Towards a Quantitative Analysis of CSI for AI/ML Based Sensing

Elena Tonini*, Francesco Gringoli*, Renato Lo Cigno*, Marco Cominelli†
*DII, University of Brescia and CNIT, Italy.
†DEIB, Politecnico di Milano, Italy.

***Abstract*—Channel State Information (CSI) sensing is now an established element of Integrated Sensing and Communication (ISAC) operations, but what is its real potential, and what are its limits? The literature focused more on sophisticated AI systems to exploit CSI variations imposed by different propagation scenarios, indeed achieving amazing results, but few, if any works tackled the topic of characterizing the long-term CSI behavior, its stability, and its stochastic properties to achieve insight in the potential and limits of CSI sensing. This work presents a first attempt in this direction, providing a framework that allows the comparison of CSIs quantifying the difference between CSI collected in different scenarios and showing that a quantitative analysis of the CSI is possible, and it can also help to explain the accuracy difference observed between distinct experiments with a CNN-based localization method taken from the literature.**

I. Introduction

Passive Integrated Sensing and Communication (ISAC), that is, without the cooperation of a device, performed through the Channel State Information (CSI) analysis, has received keen attention in the past ten years or so. Starting from seminal papers like [1], [2] to recent advances presented (among many others) in [3], [4], the community explored many techniques to extract the, presumably rich information that the ambient embeds within the propagating signal. Both IEEE and 3GPP consider CSI analysis as a fundamental staple for future Wi-Fi and beyond-5G networks: Essential as a means to enhance communications, including advanced Multiple-Input Multiple-Output (MIMO) and Multi User-MIMO (MU-MIMO) techniques, but integral also to ISAC applications.

Most works in the area of CSI-based ISAC use Artificial Intelligence (AI) techniques, either supervised or not, to extract the information from the CSI. Techniques range from Deep Neural Network (DNN) [5] to auto-encoders [6], [7], to Doppler analysis for motion recognition [8],[9], and more. Great attention has recently been paid to the potential that larger bandwidths in new Wi-Fi standards can bring to sensing. IEEE 802.11ax [10] introduced the sub-carrier spacing of $78.125\,\mathrm{kHz}$ up to $160\,\mathrm{MHz}$, while 802.11be [11] introduced 320-MHz channels and the new Extremely High Throughput (EHT) mode.

All this research leads to impressive advances and promising innovations; however, we deem that the full understanding of the potential of CSI use for ISAC is hindered by the lack of analysis and possible models of the CSI properties beyond channel coherence; Sect. I-A provides a short discussion of the few works we found dedicated specifically to this topic. This paper sets out to address the issue, at least as an initial endeavor, of quantifying the amount of information embedded in CSIs and providing the empirical evidence that different scenarios can be effectively told apart without ambiguity. Sect. II provides a brief background on CSI estimation and an overview of the CSI properties analyzed throughout the paper. The contributions (Sect. III) are the introduction of a formal representation of CSI as an information alphabet and a methodology to quantify the (dis)similarity of CSIs collected in various environments. This study confirms the difficulty of analyzing CSIs with non-AI-based approaches, but also highlights interpretative potentials. Sect. IV examines results from [12] with our proposed methodology, showing how it can help identify situations where AI-based approaches yield ambiguous results.

This work is limited to the analysis of Wi-Fi signals but, since 3GPP technologies use Orthogonal Frequency Division Multiplexing (OFDM) modulations as well, the extension of the analysis to 5G and beyond is feasible, if not straightforward.

For the sake of clarity, we will always use the term *scenario* to identify a situation that we wish to be able to sense and isolate from any other: Different rooms are different scenarios, as well as an empty room is a different scenario from a room with any number of people inside. If the application is localization, each position of a person is a specific scenario to be distinguished from the other positions, and so forth. We instead call *experiment* a single session of measurement in a *scenario*. Thus, in a localization application using Machine Learning (ML) methods, the training and testing datasets are different experiments over the same set of scenarios.

A. Related Work

CSI sensing can deal with a wide range of applications, from localization to human detection, to gesture and activity recognition. Current investigations usually employ Deep Learning (DL) to target one specific application at a time [13].

This work was partially supported by the European Union (EU) and the Italian Ministry for Universities and Research (MUR), National Recovery and Resilience Plan (NRRP), through the projects "PROSDS (CUP C89J24000270004) Spoke 4, Structural Project SUPER and EMBRACE (CUP E63C22002070006) Spoke 6, Structural Project SEXTET both part of RESTART (PE00000001)," and "CSI-Future (Joint Communication and Sensing: CSI-Based Sensing for Future Wireless Networks) PRIN 2022 PNRR (CUP D53D23016040001)".

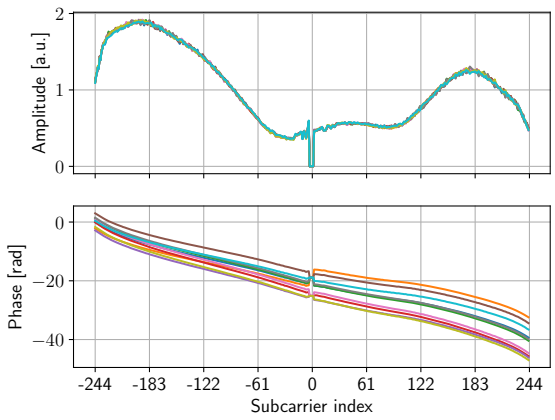

Figure 1: Amplitude (top) and phase (bottom) of 10 different CSI collected during a 10-minute experiment (code E40.0/0, see Tab. I). Amplitude values are normalized by the integral of the energy to remove the effect of the AGC.

However, even if the results are often impressive in terms of accuracy, there are few (if any) efforts to analyze what its limits are. In general, DL techniques are based on the collection of very large labeled datasets for each target application, which is often a labor-intensive and time-consuming endeavor. In this case, while cross-domain Wi-Fi sensing can mitigate difficulties in data collection activities [14], it is still unclear whether it provides reliable indications about how sensing algorithms perform in different scenarios.

To the best of our knowledge, this is the first work that seeks to characterize the stochastic properties of Wi-Fi CSI to investigate the fundamental limits of CSI sensing. In state-of-the-art works, the representation of the stochastic properties of the CSI is somewhat embedded inside the DL model [15]; however, such representation has two major shortcomings: First, it is not general by definition, as it is strongly tied to the specific DL model and not to the statistical properties of the CSI; second, and more importantly, the DL-based representation is extremely complex, and thus it cannot be easily analyzed nor explained. For this reason, performance theoretical analysis is often limited to very simple classification tasks and models [16].

The recent work in [17] uses statistical methods to understand whether some CSI subcarriers contribute more to the overall accuracy of a breath detection application. However, the goal of the analysis is to find how to improve existing algorithms and not to explore the more fundamental question posed by this work about Wi-Fi sensing opportunities and limitations.

II. CSI Estimation and Sensing Background

The CSI of a Wi-Fi frame is the representation of the channel distortion as measured at the receiver: $\mathbf{C}(n) = A_{\mathbf{C}}(n)e^{j\angle\mathbf{C}(n)}$ where $n \in [-N_{\text{SC}}, N_{\text{SC}}]$ is the subcarrier index of the OFDM modulation. Fig. 1 shows $A_{\mathbf{C}}(n)$ (top) and $\angle\mathbf{C}(n)$ (bottom) of 10 different CSIs collected at random times during experiment E40.0/0 (see Tab. I). The CSIs are remarkably similar to one another, and this similarity, together with subtler features not always easy to identify, is what DL algorithms exploit to classify different scenarios.

Albeit usually described as the 'channel response', the CSI is a local estimation of it, normally extracted with some tool from the chipset, represented in some arbitrary measure unit ([a.u.]) as IQ components of the received signal. The extraction is cast onto floating point numbers, also affected by changes in the Automatic Gain Control (AGC), but the original extraction is based on some fixed point representation of the Analog-to-Digital Converter (ADC) component of the chipset, which is unfortunately not known by the extractor[1]. Either $A_{\mathbf{C}}(n)$ (alone or with the phase $\angle\mathbf{C}(n)$), or directly the IQ values are fed to DL algorithms for training and testing purposes. The goal of sensing is the recognition of the same scenario in *collections* of CSIs $\mathbf{C}(k,n)$ measured in different moments. $k \in [1, M_{\mathbf{C}}]$ is the sequence number (ordering) in the collection, which consists in $M_{\mathbf{C}}$ CSIs. $M_{\mathbf{C}}$ is normally in the order of some thousands for training purposes and as small as possible, compatible with achieving the desired accuracy, for testing purposes.

Model-based scenario identification seems unfeasible at present; however, this is no reason to stop investigating if CSIs can be analyzed and interpreted through an appropriate lens that can give an insight into, and maybe inform and direct, the sensing process.

From the perspective of a generic observer, $\mathbf{C}(k,n),\ n \in [-N_{\text{SC}}, N_{\text{SC}}]$ is a sample of a multi-varied $(2 \cdot (2N_{\text{SC}} + 1)$ variables: amplitude and phase of each subcarrier) random variable with unknown distributions, dependent on the specific scenario S_j considered. More appropriately, a collection of CSI $\mathbf{C}(k),\ k \in [1, M_{\mathbf{C}}]$ is a realization of a Discrete-Time Discrete-Space (DTDS) random process sampled from the specific subset of the CSI population conditioned on the scenario S_j. The marginal distributions and the correlation structure of this process are what DL algorithms exploit to sense the environment, i.e., to distinguish the realizations of this process that are conditioned on different scenarios.

For the sake of simplicity, in this paper, we limit the analysis to the amplitude $A_{\mathbf{C}}(n)$, leaving the analysis of the phase for future work. Of particular interest for the rest of the paper is the process of the single subcarrier n amplitude increments:

$$\begin{aligned}\delta_{\mathbf{C}}(k,n) =& A_{\mathbf{C}}(k,n) - A_{\mathbf{C}}(k-1,n);\\ & k \in [2, M_{\mathbf{C}}], n \in [-N_{\text{SC}}, N_{\text{SC}}]\end{aligned} \tag{1}$$

and the auto-covariance of the single subcarrier n amplitude:

$$\begin{aligned}R(\tau,n) =& \mathbb{E}[A_{\mathbf{C}}(k,n), A_{\mathbf{C}}(k-\tau,n)] - \mu_n^2\\ & k \in [1, M_{\mathbf{C}}], n \in [-N_{\text{SC}}, N_{\text{SC}}], \tau \in [0, M_{\mathbf{C}}/2]\end{aligned} \tag{2}$$

assuming the DTDS of the CSI is locally (limited to the single experiment) wide-sense stationary. If this assumption does not hold, the possibility of 'sensing' exploiting the CSI properties fails. τ is limited to $M_{\mathbf{C}}/2$ to have a good estimate of $R()$, and μ_n is the sample mean of the n-th subcarrier.

[1] We use here an extension of the Nexmon CSI Extractor tool, which was presented for the first time in [18] and later extended in collaboration with the SEEMOO research group at the University of Darmstadt.

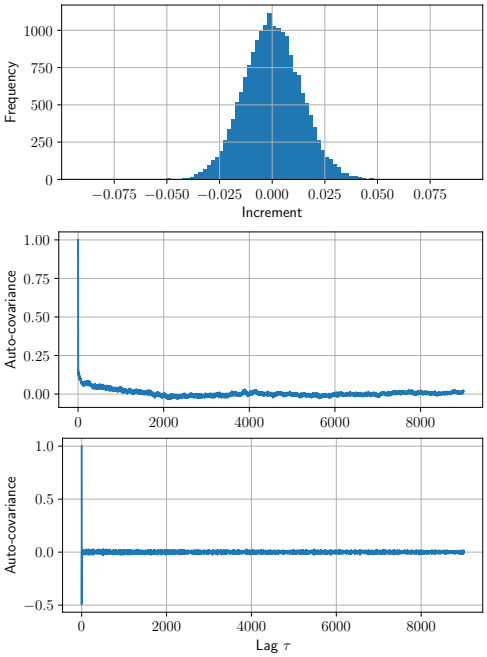

Figure 2: Fundamental properties of subcarrier amplitudes (experiment E40.0/0). Top: Distribution of the increments (Eq. (1)). Middle: Auto-covariance of the amplitude process (Eq. (2)). Bottom: Auto-covariance of the increments process.

Given the CSI extraction procedure and the differences that arise from distinct bandwidths, technologies and devices, the CSI must be properly normalized to be able to make meaningful comparisons. The detailed description of the process can be found in [19], but in brief, it entails the following three steps: *i)* Normalize the energy of each CSI to compensate the AGC variations; *ii)* Find the maximum and the minimum of $A_{\mathbf{C}}(k),\ k \in [1, M_{\mathbf{C}}]$ in each experiment; *iii)* Map the amplitudes of all the CSI in the experiment in the interval $[0, 1]$. To avoid cluttering the notation, from now on $A_{\mathbf{C}}(k,n)$ and $\delta_{\mathbf{C}}(k,n)$ refer to the output of this normalization process.

Fig. 2 reports some fundamental properties of a single subcarrier for a typical experiment after $A_{\mathbf{C}}$ has been normalized as described above. The top plot reports the distribution of the increments (Eq. (1)); the middle plot the auto-covariance of the amplitude process (Eq. (2)); and the bottom plot the auto-covariance of the increments process. The first observation is that the increments amplitude is limited, and this is supported by all experiments, where the variance of the process is always smaller than 0.015. Indeed, the analysis of the increments, which is simple and fast, allowed us to identify a few anomalies in the experiments E40.0/1,2, where some of the captured CSI were clearly a measurement anomaly: We collect only CSI of decoded frames, but these anomalies would not allow frame decoding; thus, they must be a glitch of the CSI extraction process, we will discuss later why this is indeed important and may affect DL sensing methods that do not identify these anomalies. The auto-covariance of $A_{\mathbf{C}}$ (middle plot) indicates a mild memory, extending for several hundred frames; recall that the auto-covariance does not include the mean of the process, which indeed represents the main long-term characterization of each subcarrier in the CSI. These two observations are verified quantitatively across all experiments and all subcarriers.

The bottom plot of Fig. 2 reports the auto-covariance of the increments, which instead shows an unexpected behavior: This process has a 1-step strong negative memory ($R(1,n) \simeq -0.4$), and this is verified for all n; however, this memory is scenario-dependent, and we do not have an explanation for its value. We call this value $R1$ and report it in the last column of Tab. I for the sake of completeness. Since $R1$ is very similar for multiple experiments of the same scenario, we report the mean across the experiments.

The conclusion of this first analysis, and the first takeaway of this work, is that the amplitude of a CSI can be modeled as a Markov process, but for the anomaly of $R(1,\cdot)$, which has non-marginal implications in a learning process. Modeling the amplitude of CSI subcarriers as a Markov process was a working hypothesis of [20], where it was used to define a privacy-preserving obfuscation method to prevent CSI-based sensing, these measures (partially) confirm that hypothesis.

III. CSI: An Information Alphabet

The processing presented in Sect. II gives insight into the properties of a single realization, i.e., one experiment, but does not shed light on what makes different scenarios distinguishable. A quantized CSI lends itself to an interpretation as a symbol from an alphabet carrying information on the scenario it is correlated to, and this can be exploited for further analysis.

A first, elementary consideration is that if the amplitude of each subcarrier is represented over q_{amp}, then the size of the alphabet is $2^{q_{\text{amp}} \times n}$, which indicates an interestingly large alphabet (or learning space): if $q_{\text{amp}} = 8$ and $n = 1024$ (802.11ax with $80\,\text{MHz}$ bandwidth), then the size of the alphabet is $2^{8192} \simeq 10^{2450}$. Interestingly large, as it promises to support fine and precise sensing, but also taxingly large, making it difficult to perform numerical analysis. Indeed, such a large alphabet makes it simply impossible to proceed with a brute-force probabilistic approach, because probabilities are too small to be handled numerically, thus we proceed with a simpler approach based on the distance between the CSI amplitudes, leaving a probabilistic analysis as future work.

The first step to proceed further is to understand what is the proper value for q_{amp}. We estimate the number of bits to use starting from the minimum number of bits q_{inc} necessary to properly represent the process of the increments (Eq. (1)). The increments are not bounded, but as shown in the top plot of Fig. 2, they are rather compact and, after the normalization, their absolute value is much smaller than 1. Let $\delta^\star$ be a value such that the tails of the increments distribution outside the interval $[-\delta^\star, \delta^\star]$ are negligible, then q_{inc} can be found ensuring that the difference between the quantized distribution and the measured distribution is below a given threshold. Once again, the process of finding this number of bits is described in

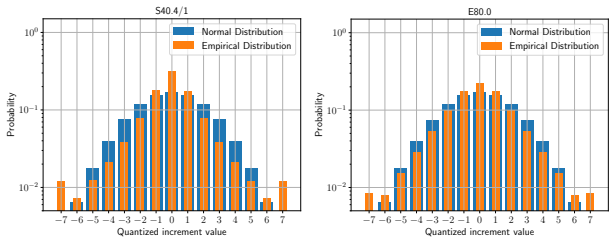

Figure 3: Comparison of the distribution of two experiments increments amplitude quantized over $q_{\text{inc}} = 4$ bits (orange) with a Gaussian distribution with the same variance, also quantized over 4 bits (blue). Values falling outside the interval $[-\delta^\star, \delta^\star]$ are accumulated on the largest possible number represented.
Left: S40.4/1, right: E80.0.

detail in [19] and it is not reported here for lack of space, but extensive tests over many experiments in different scenarios let us conclude that $q_{\text{inc}} = 4$ and $q_{\text{amp}} = 10$ are good values to proceed with the analysis. In this work, we use $\delta^\star = 3\sigma$.

Fig. 3 reports the quantized distribution of the increments amplitude in a semi-logarithmic scale, measured in two very different scenarios: S40.4 and E80.0 (see Tab. I for the experiments explanation). The distribution is measured over all the subcarriers of the CSIs of the experiment and is compared with a Gaussian distribution with the same variance. Values of the distributions falling outside the interval $[-\delta^\star, \delta^\star]$ are accumulated on the largest possible number represented instead of discarded to better represent the probability of large values. It is clear that the increment distribution is not Gaussian, but it is almost perfectly symmetric, supporting the hypothesis of a local stationary process, even if the experiments lasted several minutes. A qualitative comparison with the Gaussian distribution indicates that the increments tails are heavier than those of a Gaussian: What is the implication on sensing is not trivial to tell, but it may indicate that large, sudden, random variations are possible. As expected, the tails are heavier for S40.4, which is a more dynamic scenario, while the bandwidth does not seem to have any influence on the increments process. The analysis of all the other experiments confirms that 4 bits are enough to represent the increments and also that the empirical distribution always has a higher kurtosis, thus higher tails, compared to a Gaussian with the same variance.

After this normalization and quantization, the $A_{\mathbf{C}}$ becomes a vector of $(2N_{\text{SC}} + 1)$ integer numbers represented over $q_{\text{amp}} = 10$ bits, and the distance between two arbitrary CSI i and j, independently from the experiment is

$$W_D(A_{\mathbf{Ci}}, A_{\mathbf{Cj}}) = \sum_{n=-N_{\text{SC}}}^{N_{\text{SC}}} |A_{\mathbf{Ci}}(n) - A_{\mathbf{Cj}}(n)| \qquad (3)$$

given the two CSI have the same number of subcarriers. Eq. (3) can be used to estimate and quantify the difference between two CSI and consequently if they belong to different scenarios, at least in the hypothesis that CSIs in the same scenario are similar one-another.

To maintain a reasonably simple approach, we define as representative of a specific experiment the average of the CSI amplitudes collected in the experiment:

$$A^\star_{\mathbf{C}}(n) = \frac{1}{M_{\mathbf{C}}} \sum_{k=1}^{M_{\mathbf{C}}} A_{\mathbf{C}}(k, n) \qquad (4)$$

thus we can evaluate the similitude of a generic CSI **Ci** to one scenario S_j by computing the distance $W_D(A_{\mathbf{Ci}}, A^\star_{\mathbf{Cj}})$. Finally, we can have a compact indication of the similarity of two experiments by computing the distance of one experiment from the average of the other:

$$\overline{W_D(A_{\mathbf{Ci}}, A^\star_{\mathbf{Cj}})} = \frac{1}{M_{\mathbf{C}}} \sum_{k=1}^{M_{\mathbf{C}}} W_D(A_{\mathbf{Ci}}(k), A^\star_{\mathbf{Cj}}) \qquad (5)$$

Clearly, $\overline{W_D(A_{\mathbf{Ci}}, A^\star_{\mathbf{Cj}})} \neq \overline{W_D(A_{\mathbf{Cj}}, A^\star_{\mathbf{Ci}})}$, even if we expect the two values to be similar.

We are aware that a probabilistic approach, e.g., with a chi-square test, would be more powerful; however, as already discussed, this approach is numerically impossible; thus, we proceed to evaluate if the simple approach presented is useful to interpret CSI sensing or not.

Scenario	**# Exp.**	**Duration**	**# CSI**	$R1$
E20.0	4	10 min.	> 75000	−0.49
E40.0	3	10 min.	> 55000	−0.46
E80.0	1	10 min.	> 18000	−0.49
S20.1	4	10 min.	> 60000	−0.44
S40.1	1	10 min.	> 18000	−0.18
S40.2	1	10 min.	> 18000	−0.12
S40.3	1	10 min.	> 18000	−0.02
S40.4	2	10 min.	> 36000	−0.02
S40.5	2	10 min.	> 36000	−0.24
S80.1	1	10 min.	> 18000	−0.18
S80.2	2	10 min.	> 36000	−0.01
S80.3	1	10 min.	> 18000	−0.09
S80.4	1	10 min.	> 18000	+0.02
S80.5	1	1 min.	> 2600	+0.25
M20.4	4	10 min.	> 58000	−0.09

Table I: Summary of the experiments with 802.11ax. Each scenario is identified using the 'code' `XB.n` where `X` can be E (Empty office), S (people Sitting), or M (people Moving); `B` is the Bandwidth (20, 40 or 80 MHz); and `n` is the number of people in the room. When needed, different experiments in the same scenario are distinguished with /0,1,2,3 at the end of the scenario code. The last column reports the auto-covariance parameter of the increments as discussed in Sect. II. The dataset is available at https://doi.org/10.5281/zenodo.14651521.

IV. Experiments and Results

The results we present are based on the experiments reported in Tab. I (performed during several months) and additional localization experiments reported in [12]. All the experiments have been carried out by the Advanced Networking Systems research group[2] of the University of Brescia. The office used for the experiments is depicted in Fig. 4; the Tx and Rx are both RT-AX82U Access Points (APs) from Asus equipped with the Wi-Fi Broadcom chipset BRCM43684. The frame generation rate is around 30 frames per second (fps) in all experiments but S80.5 where the generation rate was higher. The total amount of data is huge, and space allows only for the presentation of a few selected results, which, however, have all been confirmed and are supported by the analysis of the entire dataset, which is public as indicated in Tab. I.

We start analyzing the distribution of the W_D values for some experiments, i.e., $W_D(A_{\mathbf{Ci}}, A^\star_{\mathbf{Ci}})$ for selected scenarios

[2] https://ans.unibs.it

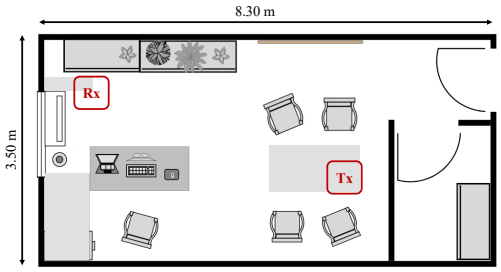

Figure 4: Office layout in the experimental testbed.

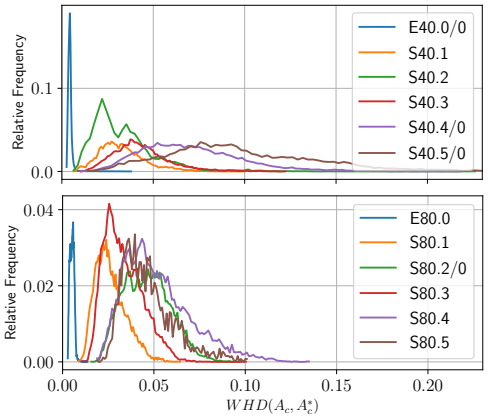

Figure 5: Distribution of $W_D(A_{\mathbf{Ci}}, A^\star_{\mathbf{Ci}})$ for 40 MHz (top) and 80 MHz (bottom) scenarios of type E and S.

S_i. In particular, scenarios with B=40 MHz and 80 MHz with the empty office or a variable number of sitting people. Fig. 5 reports results revealing that such distribution can be used to easily separate the Empty scenario from other configurations with people inside the room, regardless of the bandwidth of the CSI. In particular, the CSIs in the Empty scenario are very similar, exhibiting a very narrow peak of W_D close to 0 with both 40 MHz and 80 MHz; instead, when people are present in the room, W_D increases, meaning that the CSIs are changing *more* during the experiments. Interestingly, as more people are inside the room, both the average W_D and the standard deviation of the distributions increase, validating the intuition that when the scenario becomes more dynamic, the propagation channel becomes more variable and unpredictable. The key takeaway is that simple techniques can already tell apart an empty space from a space with people, while counting people may be more difficult.

An interesting question is whether a very compact parameter can give useful information. To this end, Tab. II reports $\overline{W_D(A_{\mathbf{Ci}}, A^\star_{\mathbf{Cj}})}$ normalized to 1, i.e., divided by $(2N_{\text{SC}}+1)\cdot(2^{q_{\text{amp}}}-1)$, for the 10 experiments at 40 MHz (the analysis for the 20 and 80 MHz confirms the results). We only report the 3 decimal numbers for conciseness (e.g., $0.0491 \rightarrow 049$). This is just the average distance, thus we cannot expect structured and rich information, however, we can do some interesting observations.

First, in all cases, the minimum distance of the CSIs distribution from the average is the one for the same experiment. Second, when we consider E40.0 and S40.5 scenarios the distance between different experiments

		$A_{\mathbf{C}}$									
		E40.0/0	E40.0/1	E40.0/2	S40.1	S40.2	S40.3	S40.4/0	S40.4/1	S40.5/0	S40.5/1
$A^\star_{\mathbf{C}}$	E40.0/0	**005**	012	016	257	181	135	146	144	139	169
	E40.0/1	009	**009**	023	251	175	131	141	138	136	165
	E40.0/2	016	025	**005**	341	270	145	158	155	150	180
	S40.1	255	249	269	**033**	180	252	247	243	245	229
	S40.2	180	174	190	177	**045**	224	195	171	208	170
	S40.3	127	124	139	251	224	**044**	128	161	091	156
	S40.4/0	130	124	144	237	189	117	**072**	094	104	144
	S40.4/1	125	117	138	228	161	149	086	**079**	125	145
	S40.5/0	132	119	126	239	206	080	108	133	**062**	116
	S40.5/1	138	132	152	210	155	129	126	130	090	**099**

Table II: Normalized $\overline{W_D(A_{\mathbf{Ci}}, A^\star_{\mathbf{Cj}})}$ for experiments at 40 MHz; in bold, the minimum distance for each experiment.

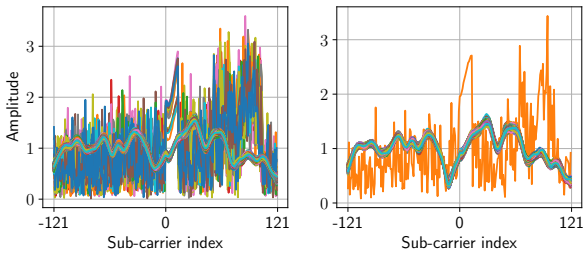

Figure 6: CSI samples from the dataset in [12] before filtering out the spurious data.

remains very small, and this is good news because it indicates that scenarios are clearly recognizable. The analysis of more complex metrics, e.g., the inter-CSI distance distribution will probably be more informative and can help select appropriate DL methods to classify scenarios.

So far, we have analyzed data collected explicitly for this work, and many more results are obtained on this dataset than we can fit in this paper, but we want to test if this methodology can also work on datasets collected for other purposes. To this end, we use a different dataset[3], collected for the work in [12]. The dataset contains CSI samples from 80-MHz 802.11ac frames, the Tx is an Ettus USRP N300 software-defined radio, and the Rx is a commercial AP (Asus RT-AC86U); thus, the technology and the devices are completely different from the other experiments. CSI samples are collected over a short time frame with a person standing in 8 different target positions (P1 ... P8) inside a room, thus corresponding to 8 scenarios in total. The captures have been repeated twice for each location because the goal in [12] was to train (and then test) a localization system based on a neural network with appropriate training and testing. Thus the dataset contains 16 experiments: 8 for training and 8 for testing.

Given the observations about Fig. 2, we verified that also this dataset contains similar anomalies, indicating that such anomalies are not related to the specific hardware used. Actually, the anomalies may also appear in many other works, albeit at this stage of the investigation, we cannot tell if they are caused by the experimental setup or are in some way intrinsic to the CSI reporting. Fig. 6 reports the amplitudes of 1000 CSI amplitudes for the training experiments in P3 and P6, mixing regular with anomalous ones, which are clearly physically impossible for a standard Wi-Fi frame. Analyzing

[3] The dataset is publicly available at https://zenodo.org/records/5885636, and while its original goal was to test an anti-sensing application to protect users' privacy using "distorted" transmissions, the repository also contains clean CSI data that suits the analysis carried out in this work.

		TESTING							
	POS	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
TRAINING $A^{\star}_{\mathbf{C}}$	**1**	**026**	050	062	076	104	050	069	072
		93.7%	0.3%	0.3%	0.3%	0%	0%	0%	0%
	2	072	**034**	037	046	093	061	052	041
		3.1%	**95.1%**	6.2%	1.8%	0%	0%	0.3%	0.7%
	3	079	054	**022**	050	094	065	067	042
		0.3%	1%	**92.7%**	0%	0%	0%	0%	1.1%
	4	086	065	058	**033**	087	063	067	059
		2.9%	0%	0%	**96.5%**	0%	0%	0%	0%
	5	143	112	102	085	**037**	113	112	109
		0%	0%	0%	0%	**100%**	0%	0%	0%
	6	067	051	051	061	080	**034**	060	067
		0%	0.4%	0%	0%	0%	**98.5%**	1.4%	0%
	7	080	051	060	058	094	058	**039**	056
		0%	1.6%	0%	0%	0%	1.5%	**98.3%**	0%
	8	089	059	040	052	103	075	062	**031**
		0%	1.6%	0.8%	1.1%	0%	0%	0%	**98.2%**

Table III: Average normalized W_D computed on the dataset in [12] (case rx1). In each cell, the top number is the W_D (first three decimal digits, lower is better), while the number below is the classification accuracy in the associated confusion matrix.

in details the results of [12] it emerges that the classification was less accurate exactly for these training sets, but the DL algorithm could obviously not identify this situation. Tab. III reports, together, the confusion matrix of the measures from [12] after they have been cleaned from the anomalies and $\overline{W_D(A_{\mathbf{Ci}}, A^{\star}_{\mathbf{Cj}})}$, where $A^{\star}_{\mathbf{Cj}}$ refers to the training and $A_{\mathbf{Ci}}$ to the testing. In each cell of the table the number on the top is the normalized $\overline{W_D(A_{\mathbf{Ci}}, A^{\star}_{\mathbf{Cj}})}$, while the percentage on the bottom is the accuracy of the confusion matrix. Also in this case the minimum distance is achieved for the same scenarios, reinforcing the confidence in the classification of the DL algorithm, and in some sense explaining or at least justifying it.

V. Conclusions and Future Work

This work opens a research thread that may help understand and interpret the potential and limitations of CSI sensing. The impossibility of using electromagnetic propagation models to explain and justify the CSI characteristics used in sensing with powerful AI techniques, together with the lack in the literature of long-term characterization (in stochastic terms) of the propagation environment has so-far limited the possibility to fully interpret CSI-sensing results, both for good and for bad, i.e., to understand why in some cases it shows amazingly good results, while in others it seems to fail. Failures are very often related to the difficulty of replicating good results in different experiments, but sometimes also to performance differences within the same experiment. We have shown that the CSI can indeed carry an enormously large amount of information, but analysis may fail because of the complexity itself or because of anomalies in the CSI extraction process, an observation that was never done before to the best of our knowledge and that may have affected many works, as AI/ML processing does not allow easy identification of anomalies.

We plan to proceed further in this research line, extending the public repository of CSI collected in scenarios described with all the detail possible, improving the analysis tools, possibly including probabilistic approaches that will improve the interpretation of results, and building generative models of CSI to augment the datasets, simplifying and streamlining future data collection campaigns.

References

[1] F. Adib and D. Katabi, "See through walls with WiFi!" In *ACM Int. Conf. of the Special Interest Group on Data Communication (SIGCOMM)*, Aug. 2013, pp. 75–86.

[2] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. Ni, "CSI-Based Indoor Localization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 7, pp. 1300–1309, Jul. 2013.

[3] H. Guan, A. Sharma, D. Mishra, and A. Seneviratne, "Experimental Accuracy Comparison for 2.4GHz and 5GHz WiFi Sensing Systems," in *IEEE ICC 2023*, May 2023, pp. 4755–4760.

[4] M. Cominelli, F. Gringoli, and F. Restuccia, "Exposing the CSI: A Systematic Investigation of CSI-based Wi-Fi Sensing Capabilities and Limitations," in *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Mar. 2023, pp. 81–90.

[5] Z. He, X. Zhang, Y. Wang, Y. Lin, G. Gui, and H. Gacanin, "A Robust CSI-Based Wi-Fi Passive Sensing Method Using Attention Mechanism Deep Learning," *IEEE Internet of Things Journal*, vol. 10, no. 19, pp. 17 490–17 499, 2023.

[6] M. Chahoushi, M. Nabati, R. Asvadi, and S. A. Ghorashi, "CSI-Based Human Activity Recognition Using Multi-Input Multi-Output Autoencoder and Fine-Tuning," *Sensors*, vol. 23, no. 7, 2023.

[7] M. Cominelli, F. Gringoli, L. M. Kaplan, M. B. Srivastava, and F. Cerutti, "Accurate Passive Radar via an Uncertainty-Aware Fusion of Wi-Fi Sensing Data," in *2023 26th IEEE Int. Conf. on Information Fusion (FUSION)*, IEEE, 2023, pp. 1–8.

[8] F. Shi, W. Li, C. Tang, P. Brennan, and K. Chetty, "Doppler Sensing Using WiFi Round-Trip Channel State Information," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2023, pp. 1–6.

[9] W. Li, M. J. Bocus, C. Tang, R. J. Piechocki, K. Woodbridge, and K. Chetty, "On CSI and Passive Wi-Fi Radar for Opportunistic Physical Activity Recognition," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 607–620, 2022.

[10] 802.11ax, *IEEE Standard for Information technology – Part 11: Wireless LAN MAC and PHY Specifications Amendment 1: Enhancements for High-Efficiency WLAN*.

[11] 802.11be, *IEEE Draft Standard for Information technology – Part 11: Wireless LAN MAC and PHY Specifications Amendment: Enhancements for Extremely High Throughput (EHT)*.

[12] M. Cominelli, F. Gringoli, and R. Lo Cigno, "AntiSense: Standard-compliant CSI obfuscation against unauthorized Wi-Fi sensing," *Elsevier Computer Communications*, vol. 185, pp. 92–103, Mar. 2022.

[13] I. Nirmal, Khamis, and et al., "Deep Learning for Radio-Based Human Sensing: Recent Advances and Future Directions," *IEEE Comm. Surv. & Tutorials*, vol. 23, no. 2, 2021.

[14] C. Chen, G. Zhou, and Y. Lin, "Cross-Domain WiFi Sensing with Channel State Information: A Survey," *ACM Comput. Surv.*, vol. 55, no. 11, Feb. 2023. [Online]. Available: https://doi.org/10.1145/3570325.

[15] J. Xue, J. Zhang, Z. Gao, and W. Xiao, "Enhanced WiFi CSI Fingerprints for Device-Free Localization With Deep Learning Representations," *IEEE Sensors Journal*, vol. 23, no. 3, pp. 2750–2759, 2023.

[16] M. Cominelli et al., "PRIVAtE: Passive Radar Interpretability using Variational Auto Encoders," in *2023 IEEE International Workshop on Technologies for Defense and Security (TechDefense)*, 2023, pp. 444–448.

[17] T. Deng, B. Zheng, R. Du, F. Liu, and T. X. Han, "A statistical sensing method by utilizing Wi-Fi CSI subcarriers: Empirical study and performance enhancement," *Journal of Information and Intelligence*, vol. 2, no. 4, pp. 365–374, 2024.

[18] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free Your CSI: A Channel State Information Extraction Platform For Modern Wi-Fi Chipsets," in *13th ACM Int. Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization (WiNTECH)*, 2019, pp. 21–28.

[19] E. Tonini, "Statistical Analysis to Support CSI-Based Sensing Methods," M.S. thesis, University of Brescia, Department of Information Engineering — Advanced Networking Systems (ANS) group, Sep. 2024.

[20] M. Cominelli, F. Gringoli, and R. Lo Cigno, "On the properties of device-free multi-point CSI localization and its obfuscation," *Elsevier Computer Communications*, vol. 189, pp. 67–78, May 2022.