BFA-Sense: Learning Beamforming Feedback Angles for Wi-Fi Sensing

Khandaker Foysal Haque*, Francesca Meneghello†, Francesco Restuccia*
*Institute for the Wireless Internet of Things, Northeastern University, Boston, USA
†Department of Information Engineering, University of Padova, Italy
{haque.k, f.restuccia}@northeastern.edu*, francesca.meneghello.1@unipd.it†

2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) | 979-8-3503-0436-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/PerComWorkshops59983.2024.10503460

***Abstract*—In this paper, we propose BFA-Sense, a completely novel approach to implement standard-compliant Wi-Fi sensing applications. Wi-Fi sensing enables game-changing applications in remote healthcare, home entertainment, and home surveillance, among others. However, existing work leverages the manual extraction of the uncompressed channel state information (CSI) from Wi-Fi chips, which is not supported by the 802.11 standard-compliant devices and hence requires the use of specialized equipment. On the contrary, BFA-Sense leverages the compressed beamforming feedback angles (BFAs) transmitted during the standard-compliant sounding procedure to characterize the propagation environment. Conversely from the uncompressed CSI, BFAs (i) can be recorded without any firmware modification, and (ii) allows a single monitor device to simultaneously capture the channels between the access point and all the stations, thus providing much better sensitivity. We evaluate BFA-Sense through an extensive data collection campaign with three subjects performing twenty different activities in three different environments. We assess the cross-domain adaptability of BFA-Sense through embedding learning for tackling unseen environments with a few samples from the new environment. The results show that the proposed BFAs-based approach achieves about 11% more accuracy when compared to CSI-based prior work.**

***Index Terms*—Wi-Fi sensing, multiple-input multiple-output (MIMO), beamforming, beamforming feedback angles.**

I. Introduction

Wi-Fi human sensing aims at detecting the changes in the propagation of Wi-Fi signals and associating them to the way a subject stays/moves in the environment, thus enabling device-free monitoring solutions [1]–[3]. To date, the vast majority of Wi-Fi sensing systems – discussed in Section II – leverage channel measurements obtained from pilot symbols as sensing primitive. Such measurements – referred to as channel state information (CSI) – describe the way the signals propagate in the environment. Despite leading to good performance, CSI-based techniques require extracting and recording the uncompressed CSI estimated by the Wi-Fi devices. However, this operation is currently not supported by commercially available IEEE 802.11 devices. To enable CSI-based sensing, researchers have introduced custom-tailored firmware modifications that allow extracting the CSI [4]–[6]. These CSI extraction tools only provide support for single-user multiple input multiple output (SU-MIMO) sensing as the channel is sounded on the link between the transmitter and the device implementing the firmware modifications. Therefore, *Wi-Fi sensing approaches relying on CSI extraction tools cannot benefit from the spatial diversity that can be gained through multi-user multiple input multiple output (MU-MIMO) transmissions*. Spatial diversity may be achieved considering multiple CSI collectors but this would increase the computation burden as synchronization among the devices would be needed. Moreover, even if CSI extraction could be supported in the future without the need for custom-tailored firmware modifications, it would require additional processing to extract data from the chip, thus increasing energy consumption. Therefore, we argue that more suitable approaches to Wi-Fi sensing should be put forward.

In this paper, we propose BFA-Sense, an entirely new approach to Wi-Fi sensing that leverages the MU-MIMO capabilities of Wi-Fi to drastically increase sensing performance while substantially reducing the related overhead. As shown in Figure 1, BFA-Sense leverages the compressed beamforming feedback information (BFI) – i.e., the beamforming feedback angles (BFAs). This is periodically transmitted for multiple-input multiple-output (MIMO) operations – to sense the propagation environment between the access point (AP) and the connected stations (STAs).

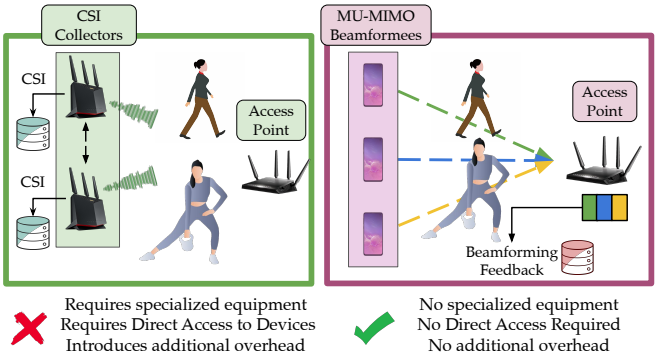


Fig. 1: CSI-based vs BFI-based Wi-Fi sensing.

In stark contrast to CSI-based sensing, BFA-Sense (i) does not need firmware modifications, (ii) does not require synchronization among different devices, since a single monitor device can capture the information about all the MIMO channels established between the AP and the STAs. Since BFAs frames are sent unencrypted to keep the processing delay below a few milliseconds, any off-the-shelf Wi-Fi device can capture and decode them without any firmware modification [7]. Unlike CSI extraction tools that focus on a single MIMO channel, while capturing the BFAs we obtain the information associated with all the STAs involved in a MU-MIMO transmission. Thus, with BFAs-based approach, multiple spatially diverse channel information is collected with a single capture. In summary, this paper provides the following contributions:

• We propose BFA-Sense, a new approach to Wi-Fi sensing where the standard-compliant BFAs routinely sent in MU-MIMO Wi-Fi networks is used to characterize the propagation environment. To the best of our knowledge, this is the first work proposing the utilization of BFAs for sensing;

• We developed an embedding learning model for activity classification to quickly adapt the classifier to new environments with few additional data points;

• We extensively evaluate BFA-Sense through a comprehensive data collection campaign, with three subjects performing twenty different activities in three different environments. We used commercial IEEE 802.11ac Wi-Fi devices to set up a MU-MIMO network. The network was synchronized with a camera-based system to record the activity's ground truth during the experiments.

We show that the proposed BFAs-based approach achieves about 10% more accuracy when compared to state-of-the-art CSI-based techniques. Moreover, leveraging embedding learning BFA-Sense reaches, on average, up to 65% of accuracy in adapting the activity classifier to new environments. For reproducibility, the entire dataset along with the code repository is shared at **https://github.com/kfoysalhaque/BFA-Sense**

II. Related Work

Over the last ten years, a lot of efforts have been made to explore wireless sensing solutions, as summarized by Liu et al. in [8]. The first Wi-Fi sensing approaches were based on the received signal strength indicator (RSSI) [9]. More recently, researchers have focused on the more fine-grained CSI information that describes how the wireless channel modifies signals at different frequencies rather than providing a cumulative metric on the signal attenuation as the RSSI does.

Limitations of CSI-based Sensing. The CSI is computed at the physical layer (PHY) and is currently not made available to the end-user in commercially available devices. To overcome this limitation, in recent years, researchers have developed some CSI extraction tools that modify the firmware of specific devices to unveil the information of interest. Two of them, namely Linux CSI and Atheros CSI, target IEEE 802.11n compliant network interface cards (NICs), for up to 40 MHz of bandwidth [4], [10]. Nexmon CSI, allows extracting the CSI from some IEEE 802.11ac compliant devices, supporting bandwidths up to 80 MHz [5]. The most recent tool, AX CSI is designed for IEEE 802.11ax devices and provides CSI measurements also on 160 MHz channels [11]. These tools, however, need non-trivial firmware modifications of the NICs.

Recent BFI-based Sensing Approaches. BFI is gaining momentum in the research community as a proxy to the CSI as it provides spatially diverse rich channel information from Wi-Fi compliant devices without any firmware modification or direct access to the hardware. Wi-BFI [12] is one of the first tools to extract BFAs and reconstruct BFI in the form of beamforming matrices $\mathbf{V_k}$.

Jiang et al. demonstrated how the design and computation of the beamforming feedback matrix affects the sensing performance [13]. Kondo et al. evaluate the performance of uni-directional and bi-directional beamforming on Wi-Fi sensing through the BFI reconstructed from BFAs [14]. The same authors leveraged the BFI for respiratory rate estimation [15]. However, all these work considers the BFI matrices reconstructed from BFAs which introduces additional preprocessing, system latency, and computational burden for the sensing system. On the contrary, BFA-Sense is based on the compressed BFAs which are directly captured from ongoing transmissions and do not need any pre-processing.

III. The BFA-Sense System

BFA-Sense leverages the channel estimation mechanism standardized in IEEE 802.11ac/ax to sound the physical environment, as summarized in Figure 2 and detailed in the next part of this section. The channel estimation is performed on the STAs (beamformees) and is reported to the AP (beamformer) that uses it to properly beamform MU-MIMO transmissions. The report is transmitted over the air in the form of BFAs in clear text. Since the AP continuously triggers the channel estimation procedure on the connected STAs, *the BFAs contains very rich and spatially diverse information*. Moreover, the reports from different beamformees *can be collected with a single capture* by any Wi-Fi-compliant device, thus reducing the system complexity.

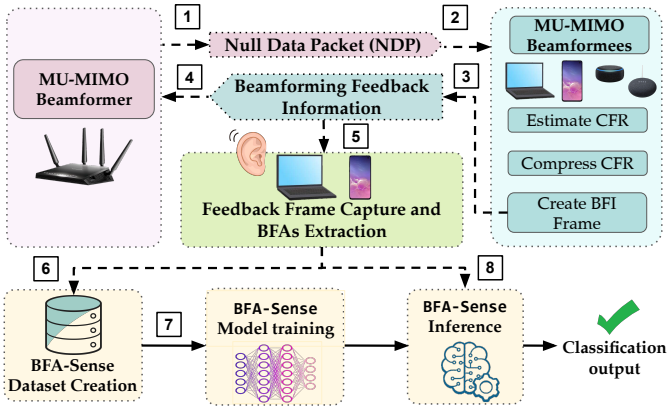


Fig. 2: The BFA-Sense Wi-Fi sensing system.

BFA-Sense Technical Challenges. We stress that BFAs represent a completely new type of data to perform Wi-Fi sensing. While CSI consists of complex I/Q-values, BFAs are expressed as integer numbers derived by compressing and quantizing the CSI information. In this respect, we needed to implement a novel data processing pipeline for the new data type that effectively performs activity classification based on BFAs data and provides adaptation capabilities for new environments. In the following, we detail the BFA-Sense sensing system.

A. BFA-Sense: A Walkthrough

The BFA-Sense sensing system consists of eight steps, as depicted in Figure 2. The process stems from the way beamforming is implemented in IEEE 802.11 networks (steps 1-4). Hence, steps 5-8 are the core of BFA-Sense and empower the

Wi-Fi system with sensing capabilities. In the following, the entire pipeline is detailed.

In MU-MIMO, the beamformer (AP) uses a precoding matrix to linearly combine the signals to be simultaneously transmitted to the different beamformees (STAs). The precoding matrix is derived from the CSI matrices $\mathbf{H}$, estimated by each of the beamformees and that describe how the environment modifies the irradiated signals in their path to the receivers. The estimation process is called *channel sounding* and is triggered by the AP which periodically broadcasts a null data packet (NDP) (**step 1** in Figure 2). The NDP contains sequences of bits – named long training fields (LTFs) – the decoded version of which is known to the beamformees. Since its purpose is to sound the channel, the NDP *is not beamformed* by the AP. This is particularly advantageous for sensing purposes since the resulting CSI estimation is not affected by inter-stream or inter-user interference. The LTFs are transmitted over the different beamformer antennas in subsequent time slots, thus allowing each beamformee to estimate the CSI of the links between its receiving antennas and the beamformer transmitting antennas.

The NDP is received and decoded by each STA (**step 2** in Figure 2) to estimate the CSI $\mathbf{H}$. The different LTFs are used to estimate the channel over each pair of transmitting (TX) and receiving (RX) antennas, for every orthogonal frequency-division multiplexing (OFDM) sub-channel. This generates a $K \times M \times N$ matrix $\mathbf{H}$ for each beamformee, where K is the number of OFDM sub-channels while M and N are respectively the numbers of TX and RX antennas. Next, the CSI is compressed – to reduce the channel overhead – and fed back to the beamformer. Using $\mathbf{H}_k$ to identify the $M \times N$ sub-matrix of $\mathbf{H}$ containing the CSI samples related to sub-channel k, the *compressed beamforming feedback* is obtained as follows ([16], Chapter 13). First, $\mathbf{H}_k$ is decomposed through singular value decomposition (SVD) and the complex-valued beamforming matrix $\mathbf{V}_k$ is defined by collecting the first $N_{\mathrm{SS}} \leq N$ columns of the $M \times M$ right unitary matrix. The beamforming matrix is referred to as the *beamforming feedback information* (BFI) and is used by the beamformer to compute the pre-coding weights for the N_{SS} spatial streams directed to the beamformer.

To reduce the spectrum usage for feedback transmission, $\mathbf{V}_k$ is compressed by converting it into polar coordinates. Using this transformation, the beamformee is only required to feed back to the beamformer the ϕ and ψ angles defining the transformation as they allow reconstructing $\mathbf{V}_k$. To further reduce the channel occupancy, the angles are quantized using $b_\phi \in \{7, 9\}$ bits for ϕ and $b_\psi = b_\phi - 2$ bits for ψ. The quantized values – $q_\phi = \{0, \ldots, 2^{b_\phi} - 1\}$ and $q_\psi = \{0, \ldots, 2^{b_\psi} - 1\}$ – are packed into the compressed beamforming frame (**step 3**) and transmitted to the AP (**step 4**). Since MU-MIMO requires fine-grained channel sounding – every around 10 milliseconds to account for user mobility, according to [17] – it is fundamental to process the BFI in a fast manner at the AP. For this reason, and since cryptography would lead to excessive delays, the angles are currently sent unencrypted. Therefore, *the BFI reports are exposed to and can be read by any device that can access the wireless channel*.

`BFA-Sense` leverages this sounding procedure for sensing by capturing the transmitted BFI report frames (**step 5**) and using the decoded BFAs as a proxy to detect the changes in the environment. This is done by using the BFAs as input for a learning-based neural network classifier (detailed in Section III-B). To reliably estimate the activity being performed by a human moving within the propagation environment, the `BFA-Sense` classifier relies on the BFAs transmitted by all the beamformees in the network and captured during a time window of W seconds.

To keep the model simple for implementation on memory- and battery-constrained devices, we decided to follow a fixed-input approach. Specifically, we consider the average number S of BFI reports counted (at training time) in windows of W seconds. Windows having less than S packets are padded with placeholder packets containing zero-valued angles while packets exceeding such threshold are discarded.

To obtain the training data, the $S \times K \times A \times U$ tensors derived from the BFI reports captured during the data collection phase are stored in a dataset, together with their associated activity label, and a timestamp (**step 6** in Figure 2). Here, S, A, and U denote the number of samples, number of BFAs in BFI report and the number of users in MU-MIMO respectively. The trained model (**step 7**) is then used for online sensing (**step 8**).

B. Embedding Learning Algorithm

Few-shot learning (FSL) aims at training models that can rapidly generalize to new tasks with only a limited number of new labeled samples. Given this, FSL is a good candidate to tackle the domain generalization problem in Wi-Fi sensing. For `BFA-Sense`, we adopt an FSL architecture named *embedding learning* [18]. Embedding learning is a technique used in machine learning to convert inputs into numerical vector representations in a way that preserves their characteristics in relation to the task. These representations, called embeddings, capture semantic similarities between input items. The process involves training a model to learn the embeddings by analyzing large amounts of data, and adjusting them so that similar items have closer numerical representations in the vector space. Thus, during the inference time, the embedding network does not need to be fine-tuned, and only a few labeled samples will be used as references for embedding mapping to classify unobserved data.

Formally, we consider the embedding network as a function $E_\theta : X \rightarrow Z$, where Z denotes the latent vector. The classifier $C_\phi : Z \rightarrow Y$ is to find a mapping between encoded features Z and labels Y. θ and ϕ are the trainable parameters of the embedding network and classifier, respectively. Hence, The overall system $F_\psi(X) = Y$ can be written as $C_\phi(E_\theta(X)) = Y$, where $\psi = \{\theta, \phi\}$ is the total trainable parameters of the whole system.

In the last decade, convolutional neural networks (CNNs) have achieved tremendous success in addressing computer vision tasks [19], [20]. The convolution layer, at the basis of CNNs, can efficiently extract relevant features from the input by

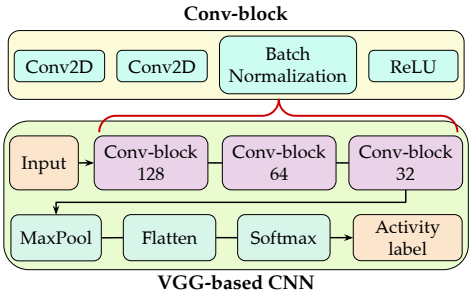


Fig. 3: BFA-Sense learning-based activity classifier.

performing convolution operations on the data through learnable kernels. Given that in this article we aim to investigate the effectiveness of BFAs-based sensing as compared to CSI-based sensing, we propose to use a VGG-based CNN architecture as the embedding learning classifier for human activities [20]. The network is depicted in Figure 3 and entails stacking three convolutional blocks (conv-block) and a max-pooling (MaxPool) layer. The Softmax activation function is applied to the flattened output to obtain the probability distribution over the activities. The three conv-blocks have 128, 64, and 32 filters, respectively. We choose a descending order of number of activation maps to reduce the model size. Moreover, features in lower layers are usually sparser and thus require extracting more activation maps to be properly captured.

IV. Performance Evaluation

A. Experimental Setup and Data Collection

We collected experimental data in three environments: a kitchen, a living room, and a classroom. We consider three human subjects and twenty different activities: *jogging, clapping, push forward, boxing, writing, brushing teeth, rotating, standing, eating, reading a book, waiving, walking, browsing phone, drinking, hands-up-down, phone call, side bends, check the wrist (watch), washing hands, and browsing laptop*. The activities are performed independently by each subject within a designated rectangular region in each of the three environments. Both BFAs and CSI data were collected for the same duration of 300 seconds for each of the twenty activities. To create the ground truth, we captured the synchronous video streams of the subjects performing the activities. The video streams are synchronized with the data to show what the subject is doing during the transmission of the NDP frame triggering the BFI computation and transmission.

MU-MIMO Setup and Equipment. We set up an IEEE 802.11ac MU-MIMO network operating on channel 153 with center frequency $f_c = 5.77$ GHz and 80 MHz of bandwidth. This allows sounding $K = 234$ OFDM sub-channels, i.e., 256 available sub-channels on 80 MHz channels minus 14 control sub-channels and 8 pilots. We use one AP (beamformer) and three STAs (beamformees), as depicted in Figure 4 in orange. The AP and the STAs were implemented through Netgear Nighthawk X4S AC2600 routers with $M = 3$ and $N = 1$ antennas enabled respectively for the AP and each of the STAs. The three STAs were served with $N_{ss} = 1$ spatial stream each and placed at three different heights and significantly spaced

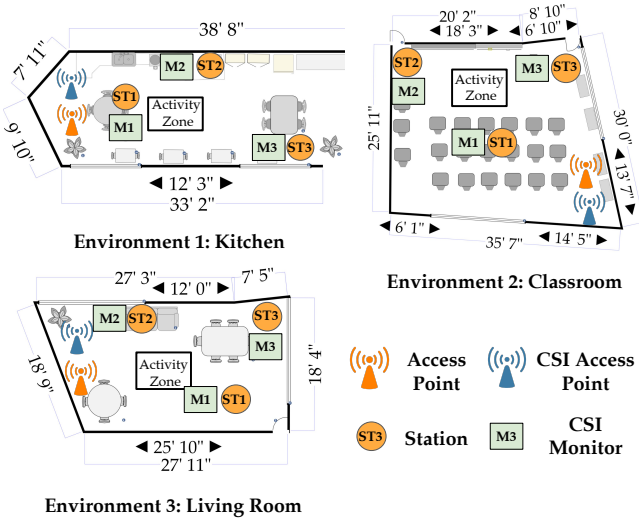


Fig. 4: Experimental setups for data collection.

from each other to form a 3×3 MU-MIMO system. According to the IEEE 802.11ac standard, four beamforming feedback angles (two ϕ and two ψ) are needed to represent each of the 3×1 channels between the AP and the STAs.

UDP data streams were sent from the AP to the STAs in the downlink direction to trigger the channel sounding. The BFAs frames were captured with the Wireshark network protocol analyzer running on an off-the-shelf laptop equipped with an Intel 9560NGW 802.11ac NIC set in monitor mode. Notice that the BFI frame-capturing device does not need any direct link with the AP or the STAs. Hence, the ϕ and the ψ angles were extracted from the captured frames for each of the STAs and used as input to the BFA-Sense learning framework. For decoding the BFAs we used the Wi-BFI tool in [12].

Network Setup and Equipment for CSI Collection. For comparative studies, CSI data has also been collected concurrently with the BFAs frame capture. For this purpose, a Wi-Fi network consisting of an AP (referred to as *CSI AP*) and three CSI monitors has been collocated with the STAs within the same environments, as depicted in Figure 4. The network operates on the IEEE 802.11ac channel 42, i.e., the center frequency is $f_c = 5.21$ GHz and the 80 MHz of bandwidth. The AP is implemented with a Netgear Nighthawk X4S AC2600 router while the CSI monitors are Asus RT-AC86U routers equipped with the Nexmon CSI extraction tool [5].

To have the same setup as in the MU-MIMO network, the CSI AP was enabled with $M = 3$ antennas whereas the CSI monitors were set up to sense the channel through $N = 1$ antenna over $N_{ss} = 1$ spatial stream each. UDP packets were broadcasted from the CSI AP to trigger the channel estimation on the three CSI monitors.

B. Performance Analysis

In the following, the results are obtained considering time windows of $W = 0.1$ s and $S = 10$ packets/sample.

1) Evaluation Varying the Number of OFDM Sub-Channel and Angle Resolution: It is known that Wi-Fi sensing performs worse when lowering the number of sub-channels considered

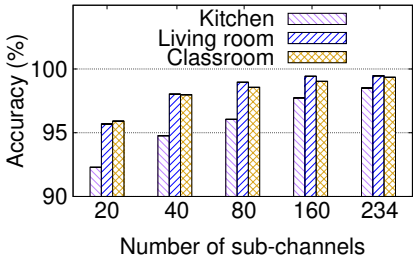


Fig. 5: BFA−Sense accuracy as a function of the no. sub-channels.

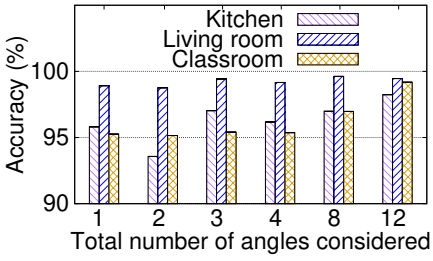


Fig. 6: BFA−Sense accuracy as a function of the number of BFAs considered.

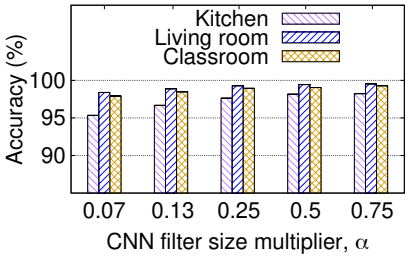


Fig. 7: BFA−Sense accuracy as a function of the number of filters in the CNN.

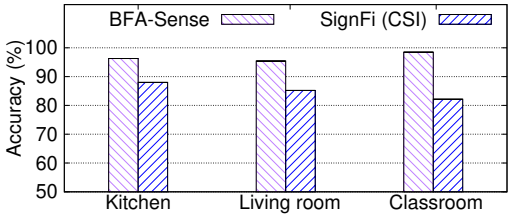
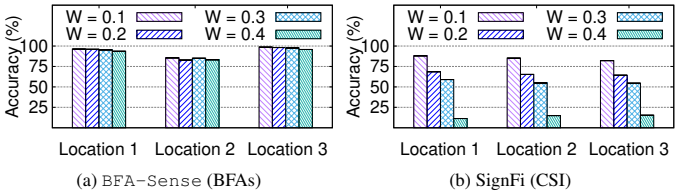


Fig. 8: BFA−Sense (BFI) vs SignFi (CSI) performance.



(a) BFA−Sense (BFAs) (b) SignFi (CSI)

Fig. 9: Performance comparison varying the window size W.

in the sensing process [2]. Higher sampling frequency or extensive feature extraction came at the cost of increasing the computational burden and intensifying pre-processing steps, as well as increasing the computational complexity of the learning process. For this reason, we investigate the impact of the number of sub-channels and angles on the sensing performance.

Figure 5 shows the accuracy of BFA−Sense as a function of the number of sub-channels utilized in the learning process. To down-sample the sub-channels, we take the first 20, 40, 80, and 160 sub-channels, to emulate sensing systems with smaller available bandwidths. The results show that the accuracy decreases by 6.31%, 3.80%, and 3.46% respectively for the kitchen, living room, and classroom when we switch from 234 to 20 sub-channels. However, notice that this operation drastically decreases the input tensor dimension from $10 \times 234 \times 12 = 28080$ to $10 \times 20 \times 12 = 2400$, implying that sub-channel resolution decreases the computational burden by 10 times while maintaining the accuracy above 92%.

Figure 6 shows BFA−Sense performance as a function of the number of BFAs considered for sensing. ST1 is considered for the results with up to four angles, ST1 and ST2 are considered for the combination of eight angles, and all three stations are considered for the combination of 12 angles. Figure 6 shows that the accuracy decreases by 1.98%, 0.16%, and 2.22% in the kitchen, living room, and classroom respectively when considering a single angle with respect to the combination of 12 angles. Even though the above results show no significant variation in performance even if the angle resolution is decreased from 12 angles combined to any individual angle, we suggest aggregating at least the angles of two spatially diverse STAs to obtain a robust algorithm.

2) Evaluation Varying the Number of Filters in the CNN: To further investigate the trade-off between computation complexity and accuracy, we use a multiplier $\alpha \in (0, 1]$ to reduce proportionally the number of filters (i.e., channels) at each layer of the CNN-based classifier in Figure 3. Specifically, applying the multiplier, the number of channels of each conv−block become $128\times\alpha$, $64\times\alpha$, and $32\times\alpha$, respectively. Figure 7 shows how the accuracy changes when varying α in $\{0.07, 0.13, 0.25, 0.5, 0.75\}$. BFA−Sense accuracy, averaged over the three environments, is 97.22%, 98.01%, 98.62%, 98.88%, and 99.02%, respectively. As the α multiplier decreases from 0.75 to 0.07, the accuracy drops marginally by 1.8%. Note that decreasing α the classifier computation complexity is reduced by α^2 roughly. Hence, the results indicate that BFA−Sense can adapt to limited computation resources and latency-sensitive cases by sacrificing little accuracy.

3) Comparison Between BFA−Sense and CSI-based Sensing: Figure 8 shows the classification accuracy of BFA−Sense compared with the state-of-the-art CSI-based SignFi algorithm [21] in the three different environments. We consider ST1, ST2, and ST3 as BFAs stations and the co-located M1, M2, and M3 as CSI monitors (see Figure 4). We evaluate the performance of BFA−Sense using the processing presented in Section III-A and CSI-based sensing using the processing presented in [21], adopting the CNN architecture in Figure 3 as the activity classifier in both the cases.

The accuracy of BFA−Sense in the kitchen, living room, and classroom is respectively 96.29%, 95.38%, and 98.47% whereas SignFi reaches 87.99%, 85.19%, and 82.14% of accuracy respectively, resulting in about 11% accuracy drop on average. This suggests that using the compressed BFAs, the activity classifier can perform well using few convolutional layers while CSI-based approaches would require deeper networks to extract representative features and achieve similar performance.

Figure 9 shows the performance of BFA−Sense and SignFi as a function of the CSI and BFAs capture location, and the window size W. We can see that for all the locations, BFA−Sense achieves better performance in comparison to SignFi following the same trend as presented in Figure 8. Moreover, in the case of SignFi, we also observe a significant performance variation when varying the window size W. For SignFi, with the variation of the window size, the accuracy varies by 39.75% on average, while the variability is only 1.36% for BFA−Sense.

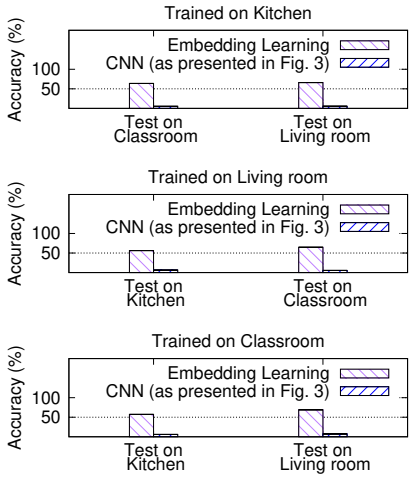


Fig. 10: BFA−Sense adaptation to new environments.

This suggests that feedback angles are a much more stable and reliable measurement than CSI.

C. Evaluation of Domain Adaptation

We compare the performance of embedding learning with baseline CNN based approach. Figure 10 shows that with only 15 s of data, embedding learning can adapt to new environments with an average accuracy of 64.94%, 60.52%, and 63.10% when trained in the kitchen, living room, and classroom respectively. CNN achieves 5.38%, 5.92%, and 6.62% respectively, which is 90.50% less than embedding learning.

V. Conclusions and Remarks

In this article, we have proposed BFA−Sense, a novel approach to Wi-Fi sensing based on MU-MIMO beamforming feedback angles (BFAs). Conversely from CSI-based approaches, (i) the BFAs can be easily recorded by off-the-shelf devices without MIMO capabilities and without any firmware modification; (ii) the BFAs allow a single device to concurrently capture all the channels between the AP and the different STAs, thus achieving a much higher sensing granularity. BFA−Sense includes an embedding learning-based classification algorithm to adapt to new environments with few additional data samples. We have evaluated BFA−Sense through an extensive data collection campaign involving three subjects performing twenty different activities in three indoor environments. The results show that compared to traditional CSI-based sensing approaches, BFA−Sense improves the accuracy by 11% on average, while the embedding learning approach achieves classification accuracy up to 65% in adapting to new environments. We hope that this work will pave the way for further research on BFAs-based sensing.

Acknowledgements

This work has been funded in part by the National Science Foundation under grants CNS-2134973, CNS-2120447, and ECCS-2229472, by the Air Force Office of Scientific Research under contract number FA9550-23-1-0261 and by the Office of Naval Research under award number N00014-23-1-2221.

References

[1] F. Meneghello, D. Garlisi, N. Dal Fabbro, I. Tinnirello, and M. Rossi, "SHARP: Environment and Person Independent Activity Recognition with Commodity IEEE 802.11 Access Points," *IEEE Transactions on Mobile Computing*, pp. 1–16, 2022.

[2] K. F. Haque, M. Zhang, and F. Restuccia, "SiMWiSense: Simultaneous Multi-Subject Activity Classification Through Wi-Fi Signals," in *2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pp. 46–55, 2023.

[3] F. Meneghello, C. Chen, C. Cordeiro, and F. Restuccia, "Toward Integrated Sensing and Communications in IEEE 802.11bf Wi-Fi Networks," *IEEE Communications Magazine*, vol. 61, no. 7, pp. 128–133, 2023.

[4] Y. Xie, Z. Li, and M. Li, "Precise Power Delay Profiling with Commodity Wi-Fi," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 2015.

[5] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free Your CSI: A Channel State Information Extraction Platform For Modern Wi-Fi Chipsets," in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, (New York, NY, USA), p. 21–28, Association for Computing Machinery, 2019.

[6] Z. Jiang, T. H. Luan, X. Ren, D. Lv, H. Hao, J. Wang, K. Zhao, W. Xi, Y. Xu, and R. Li, "Eliminating the Barriers: Demystifying Wi-Fi Baseband Design and Introducing the PicoScenes Wi-Fi Sensing Platform," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4476–4496, 2022.

[7] E. Aryafar, N. Anand, T. Salonidis, and E. W. Knightly, "Design and Experimental Evaluation of Multi-User Beamforming in Wireless LANs," in *Proc. of the 16th Annual International Conference on Mobile Computing and Networking (MobiCom)*, (New York, NY, USA), 2010.

[8] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1629–1645, 2019.

[9] C.-F. Hsieh, Y.-C. Chen, C.-Y. Hsieh, and M.-L. Ku, "Device-free indoor human activity recognition using Wi-Fi RSSI: machine learning approaches," in *2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*, pp. 1–2, IEEE, 2020.

[10] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool Release: Gathering 802.11n Traces with Channel State Information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, p. 53, 2011.

[11] F. Gringoli, M. Cominelli, A. Blanco, and J. Widmer, "AX-CSI: Enabling CSI Extraction on Commercial 802.11ax Wi-Fi Platforms," in *Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & CHaracterization*, (New York, NY, USA), p. 46–53, Association for Computing Machinery, 2022.

[12] K. F. Haque, F. Meneghello, and F. Restuccia, "Wi-BFI: Extracting the IEEE 802.11 Beamforming Feedback Information from Commercial Wi-Fi Devices," in *Proceedings of ACM WiNTECH*, (Madrid, Spain), 2023.

[13] Y. Jiang, X. Zhu, R. Du, Y. Lv, T. X. Han, D. X. Yang, Y. Zhang, Y. Li, and Y. Gong, "On the Design of Beamforming Feedback for Wi-Fi Sensing," *IEEE Wireless Communications Letters*, vol. 11, no. 10, pp. 2036–2040, 2022.

[14] S. Kondo, S. Itahara, K. Yamashita, K. Yamamoto, Y. Koda, T. Nishio, and A. Taya, "Bi-Directional Beamforming Feedback-Based Firmware-Agnostic WiFi Sensing: An Empirical Study," *IEEE Access*, vol. 10, pp. 36924–36934, 2022.

[15] T. Kanda, T. Sato, H. Awano, S. Kondo, and K. Yamamoto, "Respiratory Rate Estimation Based on WiFi Frame Capture," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*, pp. 881–884, 2022.

[16] E. Perahia and R. Stacey, *Next Generation Wireless LANs: Throughput, Robustness, and Reliability in 802.11n*. Cambridge Univ. Press, 2008.

[17] M. S. Gast, *802.11 ac: A Survival Guide: Wi-Fi at Gigabit and Beyond*. " O'Reilly Media, Inc.", 2013.

[18] Y. Tian, Y. Wang, D. Krishnan, J. B. Tenenbaum, and P. Isola, "Rethinking Few-shot Image Classification: A Good Embedding is All You Need?," in *Computer Vision – ECCV*, (Glasgow, UK), 2020.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[21] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "SignFi: Sign language recognition using WiFi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–21, 2018.