Enabling Ultra-Wideband Wi-Fi Sensing via Sparse Channel Sampling

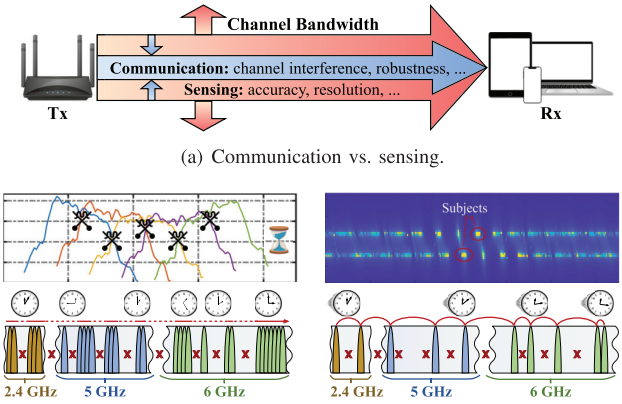
Xin Li, *Member, IEEE*, Jingzhi Hu, *Member, IEEE*, Hongbo Wang, *Graduate Student Member, IEEE*, Zhe Chen, *Member, IEEE*, and Jun Luo, *Fellow, IEEE*

***Abstract*—As a technology with ubiquitous presence in unlicensed spectrum, Wi-Fi has demonstrated prominent capabilities in both communication and sensing. However, since the bandwidth requirements for communication and sensing differ significantly, channel bandwidths excessive for communication (e.g., 160 MHz) still fail to achieve multi-person sensing. Though stitching multiple consecutive channels to expand the effective bandwidth sounds plausible, it may never reach *ultra-wideband* (UWB) in practice. To this end, we propose UWB-Fi as a novel Wi-Fi sensing framework with ultra-wide bandwidth, leveraging only discrete and irregular channel samples. We first design a fast channel hopping scheme to enable arbitrary channel sampling across 4.7 GHz bandwidth on commodity Wi-Fi hardware without interrupting default communications. As no algorithm exists to exploit such channel samples, we establish a theoretical analysis driven by *compressive sensing*, so as to enable an *explainable* deep learning model. This model transforms sparse channel samples into high-dimensional (position) spectra, effectively avoiding the *bias-variance dilemma* in parameter estimation while encoding sufficient information for general sensing. Our extensive evaluations demonstrate that UWB-Fi successfully achieves centimeter-level fine-granularity multi-person sensing.**

***Index Terms*—Wi-Fi human sensing, multi-person sensing, ultra-wideband sensing, Wi-Fi localization.**

I. Introduction

Wi-FI has gradually evolved into the most ubiquitous technology operating in unlicensed spectrum, thanks to its widespread low-cost infrastructure. With its prominent deployment scale and spectrum efficiency, Wi-Fi communication is widely applied in personal, commercial, industrial, and public services, thus becoming an indispensable part of modern life [2]. This ubiquity has in turn fueled the rapid growth of another prominent research direction: Wi-Fi's sensing capability has consistently been recognized as an essential area of development [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. Promoted by the availability of *channel state information* (CSI) [13], diversified sensing applications have been proposed over the past decade, including vital signs monitoring [8], [9], [10], gesture detection [4], [6], activity recognition [7], [11], [12], as well as localization and motion tracking [3], [5], [14].



(a) Communication vs. sensing.

(b) Continuous channel stitching.

(c) Sparse channel sampling.

Fig. 1. Communication and sensing have very different bandwidth requirements. To expand sensing bandwidth, continuous stitching is far from practical, but all its weaknesses can be overcome by sparse channel sampling.

Though promising on both sides, Wi-Fi's communication and sensing bear sharp difference in channel bandwidth requirement, as shown in Fig. 1(a). For Wi-Fi communication, the channel bandwidth is restricted from being too wide [15], [16], since excessive wide bandwidth incurs higher channel interference to the normal operation of other devices [15]. In addition, under the fixed transmission power (configured for energy efficiency) of access points (APs), excessively increasing bandwidth reduces the power per subcarrier, ultimately harming communication robustness [16]. Meanwhile, fine-granularity multi-person sensing demands a resolution of around 20 cm, considering the typical body size and inter-body distance in common indoor scenarios, which can only be offered by GHz-level bandwidth [17]. In other words, even the 320 MHz bandwidth promised by Wi-Fi 7 fails to provide sufficient spatial resolution (in both *range* and *bearing*). Although a specific device [18] or algorithm [9] can distinguish multiple people in individual applications, it cannot provide general support for fine-granularity sensing at the *physical* layer. Since GHz-level channel bandwidth has not been available to Wi-Fi, efforts was made to utilize

Received 5 November 2024; revised 14 April 2025; accepted 27 May 2025. Date of publication 30 June 2025; date of current version 6 November 2025. This work was supported in part by the National Research Foundation Singapore and the Infocomm Media Development Authority under its Future Communications Research & Development Programme and in part by MOE Tier 1 under Grant RG16/22. An earlier version of this paper was presented in part at ACM MobiSys'24 [DOI: 10.1145/3643832.3661889]. *(Corresponding author: Jun Luo.)*

Xin Li, Jingzhi Hu, Hongbo Wang, and Jun Luo are with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798 (e-mail: l.xin@ntu.edu.sg; jingzhi.hu518@gmail.com; hongbo001@ntu.edu.sg; junluo@ntu.edu.sg).

Zhe Chen is with the Institute of Space Internet, Fudan University, Shanghai 200437, China (e-mail: zhechen@fudan.edu.cn).

Digital Object Identifier 10.1109/JSAC.2025.3584510

multiple links [19], [20] or antennas [21], [22] to obtain physical *diversity*, yet they can hardly be extended beyond differentiating very few persons, significantly restricting their applicability.

Inspired by the Wi-Fi spectrum spanning 2.4 to 7.1 GHz, fully exploiting multiple channel resources to achieve *ultra-wideband* (UWB) sensing in a physical sense has emerged as a promising yet challenging research for the future of Wi-Fi. An attempted approach was *continuous channel stitching* [23], [24]: it combines multiple consecutive channels to form a wider bandwidth and indicates the theoretical performance upper bound due to the complete information. Unfortunately, this approach suffers from four major weaknesses as shown in Fig. 1(b), rendering it far from practical. Firstly, consecutive channels may not always be available for sensing since other devices might contend for or occupy some channels [25]. Secondly, it demands excessively overlapped channels to handle the boundary effect of frequency response, making the stitching process highly complex. Thirdly, the budget of channel coherence time is insufficient to scan a large number of overlapped channels needed by fine-granularity sensing. Finally, stitching approaches [23], [24] were only shown to be viable for localization purpose; it remains open if they can extend to fine-granularity general sensing.

Intuitively, one may potentially overcome all the aforementioned limitations by resorting to *sparse channel sampling*: as shown in Fig. 1(c), a wide bandwidth can be recovered from non-continuous and even irregular channel samples, totally avoiding the need for continuous bandwidth stitching.[1] This stems from the *compressive sensing* theory that sparse physical phenomena can be recovered with a small quantity of samples [27]. This strategy allows a free selection of arbitrary channels according to their availability in runtime, so it has a higher chance to gather sufficient information for "assembling" an effective spectrum for both fine-granularity and general-purpose sensing. Being sparse samples also naturally avoids the boundary effect of frequency response, since samples do not share common boundaries. Last but not least, sparse channel sampling potentially requires much fewer samples and thus can readily meet the channel coherence time budget, making accurate parameter estimation feasible.

Of course, this intuitive strategy faces its own challenges. Firstly, existing Wi-Fi CSI acquisition schemes are not able to achieve fast channel hopping, hampering the realization of sparse channel sampling. Secondly, a comprehensive theory to model general sensing based on sparse channel samples remains unestablished. Thirdly, unlike RFID [28] and radar [29] systems, each channel hopping in Wi-Fi may potentially introduce intractable random hardware-related offsets (in particular carrier phase offset or CPO) [30]; these offsets, along with various other interferences [31] in the specific Wi-Fi waveform, significantly degrade the accurate and coherent sparse recovery. Lastly, the factors that affecting the performance of sparse channel sampling remain largely unknown.

[1]Although Chronos [26] does not adopt a continuous bandwidth stitching strategy, its algorithm is limited to estimating the range between a pair of Wi-Fi devices, far from serving the general sensing purpose in our context.

To tackle these challenges, we propose UWB-Fi, a novel framework that leverages discrete and irregular channel sampling to expand Wi-Fi bandwidth and thus to achieve fine-granularity sensing. UWB-Fi first comprises a fast channel hopping scheme to perform arbitrary sampling within all Wi-Fi channels (2.4 GHz to 7.1 GHz) on commodity Wi-Fi devices (e.g., Intel AX210). As no existing signal processing algorithm can synthesize channel samples bearing random offsets into a coherent "snapshot", we innovate an explainable deep learning approach based on compressive sensing: it is guided by a ray-tracing model [32] to train a neural network, so as to map channel samples into high-dimensional spectral parameters rather than individual scalar values. As a result, our channel synthesizing pipeline successfully avoids the *bias-variance dilemma* in parameter estimation and removes offsets/interferences inherent to Wi-Fi, thanks to the diversity offered by ultra-wideband, while laying the foundation for general sensing. Finally, we implement a prototype of UWB-Fi using two laptops and conduct extensive experiments to evaluate its performance and influencing factors. In summary, our main contributions are:

- We propose UWB-Fi as the first Wi-Fi framework to achieve *physical UWB sensing* using only discrete and irregular channel samples across 4.7 GHz bandwidth.
- We design a fast channel hopping scheme for UWB-Fi to execute arbitrary channel sampling upon commodity Wi-Fi devices.
- We innovate an explainable deep learning approach based on compressive sensing to synthesize sparse channel samples bearing unknown offsets and interferences.
- We implement the first UWB sensing prototype on commodity Wi-Fi hardware, and demonstrate UWB-Fi's GHz-level effective sensing bandwidth and general sensing capabilities through extensive evaluations.

The rest of our paper is structured as follows. Section II introduces the theoretical basis and motivation of UWB-Fi. Section III elaborates on the framework design of UWB-Fi. Sections IV and V respectively explain UWB-Fi's implementation and report the extensive evaluations on UWB-Fi, followed by the conclusion of our paper in Section VI.

II. Theoretical Basis and Motivation

In this section, we first establish Wi-Fi sensing basics and analyze the impact of bandwidth on parameter estimation. We then explain why existing channel stitching approaches cannot realize fine-granularity sensing. Moreover, we point out that sparse channel sampling guided by compressive sensing theory is a promising method. Finally, we analyze the incapability of the conventional approach for sparse recovery and discuss its potential to inspire a neural network.

A. Wi-Fi Sensing Basics

Assuming a Wi-Fi sensing system with a Tx-Rx pair and multiple sensing subjects, we start by introducing a ray-tracing model to characterize a "snapshot" of sensing results; it serves as the basis for all sensing tasks, as any task is simply a concatenation of several such snapshots. The snapshot involves a

set of (range, bearing) tuples as $\{(\tau_0, \theta_0), \cdots, (\tau_k, \theta_k), \cdots\}$, where k indexes a subject (hence the introduced reflection path), while τ and θ derived by Wi-Fi sensing are actually the *time of flight* (ToF) and *angle of arrival* (AoA). Let the Rx be equipped with uniformly spaced antennas of interval d, the received CSI $[\boldsymbol{H}]_{n,m} = h_{n,m}$ can be modeled as follows:

$$\begin{aligned} h_{n,m} &= \sum_{k=1}^{K} \alpha_{n,m,k} \cdot h_{m,k}^{\mathrm{ToF}} \cdot h_{n,k}^{\mathrm{AoA}} + \epsilon \\ &= \sum_{k=1}^{K} \alpha_{n,m,k} \mathrm{e}^{-\mathrm{i}2\pi(f_{\mathrm{c}} \pm (m-1)f_{\mathrm{b}})\tau_k} \mathrm{e}^{-\mathrm{i}2\pi(n-1)d\cos(\theta_k)\frac{f_{\mathrm{c}}}{\mathrm{c}}} \\ &\quad + \epsilon, \end{aligned} \tag{1}$$

where n and m respectively index the antenna and subcarrier, α represents channel gain, f_{c} and f_{b} respectively denote channel centre frequency and subcarrier bandwidth, ϵ indicates noise, and c is the speed of light.

According to [17], the temporal resolution $\Delta\tau = 1/B$, where B denotes the total sensing bandwidth, grows linearly with B, so the *range resolution* of a snapshot, derived from the ToF term in Eqn. (1), benefits from a wider bandwidth B. Though the *bearing* (AoA) *resolution* is contingent upon antenna quantity, the estimation of AoA may still benefit from a wider B, because improving the resolution of $d\cos(\theta_k)/\mathrm{c}$ (also a temporal component) yields a higher precision in estimating AoA. These observations allow us to conclude that **the total sensing bandwidth holds pivotal significance in realizing fine-granularity sensing.**

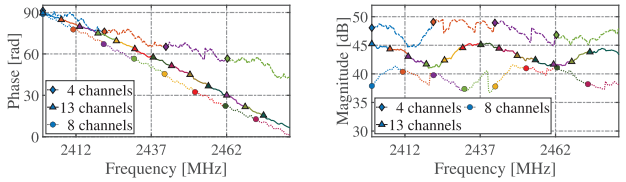
In sensing, ToF and AoA can generally be derived from the received CSI $\boldsymbol{H}$ by leveraging an estimation denoted by:

$$\boldsymbol{\Xi}(\tau, \theta) = \mathcal{G}(\boldsymbol{H}), \tag{2}$$

where $\mathcal{G}$ is a mapping operation that can be implemented by various algorithms, in which we adopt the classical MUltiple SIgnal Classification (MUSIC) [33] in this work. By constructing a matrix $\boldsymbol{E}$ from the eigenvectors of $\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}$ that correspond to eigenvalues smaller than a threshold, the ToF-AoA spectrum is $\boldsymbol{\Xi}(\tau, \theta) = \frac{1}{v^{\mathsf{H}}(\tau,\theta)EE^{\mathsf{H}}v(\tau,\theta)}$, where superscript H denotes the Hermite transpose of the matrix. Since $\boldsymbol{E}$ represents the noise vector subspace orthogonal to the signal vector subspace, the subject information will appear as peaks in $\boldsymbol{\Xi}(\tau, \theta)$ [31].

B. Infeasibility of Stitching Methods

To expand bandwidth, Splicer [23] and ToneTrack [24] conduct channel stitching to merge continuous spectra. However, the boundary effect (i.e., drastic roll-off) of the frequency response makes it impossible to stitch neighboring channels while preserving continuous phases and magnitudes. This has forced [23], [24] to use channels substantially overlapping with each other and sophisticated algorithms to align/calibrate these spectra. To demonstrate it, we perform continuous channel scanning in the 2.4 GHz band, where standard channels overlap [34]. By initially stitching 4 channel samples spanning 80 MHz, we observe significant fluctuations in both phase and magnitude, as depicted in Fig. 2. After noticing persistent fluctuations with 8 channel samples, we gradually increase the number to 13, at which point the stitched phase and magnitude



(a) Phase. (b) Magnitude.

Fig. 2. Results of continuous channel stitching.

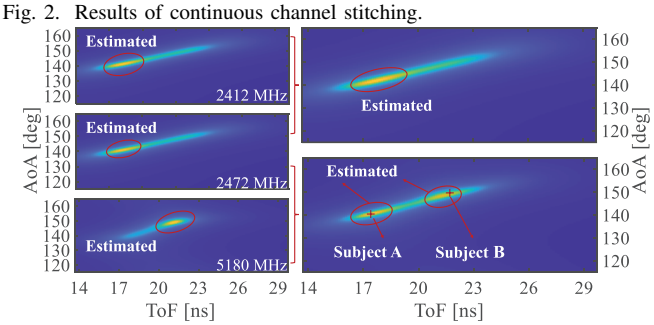


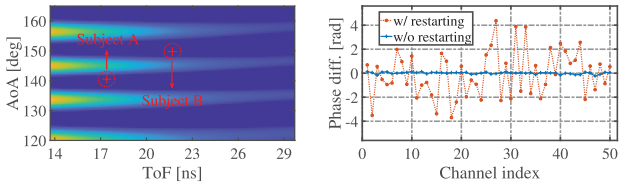
Fig. 3. Estimation results of sparse channel samples. Left: ToF-AoA spectra of channel samples with different centre frequencies; Right: the synthesized ToF-AoA spectra.

finally become smooth. Consequently, as we shall demonstrate soon, this approach is confined by two-time budget constraints, namely coherence time and real-time processing, rendering it inapplicable to real-life applications. To put the illustration into our perspective, we consider a spectrum with a bandwidth of 4.7 GHz, ranging from 2.4 GHz to 7.1 GHz, stitched from multiple overlapped 20 MHz channels shifted by 5 MHz, resulting in a total of 936 channels.

Technically, all the channel samples should be obtained within *coherence time* to guarantee effective synthesizing. According to [35], [36], and [37], the coherence time is approximately 100 ms for the Wi-Fi spectrum. Meanwhile, sampling a single channel empirically demands at least 2 ms [26]. Thus, at most $100/2 = 50$ overlapped channel samples can be captured, significantly short of the necessary amount for stitching a wide bandwidth. In addition, the alignment process introduces a substantial cost in computation due to the need for a large number of overlapped channels, rendering the feasibility of *real-time sensing* questionable. According to our experiments, each stitching performed in MATLAB takes approximately 13 ms on a 12th Gen Intel(R) Core(TM) i5-12500H processor in a Lenovo Thinkbook 14 laptop. Extrapolating this result to the stitching of 936 channels leads to an excessive duration of around 12 s. Thus, it is imperative to propose novel approaches to extend bandwidth towards UWB range in practice.

C. Compressive Sensing-Inspired Channel Sampling

Fortunately, since the physical phenomena captured by each snapshot are often sparse in nature [38], [39], it is feasible to apply the compressive sensing framework [27]. To demonstrate the sparsity, in terms of information in wideband signals, related to sensing (range) resolution, we simulate three ideal 20 MHz bandwidth channel samples distributed across the 2.4 GHz and 5 GHz based on Eqn. (1) to distinguish two subjects separated by 0.88 m, with their ToF-AoA spectra



(a) ToF-AoA spectrum. (b) Phase difference.

Fig. 4. Analysis of real-life channel samples illustrates that additional interference from channel hopping impedes parameter estimation.

illustrated in the left panels of Fig. 3. Due to insufficient resolution indicated by $\Delta\tau = 1/B$, none of these three individual channel samples can effectively differentiate the two subjects. Meanwhile, we observe that each spectrum shows different ambiguity in the subjects' information due to the estimated area varying with f_c, which inspires us to synthesize their spectra to gain higher resolution. As shown in the right panels of Fig. 3, combining two spectra from the 2.4 GHz band still fails to separate the two subjects, whereas incorporating two spectra from both the 2.4 GHz and 5 GHz bands succeeds, evidently demonstrating the sparsity: not all information across 2.7 GHz bandwidth is needed to achieve an adequate resolution. Moreover, this comparison highlights that high resolution can be acheived with sparse samples only if the samples are largely independent from each other.

Intuitively speaking, expanding the sensing bandwidth B can be interpreted as increasing channel samples, offering greater information diversity for estimation algorithms to derive both range and bearing. Assuming a wide-bandwidth signal composed of a large number of continuous 20 MHz channel samples, $\mathcal{Y} = \{\mathcal{Y}_1, \ldots, \mathcal{Y}_Q\}$, it can be expressed as $\mathcal{Y} = \Psi \boldsymbol{Y}$ using a $Q \times Q$ sparse basis matrix Ψ and a sparse representation $\boldsymbol{Y}$ containing complete spectrum information due to sparsity. Assuming a $M \times Q$ $(M < Q)$ measurement matrix Φ that $\boldsymbol{H} = \Phi\mathcal{Y}$, the compressive sensing framework is:

$$\boldsymbol{H} = \Phi\Psi\boldsymbol{Y} = \Theta\boldsymbol{Y}, \tag{3}$$

where Θ is the sensing matrix. The principle of compressive sensing [27] indicates that we can effectively acquire the fine-granularity information contained in $\boldsymbol{Y}$ across B by sampling a discrete and low-dimensional channel measurement $\boldsymbol{H}$ (even the subcarriers within them), rather than sampling every channel within B.

D. Sparse Recovery for Sensing

Compressive sensing lays the theoretical foundation for fine-granularity sensing with sparse channel samples. However, despite the theoretical feasibility being validated, attempting to estimate ToF and AoA using real-life channel samples, as shown in Fig. 4(a), does not yield meaningful subject-related information. Apart from common noise and channel fading, the reason for this failure is the unknown parameters introduced by the sparse sampling. In particular, the CPO may randomly vary for every channel hopping due to the NIC restarting. To illustrate this issue, we control Wi-Fi NICs to collect samples from the same channel within the coherence time under two conditions: restarting the NICs during every sampling and normal sampling without restarting. The phase differences between the current and the previous CSIs sampled across 50 channels within the 6 GHz band are shown in Fig. 4(b). The results indicate that, compared with normal sampling, the phase differences under the restarting exhibit significant random fluctuations. With multiple restarts during channel hopping, it is not a surprise that the accumulated randomness can totally overwhelm the information embedded in these samples. Consequently, conventional signal processing approaches are unable to derive reliable estimation outcomes of multiple reflection paths from sparse channel samples.

Since conventional tools can hardly characterize all such interferences (some rather random) to recover ToF-AoA parameter tuples from sparse channel samples, one may think that a well-trained neural network can be a competent alternative. However, such a direct recovery can be far from effective due to the bias-variance dilemma [40] (to be proven in Sections III-B and V-D) and the absence of general sensing information. Inspired by the spectra shown in Fig. 3, we consider a composite function $\tilde{\mathcal{G}}(\tilde{\Theta}^{\dagger}(\cdot)) : \boldsymbol{H} \rightarrow \boldsymbol{\Xi}$ mapping real-life channel samples $\boldsymbol{H}$ to spectrum $\boldsymbol{\Xi}$ that imply ToF-AoA parameter tuples. Nevertheless, this recovery remains rather complex, involving not only the filtering of interferences but also the estimation of spectra with general sensing capabilities; comprehensively considering all factors to derive an explicit expression by conventional approaches is extremely challenging, if not entirely impossible. Thus, we invoke the universal approximation theorem [41] to approximate the well-founded $\tilde{\mathcal{G}}(\tilde{\Theta}^{\dagger}(\cdot))$ with a neural network $\mathbf{g}$. We shall discuss how $\mathbf{g}$ is established and trained in Sections III-B and III-C.

III. UWB-Fi Framework Design

Our UWB-Fi is designed to achieve fast sparse channel sampling for meeting the coherence time budget and to recover informative spectra from sparse channel samples. As illustrated in Fig. 5, UWB-Fi consists of three major components:

- Fast channel hopping scheme: It involves a high-level strategy to reliably manage channel hopping and a low-level scheme to efficiently access CSIs.
- Sparse recovery neural network: An encoder-decoder structure is adopted to emulate both a spectrum filter and the recovery of one spectral domain to another.
- Model-driven training strategy: Instead of using limited CSIs and scalar tuples as the training set, this strategy takes model-augmented data with enriched information.

We elaborate on these components in Sections III-A to III-C, and explain how to produce fine-granularity sensing results by synthesizing the neural-inferred snapshots in Section III-D.

A. Fast Channel Hopping Scheme

In compressive sensing, the initial task is to achieve sparse channel sampling, which needs to be fast enough to fulfill channel coherence time requirements. Due to the absence of available tools, we develop a fast channel hopping scheme on commercial Wi-Fi NICs with a focus on practicality.

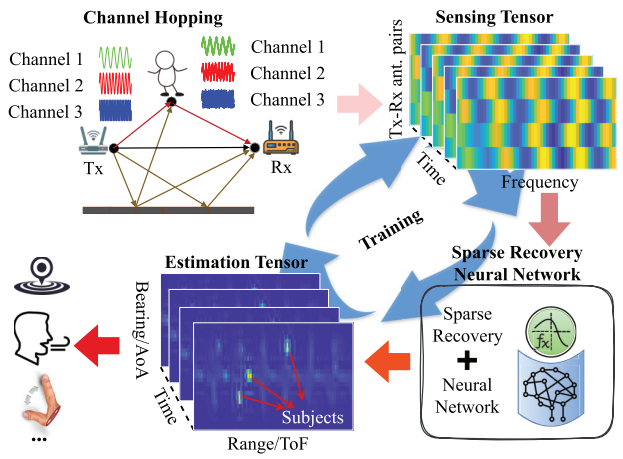


Fig. 5. UWB-Fi framework overview.

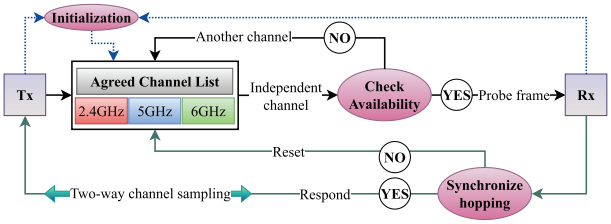
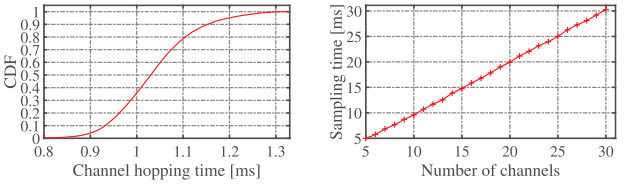


Fig. 6. Fast channel hopping scheme.

1) High-Level Control Strategy: This strategy comprises four essential stages: channel selection, synchronized hopping, data acquisition, and information extraction, as depicted in Fig. 6. We have in total 97 Wi-Fi channels of 20 MHz bandwidth as our hopping candidates across the 2.4 GHz, 5 GHz, and 6 GHz bands. Each hopping is initiated by the Tx that first checks the availability of a randomly selected channel and sends a probe frame to the Rx in the current channel so as to synchronize this hopping. Upon a successful channel hopping, the Rx promptly responds to the Tx to enable a two-way channel sampling, which then brings the Tx to select the next channel to hop; here the choice has to abide by the principle of being different from all previously visited channels. If the Rx fails to respond within the allocated time, the Tx logs a failure; in the case of persistent failures of hopping to unavailable channels, the Tx simply reverts back to the 2412 MHz "rendezvous" channel to reset. User-specified channel sampling sequences are also allowed.

Since normal Wi-Fi sensing uses only one frame to achieve a narrowband snapshot while carrying data traffic within the frame, UWB-Fi may have the same data payload segmented into all hopping frames to avoid interrupting default Wi-Fi communications. Moreover, the proportion of frames used for sensing and their transmission timing can be adjusted according to different applications, further reducing the impact on communication efficiency. In typical indoor environments with 3-5 dominant multipath components [42], we empirically set the total channel samples as 20 (also see Section V-C.3). For complex scenarios, we can increase the number of sampled channels or the sampling rate to accumulate more snapshots within the coherence time, effectively obtaining more samples.



(a) One channel hopping time. (b) Multiple channels hopping time.

Fig. 7. Channel hopping time of UWB-Fi.

2) Low-Level Access Scheme: To accelerate channel hopping (from at least 2 ms [26] down to around 1 ms), we need to decrease the latency of data acquisition caused by interactions between *user* and *kernel* spaces. We firstly leverage `debugfs`[43] to connect kernel and user spaces without kernel access constraints (e.g., `sysfs`[44] has strict one-value-per-file rules). To further speed accessing kernel space, we create two threads: `para_set`, and `data_proc`, respectively used to set hardware parameters (e.g., hopping channels) and to quickly move channel data from kernel to user space. We adopt an inter-process communication method (e.g., POSIX shared memory) to achieve fast data exchange among these two threads and upper-layer computation modules. Although setting hardware parameters may require resetting the Wi-Fi hardware, the drivers of certain hardware [45] offer direct register access and can avoid the latency caused by resetting.

To assess the low-level access scheme, Fig. 7(a) displays the cumulative distribution of the time required for one channel hopping, showing rapid and consistent performance, with an average duration of around 1 ms. Additionally, we measure the total mean time needed to sample a various numbers of channels; as depicted in Fig. 7(b), it linearly grows with the number of channels. In particular, collecting 20 channel samples (see Section III-A.1) requires only 20 ms. Given a coherence time of 100 ms [35], [36], [37], this 20 ms duration comfortably adheres to the coherence time budget. Furthermore, considering the data processing time (detailed in Section III-D) and other possible delays, the total overhead of UWB-Fi remains below 100 ms. These results underscore UWB-Fi's capability to conduct real-time sensing tasks under most circumstances.

B. Sparse Recovery Neural Network

The essential task for UWB-Fi to realize fine-granularity sensing, similar to UWB radar sensing [46], is to sharply identify multiple subjects (in terms of their respective multipath components) as clearly separated *bin*s in each snapshot. In this section, we will design the sparse recovery neural network following this purpose. As our main contribution lies in the design of an explainable model and training strategy, we select modules for approximation based on our experience in [47]. We shall elaborate on the model design in the following, but refrain from evaluating it against other possible choices later.

1) Overview of Sparse Recovery Network: As described in Section II-D, directly mapping channel samples into ToF-AoA tuples can be far from effective; even if effective, it is confined to pure localization purpose and fails to deliver

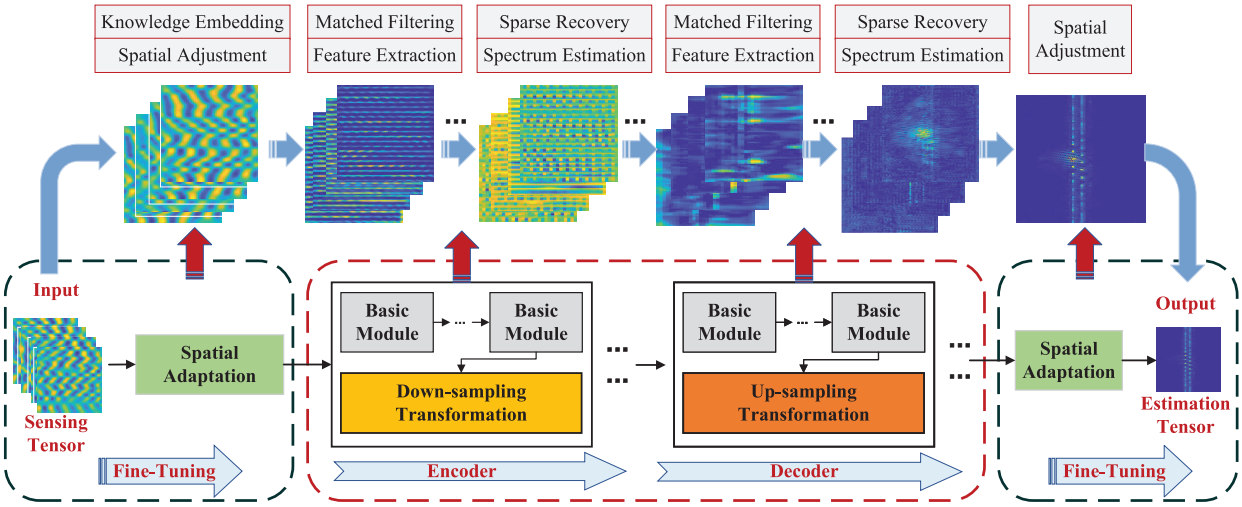


Fig. 8. SpecTrans: Spectral transformation across two spectral domains for sparse recovery.

CSI phase information in the form of bins for further fine-granularity general sensing. We prove in following Corollary 1 and Proposition 1 that, as a recovery target, the well-crafted spectrum can improve estimation performance and encode information for general sensing, respectively. Thus, we employ a neural network $\mathbf{g}$ to recover ToF-AoA spectra in the real domain from sparse channel samples in the CSI domain.

Corollary 1: Under the constraints of the bias-variance dilemma, the spectral parameter structure achieves a substantially lower MSE than the tuple parameter as the output of an estimation algorithm.

Proof: Assuming $\boldsymbol{\Xi} = \mathbf{g}(\boldsymbol{H}) + \epsilon$, where ϵ is the noise. The trained neural network is $\hat{\mathbf{g}}$, then the MSE of it is:

$$\mathbf{MSE} = \underbrace{\left(\mathbb{E}[\hat{\mathbf{g}}(\boldsymbol{H})] - \mathbf{g}(\boldsymbol{H})\right)^2}_{\mathbf{Bias}} + \underbrace{\mathbb{E}[(\hat{\mathbf{g}}(\boldsymbol{H}) - \mathbb{E}[\hat{\mathbf{g}}(\boldsymbol{H})])^2]}_{\mathbf{Variance}} + \zeta^2, \tag{4}$$

where $\zeta^2 = \mathbb{E}[(\boldsymbol{\Xi} - \mathbf{g}(\boldsymbol{H}))^2]$ is unrelated to $\hat{\mathbf{g}}$ and considered as a constant. The overall **Bias** is distributed across all spectral parameters and tends to be offloaded to non-peak regions due to its global continuity, while the local strong convexity near spectral peaks attenuates **Bias** in their estimated positions [48]. Meanwhile, as a high-dimensional tensor representation, the spectrum inherently imposes both implicit regularization and smoothness constraints, effectively suppressing **Variance** [49]. Therefore, the spectrum structure contributes to ToF-AoA estimation in both low **Bias** and **Variance**. ■

Proposition 1: The amplitude parameters of the ToF-AoA spectrum can encode the phase information of CSI, enabling fine-granularity general sensing.

Proof: The proposition is equivalent to perturbations in the CSI phase leading to amplitude fluctuations of the ToF-AoA spectrum, which can be simplified to impact the EE^{H} term w.r.t. Eqn. (2). Assuming the phase perturbation in CSI is $\delta = [e^{\mathrm{i}\Delta\omega_{n,m}}]$, then $\tilde{\boldsymbol{H}} = \boldsymbol{H} \odot \delta$, where $\odot$ represents Hadamard product. Given that $e^{\mathrm{i}\Delta\omega_{n,m}} \approx 1 + \mathrm{i}\Delta\omega_{n,m}$ and $\Delta\boldsymbol{H} = \boldsymbol{H} \odot [\mathrm{i}\Delta\omega_{n,m}]$, then $\tilde{\boldsymbol{H}} \approx \boldsymbol{H} + \Delta\boldsymbol{H}$. With $\boldsymbol{\mathcal{H}} = \boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}$ and $\tilde{\boldsymbol{\mathcal{H}}} = \tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^{\mathsf{H}}$, the difference between them is $\Delta\boldsymbol{\mathcal{H}} \approx \boldsymbol{H}\Delta\boldsymbol{H}^{\mathsf{H}} + \Delta\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}$. By constructing an orthogonal projection matrix $\boldsymbol{P}_j = \mathbf{I} - \mathcal{E}_j\mathcal{E}_j^{\mathsf{H}}$, the perturbation in the eigenvector is:

$$\Delta\mathcal{E}_j \approx \sum_{i \neq j} \frac{\mathcal{E}_i^{\mathcal{H}} \Delta\boldsymbol{\mathcal{H}} \mathcal{E}_j}{\lambda_j - \lambda_i} \mathcal{E}_i, \tag{5}$$

where λ_j and $\mathcal{E}_j$ are the j-th eigenvalue and eigenvector of $\boldsymbol{\mathcal{H}}$, respectively. The term EE^{H} will introduce the fluctuation $E\Delta E^{\mathsf{H}} + \Delta E E^{\mathsf{H}}$ with $\Delta E = [\Delta\mathcal{E}_j]$. ■

To handle the sparse recovery, we design a SpecTrans network $\mathbf{g}$ following an *encoder-decoder* (ED) architecture [50], with the fully convolutional backbone structure and trainable attention layers adopted for efficient implementation. We also note that SpecTrans in the sparse recovery framework incorporates two model-based algorithms: i) a matched filter to remove hardware-related interferences/offsets and ii) the estimation algorithm to transform sparse channel samples to ToF-AoA spectra. Network modules share similar algorithmic nature with those involved in the two model-based algorithms but their weights are variables to be trained. Based on Fig. 8, we explain each block of SpecTrans in the following.

2) Basic Module: This module acts both as a trainable matched filter to remove hardware-related interferences/offsets from sparse channel samples and as a feature extractor. Assuming the received CSI is $\boldsymbol{H} = \boldsymbol{H}_{\mathbf{s}} + \boldsymbol{H}_{\mathbf{n}}$, the matched filter applied to $\boldsymbol{H}$ is determined as $\boldsymbol{V} = \arg\max_{\boldsymbol{V}} |\boldsymbol{V} * \boldsymbol{H}_{\mathbf{s}}|^2 / \mathbb{E}|\boldsymbol{V} * \boldsymbol{H}_{\mathbf{n}}|^2$, where $\boldsymbol{H}_{\mathbf{s}}$ and $\boldsymbol{H}_{\mathbf{n}}$ are the signal and noise components, respectively. This can be efficiently implemented using convolutional layers, for which we adopt DenseNet [51] and denote it by $\mathbf{d}(\cdot)$. To further extract features, an SKNet [52], denoted as $\mathbf{s}(\cdot)$ and having an attention-intensive structure, is selected for precise estimation. Thus, the basic module is built by:

$$x_\ell^{\text{out}} = [x_\ell^{\text{in}}, \mathbf{s}_\ell(\mathbf{d}_\ell(x_\ell^{\text{in}})], \tag{6}$$

where x_ℓ^{in} and x_ℓ^{out} indicate the input and output of the ℓ-th layer, respectively. As shown in Eqn. (6), modules $\mathbf{s}$ and $\mathbf{d}$ are applied to produce part of the layer output, and their results are concatenated with the layer input, ensuring the concentration of critical channel information (for computational efficiency),

while maintaining the availability of model gradient. Additionally, to avoid variance shift [53], we specifically use batch normalization for regularization. Finally, the basic module keeps the size of the feature map unchanged to better control over the sparse recovery.

3) Cross-Domain Transformation: Given the inherent spatial transformations, this module fits the core sparse recovery and spectrum estimation processes. Based on Eqn. (3) and the compressive sensing techniques, to recover $\boldsymbol{Y}$ from $\boldsymbol{H}$, L1 relaxation and the Lagrange multiplier λ_{m} [54] are adopted to formulate an unconstrained problem:

$$\hat{\boldsymbol{Y}} = \arg\min_{\boldsymbol{Y}} \mu||\boldsymbol{Y}||_1 + \frac{\lambda_{\mathrm{m}}}{2}||\boldsymbol{H} - \Theta\boldsymbol{Y}||_2^2, \tag{7}$$

where μ is a soft threshold. As the L1 norm is not completely differentiable, proximal operator [55] is introduced to update:

$$\hat{\boldsymbol{Y}}^{q+1} = \mathbf{Prox}_{\mu||\boldsymbol{Y}^q||_1}\left(\hat{\boldsymbol{Y}}^q - \eta\lambda_{\mathrm{m}}\Theta^{\mathsf{T}}\Theta\hat{\boldsymbol{Y}}^q + \eta\lambda_{\mathrm{m}}\Theta^{\mathsf{T}}\boldsymbol{H}\right), \tag{8}$$

where $\mathbf{Prox}_{\mu||\cdot||_1}(x) = \mathrm{sign}(x)\max(|x| - \mu, 0)$ and is abbreviated as $\mathbf{Prox}$ hereafter and η is the step-size. Network modules are initially built to emulate the iterations of Eqn. (8):

$$\hat{\boldsymbol{Y}}^{q+1} = \mathbf{Prox}\left((\mathbf{I} - \eta\lambda_{\mathrm{m}}\boldsymbol{W}^{\mathsf{T}}\boldsymbol{W})\hat{\boldsymbol{Y}}^q + \eta\lambda_{\mathrm{m}}\boldsymbol{W}^{\mathsf{T}}\boldsymbol{H}\right), \tag{9}$$

where $\boldsymbol{W}$ is the weight matrix, and parameters $\{\boldsymbol{W}, \lambda_{\mathrm{m}}, \eta\}$ are all trainable. However, fully recovering continuous channel samples incurs high complexity and is unnecessary. Thus, we divide Eqn. (9) into two steps to further refine it with diverse weight matrices, along with a nonlinear function to handle various recovery targets (e.g., ToF-AoA spectrum). The first and second steps are refined as $\hat{\boldsymbol{Y}}_0^{q+1} = \mathrm{ReLU}\left(\boldsymbol{W}_1 * \hat{\boldsymbol{Y}}^q\right)$ and $\hat{\boldsymbol{Y}}^{q+1} = \mathbf{Prox}\left(\hat{\boldsymbol{Y}}_0^{q+1} + \boldsymbol{W}_2 * \boldsymbol{H}\right)$ respectively, where $\boldsymbol{W}_{1,2}$ and * denote the convolution kernels and the convolution operation. Thus, the cross-domain transformation module is stacked leveraging a two-layer structure to facilitate the sparse recovery and spectrum estimation:

$$x_\ell^{\mathrm{out}} = \mathbf{Prox}\left(\mathrm{ReLU}(\boldsymbol{W}_1 * x_\ell^{\mathrm{in}}) + \boldsymbol{W}_2 * x_0^{\mathrm{in}}\right). \tag{10}$$

Specifically, we utilize a fully convolutional structure, with strided convolutions as the *encoder* for down-sampling transformation and transposed convolutions as the *decoder* for up-sampling transformation. As the number of feature maps increases in basic modules, we avoid dense connections in transformation modules to reduce the quantity of feature maps.

4) Spatial Adaptation: Although UWB-Fi is largely immune to input variations across different environments thanks to its UWB nature (confirmed in Section V-C.2), we consider the worst case scenarios where cross-environment variations become so prominent that they overwhelm the signals in the estimation results. Thus, we introduce *spatial adaptation* (SA) in the fine-tuning stage after the completion of training the core ED network; it is achieved through convolutional layers plugged at the points of entry and egress of the core network. Fine-tuning the core network without altering its structure by spatial adaptation allows our overall design to gain sufficient versatility for adapting to diverse application scenarios.

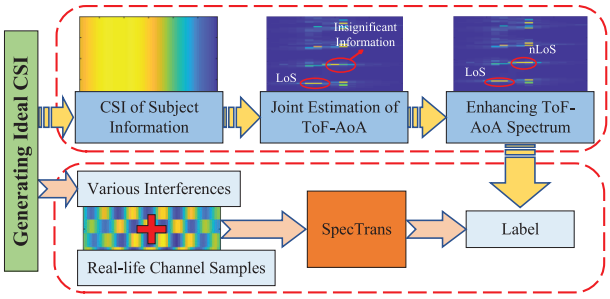


Fig. 9. Data augmentation training strategy.

C. Training With Data Augmentation

As our SpecTrans aims to effectively map sparse channel samples to the ToF-AoA spectra, it has to be trained with the same type of data: in particular, the ground truth labels have to be in the form of ToF-AoA spectra (as shown in Fig. 3), rather than ToF-AoA tuples $\{(\tau_0, \theta_0), \cdots, (\tau_k, \theta_k), \cdots\}$ that we obtained during the data collection process. However, without a well-trained SpecTrans, we cannot derive the ToF-AoA spectra from real-life channel samples using traditional methods (as already explained in Section II-D). Therefore, what we need is a procedure to convert the known ToF-AoA tuples into the corresponding spectra that may potentially yield these tuples as the estimation results. Firstly, we generate ideal CSI data with different centre frequencies based on Eqn. (1). Next, we perform joint estimation of ToF-AoA spectrum for each ideal channel sample and then synthesize the spectra. The synthesized spectrum is further enhanced through exponential lifting and Laplacian filtering [56] to obtain the final label. Moreover, training data should be endowed with diversified background multipath interference to train SpecTrans towards competent generalizability across distinct/unseen environments. To finally train the core ED network, we adopt the following objective function with an L2 penalty term on network weights:

$$\min_{\boldsymbol{W}}\left(\sum(\mathbf{g}_{\boldsymbol{W}}(\boldsymbol{H}) - \boldsymbol{\Xi})^2 + 0.5\lambda_{\mathrm{p}}\sum\boldsymbol{W}^2\right), \tag{11}$$

where, $\mathbf{g}_{\boldsymbol{W}}(\cdot)$ represents the core network with weight $\boldsymbol{W}$, $\boldsymbol{\Xi}$ denotes the spectrum label, $\boldsymbol{H}$ refers to the input sparse channel samples, and λ_{p} is the coefficient of L2 penalty. Building on this, the data augmentation training strategy is shown in Fig. 9, and the algorithms are detailed as follows.

1) Background Interference Generation: The cross-environment of Wi-Fi sensing system is typically addressed by training with large datasets, which comes with high costs and often yields unsatisfactory performance in unseen environments. Besides facilitating precise subject information capture through high spatial resolution, UWB-Fi offers an efficient solution by diversifying the training dataset with model-generated interferences. We analyze various types of background interference in typical indoor environments and estimate their reflection paths [57]. They are then transformed into CSI samples based on Eqn. (1) and integrated with the real-life samples as a diverse dataset. The details are illustrated in Algorithm 1.

2) Model-Driven Label Generation: Since we can generate channel samples by plugging the known ToF-AoA tuples into

Algorithm 1 Diversification of Background Interference

Input: Real-life dataset $\boldsymbol{H}_{\mathrm{R}}$; potential background interferences $\mathcal{T}_{\mathrm{B}}$; n, d, m, f_{s} as in Eqn. (1).
Output: Dataset with diverse interferences $\boldsymbol{H}$.

Initialization: L is the size of $\boldsymbol{H}_{\mathrm{R}}$ and $\boldsymbol{H} = \emptyset$.
for $l = 1, \ldots, L$ **do**
  Randomly select a tuple $\mathcal{T}_{\mathrm{B}_l} = (\tau_{\mathrm{B}_l}, \theta_{\mathrm{B}_l})$ from $\mathcal{T}_{\mathrm{B}}$;
  Generate α_l and ϵ_l;
  Obtain the parameter vector $\Omega_l = \{\tau_{\mathrm{B}_l}, \theta_{\mathrm{B}_l}, \alpha_l, \epsilon_l\}$;
  Extract centre frequency vector $f_{\mathrm{c}}^l = \{f_{\mathrm{c}_1}^l, f_{\mathrm{c}_2}^l, \ldots\}$ from $\boldsymbol{H}_{\mathrm{R}_l} = \{\boldsymbol{H}_{\mathrm{R}_l^1}, \boldsymbol{H}_{\mathrm{R}_l^2}, \ldots\}$;
  for $f_{\mathrm{c}_j}^l \in f_{\mathrm{c}}^l$ **do**
    $\boldsymbol{H}_{\mathrm{B}_l^j} \leftarrow$ Eqn. (1) with parameters Ω_l and $f_{\mathrm{c}_j}^l$;
    $\boldsymbol{H}_{l_j} \leftarrow \boldsymbol{H}_{\mathrm{R}_l^j} + \boldsymbol{H}_{\mathrm{B}_l^j}$;
  end
  $\boldsymbol{H} = \boldsymbol{H} \cup \boldsymbol{H}_l$.
end

Eqn. (1), we can then utilize Eqn. (2) to derive the ToF-AoA spectra as labels. Because we need only to train SpecTrans for deriving meaningful ToF-AoA spectra that can yield accurate ToF-AoA estimations, the ideal channel samples used for generating the training labels do not need to match the randomly sampled channels. Instead, we can leverage more channel samples for a more accurate spectrum estimation [14], while expecting SpecTrans to be powerful enough for performing a sparse recovery. Moreover, we should also consider the limitations of SpecTrans's fitting capacity, so we refrain from using all frequency bands as the ideal channel samples but empirically simulate 50 discrete channel samples to derive the ToF-AoA spectra labels; these spectra labels are further enhanced by exponentially amplifying the magnitude of the points pertaining to a subject. Besides, the ToF-AoA spectrum, defined by a discrete grid of $\Delta = (\Delta\tau, \Delta\theta)$, implies the existence and controllability of the quantization error bound, enabling the design of spectra that encode comprehensive information by introducing a fluctuation term $\hat{\Delta} < \Delta/2$. The whole label generation process is detailed in Algorithm 2, where the parameter estimation algorithm is denoted as $\mathcal{G}(\cdot)$.

D. From Snapshots to General Sensing

As shown in Fig. 5, UWB-Fi starts with the fast channel hopping scheme to collect sparse channel samples and assembles them into a sensing tensor (Section III-A). UWB-Fi then leverages the trained SpecTrans network (Section III-B) to derive ToF-AoA spectra, i.e., the estimation tensor, from the sensing tensor; this network is trained via model-augmented data and the corresponding training method (Section III-C), where sparse channel samples are diversified with background interference following Algorithm 1 as input, and ToF-AoA tuples are converted into spectra based on Algorithm 2 as labels. At this point, UWB-Fi follows the convention to use a thresholding technique [5] to filter out the minor peaks caused by periodic aliasing and obtain the main peaks indicating the ToF-AoA tuples of distinct subjects of interest in each ToF-AoA spectral snapshot. The outcomes are then converted

Algorithm 2 ToF-AoA Spectrum Generation

Input: Ground truth dataset $\mathcal{T}_{\mathrm{R}} = (\tau_{\mathrm{R}}, \theta_{\mathrm{R}})$; a list $f_{\mathrm{c}} = \{f_{\mathrm{c}_1}, f_{\mathrm{c}_2}, \ldots\}$ of centre frequencies and n, d, m, f_{s}, α, ϵ as in Eqn. (1); enhancement factors: γ, x, y, and σ; fluctuation $\hat{\Delta}(\Delta\hat{\tau}, \Delta\hat{\theta})$.
Output: ToF-AoA spectrum $\boldsymbol{\Xi}$.

$\tau_{\mathrm{R}} \in \{\tau_{\mathrm{R}} - \Delta\hat{\tau}, \tau_{\mathrm{R}}, \tau_{\mathrm{R}} + \Delta\hat{\tau}\}$;
$\theta_{\mathrm{R}} \in \{\theta_{\mathrm{R}} - \Delta\hat{\theta}, \theta_{\mathrm{R}}, \theta_{\mathrm{R}} + \Delta\hat{\theta}\}$;
for $f_{\mathrm{c}_j} \in f_{\mathrm{c}}$ **do**
  $\boldsymbol{H} \leftarrow$ Eqn. (1) with parameters τ_{R}, θ_{R}, f_{c_j};
  $\boldsymbol{\Xi} \leftarrow \boldsymbol{\Xi} + \mathcal{G}(\boldsymbol{H})$;
end
$\boldsymbol{\Xi} \leftarrow \mathsf{Enhance}\,(\boldsymbol{\Xi}/\mathrm{length}(f_{\mathrm{c}}), \gamma, x, y, \sigma)$;
Function $\mathsf{Enhance}(\boldsymbol{\Xi}, \gamma, x, y, \sigma)$**:**
  $\boldsymbol{\Xi} \leftarrow |\boldsymbol{\Xi}|^{\gamma}$;
  $\boldsymbol{Z}(x, y) \leftarrow \left| -\frac{1}{\pi\sigma^4}\left(1 - \frac{x^2+y^2}{2\sigma^2}\right) e^{-\frac{x^2+y^2}{2\sigma^2}} \right|$;
  $\boldsymbol{\Xi} \leftarrow \boldsymbol{\Xi} \odot (\boldsymbol{\Xi} * \boldsymbol{Z})$;
  return $\boldsymbol{\Xi}$;
end
x, y and σ are parameters of a Laplacian filter [56];
$*$ denotes convolution and $\odot$ represents Hadamard product.

from polar coordinates (i.e., ToF-AoA tuples) to Cartesian coordinates to determine the subjects' locations. Moreover, the phase information of a subject's motion is encoded into the multiple spectral amplitudes around the subject in the snapshot (confirmed in Sections III-B.1 and V-C.1). Therefore, tracking the spectral amplitudes across multiple snapshots for each subject should provide information about its motion status.

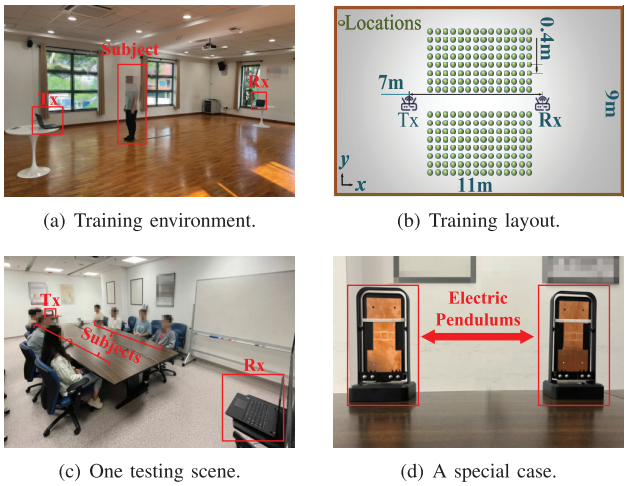
In particular, for the *k*-th subject, UWB-Fi extracts a set of tuples in its neighborhood with an empirical radius $\Delta r = 0.3$ m, $\{(\tau, \theta) | \sqrt{(\tau - \tau_k)^2 + (\theta - \theta_k)^2} \leq \Delta r\}$ to form its own *bin*, since the motion of the *k*-th subject affects its surrounding spectral values. We further extend the tuples of a bin in a snapshot at time *t*: (τ, θ) to $(\tau(t), \theta(t), \phi(t))$ where $\phi(t)$ is the spectral amplitude at $(\tau(t), \theta(t))$. Finally, UWB-Fi delivers the tuples contained in a bin across multiple snapshots to upper-layer sensing applications, including gesture recognition, vital signs monitoring, and so on. It is worth noting that UWB-Fi, being implemented with neural model, operate at the millisecond level during the inference time. This processing time (plus the millisecond-level channel sampling time discussed in Section III-A.2) is still significantly shorter than the 12 seconds required for stitching channels as explained in Section II-B, thus greatly enhancing real-time sensing capabilities.

IV. Prototype and Experiment Setup

In this section, we elaborate on UWB-Fi's implementation, and also introduce the experiment setup.

A. Framework Implementation

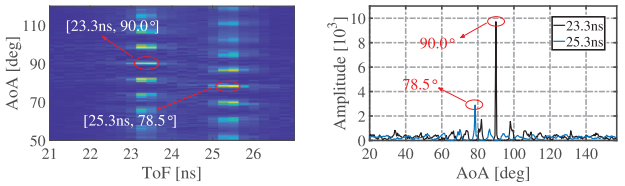
We use two Acer TravelMate laptops equipped with Intel AX210 Wi-Fi NIC to implement a UWB-Fi prototype. This



(a) Training environment. (b) Training layout.

(c) One testing scene. (d) A special case.

Fig. 10. Experiment setting examples: (a) training scene in the RR with (b) a layout containing positions marks; (c) a typical testing scene in the CR, and (d) a case with two electric pendulums as subjects.



(a) ToF-AoA spectrum. (b) AoA-Amplitude sectional views.

Fig. 11. SpecTrans output clearly differentiates two paths.

Wi-Fi NIC has two antennas and supports all channels mentioned in Section III-A.1. We implement the fast channel hopping scheme (Section III-A) in the laptops to capture the channel samples. The channel samples are then parsed and converted into the matrix structure using MATLAB. Our SpecTrans network is implemented in the PyTorch 1.7.1 environment. The encoder consists of 3 basic modules and one down-sampling transformation module, while the decoder comprises 3 basic modules and one up-sampling transformation module.

B. Experiment Setup

We recruit 8 subjects, 6 males and 2 females, aged between 20 and 30. We conduct experiments in six environments, including rehearsal room (RR), lecture room (LR), reading room (RD), conference room (CR), discussion room (DR), and private office (PO). We collect training data in RR and use data collected in other environments for testing purposes. Although our testing environments have diversified sizes and laptop distance settings (ranging from 3 m to 9 m), we set the distance between the two laptops to 7 m in the training environment, which is proven to be generalizable to other environments in Section V-C.2, thanks to the augmented data with diversified interference "background" (see Section III-C.1) and the model-based neural network design. Besides comparison and ablation studies, we mainly perform three sets of experiments on multi-person sensing applications; they have strictly followed the IRB of our institute.

1) Localization: We instruct subjects to stand or sit in different positions in each experiment. Fig. 10(a) shows the training environment in the RR and Fig. 10(b) marks the positions for data collection: we take the gravity center of a subject as its ground truth position. We adopt one continuous channel stitching scheme [24] as our comparison baseline, where overlapped channels at 2.4 GHz band are all collected for stitching a sensing bandwidth of 80 MHz. As no existing solution can span across 4.7 GHz using commodity Wi-Fi hardware to provide the same sensing bandwidth as UWB-Fi, we only get two barely comparable choices to serve as baseline [23], [24], while the other one [23] is omitted as it yields similar results in our experiments. We arrange for the testing environment of the baseline to also be within our training environment, because cross-environment capability is irrelevant to the baseline thanks to its model-driven processing nature. Subsequently, we conduct localization for the baseline according to its own proposed methodology. In order to exert the finest resolution of UWB-Fi, we arrange a testing scene with two electric pendulums as shown in Fig. 10(d).

2) Gesture Recognition: We choose one of the testing environments with 8 subjects sitting around a table in the CR (shown in Fig. 10(c)) for this purpose. The subjects are instructed to simultaneously perform six distinct hand gestures: forward-backward (FB), up-down (UD), left-right (LR), draw circle (DC), zig-zag (ZZ), and clap (CL). Each gesture is performed 20 times. These six gesture classes serve as ground truth. After obtaining the temporal sequence of ToF-AoA spectra produced by SpecTrans, we extract individual bin sequences corresponding respective subjects. Then, we employ the classifier proposed by Widar3.0 [6] to recognize gestures.

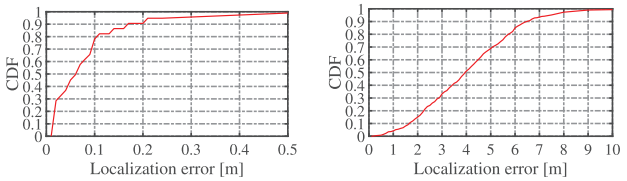
3) Respiration Monitoring: We use the "8-subject in CR" scene here again. Apart from recording channel samples, we concurrently record subjects' respiratory waveforms as ground truth, for both training the already established respiration monitoring algorithms and evaluation purposes, using NeuLog chest strap [58]. The total recording time for respiration is 80 minutes. A processing pipeline similar to the gesture recognition is adopted here, except that we employ the method proposed by MoRe-Fi [59] to predict respiratory waveforms.

V. Evaluations

In this section, we embark on a comprehensive evaluation of UWB-Fi's capabilities. Given its essential function of differentiating multiple subjects, we start with an investigation into UWB-Fi's localization resolution and accuracy: comparing UWB-Fi with the baseline stresses on its superiority in fine-granularity sensing. We also further report sensing results for multi-subject gesture recognition and respiration monitoring, demonstrating UWB-Fi's fine-granularity general sensing performance. Finally, we perform the analysis of impact factors and an ablation study on our training approach, demonstrating the robustness and rationale of our UWB-Fi framework.

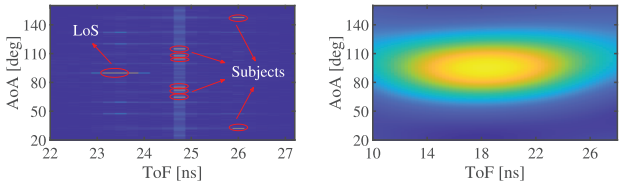
A. Localization and Comparison

1) Localization Accuracy: Let us first use an example to illustrate the localization capability of UWB-Fi. As shown in Fig. 11(a), a ToF-AoA spectrum derived from channel samples via SpecTrans clearly exhibits two peaks (after filtering out



(a) UWB-Fi. (b) Baseline.

Fig. 12. Contrasting UWB-Fi with baseline in localization accuracy: about 70 times lower in median error!.



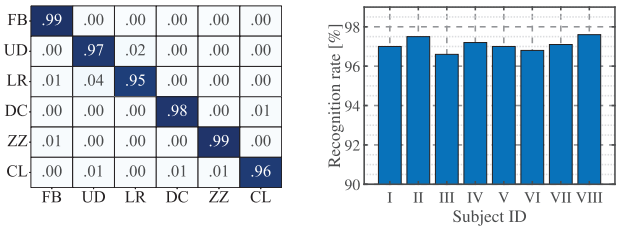
(a) UWB-Fi. (b) Baseline.

Fig. 13. Significant distinction in localization resolution between UWB-Fi and baseline.

the minor peaks caused by periodic aliasing, as explained in Section III-D) corresponding to two ToF values of 23.3 ns and 25.3 ns, indicating two different propagation paths. In the meantime, Fig. 11(b) shows two sectional views in AoA at these two ToF values; they sharply indicates two bearings of 90° and 78.5°. In fact, these two ToF-AoA tuples precisely correspond to the (direct) Tx-Rx path and the reflection path caused by a subject: the former apparently contains much higher energy. Recall that earlier proposals made great efforts to either get around or infer the information on Tx-Rx path [21], [26], [31], yet UWB-Fi can naturally obtain it and further leverages it as a reference to establish the coordinate system for localization.

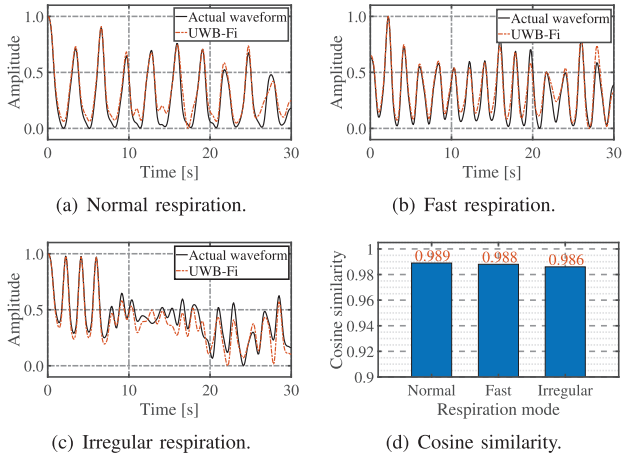
Given these precise ToF-AoA tuples and the established coordinate system, individual subjects can be located via the method described in Section III-D. Fig. 12 compares UWB-Fi with the baseline in terms of their localization accuracy. One can readily observe that UWB-Fi achieves a median error of 6 cm, about 70 times lower than the baseline's 4 m. Meanwhile, the 90th percentile tail errors for UWB-Fi and the baseline are 0.17 m and 6.5 m, respectively. Considering the typical size of an adult, UWB-Fi clearly demonstrates a sufficient accuracy for common indoor scenarios, whereas the baseline's performance is far from adequate (even considering the bias between the potential reflection surface and the gravity center), making it inapplicable to fine-granularity sensing contexts.

2) Localization Resolution: We proceed to investigate localization resolution of UWB-Fi, by having 8 subjects standing symmetrically with respect to the Tx-Rx path in the RD environment. Fig. 13(a) and Fig. 13(b) depict the ToF-AoA spectra in multi-person localization scenes for UWB-Fi and the baseline, respectively. One may clearly discern multiple peaks/bins (as specified in Section III-D) in the spectrum of Fig. 13(a), each representing distinct individuals and facilitating further sensing requirements. On the contrary, Fig. 13(b) confirms that, when employing the baseline method for localization, the ToF-AoA spectrum is apparently incapable of discerning different subjects. UWB-Fi's exceptional performance is primarily attributed to its capability in effectively



(a) Confusion matrix of gestures. (b) Recognition rates of each subject.

Fig. 14. Gesture recognition results.



(a) Normal respiration. (b) Fast respiration.

(c) Irregular respiration. (d) Cosine similarity.

Fig. 15. Recovered respiration waveforms in (a) normal, (b) fast, (c) irregular modes, and (d) their cosine similarity values.

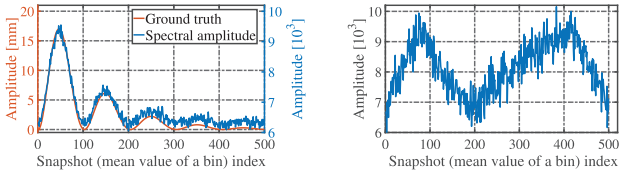
acquiring ultra-wide bandwidth and eliminating hardware-related random offsets. In particular, this capability may even compensate the hardware deficit in antenna quantity, as acquiring AoA with only two antennas were otherwise impossible due to the limited angular resolution [60] offered by two coarse-grained observations.

B. Fine-Granularity Sensing

Once subjects' bins are clearly identified, individual human sensing can be readily performed via tracking the spectral amplitude across multiple snapshots. In this section, we showcase two upper-layer applications that successfully achieve fine-granularity sensing based on UWB-Fi, without comparing it to the baseline, as the baseline is intended solely for localization.

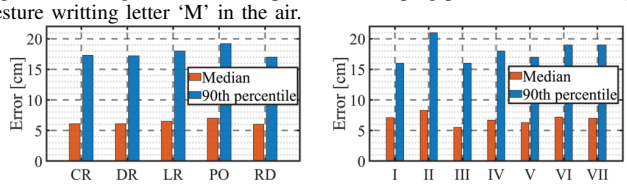
1) Gesture Recognition: Using the setup prescribed in Section IV-B.2, the performance of gesture recognition upon UWB-Fi's output can be evaluated. As shown by the confusion matrix for all subjects in Fig. 14(a), UWB-Fi achieves an overall recognition rate of 97.1%: the FB gesture has the highest recognition rate at 99.0%, while the UD and LR gestures exhibit more confusion. As both UD and LR gestures primarily involve subtle wrist movements whereas other gestures involve more extensive arm movements, their weaker performance can be largely attributed to their lower impacts on channel samples. We then proceed to analyze the recognition rate of different subjects for all gestures: as shown in Fig. 14(b), the rates of all subjects stand around a satisfactory 97%.

2) Respiration Monitoring: We hereby demonstrate how UWB-Fi's output (ToF-AoA spectrum) can be used to acquire



(a) Pendulum. (b) Finger gesture.

Fig. 16. Sensing results of (a) a pendulum damping process and (b) a finger gesture writting letter 'M' in the air.



(a) Various environments. (b) Various subjects.

Fig. 17. Localization median and 90th percentiles tail errors across various (a) environments and (b) subjects.

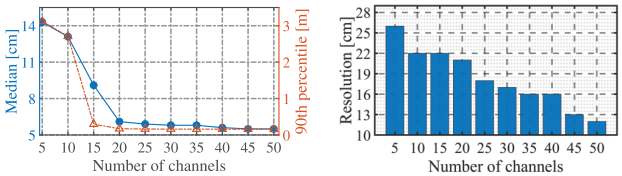
respiratory waveforms, using the setup described in Section IV-B.3. To validate waveform recovery across diverse breathing dynamics, we instruct one subject to perform normal, fast, and irregular respiration, while keeping other subjects' breathe regularly. Figs. 15(a), 15(b), and 15(c) present the recovered respiratory waveforms of these three situations; one can readily observe that the recovered waveforms by UWB-Fi effectively adapt to the variations taking place in reality. Even in such challenging circumstance of irregular respiration, the output of UWB-Fi still allows for a remarkably successful recovery in a respiratory waveform, with only a minor discrepancy observed during the 10 s to 20 s interval, corresponding to the period of extremely light respiration. Besides, all waveforms bear high morphological similarity to the ground truth according to Fig. 15(d).

C. Impact Factors

In this section, we delve deeper into the sensing principle and impact factors.

1) Sensing With Spectral Amplitude: Trained by the ToF-AoA spectrum, UWB-Fi also encodes phase information in spectral amplitude. To confirm this claim, we extract the subject's bin sequences across multiple snapshots (as explained in Section III-D) for both a pendulum and a finger gesture (writing the letter 'M'), and track their processed mean values, as depicted in Fig. 16. Since continuous tracking such motions is impossible with only signal amplitude (which only infers range variation and has centimeter-level precision even for UWB-Fi), the demonstrations in Fig. 16 strongly confirm that continuous variations in phase have been represented by the spectral amplitudes of UWB-Fi's snapshots.

2) Environments and Subjects: We evaluate the localization accuracy of all subjects across five unseen environments. The results, as depicted in Fig. 17, confirm that varying environments and subject positions have a minimal impact on the localization performance: the median and 90th percentiles tail errors remain consistent (with only minor fluctuations) around 6 cm and 17 cm, respectively; this is true even in the PO that substantially deviates from our RR training environment. It is worth noting that this generalizability, though partially coming



(a) Localization accuracy. (b) Localization resolution.

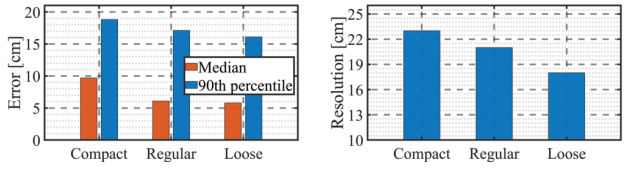
Fig. 18. Localization error and resolution vs. varying channel quantity.

from the design of SpecTrans (without even invoking the SAs in Section III-B.4), should be more attributed to the ultra-wide bandwidth of UWB-Fi and the training method: the resulting fine-granularity sensing allows UWB-Fi to accurately discern bins pertinent to specific subjects, naturally filtering inferences related to various multipath.

3) Channel Quantity: We hereby use experiment results to analyze the impact of the channel quantity on localization accuracy and resolution. For the accuracy analysis, UWB-Fi randomly collects 5 to 50 sparse channel samples and evaluates performance leveraging median and 90th percentile errors. The results, as shown in Fig. 18(a), indicate that both the errors decrease as the number of channels increases, but these changes level off after reaching 20 channel samples. We employ the same setup to analyze spatial resolution. To ensure a sufficiently small distance between the two subjects, we conduct tests on the two pendulums, as shown in Fig. 10(d). The results, shown in Fig. 18(b), indicate that resolution improves with an increase in the sampled channel quantity, reaching sub-decimeter level at 50 channels. For real-time and error considerations, UWB-Fi samples 20 channels, achieving a resolution of approximately 20 cm, which is sufficient for indoor scenarios to distinguish normal-sized adults. Of course, these results might still be affected by conditions unknown to us by far, but we have offered adaptation schemes in Section III-A.1.

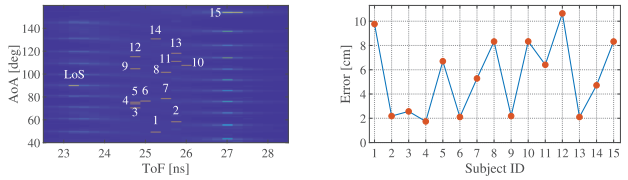
It is worth noting that, when contrasting with Figs. 12(b) and 13(b), it becomes evident that, even with only 5 channels, UWB-Fi yields much higher precision than the baseline employing 13 overlapped channels. The reason accounting for this is twofold: i) UWB-Fi's the fast channel hopping allows it to gather channel samples across up to 4.7 GHz UWB bandwidth and hence to acquire far more comprehensive information, and ii) the impressive sparse recovery and offset elimination capabilities of SpecTrans helps UWB-Fi to fully extract the useful information within the channel samples.

4) Placement of Channels: Following the same experiment setup as previously described using 20 channel samples, we proceed to investigate the impact of channel placement on accuracy and spatial resolution. We design three scenarios: randomly selected channel samples (Regular), manually selected compact channel samples (Compact), and manually selected loose channel samples spanning from 2.4 GHz to 6 GHz bands (Loose). The results, as depicted in Fig. 19, demonstrate that both accuracy and resolution achieve the best performance under the loose placement, while performing the worst under the compact scenario. The main reason is that the broader span of the loose placement provides greater diversity, enabling the recovery of a wider effective bandwidth from



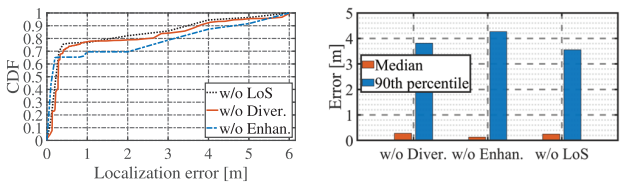
(a) Localization accuracy. (b) Localization resolution.

Fig. 19. Impact of channel placement.



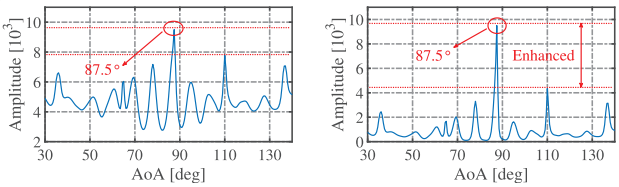
(a) ToF-AoA spectrum. (b) Localization errors.

Fig. 20. A significant number of recognizable subjects.



(a) Localization error. (b) Median and 90th percentile errors.

Fig. 21. Ablation studies on localization error.



(a) Initial estimation. (b) Enhanced estimation.

Fig. 22. AoA-Amplitude curves (a) without and (b) with enhancement: the latter allows for a much better differentiation.

sparse channel samples. Therefore, when the availability of all channels is known, manually setting a loose placement of channels can assist UWB-Fi in achieving better performance.

5) Number of Subjects: To demonstrate UWB-Fi is reliable for a significant number of subjects, we increase to 15 people, enough to fill the CR, standing randomly in various positions. The result, as shown in Fig. 20(a), clearly displays multiple distinct peaks in the ToF-AoA spectrum, enabling us to identify all 15 subjects. Furthermore, we also conduct additional analysis on the localization errors of the 15 subjects, as depicted in Fig. 20(b). The maximum error is only 10.6 cm, while the minimum error is less than 2 cm. These findings underscore the stable sensing performance of UWB-Fi, even in scenarios involving highly crowded subjects.

D. Training Approaches Ablation

In order to validate the effectiveness of our training approach proposed in Section III-C, we compare it against four distinct label generation methods that include: A) generating a pseudo-spectrum directly from ToF-AoA tuples, B) training data without background interference (w/o Diver.), C) estimating ToF-AoA spectra without enhancement (w/o Enhan.), and D) generating labels without using the Tx-Rx path as a reference (w/o LoS). Note that one should not expect a neural network to directly output ToF-AoA tuples (hence should not be trained so), the fundamental reason lies in the the bias-variance dilemma in estimation [40]. As discussed in III-B.1, directly translating channel samples to a few real-valued tuples has to strike a balance between bias and variance, leading to bad performance on either side.

1) Method A: We first take a straightforward method to generate labels by converting each ToF-AoA tuple into a Gaussian kernel placed at corresponding position in a 2D spectrum; however, training with this kind of spectrum fails. Denote the neural network taking such spectra as labels by $\mathbf{g}_{\mathrm{A}}$, we train both $\mathbf{g}_{\mathrm{A}}$ and $\mathbf{g}_{\boldsymbol{W}}$ (see Section III-C) and measure the entropy [61] of them; the results show that entropy values of 253 and 132 for $\mathbf{g}_{\mathrm{A}}$ and $\mathbf{g}_{\boldsymbol{W}}$, respectively. Due to the higher complexity of $\mathbf{g}_{\mathrm{A}}$, using labels generated by Method A for training certainly hampers convergence.

2) Method B (w/o Diver.): Since UWB-Fi's resolution helps isolate subjects from the environment, we consider removing Algorithm 1. However, as shown in Fig. 21, the median and 90th percentile errors increase to 28 cm and 3.8 m, respectively. It indicates that incorporating background interferences generated by Algorithm 1 into real-life channel samples can help SpecTrans achieve superior cross-environment capability. Besides, this algorithm is also crucial for reducing the workload associated with training data collection.

3) Method C (w/o Enhan.): While generating labels using Algorithm 2, one might wonder if the enhancement (the $\mathsf{Enhance}(\cdot)$ function) is necessary. According to Fig. 22(a), it appears that removing the enhancement could significantly affect the estimation results. Specifically, we create a set of ideal channel samples with only one target peak at 87.5°. However, as the antenna quantity is limited, the resulting aliasing causes multiple interference peaks with one prominently appearing at around 110°. Without proper enhancement to differentiate these peaks, using such labels to train SpecTrans could significantly reduce accuracy. As depicted in Fig. 21, the median error increases to around 13 cm, while the 90th percentile error even surpasses 4 m, which is obviously worse than our previous results in Fig. 12(a). With the enhancement, Fig. 22(b) shows that the target peak becomes significantly more distinguishable compared to other "fake" ones, potentially improving the training outcome.

4) Method D (w/o LoS): Since the Tx-Rx path appears to be largely irrelevant to sensing often based on reflection paths, we consider removing it from the labels for the sake of simplicity. Unfortunately, the outcomes, as depicted in Fig. 21, fall short of expectations. Despite the median error remaining reasonably small, there is a pronounced long tail in the distribution, with the 90th percentile error exceeding 3 m. This negative effect can be explained by Fig. 11(b): in reality, the Tx-Rx path signal energy is much stronger than other paths; removing it can be seen as introducing a significant perturbation, as shown in Eqn. (5). Though the Tx-Rx path signal is probably not useful for sensing reflection paths, ignoring it outright could fundamentally change the "landscape" of a snapshot rather than merely removing a peak from the spectrum, hence

negatively affecting the correspondence between training data and real-life samples. Therefore, regardless of whether the Tx-Rx path is used for sensing or not, the training labels have to emulate all paths. In fact, UWB-Fi indeed needs the Tx-Rx path as a reference point to establish a coordinate system for localization.

VI. Conclusion

We have introduced UWB-Fi to be a pioneering Wi-Fi sensing framework operating with a 4.7 GHz ultra-wide bandwidth; it achieves this bandwidth exclusively via discrete and irregular channel sampling. UWB-Fi first boasts a fast channel hopping scheme to enable sparse sampling across its UWB on commodity Wi-Fi hardware, all without disrupting default communications. In absence of suitable signal processing tools for handling such channel samples, we have innovated in a model-based deep learning approach: along with its model-drive training strategy, UWB-Fi's SpecTrans network is capable of translating sparse channel samples into parameter spectra precisely implying ToF-AoA attributes of individual subjects, effectively sidestepping the curse of bias-variance dilemma. Meanwhile, SpecTrans also effectively mitigates hardware-induced offsets inherent to Wi-Fi signals. Leveraging extensive experiments on our UWB-Fi prototype, we have demonstrated that UWB-Fi can help common Wi-Fi sensing applications successfully achieve fine-granularity sensing in indoor multi-subject scenarios. As a future direction, we plan to fully explore beamforming enabled by channel sounding mechanisms [62], [63] to further enhance the performance of UWB-Fi. The codes accompanying this paper are available via https://github.com/DeepWiSe888/UWB-Fi.

References

[1] X. Li, H. Wang, Z. Chen, Z. Jiang, and J. Luo, "UWB-fi: Pushing Wi-Fi towards ultra-wideband for fine-granularity sensing," in *Proc. 22nd Annu. Int. Conf. Mobile Syst., Appl. Services*, Jun. 2024, pp. 42–55.

[2] C. Deng et al., "IEEE 802.11 be Wi-Fi 7: New challenges and opportunities," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2136–2166, 4th Quart., 2020.

[3] J. Wang et al., "LiFS: Low human-effort, device-free localization with fine-grained subcarrier information," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2016, pp. 243–256.

[4] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proc. 15th Annu. Int. Conf. Mobile Syst., Appl., Services*, 2017, pp. 252–264.

[5] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2.0: Passive human tracking with a single Wi-Fi link," in *Proc. 16th Annu. Int. Conf. Mobile Syst., Appl., Services*, 2018, pp. 350–361.

[6] Y. Zheng et al., "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proc. 17th Annu. Int. Conf. Mobile Syst. Appl. Services*, 2019, pp. 313–325.

[7] W. Jiang et al., "Towards 3D human pose construction using WiFi," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, Apr. 2020, pp. 1–14.

[8] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf WiFi," in *Proc. 16th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2015, pp. 267–276.

[9] Y. Zeng, D. Wu, J. Xiong, J. Liu, Z. Liu, and D. Zhang, "MultiSense: Enabling multi-person respiration sensing with commodity WiFi," in *Proc. 22nd ACM UbiComp*, 2020, pp. 1–29.

[10] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst.*, Jun. 2017, pp. 1230–1239.

[11] Y. He, J. Liu, M. Li, G. Yu, and J. Han, "Forward-compatible integrated sensing and communication for WiFi," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 9, pp. 2440–2456, Sep. 2024.

[12] Y. He, G. Yu, Y. Cai, and H. Luo, "Integrated sensing, computation, and communication: System framework and performance optimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 1114–1128, Feb. 2024.

[13] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

[14] Z. Chen et al., "*M^{3}*: Multipath assisted Wi-Fi localization with a single access point," *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 588–602, Feb. 2021.

[15] K. Chintalapudi et al., "WiFi-NC: WiFi over narrow channels," in *Proc. 9th USENIX NSDI*, 2012, pp. 43–56.

[16] Y. Luo and K. Chin, "An energy efficient channel bonding and transmit power control approach for WiFi networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8251–8263, Aug. 2021.

[17] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. 11th USENIX Conf. Netw. Syst. Design Implement.*, Apr. 2014, pp. 317–329.

[18] F. Adib and D. Katabi, "See through walls with Wi-Fi!," in *Proc. 27th ACM SIGCOMM*, 2013, pp. 75–86.

[19] S. Tan, L. Zhang, Z. Wang, and J. Yang, "MultiTrack: Multi-user tracking and activity recognition using commodity WiFi," in *Proc. CHI Conf.*, 2019, pp. 1–12.

[20] J. Hu, T. Zheng, Z. Chen, H. Wang, and J. Luo, "MUSE-fi: Contactless MUti-person SEnsing exploiting near-field Wi-Fi channel variation," in *Proc. 29th Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2023, pp. 1–15.

[21] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "MD-track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, Aug. 2019, pp. 1–16.

[22] C. R. Karanam, B. Korany, and Y. Mostofi, "Tracking from one side: Multi-person passive tracking with WiFi magnitude measurements," in *Proc. 18th Int. Conf. Inf. Process. Sensor Netw.*, Apr. 2019, pp. 181–192.

[23] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity WiFi," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 53–64.

[24] J. Xiong, K. Sundaresan, and K. Jamieson, "Tonetrack: Leveraging frequency-agile radios for time-based indoor wireless localization," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 537–549.

[25] J. Huang, G. Xing, G. Zhou, and R. Zhou, "Beyond co-existence: Exploiting WiFi white space for zigbee performance assurance," in *Proc. 18th IEEE Int. Conf. Netw. Protocols*, Oct. 2010, pp. 305–314.

[26] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in *Proc. 13th Usenix Conf. Netw. Syst. Design Implement.*, 2016, pp. 165–178.

[27] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[28] Y. Ma, N. Selby, and F. Adib, "Minding the billions: Ultra-wideband localization for deployed RFID tags," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw. (ACM MobiCom)*, Oct. 2017, pp. 248–260.

[29] K. M. Cuomo, J. E. Pion, and J. T. Mayhan, "Ultrawide-band coherent processing," *IEEE Trans. Antennas Propag.*, vol. 47, no. 6, pp. 1094–1107, Jun. 1999.

[30] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, pp. 1–36, Jun. 2019.

[31] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," in *Proc. ACM Conf. Special Interest Group Data Commun.*, Aug. 2015, pp. 269–282.

[32] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: Principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015.

[33] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[34] E. Perahia and R. Stacey, *Next Generation Wireless LANs: 802.11n and 802.11ac*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[35] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, vol. 47, no. 6, pp. 957–1000, Jul. 1968.

[36] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Hoboken, NJ, USA: Wiley, 1994.

[37] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[38] J. Pegoraro, J. O. Lacruz, M. Rossi, and J. Widmer, "SPARCS: A sparse recovery approach for integrated communication and human sensing in mmWave systems," in *Proc. 21st ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, May 2022, pp. 79–91.

[39] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.
[40] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
[41] D.-X. Zhou, "Universality of deep convolutional neural networks," *Appl. Comput. Harmon. Anal.*, vol. 48, no. 2, pp. 787–794, Mar. 2020.
[42] N. Czink, X. Yin, H. Ozcelik, M. Herdin, E. Bonek, and B. Fleury, "Cluster characteristics in a MIMO indoor propagation environment," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1465–1475, Apr. 2007.
[43] Kernel Develop. Community. (2009). *DebugFS*. Accessed: Aug. 18, 2023. [Online]. Available: https://docs.kernel.org/filesystems/debugfs.html
[44] T. K. D. Community. (2009). *Sysfs*. Accessed: Aug. 18, 2023. [Online]. Available: https://www.kernel.org/doc/html/next/filesystems/sysfs.html
[45] M. Zargari et al., "A dual-band CMOS MIMO radio SoC for IEEE 802.11 n wireless LAN," *IEEE J. Solid-State Circuits*, vol. 43, no. 12, pp. 2882–2895, Dec. 2008.
[46] T. Zheng, Z. Chen, S. Ding, and J. Luo, "Enhancing RF sensing with deep learning: A layered approach," *IEEE Commun. Mag.*, vol. 59, no. 2, pp. 70–76, Feb. 2021.
[47] S. Ding, Z. Chen, T. Zheng, and J. Luo, "RF-Net: A unified meta-learning framework for RF-enabled one-shot human activity recognition," in *Proc. 18th Conf. Embedded Netw. Sens. Syst.*, 2020, pp. 517–530.
[48] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér–Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
[49] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, vol. 452. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.
[50] Y. Zhu and N. Zabaras, "Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification," *J. Comput. Phys.*, vol. 366, pp. 415–447, Aug. 2018.
[51] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 4700–4708.
[52] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 510–519.
[53] X. Li, S. Chen, X. Hu, and J. Yang, "Understanding the disharmony between dropout and batch normalization by variance shift," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2682–2690.
[54] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, Jan. 2009.
[55] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Nov. 2014.
[56] A. Wu and C. Deng, "Discriminating known from unknown objects via structure-enhanced recurrent variational AutoEncoder," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 23956–23965.
[57] E. Basar and I. Yildirim, "Reconfigurable intelligent surfaces for future wireless networks: A channel modeling perspective," *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 108–114, Jun. 2021.
[58] NeuLog.(2017). *Respiration Monitor Belt Logger Sensor NUL-236*. Accessed: Feb. 12, 2023. [Online]. Available: https://neulog.com/respiration-monitor-belt/
[59] T. Zheng, Z. Chen, S. Zhang, C. Cai, and J. Luo, "MoRe-Fi: Motion-robust and fine-grained respiration monitoring via deep-learning UWB radar," in *Proc. 19th ACM Conf. Embedded Netw. Sensor Syst.*, Coimbra, Portugal, Nov. 2021, pp. 111–124.
[60] Z. Chen et al., "AWL: Turning spatial aliasing from foe to friend for accurate WiFi localization," in *Proc. 13th Int. Conf. Emerg. Netw. EXperiments Technol.*, Nov. 2017, pp. 238–250.
[61] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
[62] M. Xu et al., "Beamforming made malicious: Manipulating Wi-Fi traffic via beamforming feedback forgery," in *Proc. 30th Annu. Int. Conf. Mobile Comput. Netw.*, Dec. 2024, pp. 908–922.
[63] Y. He, M. Xu, F. Xiao, and J. Luo, "VersaBeam: Versatile beamforming for integrated sensing and communication over commodity Wi-Fi," in *Proc. 44th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Jan. 2025, pp. 1–10.

Xin Li (Member, IEEE) received the Ph.D. degree in power machinery and engineering from Tianjin University, China, in 2022. From 2022 to 2023, he was a Senior Software Engineer with the Intelligent Automotive Solution Business Unit, Huawei Technologies Company Ltd. He is currently a Research Fellow with the College of Computing and Data Science, Nanyang Technological University. His research interests include machine learning, signal processing, and wireless sensing.

Jingzhi Hu (Member, IEEE) received the B.S. degree from the School of Electrical Engineering and Computer Science, Peking University, in 2017, and the Ph.D. degree from the School of Electronics, Peking University, in 2022. After the graduation, he is currently a Research Fellow with the College of Computing and Data Science, Nanyang Technological University. His main research interests include machine learning, Wi-Fi sensing systems, and reconfigurable intelligent surface-aided RF sensing techniques for the Internet of Things.

Hongbo Wang (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from the University of Electronic Science and Technology of China in 2020 and the M.S. degree in communications engineering from Nanyang Technological University, Singapore, in 2021, where he is a currently pursuing Ph.D. degree with the College of Computing and Data Science. He has published papers in ACM SenSys, ACM MobiCom, ACM CCS, ACM MobiSys, IEEE INFOCOM, and IEEE TRANSACTIONS ON MOBILE COMPUTING. His research interests include wireless sensing and integrated sensing and communication (ISAC).

Zhe Chen (Member, IEEE) received the Ph.D. degree in computer science from Fudan University, China. He is currently an Associate Professor with the Institute of Space Internet, Fudan University, and the Co-Founder of AIWiSe Ltd. Inc. Before joining Fudan University, he worked as a Research Fellow with NTU for several years. His research achievements, along with his efforts in launching products based on them, have thus earned him the 2021 ACM SIGMOBILE China Rising Star Award recently. During the Ph.D. degree, he received the 2019 ACM SIGCOMM China Doctoral Dissertation Award.

Jun Luo (Fellow, IEEE) received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, China, and the Ph.D. degree in computer science from Swiss Federal Technology Institute of Lausanne (EPFL), Lausanne, Switzerland. From 2006 to 2008, he was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. In 2008, he joined Nanyang Technological University, Singapore, as a Faculty Member, where he is currently an Associate Professor. His research interests include mobile and pervasive computing, wireless networking, machine learning and computer vision, applied operations research, as well as security. More information can be found at http://www.ntu.edu.sg/home/junluo