MUSE-Fi: Contactless MUti-person SEnsing Exploiting Near-field Wi-Fi Channel Variation

Jingzhi Hu[1*] Tianyue Zheng[1*] Zhe Chen[2] Hongbo Wang[1] 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
Email: {jingzhi.hu, tianyue002, hongbo001, junluo}@ntu.edu.sg, zhechen@fudan.edu.cn

ABSTRACT

Having been studied for more than a decade, Wi-Fi human sensing still faces a major challenge in the presence of multiple persons, simply because the limited bandwidth of Wi-Fi fails to provide a sufficient range resolution to physically separate multiple subjects. Existing solutions mostly avoid this challenge by switching to radars with GHz bandwidth, at the cost of cumbersome deployments. Therefore, *could Wi-Fi human sensing handle multiple subjects* remains an open question. This paper presents MUSE-Fi, the first Wi-Fi multi-person sensing system with physical separability. The principle behind MUSE-Fi is that, given a Wi-Fi device (e.g., smartphone) very close to a subject, the *near-field* channel variation caused by the subject significantly overwhelms variations caused by other distant subjects. Consequently, focusing on the channel state information (CSI) carried by the traffic in and out of this device naturally allows for physically separating multiple subjects. Based on this principle, we propose three sensing strategies for MUSE-Fi: i) uplink CSI, ii) downlink CSI, and iii) downlink beamforming feedback, where we specifically tackle signal recovery from sparse (per-user) traffic under realistic multi-user communication scenarios. Our extensive evaluations clearly demonstrate that MUSE-Fi is able to successfully handle multi-person sensing with respect to three typical applications: respiration monitoring, gesture detection, and activity recognition.

CCS CONCEPTS

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

* Both authors contributed equally to this research.

ACM MobiCom'23, October 2–6, 2023, Madrid, Spain

ACM ISBN 978-1-4503-9990-6/23/10.
https://doi.org/10.1145/3570361.3613290

KEYWORDS

Wi-Fi human sensing, multi-person sensing, ISAC, respiration monitoring, gesture detection, and activity recognition.

ACM Reference Format:
J. Hu, T. Zheng, Z. Chen, H, Wang, and J. Luo. 2023. MUSE-Fi: Contactless MUti-person SEnsing Exploiting Near-field Wi-Fi Channel Variation. In *The 29th Annual International Conference on Mobile Computing and Networking (ACM MobiCom'23), October 2–6, 2023, Madrid, Spain.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3570361.3613290

1 INTRODUCTION

Since we were first able to obtain CSI (channel state information) in certain Wi-Fi devices [21], *Wi-Fi human sensing* [25, 35, 44, 56, 58, 59, 67, 72] has been attracting significant attention from both academia and industry. During the past decade or so, many applications of Wi-Fi human sensing have been developed, notably including vital signs monitoring [35, 67], gesture detection [56, 72], activity recognition [15, 25], as well as localization and motion tracking [14, 44, 58, 68]. Whereas such sensing applications have a promising potential to be integrated with the ubiquitously deployed Wi-Fi communication infrastructures, they all face a major obstacle in conducting realistic *multi-person sensing: the limited Wi-Fi bandwidth fails to offer a sufficient range resolution to distinguish different sensing subjects.*

Because Wi-Fi communication does not seem to embrace a super-wide bandwidth due to its contention-based multi-access nature,[1] existing sensing proposals often avoid its limitation by resorting to radars with a GHz-level bandwidth [3, 70], yet radar sensing is somehow inferior to Wi-Fi sensing as it demands extra deployments. In order to continue exploiting Wi-Fi's potential in integrated sensing and communication (ISAC) [12, 22], two makeshifts are often adopted. On one hand, many distributed antennas can be used to achieve enhanced spatial resolution for separating subjects [44], at the cost of messing up with the Wi-Fi communication functions. On the other hand, signal processing techniques for separating five subjects at the CSI level have been attempted [67] without offering guaranteed separability in general [69, 70].

[1]The 320MHz bandwidth of Wi-Fi 7 [29] only yield a meter-level range resolution, yet leading to insufficient sampling rates due to frame aggregation.

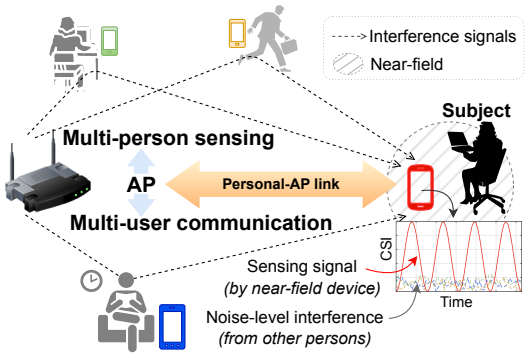


Figure 1: While each personal device *uniquely* identifies a person, the sensing signal (upon the person) offered by the identifying device within *near-field* overwhelms the interference from other persons.

In reality (as in Figure 1), each person often has its own wearable Wi-Fi devices, typically a smartphone or even a smartwatch. Although the communication link between such a personal device and the nearby Wi-Fi access point (AP) is deemed as the basic sensing media by earlier proposals on Wi-Fi human sensing, those proposals aim to leverage either a single link to perform sensing [35, 44, 59, 67] or multiple links to offer a slightly improved spatial resolution [25, 43, 58, 72]. They all neglect two fundamental factors in such a realistic multi-user communication setting shown in Figure 1: i) each personal-AP link uniquely identifies the human subject to be sensed, and ii) since the subject is within the *near-field* (less than 0.2m in range) of its own Wi-Fi device, the channel variation caused by its motions to its personal-AP link could be so strong as to push the interference from other subjects down to the noise floor. In other words, the default multi-user communication setting of Wi-Fi does offer the potential to be naturally extended to multi-person sensing, if one can properly integrate sensing into communication.

From application perspective, such near-field multi-person sensing naturally supports various functions under the pervasive deployment of Wi-Fi infrastructure. As these functions include sensing vital signals, gestures, activities, and locations, they are especially applicable to eXtended Reality (XR). In particular, integrating gestures and activities recognition into Wi-Fi communication reduces the peripheral sensors, leading to lighter and less power-consuming virtual reality (VR) and merged reality (MR) headsets, making them more desirable for long-time wearing [4, 54]. Furthermore, the environmental and human sensing results indicate key contextual and localization information of nearby human and object motions; overlaying such information on the top of real-world visions facilitates augmented reality (AR) and MR applications in intrusion detection, patient monitoring, and machine status assessing [10, 34].

Nonetheless, integrating multi-person sensing with multi-user communication is highly non-trivial, as existing practices, exploiting only artificial Wi-Fi traffic for the sensing purpose, barely offer any experience. In practice, multi-user scenarios typically cause a much lower and very irregular frame arrival rate per link, thanks to the contention-based multi-access nature of Wi-Fi. Since the CSI carried by each frame is a critical channel state sample for Wi-Fi sensing, a lower and irregular frame rate indicates a lower and irregular sampling rate, which may significantly confine the usability of Wi-Fi sensing. As most Wi-Fi sensing applications have been developed upon a high and regular frame rate (up to 1000 frame/s [25, 44, 67]), this challenge, crucial to seamless integration of multi-person sensing with multi-user communication for Wi-Fi, has never been seriously tackled.

To address these challenges, we propose MUSE-Fi as a novel MUlti-person SEnsing system leveraging Wi-Fi. To motivate MUSE-Fi, we first theoretically characterize and experimentally verify the *dominating effect* in near-field Wi-Fi sensing, upon which we develop criteria on the physical separability of multiple subjects. Based on the theoretical characterizations, we propose three sensing strategies for MUSE-Fi to be integrated with the traffic cross each personal-AP link, namely exploiting i) uplink (to AP) CSI, ii) downlink (from AP) CSI, and iii) downlink BFI (beamforming feedback information) [9]. For all strategies, we propose a *sparse recovery algorithm* (SRA) to mask the potential variance in frame rate; it aims to regulate the input samples so as to deliver a unified data flow to later processing pipelines for respective sensing functions. In addition, we study the sensing effectiveness of these strategies by contrasting the BFI-enabled compressive sensing with conventional CSI-based Wi-Fi sensing. Our key contributions can be summarized as follows:

- We propose MUSE-Fi as the first true multi-person Wi-Fi sensing system; it integrates multi-person sensing with multi-user communication in a seamless manner.
- We, for the first time, expose the dominating effect of near-field Wi-Fi sensing; it is exploited by MUSE-Fi to achieve physical separation of multiple subjects.
- We design three sensing strategies for MUSE-Fi and equip them with an SRA to mask the variance in frame rate.
- We reveal the pros and cons of BFI-enabled Wi-Fi sensing against the conventional CSI-enabled one.
- We implement MUSE-Fi prototype and evaluate it with extensive experiments. The promising results confirm that MUSE-Fi indeed supports multi-person Wi-Fi sensing under realistic scenarios.

The rest of the paper is organized as follows. Section 2 discusses the dominating effect of near-field sensing both theoretically and experimentally. Section 3 presents the sensing strategies for MUSE-Fi, along with the crucial SRA to

regulate the frame rate. Section 4 specifies how the MUSE-Fi prototype is implemented and how the application scenarios for case studies are set up. Section 5 reports the evaluation results of three case studies. Related works are briefly discussed in Section 6. Finally, Section 7 concludes our paper.

2 SENSING BY NEAR-FIELD DOMINATION

In this section, we introduce the Wi-Fi human sensing basics, and systematically study and validate the dominating signal variations in near-field sensing. Compared with conventional antenna near-field and capacitive coupling [8, 20] not developed for practical multi-person sensing, our theoretical analyses allow for characterizing the feasible region of near-field sensing and shedding insights on the upper/lower bounds of subject number and spacing.

2.1 Wi-Fi Human Sensing Basics

We start by introducing a Wi-Fi sensing system with an AP and *user equipment* (UE) pair aiming to sense the physical motion of a human *subject* denoted by $\mathcal{S}$. At time t, denote the distance between the AP and $\mathcal{S}$ by $d_{A,\mathcal{S}}(t)$ and the distance between $\mathcal{S}$ and the UE by $d_{\mathcal{S},E}(t)$. Further focusing on the influence of $\mathcal{S}$, we model the wireless channel gain between the AP and the UE as:

$$h_{A,E}(t) = h_{A,\mathcal{S},E}(t) + h^S_{A,E} + h^D_{A,E}(t), \tag{1}$$

where $h^S_{A,E}$ and $h^D_{A,E}(t)$ represent the static and dynamic channel gains between the AP and UE due to, respectively, the direct communication path and interfering motions along it, and $h_{A,\mathcal{S},E}(t)$ indicates the channel gain from the AP to UE via the reflection of $\mathcal{S}$, which can be expressed as:

$$h_{A,\mathcal{S},E}(t) = \frac{\lambda^2\sqrt{G_{A,\mathcal{S},E}}\exp(-\mathrm{i}2\pi(d_{A,\mathcal{S}}(t)+d_{\mathcal{S},E}(t))/\lambda)}{(4\pi)^2(d_{A,\mathcal{S}}(t)d_{\mathcal{S},E}(t))^{\alpha/2}}, \tag{2}$$

where λ is the carrier wavelength, $G_{A,\mathcal{S},E}$ represents the product of Tx and Rx antenna gains and the reflection coefficient of $\mathcal{S}$, and α is the path loss exponent [18]. Typically, $\alpha \in [2,4]$ with $\alpha \approx 4$ for indoor environments [47]. Therefore, Wi-Fi human sensing can be described as follows: the physical motion of a human subject results in changes of $d_{A,\mathcal{S}}(t)$ and $d_{\mathcal{S},E}(t)$, which in turn lead to the changes of channel gain $h_{A,\mathcal{S},E}(t)$ over time. Therefore, by analyzing the time series of $h_{A,\mathcal{S},E}(t)$ obtained from the CSI of the Wi-Fi frames, both AP and UE are able to sense the motion of $\mathcal{S}$.

2.2 Feasible Region for Near-field Sensing

Consider a more general scenario where two persons exist in the Wi-Fi sensing system. Without loss of generality, we let one of them be the subject $\mathcal{S}$, and refer to the other as the *interferer* denoted by $\mathcal{I}$. Consequently, the channel gain between the AP and UE in Eqn. (1) becomes:

$$\tilde{h}_{A,E}(t) = h_{A,\mathcal{S},E}(t) + h_{A,\mathcal{I},E}(t) + h^S_{A,E} + h^D_{A,E}(t), \tag{3}$$

where $h_{A,\mathcal{I},E}(t)$ is the channel gain from the AP to UE via the reflection of $\mathcal{I}$; it can be modeled in a similar manner as Eqn. (2). Eqn. (3) seems to suggest that it is hard to separate the channel influences imposed by $\mathcal{S}$ and $\mathcal{I}$ since their channel gains get mixed up. Nevertheless, we point out that, in the Wi-Fi sensing scenarios where $\mathcal{S}$ is close to or in the near-field of UE (i.e., distance below 0.2 m, empirically), the variation of the channel gain is dominated by the $\mathcal{S}$'s physical motion; in other words, $\partial|h_{A,\mathcal{S},E}(t)|/\partial t \gg \partial|h_{A,\mathcal{I},E}(t)|/\partial t$. We term this phenomenon **near-field domination** effect, and we provide its theoretical analysis as follows.

Firstly, to quantify the variation of $h_{A,\mathcal{S},E}(t)$, we evaluate it by the squared amplitude of the partial derivative of $h_{A,\mathcal{S},E}(t)$ w.r.t. t, which is referred to as the *power of channel variation*. To simplify the analysis, we assume $\partial d_{A,\mathcal{S}}(t)/\partial t = \partial d_{\mathcal{S},E}(t)/\partial t = v_{\mathcal{S}}$. The value of $v_{\mathcal{S}}$ can be interpreted as the *intensity* of $\mathcal{S}$'s motion in terms of speed. The power of channel variation of $\mathcal{S}$ can then be calculated as:

$$\begin{aligned} P_{\mathcal{S}} &= \left|\frac{\partial h_{A,\mathcal{S},E}(t)}{\partial t}\right|^2 \\ &\approx \frac{G_{A,\mathcal{S},E}\lambda^4 v_{\mathcal{S}}^2}{(4\pi)^4(d_{A,\mathcal{S}}d_{\mathcal{S},E})^{\alpha}}\left[\frac{\alpha^2}{4}\left(\frac{d_{A,\mathcal{S}}+d_{\mathcal{S},E}}{d_{A,\mathcal{S}}d_{\mathcal{S},E}}\right)^2 + \frac{16\pi^2}{\lambda^2}\right] \\ &\overset{(\star)}{\approx} \tilde{G}_{A,\mathcal{S},E}\cdot v_{\mathcal{S}}^2\cdot(d_{A,\mathcal{S}}d_{\mathcal{S},E})^{-\alpha}, \end{aligned} \tag{4}$$

where we omit symbol t in the distance notations and let $\tilde{G}_{A,\mathcal{S},E} = (\lambda/4\pi)^2 G_{A,\mathcal{S},E}$ for the sake of brevity. In the second row of Eqn. (4), the first term inside the bracket is caused by the amplitude variation of the channel gain and the second term results from the phase variation of the channel gain. ($\star$) holds because, in typical 5 GHz Wi-Fi sensing systems with $\mathcal{S}$ in the near-field of UE (e.g., $d_{A,\mathcal{S}} \sim 3$ m, $d_{\mathcal{S},E} \sim 0.1$ m, and $\lambda \sim 0.06$ m), the first term in the bracket is much smaller than the second term and thus can be omitted, implying that the channel variation is mainly due to that $\mathcal{S}$'s motion changes the phase of the channel.

The power of variation of $h_{A,\mathcal{I},E}(t)$ can be similarly derived for $\mathcal{I}$ as $P_{\mathcal{I}} = \tilde{G}_{A,\mathcal{I},E}v_{\mathcal{I}}^2(d_{A,\mathcal{I}}d_{\mathcal{I},E})^{-\alpha}$, with $d_{A,\mathcal{I}}$ and $d_{\mathcal{I},E}$ being the distance between the AP and $\mathcal{I}$ and between $\mathcal{I}$ and the UE, respectively, and $v_{\mathcal{I}}$ being the intensity. Consequently, the near field domination effect can be interpreted as $P_{\mathcal{S}}$ being significantly larger than $P_{\mathcal{I}}$, thanks to $d_{\mathcal{S},E}$ being much smaller than $d_{\mathcal{I},E}$, given $d_{A,\mathcal{S}} \approx d_{A,\mathcal{I}}$ and $v_{\mathcal{S}} \approx v_{\mathcal{I}}$. It is worth noting that assuming $v_{\mathcal{S}} \approx v_{\mathcal{I}}$ may not be practical, because human sensing to different targets may have distinct meaning (e.g., respiration monitoring against gesture detection). Though our following derivation shall stick to this symmetric assumption for the ease of exposition, we will experimentally validate that the near-field domination

still holds even with asymmetric intensities, as far as there is a sufficient discrepancy between $d_{\mathcal{S},E}$ and $d_{\mathcal{I},E}$.

In order to concretely characterize the interferer's feasible region that maintains the domination of $P_{\mathcal{S}}$ at the UE, we propose an novel metric **variation to interference ratio** (VIR); it evaluates the variation power ratio between $h_{A,\mathcal{S},E}(t)$ and the sum of $h_{A,\mathcal{I},E}(t)$ and dynamic channel $h^D_{A,E}(t)$. Based on [59], $h^D_{A,E}(t)$ can be also treated as an interference, whose power P_d is in linear proportion to that of the static channel gain. Therefore, assuming a LoS path between the AP and UE, we have $P_d = \eta\lambda^2 {d_{A,E}}^{-\alpha} + b$, where η and b are fixed for a given pair of AP and UE. Then, we have:

$$\mathrm{VIR}_{\mathcal{S}} = \frac{P_{\mathcal{S}}}{P_{\mathcal{I}}+P_d} = \frac{v_{\mathcal{S}}^2 \tilde{G}_{A,\mathcal{S},E}(d_{A,\mathcal{S}} d_{\mathcal{S},E})^{-\alpha}}{\eta\lambda^2 d_{A,E}^{-\alpha} + b + v_{\mathcal{I}}^2 \tilde{G}_{A,\mathcal{I},E}(d_{A,\mathcal{I}} d_{\mathcal{I},E})^{-\alpha}}. \quad (5)$$

Intuitively, the feasible region of $\mathcal{I}$ is indicated by $\mathrm{VIR}_{\mathcal{S}}$ value being greater than a threshold γ_{th}.

To deliver more visual insights, we illustrate the feasible region of $\mathcal{I}$ for $\gamma_{\mathrm{th}} = 50$ in Figure 2, given $v_{\mathcal{S}}, v_{\mathcal{I}}, \gamma, b, \tilde{G}_{A,\mathcal{S},E}$, and $\tilde{G}_{A,\mathcal{I},E}$ being normalized to 1 and $\alpha = 4$. It can be observed (e.g., by the small infeasible circular regions around $\mathcal{S}$ and the AP) that the separation distance between $\mathcal{S}$ and $\mathcal{I}$ can be potentially short without resulting in poor VIR for $\mathcal{S}$, given $\mathcal{I}$ is not close to the AP, too. Moreover, as $\mathcal{I}$ is also a potential subject of the Wi-Fi sensing system, it needs to be treated symmetrically to $\mathcal{S}$, i.e., $\mathcal{I}$ is also interfered by $\mathcal{S}$, so $\mathrm{VIR}_{\mathcal{I}}$, similarly characterized as $\mathrm{VIR}_{\mathcal{S}}$ in Eqn. (5), should also be dominating at $\mathcal{I}$'s UE, meaning $\mathrm{VIR}_{\mathcal{I}} < \gamma_{\mathrm{th}}$ when $d_{A,\mathcal{I}}$ being large is infeasible too. Therefore, besides the small regions mentioned earlier, a bigger one around the whole system indicates the whole boundary of $\mathcal{I}$'s feasible region.

Although these results are obtained numerically and serve for indicative purpose only, the resulting feasible region for a single interferer $\mathcal{I}$ establishes the physical guarantee that the channel variation due to $\mathcal{S}$ can be well separated from that due to $\mathcal{I}$, and vice versa. It can also be observed from Figure 2 that the contours of $\mathrm{VIR}_{\mathcal{S}}$ constitute a set of

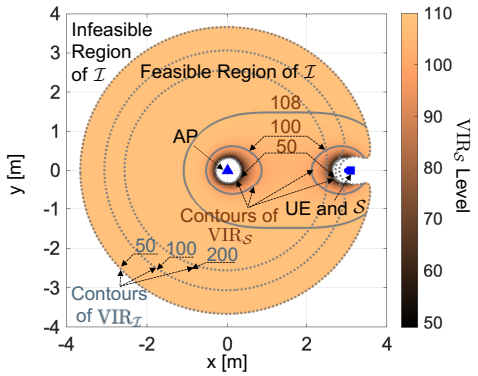


Figure 2: Feasible region of $\mathcal{I}$ and the contours of VIR levels for both $\mathcal{S}$ and $\mathcal{I}$.

Cassini ovals around the AP and $\mathcal{S}$, which may appear to be similar to the experimental results in [59]. Nevertheless, our VIR is derived from the perspective of channel variation rather than from the SNR metric adopted in [59], which evaluates the ratio between the power of signals reflected by the subject and the noises. We strongly believe our channel variation analysis is far more relevant to sensing applications, as the physical information of a subject is represented by the channel variation rather than the signal strength.

2.3 Insights into Multi-person Scenarios

Now we are ready to extend the analysis in Section 2.2 to multi-person scenarios, where N subjects ($N \geq 3$) are using the Wi-Fi sensing system and each of them stays in the near-field of its UE. Though we could extend Eqn. (5) to multi-person case by putting N-fold interference in the denominator, the resulting characterization would be too general to shed any immediate lights on sensing performance, because the distribution patterns of these subjects have infinite possibilities. To this end, we consider two symmetric distribution cases, aiming to address two important questions separately: i) how many subjects can a Wi-Fi sensing system support? and ii) how close can adjacent subjects be? To simplify the presentation, we no longer distinguish between the positions of the subject and its UE.

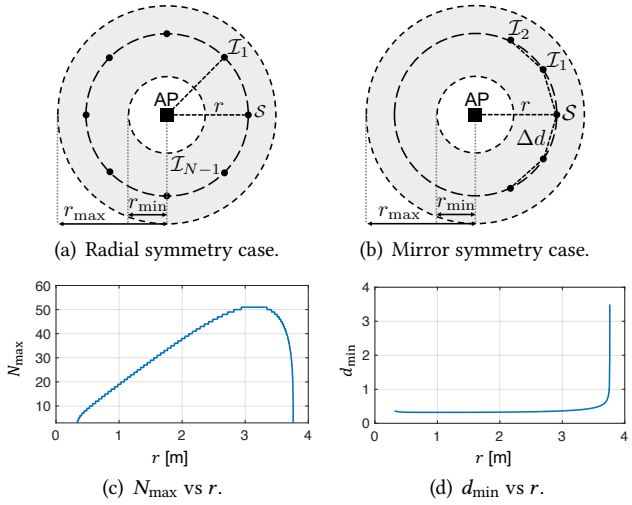
To answer question i), we analyze the case where N subjects stay at distance r from the AP and are uniformly spaced, as shown in Figure 3(a). Due to the radial symmetry of their positions, we can focus on analyzing any of them, which is again named $\mathcal{S}$. Extending Eqn. (5), we have:

$$\mathrm{VIR}^{(\mathrm{i})}_{\mathcal{S}} = \frac{\tilde{G}(\Delta r)^{-\alpha}}{\eta\lambda^2 + br^{\alpha} + (2r)^{-\alpha}\tilde{G}\cdot\sum_{j=1}^{N-1}\sin^{-\alpha}(j\cdot\pi/N)}, \quad (6)$$

where $\tilde{G}$ represents the identical values of $\tilde{G}_{A,\mathcal{S},E}$ and $\tilde{G}_{A,\mathcal{I}_j,E}$ ($\forall j \in \{1,\ldots N-1\}$), and Δr denotes the short distance from $\mathcal{S}$ to its UE. Generally, the series summation in Eqn. (6), $\sum_{j=1}^{N-1}\sin^{-\alpha}(j\cdot\pi/N)$, has no closed-form expression w.r.t. N. Fortunately, given $N \in [3, 60]$ and $\alpha \in [2, 4]$, the series summation can be numerically fitted by a function in the form of $p_1 N^{p_2} + p_3$ with R-square ≈ 1, where parameters p_1, p_2, and p_3 are dependent on α: $p_1 = 0.0230$, $p_2 = 3.99$, $p_3 = 38.0$ for $\alpha = 4$. Now given γ_{th}, the **upper bound** on the number of subjects that can be accommodated becomes:

$$N_{\max} = \left\lfloor \left(\frac{(2r)^{\alpha}}{p_1} \cdot \frac{\tilde{G}(\Delta r)^{-\alpha} - \eta\lambda^2\gamma_{\mathrm{th}} - br^{\alpha}\gamma_{\mathrm{th}}}{\tilde{G}\gamma_{\mathrm{th}}} - \frac{p_3}{p_1} \right)^{1/p_2} \right\rfloor. \quad (7)$$

Moreover, based on (7), we can also derive the minimum and maximum distances (resp. $r_{\min}$ and $r_{\max}$) between $\mathcal{S}$ and the AP for the considered case to be feasible by solving the inequality $N_{\max} \geq 3$ in the field of real number. As shown



(a) Radial symmetry case. (b) Mirror symmetry case.

(c) $N_{\max}$ vs r. (d) $d_{\min}$ vs r.

Figure 3: Two symmetric cases considered for multi-person sensing scenarios. (a)&(c) N subjects uniformly spaced and (b)&(d) $2K+1$ subjects closely located.

in Figure 3(c), $N_{\max}$ first increases and then decreases in r; it reaches its maximum 51 when $r \in [2.94, 3.35]$ m.

To answer question ii), we consider the case as shown in Figure 3(b), where $N = 2K+1, K = 1, 2, 3, \ldots$. Denote the distance and angle between each pair of neighboring subjects by Δd and ϕ, we have $\phi = 2 \cdot \arcsin(\Delta d/(2r))$. It is clear that the middle subject $\mathcal{S}$ suffers from the worst interference (among all subjects) when $2\pi - (2K+1)\phi > 0$, equivalent to $\Delta d < 2r\sin(\pi/(2K+1))$. Therefore, given $\Delta d < 2r\sin(\pi/(2K+1))$, the condition for $\mathrm{VIR}_{\mathcal{S}} > \gamma_{\mathrm{th}}$ determines the low bound of Δd. Using Eqn. (5) again, we have:

$$\mathrm{VIR}_{\mathcal{S}}^{(\mathrm{ii})} = \frac{\tilde{G}\Delta r^{-\alpha}}{\eta\lambda^2 + br^{\alpha} + 2\tilde{G}(2r)^{-\alpha}\sum_{j=1}^{K}\sin^{-\alpha}(j\phi/2)}. \quad (8)$$

Again, the series summation $\sum_{j=1}^{K}\sin^{-\alpha}(j\phi/2)$ can be fitted by function $q_1(\sin(\phi/2))^{q_2} + q_3$ given $\alpha \in [2,4]$, $K \in [1,10]$, and $\phi \in [\pi/180, \pi/(2K+1)]$ with R-square ≈ 1, where parameters q_1, q_2, and q_3 depend on K and α: $q_1 = 1.06$, $q_2 = -4$, and $q_3 = 6.57$ for $\alpha = 4$ and $K = 2$. Consequently, we obtain the **lower bound** of Δd as follows:

$$\Delta d_{\min} = 2r\left(\frac{(2r)^{\alpha}}{q_1}\cdot\frac{\tilde{G}(\Delta r)^{-\alpha} - \eta\lambda^2\gamma_{\mathrm{th}} - br^{\alpha}\gamma_{\mathrm{th}}}{2\tilde{G}\gamma_{\mathrm{th}}} - \frac{q_3}{q_1}\right)^{\frac{1}{q_2}}. \quad (9)$$

By solving $\Delta d_{\min} \le 2r\sin(\pi/(2K+1))$ in the field of real number, we can obtain the boundary for the distance between the AP and the subjects, i.e., $r_{\min}$ and $r_{\max}$. Figure 3(d) shows that $\Delta d_{\min}$ remains around 0.34 m for $r \in [0.32, 3.30]$ m but increases steeply to 3.49 m for $r \in [3.30, 3.76]$ m.

2.4 Proof-of-Concept Pre-Experiments

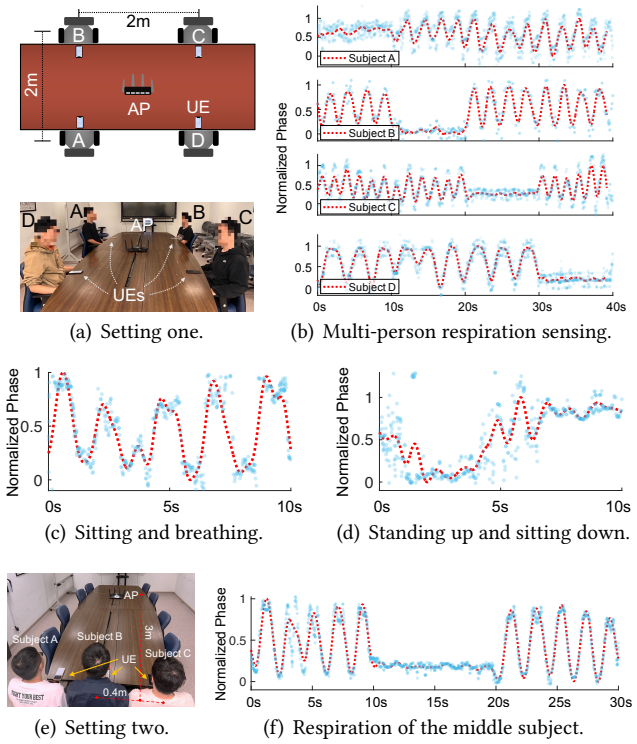
We first conduct two preliminary experiments to briefly validate the theoretical analysis in Eqn. (7). We deploy a Wi-Fi 5 AP in the middle of a table, serving 4 users spaced 2 m apart around the table, as depicted in Figure 4(a). Each *user* in this paper involves a UE and a subject placed 15 cm apart from each other. We leverage the iPerf3 tool [55] to emulate the 4 UEs and the AP constantly exchanging data frames with each other for 40 seconds, during which the four subjects take turns holding their breath. Their irregular CSI sequences (blue points) are obtained by the uplink traffic from the UEs to the AP, and the outcomes after the processing of MUSE-Fi are shown as red (dotted) curves in Figure 4.

Our results in Figure 4(b) show that the respirations of the 4 subjects never interfere with each other, indicating an effective multi-person sensing. Though the theoretical results in Figure 3(c) suggest that up to 25 users can be supported with $r = 1.41$ m, our preliminary experiments are conducted in a more conservative manner, as the theoretical results are obtained under ideal conditions without user asymmetry. This setting also leaves room for us to validate asymmetric sensing, where we let three subjects breathe normally while the Subject D performs the activity of standing up and sitting down. For brevity, we only show the sensing results for one of the breathing subjects and Subject D in Figures 4(c) and 4(d), respectively; these results clearly demonstrate that multi-person sensing can still be successfully achieved even under asymmetry scenarios. In our case studies later, more users will be involved to better confirm the effectiveness of near-field domination for MUSE-Fi.

We conduct another experiment to validate the case of close proximity between users in Eqn. (9). We let three subjects sit side by side with 40 cm distance between the centers of gravity of the neighboring ones, as shown in Figure 4(e). Our results in Figure 4(f) focus on the centered subject (the most interfered one), whose respiratory (held or not) can be clearly sensed regardless of the behaviors of others. These results evidently validate the effectiveness of feasibility of near-field domination on Wi-Fi multi-person sensing even under close proximity between neighboring subjects.

3 SHAPING-UP MUSE-FI

Given the physical guarantees on multi-person separation through near-field domination, we hereby introduce the detailed components and sensing strategies of MUSE-Fi. In hardware, MUSE-Fi is comprised of a commodity AP and multiple UEs owned by subjects demanding sensing services from the system; all Wi-Fi devices follow the prevalent Wi-Fi standard and their CSIs can be readily obtained from the received frames [19, 26]. In the following, we first introduce three sensing strategies adopted by MUSE-Fi, along with



(a) Setting one. (b) Multi-person respiration sensing.

(c) Sitting and breathing. (d) Standing up and sitting down.

(e) Setting two. (f) Respiration of the middle subject.

Figure 4: Preliminary experiments. (a) Setting for the first two experiments. (b) Multi-person respiration sensing in action. (c) and (d) Multi-person asymmetric sensing with both respiration and activity. (e) Setting for the 3rd experiment. (f) Respiration (held or not) from the middle subject, i.e., Subject B in (e).

their practical issues. We then specifically investigate two critical issues faced by these strategies.

3.1 Three Sensing Strategies for MUSE-Fi

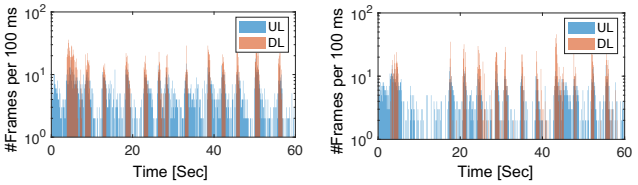
Since CSIs are carried by Wi-Fi frames, MUSE-Fi has three sensing strategies based on the traffic direction and how CSIs are carried. In particular, they are:

- **UL-CSI Sensing**: The uplink (UL) traffic and the CSI, obtained from the long training sequence (LTS) carried in the preamble of a frame, are adopted as sensing primitives.
- **DL-CSI Sensing**: The downlink (DL) traffic and the carried CSI are adopted as sensing primitives.
- **UL-BFI Sensing**: The UL traffic and the carried BFI are adopted as sensing primitives.

Apparently, the entity to handle sensing (data processing) depends on the direction of data flow: both UL-CSI and UL-BFI require sensing to take place on the AP side, while DL-CSI demands the UE to handle sensing. Here, BFI is actually a compressed version of the DL-CSI but carried by UL traffic for the AP to be aware of the DL channel conditions, so as to fine-tune its MIMO precoding; it only becomes available since IEEE 802.11ac standard [37]. As BFI is transmitted with *action frames* (part of UL traffic) in plain form, sensing can also take place at Wi-Fi devices capable of overhearing the traffic; the incurred privacy issue will be further discussed in one of our companion works.

3.1.1 User Registration. When a *user* (a subject with its UE) demands access to MUSE-Fi, it should first announce its presence along with the sensing application it requires. This user registration process is necessary for three reasons: i) it lets MUSE-Fi be aware of the number of users and their respective motion types, so as to coordinate users more accurately (e.g., reject users if system capacity is reached), ii) it prepares MUSE-Fi with prior information to fine-tune its processing pipelines, such as selecting different filters according to motion intensity or involving algorithmic modules to handle excessive interference if the average motion intensity is too high, and iii) it preserves privacy for normal Wi-Fi users with no intention to access MUSE-Fi. In the following, we will focus only on the registered users, leaving detailed registration procedure in our extended report.

3.1.2 Practical Issues. In stark contrast to existing Wi-Fi sensing systems working with artificially generated continuous traffic and mostly with only a single link, the most prominent challenge for MUSE-Fi is to handle the bursty and intermittent traffic in practice. Specifically, due to the multi-user communication infrastructure and the contention-based medium access mechanism on which MUSE-Fi is based, both UL and DL traffic exhibit bursty and intermittent characteristics, leading to CSI time series being sparse with many discontinuous parts. To illustrate this, we depict the frame arrival rates for both UL and DL when one or two users watch 1080p videos using their respective UEs in Figure 5. Both UL and DL traffics already exhibit bursty and intermittent patterns for one user, caused by upper-layer protocols' data caching and rate control [62]; these are exacerbated by channel contentions even with only one additional user, as shown in Figure 5(b). Moreover, the BFI is contained only in a small portion of UL traffic, and its sample rate is about 1/10 of DL frames, peaking at roughly 10 frames per second,



(a) One User. (b) Two Users.

Figure 5: Frame arrival rates in terms of number of frames per 100 ms versus the observation time when (a) one or (b) two users stream 1080p videos.

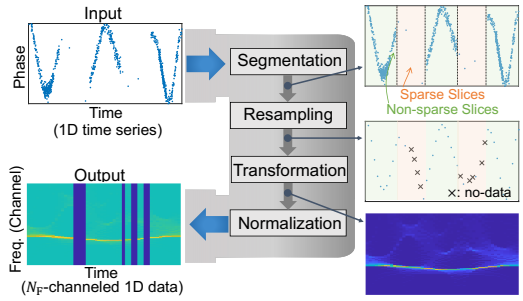


Figure 6: Data transformation pipeline of MUSE-Fi.

which means the UL-BFI sensing strategy faces the severest data sparsity. Consequently, MUSE-Fi needs to be capable of recovering continuous channel variation from a CSI time series with sparse samples.

3.2 Sparse Recovery Algorithm

We propose an SRA for MUSE-Fi to recover continuous channel variation from the intermittently sparse samples due to realistic traffic. The SRA is comprised of two components: a *data transformation pipeline* to pre-process a sparse CSI sequence sampled under realistic traffic, and a *self-supervised data recovering network* to recover the densely sampled CSI sequence. The core novelty of SRA lies in eliminating the extensive label collection, by interpreting the correlations between sparse and non-sparse data slices. Without loss of generality (of three sensing strategies), we denote the input data to SRA by a 1D time series $\{x_t\}$ drawn from the CSI of specific antenna pair and subcarrier, where t is the sampling time and $x_t \in \mathbb{R}$ is the phase of a corresponding CSI sample. SRA outputs $\{\boldsymbol{y}_t\}$ as an evenly and densely sampled multi-channel sequence where $\boldsymbol{y}_t \in \mathbb{R}^{N_F}$ represents the N_F frequency components of the CSI for time t. This output can be used directly as the sensing result, or be further processed to recognize the activity or gesture of the subject. We elaborate on the two SRA components in the following, .

3.2.1 Data Transformation Pipeline. This four-step pipeline transforms $\{x_t\}$ into an evenly resampled output sequence $\{\hat{\boldsymbol{x}}_n \in [-1,1]^{N_F}\}_{1\le n\le N_s}$ with length N_s as the total number of resampled time instants, which is shown in Figure 6.

Segmentation. We first segment the time series into two types of slices, i.e., *sparse slices*, where the samples are sparse in time, and *non-sparse slices*, for them to be treated differently. This segmentation is done by using a sliding window of length Δt in time to check whether it contains a sufficient number of samples. In particular, time slices with more than N_{nsp} samples are marked as *non-sparse*; otherwise as *sparse*. Here Δt and N_{nsp} are parameters specified by sensing applications and are empirically set in Section 4.

Resampling. The time series is resampled so that the samples can be evenly spaced in time, facilitating further denoising and sparse recovery. Within each non-sparse slice, the outliers are removed, and an interpolation is performed to meet a resampling frequency f_{rs}. Each sparse slice is resampled with frequency f_{rs}, with samples moved to their nearest resampled time instants; those time instants without data are tagged as "`no-data`" and filled linearly. The result is denoised through a low-pass filter with cut-off frequency f_{cut}. Both f_{rs} and f_{cut} are empirically specified in Section 4.

Transformation. This step transforms the resampled time series into its *spectrogram* $\{\tilde{\boldsymbol{x}}_n \in \mathbb{R}^{N_F}\}_{1\le n\le N_s}$ with N_F frequency components. The reason behind this is that motions of subjects generally lead to channel variations whose patterns are environment- and subject-specific and hardly recognizable in the time domain. By transforming it into a spectrogram, the impact of subject's motion on the channel becomes more apparent, facilitating effective sparse recovery.

Normalization. Mapping each $\tilde{\boldsymbol{x}}_n$ into $\hat{\boldsymbol{x}}_n \in [0,1]^{N_F}$ via min-max normalization allows for focusing on the relative variation pattern while eliminating the magnitude difference of channel variation potentially caused by subject positions. Besides, the frequency components of time instants with `no-data` tags are assigned value -1, making them distinct from those with data and clearly indicating the data sparsity.

3.2.2 TCN-based Sparse Slice Recovering. Rather than employing a heavy neural network like U-Net for audio inpainting applications [28], we adopt a temporal convolutional network (TCN) based autoencoder (AE) to achieve sparse recovery, involving fewer parameters to make it efficient to train and deploy in resource-limited devices as UEs and APs. TCNs are superior to other types of neural networks (e.g., LSTMs) as they exploit convolutional layers with dilated kernels to capture ultra long-range dependencies in samples while maintaining a manageable number of parameters [7].

Network Structure. As shown in Figure 7, the designed TCN-based AE consists of an input layer, 4 TCN blocks, a 1D convolutional AE module, and an output layer. The core of the network is the TCN blocks, whose components are featured by the dilated convolutional layers and a residual connection. In particular, taking the first dilated convolution layer as an example, it takes $\{\hat{\boldsymbol{x}}_n\}_{1\le n\le N_s}$ as input and applys dilated convolution to it with N_{ch} 1D-kernels to obtain a N_{ch}-channeled output $\{\boldsymbol{z}_n\}_{1\le n\le N_s}$ for the next layer. For the k-th channel of the output ($k = 1, ..., N_{ch}$), given the 1D-kernel $\boldsymbol{F}_k = (\boldsymbol{f}_{k,1}, ..., \boldsymbol{f}_{k,L})$ with $\boldsymbol{f}_{k,l} \in \mathbb{R}^{N_F}$ ($l = 1, ..., L$) and L being the kernel size, the dilated convolution can be expressed as:

$$z_{k,n} = \sum_{i=0}^{L-1} \boldsymbol{f}_{k,i+1}^{\top} \hat{\boldsymbol{x}}_{n-\chi\cdot i},\ \forall n \in \{1, ..., N_s\}, \tag{10}$$

where $\hat{\boldsymbol{x}}_n$ is zero-padded for $n < 1$, $(\cdot)^{\top}$ is the transpose operator, and $\chi \in \mathbb{Z}^+$ denotes the *dilation factor* used to expand the receptive field of the output element.

According to Eqn. (10), the operation with a small dilation factor (e.g., $\chi = 1$) degenerates to a traditional convolution

for extracting the features of local context around each element of the input. As χ increases, the mutual dependency of local features can be captured by the kernel, and each element of the output can represent the local features of input in a wider range. Therefore, by utilizing a stack of dilated convolutional layers with exponentially increased χ as shown in Figure 7, the local features are gradually extracted and collected, enabling each node of the output layer to take into account well-represented local features for almost the entire spectrogram. Finally, with the results of the TCN blocks, the AE can effectively predict and recover the missing data. Representing the TCN-AE network as a $\boldsymbol{w}$-parameterized function $\mathcal{F}_{\boldsymbol{w}}$, re-arranging $\{\hat{\boldsymbol{x}}_n\}_{1\le n\le N_s}$ into matrix form $\hat{\boldsymbol{X}} \in \mathbb{R}^{N_F\times N_s}$, and denoting the output by $\tilde{\boldsymbol{Y}} \in \mathbb{R}^{N_F\times N_s}$, the recovering process can be represented as $\mathcal{F}_{\boldsymbol{w}} : \hat{\boldsymbol{X}} \to \tilde{\boldsymbol{Y}}$.

Self-supervised Training. Generally, the pre-collected training dataset needs to contain data-label pairs: spectrogram with sparse slices as the data and corresponding ground truth with no sparsity as the label. Unfortunately, collecting the ground truth directly is almost impossible, as sparse slices are caused by the lack of frames (hence losing the carried ground truth samples) during certain periods. To overcome this impossibility, we propose a self-supervised training method; it leverages only the non-sparse slices for training the TCN-AE, aiming to restore the hypothetical non-sparse data that facilitate various downstream sensing tasks. Specifically, we collect the spectrograms of non-sparse time slices to form the training label set, while obtaining the corresponding input data by randomly assigning `no-data` tags to the elements for creating artificial data sparsity. We note that this tag assignment needs to preserve the bursty and random patterns in the occurrence of `no-data` elements.

Moreover, to augment the training dataset, each expanded non-sparse spectrogram data in the label set are reused for multiple times with random tag assignments. Consequently, we obtain a completely labeled training dataset, denoted by $\mathcal{D}_{\text{train}} = \{\mathcal{T}(Y), Y\}$, without resorting to the impossible ground truth collection process. Here $Y \in \mathbb{R}^{N_F\times N_s}$ represents the spectrogram data of a non-sparse time slice, and $\mathcal{T}(\cdot)$

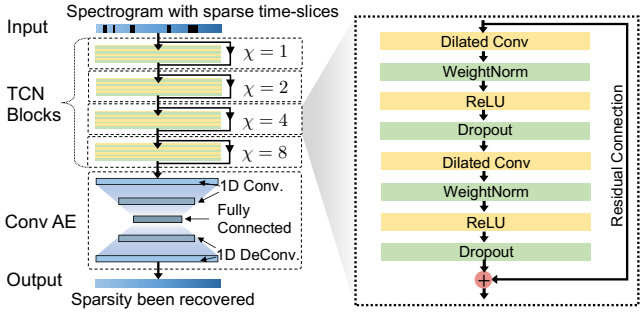


Figure 7: The structure of TCN-AE in MUSE-Fi.

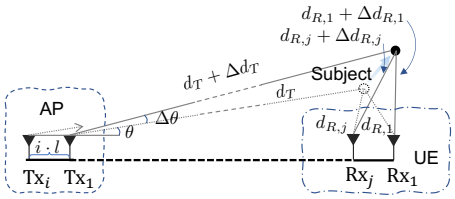


Figure 8: Channel variation due to subject motion: dashed and filled circles respectively represent the original subject position and that after motion.

is the random tag assignment. Finally, we can express the training process by the following optimization problem, to minimize the expected mean-squared error (MSE) between the recovered spectrogram and its label:

$$\min_{\boldsymbol{w}} \mathop{\mathbb{E}}_{(\mathcal{T}(Y),Y)\in\mathcal{D}_{\text{train}}} \|\mathcal{F}_{\boldsymbol{w}}(\hat{\boldsymbol{X}}) - Y\|_2^2, \text{ s.t. } \hat{\boldsymbol{X}} = \mathcal{T}(Y). \quad (11)$$

The overheads of the training is minor because no online labeling process is needed, and thus MUSE-Fi can collect the dataset automatically and conduct the training offline without incurring any real-time overheads.

3.3 To Compress or Not to Compress?

In this section, we specifically study the effectiveness of BFI-enabled sensing. Consider a conventional CSI matrix $\boldsymbol{H} \in \mathbb{C}^{N_{\text{rx}}\times N_{\text{tx}}}$ for a given subcarrier of a DL link with N_{tx} antennas for Tx at AP and N_{rx} antennas for Rx at UE. Instead of directly feedbacking $\boldsymbol{H}$ to the AP, the UE piggybacks a compressive form of $\boldsymbol{H}$ (i.e., BFI) onto the UL traffic, containing only the necessary information for Tx beamforming. Consider a channel state represented by $\boldsymbol{H}_0$, the UE first obtains the Tx beamforming matrix $\boldsymbol{V}$ by conducting the singular value decomposition (SVD) on $\boldsymbol{H}_0$, i.e., $\boldsymbol{H}_0 = \boldsymbol{USV}^*$ where $\boldsymbol{U} \in \mathbb{C}^{N_{\text{rx}}\times N_{\text{rx}}}$ and $\boldsymbol{V} \in \mathbb{C}^{N_{\text{tx}}\times N_{\text{tx}}}$ are unitary matrices, $\boldsymbol{S} \in \mathbb{R}^{N_{\text{rx}}\times N_{\text{tx}}}$ is a rectangular diagonal matrix with non-negative real values on the diagonal, and $(\cdot)^*$ denotes the conjugate transpose. The UE then compresses the channel state by converting $\boldsymbol{V}$ into BFI, which is represented by a series of real angles, and sends it to the AP. With the BFI received at the AP, a *reconstructed beamforming matrix* $\tilde{\boldsymbol{V}}$ is obtained, whose column vectors approximate those of $\boldsymbol{V}$ except for the column-wise phase-shifts that enforce the elements in the last row real-valued [16].

Based on the premise above, the BFI-enabled sensing under MUSE-Fi's context can be analyzed, following the illustration in Figure 8, where a subject is in the near-field of the UE while far from the AP. After a displacement of the subject, the altered channel condition changes the CSI matrix to $\boldsymbol{H}_1$ and also affects the SVD result by $\boldsymbol{H}_1 = \boldsymbol{Q}_{\text{rx}}\boldsymbol{H}_0\boldsymbol{Q}_{\text{tx}} = \boldsymbol{U'S'V'}^*$, where $\boldsymbol{Q}_{\text{rx}} = \mathbf{diag}(\rho_1 e^{-\mathrm{i}\frac{2\pi}{\lambda}\Delta d_{R,1}}, ..., \rho_{N_{\text{rx}}} e^{-\mathrm{i}\frac{2\pi}{\lambda}\Delta d_{R,N_{\text{rx}}}})$ and $\boldsymbol{Q}_{\text{tx}} = \mathbf{diag}(e^{-\mathrm{i}\frac{2\pi}{\lambda}\Delta d_T}, ..., e^{-\mathrm{i}\frac{2\pi}{\lambda}[\Delta d_T - (N_{\text{tx}}-1)\ell\Delta\theta\sin(\theta)]})$ with

ℓ being the distance between adjacent Tx antennas, and ρ_j being the amplitude ratio between two channel gains respectively from the subject at new position and original position to the j-th Rx antenna. We can observe that $\boldsymbol{V}' = \boldsymbol{Q}_{\text{tx}}^{*}\boldsymbol{V}$. Thus, based on the relationship between $\boldsymbol{V}'$ and $\boldsymbol{V}$, the reconstructed beamforming matrix after motion becomes:

$$\tilde{\boldsymbol{V}}' = \mathbf{diag}(e^{-\mathrm{i}\frac{2\pi}{\lambda}(N_{\text{tx}}-1)l\Delta\theta\sin(\theta)}, e^{-\mathrm{i}\frac{2\pi}{\lambda}(N_{\text{tx}}-2)l\Delta\theta\sin(\theta)}, ..., 1)\tilde{\boldsymbol{V}}.$$

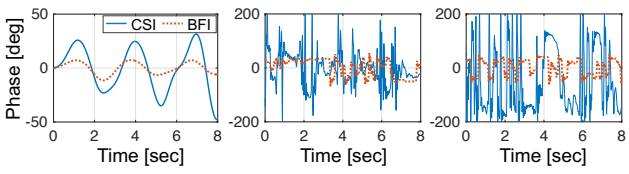
Apparently, the BFI variation from $\tilde{\boldsymbol{V}}$ to $\tilde{\boldsymbol{V}}'$ depends only on the change of the relative direction from the subject to the AP, i.e., $\Delta\theta$, which does not concern the UE at all. As shown in Figure 9, this compressive sensing brings **stability** to the sensing signal, yet at a cost of **reduced sensitivity** compared with CSI-based sensing, because BFI-sensing is almost insensitive to the relative motion between the subject and UE. For cases where the subject has rapid movements as shown in Figures 9(b) and 9(c), BFI sensing is more preferable as it produces results that are more stable, compared with CSI sensing that often causes drastic changes blended with noise and outliers. However, BFI sensing can be rather insensitive to micro-motions (albeit still viable), as demonstrated by Figure 9(a). Therefore, whether to use CSI or compressed BFI depends on the specific application and the trade-offs between stability and sensitivity. Note that our analysis assumes that the subject is off the LoS path, which does not account for cases where, for example, the subject's hands are operating on a (smartphone) UE.

4 PROTOTYPING & EXPERIMENT SETUP

In this section, we first elaborate on MUSE-Fi's implementation, then we introduce the experiment setup.

4.1 Implementing MUSE-Fi

MUSE-Fi consists of an AP and multiple UEs owned by subjects seeking sensing services. The AP is a Netgear Nighthawk X10 router [38], and the UEs include smartphones such as iPhone 13 [5] and OnePlus 10T [40], as well as Acer TravelMate laptops [1]. The Wi-Fi NICs adopted by MUSE-Fi employ 802.11b/g/n/ac for both UL-CSI and DL-CSI sensing, but utilize only 802.11ac for UL-BFI sensing (as BFI is available there only). The retrieval of CSIs is achieved via both Nexmon [19] and PicoScenes [26], while Wireshark [41] is sufficient to obtain cleartext BFI information from `Action No-ACK` frames. The obtained CSI and BFI information is analyzed using Matlab.



(a) Respiration. (b) Gesture (front-back). (c) Activity (jumping).

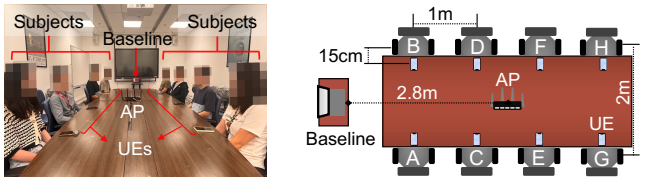
Figure 9: CSI vs. BFI in the time domain.

For training the SRA, after the sensing signals are passed through the pipeline, non-sparse slices in the spectrogram with a duration greater than 4s are picked for self-supervised training. To be specific, the non-sparse slices are used as ground truth (labels), and we then perform a random `no-data` tag assignment to them, thus generating sparse slices as the corresponding training inputs. We use 70% of the slices for training TCN-AE and the remaining 30% for testing. The parameters for sparse recovery are set as follows: $\Delta t = 0.1$s, $N_{\text{nsp}} = 2, L = 5, N_{\text{F}} = 32, N_{\text{ch}} = 64, f_{\text{rs}} = 64$Hz, and $f_{\text{cut}} = 1$Hz for respiration monitoring $f_{\text{cut}} = 20$Hz for other cases.

4.2 Experiment Setup

We first conduct micro-benchmark studies with a real-time video conference application, then we perform three case studies for realistic sensing applications. The setups for the case studies share three commonalities: i) each UE is placed in the near field of its associated subject, and it connects to the AP and continuously streams 1080p videos to emulate daily network usage, ii) all subjects perform specified activities simultaneously to test MUSE-Fi's ability in performing multi-person sensing, and iii) each experiment is conducted in a typical indoor meeting room with a different interior furniture arrangement. We also compare MUSE-Fi with a *non-near-field baseline* that employs another Wi-Fi device placed on the LoS path of the AP but not in the near-field of any subjects to collect CSI and BFI. Figure 10(a) illustrates our experiment setup for case studies, where the AP, subjects, UEs, and baseline device are all exhibited and annoted in Figure 10(b).

Respiration Monitoring. We let 8 subjects breathe simultaneously, and use NeuLog chest belts [39] to obtain the ground truth. The total respiration recording period is 80-minute. During the *Transformation* step of the SRA, the short-time Fourier transform (STFT) is employed to focus on the low-frequency components of respiration We employ a 3-layer convolutional neural network (CNN) to extract respiratory rate from the spectrogram.



(a) Testing scene. (b) Layout.

Figure 10: Experiment scene (a) and layout of subject arrangement (b) for all three case studies. The 15cm UE-subject distance is only meant to indicate a near-field layout rather than be fixed to that given value.

Gesture Detection. We let 8 subjects simultaneously perform six gestures, namely circle (CR), front-back (FB), slide (SL), star (ST), wave (WV), and zig-zag (ZZ). Each activity is performed 500 times, resulting in 24,000 CSI time series each containing 256 samples. We adopt the wavelet synchrosqueezed transform (WSST) [2] in the *Transformation* step, as it is highly effective in interpreting gesture signals that are non-stationary and contain complex frequency components. Besides, we employ the same classifier as in Widar3.0 [72] to achieve gesture detection from the spectrogram.

Activity Recognition. We let 8 subjects simultaneously perform six daily observed human activities: bending (BD), jumping (JM), rotating (RT), sitting down (SD), standing up (SU), and walking (WL). Each activity is performed 200 times, resulting in 9,600 CSI time series each containing 256 samples. Similar to gesture recognition, we employ WSST for transforming a time series into a spectrogram. To classify these activities, we utilize the same classifier as in RF-Net [15].

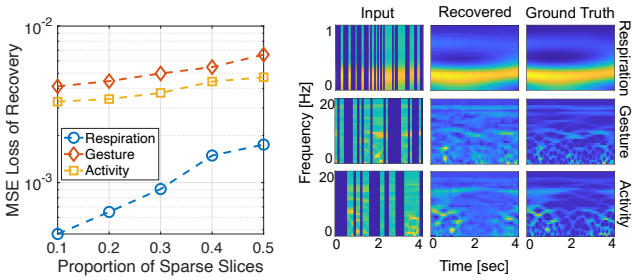
All evaluations focus on demonstrating MUSE-Fi's capability of multi-person sensing with commodity Wi-Fi devices; they, by no means, aim to show competitive performance against existing single-person monitoring systems. Instead, our objective is to validate MUSE-Fi's physical separability and quantify its benefits over non-near-field sensing. The comparisons between them are done by contrasting the sensing accuracy results of the former for an arbitrary subject against those of the latter. Our experiments have strictly followed the IRB of our institute.

5 EVALUATIONS

In this section, we begin with two micro-benchmark studies, verifying the effectiveness of SRA and further testing the differences in sensing via BFI vs CSI. This is then followed by the three case studies specified in Section 4.2.

5.1 Micro-benchmark Studies

5.1.1 Effectiveness of Sparse Recovery. To demonstrate the effectiveness of SRA, we collect CSI time series for one subject performing respiration, gesture, and activity. We use the MSE loss between the recovered and ground truth spectrograms as in Eqn. (11) to evaluate the performance of SRA. Figure 11(a) displays how the MSE losses of the recovery for all three categories vary with the amount of missing slices, clearly showing the MSE losses for respiration, gesture, and activity as below 2×10^{-3}, 5×10^{-3}, and 7×10^{-3}, respectively. Given the normalized spectrogram data, these resulting MSE values are sufficiently low to indicate successful recovery, thus validating the effectiveness of SRA. We also provide, in Figure 11(b), examples of recovered spectrograms. It is evident that SRA successfully recovers a significant portion of the input spectrogram, albeit miss a few minor details.



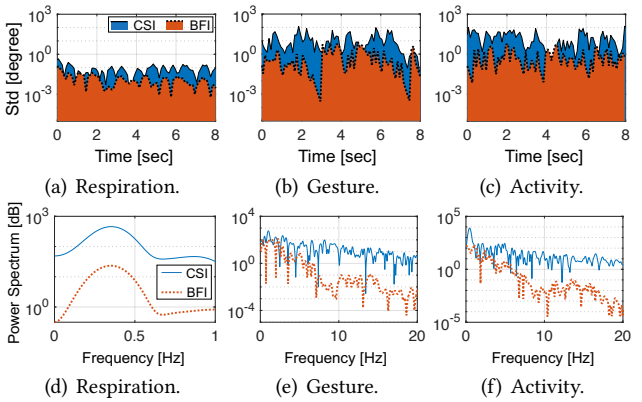
(a) MSE Loss. (b) Illustration of sparse recovery.

Figure 11: Performance of sparse recovery.

Upon further inspection, it is noticeable that gestures induce the largest MSE loss in recovery. This is likely because the hands of the subject, when compared with the subject's body, are closer to the UE, making the sensing results more sensitive to hand movements. The higher sensitivity to gestures introduces more complicated time-frequency patterns to the input spectrogram, naturally lowering the accuracy of the sparse recovery. On the contrary, respiration induces the least MSE loss because of its relatively stable and periodic style, which results in more regular patterns in the spectrogram and hence facilitates sparse recovery.

5.1.2 Comparison between CSI and BFI. Since UL-CSI and DL-CSI are symmetric, we refrain from comparing them but rather combine their outcomes and analysis in the following. To further analyze the brief observations made in Section 3.3, we perform sensing on a subject carrying out Since the time-domain results are consistent with those shown in Figure 9, we do not show such results again for brevity.

We further investigate the fluctuations of the BFI and CSI signals by calculating the standard deviations of the detrended signals over periods of 0.1 seconds, as shown in Figures 12(a) to 12(c). The figure reveals that BFI signals are more stable than CSI signals, but this stability comes at the cost of reduced sensitivity. We also examine the power



(a) Respiration. (b) Gesture. (c) Activity.

(d) Respiration. (e) Gesture. (f) Activity.

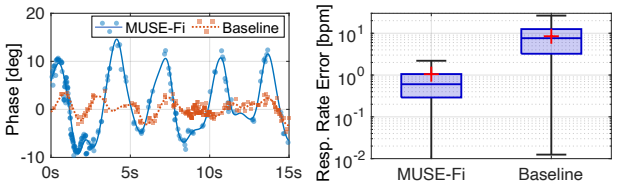
Figure 12: Comparing CSI and BFI in terms of standard deviations (a)-(c) and power spectrum (d)-(f).

spectrum of the CSI and BFI in Figures 12(d) to 12(f). One may readily observe that while CSI preserves the respiration signal and presents a smooth spectrum, it is too sensitive to rapid and large-scale motions and results in excessive power in high frequencies for gesture and activity. In comparison, BFI effectively suppresses high-frequency components, while its response to low-frequency subtle movements (e.g., respiration) is less pronounced.

To explain these phenomena, it is worth noting that the phase of CSI is directly related to relative displacement, making it sensitive to small-scale movements (e.g., respiration). However, large-scale movements cause abrupt phase changes that cannot be captured by insufficient sampling, resulting in irregularities in the CSI signal. As explained in Section 3.3, the BFI-based sensing strategy only captures the relative directional changes from the subject to the AP: if one deems the conversion from CSI to BFI as "low-pass" filtering, it would be natural to expect fewer variations but also lowered strength in the resulting signals. This property is particularly beneficial for a future study on subjects carrying smartphones on their bodies for continuous vital signs monitoring [13, 71], as BFI sensing may filter out body movement interference.

5.2 Case-I: Respiration Monitoring

We conduct experiments to monitor multi-person respiration using the setup described in Section 4.2. After obtaining sparse recovery results, we use a 3-layer CNN to extract the respiration rate and measure the respiration rate error as $|R_E - R_A|$, where R_E is the estimated respiration rate and R_A is the actual respiration rate. Figure 13(a) showcases the respiration waveforms obtained by MUSE-Fi and the baseline method. One may clearly observe that MUSE-Fi recovers the respiration waveforms effectively, whereas the baseline method only captures a noisy signal mixture contributed by multiple subjects. We further assess the accuracy of respiration rate estimation of both MUSE-Fi and the baseline in Figure 13(b); the results reveal that MUSE-Fi achieves accurate respiration monitoring with both median and mean

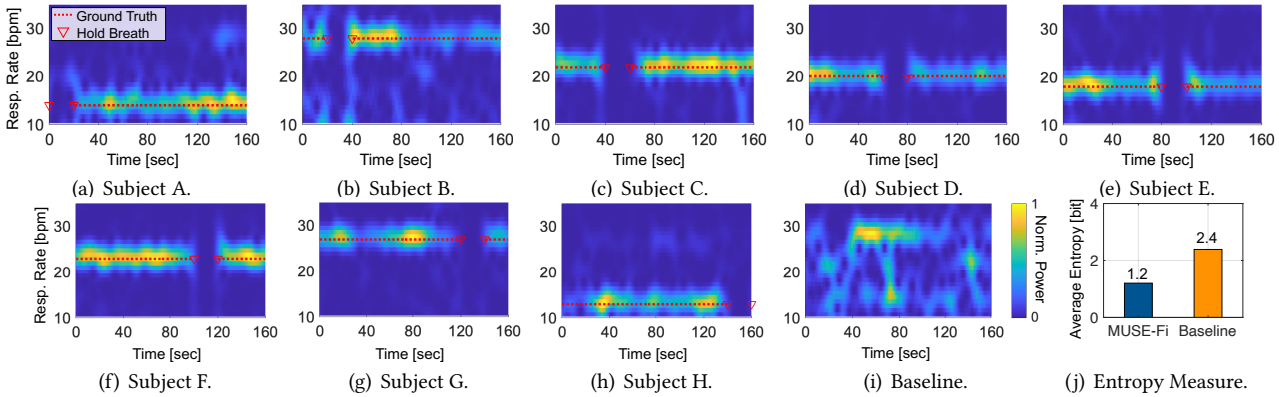


(a) Respiratory waveform. (b) Sensing error.

Figure 13: Comparison between MUSE-Fi and the baseline in terms of respiration sensing.

respiration rate errors less than 1 bpm. In contrast, the baseline exhibits a median and mean respiration rate error of 7 and 8 bpm, respectively, making it almost useless in multi-person scenarios.

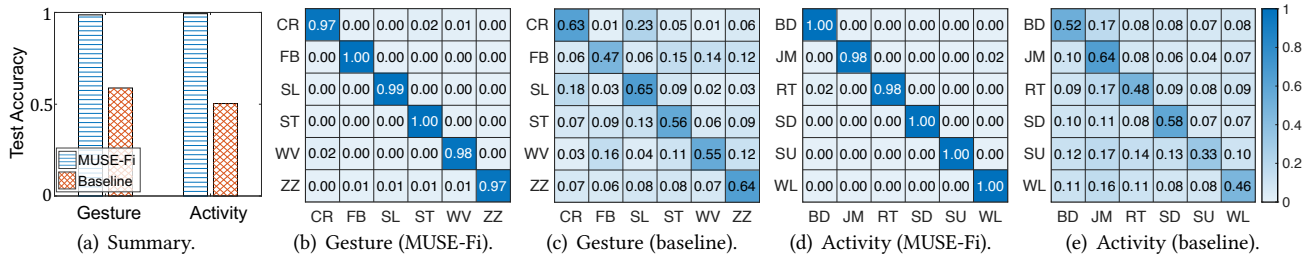
To further understand the performance difference, we present the spectrograms of MUSE-Fi and the baseline method in Figures 14(a)-14(h) and 14(i), respectively. We can observe that MUSE-Fi recovers a clear signal around the ground truth frequency thanks to the near-field domination. Moreover, we let the subjects sequentially hold their breath for 20 s, and the correspondence between the breath-holding periods and signal interruption (whose boundary is denoted by two triangular markers) on the spectrograms firmly proves that the respiration signals from different subjects are well separated. On the contrary, the baseline method fails to distinguish respiration from multiple subjects, resulting in a noisy spectrogram where no accurate respiration rate can be obtained. We further employ the average spectral entropy [48] of the normalized spectrogram to measure the residual uncertainty in determining respiration rate. Specifically, Figure 14(j) indicates a 2.4 bit entropy for the baseline, much higher than the 1.2 bit entropy of MUSE-Fi. Intuitively, the variety of potential respiration rates represented by the spectrogram increases exponentially with its spectral entropy. Therefore, this halved entropy value implies that the respiration rate decision of MUSE-Fi can be much more precise than that of the baseline, thus explaining our result in Figure 13(b). All



(a) Subject A. (b) Subject B. (c) Subject C. (d) Subject D. (e) Subject E.

(f) Subject F. (g) Subject G. (h) Subject H. (i) Baseline. (j) Entropy Measure.

Figure 14: Comparison and analysis on MUSE-Fi and the baseline, in terms of the respiration spectrograms.



(a) Summary. (b) Gesture (MUSE-Fi). (c) Gesture (baseline). (d) Activity (MUSE-Fi). (e) Activity (baseline).

Figure 15: Comparison between MUSE-Fi and baselines for gesture and activity recognition.

these results provide conclusive evidence of the efficacy of MUSE-Fi in respiration monitoring.

5.3 Case-II: Gesture Detection

We also conduct experiments on gesture detection, and summarize the statistics in Figure 15(a). One may readily observe that MUSE-Fi achieves a mean test accuracy of more than 98%, while that of the baseline is only 57%. We further inspect the confusion matrices of MUSE-Fi and the baseline in Figures 15(b) and 15(c), respectively. The confusion matrices reveal that MUSE-Fi can correctly classify most gestures, while the baseline often confuses one gesture with others. Specifically, the circle (CR) and slide (SL) gestures are the most confusing pair for the baseline, as they both involve moving one's hands smoothly over the phone, which may appear similar to the baseline in the far field, but are readily differentiable by the near-field MUSE-Fi. The baseline's inferior performance can also be attributed to its inability to disentangle signals from interference caused by the moving hands of multiple people, eventually causing its unacceptable detection behavior (i.e., 39% lower than MUSE-Fi). These results evidently confirm MUSE-Fi's effectiveness in resolving multi-person gesture detection for real-world applications.

5.4 Case-III: Activity Recognition

We further conduct experiments on activity recognition, with statistics summarized in Figure 15(a); MUSE-Fi's mean accuracy of more than 98% is doubled of the baseline's that drops by 8% compared with the gesture detection task. This performance degradation is attributed to the greater interference induced by large-scale and rapid human activities. We further inspect the confusion matrices of MUSE-Fi and the baseline in Figures 15(d) and 15(e), respectively. One may readily observe that the accuracy for all 6 activities is above 0.98 for MUSE-Fi, while the baseline's accuracy is all less than 0.52 (with SU bearing the worst accuracy of 0.33). The results from all three case studies have successfully demonstrated a great potential to realize a long-standing vision for Wi-Fi human sensing: multiple people sitting around a table (e.g., holding a meeting), while leveraging contactless sensing to accomplish diversified tasks with the support of their respective smartphones and only one Wi-Fi AP.

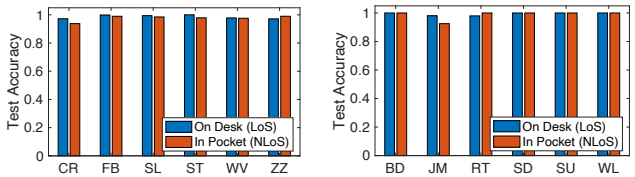
5.5 Extended Experiments and Discussions

To prove the generalizability of MUSE-Fi, we evaluate it in another practical scenario, where the subjects carry their smartphones inside their pockets, hence the LoS paths between the AP and UEs are blocked. Here we focus on the gesture detection and activity recognition tasks, given their relevance to enabling the computer-human interfacing for XR applications. In Figure 16, we compare the sensing accuracy of MUSE-Fi in two scenarios: 1) the on-desk scenario with LoS condition as in Figure 10, and 2) the in-pocket scenario with non-LoS (NLoS) condition, where the sensing accuracy of MUSE-Fi are shown to be similar and higher than 92% for the both scenarios. This is because the Wi-Fi signals can diffract and bypass the boundary of body, while the condition for near-field domination effect still holds.

Based on the above case studies and extended experiments, we make the following discussions on MUSE-Fi's generalizability, potential applications, and key factors.

Generalizability. MUSE-Fi is capable of generalizing beyond current experimental setup because, firstly, environment dynamics have a small impact on MUSE-Fi due to the near-field domination effect; and secondly, environment layout changes typically manifests as additional biases to the CSI and can be removed during the normalization of SRA.

Potential Applications. With the physical separability guaranteed by the near-field domination effect, MUSE-Fi is scalable to ubiquitous Wi-Fi networks in daily life and can provide solutions for XR. Based on respiration monitoring results in Section 5.2 MUSE-Fi can track subtle motions of



(a) Gesture (b) Activity.

Figure 16: Comparison between the sensing accuracy of MUSE-Fi under LoS and NLoS conditions.

subjects near distributed UEs, which can extend to AR/MR solutions for visualizing intrusion, vital signs of human, and operation status of machines. Besides, the results in Sections 5.3, 5.4, and 5.5 indicate its capability to accurately recognize individuals' gestures and activities. This means MUSE-Fi enables the sensing functionality of gesture detection and activity recognition to be integrated into Wi-Fi modules naturally carried by each individual, potentially reducing the cost, weight, and power consumption of VR/MR headsets for computer-human interaction.

Key Factor Analysis. The near-field domination effect assumes the existence of LoS UE-AP paths, yet MUSE-Fi is also robust to the LoS blockage by the subject's body as shown in Figure 16. Therefore, MUSE-Fi is effective in common practice where APs are located above the subjects. If LoS paths are blocked, and the signals travel via NLoS paths involving reflection in the environment, then the near-field domination condition can be naturally extended to the shortest NLoS paths, provided that these paths are clear of environment dynamics (e.g., other irrelevant motions) to avoid injecting interference directly into the received signals. In addition, MUSE-Fi efficiently handles the sparsity of realistic data traffic for online videos and meetings etc, while the highly sparse data traffic for idle UEs may be beyond recovery and lead to invalid sensing results. Finally, sensing security [23, 36] and co-existing with other co-channel communication systems [32, 33, 63] should be handled in the future.

6 RELATED WORK

Our work is closely related to contactless human sensing [3, 13, 15, 25, 35, 42, 44, 50, 56, 58, 59, 64, 66, 67, 72] that has experienced significant advancements in the past decade. Of all these solutions, Wi-Fi human sensing [6, 25, 35, 44, 56, 58, 59, 67, 72] gains particular popularity because it is compatible with existing communication hardware, thus not incurring additional deployment cost. Specifically, Wi-Fi human sensing exploits CSI [21] retrieved from the received signals of Wi-Fi communications. The ubiquity of Wi-Fi infrastructure (e.g., Wi-Fi APs, laptops, and smartphones) has led to a multitude of research studies for various applications, including vital sign monitoring [35, 59], gesture detection [56, 72], activity recognition [15, 25], localization [30, 49], and motion tracking [51, 58], which, at a higher level, are driving the vision of smart home [3] and digital healthcare [17, 53].

While the above proposals mainly focus on single-person scenarios due to the limited range resolution of existing Wi-Fi technology, recent research has explored the use of next-generation Wi-Fi technologies to overcome this challenge. ViMo [57] uses 60 GHz 802.11ad devices [45] with high bandwidth and a 32-element phased array to emulate a radar. Similarly, mmTrack [61] employs the same 802.11ad device for multi-person localization. However, the limited adoption and high cost of 802.11ad devices may hinder the goal of multi-person monitoring. It should be noted that MUSE-Fi shares part of its name with MUSE [52], but these two systems bear distinct objectives: whereas MUSE focuses on communication scheduling under MU-MIMO, MUSE-Fi targets multi-person sensing.

Other techniques than future Wi-Fi hardware may also help enhance human sensing. Widar2.0 [44] provides partial support for multi-person sensing by using multiple antennas to improve spatial resolution. Karanam *et al.* [27] use the magnitude measurements from an array of receivers to perform multi-person tracking. Lan *et al.* [31] employ metasurface antennas with varying beam patterns to perform multi-person activity recognition. Liu et al. [35] estimate multi-person respiration rates by analyzing CSI's power spectral density. PhaseBeat [60] and TR-BREATH [11] leverage root-MUSIC [46] to separate multi-person sensing signals, while Yang et al. [65] optimize transceiver deployment using the Fresnel zone model to reduce interference, but with the requirement of accurate subject location and fixed transceiver placement. MultiSense [67] treats multi-person sensing as a blind source separation problem and uses ICA [24] to extract waveforms. Last but not least, SPARCS [42] recovers the micro-doppler spectrum by using intrinsic sparsity of wideband mmWave channels. However, it does not fit for narrowband Wi-Fi systems operating at microwave band.

7 CONCLUSION

Taking an important step towards ubiquitous human sensing, MUSE-Fi has innovated in Wi-Fi multi-person sensing by addressing the major challenge of physically separating multiple subjects. Leveraging the near-field channel variation caused by a subject in close proximity to a Wi-Fi device, MUSE-Fi has demonstrated successful handling of multi-person sensing for respiration monitoring, gesture detection, and activity recognition. This success also stems from our two technical developments: i) an SRA to cope with realistic (intermittent) Wi-Fi traffic under multi-user scenarios, and ii) a study on the difference between CSI and BFI sensing. Our extensive evaluations have evidently confirmed that MUSE-Fi is a cost-effective alternative to radar-based systems that often require extra deployments. Moving forward, we believe that MUSE-Fi has significant potential to be extended into various applications, including healthcare, smart homes, and even security; we are also planning to deploy MUSE-Fi on larger scales so as to evaluate its performance in more diversified scenarios.

ACKNOWLEDGEMENT

This research is support by National Research Foundation (NRF) Future Communications Research & Development Programme (FCP) grant FCP-NTU-RG-2022-015.

REFERENCES

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

[2] Paul S. Addison. 2017. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance.* CRC Press.

[3] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2015. Smart Homes That Monitor Breathing and Heart Rate. In *Proc. of the 33rd ACM CHI.* 837–846.

[4] Ian F Akyildiz and Hongzhi Guo. 2022. Wireless Extended Reality (XR): Challenges and New Research Directions. *ITU J. Future Evol. Technol* 3, 2 (2022), 1–15.

[5] Apple Inc. 2023. Buy iPhone 13. https://www.apple.com/sg/shop/buy-iphone/iphone-13. Online; accessed 12 February 2023.

[6] Martin Azizyan, Ionut Constandache, and Romit Roy Choudhury. 2009. SurroundSense: Mobile Phone Localization via Ambience Fingerprinting. In *Proc. of the 15th ACM MobiCom.* 261–272.

[7] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv preprint arXiv:1803.01271* (2018).

[8] Constantine A Balanis. 2016. *Antenna theory: Analysis and design.* John Wiley & Sons.

[9] Oscar Bejarano, Edward W. Knightly, and Minyoung Park. 2013. IEEE 802.11ac: From Channelization to Multi-User MIMO. *IEEE Communications Magazine* 51, 10 (2013), 84–90.

[10] Julie Carmigniani, Borko Furht, Marco Anisetti, Paolo Ceravolo, Ernesto Damiani, and Misa Ivkovic. 2011. Augmented Reality Technologies, Systems and Applications. *Multimed. Tools Appl.* 51 (2011), 341–377.

[11] Chen Chen, Yi Han, Yan Chen, Hung-Quoc Lai, Feng Zhang, Beibei Wang, and K. J. Ray Liu. 2018. TR-BREATH: Time-Reversal Breathing Rate Estimation and Detection. *IEEE Transactions on Biomedical Engineering* 65, 3 (2018), 489–501.

[12] 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.

[13] 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.

[14] 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 Transactions on Mobile Computing* 20, 2 (2021), 588–602.

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

[17] Yao Ge, Ahmad Taha, Syed Aziz Shah, Kia Dashtipour, Shuyuan Zhu, Jonathan Cooper, Qammer H Abbasi, and Muhammad Ali Imran. 2022. Contactless WiFi Sensing and Monitoring for Future Healthcare-Emerging Trends, Challenges, and Opportunities. *IEEE Reviews in Biomedical Engineering* 16 (2022), 171–191.

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

[19] Francesco Gringoli, Matthias Schulz, Jakob Link, and Matthias Hollick. 2019. Free Your CSI: A Channel State Information Extraction Platform For Modern Wi-Fi Chipsets. In *In Proc. of the 13th ACM WiNTECH.* 21–28.

[20] Tobias Grosse-Puppendahl, Sebastian Herber, Raphael Wimmer, Frank Englert, Sebastian Beck, Julian Von Wilmsdorff, Reiner Wichert, and Arjan Kuijper. 2014. Capacitive Near-field Communication for Ubiquitous Interaction and Perception. In *Proc. of the UbiComp'14.* 231–242.

[21] 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.

[22] Yinghui He, Jianwei Liu, Mo Li, Guanding Yu, Jinsong Han, and Kui Ren. 2023. SenCom: Integrated Sensing and Communication with Practical WiFi. In *Proc. of the 29th ACM MobiCom.* 1–16.

[23] Jingyang Hu, Hongbo Wang, Tianyue Zheng, Jingzhi Hu, Zhe Chen, Hongbo Jiang, and Jun Luo. 2023. Password-Stealing without Hacking: Wi-Fi Enabled Practical Keystroke Eavesdropping. In *Proc. of the 30th ACM CCS.* 1–14.

[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] Zhiping Jiang, Tom H. Luan, Xincheng Ren, Dongtao Lv, Han Hao, Jing Wang, Kun Zhao, Wei Xi, Yueshen Xu, and Rui Li. 2021. Eliminating the Barriers: Demystifying Wi-Fi Baseband Design and Introducing the PicoScenes Wi-Fi Sensing Platform. *IEEE Internet of Things Journal* (2021), 1–21.

[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] Mikolaj Kegler, Pierre Beckmann, and Milos Cernak. 2020. Deep Speech Inpainting of Time-frequency Masks. In *Proc. of the 21st ISCA INTERSPEECH.* 3276–3280.

[29] Evgeny Khorov, Ilya Levitsky, and Ian F Akyildiz. 2020. Current Status and Directions of IEEE 802.11be, the Future Wi-Fi 7. *IEEE Access* 8 (2020), 88664–88688.

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

[31] 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.

[32] 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.

[33] 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.

[34] Jian Liu, Hongbo Liu, Yingying Chen, Yan Wang, and Chen Wang. 2020. Wireless Sensing for Human Activity: A Survey. *IEEE Commun. Surv. Tutor.* 22, 3 (2020), 1629–1645.

[35] 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.

[36] Jun Luo, Hangcheng Cao, Hongbo Jiang, Yanbing Yang, and Zhe Chen. 2024. MIMOCrypt: Multi-User Privacy-Preserving Wi-Fi Sensing via MIMO Encryption. In *Proc. of the 45st IEEE S&P.* 1–19.

[37] Matthew S. Gast. 2013. 802.11ac: A Survival Guide. https://www.oreilly.com/library/view/80211ac-a-survival/9781449357702/ch03.html. Online; accessed 26 February 2022.

[38] Netgear. 2023. Nighthawk X10 Smart WiFi Router (AD7200). https://www.netgear.com/sg/home/wifi/routers/ad7200-fastest-router/. Online; accessed 12 February 2023.

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

[40] OnePlus. 2023. OnePlus 10T 5G. https://www.oneplus.com/sg/10t. Online; accessed 12 February 2023.

[41] Angela Orebaugh, Gilbert Ramirez, and Jay Beale. 2006. *Wireshark & Ethereal Network Protocol Analyzer Toolkit.* Elsevier.

[42] 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.

[43] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, and Kyle Jamieson. 2017. Widar: Decimeter-Level Passive Tracking via Velocity Monitoring with Commodity Wi-Fi. In *Proc. of the 18th ACM MobiHoc*. 6:1–10.

[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] Qualcomm Technologies, Inc. . 2022. Qualcomm 802.11ad 60GHz WiFi. https://www.qualcomm.com/products/features/80211ad. Accessed: 2022-05-28.

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

[47] Theodore S Rappaport. 2010. *Wireless Communications: Principles and practice.* Pearson Education India.

[48] Claude Elwood Shannon. 2001. A Mathematical Theory of Communication. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* 5, 1 (2001), 3–55.

[49] Elahe Soltanaghaei, Avinash Kalyanaraman, and Kamin Whitehouse. 2018. Multipath Triangulation: Decimeter-level WiFi Localization and Orientation with a Single Unaided Receiver. In *Proc. of the 16th ACM MobiSys*. 376–388.

[50] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: Monitoring Human Lung Function via Acoustic Sensing on Commodity Smartphones. In *Proc. of the 26th ACM MobiCom*. 1–14.

[51] Li Sun, Souvik Sen, Dimitrios Koutsonikolas, and Kyu-Han Kim. 2015. WiDraw: Enabling Hands-free Drawing in the Air on Commodity WiFi Devices. In *Proc. of the 21st ACM MobiCom*. 77–89.

[52] Sanjib Sur, Ioannis Pefkianakis, Xinyu Zhang, and Kyu-Han Kim. 2016. Practical MU-MIMO User Selection on 802.11ac Commodity Networks. In *Proc. of the 22nd ACM MobiCom*. 122–134.

[53] Bo Tan, Qingchao Chen, Kevin Chetty, Karl Woodbridge, Wenda Li, and Robert Piechocki. 2018. Exploiting WiFi Channel State Information for Residential Healthcare Informatics. *IEEE Communications Magazine* 56, 5 (2018), 130–137.

[54] Sheng Tan, Yili Ren, Jie Yang, and Yingying Chen. 2022. Commodity WiFi Sensing in Ten Years: Status, Challenges, and Opportunities. *IEEE Internet Things J.* 9, 18 (2022), 17832–17843.

[55] Ajay Tirumala. 1999. iPerf: The TCP/UDP Bandwidth Measurement Tool. *http://dast. nlanr. net/Projects/Iperf/* (1999).

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

[57] Fengyu Wang, Feng Zhang, Chenshu Wu, Beibei Wang, and K. J. Ray Liu. 2021. ViMo: Multiperson Vital Sign Monitoring Using Commodity Millimeter-Wave Radio. *IEEE Internet of Things Journal* 8, 3 (2021), 1294–1307.

[58] 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.

[59] 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.

[60] 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.

[61] Chenshu Wu, Feng Zhang, Beibei Wang, and KJ Ray Liu. 2020. mmTrack: Passive Multi-person Localization using Commodity Millimeter Wave Radio. In *Proc. of the 39th IEEE INFOCOM*. IEEE, 2400–2409.

[62] Dapeng Wu, Y.T. Hou, Wenwu Zhu, Ya-Qin Zhang, and J.M. Peha. 2001. Streaming Video Over The Internet: Approaches and Directions. *IEEE Trans. Circuits Syst. Video Technol.* 11, 3 (2001), 282–300.

[63] 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.

[64] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. BreathListener: Fine-Grained Breathing Monitoring in Driving Environments Utilizing Acoustic Signals. In *Proc. of the 17th ACM MobiSys*. 54–66.

[65] Yanni Yang, Jiannong Cao, Xuefeng Liu, and Kai Xing. 2018. Multi-person Sleeping Respiration Monitoring with COTS WiFi Devices. In *Proc. of the 15th IEEE MASS*. 37–45.

[66] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. 2017. Strata: Fine-grained Acoustic-based Device-free Tracking. In *Proc. of the 15th ACM MobiSys*. 15–28.

[67] Youwei Zeng, Dan Wu, Jie Xiong, Jinyi Liu, Zhaopeng Liu, and Daqing Zhang. 2020. MultiSense: Enabling Multi-Person Respiration Sensing with Commodity WiFi. In *Proc. of the 22nd UbiComp*. 102:1–29.

[68] 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.

[69] Shujie Zhang, Tianyue Zheng, Zhe Chen, and Jun Luo. 2022. Can We Obtain Fine-grained Heartbeat Waveform via Contact-free RF-sensing?. In *Proc. of the 41st IEEE INFOCOM*. 1759–1768.

[70] Shujie Zhang, Tianyue Zheng, Hongbo Wang, Zhe Chen, and Jun Luo. 2022. Quantifying the Physical Separability of RF-based Multi-Person Respiration Monitoring via SINR. In *Proc. of the 20th ACM SenSys*. 47–60.

[71] 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.

[72] 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.