UWB-Fi: Pushing Wi-Fi towards Ultra-wideband for Fine-Granularity Sensing

Xin Li[1] Hongbo Wang[1] Zhe Chen[2*] Zhiping Jiang[3] Jun Luo[1*]
[1]School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore
[2]Intelligent Networking and Computing Research Center and School of Computer Science, Fudan University, China
[3]School of Computer Science and Technology, Xidian University, China
Email: {l.xin, hongbo001, junluo}@ntu.edu.sg, zhechen13@fudan.edu.cn, zpj@xidian.edu.cn

ABSTRACT

The limited bandwidth of Wi-Fi severely confines the granularity (especially in differentiating multiple subjects) of Wi-Fi sensing, posing a significant challenge for its wide adoption. Though utilizing multiple channels to expand the effective bandwidth sounds plausible, continuous spectrum stitching towards *ultra-wideband* (UWB) is far from practical given various constraints (e.g., the runtime channel availability and inconsistent channel responses across a wide bandwidth). To this end, we propose UWB-Fi as a novel Wi-Fi sensing system with ultra-wide bandwidth, leveraging only discrete and irregular channel sampling. We first design a fast channel hopping scheme to perform arbitrary sampling across 4.7 GHz (i.e., 2.4 to 7.1 GHz) bandwidth on commodity Wi-Fi hardware without interrupting default communications. As no signal processing tool is available to handle such channel samples, we innovate in a model-based deep learning approach that translates discrete channel samples to high-dimensional spectral parameters; this method successfully avoids the *bias-variance tradeoff* in parameter estimation, while filtering out hardware-related offsets inherent to Wi-Fi. Through extensive evaluations, we demonstrate that UWB-Fi successfully achieves fine-granularity sensing, enabling centimeter-level resolution for indoor multi-person sensing.

CCS CONCEPTS

• **Hardware → Signal processing systems**; • **Human-centered computing → Ubiquitous and mobile computing**; • **Computing methodologies → Machine learning**.

KEYWORDS

Wi-Fi human sensing, multi-person sensing, ISAC, spatial resolution, Wi-Fi localization.

ACM Reference Format:
X. Li, H. Wang, Z. Chen, Z. Jiang, and J. Luo. 2024. UWB-Fi: Pushing Wi-Fi towards Ultra-wideband for Fine-Granularity Sensing. In *The 22nd Annual International Conference on Mobile Systems, Applications and Services (MobiSys'24), June 3–7, 2024, Minato-ku, Tokyo, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3643832.3661889

* Zhe Chen and Jun Luo are both the corresponding authors.

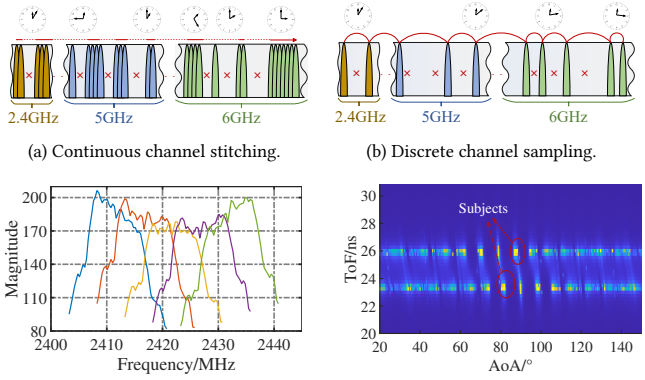
MobiSys'24, June 3–7, 2024, Minato-ku, Tokyo, Japan

ACM ISBN 979-8-4007-0581-6/24/06.
https://doi.org/10.1145/3643832.3661889

1 INTRODUCTION

Wi-Fi sensing [25, 37, 44, 53, 54, 57, 71, 76] has garnered attention from both academia and industry in recent years, thanks to the wide deployment of Wi-Fi infrastructure. Promoted by the ability of obtaining *channel state information* (CSI) [17], diversified sensing applications have been proposed and implemented, including vital signs monitoring [37, 71], gesture detection [53, 76], activity recognition [15, 25], as well as localization and motion tracking [10, 44, 54, 72]. Despite the promising outcomes demonstrated by these applications, one major challenge still persists as Wi-Fi sensing continues to evolve: the limited Wi-Fi bandwidth fails to offer a sufficient spatial resolution (in both *range* and *bearing*), hindering the realization of fine-granularity sensing to precisely differentiate multiple subjects. Since fine-granularity sensing demands centimeter-level resolution offered only by a GHz-level bandwidth, fully exploiting the capabilities of commodity Wi-Fi hardware to achieve *ultra-wideband* (UWB) sensing in a physical sense has emerged as a necessary yet challenging research direction under the *integrated sensing and communication* (ISAC) ambition [7].

An attempted solution to overcome the bandwidth limitation of Wi-Fi sensing is *continuous channel stitching* [65, 67]; it combines multiple consecutive channels to form a wider bandwidth, as shown in Figure 1a. This approach, albeit indicating the performance upper bound, faces four major weaknesses. Firstly, consecutive channels may not always be available for sensing, as these channels can be occupied or contended by other co-channel devices [22, 28], while certain channels are meant only for special purposes and hence not accessible by civilian Wi-Fi applications [59]. Secondly, the



(a) Continuous channel stitching. (b) Discrete channel sampling.

(c) Boundary effect needs overlaps. (d) ToF-AoA spectrum.

Figure 1: Continuous stitching suffers availability, boundary effect, and time constraint (a) and (c), yet all these can be overcome by discrete sampling (b) and (d).

boundary effect of frequency response, resulting in incompatibility between adjacent channels, demands a lot more overlapped channels and hence makes the stitching process very cumbersome (if not impossible), as shown in Figure 1c. Thirdly, the channel coherence time budget is not sufficient to scan a large number of overlapped channels needed by UWB sensing. Finally, the continuous channel stitching proposals [65, 67] are only shown to be viable for localization purpose; it is highly doubtful if they may support general sensing applications. Therefore, continuous channel stitching is apparently infeasible, hence calling for novel solutions to enable UWB sensing on Wi-Fi.

Since UWB sensing relies on a wider bandwidth to enhance *sensing diversity*, continuous bandwidth stitching is not a must. Intuitively, one may potentially overcome the aforementioned limitations by resorting to *discrete channel sampling*: as shown in Figure 1b, a wide bandwidth can be scanned in a non-continuous and even irregular manner.[1] This strategy allows a free selection of arbitrary channels subject to their availability in runtime, so it has a higher chance to gather sufficient information for "assembling" an effective UWB spectrum. Being discrete samples also naturally avoids the boundary effect of frequency response (a major "curse" on continuous stitching), since samples do not share common boundaries. Last but not least, discrete channel sampling potentially requires much fewer samples thanks to the principle of *compressive sensing* [60, 62, 63] and the sparse nature of most physical phenomena. Therefore, meeting the channel coherence time budget while achieving accurate parameter estimation (see Figure 1d) can become feasible.

Of course, this intuitively plausible strategy poses its own challenges. First, existing Wi-Fi CSI acquisition schemes are not able to achieve fast channel hopping, hampering the realization of discrete channel sampling. Second, since each channel hopping may potentially introduces unknown interference, how to handle these interferences to ensure accurate and coherent data processing remains open. Third, though compressive sensing demands fewer samples, one still needs to figure out what is the sufficient sample quantity for a specific sensing application. Last but not least, the accuracy of Wi-Fi sensing suffers from various hardware-related random offsets (e.g., CPO or carrier phase offset) [39]; it is well-known that these offsets may significantly damage the quality of Wi-Fi sensing [29].

To tackle these challenges, we propose UWB-Fi, a novel system that leverages discrete and irregular channel sampling to expand Wi-Fi bandwidth and thus to achieve fine-granularity sensing. UWB-Fi first boasts a fast channel hopping scheme to realize arbitrary sampling within all Wi-Fi channels (2.4 GHz to 7.1 GHz)[2] on up-to-date 802.11ax commodity Wi-Fi devices (e.g., Intel AX210). As no existing signal processing technique can synthesize channel samples bearing unknown interference into a coherent "snapshot", we innovate in a model-based deep learning approach: it is guided by a ray-tracing model [69] 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 tradeoff* in parameter estimation, while removing hardware-related offsets inherent to Wi-Fi, thanks to the diversity offered by UWB sensing. We also calibrate the number of channel samples required by the compressive sensing principle under typical Wi-Fi sensing applications via extensive experiments. Finally, we implement a prototype of UWB-Fi using two laptops and conduct extensive experiments to evaluate its performance. In summary, our main contributions are:

- We propose UWB-Fi as the first Wi-Fi system to achieve *physical UWB sensing* using only discrete and irregular channel samples across nearly 5 GHz bandwidth.
- We design a fast channel hopping scheme for UWB-Fi to execute arbitrary channel sampling upon commodity Wi-Fi devices.
- We innovate in a model-based deep learning approach to address the challenge of synthesizing discrete channel samples bearing unknown interferences.
- We implement the first UWB sensing prototype on commodity Wi-Fi hardware, and conduct extensive evaluations on it to demonstrate UWB-Fi's centimeter-level fine-granularity sensing capability.

The rest of our paper is structured as follows. Section 2 introduces the background and motivation of UWB-Fi. Section 3 elaborates on the system design of UWB-Fi. Sections 4 and 5 respectively explain UWB-Fi's implementation and report the extensive evaluations on UWB-Fi. Related works and discussions of UWB-Fi's limitations and potentials are briefly captured in Section 6, followed by the conclusion of our paper in Section 7.

2 BACKGROUND AND MOTIVATION

In this section, we first establish Wi-Fi sensing basics on how bandwidth affects sensing granularity. We further explain why existing channel stitching approaches cannot realize UWB sensing. Finally, we point out that both the challenges and potentials of synthesizing discrete channel samples for UWB sensing: conventional signal processing techniques, though certainly incapable of directly synthesizing channel samples, may still help train a neural model for the same purpose.

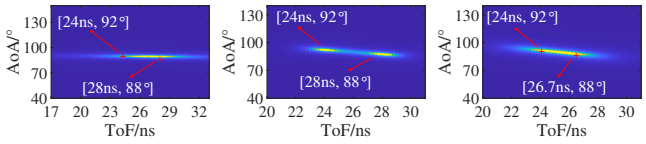
2.1 Wi-Fi Sensing Basics

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 distributed antennas of interval d, the received CSI $\boldsymbol{H} = [h_{n,m}]$ can be modeled as follows:

$$\begin{aligned} h_{n,m} &= \textstyle\sum_{k=1}^{K} \alpha_{n,m,k} \cdot h_{m,k}^{\mathrm{ToF}} \cdot h_{n,k}^{\mathrm{AoA}} \\ &= \sum_{k=1}^{K} \alpha_{n,m,k} \left(\mathrm{e}^{-j2\pi(f_c \pm (m-1)f_b)\tau_k}\right)\left(\mathrm{e}^{-j2\pi(n-1)d\cos(\theta_k)f_c/c}\right), \end{aligned} \tag{1}$$

[1] Leveraging frequency hopping to sample channels has been adopted by Chronos [51], yet its algorithm can only estimate the range between a pair of Wi-Fi devices, far from the general sensing purpose in our context.

[2] Specifically, the whole Wi-Fi 6 spectrum encompasses 2.4GHz band (2412 to 2484MHz), 5GHz band (5160 to 5885MHz), and 6GHz band (5935 to 7115MHz) [59].



(a) 20 MHz, 1.2 m. (b) 320 MHz, 1.2 m. (c) 320 MHz, 0.8 m.

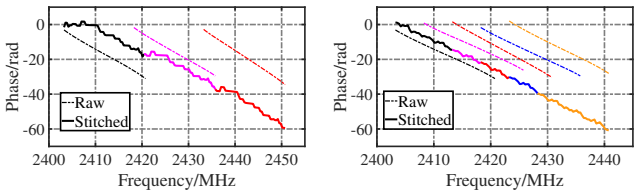
Figure 2: Range resolution performance for two subjects at different distances under various Wi-Fi bandwidths.

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, and c is the speed of light. According to [4], the temporal resolution $\Delta\tau = 1/B$, with B being the total sensing bandwidth, grows linearly in 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 $d\cos(\theta_k)/c$ (also a temporal component) yields a higher precision in estimating AoA. Therefore, these observations allow us to conclude that **the total sensing bandwidth assumes pivotal significance in realizing fine-granularity sensing.**

To illustrate this significance, we simulate the impact of Wi-Fi bandwidth on the ability to differentiate two subjects at varying distances, as depicted in Figure 2. With $B = 20$MHz and a separation of 1.2 m (Figure 2a), the ToF-AoA spectra are totally mixed without indicating the subjects' positions. Augmenting the bandwidth to 320 MHz (Figure 2b) allows the ToF-AoA spectra to exhibit two distinct peaks in ToF, thus enabling a sharp differentiation of the two subjects. However, the efficacy of 320 MHz sensing bandwidth wanes upon further reducing the separation between two subjects to 0.8 m (Figure 2c): as $B = 320$MHz only offers a range resolution of around 1 m, the ToF spectra get mixed again. Therefore, expanding the bandwidth to UWB range (GHz-level) is critical to achieve fine-granularity Wi-Fi sensing.

2.2 Infeasibility of Stitching Methods

To obtain a wider bandwidth, Splicer [65] and ToneTrack [67] conduct channel stitching attempting to merge continuous spectra. Unfortunately, the boundary effect (i.e., drastic roll-off) of the frequency response makes it impossible to stitch neighboring channels while preserving continuous phases, as depicted in Figure 3a. This has forced [65, 67] to use channels substantially overlapping with each other and sophisticated algorithms to align/calibrate these spectra, as shown in Figure 3b. However, as we shall demonstrate



(a) Adjacent channels. (b) Overlapped channels.

Figure 3: Stitching results given two channel arrangements: (a) adjacent channels and (b) overlapped channels.

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 4.7GHz spectrum from 2.4GHz to 7.1GHz stitched from multiple overlapping 20 MHz channels shifted every 5 MHz, which requires a total of 936 channels.

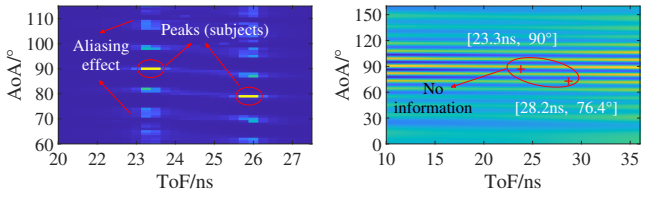
Technically speaking, all the channel samples should be acquired within *coherence time* to guarantee that they can be effectively synthesized. According to Chronos [51], the coherence time is approximately 84 ms. Meanwhile, sampling a single channel empirically demands at least 2ms. Consequently, at most $84/2 = 42$ 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 experience, stitching 5 channels, as shown in Figure 3b, takes over 64 ms on a 12-th Gen Intel(R) Core(TM) i5-12500H processor in a Lenovo Thinkbook 14 laptop. Extrapolating this result to the stitching of 936 channels results in an excessive duration of around 12 s. Therefore, it is imperative that innovative approaches are in place in order to successfully extend bandwidth towards UWB range.

2.3 Discrete Channel Sampling for Sensing

Expanding the sensing bandwidth B can be interpreted as increasing the number of subcarriers in Eqn. (1), offering greater information diversity for algorithms such as MUltiple SIgnal Classification (MUSIC) [47] to derive estimation of both range and bearing. In reality, as the physical phenomena captured by each snapshot are often sparse, the principle of compressive sensing [60] indicates that, instead of sampling every subcarrier within B, we can effectively acquire the fine-granularity sensing offered by B by sampling a few channels (hence the subcarriers within them). As a demonstration, we arbitrarily simulate 20 channel samples for two subjects, via Eqn. (1), within the range of 2.4–7.1 GHz and derive their respective ToF-AoA spectra[3] using MUSIC. Synthesizing these spectra in a maximum likelihood sense allows us to infer both ToFs and AoAs of the two subjects in Figure 4a.

Unfortunately, attempting to estimate ToF and AoA using real-life channel samples, as shown in Figure 4b, does not yield meaningful subject-related information. Apart from common noise and channel fading, the reason causing this failure is the unknown parameters introduced by the discrete sampling. In particular, the carrier phase offset (CPO) may randomly vary for every channel hopping. With a few tens of channel samples, it is not a surprise that the accumulated randomness can totally overwhelm the information embedded in these samples. Although a seemingly similar idea of channel hopping appeared in [51], the algorithm there only works for estimating parameters for the direct path between Tx and Rx. Therefore, it is surely incapable of handling our general UWB sensing tasks where multiple reflection paths (from sensing subjects) can be involved. In a nutshell, conventional signal processing approaches are quite unlike to be able to derive reliable estimation outcomes from discrete channel samples.

[3]The ToF-AoA spectrum, firstly adopted by SpotFi [29], is often the byproduct of maximum likelihood estimation on locations specified under polar coordinates.



(a) Ideal samples. (b) Real-life samples.

Figure 4: ToF-AoA spectra of discrete channel samples; the periodic pattern for each subject is caused by aliasing.

The main reason for the failure of conventional signal processing is that real-life channel samples contain far more parameters (some rather random) than ideal cases. Therefore, unless we have a tool that fully characterizes all such parameters, directly "translating" channel samples to ToF-AoA parameter tuples is almost impossible. One may think that a well-trained neural network (capable to approximate any function by the universal approximation theorem [19, 77]) can be a competent alternative, but Section 5.3 will prove that such a direct translation can be far from effective due to the bias-variance tradeoff in estimation [40]. Inspired by the fact that MUSIC outputs not scalar parameter tuples but rather parameter spectra shown in Figure 4a, we consider a function $\tilde{\mathbf{g}}(\cdot) : \hat{\boldsymbol{H}} \rightarrow \Xi$ mapping real-life channel samples $\hat{\boldsymbol{H}}$ to parameter spectra Ξ that imply ToF-AoA parameter tuples. Of course, since this function is still rather complex, comprehensively considering all factors to derive an explicit expression by conventional approaches is extremely challenging, if not totally impossible. Fortunately, we may now invoke the universal approximation theorem to approximate $\tilde{\mathbf{g}}$ with a neural network $\mathbf{g}$. In the following, we shall discuss how $\mathbf{g}$ is designed and trained in Section 3.2 and 3.3.

3 UWB-FI SYSTEM DESIGN

Our UWB-Fi is specifically designed to achieve fast channel hopping for meeting the coherence time budget and to cope with the inconsistency/interference introduced by discrete channel sampling. As illustrated in Figure 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.
- Model-based neural network: An encoder-decoder structure is adopted to emulate both a spectrum filter and a transformer from one spectral domain to another.
- Model-driven training strategy: Instead of using scalar tuples as ground truth labels, this strategy takes model-augmented labels with lifted dimension.

We present these components in Sections 3.1 to 3.3, and explain how to produce UWB sensing results by synthesizing the neural-inferred snapshots in Section 3.4.

3.1 Fast Channel Hopping Scheme

3.1.1 High-level Control Strategy. This strategy comprises four essential stages: channel selection, synchronized hopping, data acquisition, and information extraction. We have in total 97 Wi-Fi channels of 20 MHz bandwidth as our hopping candidates; they

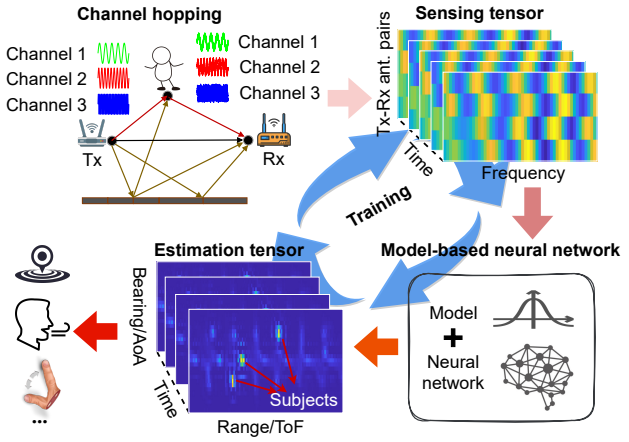
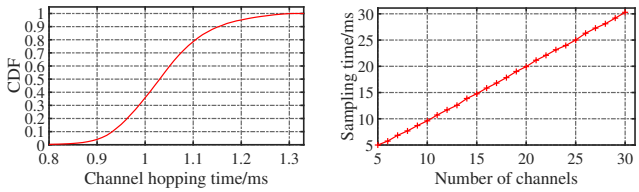


Figure 5: UWB-Fi system overview.

include 13 channels between 2412-2472 MHz, 8 channels from 5180-5320 MHz, 12 channels spanning 5500-5720 MHz, 5 channels within 5745-5825MHz, and 59 channels from 5955-7115MHz. Each hopping is initiated by the Tx that first checks the availability (not occupied by other co-channel devices or used for special purposes) 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 *independent* of 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. In typical indoor environments where 3-5 dominant multipath components (subjects) are commonly observed [14], we empirically set the total channel samples as 20 (also see Section 5.3.2). If there indeed exist more subjects, we can either increase the number of sampled channels progressively or increase the sample rate to accumulate more snapshots within the coherence time to equivalently obtain more samples.

3.1.2 Low-level Access Scheme. To accelerate channel hopping (from at least 2 ms [51] 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` [49], a special RAM-based file system in Linux kernel, to connect kernel and user spaces without kernel access constraints (e.g., `sysfs` [50] has strict one-value-per-file rules). To further lift the speed of accessing kernel space, we create two threads: `para_set`, and `data_proc`, respectively used to set hardware parameters (e.g., hopping channels introduced in Section 3.1.1) and to quickly move channel data from kernel to user space. We adopt an inter-process communication



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

Figure 6: Channel hopping time of UWB-Fi.

method (e.g., POSIX shared memory) to achieve fast data exchange among these two threads and upper-layer computation modules that use the retrieved channel samples for deriving sensing results (to be elaborated in Section 3.2). Although setting hardware parameters often requires resetting the Wi-Fi hardware (thus inevitably introducing a non-negligible latency), the drivers of certain Wi-Fi hardware [2, 23, 46, 70] offer direct register access and can hence avoid the latency caused by resetting.

To assess the effectiveness of the low-level access scheme, Figure 6a displays the cumulative distribution of the time required for one channel hopping, demonstrating rapid and consistent performance, with an average duration of around 1 ms. Additionally, we measure the total time needed to sample a various numbers of channels; as depicted in Figure 6b, this time linearly grows with the number of channels. In particular, collecting 20 channel samples (as specified in Section 3.1.1) requires only 20 ms. Given a coherence time of 84 ms [51], this 20 ms duration comfortably adheres to the coherence time budget. Furthermore, considering the data processing time (detailed in Section 3.4) 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.

3.2 Cross-Domain Spectral Transformation

The essential task for UWB-Fi to realize fine-granularity sensing, very similar to UWB radar sensing [9, 74], is to sharply identify multiple subjects (in terms of their respective multipath components) as clearly separated *bin*s in each snapshot. As we have explained in Section 2.3, directly mapping channel samples into ToF-AoA tuples can be infeasible; even if it is feasible, it is confined to pure localization purpose and fails to deliver outputs in the form of bins to further enable general sensing. Therefore, we choose to employ a neural network $\mathbf{g}$, aiming to transform discrete channel spectra in CSI domain towards ToF-AoA spectra in real domain. As our main contribution lies in training $\mathbf{g}$, we simply assemble a neural model $\mathbf{g}$ based on our experience [5, 15]. Consequently, we shall elaborate on the model design in the following, but refrain from evaluating it against other possible choices later.

3.2.1 Network Architecture. To handle cross-domain spectral transformation, our SpecTrans network $\mathbf{g}$ follows a classical *encoder-decoder* (ED) framework [11, 78]; it involves reducing data from an original domain into a lower-dimensional space and then reconstructing it into a new domain. While a fully convolutional structure is adopted to have an efficient implementation, we also bear in mind that SpecTrans design should be inspired by two model-based algorithms: i) a matched filter to remove hardware-related interferences/offsets and ii) the MUSIC algorithm to convert clean channel samples to ToF-AoA spectra. Therefore, we make use of the novel *transformer* architecture [52] to enhance the ED framework with trainable attention modules: these modules share similar algorithmic nature with those involved in the two model-based algorithms (e.g., the correlations used by MUSIC) but their weights are variables to be trained. Based on the illustration in Figure 7, we explain respective building block of SpecTrans in the following.

3.2.2 Basic Module. This module acts as a feature extractor and a trainable matched filter to remove hardware-related interferences/offsets from the extracted features. It involves a DenseNet [21] for feature extraction, and a SKNet [34] with attention-intensive structure as the matched filter to ensure precise estimation. Denoting DenseNet by $\mathbf{d}(\cdot)$ and SKNet by $\mathbf{s}(\cdot)$, the basic module can be represented as:

$$x_\ell^{\text{out}} = [x_\ell^{\text{in}}, \mathbf{s}_\ell(\mathbf{d}_\ell(x_\ell^{\text{in}})], \tag{2}$$

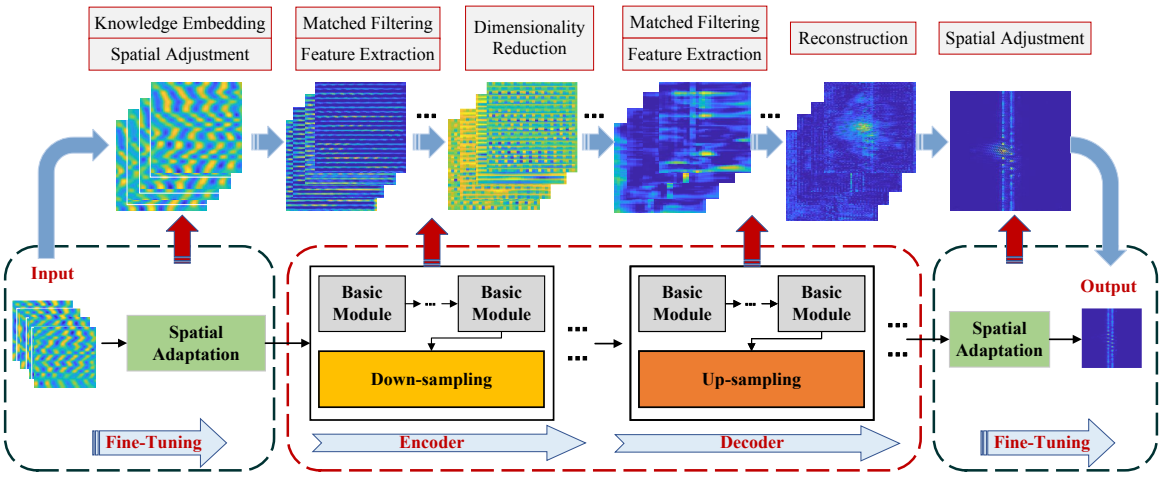


Figure 7: SpecTrans: Spectral transformation across two spectral domains.

where x_ℓ^{in} and x_ℓ^{out} indicate the input and output of the ℓ-th layer, respectively. As both modules $\mathbf{s}$ and $\mathbf{d}$ are only applied to produce part of the layer output, combining these filtered results with the layer input in the output ensures the concentration of critical channel information (for computational efficiency) while maintaining the availability of model gradient. To further avoid variance shift [33] during feature extraction and filtering, we specifically use batch normalization for regularization. Finally, for better control over the cross-domain transformation within the deep structure, the basic module keeps the size of the feature map intact.

3.2.3 Down- and Up-sampling. We utilize a fully convolutional structure for down- and up-sampling, aiming to fit the MUSIC process while avoiding information loss and hindered pixel-level regression caused by pooling layers. We utilize strided convolutions for down-sampling, which, combined with several basic modules, constitutes an *encoder*. The overall purpose of the encoder is to integrate features and reduce noise in channel samples, effectively mapping them into a lower-dimensional space. The *decoder* achieves up-sampling predominantly through transposed convolutions; they work in tandem with several basic modules to form the decoder. The decoder targets denoising and integrating data within the lower-dimensional space, while gradually reconstructing the higher-dimensional ToF-AoA spectra space. As the number of feature maps increases in the basic module, we avoid using dense connections in sampling in order to consistently reduce the quantity of feature maps.

3.2.4 Spatial Adaptation. Although UWB-Fi is largely immune to input variations across different environments thanks to its UWB nature (confirmed in Section 5.1.1), we consider the worst case scenarios where cross-environment variations become so prominent that they overwhelm the signals in the estimation results. To this end, we introduce *spatial adaptation* (SA) trained 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. As the sensing tensor (of channel samples) passes through the SA layer with a large quantity of feature maps at the entry point, it experiences spatial adjustments and incorporates task-specific information. On the contrary, another SA layer at the egress point is designed with only one feature map, primarily meant for fine-tuning the mapping relationship between the core network's output and the estimation tensor (i.e., ToF-AoA spectra). Fine-tuning the core network without altering its structure by spatial adaptation allows our overall design to gain sufficient versatility for adapting to diversified application scenarios.

3.3 Training with Data Augmentation

As our SpecTrans aims to effectively map 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 Figure 4a), 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 2.3). 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. Fortunately, as explained in Section 2.1, we can generate ideal channel samples by plugging the known ToF-AoA tuples into Eqn. (1), and then utilize MUSIC to derive the ToF-AoA spectra as labels. Moreover, training data should be endowed with diversified background multipath interference to train SpecTrans towards competent generalizability across distinct/unseen environments. Given limited real-life channel samples, we can augment them with model-generated multipath interference as background to significantly increase the dataset size (see Figure 8 bottom). In the following, we focus only on label generation, as generating background is rather straightforward.

According to Section 3.1.1, the fast channel hopping of UWB-Fi results in discrete and irregular channel samples, which creates another dilemma: how to produce ideal channel samples (for deriving ToF-AoA spectra labels) exactly corresponding to these random samples? In fact, 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 [10, 64], while expecting SpecTrans to be powerful enough for performing a sparse recovery. Of course, 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.

The whole label generation process is detailed in **Algorithm 1**, where the MUSIC algorithm is denoted as $\mathsf{MUSIC}(\cdot)$. Based on this, we further illustrate the self-supervised training strategy in Figure 8. Firstly, we generate ideal CSI data with different carrier frequencies based on Eqn. (1). We then employ MUSIC to perform joint estimation of ToF-AoA for each ideal channel sample and then fuse the spectra. The fused spectrum is further enhanced through exponential lifting and Laplacian filtering to obtain the final labels. To finally train the core ED network, we adopt the following objective function with an L2 penalty term on network weights:

$$\min_{w} \left(\sum \left(\mathrm{g}_w(\boldsymbol{H}) - \Xi \right)^2 + 0.5\lambda \sum w^2 \right), \tag{3}$$

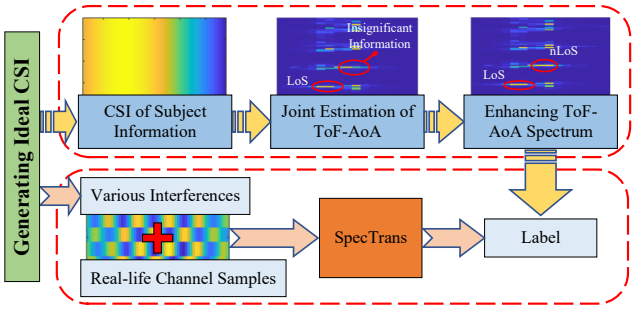


Figure 8: Self-supervised training strategy.

Algorithm 1: ToF-AoA Spectrum Generation

Input: Channel parameters $[\theta_k]$, $[\tau_k]$; a list $[f_i]$ of carrier frequencies and n, d, m, f_s, α as in Eqn. (1); enhancement factors: γ, x, y, and σ.
Output: ToF-AoA spectrum Ξ.

for $f_c \in [f_i]$ **do**
  $\boldsymbol{H} \leftarrow$ Eqn.(1) with channel parameters $[\theta_k]$, $[\tau_k]$;
  $\Xi_{\text{MUSIC}} \leftarrow \text{MUSIC}(\boldsymbol{H})$;
end
$\Xi \leftarrow \text{Enhance}\,(\Xi_{\text{MUSIC}}/\text{length}([f_i]), \gamma, x, y, \sigma)$;

Function $\text{Enhance}(\Xi, \gamma, x, y, \sigma)$**:**
  $\Xi \leftarrow |\Xi|^{\gamma}$;
  $z \leftarrow \left| -\frac{1}{\pi\sigma^4}\left(1 - \frac{x^2+y^2}{2\sigma^2}\right) e^{-\frac{x^2+y^2}{2\sigma^2}} \right|$;
  $\Xi \leftarrow \Xi \odot (\Xi * z)$;
  return Ξ;
end
Remarks: x, y and σ are parameters of a Laplacian filter [61], $*$ denotes convolution, and $\odot$ represents Hadamard product.

where, $\mathbf{g}_w(\cdot)$ represents the core network with weight w, Ξ denotes the labels, $\boldsymbol{H}$ refers to the input discrete channel samples, and λ is the coefficient of L2 penalty.

3.4 From Snapshots to General Sensing

As shown in Figure 5, UWB-Fi starts with the fast channel hopping scheme to collect discrete channel samples and assembles them into a sensing tensor (Section 3.1). UWB-Fi then leverages the trained SpecTrans network to estimate ToF-AoA spectra (Section 3.2) from the collected channel samples; this network is trained via model-augmented ToF-AoA spectra as labels and the corresponding training method (Section 3.3). At this point, UWB-Fi follows the convention to use a thresholding technique [44] 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 from polar coordinates (i.e., ToF-AoA tuples) to Cartesian coordinates to determine the subjects' locations [26]. Moreover, the phase information of a subject's motion is encoded into the multiple spectral amplitudes around the subject in the snapshot (confirmed in Section 5.3.1). 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 fine-granularity sensing applications with already established data processing procedures; these include gesture recognition (e.g., [36, 76]), activity recognition (e.g., [5, 15]), and vital signs monitoring (e.g., [8, 75]). 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 3.1.2) is still significantly shorter than the 12 seconds required for stitching channels as explained in Section 2.2, thus greatly enhancing real-time sensing capabilities.

4 PROTOTYPE AND EXPERIMENT SETUP

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

4.1 System Implementation

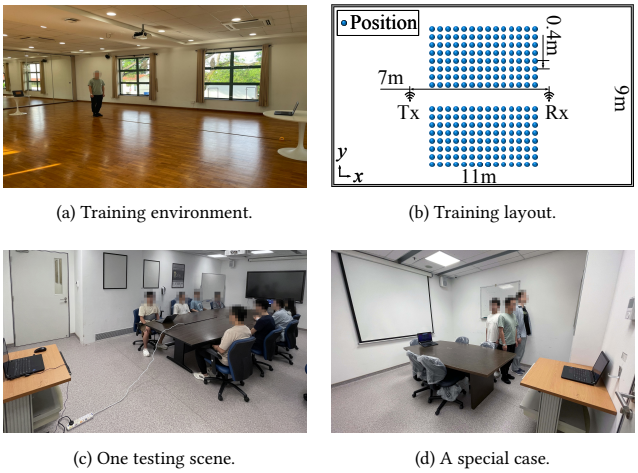
We use two Acer TravelMate laptops [3] equipped with Intel AX210 Wi-Fi NIC [12] to implement a UWB-Fi prototype. This Wi-Fi NIC has two antennas and supports all channels mentioned in Section 3.1.1. We implement the fast channel hopping scheme (Section 3.1) in the laptops to capture the channel samples. The captured channel samples are then parsed and converted into the matrix structure using MATLAB. Our SpecTrans network is implemented in the PyTorch 1.7.1 [42] environment. The encoder consists of 3 basic modules and one down-sampling module, while the decoder comprises 3 basic modules and one up-sampling module.

- A basic module contains a DenseNet with convolutional kernels of size 3, stride 1, and padding 1, as well as a SKNet with convolutional kernels of various sizes (1, 3, 5, and 7) to achieve multi-scale feature extraction.
- A sampling module consists of two non-densely connected convolutional layers. The first convolutional layer (shared by both down- and up-sampling modules) uses convolutional kernels of size 3, stride 1, and padding 1 to preserve the feature map size. The second convolutional layer leverages convolutional kernels of size 3, stride 2, and padding 1 and transposed convolutional kernels of the same parameter settings for up- and down-sampling modules, respectively.

The first SA module adopts a convolutional layer with kernels of size 5, stride 2, padding 2, and dilation 2, but another SA module following the decoder module takes a convolutional layer with kernels of size 3, stride 1, and padding 1 to ensure compatibility with the ToF-AoA spectra dimensions.

4.2 Experiment Setup

We recruit 8 subjects, 6 males and 2 females, aged between 20 and 30. We conduct experiments in six environments, including dance studio (DS), classroom (CR), library (LB), meeting room (MR), office (OF), and self-study room (SR). We collect our training data in DS and use data collected in other environments for testing purposes. Although our testing environments have diversified sizes, we set the distance between the two laptops to 7m in the training environment, which is proven to be generalizable to other environments in Section 5.1.1, thanks to the augmented data with diversified interference "background" (see Section 3.3). Besides comparison and ablation studies, we mainly perform three sets of experiments on multi-person sensing applications; these experiments have strictly followed the IRB of our institute.



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

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

Figure 9: Experiment setting examples: (a) training scene in the DS with (b) a layout containing positions marks; (c) a typical testing scene in the MR, and (d) a case with extremely short inter-subject distances in the SR.

Localization. We instruct one or multiple subjects to stand or sit in different positions in each experiment. Figure 9a shows the training environment in the DS and Figure 9b marks different positions for data collection: we take the gravity center of a subject as its ground truth position. We adopt one continuous channel stitching scheme [67] as our comparison baseline, where overlapped channels at 2.4 GHz band [1] are all collected for stitching a sensing bandwidth of 80 MHz. As no existing solution has the same UWB sensing bandwidth as UWB-Fi, we only get two barely comparable choices to serve as baseline [65, 67], while the other one [65] 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-based data processing nature. Subsequently, we conduct localization for the baseline according to its own proposed methodology. In order to exert the finest sensing granularity of UWB-Fi, we arrange a testing scene with three subjects standing within 0.4 m as shown in Figure 9d.

Respiration Monitoring. We choose one of the testing environments with 8 subjects sitting around a table in the MR (shown in Figure 9c) for this purpose. 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 [41]. The total recording time for respiration is 80 minutes. After obtaining the temporal sequence of ToF-AoA spectra produced by SpecTrans, we extract individual bin sequences corresponding respective subjects. Then we leverage the method proposed by MoRe-Fi [75] to predict respiratory waveforms.

Gesture Recognition. We again leverage the"8-subject in MR" scene here: the subjects are instructed to simultaneously perform six distinct hand gestures: push-pull (PP), up-down (UD), sweep (SW), draw circle (DC), draw zig-zag (DZ), and clap (CL). These six gesture classes serve as ground truth. Each gesture is performed 20 times. A data processing pipeline similar to the respiration monitoring task is adopted here, except that we employ the classifier proposed by Widar3.0 [76] to recognize gestures.

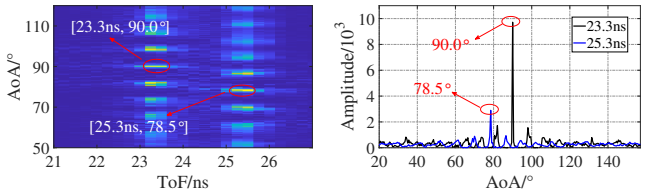
5 EVALUATIONS

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 verify UWB-Fi's cross-environment generalizability, and further report sensing results for multi-subject respiration monitoring and gesture recognition, demonstrating UWB-Fi's fine-granularity sensing performance. Finally, we perform an ablation study on our training approach, proving the effectiveness of our model-driven data augmentation process.

5.1 Localization and Comparison

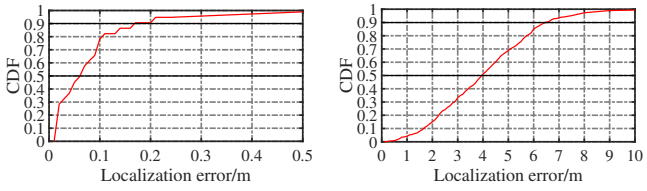
5.1.1 Localization Accuracy. Let us first use an example to illustrate the localization capability of UWB-Fi. As shown in Figure 10a, a ToF-AoA spectrum derived from channel samples via SpecTrans clearly exhibits two peaks (after filtering out the minor peaks caused by periodic aliasing, as explained in Section 3.4) corresponding to two ToF values of 23.3 ns and 25.3 ns, indicating two different propagation paths. In the meantime, Figure 10b 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 [16]. Recall that earlier proposals made great efforts to either get around or infer the information on Tx-Rx path [29, 51, 66], 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 3.4. Figure 11 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 (making it inapplicable to fine-grained sensing contexts). UWB-Fi's exceptional performance is primarily attributed to its capability in effectively acquiring ultra-wide bandwidth and eliminating hardware-related



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

Figure 10: SpecTrans output clearly differentiates two paths.

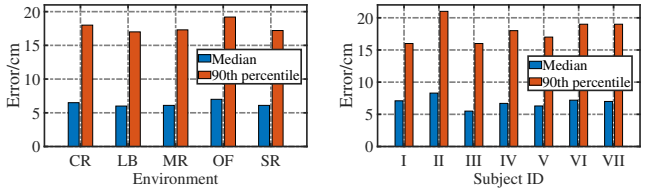


(a) UWB-Fi. (b) Baseline.

Figure 11: Contrasting UWB-Fi with baseline in localization accuracy: about 70 times lower in median error!

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 [6] offered by two coarse-grained observations.

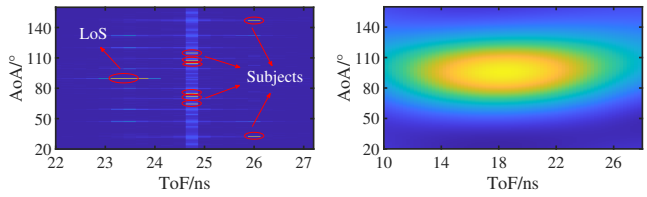
To showcase UWB-Fi's cross-environment and -subject generalizability, we evaluate the localization accuracy of all subjects across five unseen environments. The results, as depicted in Figure 12, confirm that varying environments and subject positions have a minimal impact on the localization performance: 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 SR that substantially deviates from our DS training environment. It is worth noting that this generalizability, though partially coming from the design of SpecTrans (without even invoking the SAs in Section 3.2.4), 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.



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

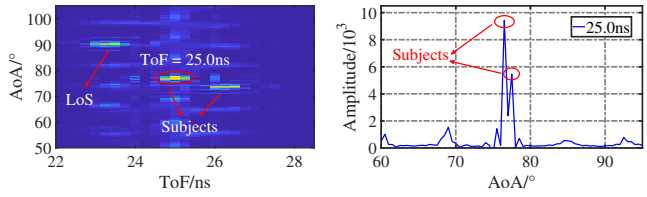
Figure 12: Localization median and 90th percentiles tail errors across various (a) environments and (b) subjects.

5.1.2 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 LB environment. Figures 13a and 13b 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 3.4) in the spectrum of Figure 13a, each representing distinct individuals and facilitating further sensing requirements. On the contrary, Figure 13b confirms that, when employing the baseline method for localization, the ToF-AoA spectrum is apparently incapable of discerning different subjects. To demonstrate the finest sensing granularity of UWB-Fi, we deliberately position three subjects in close proximity (see Figure 9d). The resulting ToF-AoA spectrum (shown in Figure 14a) reveals distinct peaks in ToF suggesting at



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

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



(a) ToF-AoA spectrum. (b) AoA-Amplitude curve.

Figure 14: Exerting the finest localization resolution.

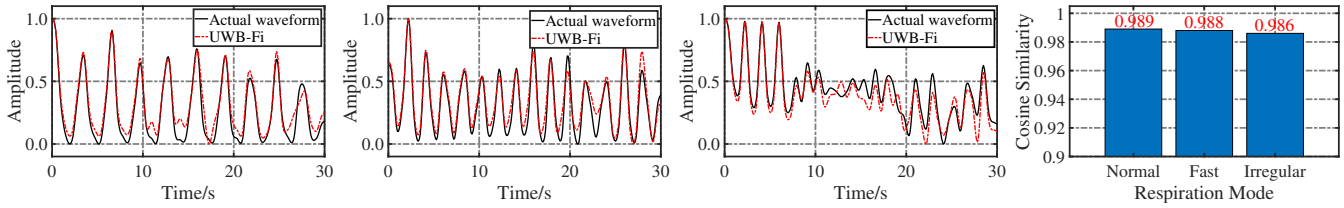
least two reflection paths. Further zooming in the sectional view at ToF = 25 ns (plotted in Figure 14b) indicates the presence of two separate AoA peaks allowing the distinction of three subjects. This result further demonstrates that UWB-Fi achieves a resolution of at least 1.25 ns for ToF and 1° for AoA, confirming its full capability in conducting fine-granularity multi-person sensing.

5.2 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.

5.2.1 Respiration Monitoring. We hereby demonstrate how UWB-Fi's output (ToF-AoA spectrum) can be used to acquire respiratory waveforms, using the setup described in Section 4.2. 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. Figures 15a and 15b present the recovered respiratory waveforms of these two situations; one can readily observe that the recovered waveforms by UWB-Fi effectively adapt to the variations taking place in reality, while all waveforms bear high morphological similarity to the ground truth according to Figure 15d.

To further explore the efficacy of waveform recovery under more extreme respiration situations, Figure 15c showcases waveform recovery in the presence of induced irregular variations in a subject's respiration rhythm. Even in such challenging circumstances, 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. This minor disparity can be attributed to reduced chest movements during this period that potentially



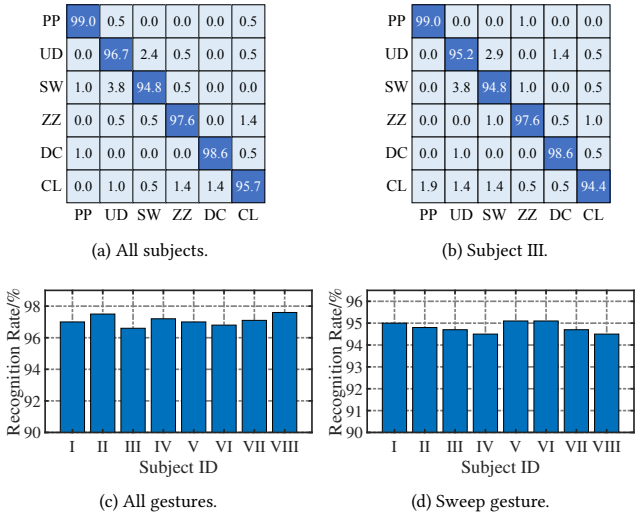
(a) Normal respiration. (b) Fast respiration. (c) Irregular respiration. (d) Cosine similarity.

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

lead to a low signal strength, which causes a tiny degradation in morphological similarity, as shown in Figure 15d.

5.2.2 Gesture Recognition. Again using the setup prescribed in Section 4.2, the performance of gesture recognition upon UWB-Fi's output can be evaluated. As shown by the confusion matrix for all subjects in Figure 16a, UWB-Fi achieves an overall recognition rate of 97.1%: the PP gesture has the highest recognition rate at 99.0%, while the UD and SW gestures exhibit more confusion. As both UD and SW 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 Figure 16c, the rates of all subjects stand around 97%.

Interestingly, subject III displays a slightly lower recognition rate of 96.6%; this brings us into a further analysis on subject III's recognition rates of individual gestures in Figure 16b. Different from Figure 16a, we find that the recognition rate for the CL gesture is lowest for subject III, rather than the expected UD and SW. This disparity may be attributed to the subject's physique that causes smaller ranges of motion. Finally, we specifically examine the easily confused SW gesture, with the results shown in Figure 16d. Since the recognition rates across various subjects hover between 94% and



(a) All subjects. (b) Subject III. (c) All gestures. (d) Sweep gesture.

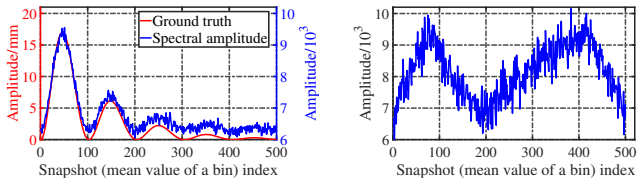
Figure 16: Confusion matrices for all gestures from (a) all subjects and (b) subject III, as well as recognition rates for all subjects in terms of (c) all gestures and (d) sweep gesture.

95%, the consistent lower rates appear to be irrelevant to individual subjects' inconsistent execution of the gesture.

5.3 Calibration and Ablation Study

In this section, we delve deeper into the sensing principle, impact of channel quantity, and training methods.

5.3.1 Sensing with Spectral Amplitude. Similar to real UWB sensing [9], 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 3.4) for both a pendulum and a finger gesture (writing the letter 'M'), and track their mean values, as depicted in Figure 17. Since continuous tracking such motions is impossible with only signal amplitude (which only infers range variation and has centimeter-level resolution even for UWB-Fi), the demonstrations in Figure 17 strongly confirm that continuous variations in phase have been represented by the spectral amplitudes of UWB-Fi's snapshots.

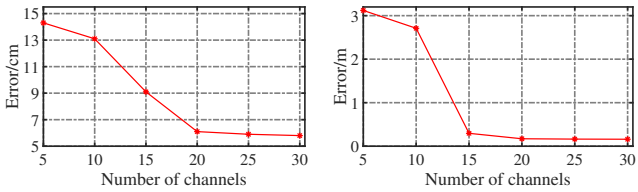


(a) Pendulum. (b) Finger gesture.

Figure 17: Sensing results of (a) a pendulum damping process and (b) a finger gesture writting letter 'M' in the air.

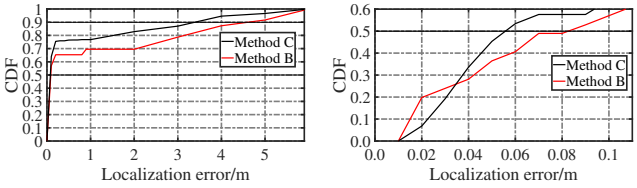
5.3.2 Channel Quantity Calibration and Channel Hopping Time. We hereby use experiment results to confirm our choice of 20 channel samples for UWB-Fi. We arrange 8 subjects arbitrarily standing in the DS and run UWB-Fi with the sampled channel quantity increasing from 5 to 30. As respectively shown in Figures 18a and 18b, the decreasing trends of median and the 90th percentile localization errors get saturated at around 20 sampled channels, confirming the rationale of our empirical choice. In fact, our experience with the same quantity in all testing environments further proves its efficacy, especially for the special case with close proximity among subjects (see Figure 14). Of course, these results might still be affected by conditions unknown to us by far, but we have offered adaptation schemes in Section 3.1.1.

It is worth noting that, when contrasting with Figure 11b, it becomes evident that, even with only 5 channels, UWB-Fi yields much higher accuracy than the baseline employing 13 overlapped channels. The reason accounting for this is twofold: i) UWB-Fi's



(a) Median errors. (b) 90th percentile errors.

Figure 18: Localization error vs. varying channel quantity.



(a) Localization error. (b) Zoom-in view.

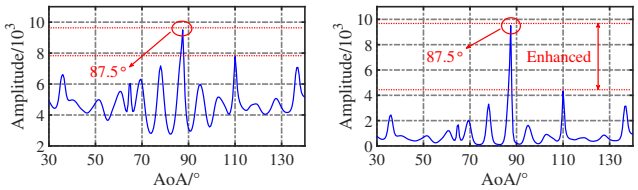
Figure 20: Method B and C's (a) localization error CDFs and their (b) zoom-in views.

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.

5.3.3 Training Approaches Ablation. In order to validate the effectiveness of our training approach proposed in Section 3.3, we compare it against three distinct label generation methods that include: A) generating a pseudo-spectrum directly from ToF-AoA tuples, B) estimating ToF-AoA spectra without enhancement, and C) generating labels without using the Tx-Rx path as a reference. 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 tradeoff in estimation [40]. Essentially, 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. Using higher-dimensional spectra as output (hence such labels for training) would allow for offloading the large variance to irrelevant parameters while achieving unbiased estimation with small variance on ToF-AoA pairs.

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. Denote the neural network taking such spectra as labels by $\mathbf{g}_A$, we train both $\mathbf{g}_A$ and $\mathbf{g}_w$ (see Section 3.3) and measure the entropy [48] of resulting networks; the results show that entropy values 253 and 132 for $\mathbf{g}_A$ and $\mathbf{g}_w$, respectively. Due to the higher complexity of $\mathbf{g}_A$, using labels generated by Method A for training certainly hampers convergence.

Method B. While generating labels using Algorithm 1, one might wonder if the enhancement (the Enhance$(\cdot)$ function) is necessary. According to Figure 19a, it appears that removing the enhancement



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

Figure 19: AoA-Amplitude curves (a) without and (b) with enhancement: the latter allows for a much better differentiation.

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 Figure 20a, the median error increases to around 8 cm, while the 90th percentile error even surpasses 3 m, which is obviously worse than our previous results in Figure 11a. With the enhancement, Figure 19b shows that the target peak becomes significantly more distinguishable compared to other "fake" ones, potentially improving the training outcome.

Method C. 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 Figure 20, 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 4 m. This negative effect can be explained by Figure 10b: in reality, the Tx-Rx path signal energy is much stronger than other paths; though it is probably not useful for sensing reflection paths, ignoring it outright could fundamentally change the "landscape" of a snapshot, 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.

6 RELATED WORK AND DISCUSSION

We discuss existing Wi-Fi sensing proposals by grouping them into two categories, based on whether a subject should carry a Wi-Fi device or not.

Device-based Sensing. As a seminal proposal, SpotFi [29] pioneered in the realm of device-based sensing. It leveraged MUSIC to estimate AoA rather accurately but offloaded estimation errors to ToF. In contrast, Chronos [51] estimated ToF by leveraging Wi-Fi channel hopping across both 2.4GHz and 5GHz bands. Its primary focus, however, lied in estimating parameters pertinent to the Tx-Rx path. Meanwhile, other endeavors, such as Splicer [65] and ToneTrack [67], also incorporated the information of multiple channels to enhance the accuracy of power delay profiles and relative ToFs resolution, respectively. It is worth noting, however, that the aforementioned device-based methods are predominantly centered around single-person sensing. The only exception is the recently

published MUSE-Fi [20], which capitalized on the near-field dominance of Wi-Fi sensing to achieve physical separation of multiple subjects, without the need for resorting to a ultra-wide bandwidth.

Device-Free Sensing. LiFS [54] pioneered in a device-free localization approach, making use of fine-grained CSI subcarrier. A later work PhaseBeat [58] harnessed root-MUSIC [45] to segregate signals multi-person sensing signals. MultiSense [71] treated multi-person sensing as a blind source separation issue, utilizing ICA [24] for waveform extraction. SPARCS [43] recovered micro-doppler spectra by exploiting the intrinsic sparsity of wideband mmWave channels. In addition to the above algorithmic attempts, other studies considered leveraging expanded hardware deployments for the sake of multi-person separation. Widar2.0 [44] extended partial support for multi-person sensing through the implementation with multiple antennas to enhance spatial resolution. Karanam *et al.* [27] exploited magnitude measurements derived from an array of receivers to execute multi-person tracking. Last but not least, Lan *et al.* [30] utilized metasurface antennas with varying beam patterns for multi-person activity recognition.

Discussions. Our UWB-Fi distinguishes itself as a system capable of accommodating both device-based and -free sensing modalities, thanks to its physically "widened (or diversified) view" in bandwidth with discrete channel samples. This capacity underscores the versatility of UWB-Fi and its capacity to adapt to diversified sensing contexts, strongly certifying its real-world applicability, especially towards leveraging Wi-Fi to perform sensing in a "crowd" manner [18, 55, 56, 73].

Although existing proposals have attempted to handle discrete channels for achieving wider bandwidth (e.g., [35, 38] for RFID and [13] for radar signals), Wi-Fi signals are fundamentally different in waveform and the existence of various hardware-related random offsets. Consequently, these solutions are not applicable to fuse discrete Wi-Fi channels. Additionally, these works typically focus on localization applications and lack the capability required for general sensing tasks supported by UWB-Fi.

As a system designed for general fine-granularity sensing applications, UWB-Fi can accommodate any number of subjects and exhibits sufficient cross-environment and cross-subject generalizability. This capability stems from UWB-Fi's ultra-wide bandwidth and the training procedure that aims to train UWB-Fi for separating a subject out of interference background, achieved by its model-based label generation. Because AX210 NIC is the only contemporary hardware available for 802.11ax-format CSI extraction in 6G channels, UWB-Fi has not be tested for its cross-hardware capability, which we leave as a future exploration with the evolution of Wi-Fi hardware.

Since UWB-Fi pioneers the physical UWB Wi-Fi sensing, there remains ample room for further exploration and refinement beyond the aforementioned cross-hardware generalizability. Our approach to discrete channel selection involves random sampling; though it is often deemed as the most natural strategy (and working fine according to our evaluation results), utilizing different channel selection strategies may further enhance UWB-Fi's performance. Therefore, identifying better channel placements could be further studied, especially when Wi-Fi has to co-exist with other wireless technologies [31, 32, 68]. In addition, as explained in Section 3.2, our SpecTrans model is empirically assembled (rather than optimally constructed) as a trainable function for fusing a UWB spectrum. Therefore, exploring neural model designs to improve UWB-Fi's fitting ability remains a valid future study.

7 CONCLUSION

We have introduced UWB-Fi to be a pioneering Wi-Fi sensing system operating with a 4.7GHz 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 versatile 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 discrete channel samples into parameter spectra precisely implying ToF-AoA attributes of individual subjects, effectively sidestepping the curse of bias-variance tradeoff. 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. The *research artifact*s accompanying this paper are available via http://doi.org/10.5281/zenodo.11093811.

ACKNOWLEDGEMENT

We are grateful to anonymous reviewers for their constructive suggestions. This research is supported in part by National Research Foundation (NRF) Future Communications Research & Development Programme (FCP) grant FCP-NTU-RG-2022-015 and MOE Tier 1 grant RG16/22.

REFERENCES

[1] 2021. *IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks–Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Redline.* IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016) - Redline. IEEE.

[2] Shahram Abdollahi-Alibeik, David Weber, Hakan Dogan, William W Si, Burcin Baytekin, Abbas Komijani, Richard Chang, Babak Vakili-Amini, MeeLan Lee, Haitao Gan, et al. 2011. A 65nm Dual-band 3-stream 802.11n MIMO WLAN SoC. In *2011 IEEE International Solid-State Circuits Conference*. IEEE, 170–172.

[3] Acer. 2023. Acer TravelMate. https://www.acer.com/sg-en/laptops/travelmate. Online; accessed 12 February 2023.

[4] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C Miller. 2014. 3D Tracking via Body Radio Reflections. In *Proc. of the 11th USENIX NSDI*. 317–329.

[5] Zhe Chen, Chao Cai, Tianyue Zheng, Jun Luo, Jie Xiong, and Xin Wang. 2023. RF-Based Human Activity Recognition Using Signal Adapted Convolutional Neural Network. *IEEE Trans. on Mobile Computing* 22, 1 (2023), 487–499.

[6] Zhe Chen, Zhongmin Li, Zhang Xu, Guorong Zhu Zhu, Yuedong Xu, Jie Xiong, and Xin Wang. 2017. AWL: Turning Spatial Aliasing From Foe to Friend for Accurate WiFi Localization. In *Proc. of the 13th ACM CoNEXT*. 238–250.

[7] Zhe Chen, Tianyue Zheng, Chao Hu, Hangcheng Cao, Yanbing Yang, Hongbo Jiang, and Jun Luo. 2023. ISACoT: Integrating Sensing with Data Traffic for Ubiquitous IoT Devices. *IEEE Communications Magazine* 61, 5 (2023), 98–104.

[8] Zhe Chen, Tianyue Zheng, and Jun Luo. 2021. MoVi-Fi: Motion-Robust Vital Signs Waveform Recovery via Deep Interpreted RF Sensing. In *Proc. of the 27th ACM MobiCom*. 392–405.

[9] Zhe Chen, Tianyue Zheng, and Jun Luo. 2021. Octopus: A Practical and Versatile Wideband MIMO Sensing Platform. In *Proc. of the 27th ACM MobiCom*. 601–614.

[10] Zhe Chen, Guorong Zhu, Sulei Wang, Yuedong Xu, Jie Xiong, Jin Zhao, Jun Luo, and Xin Wang. 2021. M^3: Multipath Assisted Wi-Fi Localization with a Single Access Point. *IEEE Trans. on Mobile Computing* 20, 2 (2021), 588–602.

[11] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. arXiv:1406.1078 [cs.CL]

[12] Intel Corporation. 2023. Intel® Wi-Fi 6 AX201. https://www.intel.sg/content/www/xa/en/products/sku/130293/intel-wifi-6-ax201-gig/specifications.html. Online; accessed 25 March 2023.

[13] Kevin M Cuomo, Jean E Pion, and Joseph T Mayhan. 1999. Ultrawide-Band Coherent Processing. *IEEE Trans. on Antennas and Propagation* 47, 6 (1999), 1094–1107.

[14] Nicolai Czink, Xuefeng Yin, Huseyin Ozcelik, Markus Herdin, Ernst Bonek, and Bernard H Fleury. 2007. Cluster Characteristics in a MIMO Indoor Propagation Environment. *IEEE Trans. on Wireless Communications* 6, 4 (2007), 1465–1475.

[15] Shuya Ding, Zhe Chen, Tianyue Zheng, and Jun Luo. 2020. RF-Net: A Unified Meta-Learning Framework for RF-enabled One-Shot Human Activity Recognition. In *Proc. of the 18th ACM SenSys*. 517–530.

[16] Andrea Goldsmith. 2005. *Wireless Communications*. Cambridge University Press, Cambridge, U.K.

[17] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool Release: Gathering 802.11n Traces with Channel State Information. *ACM SIGCOMM Comput. Commun. Rev.* 41, 1 (2011), 53.

[18] Kai Han, Chi Zhang, Jun Luo, Menglan Hu, and Bharadwaj Veeravalli. 2016. Truthful Scheduling Mechanisms for Powering Mobile Crowdsensing. *IEEE Trans. on Computers* 65, 1 (2016), 294–307.

[19] Andreas Heinecke, Jinn Ho, and Wen-Liang Hwang. 2020. Refinement and Universal Approximation via Sparsely Connected ReLU Convolution Nets. *IEEE Signal Processing Letters* 27 (2020), 1175–1179.

[20] Jingzhi Hu, Tianyue Zheng, Zhe Chen, Hongbo Wang, and Jun Luo. 2023. MUSE-Fi: Contactless MUti-person SEnsing Exploiting Near-field Wi-Fi Channel Variation. In *Proc. of the 29th ACM MobiCom*. 75:1–15.

[21] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely Connected Convolutional Networks. In *Proc. of the 30th IEEE/CVF CVPR*. 4700–4708.

[22] Jun Huang, Guoliang Xing, Gang Zhou, and Ruogu Zhou. 2010. Beyond Co-existence: Exploiting WiFi White Space for Zigbee Performance Assurance. In *Proc. of the 18th IEEE ICNP*. 305–314.

[23] Paul J Husted and Tao-fei Samuel Ng. 2006. Receiving and Transmitting Signals Having Multiple Modulation Types using Sequencing Interpolator. US Patent 7,098,821.

[24] Aapo Hyvärinen and Erkki Oja. 2000. Independent Component Analysis: Algorithms and Applications. *Neural Networks* 13, 4-5 (2000), 411–430.

[25] Wenjun Jiang, Hongfei Xue, Chenglin Miao, Wang Shiyang, Lin Sen, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun, and Lu Su. 2020. Towards 3D Human Pose Construction Using WiFi. In *Proc. of the 26th ACM MobiCom*. 23:1–14.

[26] Simon J Julier and Jeffrey K Uhlmann. 1997. Consistent Debiased Method for Converting Between Polar and Cartesian Coordinate Systems. In *Acquisition, Tracking, and Pointing XI*, Vol. 3086. Spie, 110–121.

[27] Chitra R. Karanam, Belal Korany, and Yasamin Mostofi. 2019. Tracking from one Side: Multi-person Passive Tracking with WiFi Magnitude Measurements. In *Proc. of the 18th IEEE IPSN*. 181–192.

[28] Katarzyna Kosek-Szott, Janusz Gozdecki, Krzysztof Loziak, Marek Natkaniec, Lukasz Prasnal, Szymon Szott, and Michal Wagrowski. 2017. Coexistence Issues in Future WiFi Networks. *IEEE Network* 31, 4 (2017), 86–95.

[29] Kotaru, Manikanta and Joshi, Kiran and Bharadia, Dinesh and Katti, Sachin. 2015. SpotFi: Decimeter Level Localization Using WiFi. In *Proc. of 29th ACM SIGCOMM*. 269–282.

[30] Guohao Lan, Mohammadreza F Imani, Philipp Del Hougne, Wenjun Hu, David R Smith, and Maria Gorlatova. 2020. Wireless Sensing using Dynamic Metasurface Antennas: Challenges and Opportunities. *IEEE Communications Magazine* 58, 6 (2020), 66–71.

[31] Feng Li, Jun Luo, Gaotao Shi, and Ying He. 2013. FAVOR: Frequency Allocation for Versatile Occupancy of Spectrum in Wireless Sensor Networks. In *Proc. of the 14th ACM MobiHoc*. 39–48.

[32] Feng Li, Jun Luo, Gaotao Shi, and Ying He. 2017. ART: Adaptive fRequency-Temporal Co-Existing of ZigBee and WiFi. *IEEE Trans. on Mobile Computing* 16, 3 (2017), 662–674.

[33] Xiang Li, Shuo Chen, Xiaolin Hu, and Jian Yang. 2019. Understanding the Disharmony Between Dropout and Batch Normalization by Variance Shift. In *Proc. of the 32nd IEEE/CVF CVPR*. 2682–2690.

[34] Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang. 2019. Selective Kernel Networks. In *Proc. of the 32nd IEEE/CVF CVPR*. 510–519.

[35] Bo Liang, Purui Wang, Renjie Zhao, Heyu Guo, Pengyu Zhang, Junchen Guo, Shunmin Zhu, Hongqiang Harry Liu, Xinyu Zhang, and Chenren Xu. 2023. RFChord: Towards Deployable RFID Localization System for Logistic Networks. In *Proc. of the 20th USENIX NSDI*. 1783–1799.

[36] Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: Ubiquitous Gesture Sensing with Millimeter Wave Radar. *ACM Trans. on Graphics* 35, 4 (2016), 1–19.

[37] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. 2015. Tracking Vital Signs During Sleep Leveraging Off-the-Shelf WiFi. In *Proc. of the 16th ACM MobiHoc*. 267–276.

[38] Yunfei Ma, Nicholas Selby, and Fadel Adib. 2017. Minding the Billions: Ultra-Wideband Localization for Deployed RFID Tags. In *Proc. of the 23rd ACM MobiCom*. 248–260.

[39] Yongsen Ma, Gang Zhou, and Shuangquan Wang. 2019. WIFI Sensing With Channel State Information: A Survey. *ACM Computing Surveys (CSUR)* 52, 3 (2019), 1–36.

[40] Kevin P Murphy. 2012. *Machine Learning: A Probabilistic Perspective*. MIT press.

[41] NeuLog. 2017. Respiration Monitor Belt Logger Sensor NUL-236. https://neulog.com/respiration-monitor-belt/. Online; accessed 12 February 2023.

[42] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv preprint arXiv:1912.01703* (2019).

[43] Jacopo Pegoraro, Jesus O. Lacruz, Michele Rossi, and Joerg Widmer. 2022. SPARCS: A Sparse Recovery Approach for Integrated Communication and Human Sensing in mmWave Systems. In *Proc. of the 21st ACM/IEEE IPSN*. 79–91.

[44] Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, Zheng Yang, and Yunhao Liu. 2018. Widar2.0: Passive Human Tracking with a Single Wi-Fi Link. In *Proc. of the 16th ACM MobiSys*. 350–361.

[45] B.D. Rao and K.V.S. Hari. 1989. Performance Analysis of Root-Music. *IEEE Trans. on Acoustics, Speech, and Signal Processing* 37, 12 (1989), 1939–1949.

[46] Sundar G Sankaran, Brian J Zargari, Lalitkumar Y Nathawad, Hirad Samavati, Srenik S Mehta, Alireza Kheirkhahi, Phoebe Chen, Ke Gong, Babak Vakili-Amini, Justin A Hwang, et al. 2009. Design and Implementation of a CMO 802.11n SoC. *IEEE Communications Magazine* 47, 4 (2009), 134–143.

[47] R. Schmidt. 1986. Multiple Emitter Location and Signal Parameter Estimation. *IEEE Trans. on Antennas and Propagation* 34, 3 (1986), 276–280.

[48] Claude Elwood Shannon. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27, 3 (1948), 379–423.

[49] The Kernel Development Community. 2009. DebugFS. https://docs.kernel.org/filesystems/debugfs.html. Online; accessed 18 August 2023.

[50] The Kernel Development Community. 2009. Sysfs. https://www.kernel.org/doc/html/next/filesystems/sysfs.html. Online; accessed 18 August 2023.

[51] Deepak Vasisht, Swarun Kumar, and Dina Katabi. 2016. Decimeter-Level Localization with a Single WiFi Access Point. In *Proc. of the 13th USENIX NSDI*. 165–178.

[52] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Proc. of the 31st ACM NIPS*. 6000–6010.

[53] Aditya Virmani and Muhammad Shahzad. 2017. Position and Orientation Agnostic Gesture Recognition Using WiFi. In *Proc. of the 15th ACM MobiSys*. 252–264.

[54] Ju Wang, Hongbo Jiang, Jie Xiong, Kyle Jamieson, Xiaojiang Chen, Dingyi Fang, and Binbin Xie. 2016. LiFS: Low Human-Effort, Device-Free Localization with Fine-Grained Subcarrier Information. In *Proc. of the 22nd ACM MobiCom*. 243–256.

[55] Jin Wang, Jun Luo, Sinno Jialin Pan, and Aixin Sun. 2019. Learning-Based Outdoor Localization Exploiting Crowd-Labeled WiFi Hotspots. *IEEE Trans. on Mobile Computing* 18, 4 (2019), 896–909.

[56] Jin Wang, Nicholas Tan, Jun Luo, and Sinno Jialin Pan. 2017. WOLoc: WiFi-only outdoor localization using crowdsensed hotspot labels. In *Proc. of the 36th IEEE INFOCOM*. 1–9.

[57] Xuanzhi Wang, Kai Niu, Jie Xiong, Bochong Qian, Zhiyun Yao, Tairong Lou, and Daqing Zhang. 2022. Placement Matters: Understanding the Effects of Device Placement for WiFi Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 1 (2022), 32:1–25.

[58] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. PhaseBeat: Exploiting CSI Phase Data for Vital Sign Monitoring with Commodity WiFi Devices. In *Proc. of the 37th IEEE ICDCS*. 1230–1239.

[59] Wikipedia. 2023. List of WLAN Channels. https://en.wikipedia.org/wiki/List_of_WLAN_channels. Online; accessed 23 July 2023.

[60] John Wright and Yi Ma. 2010. Dense Error Correction Via ℓ^1-Minimization. *IEEE Trans. on Information Theory* 56, 7 (2010), 3540–3560.

[61] Aming Wu and Cheng Deng. 2023. Discriminating Known From Unknown Objects via Structure-Enhanced Recurrent Variational AutoEncoder. In *Proc. of the 36th IEEE/CVF CVPR*. 23956–23965.

[62] Liu Xiang and Jun Luo. 2011. Compressed Data Aggregation for Energy Efficient Wireless Sensor Networks. In *Proc. of the 8th IEEE SECON*. 46–54.

[63] Liu Xiang, Jun Luo, and Catherine Rosenberg. 2013. Compressed Data Aggregation: Energy-Efficient and High-Fidelity Data Collection. *IEEE/ACM Trans. on Networking* 21, 6 (2013), 1722–1735.

[64] Xiufeng Xie, Eugene Chai, Xinyu Zhang, Karthikeyan Sundaresan, Amir Khojastepour, and Sampath Rangarajan. 2015. Hekaton: Efficient and Ractical Large-Scale MIMO. In *Proc. of the 21st ACM MobiCom*. 304–316.

[65] Yaxiong Xie, Zhenjiang Li, and Mo Li. 2015. Precise Power Delay Profiling with Commodity Wi-Fi. In *Proc. of the 21st ACM MobiCom*. 53–64.
[66] Yaxiong Xie, Jie Xiong, Mo Li, and Kyle Jamieson. 2019. mD-Track: Leveraging Multi-Dimensionality for Passive Indoor Wi-Fi Tracking. In *Proc. of the 25th ACM MobiCom*. 8:1–16.
[67] Jie Xiong, Karthikeyan Sundaresan, and Kyle Jamieson. 2015. ToneTrack: Leveraging Frequency-Agile Radios for Time-Based Indoor Wireless Localization. In *Proc. of the 21st ACM MobiCom*. 537–549.
[68] Ruitao Xu, Gaotao Shi, Jun Luo, Zenghua Zhao, and Yantai Shu. 2011. MuZi: Multi-Channel ZigBee Networks for Avoiding WiFi Interference. In *Proc. of the 4th IEEE/ACM CPSCOM*. 323–329.
[69] Zhengqing Yun and Magdy F Iskander. 2015. Ray Tracing for Radio Propagation Modeling: Principles and Applications. *IEEE Access* 3 (2015), 1089–1100.
[70] Masoud Zargari, Lalitkumar Y Nathawad, Hirad Samavati, Srenik S Mehta, Alireza Kheirkhahi, Phoebe Chen, Ke Gong, Babak Vakili-Amini, Justin A Hwang, Shuo-Wei Mike Chen, et al. 2008. A Dual-band CMOS MIMO Radio SoC for IEEE 802.11n Wireless LAN. *IEEE Journal of Solid-State Circuits* 43, 12 (2008), 2882–2895.
[71] Zeng, Youwei and Wu, Dan and Xiong, Jie and Liu, Jinyi and Liu, Zhaopeng and Zhang, Daqing. 2020. MultiSense: Enabling Multi-Person Respiration Sensing with Commodity WiFi. In *Proc. of the 22nd UbiComp*. 102:1–29.
[72] Chi Zhang, Feng Li, Jun Luo, and Ying He. 2014. iLocScan: Harnessing Multipath for Simultaneous Indoor Source Localization and Space Scanning. In *Proc. of the 12th ACM SenSys*. 91—104.
[73] Chi Zhang, Jun Luo, and Jianxin Wu. 2014. A Dual-Sensor Enabled Indoor Localization System with Crowdsensing Spot Survey. In *Proc. of the 10th IEEE/ACM DCOSS*. 75–82.
[74] Tianyue Zheng, Zhe Chen, Shuya Ding, and Jun Luo. 2021. Enhancing RF Sensing with Deep Learning: A Layered Approach. *IEEE Communications Magazine* 59, 2 (2021), 70–76.
[75] Tianyue Zheng, Zhe Chen, Shujie Zhang, Chao Cai, and Jun Luo. 2021. MoRe-Fi: Motion-robust and Fine-grained Respiration Monitoring via Deep-Learning UWB Radar. In *Proc. of the 19th ACM SenSys*. 111–124.
[76] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. 2019. Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi. In *Proc. of the 17th ACM MobiSys*. 313–325.
[77] Ding-Xuan Zhou. 2020. Universality of Deep Convolutional Neural Networks. *Applied and Computational Harmonic Analysis* 48, 2 (2020), 787–794.
[78] Yinhao Zhu and Nicholas Zabaras. 2018. Bayesian Deep Convolutional Encoder–Decoder Networks for Surrogate Modeling and Uncertainty Quantification. *J. Comput. Phys.* 366 (2018), 415–447.