Plan Ceibal [9], a major education service provider, which runs Uruguay's nation-wide one-to-one computing program. Among several challenges, one of its most relevant responsibilities is to provide connectivity to all educational centers throughout the country, mostly primary and secondary schools, and thus corresponding to indoor scenarios. This makes it one of the nation's largest Internet provider, reaching a number of devices comparable to the number of subscribers of mobile network operators.

It is important to highlight that there is an enormous variety of buildings, ranging from centennial constructions with several stories and hundreds of students to small rural schools with just a few tens. This may be seen in Fig. 1, which shows the histogram of the number of APs per building (i.e. network): each building is typically covered by 5 APs, although approximately 20% of the buildings required more than 10 APs. Moreover, there are no special regulatory aspects regarding interferences from external networks (e.g. from other APs) into these educational buildings.

For some years now, Plan Ceibal has incorporated high-end Wi-Fi solutions, which allows a relatively complete and continuous monitoring of the state of the network, which we leverage in this study. In particular, most of the connectivity solution is administered by two Cisco Flex 7500 Wireless LAN Controllers (WLCs), each of them supporting up to 6000 APs. The complete list of AP models and their most important parameters are included in Table I. All APs use 20 MHz channels in 2.4 GHz and 40 MHz in the 5 GHz band (and only non-overlapping channels on both bands).

As part of the Radio Resource Management (RRM) algorithm executed by the WLCs [10], each AP in the network periodically sends a so-called NDP (Neighbor Discovery Protocol) packet on every channel and band possible. These

TABLE I
AP models and how many are present in the complete network

AP model	Number of APs	Standard (2.4 GHz / 5 GHz)
AIR-CAP2702I-A-K9	5098	802.11n / 802.11ac-Wave 1
AIR-CAP1702I-A-K9	2681	802.11n / 802.11ac-Wave 1
AIR-AP1832I-A-K9	862	802.11n / 802.11ac-Wave 2
AIR-CAP2602I-A-K9	759	802.11n / 802.11n

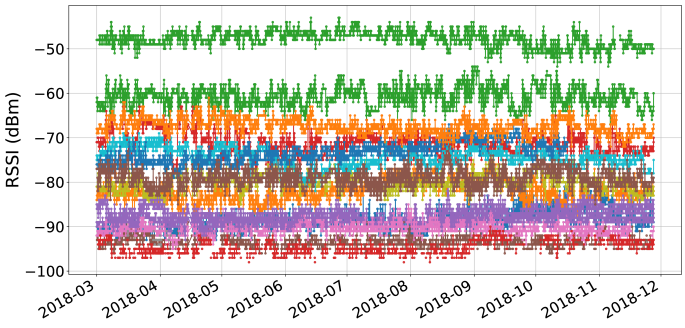

Fig. 2. A typical sequence of RSSI measurements showing how one AP is seen from all its neighbours in the 2.4 GHz.

broadcast messages are sent at the maximum allowed power for the channel/band, at the lowest supported data rate and using a single radio chain (meaning no beamforming is applied in their transmission). By default, an NDP packet is sent over all channels every 180 seconds (i.e. an AP goes off channel roughly every 16 seconds to send an NDP packet over each of the 11 channels in the 2.4 GHz band, and every 8 seconds for the 22 channels in the 5 GHz band). All received NDP packets, along with the corresponding RSSI (expressed in dBm and with a resolution of 1 dBm) and channel, are forwarded to the WLC. These values are averaged by the WLC over a period of 15 minutes (the so-called pruning interval), corresponding to 5 measurements per neighbor.

We have setup a system where every hour we query (via SNMP) the WLCs about the last measurement corresponding to all the APs (and thus the averages do not overlap). In particular, we are interested in the RSSI each AP *sees* every other AP of the network, for both frequency bands. The timescale was chosen in order to minimize the effect on the operational network. A typical sequence of RSSI measurements is shown in Fig. 2, where we have all the values corresponding to how certain AP is seen from all its neighbours in the 2.4 GHz band. The period is restricted to the school year (from March to December) and missing data (in the example, note the small gap in mid-April) is mostly due to holidays (when the equipment might be turned off at schools), although since this data is not stored by neither the APs nor the WLC, any problem in the connection or the request between our system and the WLC will result in a missing measurement.

We have also gathered from the WLCs the configured transmission (Tx) power and frequency channel of all the APs for both bands. The Tx power is actually reported by means of the so-called power level, a Cisco terminology for this parameter. With this value, as well as the frequency channel

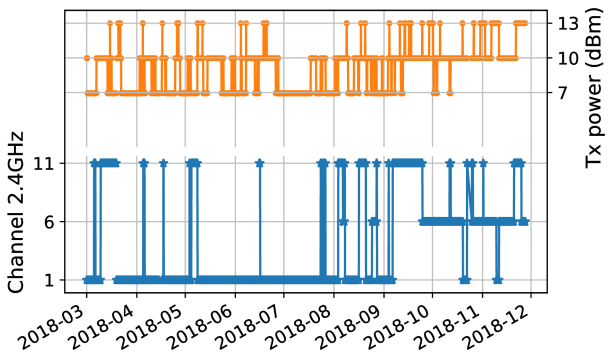

Fig. 3. The configured transmission power and channel of the same AP as in Fig. 2.

and the AP model (also obtained from the WLCs) it is possible to calculate the Tx power of each AP in dBm. As indicated in the Cisco documentation [11] a power level of 1 corresponds to the maximum power setting for the AP (which depends on the frequency, since the regulation is different in each band). Each subsequent power level (up to 8) represents a 3 dBm reduction in transmit power. Note that, since they stem from NDP packets, our RSSI measurements correspond to a power level of 1 on the receiver's channel. This means, for instance, 22 dBm in the 2.4 GHz band for all models, or 14 dBm for the 2602 model in channels 36 to 48 in the 5 GHz band.

An example sequence corresponding to the same AP as in Fig. 2 is shown in Fig. 3. All this data is timestamped with the time at which the whole "round" of SNMP queries started and saved into a database. Note that although the timestamp is the same, the actual moment the measurements were taken to compute the resulting average for every AP will not be exactly the same, but the maximum difference never exceeds 180s.

III. Model and Related Work

Let us present the notation we will use throughout the paper. Let i and j be a pair of APs.[1] We denote the RSSI on node i from the transmission of j at timestamp t by $RSSI_t(i,j)$. Moreover, we will use $f_t(i)$ to indicate the frequency/channel used by node i at t. Finally, let $P_t(j)$ be the transmission power used by j on timestamp t (in our case the maximum allowed power on $f_t(i)$, the receiver's channel).

Mathematical models for RSSI values are typically expressed as:

$$RSSI_t(i,j) = P_t(j) - L(i,j) + \zeta_t(i,j), \qquad (1)$$

where $L(i,j)$ is the path-loss, which is deterministic and depends on the propagation path between i and j (such as distance between them, or the number of walls and their material), and $\zeta_t(i,j)$ is a zero-mean random variable called fading (or shadowing), which models random variations around this path-loss.

As we discussed in the previous section, in our particular case $RSSI_t(i,j)$ is the result of averaging the RSSI of 5 NDP packets sent by j and received at i on frequency $f_t(i)$. In such case the random fluctuations around the mean are called *long-term fading* as they are produced at a time-scale several orders of magnitude larger than the symbol-time (it is thus not possible to study *short-term fading* with our dataset).

In the past years there has been several measurement studies aiming at understanding the behaviour of both $L(i,j)$ and $\zeta_t(i,j)$ in different scenarios. Naturally, the first such studies were carried out during the deployment of the first mobile telecommunication systems [12], and involved mostly outdoor propagation (or outdoor to indoor). With the advent of Wi-Fi and its immense popularity, indoor propagation gained interest.

For instance, [13] concludes that $\zeta_t(i,j)$ may be modeled by a Weibull or Nakagami random variable, although the Rice distribution acts as a reasonable approximation. However, the transmitted signal is narrowband (corresponding to a DECT transmitter). Pertaining specifically to Wi-Fi, both [14], [15] conclude that the log-distance path-loss (where the attenuation depends on the logarithm of the distance between i and j) and log-normal shadowing (i.e. $\zeta_t(i,j)$ follow a gaussian distribution) are reasonable models for an indoor propagation under 802.11b, whereas [16] concludes that Rician distribution is more accurate when we have Line-of-Sight (LOS) with the AP. On the other hand, the authors of [17] conclude that the best model for indoor attenuation is actually the multi-wall-floor model [18], and that fading is best characterized by a log-normal distribution.

All of these works focus on the 2.4 GHz band. With several more channels available, the 5 GHz band is increasingly being used despite larger attenuations, and virtually all new terminals are dual-band. One of the first reports regarding this band was [19], although the authors used narrowband signals and focused on outdoor propagation. Indoor propagation in this band was studied in [20] for corridor-like environments, and concluded that the attenuation was very similar to lower frequencies (as low as 900 MHz) except for the theoretical difference in free space loss. However, for instance [21] report that different materials present significant differences in attenuation depending on the frequency.

Wi-Fi indoor localization systems use the received RSSI as an input to estimate the position of a device [1], and thus several studies are from this community [22], [23]. For instance, [24] discusses how the distribution of fading may affect the positioning's system precision. A very interesting discussion may be found in [25] regarding the characterization of the RSSI and its usefulness for positioning. In particular, the authors conclude that the best distribution for $\zeta_t(i,j)$ is Weibull, and discuss the presence of an RSSI offset, i.e. a fixed difference between the measured RSSI and the actual signal's power (also observed in [26], [27] when characterizing symmetry).

It is important to highlight that all of these works focus on single or very few links. That is to say, a small number of works study complete networks. Measurements over operational networks is typically referred to as "in the wild" [28]. For instance, the authors of [29] study a large

[1] In the sequel, we will use the terms *AP* and *node* indistinctively.

network comprising ten thousand APs during two weeks. However, their focus was on the client's usage and not on the wireless channel. Similarly, the authors of [30] study the connectivity pattern of users during a month comprising over eight million APs. Another example is [31], which studies some 50 billion packets across a hundred thousand APs and proposes variations to TCP and the channel assignment algorithm in APs.

With several months' worth of RSSI measurements, including over 1400 networks comprising about 9500 APs in total, and resulting in approximately 70.000 and 37.000 "links" in the 2.4 GHz and 5 GHz band respectively, the dataset collected for the present article is to the best of our knowledge the largest RSSI dataset to date (even considering those included in the popular Crawdad repository, https://crawdad.org/). In the following sections we present some of the conclusions we have reached during our ongoing study of this dataset. In particular, we will focus on single-link analysis, but we will also discuss some important conclusions we have reached regarding complete networks.

IV. Link-Level Measurements

A. Stationarity

We will begin our discussion by analyzing an often overlooked but important aspect of the RSSI measurements: its stationarity. For example, this property is assumed by virtually all models used in performance evaluation (e.g. if in (1) we assume that $L(i,j)$ is constant). Furthermore, a certain level of stationarity is assumed in all Wi-Fi indoor positioning systems. For instance, so-called Wi-Fi fingerprinting-based indoor positioning techniques [1] use a set of RSSI measurements together with the corresponding zone (e.g. room) to train a machine learning algorithm that learns to map between RSSI measurements and zone. In this case, a change in the underlying distribution of RSSI measurements (or different from the one used during training) will result in a very poor performance of the positioning system.

However, such situation is relatively common. For instance, Figure 4 shows a one week long example from our measurements where stationarity does not hold. Approximately the first day presents a mean clearly higher than the rest of the week. Please recall that APs have a static position, and thus these changes may stem either from the surroundings (such as new large objects between APs or interferers near the receiver using the same channel) or from changes in the equipment itself (such as in the antenna orientation).[2] This last case, although possible since APs are not under our control, should be rare, as APs are not easily accessible. The former case is important, since a change in the surroundings will affect not only the transmission between this particular pair of APs, but all nodes in the vicinity.

We have thus studied how often this non-stationarity manifests, and for how long may the propagation environment be considered stationary. Some important remarks regarding how we proceeded follows. First, as the channel used by the AP is dynamically set by the WLC, and to avoid fluctuations due to large variations in the used frequency (see Sec. IV-C for further discussion on this aspect), we have analyzed $RSSI_t(i,j)$ for those t where $f_t(i)$ is static. That is to say, if $f_t(i)$ changes n times during the complete series, then we will study the $n+1$ resulting sub-series. The same procedure is applied when the series has missing values for certain timestamps t (cf. the small gap in mid-April in Fig. 2). However, we will only consider those sub-series longer than 100 measurements (roughly four days), to make the result statistically meaningful.

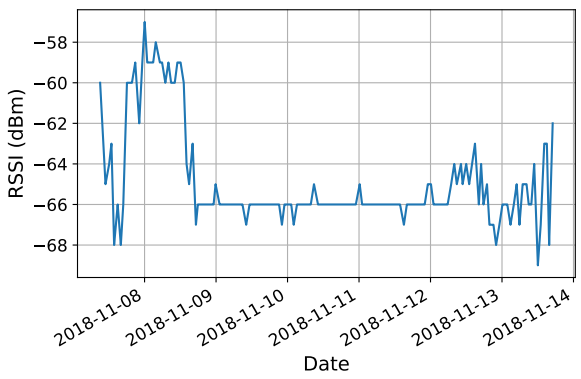

Fig. 4. An example of a non-stationary RSSI measurement series corresponding to a fixed channel in the 2.4 GHz band.

In order to analyze the stationarity of a given time-series, and since visual inspection is out of the question in our case, we have resorted to classic unit-root tests such as a combination of the augmented Dickey-Fuller (ADF) [32] and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests [33] (both implemented in the `statsmodel` library [34]). Only if both tests do not reject the stationarity hypothesis at a significance level of 0.05 we will conclude that the sub-series is stationary.

Results are shown in Table II, where we present the portion of sub-series that were actually stationary (for both frequency bands), as a function of their length in weeks (to be more precise, its length in hours integer divided by 168). A first result worth highlighting is that only half of the considered sequences are stationary. Moreover, note how the behaviour is approximately the same for all lengths and both bands.

TABLE II
Portion of RSSI sub-series that are stationary

Length (weeks)	**2.4 GHz Band**	**5 GHz Band**
0	49%	49%
1	55%	54%
2	56%	50%
3	52%	51%
4	51%	46%

As we mentioned before, this result is of major importance

[2] As discussed in Sec. II, the NDP packets are always sent at maximum power.

for positioning systems. It means that if the measurement phase of a Wi-Fi indoor positioning system is performed, for instance, during a single day, then the resulting precision may significantly decrease some days later. However, since the duration of the sub-series has little impact on the portion of stationary sub-series, a change on the underlying distribution may take some weeks. All in all, this result is an important argument in favor of positioning systems that include maintenance (i.e. updates on the mapping between RSSI measurements and location), such as [35].

B. Channel Symmetry

Let us now focus on one of the assumptions most commonly used when modeling and designing wireless systems: the symmetry of the channel. For instance, the classic CSMA algorithm assumes that if a node sees a frame from another node, its own transmission with the same power will interfere with the ongoing transmission in the same degree. More recently, most power-control algorithms on Wi-Fi make similar assumptions (even those discussed for 802.11ax [36]): that a device may calculate its transmission power based on the AP's informed power and the received RSSI.

However, an offset between the received power and the reported RSSI is generally present. This is discussed for instance in [25], where this offset is assumed fixed and due to miscalibration. Our measurements indicate that this is not always the case, and that interference plays an important role. In any case, as we will shortly present, this offset may exceed 3 dBm. The immediate practical consequence is that power control algorithms should always favor a closed-loop scheme (e.g. one where the AP indicates the stations either to increase or decrease their transmission power) instead of an open-loop one (where stations base their decision on local RSSI measurements). This is specially important for the new UL MU OFDMA access scheme in 802.11ax [36], where the AP should receive signals from different devices at almost the same power level.

To study and quantify this offset we will take all AP pairs i, j that see each other, are the same model (to avoid differences in transmission powers), transmit on the same channel, and study the difference in RSSI between the direct and reverse paths (i.e. $\Delta RSSI_t(i,j) = RSSI_t(i,j) - RSSI_t(j,i)$). Similar to the previous section, the complete time series is broken into sub-series when missing values are present (which in this case are further generated when one of the two nodes changes its channel), and we will only consider sub-series longer than 100 samples.

Note that if the difference between the direct and reverse paths stems from a fixed offset only, then the sub-series $\Delta RSSI_t(i,j)$ would be stationary, and the corresponding mean equal to this offset. However, only 60% of the sub-series on the 5 GHz band are actually stationary according to the ACF+KPSS tests (again, at significance level equal to 0.05). Results are even lower for the 2.4 GHz band, where only half of the sub-series are stationary. In order to study what may be causing these non-stationarities, we have also tested the corresponding direct and reverse sub-series, and obtained the results in Table III.

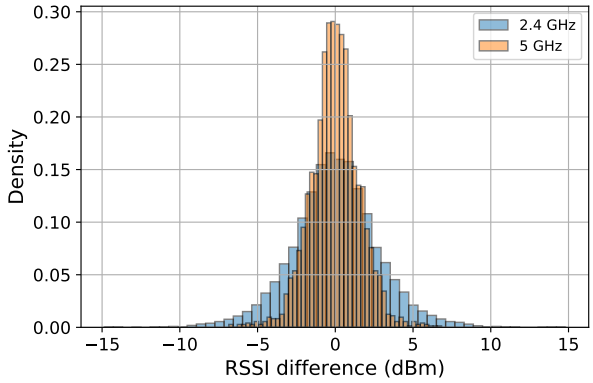

Fig. 5. Histogram of average $\Delta RSSI_t(i,j)$.

TABLE III
SYMMETRY STATIONARITY

		Reverse/Direct path		
	$\Delta RSSI_t(i,j)$	Both Stationary	Both Non-stationary	Mixed
5 GHz Band	Stationary	29%	13%	17%
	Non-stationary	7%	13%	21%
2.4 GHz Band	Stationary	21%	12%	19%
	Non-stationary	5%	18%	26%

Note that most of the non-stationary cases occur when one of the paths is stationary and the other one is not (e.g. 21% of the total sub-series for the 5 GHz band). A change in the distribution of the RSSI seen by only one of the two APs is indicative of a local interference (as opposed to a change in the propagation path between both APs). Note that an important portion of the stationary sub-series also stem from reverse and direct paths with different behaviours, although the resulting $\Delta RSSI_t(i,j)$ is still stationary (17% in the 5 GHz band and 19% in the 2.4 GHz band). This could be generated by an interferer closer to one of the two ends (resulting in one of them only marginally affected), so that the tests cannot reject stationarity for the difference $\Delta RSSI_t(i,j)$. All in all, interference appears as a very important factor when considering channel asymmetry.

Regarding the actual difference between direct and reverse paths, Fig. 5 shows the histogram of the mean of $\Delta RSSI_t(i,j)$ (for those considered as stationary) in both bands. Note how the 2.4 GHz band is prone to much more asymmetry than the 5 GHz one. This is in line with the previous discussion, since the 2.4 GHz band is subject to much more interference (either from other Wi-Fi equipment or the myriad of technologies that use this band). Note however, that even in the 5 GHz band, differences of more than 3 dB are not rare and amount to approximately 10%.

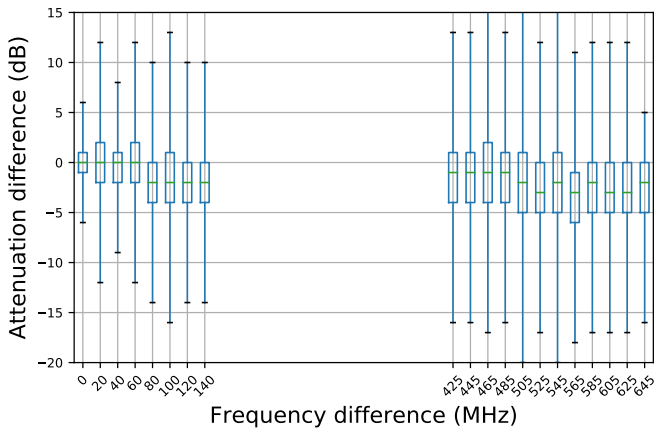

Fig. 6. The difference $L_t(i,j) - L_t(j,i)$. Each boxplot corresponds to a group of measurements with the frequency difference indicated on the x-axis.

C. Dependence on the Frequency

Given that our dataset includes the RSSI received by a certain AP from all other APs in its operating channel, we may study the "instantaneous" attenuation between pairs of nodes using different frequencies (i.e. $P_t(j) - RSSI_t(i,j)$ in Eq. 1, which we will now denote as $L_t(i,j)$).[3] That is to say, for a given timestamp, we can compute the attenuation between a pair of nodes, generally on both directions. We will now study the difference between these two attenuations as a function of the difference in frequency of the channel used by either nodes.

We thus now limit ourselves to those pairs that are the same model and study how does $L_t(j,i) - L_t(i,j)$ depends on $f_t(i) - f_t(j)$. Figure 6 shows the boxplots of these differences for the 5 GHz band. We have grouped the data according to $f_t(i) - f_t(j)$, which is used as the x-axis, and we are only showing its positive half as the data is completely symmetrical. Moreover, we are omitting the outliers for ease of visualization.

Some observations are in order. Firstly, there exists a gap in the frequency differences as Uruguay's regulations only permits operation on the 5.2, 5.3 and 5.8 GHz sub-bands (UNII-1, 2 and 3 according to the FCC terminology). This means that the right-most group of boxplots corresponds to values of $L_t(j,i) - L_t(i,j)$ where $f_t(i)$ and $f_t(j)$ belong to the upper and lower portion of the 5 GHz band respectively. Secondly, and most interestingly, there is a clear tendency to higher attenuations with higher frequencies. Note, for instance, how in the UNII-3 channels attenuation is between 0 and 5 dB larger than on the lowest frequencies.

Typical attenuation models (in dB) are linear on the logarithm of the frequency. We have conducted a least square fit on the curve in Figure 6, resulting in a coefficient equal to 5.6, somewhat larger than the typical outdoor coefficient between 2 (free space) and 4.

[3]Please recall that $P_t(j)$ depends on the channel used by i (i.e. $f_t(i)$) and the model of j.

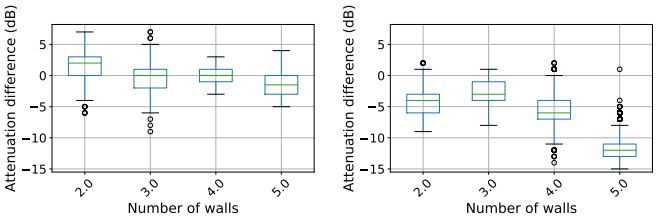

Fig. 7. $\Delta L_t(i,j)$ as the number of walls between APs i and j vary. Measurements correspond to pair of APs where $0 \leq f_t(i) - f_t(j) < 150$ MHz (left) and $f_t(i) - f_t(j) > 400$ MHz (right).

The explanation for this difference lies on the materials and structure between the APs, a possibility further supported by the great variance in each group in Fig. 6. To further justify this claim, we have obtained the plan (and AP's positions) of a school building, and studied how $\Delta L_t(i,j)$ varies with the number of walls between i and j. The results are shown in Fig. 7 where we have separated the measurements depending on $f_t(i) - f_t(j)$: the left boxplots show the results corresponding to a small difference in frequency, and the right ones to a large difference.

Please note how the group with $0 \leq f_t(i) - f_t(j) < 150$ MHz presents a behaviour very similar to the case where $f_t(i) = f_t(j)$, with $\Delta L_t(i,j)$ rarely exceeding ± 5 dB independently of the number of walls between i and j. The trend is very different for the case where $f_t(i) - f_t(j) > 400$ MHz, and clearly evidences that the 5.8 GHz sub-band presents higher attenuations than the rest of the band, and that an important factor is the presence of walls (supporting the study in [21]).

The most important practical consequence of this observation is that this sub-band may be used to decrease interference between APs in large density deployments. This may be understood as a second step in a relatively common practice for this kind of deployments: disabling the 2.4 GHz radios (or enable it only in some APs) to obtain a greater "cell" isolation.

D. Independence

In this last section regarding link-level measurements, we will focus on a second hypothesis that is assumed by all studies regarding $RSSI_t(i,j)$: that they correspond to a sequence of independent and identically distributed (iid) random variables. We have already shown that roughly half of the sequences are not identically distributed, so let us focus on the other half and study whether they are independent. To this end, we will analyze the stationary sub-series as in Sec. IV-A and apply them the Ljung-Box test (with a 0.05 significance level). Quite interestingly, only 2% of the stationary time-series form an iid sequence according to this test.

To gain understanding into the statistical correlation between timestamps, we show two example sequences in Fig. 8 along with their autocorrelation function (ACF). Please recall that each value of $RSSI_t(i,j)$ corresponds to the average of five NDP packets (over approximately 15 minutes), but we are extracting a measurement every hour, so these averages do

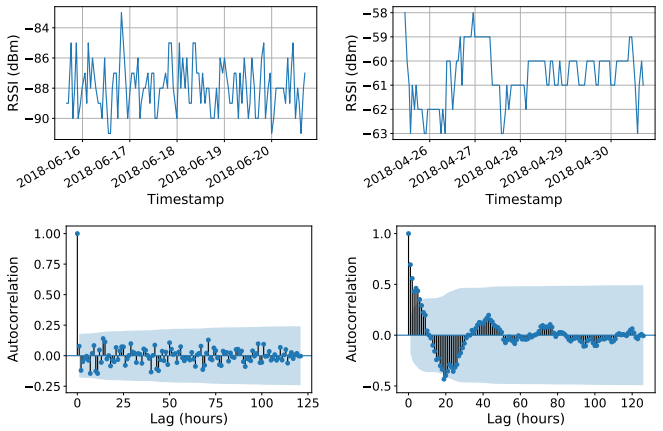

Fig. 8. Example of two $RSSI_t(i,j)$ sequences and their ACF: an iid (left) and a non-iid (right) sequence.

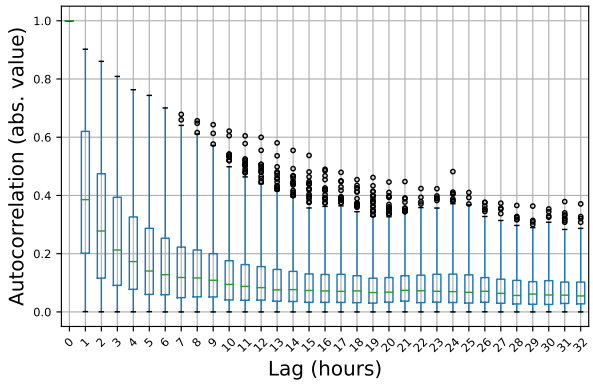

Fig. 9. Boxplot of the absolute value of the ACF of $RSSI_t(i,j)$ for up to 32 hours (corresponding to the complete school year on a certain school).

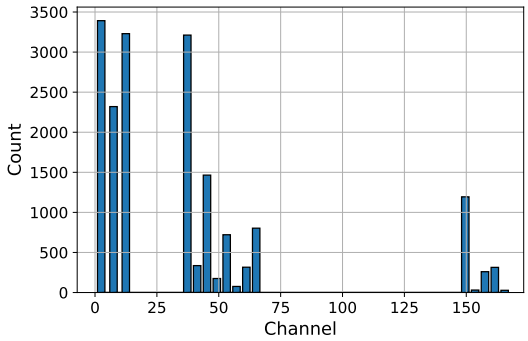

Fig. 10. Number of APs that use a certain channel. Both bands are shown.

not overlap (and thus it does not explain the strong correlation observed in Fig. 8).

From a practical point of view, the consequence of this correlation is that in order to obtain sufficiently varied data for a Wi-Fi localization system, measurements should span several hours. This may be understood as a complementary result to the one obtained in Sec. IV-A: even when the RSSI measurements are stationary, if measurements are not taken sufficiently separated in time the dataset may not reflect the actual variability in the data.

This is further supported by Fig. 9, which shows the boxplot of the absolute value of the ACF (for up to 36 hours lags) of all the sub-series corresponding to a certain school on the 5 GHz band (very similar results are obtained for the 2.4 GHz band). Note how the median autocorrelation reaches a plateau only after approximately the 10 hours lag. Furthermore, note that the 24 hours lag (i.e. the correlation at the same hour during different days) does increase, but very little. This means that it is not so important to take measurements at different times of the day, as it is to consider a long measurement period. In other words, this result indicates that the mobility of people (in this case teachers and students) does not have much influence on the attenuation variations.

V. Network-Level Measurements

We will conclude the article by presenting measurements regarding the networks. In our case, they correspond to professionally designed networks, so the topological characteristics we discuss here should generalize to other similar environments (e.g. indoor scenarios such as hospitals or offices). Moreover, as it is operated by a WLC, it may be considered a "best case-scenario" regarding co-channel interference (CCI), although the RRM algorithm optimality should be analyzed, which is part of our ongoing work. Finally, throughout this section we have considered a snapshot of the networks: in particular all measurements taken at 3 P.M. on August 16th 2018. We have considered several other dates and times (as well as summarizing certain indicators), but the observations we present here do not change qualitatively.

Let us begin by analyzing how are channels used. Figure 10 shows the results. Note how the RRM algorithm evenly distributes the channel usage among the three possibilities in the 2.4 GHz band, although in the 5 GHz band it tends to choose channel 36. More on this aspect is discussed later in this section. Regarding power levels, results are shown in Fig. 11. As expected from the previous analyses, the power used in the 2.4 GHz band is far less than in the 5 GHz (recall that a power level of 1 is the maximum, and each increase corresponds to halving the transmission power).

Both the used channel and power dictate the coverage and interference in the network. An indicator of the latter may be obtained by analyzing how many APs are sensed (and at what RSSI) by each AP (i.e. the number of neighbors of each AP). These results for both bands are shown in Fig. 12. We have divided the neighbors into those that use the same channel as the AP (and thus interfere) and all APs sensed by each AP. Note how the former may exceed five interfering neighbors in the 2.4 GHz band. This is the result of a combination of both the few channels available and the lower attenuation. This is reflected in the 5 GHz band, where the number of interfering nodes are virtually always below 3 APs.

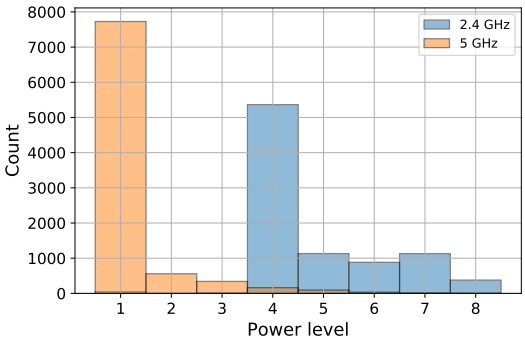

Fig. 11. Number of APs that use a certain power level (recall that a power level of 1 is equivalent to the maximum power). Both bands are shown.

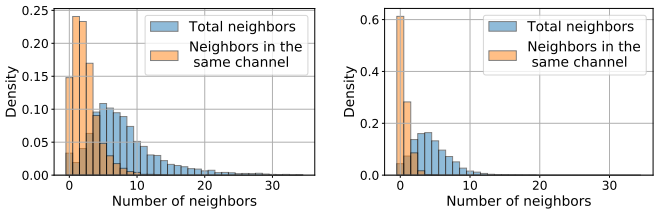

Fig. 12. Density of number of neighbors in each AP for the 2.4 GHz band (left) and 5 GHz band (right).

The question remains on how much power of these interferers reaches the AP. To study this, and to gain insight into the operation of the RRM algorithm, we took $RSSI_t(i,j)$, and corrected it based on the power level configured on j in order to calculate what interference to expect on i from j during operation (which we will denote as $P_t^{rx}(i,j)$).[4] Figure 13 shows the results for all the values of $P_t^{rx}(i,j)$ considering all neighbors or only those operating on the same channel as i.

Note how the density corresponding to the interference generated by neighbors on the same channel as i has a smaller mean, which is naturally the objective of the channel selection algorithm: avoiding strong interferers. Ideally, one would expect that, from all the interferers on node i, those operating on the same channel are received with lower power. In particular, CCI should be below the RSSI threshold used

[4]Recall that our measurements correspond to NDP packets, that are sent at maximum power, which is not necessarily equal to $P_t^{rx}(i,j)$.

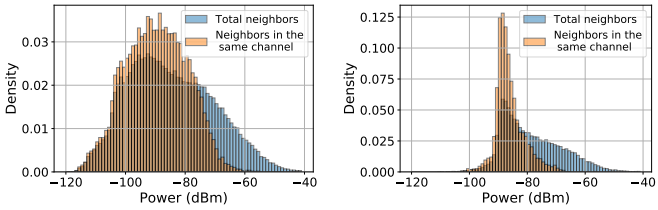

Fig. 13. Density of the power received by each AP from its neighbors ($P_t^{rx}(i,j)$) for the 2.4 GHz band (left) and 5 GHz band (right).

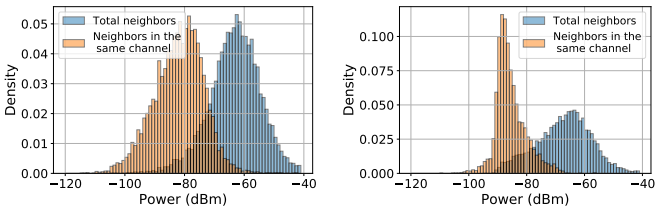

Fig. 14. Density of the maximum power received by each AP from all its neighbors and the maximum of those neighbors that operate on the same channel; for the 2.4 GHz band (left) and 5 GHz band (right).

by the standard CSMA/CA carrier sensing, typically around -80 dBm. We can see that this is mostly achieved for the 5 GHz band in Figure 14, where we are plotting the density of the maximum power received by node i from all its neighbors, along with the density of the maximum power received by i from those neighbors operating on the same channel. However, the situation is quite different for the 2.4 GHz band, where the CCI is above -80 dBm for almost half of the APs.

Figure 14 evidences that the 2.4 GHz band is saturated in its spatial reuse, where even optimally allocating channels and transmission power (according to the RRM algorithm), we cannot avoid a CCI smaller than -80 dBm around 50% of the times. In contrast, the result for the 5 GHz band is complementary to the one we obtained in Fig. 10, which showed that the WLC assigned mostly channel 36, precisely because it generally does not need to use other channels to avoid high interference. Although the key factor to achieve this is the larger number of channels available on 5 GHz, the propagation properties at this band also helps to have more isolation between APs, as we discussed in Sec. IV-C. Finally, it is worth to note that in both cases there is a mean difference of 20 dB between the maximum RSSI for any neighbour and those in the same channel, product of the allocation made by the RRM algorithm. As we mentioned before, part of our ongoing research is studying the RRM algorithm, its resulting performance and optimality.

VI. Conclusions and Future Work

We have presented one of the largest Wi-Fi RSSI dataset of an operational network to date. We have first focused on link-level measurements and have reached several important conclusions. Firstly, approximately half of the RSSI time-series may not be considered as stationary, and even those that are so, present important time correlations for several hours. Moreover, the symmetry of the channel does not hold: average RSSI (considering only APs using the same transmission power and channel) with differences of 3 dBm between both directions are not rare (particularly in the 2.4 GHz band). Our analysis shows that interference plays an important role on this asymmetry. Finally, we have observed a larger attenuation on the UNII-3 sub-band when compared to the UNII-1 sub-band, and have shown that this is due to walls between APs. As we discussed throughout the article, all of these conclusions

are (particularly) important for positioning systems as well as power control algorithms and radio planning.

We have also presented some import network-level indicators. In particular the number of neighbors per AP and the interference level they suffer should be useful for simulating planned and controlled networks such as the ones discussed here. Instead of considering pairs of APs, the next natural step is to enrich this study by considering the graph of APs. That is to say, for each timestamp we may compute a graph indicating in its edges the RSSI (or attenuation). This will allow us, for instance, to verify some assumptions made by the research community on this (random) graphs. Furthermore, we may perform spatial/temporal analysis on these graphs to detect anomalies (such as a change in the building's infrastructure), or to study the RRM (propietary and closed) algorithms and verify their optimality.

Finally, it is our intention to publish the dataset. The process of obtaining the administrative rights to do so is underway, but not finished yet. In particular, we are undertaking the anonymization procedure required.

REFERENCES

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

[2] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part i - carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Transactions on Communications*, vol. 23, no. 12, pp. 1400–1416, December 1975.

[3] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part ii - the hidden terminal problem in carrier sense multiple-access and the busy-tone solution," *IEEE Transactions on Communications*, vol. 23, no. 12, pp. 1417–1433, December 1975.

[4] V. P. Mhatre, K. Papagiannaki, and F. Baccelli, "Interference mitigation through power control in high density 802.11 wlans," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, May 2007, pp. 535–543.

[5] D. D. Coleman and D. A. Westcott, *CWNA: Certified Wireless Network Administrator Official Study Guide Exam CWNA-107*, 5th ed. Alameda, CA, USA: SYBEX Inc., 2018.

[6] F. Baccelli and B. Blaszczyszyn, "Stochastic geometry and wireless networks, volume 1: Theory," *Foundations and Trends in Networking*, vol. 3, no. 3-4, pp. 249–449, 2009.

[7] ——, "Stochastic geometry and wireless networks, volume 2: Applications," *Foundations and Trends in Networking*, vol. 4, no. 1-2, pp. 1–312, 2009.

[8] D. Kotz, C. Newport, R. S. Gray, J. Liu, Y. Yuan, and C. Elliott, "Experimental evaluation of wireless simulation assumptions," in *Proceedings of the 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWiM '04. New York, NY, USA: ACM, 2004, pp. 78–82.

[9] Plan Ceibal, "About Plan Ceibal," https://www.ceibal.edu.uy/en/institucional, 2018, [Online; accessed 23-Dic-2019].

[10] Cisco, "Radio Resource Management White Paper," https://www.cisco.com/c/en/us/td/docs/wireless/controller/technotes/8-3/b_RRM_White_Paper.html, 2016, [Online; accessed 28-Nov-2019].

[11] ——, "Channels and Maximum Power Settings for Cisco Aironet Lightweight Access Points," https://www.cisco.com/c/en/us/td/docs/wireless/access_point/channels/lwapp/reference/guide/lw_chp2/1100_chp.html, 2018, [Online; accessed 9-May-2019].

[12] C. Phillips, D. Sicker, and D. Grunwald, "A survey of wireless path loss prediction and coverage mapping methods," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 255–270, First 2013.

[13] F. Babich and G. Lombardi, "Statistical analysis and characterization of the indoor propagation channel," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 455–464, March 2000.

[14] D. B. Faria *et al.*, "Modeling signal attenuation in ieee 802.11 wireless lans-vol. 1," *Computer Science Department, Stanford University*, vol. 1, 2005.

[15] R. Akl, D. Tummala, and X. Li, "Indoor propagation modeling at 2.4 ghz for ieee 802.11 networks." in *Wireless and Optical Communications*, 2006.

[16] T. A. Wysocki and H.-J. Zepernick, "Characterization of the indoor radio propagation channel at 2.4 ghz," *Journal of telecommunications and information technology*, pp. 84–90, 2000.

[17] T. Chrysikos, G. Georgopoulos, and S. Kotsopoulos, "Wireless channel characterization for a home indoor propagation topology at 2.4 ghz," in *2011 Wireless Telecommunications Symposium (WTS)*, April 2011.

[18] M. Lott and I. Forkel, "A multi-wall-and-floor model for indoor radio propagation," in *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001.*, vol. 1, 2001, pp. 464–468.

[19] G. Durgin, T. S. Rappaport, and Hao Xu, "Measurements and models for radio path loss and penetration loss in and around homes and trees at 5.85 ghz," *IEEE Transactions on Communications*, vol. 46, no. 11, pp. 1484–1496, Nov 1998.

[20] J. Medbo and J. . Berg, "Simple and accurate path loss modeling at 5 ghz in indoor environments with corridors," in *Vehicular Technology Conference Fall 2000. IEEE VTS Fall VTC2000. 52nd Vehicular Technology Conference*, vol. 1, 2000, pp. 30–36.

[21] G. Tesserault, N. Malhouroux, and P. Pajusco, "Determination of material characteristics for optimizing wlan radio," in *2007 European Conference on Wireless Technologies*, Oct 2007, pp. 225–228.

[22] Atreyi Bose and Chuan Heng Foh, "A practical path loss model for indoor wifi positioning enhancement," in *2007 6th International Conference on Information, Communications Signal Processing*, 2007.

[23] X. Zhao, Z. Xiao, A. Markham, N. Trigoni, and Y. Ren, "Does btle measure up against wifi? a comparison of indoor location performance," in *European Wireless 2014; 20th European Wireless Conference*, 2014.

[24] W. Xue, W. Qiu, X. Hua, and K. Yu, "Improved wi-fi rssi measurement for indoor localization," *IEEE Sensors Journal*, vol. 17, no. 7, pp. 2224–2230, April 2017.

[25] A. Zanella, "Best practice in rss measurements and ranging," *IEEE Communications Surveys Tutorials*, vol. 18, no. 4, pp. 2662–2686, Fourthquarter 2016.

[26] D. Kotz, C. Newport, and C. Elliott, "The mistaken axioms of wireless-network research," *Dartmouth College Computer Science Technical Report TR2003-467*, 2003.

[27] K. Srinivasan, P. Dutta, A. Tavakoli, and P. Levis, "An empirical study of low-power wireless," *ACM Trans. Sen. Netw.*, vol. 6, no. 2, Mar. 2010.

[28] F. Larroca and F. Rodríguez, "An overview of wlan performance, some important case-scenarios and their associated models," *Wireless Personal Communications*, vol. 79, no. 1, pp. 131–184, Nov 2014.

[29] S. Biswas, J. Bicket, E. Wong, R. Musaloiu-e, A. Bhartia, and D. Aguayo, "Large-scale measurements of wireless network behavior," in *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4. ACM, 2015, pp. 153–165.

[30] L. Zhang, L. Zhao, Z. Wang, and J. Liu, "Wifi networks in metropolises: From access point and user perspectives," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 42–48, May 2017.

[31] A. Bhartia, B. Chen, F. Wang, D. Pallas, R. Musaloiu-E, T. T.-T. Lai, and H. Ma, "Measurement-based, practical techniques to improve 802.11 ac performance," in *Proceedings of the 2017 internet measurement conference*. ACM, 2017, pp. 205–219.

[32] S. E. Said and D. A. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, vol. 71, no. 3, pp. 599–607, 12 1984.

[33] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of Econometrics*, vol. 54, no. 1, pp. 159 – 178, 1992.

[34] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.

[35] S. Jung and D. Han, "Automated construction and maintenance of wi-fi radio maps for crowdsourcing-based indoor positioning systems," *IEEE Access*, vol. 6, pp. 1764–1777, 2018.

[36] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A Tutorial on IEEE 802.11ax High Efficiency WLANs," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2018.