A CSI Dataset for Wireless Human Sensing on 80 MHz Wi-Fi Channels

Francesca Meneghello, Nicolò Dal Fabbro, Domenico Garlisi, Ilenia Tinnirello, and Michele Rossi

The authors provide a dataset of IEEE 802.11ac channel measurements over an 80 MHz bandwidth channel featuring notable domain diversity, through measurement campaigns that involved thirteen subjects across different environments, days, and with different hardware.

ABSTRACT

In the last years, several machine learning-based techniques have been proposed to monitor human movements from Wi-Fi channel readings. However, the development of domain-adaptive algorithms that robustly work across different environments is still an open problem, whose solution requires large datasets characterized by strong domain diversity, in terms of environments, persons and Wi-Fi hardware. To date, the few public datasets available are mostly obsolete — as obtained via Wi-Fi devices operating on 20 or 40 MHz bands — and contain little or no domain diversity, thus dramatically limiting the advancements in the design of sensing algorithms. The present contribution aims to fill this gap by providing a dataset of IEEE 802.11ac channel measurements over an 80 MHz bandwidth channel featuring notable domain diversity, through measurement campaigns that involved thirteen subjects across different environments, days, and with different hardware. Novel experimental data is provided by blocking the direct path between the transmitter and the monitor, and collecting measurements in a semi-anechoic chamber (no multi-path fading). Overall, the dataset — available on IEEE DataPort [1] — contains more than thirteen hours of channel state information readings (23.6 GB), allowing researchers to test activity/identity recognition and people counting algorithms.

INTRODUCTION

In recent years, spurred by the pervasiveness of Wi-Fi devices, wireless human sensing strategies that rely on movement-induced modifications to the propagation of radio signals have been widely investigated [2]. Being a passive — and thus non-intrusive — approach, Wi-Fi-based human sensing (WHS) has attracted considerable attention for different applications such as, for example, device-free monitoring [3] and activity recognition (AR) [4]. Moreover, by leveraging existing communication infrastructures, WHS grants an easy and low-cost deployment.

In this work, we present a large dataset for WHS, which aims at providing researchers with a means to develop new algorithms and assess their performance on common data, thus allowing for fair comparisons. The availability of large-scale and comprehensive datasets is especially necessary when considering data-driven algorithms exploiting machine/deep learning techniques. These methods are adopted by the vast majority of experimental WHS research, as model-based solutions are of difficult implementation in a wireless context, given the complexity of radio signal propagation. However, public WHS datasets are still very few and often outdated: many of the measurement campaigns that were made available in recent years were performed on 20 or 40 MHz frequency bands, and do not contain sufficient *domain diversity*.

To bridge this gap, the dataset we made publicly available on IEEE DataPort [1] has been collected through IEEE 802.11ac devices operating on 80 MHz bands. Moreover, in contrast with currently available datasets, it contains data featuring notable domain diversity across measurement days, environments, people and hardware. This will allow researchers to assess the performance of their algorithms in completely different situations than those considered at training time, and, in turn, to develop strategies to address the domain adaptation problem in WHS. In fact, although some preliminary WHS approaches were successful, we advocate that some fundamental challenges are still open and, among these, one of the most relevant is the design of robust algorithms, that is, able to generalize and adapt to *new domains*. The dataset contains data for developing AR, person identification (PI) and people counting (PC) applications, providing Wi-Fi channel measurements for up to seven activities, ten people, and seven environments. Data collected in a semi-anechoic chamber, emulating an open-area-test site, are also included in the dataset. Moreover, we provide for the first time data collected when the direct path between transmitter and monitor is completely obstructed. Part of the data has been used to validate SHARP, our recent algorithm for environment- and person-independent Wi-Fi-based AR [4].

BACKGROUND AND EXISTING DATASETS

In September 2020, the IEEE 802.11bf working group was established to empower Wi-Fi devices with sensing capabilities [2]. The goal is to allow Wi-Fi routers to perform the dual role of communication access points (AP) and monitoring devices, leveraging ongoing Wi-Fi traffic as well as ad-hoc packets to deliver sensing services. Note that the new IEEE 802.11bf standard is not expected to

Francesca Meneghello, Nicolò Dal Fabbro, and Michele Rossi are with the University of Padova, Italy; Domenico Garlisi and Ilenia Tinnirello are with the University of Palermo, Italy; and also with CNIT, Italy.

define specific use cases or algorithms for sensing that, in turn, remain open to implementation and require further investigation by the research community [5]. Wi-Fi-based sensing exploits the fact that the presence and the movement of objects in the propagation environment affect the Wi-Fi signal (multi-path) propagation, and these modifications can be estimated via dedicated signal processing on the Wi-Fi channel frequency response (CFR) — also referred to as channel state information (CSI). Wi-Fi systems adopt orthogonal frequency division multiplexing (OFDM) by transmitting over partially overlapping and orthogonal sub-channels, and the CFR is continuously estimated for all of them. Thus, for each pair of transmitting and receiving antennas, the CFR consists of a vector of complex numbers specifying the attenuation and phase shift experienced by the signal over each OFDM sub-channel. Interested readers can find a complete description of the Wi-Fi channel in [4].

Comparison Against Other Public Datasets

Although all commercial Wi-Fi chipsets estimate the CFR, manufacturers do not make this data easily accessible for any other use different from communication. To overcome this, over the years, researchers have designed and implemented three tools to extract the (estimated) CFR. Two of them, namely Linux CSI [6] and Atheros CSI [7], target network interface cards (NICs) implementing IEEE 802.11n (operating over 20 or 40 MHz frequency bands). The third one, Nexmon CSI [8], allows extracting the CFR from specific Wi-Fi chipsets implementing IEEE 802.11ac. Leveraging such tools, several experimental campaigns have been realized in recent years, and have been used by researchers to develop different WHS algorithms. However, releasing the collected data has not been a common practice so far and, in turn, public datasets for WHS are still very few.

In [9, 10], two datasets populated with IEEE 802.11n CFR data captured over thirty OFDM sub-channels have been released. Six people carrying out six activities across a single environment, and ten volunteers, sixteen activities, and three environments are respectively considered in [9, 10]. The work in [11] provides a dataset collected over 114 OFDM sub-channels as a single person performs six activities in three environments. In [12], CFR data was collected using Raspberry-Pi with Nexmon, for 242 OFDM sub-channels (on 80 MHz bands), while a subject performed activities in four rooms. Asus RT-AC86U routers empowered with Nexmon CSI are used in [13], where a single subject performs activities across three environments. In [14], a dataset featuring a single subject moving in three environments has been collected using a Raspberry-Pi and an Asus RT-AC86U router. As for data enabling people counting applications, the dataset presented in [15] provides measurements of up to seven people. However, it only contains measurement via devices working with the IEEE 802.11n standard over a 20 MHz channel. Finally, to the best of our knowledge, the only publicly available dataset for human identification through Wi-Fi was released as part of the gait recognition study in [3]. CFR data was collected from eleven volunteers in two environments over a 20 MHz channel.

In this article, we present a comprehensive dataset providing experimental IEEE 802.11ac CFR data collected across several different environments (space diversity), persons, and transmission hardware (hardware diversity). Our dataset is the first that includes data collected in the same environment on different days (time diversity). This is of high value, as it enables the assessment of the impact that small environmental changes — which are likely to happen in real-world scenarios — have on sensing applications (as detailed below). The dataset has been collected over an 80 MHz frequency band to allow designing algorithms that can be integrated into modern Wi-Fi devices, as those currently deployed in homes and buildings. Moreover, since the sensing resolution increases with the bandwidth [9], our dataset allows developing more advanced sensing algorithms with respect to previous datasets available in the literature. We also remark that, by selecting the CFR values of sub-sets of the collected 256 OFDM sub-channels at 80 MHz, as specified in the IEEE 802.11ac standard, one can still design sensing algorithms working on the 40 MHz and 20 MHz sub-bands used by older Wi-Fi systems. To our knowledge, this is the first dataset providing data for PC and PI from an 80 MHz bandwidth communication link, and the first to provide a high domain diversity for AR. For the first time, we also provide data for measurement campaigns performed when the direct path between the transmitter and the monitor is completely obstructed, and AR measurements in a semi-anechoic chamber where the multi-path fading is strongly reduced, to enable the evaluation of the impact of obstacles in the sensing environment and interference on the performance of WHS algorithms. A detailed comparison among our dataset and existing AR, PI, and PC datasets is provided in Table 1.

Experimental Setup

In the following, we summarize the methodology used to collect data. The complete workflow is in [1] for replicability.

Wi-Fi Network Setup and CFR Data Collection

We considered a Wi-Fi network consisting of a single communication link between two routers, one acting as the AP and the other as a station connected to it. This choice does not imply any loss of generality, as wireless sensing can be performed by the monitor node considering a single channel at a time. In the presence of multiple radio links, the monitor station may scan the available links according to some time-division strategy. We used Wi-Fi routers as they allow for better control of the experimental setup with respect to implementing the `Tx` and the `Rx` with other Wi-Fi-enabled devices (e.g., computers or smartphones). The network was set up using `OpenWrt` to operate on the IEEE 802.11ac channel number 42, with a central frequency of 5,210 MHz and 80 MHz of bandwidth.

Wi-Fi traffic was generated through the `iPerf3` tool setting the packet transmitting rate to 173 packets per second, obtaining a new channel estimate every $T_c \cong 6 \times 10^{-3}$ s, which represents a good channel sampling interval for sensing. We used modulation and coding scheme

dataset	no. days per environment	no. environments	no. involved people	no. concurrent people	no. activities	no. different Tx hardware	obstructed direct path	standard	bandwidth
AR [9]	1	1	6	1	6	1	✗	802.11n	20MHz
AR [10]	1	3	10	1	16	1	✗	802.11n	20MHz
AR [11]	1	3	1	1	7	1	✗	802.11n	40MHz
AR [12]	1	4	1	1	11	1	✗	802.11ac	80MHz
AR [13]	1	3	1	1	4	1	✗	802.11ac	80MHz
AR [14]	1	3	1	1	5–8	2	✗	802.11ac	40–80MHz
PC [15]	1	3	7	1–7	1	1	✗	802.11n	20MHz
PI [3]	1	2	11	1	1	1	✗	802.11n	20MHz
AR-PC-PI [1]	1–5	7 (including a semi-anechoic chamber)	13	1–10	1–7	3	✓	802.11ac	80MHz

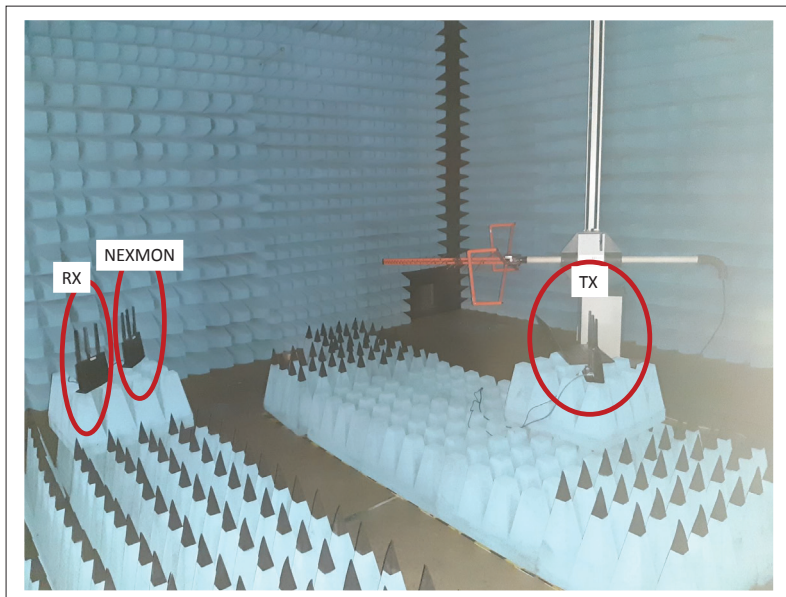TABLE 1. Public datasets for Wi-Fi-based AR, PC and PI.

FIGURE 1. Semi-anechoic chamber. The `Tx`, `Rx` and monitor Asus routers used for data transmission and CFR collection are indicated in the picture.

(MCS) 4 without frame aggregation. A single antenna was purposely enabled on the `Tx` and the `Rx` to enforce the communication over a *single spatial stream*. This conforms to the measurement procedure adopted in all CSI datasets for WHS and leads to better CFR estimates at the monitor, as there is no cross-interference among transmitting and receiving antennas, which would be present in a multi-antenna configuration. It follows that the resulting dataset can be used to train learning algorithms for sensing without having to apply any interference cancellation.

The monitor (M) continuously estimated the CFR of the `Tx-M` link by capturing the packets transmitted over-the-air by the `Tx`. In all the environments but the semi-anechoic chamber, other Wi-Fi networks coexisted with our setup, causing radio interference beyond our control. As IEEE 802.11ac networks leverage the carrier sense multiple access with collision avoidance (CSMA/CA), once interference occurs, no sensing packets are collected. The lack or delay in the acquisition of CFR samples mirrors a real-world scenario and allows evaluating the impact of the phenomenon.

HARDWARE SPECIFICATIONS

Three different setups for the `Tx-Rx` routers were considered. Two Netgear X4S AC2600 IEEE 802.11ac routers were used as `Tx-Rx` in the first setup, whereas two Asus RT-AC86U IEEE 802.11ac routers were deployed for the second. The third setup consisted of a Netgear X4S AC2600 and a TP-Link AD7200 IEEE 802.11ac/ad router. An Asus RT-AC86U router equipped with $N_{ant} = 4$ antennas was used as monitor device in all the setups. The dataset was collected using the Nexmon-CSI extraction tool [8].

MEASUREMENT SETUP

The dataset has been collected by deploying the experimental Wi-Fi network (`Tx-Rx` pair and monitor) in seven different environments, that is, a bedroom, a living room, a kitchen, a university laboratory, a university office, a semi-anechoic chamber, and a meeting room, by also changing the respective positions of `Tx`, `Rx` and monitor devices. We considered two types of direct-path obstructions to enable the evaluation of sensing applications in non-line-of-sight scenarios: a wood bookcase and a concrete block wall.

As an example of the data collection campaign, Fig. 1 shows the experimental setup deployed in the semi-anechoic chamber and consisting of three Asus routers. The walls of the semi-anechoic chamber are fully covered by 40–50 cm long pyramidal radio-absorbing panels which guarantee an absorbing factor of 110 dB at the considered frequencies (1–10 GHz). Movable radio-absorbing panels partially cover the floor while keeping enough uncovered space for users' movement. The semi-anechoic chamber did not contain any objects and no reflectors were used during the experiments.

DATA DESCRIPTION

Our dataset contains more than thirteen hours of CFR collections, resulting in 23.6 GB of data.

In the following, the sequence of CFR vectors associated with an acquisition (one `.mat` file in the dataset) is referred to as *CFR trace*. Each *CFR trace* is saved as a $(N * N_{ant}) \times M$ dimensional complex matrix, where each row is a CFR vector and M is the number of monitored OFDM sub-channels. Each trace contains data collected during 40–300 seconds, resulting in around

set	campaigns	environment	w × l × h [m]	obst.	devices pos.	Tx	Rx	person, Pi	furniture
AR-1	a-b-c-d-e	bedroom	5 × 6 × 4	—	M1-Tx-Rx	Netgear	Netgear	P1	bookcase, 2 beds, desk, chairs
AR-2	a	bedroom	5 × 6 × 4	—	M1-Tx-Rx	Netgear	Netgear	P2	
AR-3	a-b	bedroom	5 × 6 × 4	✓	M2-Tx-Rx	Netgear	Netgear	P1	
AR-4	a	bedroom	5 × 6 × 4	✓	M2-Tx-Rx	Netgear	Netgear	P2	
AR-5	a-b	living room	5 × 6 × 4	—	M3-Tx1-Rx1	Netgear	Netgear	P1	armchair, TV cabinet, table, chairs, 2 sofas
AR-6	a	kitchen	3.5 × 3 × 3.2	—	M3-Tx1-Rx1	Netgear	Netgear	P1	hob, 2 cabinets
AR-7	a	laboratory	7.5 × 3.5 × 2.9	—	M3-Tx1-Rx1	Netgear	Netgear	P3	4 desks, 6 workstations, 6 displays
AR-8	a-b	office	4 × 6 × 3	—	M3-Tx1-Rx1	Asus	Asus	P4	desk, workstation, display
AR-9	a-b-c	semi-anechoic	9 × 7 × 3.4	—	M3-Tx1-Rx1	Asus	Asus	P4	no furniture
PI-1	a	meeting room	7 × 7.5 × 3.5	—	M3-Tx1-Rx1	Netgear	Netgear	P3, P5-P13	7 desks, chairs
PI-2	a	meeting room	7 × 7.5 × 3.5	✓	M3-Tx2-Rx2	Netgear	TP-Link	P3, P5-P13	
PI-3	a	meeting room	7 × 7.5 × 3.5	—	M4-Tx1-Rx1	Netgear	Netgear	P3, P5-P13	
PI-4	a	meeting room	7 × 7.5 × 3.5	✓	M4-Tx2-Rx2	Netgear	TP-Link	P3, P5-P13	
PC-1	a	meeting room	7 × 7.5 × 3.5	—	M3-Tx1-Rx1	Netgear	Netgear	P3, P5-P13	7 desks, chairs
PC-2	a	meeting room	7 × 7.5 × 3.5	✓	M3-Tx2-Rx2	Netgear	TP-Link	P3, P5-P13	
PC-3	a	meeting room	7 × 7.5 × 3.5	—	M4-Tx1-Rx1	Netgear	Netgear	P3, P5-P13	
PC-4	a	meeting room	7 × 7.5 × 3.5	✓	M4-Tx2-Rx2	Netgear	TP-Link	P3, P5-P13	

TABLE 2. Measurement conditions. For each set, we specify the number of campaigns performed, the monitored environment and its dimensions (width × length × height), the presence of an obstacle (obst.) blocking the direct path between the transmitter and the monitor, the position of the devices (`Mj-Txj-Rxj`) and the monitored area (identified by a color as in Fig. 2), the `Tx-Rx` routers brand, the person/people (`Pi`) involved, and the environmental furniture.

N = 6,600–50,000 CFR vectors per monitor antenna per trace. As our dataset is collected considering transmissions on an 80 MHz band, 256 sub-channels can, in principle, be monitored. However, Nexmon only returns the CFR for the M = 242 data sub-channels while no information is provided for the control sub-channels. As part of our previous work, we made available a Python script to further process the data [4].

DATASET DOMAIN DIVERSITY

The campaigns have been performed on different days through several months (April-December, 2020 and January-September, 2022), ensuring considerable time diversity. Thirteen volunteers were involved: ten males, in the following, referred to as `P1`, `P4`, `P6-P13` and three females, indicated with `P2`, `P3`, and `P5`. Four volunteers performing activities within six different environments were considered for AR and two different Wi-Fi network configurations were deployed to include hardware diversity in the data. For what concerns the PI and PC applications, we provide diversity in terms of people involved (ten), and Wi-Fi hardware (four different setups). For all three applications, domain diversity in terms of LOS/NLOS conditions for the transmitter-monitor link is included.

DATASET ORGANIZATION

The dataset is structured into 26 sub-folders, as summarized in Table 2. Each sub-folder is identified by a prefix, a number and a letter. The prefix indicates the target application: AR, PI or PC. The number indicates a specific combination of monitored environment, hardware type and positions, measurement day and involved person/people, as described next. The letter identifies the different measurement campaigns performed for the same setting.

The first eighteen sub-folders contain CFR data intended for AR applications. The data was collected when a single person (`P1-P4`) is present in the monitored space. The name of each file con-
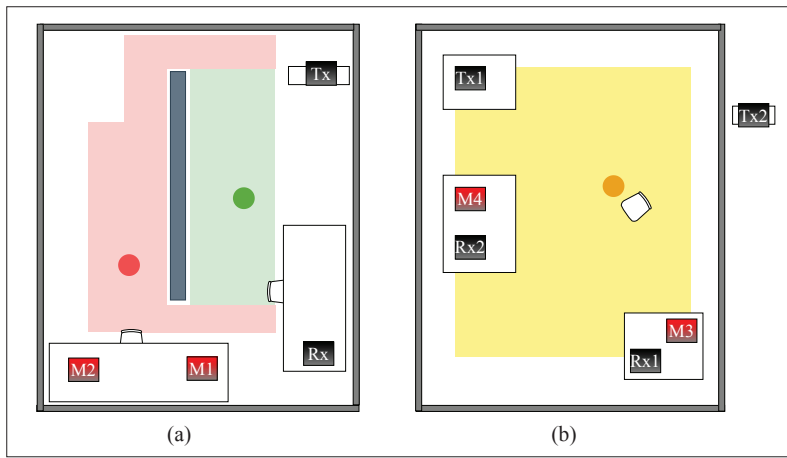
FIGURE 2. Device and user's positions in the monitored environment for: a) bedroom with a wood bookcase in the middle; b) living room, kitchen, laboratory, office, semi-anechoic chamber, and meeting room. `Txj`, `Rxj` and `Mj`, with $j \in \{1, 2, 3, 4\}$, denote the transmitter, the receiver and the monitor positions, respectively. The activities were performed within the colored areas as indicated in Table 2. The chairs have been used for the still and sitting down/standing up activities. The other in-place activities were performed in the position of the dark-colored circles.

tains a letter indicating the activity: `W` for walking, `R` for running, `J` for jumping, `L` for sitting still, `S` for standing still, `C` for sitting down/standing up, and `G` for arm exercises. The CFR when no people were present in the environment is also included for each of the scenarios (label E). When the activity was performed multiple times, a sequential number is present after the activity-related letter.

The last eight sub-folders contain data for PI and PC, and were collected from ten volunteers `P3`, `P5-P13`. The `PI` sub-folders contain one CFR trace for each volunteer collected as they *moved freely* in the environment. The person identifier is included in the name of the file by a suffix starting with `p`. `PC` sub-folders consist of different traces, each associated with a different number of people (from 1 to 10) moving in the environment, as indicated in the name of each file after letter n.

The suffixes "`p00`" and "`n00`" indicate the traces associated with the empty environment, that is, without people. The `PC` sub-folders do not contain data for the single person case, as this situation is exactly the one considered for `PI` and, in turn, data in the `PI` sub-folder associated with the same measurement conditions (`Tx-Rx-M` hardware and positions) can be used.

For each set, Table 2 provides information about the monitored environment and the person/people involved (`Pi`), together with the position and hardware (brand) of the transmitter (`Txj`), receiver (`Rxj`) and monitor (`Mj`). The positions of the devices are shown in Fig. 2a for the bedroom and in Fig. 2b for the other environments. The colors in Table 2 and Fig. 2 indicate the areas where the activities were performed. The volunteer moved freely within the colored areas for the walking and running experiments. The sitting still and sitting down/standing up activities were performed where the chairs are positioned. The other in-place activities, that is, jumping, standing still, and doing arm exercises, were performed in the location indicated by the dark-colored circles.

For the bedroom we considered two configurations to allow evaluating the impact of an

obstruction in the environment on the sensing performance. In the former, the monitor was in position `M1` and, in turn, there existed a direct path between the transmitter and the monitor (sets `AR-1`, `AR-2`). As for the latter, the monitor was in position `M2` so that the direct path (`Tx-M2`) was occluded by the bookcase in the middle of the room (gray rectangle in Fig. 2a), and the person was required to move in both the green and the red areas (sets `AR-3`, `AR-4`). Two scenarios were also considered for the meeting room. The data was concurrently collected from two Wi-Fi networks, entailing `Tx1-Rx1` and `Tx2-Rx2` respectively, through two monitor devices (`M3` and `M4`). This configuration generated four simultaneous acquisitions for each `PI` (`PC`) experiment: the files with the same suffix in different sub-folders refer to simultaneous collections. The `Tx2-M3` and `Tx2-M4` paths were obstructed by the 20-cm thick concrete block wall of the meeting room (sets `PI-2`, `PI-4`, `PC-2`, `PC-4`). As the data with and without the obstruction were concurrently collected, the impact of the obstruction can be deeply evaluated.

Note that the dataset is intended for the development of AR, PI, and PC applications based on the analysis of the CFR, as shown in, for example, [4, 11, 12]. The precise location of the subject within the environment was not recorded making the dataset not suitable for localization and tracking purposes.

EXAMPLES OF USE

The dataset enables the effective design of several smart building applications. The AR sets can be used to design applications for assisted living and smart entertainment. The PI sets can be used to design solutions for intrusion detection and keyless access to access-restricted areas. Finally, the PC sets enable controlling the number of persons that can concurrently share the same environment to guarantee safe inter-person distance. The dataset allows implementing algorithms that generalize over different domains such as the person, the environment, and the day of measurement. This feature is of paramount importance for the integration of sensing functionalities within commercial Wi-Fi devices, to provide plug-and-play sensing solutions. Note that the characteristics of the Wi-Fi channel strictly depend on the environment, and the channel variability is person-specific [5]. Moreover, small changes in the displacement of furniture make the CFR vary over time. In Fig. 3 we report two examples of CFR collected in the empty bedroom on two different days and the Pearson correlation coefficient [8] computed for each pair of traces acquired in the empty bedroom on different days and averaged over the monitoring antennas. The Pearson coefficient takes values in the range [–1, 1] where zero means that the traces are uncorrelated, while an absolute value of 1 means that the traces are linearly dependent. Both the qualitative plots and the quantitative evaluation show that the CFR traces are almost uncorrelated, confirming the need for domain robustness. The processed data and Python code to reproduce this analysis are publicly available at https://codeocean.com/capsule/4811042/tree.

The unprecedented domain diversity offered

by our dataset enables the analysis of the statistics of the Wi-Fi CFR under many different domains; and the design and evaluation of the effectiveness of machine and deep learning algorithms for wireless sensing activity recognition, person identification and people counting that leverage domain-independent features, like, for example, SHARP [4], or domain-adaptive techniques, like, for example, one-shot or few-shot learning [12]. Specifically, we identified seven different use cases as detailed below.

WHS Robustness Over Time

CFR measurements collected on many different days in the same environment and involving the same subject are included in the dataset, e.g, `AR-1a-e`, `AR-5a-b`. The analysis of this data allows understanding how small environmental changes impact the Wi-Fi channel statistics, enabling the design of algorithms that are robust to such changes.

WHS Robustness Over Environments and People

Channel recordings have been collected by having the same subject performing activities in different environments, for example, `AR-1a-e`, `AR-5a-b`, and `AR-6a`, and having different subjects performing activities in the same (`AR-1a` and `AR-2a`) or different environments (`AR-7a` and `AR-8a`). This notable diversity is very useful to test AR algorithms that are to be integrated into commercial devices, to offer plug-and-play sensing functionalities and environment/person adaptability.

WHS Robustness Over Wi-Fi Devices

Our dataset allows evaluating the performance of algorithms trained on CFR collected when data traffic is generated by Netgear routers on CFR collected when Asus routers are transmitting and vice versa. This allows studying methodologies to enforce transmission hardware robustness. Preliminary results on our SHARP algorithm [4] indicate that these domain changes (hardware) deserve attention and further research.

WHS Domain Adaptation

The high domain diversity in the dataset allows designing domain adaptive algorithms. This is usually accomplished through learning-based approaches and entails setting initial model parameters based on a specific environment and refining them in case domain changes occur. The refinement is performed by retraining some parameters based on a few examples of data from the new domain, as proposed in, for example, [13].

Impact of Multi-Path and Interference

The measurements collected in the semi-anechoic chamber allow gauging how the same person modifies the multi-path from a controlled and almost multi-path-free semi-anechoic chamber (`AR-9a-c`) to a much more multi-path affected office environment (`AR-8a-b`). Moreover, the dataset has been mostly collected in environments where other Wi-Fi devices were operating beyond our control, thus causing radio interference. The impact of such interference can be evaluated by comparing the traces collected in the semi-anechoic chamber (no interference) with those collected in the other environments.
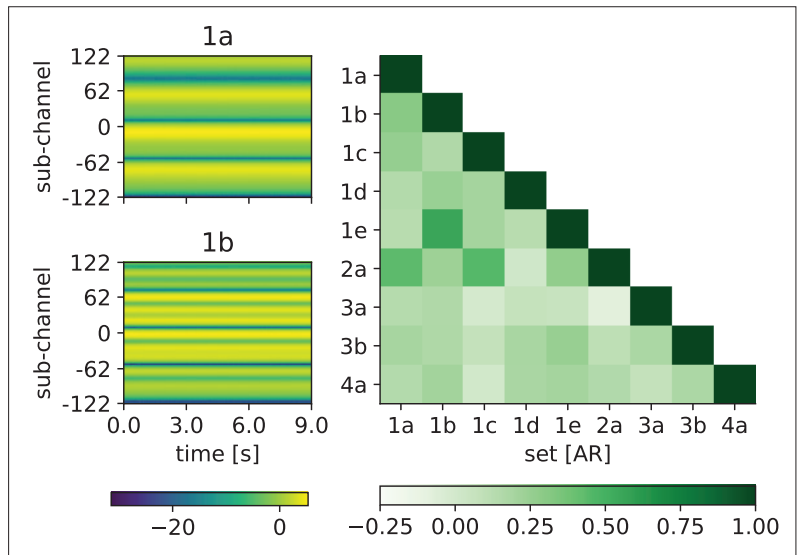
FIGURE 3. Example of CFR amplitude collected in the empty bedroom on two different days (`AR1a_E`, `AR1b_E`) and Pearson coefficient computed between each pair of traces collected in the empty bedroom on different days (`AR` sets).

Impact of Obstructions

Our dataset allows evaluating the robustness of WHS techniques when the path between the transmitter and the monitor is obstructed. This can be done for AR (`AR-3a-b`, `AR-4a`), PI (`PI-2a`, `PI-4a`) and PC (`PC-2a`, `PC-4a`) applications.

Impact of Tx and M Router Locations

The dataset includes traces *simultaneously* collected by two different devices (`M3` and `M4`) that concurrently monitor two transmission links (sets `PI`, `PC`). This allows evaluating how the position of the `Tx-M` link affects the performance of WHS algorithms.

Impact of Time, Frequency, and Spatial Diversity

New research avenues include analyzing the impact of the sampling time on the sensing accuracy by sub-sampling the CFR traces and using them as input for sensing algorithms. A second aspect regards analyzing whether sensing can benefit from selecting a sub-set of the collected OFDM sub-channels that may be more relevant for specific tasks (e.g., [5]). Lastly, the impact of the number of monitoring antennas can be assessed (e.g., [4]).

The provided examples are only some of the possible ways to use the dataset. We trust that the variety of available settings will spur new reproducible research.

Conclusions

In this article, we have presented a new comprehensive dataset that allows developing robust and domain-adaptive learning-based models for human sensing, making it possible to train and test the algorithms over different conditions in terms of deployed Wi-Fi devices, person/people involved, and/or environment. This allows tackling a common pitfall of existing techniques, being that they *do not generalize well* as the online working conditions change with respect to those considered in their training phase. Our dataset provides a common ground for performance

assessment, allowing for reproducible research. Future work will include pairing Wi-Fi CFR records obtained through equally spaced antenna arrays and multiple devices with webcam data that provides the ground truth for the subjects' location, thus providing support for developing localization and tracking applications.

Acknowledgment

References

[1] F. Meneghello et al., "CSI Dataset for Wireless Human Sensing on 80 MHz Wi-Fi Channels," 2022, accessed on Oct. 31, 2022; available: https://dx.doi.org/10.21227/xbhv-f125.
[2] C. Chen et al., "Wi-Fi Sensing Based on IEEE 802.11bf," IEEE Commun. Mag., vol. 61, no. 1, 2023.
[3] Y. Zhang et al., "GaitSense: Towards Ubiquitous Gait-Based Human Identification with Wi-Fi," ACM Trans. Sensor Networks, vol. 18, no. 1, 2021.
[4] F. Meneghello et al., "SHARP: Environment and Person Independent Activity Recognition with Commodity IEEE 802.11 Access Points," IEEE Trans. Mobile Computing, 2022.
[5] F. Meneghello et al., "Toward Integrated Sensing and Communications in IEEE 802.11bf Wi-Fi Networks," IEEE Commun. Mag., 2023.
[6] D. Halperin et al., "Tool Release: Gathering 802.11n Traces with Channel State Information," ACM SIGCOMM Computer Commun. Review, vol. 41, no. 1, 2011.
[7] Y. Xie, Z. Li, and M. Li, "Precise Power Delay Profiling with Commodity Wi-Fi," Proc. ACM MobiCom, 2015.
[8] F. Gringoli et al., "Free Your CSI: A Channel State Information Extraction Platform For Modern Wi-Fi Chipsets," Proc. ACM WiNTECH, 2019.
[9] S. Yousefi et al., "A Survey on Behavior Recognition Using Wi-Fi Channel State Information," IEEE Commun. Mag., vol. 55, no. 10, 2017.
[10] L. Guo et al., "Wiar: A Public Dataset for Wi-Fi-Based Activity Recognition," IEEE Access, vol. 7, 2019.
[11] A. Zhuravchak, O. Kapshii, and E. Pournaras, "Human Activity Recognition based on Wi-Fi CSI Data — A Deep Neural Network Approach," Procedia Computer Science, vol. 198, 2022.
[12] G. Forbes, S. Massie, and S. Craw, "Wi-Fi-Based Human Activity Recognition using Raspberry Pi," Proc. IEEE ICTAI, 2020.
[13] N. Bahadori, J. Ashdown, and F. Restuccia, "ReWiS: Reliable Wi-Fi Sensing Through Few-Shot Multi-Antenna Multi-Receiver CSI Learning," Proc. IEEE WoWMoM, 2022.
[14] J. Schäfer et al., "Human Activity Recognition Using CSI Information with Nexmon," Applied Sciences, vol. 11, no. 19, 2021.
[15] S. Di Domenico et al., "A Trained-Once Crowd Counting Method using Differential Wi-Fi Channel State Information," Proc. ACM WPA, 2016.

Biographies

FRANCESCA MENEGHELLO is an Assistant Professor at the University of Padova, Italy. Her research interests include deep-learning architectures and signal processing with application to wireless communications and sensing.

NICOLÒ DAL FABBRO is a Ph.D. student at the University of Padova, Italy. His research interests are in optimization over networks and wireless sensing.

DOMENICO GARLISI is an Assistant Professor at the University of Palermo, Italy. He has been a Visiting Researcher at the Department of Computer Science, UCLA. His main research is related to wireless networks, software-defined radio, sensor networks, IoT and wireless systems for user localization and testing.

ILENIA TINNIRELLO is a Full Professor at the University of Palermo, Italy. She has been Visiting Researcher at the Seoul National University, Korea, and the Nanyang Technological University of Singapore. Her research activities focus on wireless networks, and in particular on the design and prototyping of protocols and architectures for emerging reconfigurable wireless networks. She has been involved in several European research projects.

MICHELE ROSSI is a Full Professor at the University of Padova, Italy. His research interests include wireless sensing, green mobile networks, edge and wearable computing and the use of machine and deep learning algorithms for signal processing. He currently coordinates the EU ITN project GREENEDGE (GA no. 953775) on green edge computing. He currently serves on the Editorial Boards of the IEEE Trans. Mobile Computing, and of the IEEE Open J. the Communications Society.