RAPID: Retrofitting IEEE 802.11ay Access Points for Indoor Human Detection and Sensing

Jacopo Pegoraro[†§] , Jesus O. Lacruz[*] , Francesca Meneghello[†] ,
Enver Bashirov[†] , Michele Rossi[†] , and Joerg Widmer[*]

Abstract—In this work we present RAPID, the first joint communication and radar system based on next-generation IEEE 802.11ay WiFi networks operating in the 60 GHz band. Unlike existing approaches for human sensing at millimeter-wave frequencies, which rely on special-purpose radars, RAPID achieves radar-level sensing accuracy with IEEE 802.11ay access points, thus avoiding the burden of installing ad-hoc sensors. RAPID enables contactless human sensing applications, such as people tracking, Human Activity Recognition (HAR), and person identification without requiring modifications to the standard packet structure. Specifically, we leverage IEEE 802.11ay beam training to accurately localize and track multiple individuals within the same environment. Then, we propose a new way of using beam tracking to extract micro-Doppler signatures from the time-varying Channel Impulse Response (CIR) estimated from *reflected* packets. Such signatures are fed to a deep learning classifier to perform HAR and person identification. RAPID is implemented on a cutting-edge IEEE 802.11ay-compatible FPGA platform with phased antenna arrays, and evaluated on a large dataset of CIR measurements. It is robust across different environments and subjects, and outperforms state-of-the-art sub-6 GHz WiFi sensing techniques. Using two access points, RAPID reliably tracks multiple subjects, reaching HAR and person identification accuracies of 94% and 90%, respectively.

Index Terms—Joint communication-radar, mmWave, IEEE 802.11ay, micro-Doppler, wireless sensing, people tracking, human activity recognition (HAR), person identification.

---

1 INTRODUCTION

In this work, we design RAPID, a pervasive Joint Communication & Radar (JCR) system that extends the capabilities of upcoming WiFi technology operating in the 60 GHz Millimeter-Wave (mmWave) spectrum to integrate sensing functionalities into wireless networks. The joint provisioning of communication and sensing services is of great value to pave the way toward advanced smart-home applications without the need for deploying dedicated sensing hardware. In this regard, our target is to retrofit IEEE 802.11ay hardware so as to natively offer human and environment sensing services to end users, in addition to high-throughput communication.

Thanks to their large available bandwidth, mmWave signals allow performing localization and tracking with decimeter-level accuracy, making them the preferred solution for contactless sensing through radio waves. Most emerging mmWave sensing systems are based on *dedicated* mmWave radar devices and estimate the micro-Doppler (μD) effect induced by human motion (*signature*) with high accuracy via specifically designed bursts of phase coherent chirp signals [1], [2]. Radar μD signatures contain detailed information about the movement velocity of the different human body parts across time, and enable fine-grained sensing applications such as person identification from gait features [1], [3], [4], [5], Human Activity Recognition (HAR) [6], gait disorder diagnosis [2] and fall detection [7], among others. However, solutions based on mmWave radars come with the drawback of the need for installing bespoke sensors, which limits their scalability and ease of deployment in practical scenarios (e.g., smart buildings, offices, etc.).

In this respect, the ubiquitous deployment of WiFi devices has sparked research interest towards developing joint communication and Radio Frequency (RF) sensing technology, to avoid the cost of installing dedicated hardware while at the same time benefiting from communication capabilities. The effort of enhancing WiFi devices with environment sensing features has recently led to the establishment of the IEEE 802.11bf standardization group [8], aimed at integrating sensing functionalities into WiFi-enabled devices. While legacy WiFi technology based on IEEE 802.11n/ac/ax (sub-6 GHz bands) standards provides a viable means for environment and human sensing [9] and HAR [10], [11], it suffers from intrinsic limitations due to the relatively low bandwidth available in the sub-6 GHz license-exempt portion of the radio spectrum. This prevents highly accurate distance measurements and multi-person localization and tracking in realistic scenarios. Moving to the mmWave spectrum, previous works based on the IEEE 802.11ad standard exploit the Channel Impulse Response (CIR) estimation procedure for localizing people [12], [13], but they are not fully compliant with the communication packet structure specified by the standard and cannot match the sensing accuracy of radars, as no μD information is captured. Overall, the extraction of μD signatures is difficult using standard communication devices and protocols, due to the lack of specifically designed waveforms and transmission modes. Extracting Doppler information from sequences of subsequent packets, as done in radars, is highly non-trivial due to the random and time varying phase offsets between the transmitter and the receiver [14]. In fact, the offsets destroy

---

§ *Corresponding author e-mail:* `pegoraroja@dei.unipd.it`
† *These authors are with the Department of Information Engineering, University of Padova, Italy.*
* *These authors are with the IMDEA Networks Institute, 28918 Madrid, Spain.*

the phase coherence across different packets, preventing the extraction of μD signatures which require a phase analysis across long sequences of subsequently transmitted signals.

RAPID is the first system that successfully extracts μD signatures of human movements using standard WiFi transmission technology working on the mmWave spectrum, and achieves radar-level accuracy in sensing. It works *without modifying the packet structure* by leveraging the *in-packet* beam training and beam tracking features of IEEE 802.11ay. This leads to very low implementation and deployment cost, and allows for a highly accurate extraction of human movement information from the radio signals.

IEEE 802.11ay uses highly directional antennas to shape precise beams for communication. For that, the standard specifies efficient in-packet beam training and tracking procedures [15], based on training (TRN) fields consisting of repetitions of complementary Golay sequences [16]. The fields are transmitted with different beam patterns, which allows determining which of them is best for communication. By exploiting *beam training* packets, RAPID accurately localizes multiple human subjects within the same indoor space. Then, the μD signature associated with the movement of each subject is extracted by relying on the TRN units embedded in the data packets used for *beam tracking*, analyzing the phase differences of the CIR across subsequent packets that are reflected back by the environment. The obtained μD spectrograms are processed using deep learning classifiers to carry out continuous HAR and person identification.

Thanks to the intrinsic superior ranging resolution of the mmWave spectrum and our advanced signal processing, RAPID outperforms state-of-the-art human sensing technology based on sub-6 GHz WiFi systems. RAPID allows individually tracking multiple moving subjects, separating their signal reflections and, in turn, obtaining large improvements in terms of accuracy, robustness and generalization across environments and subjects. In addition, multiple RAPID-Access Points (APs) can be seamlessly integrated to boost detection and tracking performance. This also increases HAR and person identification accuracy by combining the information from different viewpoints.

In this work, RAPID-APs are implemented using an FPGA-based Software Defined Radio (SDR) platform equipped with phased antenna arrays, which transmits IEEE 802.11ay-compliant packets and operates in a full-duplex fashion. RAPID IEEE 802.11ay APs enable their transmit and receive chains simultaneously, avoiding the problem of random phase offsets as transmitter and receiver share the same local oscillator. Note that this does not require complex self-interference cancellation for full-duplex communication, since the receiver needs to only detect the highly robust Golay sequences of the TRN fields.

We stress that RAPID is not simply about applying radar signal processing to a different domain. Reusing standard-compliant IEEE 802.11ay signals requires developing new processing steps to obtain range, angle, and μD information, while taking care of JCR-specific problems that do not arise in radar systems. While radars typically estimate the channel using ad-hoc *chirp* waveforms, whose parameters can be *tuned* to meet the specific sensing requirements, RAPID re-uses standard-compliant Golay sequences. Therefore, the sensing resolution can not be adapted to the considered scenario, and the person detection and range estimation steps have to be entirely re-designed to be robust under such constraints. For what concerns the Angle of Arrival (AoA) estimation, mmWave radars are usually equipped with Multiple Input Multiple Output (MIMO) antenna arrays that ease the estimation of the AoA by analyzing the phase change across the spatial dimension. On the contrary, IEEE 802.11ay APs typically mount cheaper phased array antennas, therefore a different approach has to be designed to obtain the AoA by analyzing the CIR estimated through different beam patterns (Section 3.3.3). Lastly, the μD computation is challenging as it involves *(i)* striking a good balance between the packet transmission rate and the Doppler frequency resolution required to capture the μD of human movement, while *(ii)* ensuring sufficient phase coherence across adjacent packets (see Section 3.4.3). Moreover, Golay sequences are known to have low Doppler resolution [17], and no existing study has evaluated the feasibility of using them to extract fine-grained human μD signatures.

To summarize, the main contributions of our work are:

1) We design and implement RAPID, a fully standard compliant JCR system that exploits IEEE 802.11ay TRN fields to achieve radar-like human sensing, including simultaneous multi-person tracking, HAR and person identification. RAPID reuses existing fields in the communication packets and avoids the need for a dedicated sensing infrastructure. RAPID can leverage data from a single AP or combine information from multiple APs for improved performance.

2) We propose a novel method to extract μD signatures of human movement from IEEE 802.11ay CIR estimates obtained from a sequence of IEEE 802.11ay data packets with added beam tracking fields, exploiting the Golay sequences specified in the standard. To the best of our knowledge, this is the first work to do so.

3) We implement RAPID on a novel FPGA-based testbed including multiple IEEE 802.11ay-compliant APs which support full-duplex operation, so that each AP can listen to its own transmitted signal and act as a monostatic JCR device.

4) We conduct an extensive indoor measurement campaign to evaluate the proposed system and compare it to sub-6 GHz WiFi systems. To this end, we build a unique dataset including simultaneous IEEE 802.11ay and IEEE 802.11ac CIR estimates. RAPID achieves continuous tracking of up to 5 concurrently moving subjects, with HAR accuracy of 94% and person identification accuracy of 90%. Moreover, it outperforms state of the art sub-6 GHz WiFi sensing, showing superior accuracy and robustness to different environments and subjects.

The paper is organized as follows. The related work is summarized in Section 2. RAPID is introduced in Section 3, presenting its constituent processing blocks. A summary of how IEEE 802.11ay can be used for environment sensing is given in Section 4, while in Section 5 the implementation of RAPID on FPGA hardware is discussed. A thorough performance analysis of RAPID on real measurements is presented in Section 6. Section 7 concludes the discussion.

2 RELATED WORK

Sub-6 GHz sensing. Legacy WiFi technologies such as IEEE 802.11n and IEEE 802.11ac, respectively working at 2.4 or 5 GHz, have been extensively used for human sensing, including activity/gesture recognition [10], [11], [18], [19], vital sign monitoring [20] and person identification [21]. Due to the rich multipath environment at lower frequencies, existing approaches have reached good accuracies by leveraging OFDM transmission and analyzing the CIR amplitude obtained at the different subcarriers, as done in [10]. The performance of such systems can be further improved by exploiting the phase components of the CIR [18], [19], but this entails using complex algorithms for the removal of random phase offsets.

Although there is a large body of work that exploits these technologies, they have two main drawbacks: *(i)* they are effective for single-person scenarios, as the small available bandwidth only allows for coarse localization and tracking of the subjects, and *(ii)* they are highly sensitive to changes in the environment and hardly generalize to new scenarios (never seen at system calibration/training time), which can significantly worsen their performance. Addressing problem *(i)*, in [21], multi-person identification using IEEE 802.11n is achieved in a through-the-wall setting, but the subjects still need to be well separated in space (e.g., by at least $20°$ in azimuth angle at a distance of several meters). To mitigate the dependence on the environment, more elaborate deep learning and optimization approaches have been proposed in [19], [22], [23]. However, they still are not able to approach radar-like sensing accuracy.

mmWave radar sensing. mmWave frequencies offer a natural solution to the above issues, by providing decimeter-level accuracy in distance measurements and high sensitivity to the μD effect, due to their small transmission wavelength. In addition, due to the sparsity of the mmWave channel, higher robustness to environmental changes is achieved. mmWave radars have been intensively studied in the past few years as an effective means to achieve fine grained environment sensing [24]. Typical operating frequencies for these devices are the 60 or the 77 GHz bands. Centimeter-level accuracy in measuring distances is achieved thanks to the use of very large transmission bandwidths, up to 4 GHz, as dedicated radar devices are not constrained by communication requirements. Radars allow accurate HAR [6], [25] and have been used to perform person identification on small to medium-sized groups of people (up to a few tens), due to their very high resolution in obtaining the μD signatures of the subjects [4], [26]. In these works, the separation of the reflections from subjects concurrently moving in the environment is achieved through MIMO radars, which enable high angular resolution and allow tracking the users with errors below 0.2 m even in realistic scenarios where people walk and move freely [5]. However, these results are obtained within relatively small distances from the radar, ranging from 4 [5] to $6-7$ m [4].

Despite the advanced sensing capabilities, mmWave radars entail high deployment costs to cover large indoor areas, even more considering their limited working range. For this reason, multi-radar networks to cover wider areas and avoid occlusions are seldom considered in the literature.

Reusing existing mmWave communication links, as we do in this work, allows avoiding the costly deployment of additional hardware, while maintaining radar-like human sensing and detection performance.

802.11ad 60 GHz sensing. Commodity 60 GHz radios have been utilized for client device localization [27], people tracking [12], fine-grained human gesture recognition [28], [29], vital sign monitoring [30] and RF imaging [13]. Among them, in [28], pulsed radar-like operations are performed to detect and track a human hand, reconstructing handwriting with centimeter-level accuracy. Notably, [29] performs similar processing using the IEEE 802.11ad CIR estimated by a mobile device for gesture classification. In [13], a commodity 60 GHz radio equipped with a 6×6 antenna array is used to obtain the silhouette of a person moving directly in front of the device. This is achieved with an angular super-resolution algorithm derived from MUSIC [31]. However, the device needs to be operated in a *radar mode* for transmission, which may not comply with the communication standard. In [12], the estimated CIR amplitude is used along with receiver beamforming to localize and track multiple people, achieving a median localization error of 9.9 cm. This work does not exploit the phase of the CIR to extract the μD signature of the subjects, which is necessary to carry out HAR and person identification tasks. Moreover, the extension to the case of multiple APs is not considered. Overall, the research addressing human sensing through the IEEE 802.11ad standard typically does not consider the *joint* communication and sensing aspect, which requires to reuse the packet structure specified by the communication standard.

802.11ay 60 GHz sensing. To the best of our knowledge, RAPID is the first system that extracts radar-like μD signatures of human movement from IEEE 802.11ay 60 GHz APs, by retrofitting them with human sensing and μD extraction capabilities. This is obtained by preserving the IEEE 802.11ay packet structure, thus obtaining a *joint radar-communication platform that is fully standard compliant*.

3 RAPID SENSING SYSTEM

RAPID enables indoor human sensing in IEEE 802.11ay networks, by leveraging the network *in-packet* beam training and beam tracking fields. In the following, the system is presented by detailing the processing blocks that allow performing people localization and tracking, HAR, and person identification. The novel algorithms specifically designed to extract range, AoA and Doppler estimates from the CIR obtained though standard-compliant Golay sequences are deepened. The mathematical models of the CIR and the CIR phase are included to make the analysis self-comprehensive.

3.1 System overview

From a high-level perspective, RAPID performs the following operations, as shown in Fig. 1.
(1) IEEE 802.11ay CIR estimation: 802.11ay specifies the transmission of a variable number of TRN units for in-packet beam training, each using a (possibly) different Beam Pattern (BP). From the CIR, which is estimated from each TRN unit (see Section 4), RAPID obtains a scan of the whole
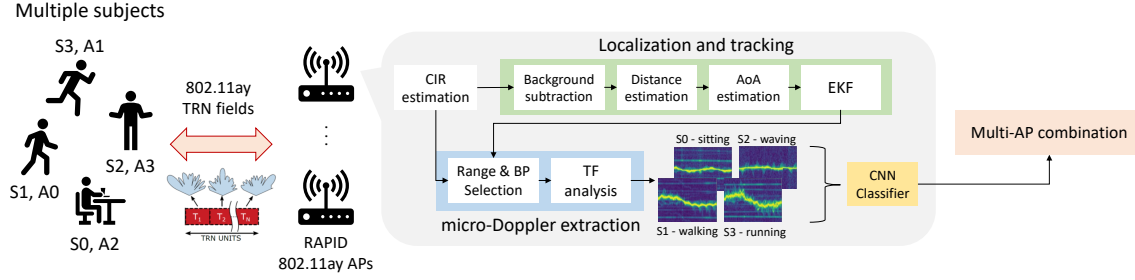
Figure 1: Overview of the RAPID system.

angular Field-of-View (FOV), which contains accurate information about all the surrounding objects and people.

(2) People localization and tracking: the individuals are detected by performing background subtraction from the CIR amplitude and applying a thresholding algorithm to detect candidate reflection paths from humans, see Section 3.3.1 and Section 3.3.2, respectively. Subsequently, a correlation based algorithm is utilized to estimate the angular position of the subjects, as described in Section 3.3.3, and an Extended Kalman Filter (EKF) is exploited to sequentially track and refine the positions of the individuals across time (Section 3.3.4). By combining more than one AP, RAPID can boost its human detection capabilities, while effectively coping with occlusion problems, as quantified in Section 6.2.

(3) μD spectrum extraction: here, the μD spectrum of each detected person is extracted. This is implemented by utilizing the CIR model as a radar return signal, and using the estimated positions from point (2) to single out the CIR portions (the paths and the BPs) containing the contributions of each subject, see Section 3.4.3. The μD signature of each individual's movement is then extracted by computing the power spectrum of the corresponding complex-valued portion of the CIR over windows of suitable length, employing Time-Frequency (TF) analysis.

(4) HAR and person identification: the spectrograms from step (3) are fed to a deep learning classifier based on a residual Convolutional Neural Network (CNN) [32] for HAR. Thanks to the separation of the CIR, and to the subsequent computation of the μD for each individual, RAPID is capable of recognizing the different activities performed by multiple subjects within the same indoor space. Moreover, through a second CNN module, it is also able to identify a person, by extracting and analyzing their gait features from the μD signature. With multiple APs, the classifications are refined by selecting the best AP to make the decision, according to the confidence of the classifier output.

In this work, we aim at localizing and tracking people within a given physical space, by identifying which person is performing which activity. This requires person identification, tracking and HAR capabilities. The person identification task is carried out by extracting and analyzing the μD associated with the human gait, as this is an effective (soft) biometric signature, which has been successfully used in many works [33]. Hence, we first detect when a person is walking, then we get his/her identity from the μD gait signature and, finally, we keep tracking the person by also recognizing their activities. This also works the other way around, i.e., if a person is at first sitting and doing other activities, and then starts walking later on; as long as tracking

works, we can later determine who was sitting earlier on. This also explains why tracking a person is critical, so that it is still clear which person is where, even when he/she performs other activities than walking.

We now present in detail each RAPID processing function, following the workflow of Fig. 1.

3.2 CIR estimation

CIR estimation is a key component of most communication systems and is used to obtain information about the environmental reflections of the signal – such as their associated angle of arrival and delay at the receiver – to properly set the data transmission parameters and decode received packets. RAPID leverages this process for sensing purposes. A key aspect to our design is that the large transmission bandwidth of mmWave systems leads to CIR containing fine-grained information about the environment. In our system, the transmitter and the receiver units are co-located: the signal sent by the former, after bouncing off nearby reflectors (objects or humans), is collected at the receiver that retrieves information for each reflector, such as its distance and angular position with respect to the device, its moving velocity and micro-Doppler.

The CIR is represented as a vector of complex channel gains, also referred to as *paths* in the following, and indicized through letter ℓ. Due to the finite delay resolution of the system, the CIR vector can only represent a discrete grid of paths, with corresponding propagation delays $\tau_\ell = \ell/B, \ell = 0, \ldots, L-1$, where B is the transmission bandwidth and $1/B$ is the delay resolution. The components of the CIR vector, which represent the complex gains for the L paths, are obtained by correlating the received signal with pre-defined Golay sequences, using standard techniques [34], see also Section 4. Path ℓ is mapped onto the corresponding reflector distance using $d_\ell = c\tau_\ell/2$, with c being the speed of light. The vector containing all the distances of interest is defined as $\mathbf{d} = [d_0, d_1, \ldots, d_{L-1}]^T$, with L being the number of paths in the CIR. If multiple CIR estimations are performed over a single packet, using different BPs, the reflections from the environments are amplified differently. This is due to the different BP shapes, as each BP steers the transmission signal towards a specific direction (beam steering). In addition, the CIR estimation is repeated for each packet k, which can be seen as sampling the CIR in time, with sampling period corresponding to the inter-packet transmission time T_c. The expression of the ℓ-th CIR component, having delay τ_ℓ, obtained using beam-

pattern p at time (packet) k is

$$h_{\ell,p}(k) = a_{\ell,p}(k)e^{j\phi_\ell(k)}, \tag{1}$$

where $a_{\ell,p}(k)$ and $\phi_\ell(k)$ are the complex gain of path ℓ at time k and its phase, respectively. The path gain depends on the contribution of the BP used for the transmission and on the reflectivity of the target, whereas the phase depends on the delay τ_ℓ. Note that $h_{\ell,p}(k)$ is a *time domain* quantity, depending on the propagation delay index ℓ and on the time. In Eq. (1) we used index k as a shorthand notation for the discrete time instants $kT_c = 0, T_c, 2T_c, \ldots$.

3.3 People localization and tracking

RAPID leverages the CIR estimates collected over time to continuously perform localization and tracking. The process develops in four steps: *(i)* background subtraction, to remove the reflected paths due to static objects, *(ii)* estimation of the subjects' distances, *(iii)* estimation of the angular positions of the subjects with respect to the device, and *(iv)* joint processing of distance and angle information using a Kalman filter to track each person's trajectory across time.

RAPID computes estimates at different rates, according to the specific resolution that is required by each task. Localization and tracking information are updated by RAPID every $\Delta t > T_c$ seconds, where index t denotes the localization/tracking time-steps, whereas HAR and identification require CIR readings at a rate $1/T_c$. The choice of setting $\Delta t > T_c$ stems from the fact that performing localization and tracking for every transmitted packet is unnecessary, as the packet transmission rate $1/T_c$ is much larger than the speed of human motion. This allows for additional flexibility in the selection of the type of BPs that are used for each packet: as we explain shortly below in Section 4 and Section 5, we can modulate how many TRN units are included in a packet according to the type of sensing function that is being performed, i.e., localization/tracking versus activity/identity recognition.

3.3.1 Background subtraction

To infer the positions of the subjects it is key to remove the reflections due to static (background) objects, as these typically have a much higher intensity than those generated by humans and may impact the localization accuracy. The background-related CIR is estimated by computing the time average of the CIR amplitude within a window of K_{static} samples, as static reflections are constant across time,

$$\bar{h}_{\ell,p} = \frac{1}{K_{\text{static}}} \sum_{k=0}^{K_{\text{static}}-1} |h_{\ell,p}(k)|. \tag{2}$$

Then, the foreground CIR amplitude component is obtained as $|\tilde{h}_{\ell,p}(t)| = \max\left(|h_{\ell,p}(t)| - \bar{h}_{\ell,p}, 0\right)$, i.e., removing the amplitude of the static paths and setting to zero the amplitude of those paths that would be present in the reference background CIR, but that are shielded by the presence of a person. We remark that, through different BPs, we perform beam steering at the transmitter. Hence, the peaks in $|\tilde{h}_{\ell,p}|$ correspond to the strongest propagation paths, as seen at the receiver when beam-pattern p is used at the TX side. Changing the BP p allows scanning the environment by

varying the transmission angle and, in turn, sweeping the whole field of view. We use this to infer the distance and the angular position of each individual, as described next.

3.3.2 Distance estimation

The distance of each subject is obtained by applying a threshold on $|\tilde{h}_{\ell,p}|$ (the time index is omitted for better readability), selecting the strongest paths across all the used BPs. First, for each reflected path ℓ, we consider vector

$$\mathbf{h}_\ell = \left[|\tilde{h}_{\ell,0}|, |\tilde{h}_{\ell,1}|, \ldots, |\tilde{h}_{\ell,N_p-1}|\right]^T, \tag{3}$$

containing the CIR values of path ℓ for each of the N_p BPs that are used at the transmitter. We collect the L_2-norms of $\mathbf{h}_\ell$, with $\ell = 0, 1, \ldots, L-1$, obtaining a new vector $\mathbf{h}$, as

$$\mathbf{h} = [||\mathbf{h}_0||_2, ||\mathbf{h}_1||_2, \ldots, ||\mathbf{h}_{L-1}||_2]^T, \tag{4}$$

containing the strengths of each path at the receiver. We locate the local maxima in $\mathbf{h}$, denoting them by $h'_0, h'_1, \ldots, h'_{n_{\text{peaks}}-1}$. Hence, we discard those peaks with amplitude smaller than a dynamic threshold A_{th} computed from the maximum and average power of the paths in the current CIR. We introduce the following coefficients α_{max}, α_{mean}, and compute the threshold value A_{th}, as

$$A_{\text{th}} = \max\left\{\alpha_{\text{max}} \cdot \max_i h'_i, \alpha_{\text{mean}} \cdot \bar{h}'\right\}, \tag{5}$$

with $\bar{h}' = \sum_i h'_i / n_{\text{peaks}}$. A thorough evaluation of suitable values for α_{max} and α_{mean} is provided in Section 6.4. With Eq. (5) the threshold is computed dynamically, proportionally to the maximum between the average and the maximum value of the CIR. The peaks that exceed the threshold are selected as candidate targets of interest and used for the subsequent AoA estimation. Denoting by $\ell_1, \ell_2, \ldots, \ell_{N_s}$ the indices of the selected (candidate) paths ($0 \le \ell_j \le L-1$), the corresponding distances are obtained as $d_{\ell_j} = c\tau_{\ell_j}/2$.

3.3.3 Angular position estimation

The following procedure is applied to each of the N_s candidate paths. Let vector $\mathbf{s}_{\ell_j} \in \mathbb{R}^{N_p}$ contain the squared CIR amplitudes from one of such paths, ℓ_j, for all used beam patterns, i.e., $\mathbf{s}_{\ell_j} = \left[|\tilde{h}_{\ell_j,0}|^2, |\tilde{h}_{\ell_j,1}|^2, \ldots, |\tilde{h}_{\ell_j,N_p-1}|^2\right]^T$. $\mathbf{s}_{\ell_j}$ is normalized by dividing it by its L_2-norm $||\mathbf{s}_{\ell_j}||_2$, then a correlation measure is used to estimate the angular position of the target by exploiting the gains of each beam pattern along the azimuth angular FOV θ. Specifically, denoting by $g_p(\theta) \in [0,1]$ the normalized gain of beam pattern p along direction θ (see Fig. 5b), the angular position for candidate path ℓ_j is estimated as

$$\theta_{\ell_j} = \arg\max_\theta \sum_{p=0}^{N_p-1} g_p(\theta) \frac{|\tilde{h}_{\ell_j,p}|^2}{||\mathbf{s}_{\ell_j}||_2^2}. \tag{6}$$

The rationale behind Eq. (6) is that if $|\tilde{h}_{\ell_j,p}|$ originates from the signal reflected off a subject, the corresponding angular direction is the one leading to the highest correlation between the CIR squared amplitude and the set of beam pattern gains. This is because each BP amplifies path ℓ_j differently, depending on the beam pointing direction.

Upon obtaining the distance and the angle estimates, an Extended Kalman filter is utilized to track the subjects' positions over time.

3.3.4 People tracking - extended Kalman filter

After the localization step, the candidate positions of the subjects are known in polar coordinates, and constitute our *observations* of the positions of the subjects, which we denote by $\mathbf{z}_t^j = [d_{\ell_j}, \theta_{\ell_j}]^T, \forall j = 1, 2, \ldots, N_s$. We employ an EKF [35] to track the physical position of each individual in the Cartesian space. Specifically, we define the true state of subject j at time t as vector $\mathbf{x}_t^j = \left[x_t^j, y_t^j, \dot{x}_t^j, \dot{y}_t^j \right]^T$, containing the coordinates along the $x - y$ horizontal plane and the movement velocity components along the same axes. We approximate the motion of the subjects with a constant velocity (CV) model [36]. As the observations $\mathbf{z}_t^j$ become available, we apply the predict and update steps of the EKF to follow the movement trajectories of the subjects [35]. The association between the observations from time $t+1$ and the states from time t is done using the nearest-neighbors joint probabilistic data association algorithm (NN-JPDA) [37].

Using the EKF estimates $\hat{\mathbf{x}}_t^j$ of each person's state across subsequent time steps allows retrieving the path and the BPs in the CIR which contain his/her μD signature.

3.4 micro-Doppler extraction

3.4.1 CIR phase model

The CIR model in Eq. (1) is here expanded and related to radar theory [38]. Using a typical radar terminology, we refer to the CIR samples $\ell = 0, 1, \ldots, L - 1$ as the *fast-time* sampling dimension, as they are obtained at the highest available sampling rate. The CIR samples collected across different packets are instead referred to as the *slow-time* samples, indicized by variable k as in Section 3.2.

Next, we consider a moving object within the monitored indoor space; the transmitted signal is reflected off the object and the corresponding contribution is retrieved at the receiver in the ℓ-th path of the CIR. To extract the μD effect caused by the movement of this object, we analyze the phase of the ℓ-th path across time. The time-dependent phase term in Eq. (1) can be expressed as follows

$$\phi_\ell(k) = -2\pi f_o \frac{2(d_\ell - v_\ell k T_c)}{c} = -2\pi f_o \bar{\tau}_\ell + 4\pi f_o \frac{v_\ell}{c} k T_c. \quad (7)$$

Here, $\bar{\tau}_\ell$ is the delay of the ℓ-th path due to the distance of the corresponding reflector from the device. v_ℓ is the radial velocity of the reflector with respect to the device, which is assumed to be *slowly* time-varying, i.e, we can consider it constant during a μD spectrum processing interval (see Section 3.4.2). From Eq. (7) it can be seen that the velocity of the object at distance d_ℓ, if greater than zero, modulates CIR phase across the slow time dimension. Following a common convention [38], in this work objects moving away from the transmitter (AP) have *positive* velocity, while incoming objects have *negative* velocity.

The human body contains multiple moving parts that have different velocities and follow different trajectories. Thanks to the small wavelength of mmWaves, in the μD we can observe these different contributions via TF analysis, as detailed in the next Section 3.4.2.

3.4.2 micro-Doppler spectrum

Human movement causes a frequency modulation on the reflected signal due to the small-scale Doppler effect produced by the different body parts. Using TF analysis of the received signal, it is possible to distinguish between different actions performed by a person or identify the individual based on his/her way of walking (*gait*) [1], [2]. mmWave radios are particularly suited for this, as their frequencies are sensitive to the μD effect due to their small wavelengths.

From Eq. (7), the μD effect of human movement can be extracted from subsequent estimates of the CIR, computed every T_c seconds. Specifically, one can compute the short-time Fourier transform (STFT) of $h_{\ell,p}(k)$, across slow-time, for each path ℓ and each beam pattern p as

$$H_{\ell,p}(n, i) = \sum_{m=0}^{M-1} h_{\ell,p}(m + n\sigma) w(m) e^{-j2\pi \frac{im}{M}}, \quad (8)$$

where n is the time index, $i = 0, 1, \ldots, N_D - 1$ is the frequency index, M is the (fixed) window length, w is a Hann window of dimension M and σ is the time granularity of the STFT. The power spectrum of $h_{\ell,p}(k)$, computed as $\mu_{\ell,p}(n, i) = |H_{\ell,p}(n, i)|^2$, contains information on the phase modulation due to the velocity v_ℓ, and can be used to analyze its evolution across subsequent windows.

Eq. (8) can not be used directly to extract the μD signature of a moving human in our setup, as it refers to a single fast time bin (a single path in the CIR) and a single BP, while people can be located in different positions across time. In addition, it would be inefficient to compute the STFT for all the paths and all the BPs. Instead, the computation should only be performed for those physical locations where a person is detected. In the following, we leverage the localization and tracking process described in Section 3.3 to only extract the CIR portions that contain useful μD information.

3.4.3 μD separation

Assume that we want to extract the μD of a person that was detected and located by the previous algorithms at a certain distance and angle with respect to the device. Hence, we extract the CIR samples from the most useful BP, i.e., the one that points in the direction of the person and that, in turn, emphasizes the most the reflection from this target.

From the estimated state of this person (Section 3.3.4), their angular position is obtained as $\hat{\theta} = \arctan(\hat{y}_t/\hat{x}_t)$ and their distance from the device, as $\hat{R} = \sqrt{\hat{x}_t^2 + \hat{y}_t^2}$. The BP approximately pointing in the direction of this person, denoted by p^*, is thus selected as the BP having the highest gain along $\hat{\theta}$, that is

$$p^* = \arg\max_p g_p(\hat{\theta}). \quad (9)$$

Moreover, due to the high ranging accuracy of mmWaves, humans typically produce reflections that influence more than a single CIR path. The CIR paths of interest are those that correspond to a neighbourhood of $\hat{R}$. In our analysis, we take the size of this neighborhood constant across all subjects, denoting it by Q. Specifically, we first select the path ℓ^* that best matches the subject's distance $\hat{R}$

$$\ell^* = \arg\min_\ell |d_\ell - \hat{R}|. \quad (10)$$

Then, from the original complex-valued CIR, we extract a window containing $Q + 1$ samples along the fast-time

dimension, centered on ℓ^*, and use Eq. (8) to compute the μD spectrum components for our target at time n as

$$\mu_i(n) = \sum_{\ell=\ell^*-Q/2}^{\ell^*+Q/2} |H_{\ell,p^*}(n,i)|^2, \quad i = 0, 1, \dots, N_D - 1.$$
(11)

These μD spectrum components are collected through vector $\boldsymbol{\mu}(n) = [\mu_0(n), \mu_1(n), \dots, \mu_{N_D-1}(n)]^T$.

To capture the human movement evolution across time, we compute the μD vectors for a window of $N_{\mu D}$ subsequent time-steps and concatenate them into a spectrogram representing the μD signature of the target up to time n, as

$$\boldsymbol{\Upsilon}_n = [\boldsymbol{\mu}(n - N_{\mu D} + 1), \boldsymbol{\mu}(n - N_{\mu D} + 2), \dots, \boldsymbol{\mu}(n)]. \quad (12)$$

The procedure described in this section is repeated for all the detected subjects.

3.4.4 Human μD range and resolution

During everyday movement, the limbs of a person usually have velocities of up to $3 - 4$ m/s [1], [2]. To fully capture the μD signature of the subjects, we must ensure that our systems achieves a sufficient resolution. Recalling Eq. (7), we know that the Doppler frequency shift induced by a moving object on the ℓ-th path is $f_\ell^D = 2f_o v_\ell/c$. Using TF analysis to estimate the Doppler spectrum as in Eq. (8), the resolution that can be obtained on the Doppler frequency is $\Delta f^D = 1/(MT_c)$. The maximum measurable Doppler frequency is instead $f_{\max}^D = 1/(2T_c)$. These quantities can be mapped onto the velocity estimate resolution and the maximum measurable velocity as

$$\Delta v = \frac{c}{2f_o M T_c}, \quad v_{\max} = \frac{c}{4f_o T_c}. \quad (13)$$

Given that we sample the CIR on a per-packet basis, to capture the μD effect of human motion we must ensure that the time T_c between the packets used in the μD estimation allows capturing the range of velocities of interest. See also Section 6 for the chosen values of M and T_c.

3.5 Activity recognition and person identification

The μD signature, obtained as in Eq. (12), contains information about the type of movement performed by the person.

To perform HAR and person identification, we use a deep neural network to classify each spectrogram. Specifically, once the μD signatures of each person have been separated, RAPID performs the following tasks: *(i)* it classifies the activity carried out by the subject into *walking* (A0), *running* (A1), *sitting down* (A2), *waving hands* (A3) and *standing still* (A4) and *(ii)* it recognizes the subject's identity during a walking phase, among a known set of individuals, denoted by S0, S1, etc. In Fig. 2, we show μD signature examples for activities A0 − 3, concurrently performed by 4 subjects within the same environment.

As human μD is highly variable across different subjects, and we seek robustness to different environment conditions and noise, we employ deep learning to classify the μD signatures. Referring to a single subject, the μD spectrum $\boldsymbol{\Upsilon}_n$ is represented as an image and processed by two separate CNNs for HAR and person identification, respectively. The two classifiers share the same architecture, as shown in Fig. 3, but are trained separately and have different weights as they perform different tasks. As the subjects are continuously tracked over time, we adopt a sliding window approach, selecting μD spectrograms with $N_{\mu D}$ μD spectrum samples for each window (matrix $\boldsymbol{\Upsilon}_n$). Subsequent windows partially overlap to increase the reactiveness of RAPID in obtaining predictions. Both CNNs are trained to extract features from the μD spectrograms and to classify the activity performed by or identity of the person, by learning a function $\mathcal{F}(\cdot)$ that maps a μD window, $\boldsymbol{\Upsilon}_n$, of size $N_D \times N_{\mu D}$, onto a vector $\mathbf{c}_n$ containing the HAR (identification) class probabilities, i.e., $\mathbf{c}_n = \mathcal{F}(\boldsymbol{\Upsilon}_n)$. The dimension of the final probability vector $\mathbf{c}_n$ is different in case of HAR or identification depending on the dimension of the classification problem. The second CNN, used for person identification, is only trained on walking spectrograms, as human gait is well known to be a soft biometric identifier [33]. Hence, during the system operation, the identification classifier is only applied on the input μD spectrogram when the activity is classified as "walking" by the HAR classifier, see Fig. 3.

3.5.1 μD spectrogram pre-processing

Prior to feeding it to the CNN classifier, the μD spectrogram is pre-processed by removing the contributions from static reflections and normalizing it.

Static reflection removal. A customary step when processing human μD signatures is the removal of static reflections, which appear as a strong power peak around the 0 m/s velocity bin. This can be done by either applying a high-pass filter to the signal or, if deep learning methods are used for classification, by directly removing the Doppler bins containing unwanted contributions, as done in [1], [3]. We adopt the latter method to remove the Doppler bins corresponding to the velocities in the interval $[-0.28, 0.28]$ m/s, as they contain very low, non-informative velocities.

Normalization. To compensate for differences in the strength of the reflections when subjects are far from the APs, we normalize each column of $\boldsymbol{\Upsilon}_n$, $\boldsymbol{\mu}(j), j = 0, 1, \dots, N_{\mu D} - 1$ in the range $[0, 1]$.

3.5.2 Deep learning classifier

We use the same CNN architecture, based on deep residual networks [32], for HAR and person identification, with the only difference being the dimension of the last classification layer. This network consists of 4 consecutive residual blocks. Each residual block has two convolutional layers [39], the first of which includes a down-sampling by a factor of 2 (*stride*). Each convolution is followed by an Exponential-Linear Unit (ELU) activation function [40] and batch normalization [41]. The output of the convolution is summed to the input (*skip connection*) and passed through another ELU activation and batch normalization. The 4 residual blocks use 8, 16, 32 and 64 filters, respectively, all having a kernel of size 3×3. After the last residual block, we apply Dropout [42] with a ratio of 0.5, and a fully-connected (or *dense*) layer with 64 units, then, a second Dropout operation with ratio of 0.2. Finally, the classification probabilities for HAR or person identification are computed via a Softmax activation function [39]. The network architecture is shown in Fig. 3.
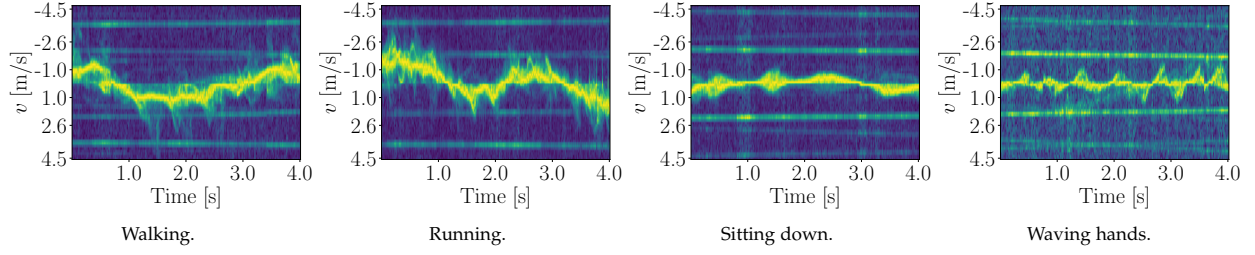
Figure 2: Example 4 s long μD spectrograms obtained by RAPID from 4 subjects. The yellow and blue colors respectively represent high and low power in the corresponding Doppler velocity bins (y axis).
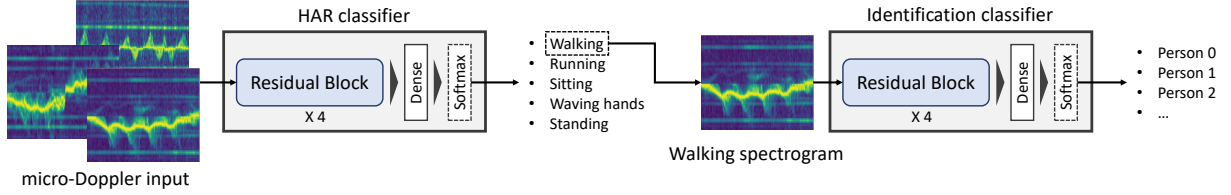

Figure 3: Block diagram of the CNN classifiers used by RAPID for HAR and person identification.

3.5.3 Combining multiple APs

Using the different points of view provided by the different APs, RAPID can improve its HAR and person identification performance. Assume that a person is independently detected and tracked by 2 or more APs concurrently. A slightly different μD signature of the person is obtained by each AP, according to the angular position and the distance of the device with respect to the person. At each time instant n, we adopt a simple decision fusion scheme including the following steps: *(i)* if a single AP detects the person, the decision made by the classifier on the corresponding μD signature is used, i.e., $\arg\max_j c_{n,j}$, where $c_{n,j}$ is element j of vector $\mathbf{c}_n$, *(ii)* if multiple APs detect the person, denote by $\mathbf{c}_n^a$ the probability vector predicted by AP a. The final decision is made by the AP that is most confident about its classification, i.e., the one that assigns the highest probability to the predicted class: $\arg\max_j \left\{ \max_a c_{n,j}^a \right\}$.

4 ENABLING SENSING IN IEEE 802.11AY

The high bandwidth of IEEE 802.11ay [43] not only provides high data throughput but also offers excellent accuracy for sensing applications. RAPID is able to extract highly accurate range, angle and μD information from CIR measurements. For this, we take advantage of the beam training and beam tracking mechanisms of IEEE 802.11ay systems.

Range and angle information are extracted from the CIR obtained via the Channel Estimation Fields (CEFs) of standard beacon frames that are frequently sent by the AP or the beam training frames sent during a Sector Level Sweep (SLS). The SLS is a two-step procedure: first, one device sends training frames using the available antenna configurations, while the second device listens using a quasi omnidirectional BP. Then, the devices exchange their roles to train the other device. After sending feedback, the devices can select the *best* combination of BP on both sides of the link. IEEE 802.11ay also introduces the concept of *in-packet* beam tracking [15], where different antenna configurations can be tested within a single packet, allowing for much quicker BP changes. This is done by appending a TRN field
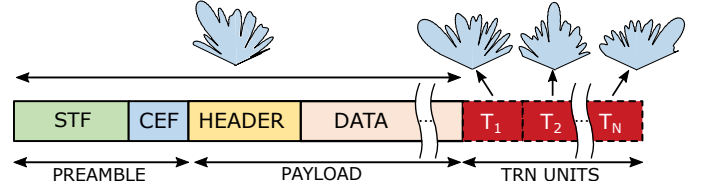

Figure 4: IEEE 802.11ay *in-packet* training packet.

to the packet as shown in Fig. 4. A TRN field is composed of multiple (variable) TRN units formed by 6 complementary Golay sequences of type a ("Ga") and b ("Gb") with length 128 samples each:

$$\{+Ga_{128}; -Gb_{128}; +Ga_{128}; +Gb_{128}; +Ga_{128}; -Gb_{128}\}. \quad (14)$$

The excellent autocorrelation properties of the complementary Golay sequences and the availability fast hardware structures for the correlation [44] make them ideal for CIR estimation [16]. The high bandwidth (1.76 GHz) results in a range resolution of ~ 8.5 cm directly from the CIR estimate. Considering the different BP shapes used during beam training, possible targets located in the FOV of the devices are *illuminated* by the respective BPs that focus energy in that direction and they appear as multi-path components in the CIR (Fig. 5d). Furthermore, we take advantage of the different amplification factors in the multi-path components (given by the different BPs) to estimate the angular positions of the subjects. For this purpose, we apply the correlation based approach explained in Section 3.3.3 to the different channel multi-path components in the channel. Considering the common speed of human motion, carrying out beaconing or a beam training procedure every, e.g., 100 ms allows accurately locating human targets in the FOV of the APs. Note that, as we show in Section 6, a *full* beam training, that scans all the available BPs, is in fact not needed and we may use a much smaller subset of BPs.

Extracting μD signatures from the CIR requires fine-grained frequency resolution, as detailed in Section 3.4.4. This cannot be achieved with the CIR estimates obtained from beacons or beam training packets only, as sampling

the CIR with $T_c = 100$ ms would lead to an insufficient maximum measurable Doppler velocity of $6.25 \cdot 10^{-3}$ m/s (see Eq. (13)). We address this by exploiting the *beam tracking* procedure defined in the standard [43]. It allows to add a configurable number of TRN units to data packets to test different BP configurations to *quickly* correct possible misalignment without requiring a full beam training procedure.

After identifying the subjects' ranges and angles using beam training packets, we include a TRN field in *data packets* with a sufficient number of TRN units to illuminate all the subjects in the scene; each TRN unit uses a suitable BP that specifically points in the direction of a person. This steers the energy of the transmitted signal so as to best capture the μD signatures of the subjects, while maintaining low additional overhead for the data packets. Considering that data packets are sent much more frequently than beam training packets, our approach can sample the CIR with a sufficiently low T_c to capture the desired range of frequencies for human movement analysis.

5 IMPLEMENTATION

The available mmWave Commercial-Off-The-Shelf (COTS) devices support IEEE 802.11ad and offer very limited access to physical layer information [45]. To the best of the authors' knowledge, there are no COTS solutions for the new IEEE 802.11ay standard available yet. To address the lack of hardware, we turn a mmWave SDR system into a JCR experimentation platform. Here we cover the design decisions made to implement RAPID on such platform.

5.1 Hardware components

As a baseline to implement a RAPID AP, we use the mm-FLEX experimental platform [46]. This open platform is composed of a baseband processor including a Xilinx Kintex Ultrascale FPGA plus high-speed AD/DA converters and DDR memory banks. Besides, it is connected through a PCIe interface to a Core i7 processor card co-located within the same hosting chassis. The latter implements configuration and control tasks for the whole system.

The baseband processor is configured to fulfill the bandwidth requirements of IEEE 802.11ad/ay standards (1.76 GHz), using a sampling frequency of 3.52 GSPS for both AD/DA converters, with 2 samples per symbol.

The RF front-end includes a 60 GHz up/down converter and phased antenna arrays from Sivers [47]. The device is able to operate on all the channels defined in the IEEE 802.11ad/ay standards [43], [48]. As shown in Fig. 5a, the device integrates two independent 16-element linear antenna arrays, one used for transmission and one for reception. The codebook of BPs for both arrays can be freely configured. The system is controlled in real-time using USB and SPI interfaces, as well as GPIO pulses for the quick BP changes required for beam training and tracking.

5.2 Full-duplex operation

To bring radar capabilities to the experimentation platform, it is necessary to support simultaneous operation of the TX and RX chains. This is achieved by concurrently enabling transmit and receive sub-systems in the RF front-end, and by enhancing the functionality of the baseband processor.

The 60 GHz front-ends used in this work [47] are laboratory equipment designed for early stage proof-of-concept communication systems. The carrier frequency is generated from a 45 MHz clock, which introduces significant Carrier Frequency Offset (CFO) and destroys the phase coherence between the CIR estimates obtained from consecutive packets. This would make the extraction of μD signatures infeasible with two independent co-located antennas. Instead, by using both transmit and receive arrays from the same RF front-end (see Fig. 5a), up and down conversion subsystems are fed by the same local oscillator which keeps CFO levels in the range of $[-40, 40]$ Hz, as shown in Fig. 5c. Although transmit and receive arrays are directly next to each other, *no complex analog or digital self-interference cancellation techniques are required*. Thanks to the directional BPs and the robustness of the Golay Sequences of the TRN units, the system only requires some transmit power control to avoid saturating the receive antennas and down-conversion stages. In Fig. 5d, we show the CIR measurements obtained from multiple BPs within a packet, by marking the self interference path and the reflections from the test room, where the different amplitudes correspond to the different BP shapes towards the direction of the reflectors.

In the baseband processor, we implement a statemachine on the FPGA logic which controls the transmit and receive data-paths. Specifically, it handles the DDR memory that stores the transmit frames, performs multiple real-time antenna reconfigurations over the TRN field of the packet, triggers the DDR memory on the receive data-path, and sets the inter-frame spacing between multiple transmitted packets. While here we focus on an AP-centric design, the same procedure can be applied to implement RAPID on any station in the network.

Since our RAPID AP operates in a mono-static configuration, we perform CIR extraction without requiring the use of packet detection and synchronization circuits. To do this, it is important to ensure deterministic latency between the transmit and receive data-paths. Considering that transmit and receive data-paths have their own independent clock structure, we use clock-domain crossing techniques to send the state machine signals across transmit and receive domains. Besides, latencies in the DDR controllers are variable, which requires the use of FIFO queues at the output/input of the TX/RX DDRs. Together, these solutions help to achieve the desired deterministic latency.
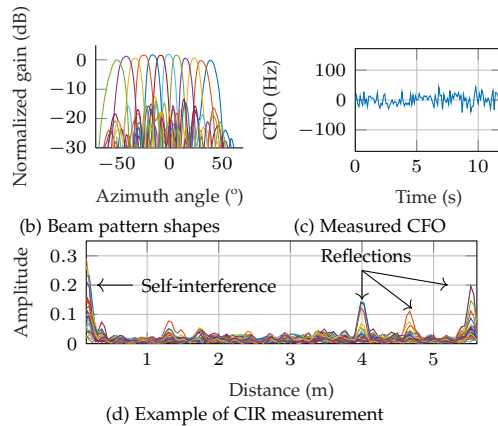
5.3 Multi-AP system

Since IEEE 802.11ay networks typically involve many APs and dense deployments, we extend the aforementioned testbed capabilities to handle multi-AP scenarios. To this end, we integrate a second baseband processor in the hosting chassis which is connected to an independent 60 GHz front-end. Each AP has their its clocking structure, i.e., APs are not synchronized. Each AP can be freely configured with its own parameters. For the sake of simplifying the system management, we use different communication channels (58.32 and 60.48 GHz) for each RF front-end, avoiding cross interference. It is worth mentioning that the channels can
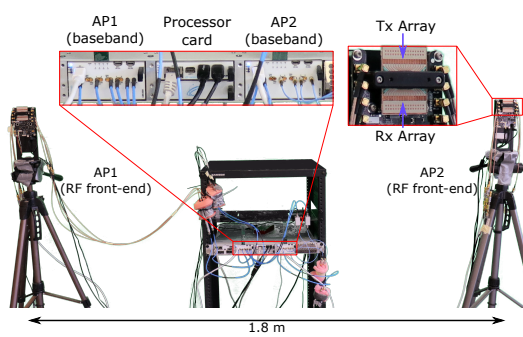
(a) Two RAPID APs deployment

(b) Beam pattern shapes   (c) Measured CFO

(d) Example of CIR measurement
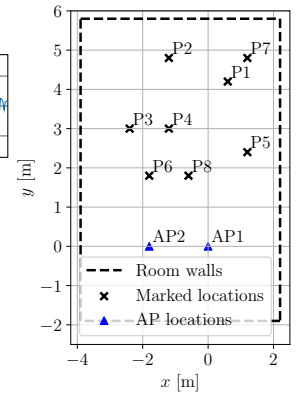
Figure 5: RAPID implementation

Figure 6: Schematic representation of E1.

Figure 7: The two environments: E1 (left) and E2 (right).

be freely configured, making it possible to operate the two RAPID APs so that they share the same frequency band, by implementing carrier sensing mechanisms.

6 EXPERIMENTAL RESULTS

In this section, we discuss the results of our extensive measurement campaign. Motivated by the discussion in Section 3.4.4 and Section 4, for the μD estimation we consider data packets (with TRN fields) spaced by $T_c = 0.27$ ms. This allows capturing velocities in the range $[-4.62, 4.62]$ m/s and leads to a resolution of $\Delta v = 0.14$ m/s when using a window of $M = 64$ samples in the DFT computation, see Eq. (13). These values are comparable to the ones achieved with radar devices [1], [3], [5]. Note that the even spacing of packets is just for convenience but is not a requirement, i.e., estimation can be done with random bursts of data packets with sufficiently small spacing. Moreover, we set to $Q = 9$ the size of the fast time window used to capture the contribution of the subjects in the CIR (see Section 3.4.3). The EKF time-step duration is set to $\Delta t = 32T_c$, which is also the time-granularity at which we obtain μD spectrum vectors. To extract range and angle information, we use in-packet beam training frames with 12 TRN units, using antenna beams covering a FOV range from $-45°$ to $45°$. With this configuration we achieve a mean accuracy of $2°$ for the angular position of a person standing in the room. We verify that this allows tracking multiple subjects reliably and with low localization error, as detailed in the following. In order to implement the angle estimation method from Section 3.3.3, we measured the BP shapes from the codebook using a motorized pan-tilt platform. In Fig. 5b, we show the 12 BPs we used to perform the experiments.

6.1 Experiment setup

We test RAPID in two different rooms, as shown in Fig. 7. The two environments are research laboratories, denoted by E1, of dimensions 6.1×7.7 m and E2, of dimensions 6×10.7 m (E2), and containing whiteboards, windows, tables, computers and equipment, making them challenging multi-path environments with a number of potential reflectors. Most of our experiments, including the collection of the training data for the NN classifier, have been carried out in E1, while we used E2 to assess the robustness of the proposed method to unknown environments. For the tests involving multiple APs, we deploy two RAPID APs as shown in Fig. 5a close to the wall, separated by 1.8 m.

To test the localization and tracking capabilities of RAPID, we mark specific known positions across E1 to determine the ground truth location as shown in Fig. 6, and perform our tests by having subjects move across these positions. The markers are denoted by Px, with x ranging from 1 to 8, while APs are represented as blue triangles. The room walls are represented with a black dashed line.

6.2 Baseline experiments

We first report the results obtained in two simple baseline experiments to verify the capability of RAPID to extract the μD signature of a moving person in an indoor scene. Here, we only use AP 1 and a single subject, performing different activities at different locations in E1.

Fig. 8 shows the EKF estimated trajectory of the subject walking along the trajectory P2-P3-P4-P5-P8-P6 together with the corresponding μD spectrogram. The light grey points represent the raw measurements (observations) obtained as explained in Section 3.3, using Cartesian coordinates. The trajectory is correctly reconstructed with remarkable accuracy. The μD signature is extracted successfully and shows the different contributions of the torso and the limbs. The former reflects more power and follows a slightly oscillating motion, which is coherent with the direction changes in the walking trajectory, while the latter are responsible for the higher velocity peaks.

Next, we test RAPID on a subject sitting down at the marker P2, as shown in Fig. 9. Also in this case, RAPID correctly estimates the location of the subject, and the μD spectrum is coherent with the sitting down activity. This
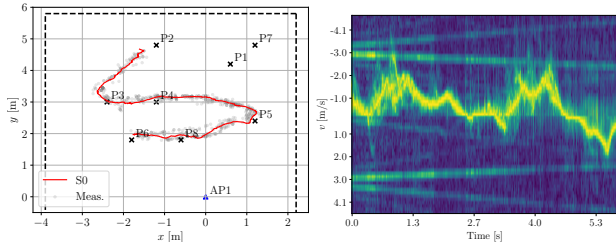
Figure 8: Subject walking trajectory (left) and a portion of the corresponding μD signature (right) extracted by RAPID.
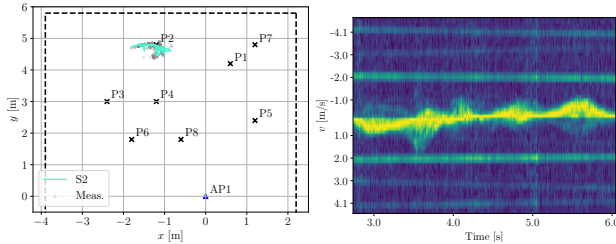

Figure 9: Estimated position of a subject sitting down (left) and a portion of the corresponding μD signature (right) extracted by RAPID.
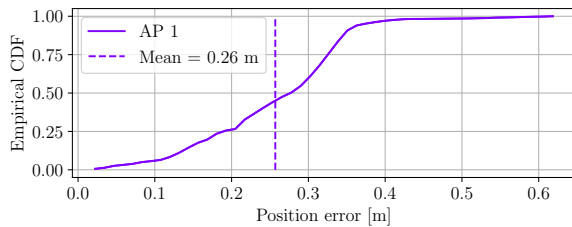

Figure 10: Empirical CDF of the positioning error for a subject sitting down in correspondence of marker P2.

is non-trivial, given that P2 is located at the edge of the experiment room. The empirical Cumulative Distribution Function (CDF) of the positioning error of the subject in Fig. 10 shows that we achieve a good localization accuracy. In this analysis, we included around 2000 position estimates made by the EKF. The median error is 26 cm, and the probability of the error being lower than 40 cm is close to 1. We stress that the subject in this case is not static, as the person alternates between sitting down and standing up. This causes the estimated position to change slightly across time-steps, increasing the localization error.

Our baseline experiments empirically prove that IEEE 802.11ay Golay sequences are adequate for human tracking and μD extraction. This is not trivial, as: *(i)* such sequences are not designed for sensing purposes and they have low Doppler resolution [17]; *(ii)* humans are believed to be poor reflectors of mmWave signals, while we showed that a background subtraction step followed by AoA estimation can reliably identify their contribution to the CIR. While it is well known that human sensing can be performed with mmWave radars employing frequency modulated *chirp* signals [1], [3], [4], RAPID is the first system to do so with mmWave communication waveforms.

6.3 Multi-person multi-AP tracking scenario

In this section, we extend the scenario to analyze the impact of multiple subjects present on the scene, which we tackle
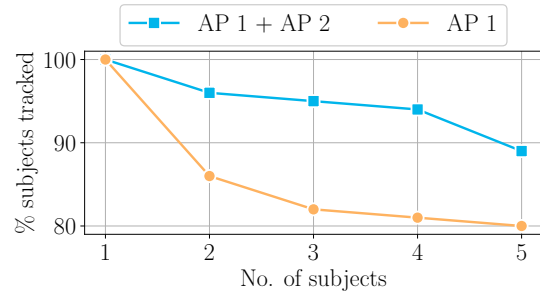

Figure 11: Rate of detection with a varying number of subjects using only AP1 and the combination of AP1 and AP2.

using multiple APs. Here, all measurements are performed using AP1 and AP2 in E1. We first consider the results obtained solely by AP1, and then we combine AP1 and AP2. Several experiments are carried out with 2 to 5 subjects, performing different activities. In total, we collect 28 such sequences each with duration ~ 10 s, of which 13 include 2 subjects, 5 include 3 subjects, 6 include 4 subjects and 4 include 5 subjects. These measurements are collected across different days, spanning a total of 3 weeks.

Presence of multiple subjects. Fig. 12 shows some example trajectories estimated by the EKF using the measurements from AP1. RAPID is able to successfully track the users with considerable accuracy in most cases, even for 5 subjects (see Fig. 12d). Note that this setup is extremely challenging, especially when more than 3 subjects are present, due to the small dimensions of the environment that lead to a high probability of occlusion happening, i.e., one subject covers the Line-of-Sight (LOS) path between the AP and another individual. mmWave signals do not propagate through the human body, and occlusions may cause missed detection and tracking errors. On the other hand, in real-life scenarios occlusions may happen frequently, and the system must be robust to these events. In Fig. 11, we report a quantitative analysis of the effect of increasing the number of subjects in terms of the percentage of subjects that are correctly detected and tracked by RAPID. Using only AP1 we observe that, despite achieving adequate tracking performance, the system capability of detecting the subjects decreases significantly as their number increases. In particular, on average one subject goes undetected when 5 individuals are present.

Improvement with multiple APs. Combining the FOVs of AP1 and AP2 effectively decreases the probability of occlusion events happening, as when the LOS between an AP and a subject is blocked, the other AP can exploit its own LOS path to detect the person. In Fig. 13 we report a qualitative example of this, showing that RAPID can effectively deal with occlusions by combining the FOVs of the 2 APs (in this case, 3 subjects are present in the environment). The EKF estimated trajectories from AP1 are shown in Fig. 13a: subjects S1 and S2 are successfully detected and tracked, while S3, who is waving hands in P3, is not. This is due to a combination of the occlusion caused by S1 and the fact that P3 is placed at the edge of the FOV of AP1. However, the position of AP2 enables it to detect S3 successfully, while the trajectory of S1 can only be partially reconstructed. Considering the trajectories estimated by both APs, RAPID
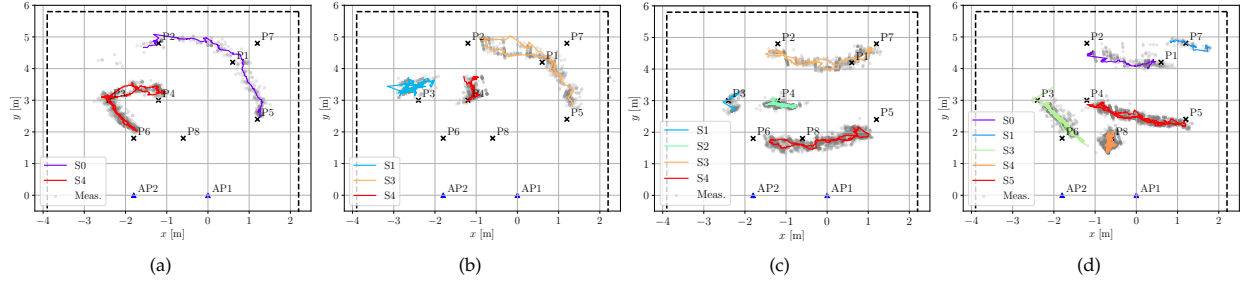
Figure 12: EKF trajectories obtained in the multiperson scenario. Here a single AP is used (AP1). We show four successful cases in which RAPID is able to reconstruct the movement trajectories of 2 **(a)**, 3 **(b)**, 4 **(c)** and 5 **(d)** people moving the the room.

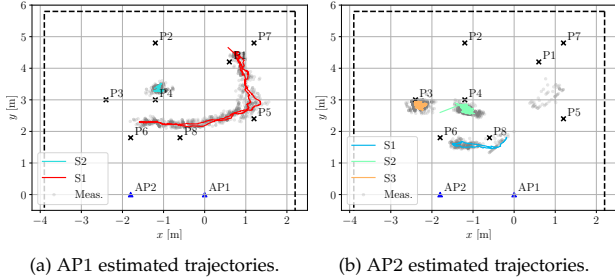(a) AP1 estimated trajectories. (b) AP2 estimated trajectories.

Figure 13: Impact of using multiple APs on the occlusion problem. Here, AP1 fails to detect and track S3, while AP2 can only partially reconstruct the trajectory of S1. The combination of the 2 APs successfully detects and tracks all subjects.
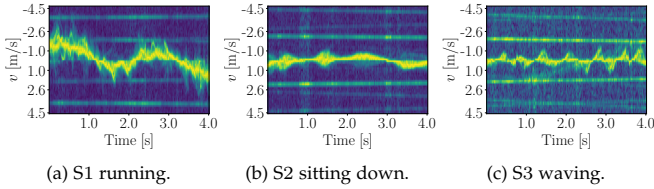

(a) S1 running. (b) S2 sitting down. (c) S3 waving.

Figure 14: Extracted μD signatures of the subjects in Fig. 13.

can detect and track all subjects, successfully extracting their μD signatures, which are reported in Fig. 14.

The subject detection rate is also significantly improved by using multiple APs, as shown in the blue curve in Fig. 11. Despite AP1 and AP2 being placed along the same axis (x), and only 1.8 m apart, this is sufficient to increase subject detection probability by 11%, 16%, 16% and 11% for the cases of 2, 3, 4 and 5 subjects, respectively.

Finally, we show the impact of *averaging* the positions estimated by the two different APs, see Fig. 15. We repeat the experiment described in Section 6.2 with a single subject sitting down in position P2. Even using this simple fusion method, RAPID achieves a significant gain in the tail of the localization error distribution. A subject positioned in P2 represents a worst-case for this kind of analysis in our setting, as the locations of the APs with respect to this point are very similar in terms of distance and angle. The same experiment is repeated for position P4, showing a larger improvement from combining the APs. In this case, RAPID goes from an average localization error of 0.35 m using the single APs independently, down to an error of 0.08 m by averaging their estimates. This is due to the more favorable positions from which P4 is illuminated by the APs.

6.4 Impact of furniture and detection parameters

In this section we analyze the impact of varying the main parameters of the proposed peak detection algorithm, $\alpha_{\max}$ and α_{mean}. To do so, we introduce the following two metrics: the tracking rate (TR) and the false tracks rejection rate (FR). TR is defined as the fraction of time during which RAPID correctly tracks the subject. We consider a subject to be correctly tracked if the EKF outputs a track that has an average tracking error with respect to the reference trajectory lower than 0.4 m. FR is defined as the ratio $1/(N_f + 1)$, where N_f is the number of spurious tracks outputted by the EKF, i.e., those tracks not corresponding to the desired subjects. These can be generated due to false detections and/or reflections on background objects and furniture. In Fig. 17, we report the average TR and FR obtained by varying the parameters of the detection algorithm $\alpha_{\max}$ and α_{mean} from 0.1 to 0.45 and from 1 to 6, respectively. The average is computed over 12 measurement sequences, acquired on two different days, with a subject walking in the room along different trajectories. To evaluate the impact of furniture and obstacles between the RAPID AP and the subject, we placed a table with a monitor, electronic equipment, and two chairs in the measurement space of E1, as shown in Fig. 17. The subject was instructed to walk around and *behind* the table across the four markers shown in Fig. 17b. Fig. 16a and Fig. 16b contain the TR without (w/o) and with furniture (w/). Notice how the presence of obstacles reduces the range of parameters that lead to good TR. To select adequate $\alpha_{\max}$ and α_{mean}, one has to strike a balance between a high TR, which ensures the target is reliably detected and tracked, and high FR, which indicates that the number of false tracks created is low. Lowering the detection parameters yields high TR, but leads to the creation of more undesired tracks, as the sensitivity of the detection is increased. This is shown in Fig. 16c, where we plot the average FR varying the detection parameters in a setup with furniture. Combining the three heatmaps in Fig. 16, one can see that suitable values of $\alpha_{\max}$ are between 0.1 and 0.3, while for α_{mean} we suggest 3 or 4 to avoid generating too many spurious tracks. In the following results, we used $\alpha_{\max} = 0.15$, $\alpha_{\text{mean}} = 3$.

Next, we compute the absolute tracking error between the EKF output trajectory and the ground truth path passing through the four markers. These values are reported in Tab. 1 in the case of no furniture in the room (w/o) and with furniture (w/), along with the corresponding standard deviations. Occlusions due to furniture only slightly degrade the tracking accuracy (5 cm higher error). This is is
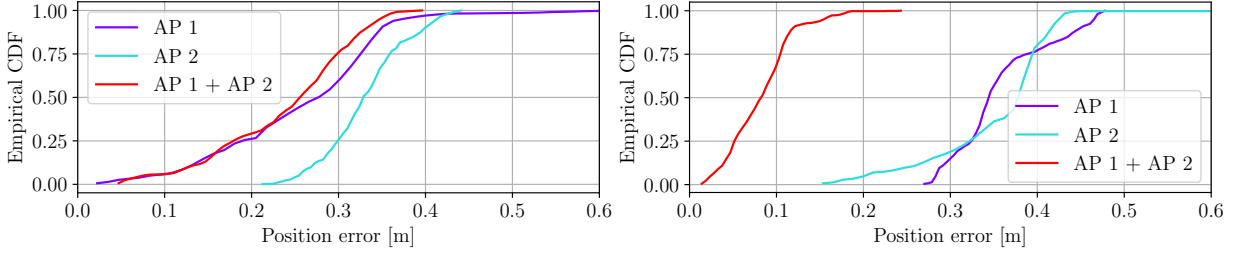
Figure 15: Localization error CDFs for a subject sitting dwon in P2 (left) and in P4 (right). Combining multiple APs brings the largest improvement when their point-of-view on the subject is the most diverse.
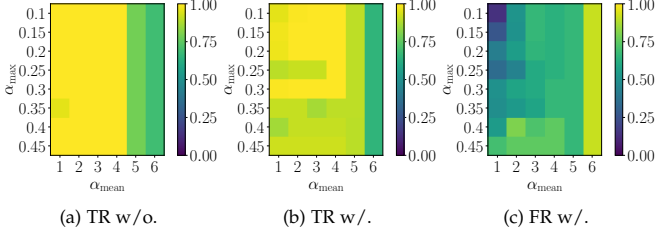

(a) TR w/o.   (b) TR w/.   (c) FR w/.

Figure 16: Mean tracking rate (TR) and false track rejection rate (FR) obtained in a setup without furniture (w/o) and with a table (w/) placed in the sensing area. The heatmap shows the performance for different values of the parameters $\alpha_{\max}$ and $\alpha_{\mean}$.

Table 1: Tracking RMSE without (w/o) and with (w/) furniture, and under complete occlusion of the subject.

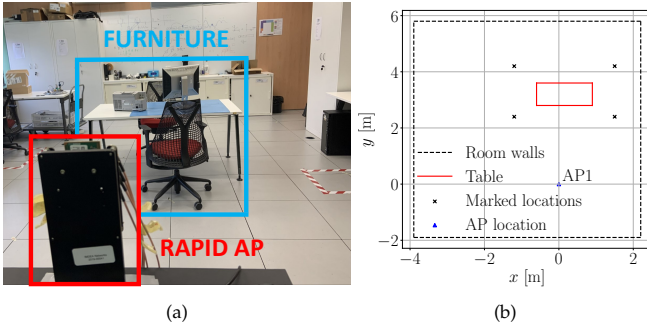|	w/o furniture	w/ furniture	compl. occlusion
RMSE [cm]	20.0 ± 6.3	25.1 ± 8.0	40.9 ± 11.9

(a)   (b)

Figure 17: E1 with furniture (a) and a schematic representation of the table location (b).

due to the fact that, even when furniture is present, RAPID can at least detect the main reflection from the subject's torso most of the time, obtaining precise estimates for the distance between the AP and the person. These are then smoothed across time by the EKF, yielding an accurate trajectory. The last column in Tab. 1 reports the tracking error limited to the part of the estimated trajectory where the subject is completely occluded, so the EKF outputs linear predictions based on past measurements until the subject becomes detectable again. This leads to a noticeable (but still contained) degradation of the tracking performance, which is however expected to drop even further in case of more complex, non-linear movement trajectories.
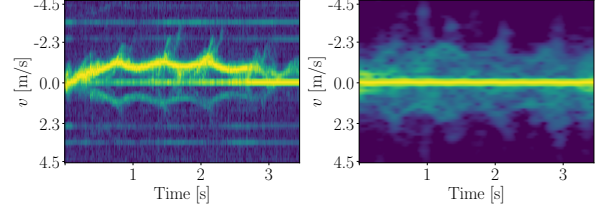

Figure 18: Walking spectrogram concurrently obtained with RAPID at 60 GHz (left) and with sub-6 GHz sensing (right).

6.5 Human activity recognition

Next, we evaluate the HAR performance of RAPID, comparing it to legacy sub-6 GHz WiFi systems. For all the experiments in this section, unless stated otherwise, we used a unique labeled training dataset of simultaneous IEEE 802.11ay CIR (at 60 GHz) and IEEE 802.11ac Channel Frequency Response (CFR) (at 5 GHz) sequences, which we collected in E1, with a single subject performing the 5 different activities A0 − 4. We used a single RAPID AP and a pair of transmitter/receiver IEEE 802.11ac routers with 4 antenna elements (ASUS RT-AC86U implementing the Nexmon-CSI firmware modifications [49]). The estimates are obtained with the two systems operating *(i) concurrently*, i.e., each training/testing sequence for the same activity of the subject is collected with both the RAPID mmWave AP and the sub-6 GHz system, and *(ii)* with the same μD frequency range and resolution. The latter is achieved by tuning the IEEE 802.11ac system inter-packet transmission time using a slight modification of Eq. (13) for the case of non co-located transmitter and receiver, i.e., $\Delta v = c/(f_o^{\mathrm{ac}} M T_c^{\mathrm{ac}})$ with $f_o^{\mathrm{ac}} = 5$ GHz. Therefore, the IEEE 802.11ac inter-packet transmission time is computed as $T_c^{\mathrm{ac}} = 2T_c f_o/f_o^{\mathrm{ac}} \approx 6$ ms. The data are obtained in sequences of approximately 10 s, for a total of around 6 minutes of CIR/CFR measurements per activity. Those sequences are gathered on multiple days over the course of one month. Next, the μD spectrograms are obtained from the collected data. To do this in the sub-6 GHz system, we adopt the pre-processing steps proposed in [50], to which we refer for additional details.

The resulting μD spectrograms are split into partially overlapping windows of 1.728 s, which are the input to the CNN. For RAPID, we use windows containing $N_{\mu D} = 200$ time-steps while for the sub-6 GHz setup each window consists of 287 samples. In Fig. 18 we show an example of the μD signatures obtained by RAPID and by the sub-6 GHz system for the same measurement sequence of a walking person. We use the CNN model detailed in Section 3.5.2 for both mmWave and sub-6 GHz spectrograms. The CNN

Table 2: Confusion matrix and F1-scores for the baseline case. Grey/white rows refer to RAPID and sub-6 GHz, respectively.

True [%]	Predictions [%]				
	Walking	Running	S. down	Waving	Still
Walking	**97.7**	0	2.3	0	0
	61.4	18.3	0	0	20.3
Running	0	**100**	0	0	0
	0.6	87.1	0	0	12.3
S. down	0	0	95.9	0	4.1
	0	0	**100**	0	0
Waving	0	0	0	**100**	0
	0	0	0.1	85.9	14.0
Still	0	0	0	0	100
	0	0	0	0	100
F1-score [%]	**98.8**	**100**	96.3	**100**	**98.4**
	75.8	91.2	**99.9**	92.4	85.4

Table 3: HAR performance under interference from another subject in the training dataset.

F1-score [%]	Walking	Running	S. down	Waving	Still
RAPID	**98.5**	**99.9**	93.2	**100**	**96.6**
Sub-6 GHz	72.2	92.4	**97.8**	58.0	77.8

is trained using the cross-entropy loss function [39] and the Adam optimizer [51], with learning rate 10^{-4}, until convergence of the loss function on a subset of the training data, used as validation set. We evaluate the performance of the classifier with a weighted average of the per-class F1-score metric, based on the number of samples per class. The F1-score is defined as $\texttt{tp}/[\texttt{tp}+0.5(\texttt{fp}+\texttt{fn})]$, where $\texttt{tp}$, $\texttt{fp}$ and $\texttt{fn}$ are the predicted true positives, false positives and false negatives, respectively.

Single person, single AP scenario. In Tab. 2 we report the confusion matrix and per-class F1-scores obtained by RAPID (grey rows) and by the IEEE 802.11ac system (white rows) on test sequences containing data from the same subject present in the training set, collected in E1. This evaluation is also referred to as our *baseline* HAR experiment in the following. Comparing the two systems, one can see that RAPID accurately classifies all activities, only showing slightly lower performance on A2, sitting down, as this mostly involves body movements directed along an orthogonal direction with respect to the receiver (along the vertical axis). Indeed, the motion-induced μD phase displacement is only measurable in the *radial* direction as we rely on the direct path between the subject and the AP. Sub-6 GHz, instead, benefits from a richer multipath environment and better recognizes A2, but confuses the other activities, especially walking with running and standing still. This is due, in part, to the low resolution of the μD obtained at 5 GHz, which contains coarser-grained information (see Fig. 18).

Impact of unknown environment and subject. Next, we further evaluate the HAR robustness of the two systems in more complex settings, involving a different room than the one used for the training data collection (E2), and a different subject performing the activities. Fig. 20 reports the weighted average of the per-class F1-scores obtained with RAPID and the sub-6 GHz system: (a) in the baseline scenario, (b) in a different room, E2, on the same subject (c) with a different subject, in the same environment (E1) and (d) in a different environment (E2) and on a different subject. The results show that RAPID outperforms the sub-6 GHz counterpart in generalizing to new environments and subjects, showing much lower performance degradation when moving to an unknown room or testing on a

different person. In scenario (d) the sub-6 GHz HAR system completely fails, obtaining a very low F1-score, due to the challenging combination of a different room and a different subject. Conversely, RAPID still achieves good performance. We stress that here the training data contain measurements from only one subject. Therefore, the CNN classifier must possess great generalization capabilities to correctly classify the activities performed by another person, as they may have slightly different features.

In addition, we test the two systems under *interference* from another subject in one of the activities of the training set, as shown in Tab. 3. For this, we use the same setting as in the baseline, but we replace the training data for A3, waving hands, with new measurements where another person, termed *interfering subject*, is present in the room besides the subject performing A3. The interfering subject performs a different, randomly selected, activity in each measurement sequence, in a position close to the intended subject, thus possibly disturbing the useful signal reflections. RAPID, thanks to the separation between different subjects enabled by the high ranging accuracy of mmWaves and the tracking process, is highly robust to the presence of other people. Sub-6 GHz sensing, instead, suffers from its low ranging resolution (~ 4 m) and is greatly affected by the interference.

Multi-person, multi-AP scenario. Next, we evaluate RAPID's HAR performance degradation when multiple subjects are concurrently present in the environment, each performing, in general, a different activity. The aim here is to assess the effectiveness of RAPID in the separation of μD signatures associated with different targets. In this evaluation, we do not consider the sub-6 GHz system, as the intrinsic limits in terms of ranging (~ 4 m) and angular ($\sim 20°$) resolutions prevent people tracking in crowded indoor scenarios such as the ones under study [21], thus making the separation of the multiple subjects infeasible.

We collect a labeled training dataset including 6 subjects performing the 5 different activities A0 − 4 using a single RAPID-AP. The data are obtained in sequences of approximately 10 s, and the resulting μD spectrograms are split into windows of 1.728 s as in the single target case. In total, this dataset contains around 2 minutes per activity *per subject* split into multiple captures. These are acquired on different days, over the course of 3 weeks.

By training on different subjects, we aim at mitigating the HAR performance reduction due to the difficulty of generalizing to different people, to better gauge the sole effect of μD separation. We test the trained model on the same multi-person sequences used in Section 6.3, adding 6 additional sequences with a single subject, for a total of 34 sequences. We use the RAPID processing steps to extract the μD signatures of each subject's movement; when using 2 APs, we use the decision fusion scheme from Section 3.5.3.

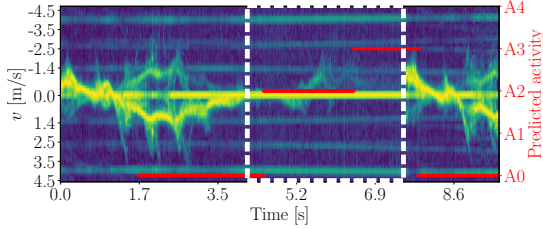Tab. 4 shows the F1-score of RAPID for a varying

Figure 19: μD signature of a walking subject (A0) under temporary occlusion due to the presence of furniture. The dashed white rectangle highlights the occlusion period, while red lines indicate RAPID's predicted activity in each frame.

Table 4: HAR F1-score and detection rate vs. no. of concurrent users.

APs	Metric	1 subj.	2 subj.	3 subj.	4 subj.	5 subj.
1	F1	99.9	99.3	97.9	95.3	94.4
	Det. rate	100	86.1	82.9	81.3	80.0
1 & 2	F1	100	99.4	99.4	95.4	94.4
	Det. rate	100	96.7	95.5	94.5	89.2

number of people in the scene, and the gain obtained by combining the 2 APs with respect to using only AP1. In addition, we also report the corresponding detection rate, previously shown in Fig. 11, for completeness. We observe that the F1-score only slightly decreases when moving from 2 to 5 subjects. This shows that the proposed μD extraction process can reliably separate the contributions of the different individuals. In addition, combining multiple APs can bring a slight improvement in some cases, by exploiting the different illumination angles of the devices.

Impact of occlusions due to furniture. Finally, we evaluate the impact of the presence of furniture on the HAR task, using the same experimental setup described in Section 6.4 (see Fig. 17). We showed previously that RAPID's human tracking and detection is only slightly affected by such occlusions, as the main CIR peaks are still detectable. However, HAR is much more challenging as the quality of the μD signature may be significantly degraded even by *partial* body occlusions, as the key contribution of some body parts (e.g., legs and/or arms) may not be visible in the spectrogram. In Fig. 19, we show the μD obtained from a subject walking (A0) around the table in Fig. 17. The additional y-axis on the right and the red lines represent the activity predicted by RAPID, obtained by sweeping the $N_{\mu D}$ frames long CNN input window over the μD. After an initialization time needed to collect the first window, RAPID correctly classifies activity A0. During the subsequent occlusion event (enclosed in the dashed white rectangle), the torso reflection in the μD becomes much fainter, while the contributions of the other body parts disappear, causing misclassifications (A2 and A3). We stress that this is due to an intrinsic limitation of mmWave signals. In fact, in case of occlusion the information about the different body parts is mostly *undetectable* at the receiver, hence the available information about the movement is insufficient to correctly classify it. However, RAPID promptly re-establishes the correct classification when the subject becomes visible again. Note that RAPID is not trained with samples including occlusion events in the training set, which makes this test even more challenging.
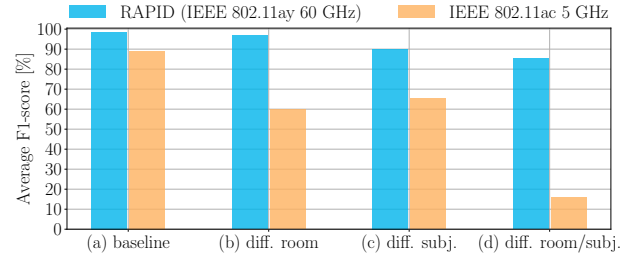

Figure 20: Comparison between the HAR F1-score obtained by RAPID and by standard IEEE 802.11ac sensing at 5 GHz for various scenarios.

6.6 Person identification

In this section we test the performance of RAPID on person identification, by building a dataset including the gait μD spectrograms of 7 subjects, collected in E1. We collect from 3 to 5 minutes of training data per subject, split into multiple captures acquired over the course of 10 days. The data from each subject is the collection of captures obtained on different days, to avoid slight daily variations in the gait to bias the dataset. The input samples for the classifier are obtained using μD windows of the same length as for HAR, i.e., 1.728 s. The CNN classifier is trained using the same parameters and loss function used for HAR.

Person identification accuracy. First, we evaluate the accuracy of person identification on a varying number of subjects to recognize. In Tab. 5 we report the accuracy values obtained by RAPID when increasing the number of subjects from 2 to 7. The obtained values are not significantly lower from those obtained with mmWave radars, and in some cases even superior, e.g., the 79% on 5 subjects in [1], the 98% with 4 subjects in [3] or the 89% with 12 subjects in [5]. This is even more valuable considering the few available training data and the short duration of the observation window used, compared to the windows used in the mentioned papers which vary between 2 and 3 s.

Continuous HAR and person identification. Finally, we show that RAPID is capable of simultaneously *(i)* tracking subjects, *(ii)* recognizing their activities, and *(iii)* identifying who is performing each activity from their gait. We perform several tests in which 2 subjects, *concurrently* present in the room, perform various activities sequentially, e.g., walking then sitting, etc. In this scenario, people tracking is of key importance to collect the temporal evolution of each subject's μD, so that all the activities performed by a person can be associated to that person's identity, obtained by RAPID when he/she is walking.

In Fig. 21 we show the results obtained by RAPID with 2 subjects, S0 and S1, behaving as follows. S0 enters the scene walking, then after approximately 3.5 s S0 stops and starts waving hands, while S2 is sitting down and then starts walking after 3.5 s. We report the μD signature extracted after successfully tracking the subjects, along with the predicted activity using our moving window approach. We observe that RAPID detects the change in the activity performed by each subject; moreover, by applying the identification CNN to the spectrogram portion where the subjects are walking, it successfully identifies them as S0 and S1 among the 7 subjects in the training set.

Table 5: Identification accuracy vs. number of subjects.

	2 subj.	3 subj.	4 subj.	5 subj.	6 subj.	7 subj.
Acc. [%]	97.8	95.9	94.6	94.1	92.7	90.0

(a) S1 walking-waving. (b) S0 sitting down-walking.
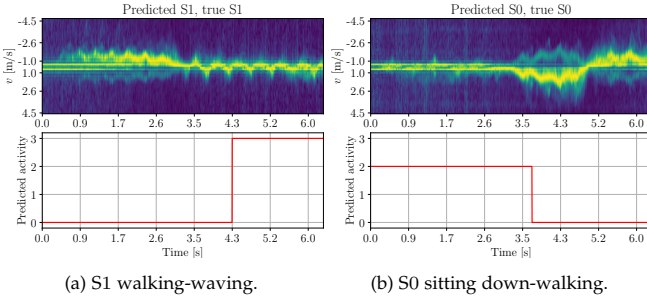
Figure 21: μD signature and corresponding CNN output when subject 0 is sitting down (A2), then starts walking (A0), while subject 1 is walking and then starts waving hands (A3).

6.7 Overhead considerations

The sensing operations performed by RAPID add a certain overhead to the communication process, due to: *(i)* appending TRN units to the communication packets, which entails transmitting redundant bits that do not carry information; *(ii)* accessing and occupying the channel for sensing, which may interfere with other communication links in the proximity of the AP. In this section, we discuss point *(i)*, while we already addressed point *(ii)* in a separate work [52]. There, we proposed a method to reconstruct μD signatures from the irregular and sparse CIR estimates obtained from communication packets, so as to minimize the need to access the channel for the sole purpose of sensing.

We can assess the overhead of RAPID by comparing the PHY layer packet size in IEEE 802.11ay to the size of TRN fields used for sensing. As shown in Fig. 4, physical layer Protocol Data Units (PDUs) include the Short Training Field (STF), the CEF and the PHY layer header, including $\text{STF}_l = 2176$, $\text{CEF}_l = 1152$ and $\text{PHY}_l = 1024$ samples, respectively [43]. Each TRN field includes 6 complementary Golay sequences, for a total of $\text{TRN}_l = 768$ samples. Therefore, the overhead introduced by appending ξ TRN fields to a packet is

$$O = \frac{\text{TRN}_l \cdot \xi}{\text{STF}_l + \text{CEF}_l + \text{PHY}_l + \text{DATA}_l + \text{TRN}_l \cdot \xi}, \quad (15)$$

where DATA_l is the length of the data portion of the packet. We recall that, with RAPID, it is sufficient to illuminate a person with *one* BP to apply the extraction of the μD spectrum, and that we can use one BP per TRN field, so ξ can be selected equal to the number of subjects tracked by RAPID. In order to reduce the inefficiency of the MAC layer and achieve Gigabit data rates, in IEEE 802.11ay large packet aggregation is permitted, allowing PHY layer PDUs to contain up to 4 MB of data. For this, multiple MAC layer PDUs of 1.5 kB are encapsulated into a single PHY layer packet. Compared to these large packet sizes, the TRN fields used by RAPID add a limited amount of overhead. To see this, consider that, e.g. Modulation and Coding Scheme (MCS) 8 is used, and that the data size is 20 kB (note that is a very small fraction of the maximum allowed aggregated packet size). Even in this conservative example

we get $\text{DATA}_l = 126784$ samples (due to the MCS used) [43], leading to $O = 0.6 \cdot \xi\%$. Moreover, RAPID does not require the TRN fields to be appended to *every* PHY layer packet, but only to one every T_c seconds. With $T_c = 0.27$ ms as in our implementation, considering the same data size used above and the IEEE 802.11ay sample rate of 1.76 Gsps, we get that the TRN fields need to be added to only one out of $3 - 4$ PHY layer PDUs, further reducing the overhead.

As a final note, we stress that RAPID performs sensing using reflections of standard-compliant packets, i.e., the transmitted packets are not designed for sensing, but we rather exploit some properties of the standard itself to enable JCR. While the header and payload of the packet are transmitted with the BP that maximizes the communications quality towards the intended receiver, the appended TRN fields can be transmitted with an arbitrary BP. To obtain a signal reflection to be used for sensing, we use the BP that illuminates the target. Therefore, the sensing operations in RAPID do not interfere with communication besides the addition of TRN fields, which have a small impact on the overall throughput, as discussed above.

7 CONCLUDING REMARKS

In this paper, we have designed and implemented RAPID, the first mmWave JCR system performing high-resolution sensing of human μD signatures through standard-compliant IEEE 802.11ay packets. RAPID uses the in-packet TRN fields, as specified by the 802.11ay standard, to estimate the channel impulse response. This makes it possible to perform joint tracking and localization of multiple people freely moving in an indoor environment. In addition, their μD signatures are extracted by analyzing the phase difference between subsequent packets, which enables advanced sensing tasks such as continuous HAR and person identification, with radar-level accuracy. RAPID successfully combines the high-resolution sensing capabilities of mmWave radars with the scalability and ease of deployment of existing communication hardware, allowing the seamless integration of multiple APs. We implemented two RAPID APs with full-duplex capabilities on an FPGA-based SDR platform equipped with phased antenna arrays, and we have thoroughly evaluated the system performance through an extensive measurement campaign. Our results show that 2 combined RAPID-APs can track up to 5 subjects concurrently moving in an indoor environment, achieving accuracies of up to 94% and 90% for HAR and person identification, respectively. Moreover, in HAR, RAPID performs significantly better than standard sub-6 GHz sensing, showing better capability of distinguishing similar activities and generalizing to new environments and unkwnown subjects.

Future research directions includes the combination of our system with sub-6 GHz radios, to benefit from the points of strength of both frequency domains: while mmWave signals are ideal for localization, tracking and μD extraction, systems operating at lower frequencies can improve the recognition of movements that do not involve a large displacement in the radial direction with respect to the receiver (e.g., sitting down), thanks to their richer multipath environment. Other research avenues include *(i)* extending the RAPID system to bistatic network configurations, where

sending and receiving units do not share a common phase reference, *(ii)* devising additional data fusion strategies for multiple APs and *(iii)* assessing the sensing performance limits at mmWaves as a function of number and location of the APs, size of the indoor space to be monitored, people density and number and type of objects in the environment, which may lead to occlusions and spurious reflections.

REFERENCES

[1] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. De Neve, and T. Dhaene, "Indoor person identification using a low-power FMCW radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, 2018.

[2] A.-K. Seifert, M. G. Amin, and A. M. Zoubir, "Toward unobtrusive in-home gait analysis based on radar micro-doppler signatures," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2629–2640, 2019.

[3] J. Pegoraro, F. Meneghello, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 2994–3009, 2021.

[4] Z. Meng, S. Fu, J. Yan, H. Liang, A. Zhou, S. Zhu, H. Ma, J. Liu, and N. Yang, "Gait Recognition for Co-Existing Multiple People Using Millimeter Wave Sensing," in *AAAI Conference on Artificial Intelligence*, (New York, New York, USA), Feb 2020.

[5] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "mID: Tracking and Identifying People with Millimeter Wave Radar," in *15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, (Santorini Island, Greece), May 2019.

[6] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, pp. 51–56, 2019.

[7] F. Jin, A. Sengupta, and S. Cao, "mmFall: Fall Detection Using 4-D mmWave Radar and a Hybrid Variational RNN AutoEncoder," *IEEE Transactions on Automation Science and Engineering*, 2020.

[8] C. Chen, H. Song, Q. Li, F. Meneghello, F. Restuccia, and C. Cordeiro, "Wi-fi sensing based on ieee 802.11 bf," *IEEE Communications Magazine*, 2022.

[9] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.

[10] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.

[11] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, 2018.

[12] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mmTrack: Passive multi-person localization using commodity millimeter wave radio," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, pp. 2400–2409, IEEE, 2020.

[13] F. Zhang, C. Wu, B. Wang, and K. R. Liu, "mmEye: Super-Resolution Millimeter Wave Imaging," *IEEE Internet of Things Journal*, 2020.

[14] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath, Z. Feng, L. Zheng, and A. Petropulu, "An Overview of Signal Processing Techniques for Joint Communication and Radar Sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, pp. 1295–1315, Nov 2021.

[15] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly, "IEEE 802.11ay: Next-Generation 60 GHz Communication for 100 Gb/s Wi-Fi," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 186–192, 2017.

[16] J. O. Lacruz, R. Ruiz, and J. Widmer, "A Real-Time Experimentation Platform for sub-6 GHz and Millimeter-Wave MIMO Systems," in *ACM MobiSys'21*, 2021.

[17] P. Kumari, N. Gonzalez-Prelcic, and R. W. Heath, "Investigating the IEEE 802.11ad Standard for Millimeter Wave Automotive Radar," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, pp. 1–5, 2015.

[18] H. Li, X. He, X. Chen, Y. Fang, and Q. Fang, "Wi-motion: A robust human activity recognition using WiFi signals," *IEEE Access*, vol. 7, pp. 153287–153299, 2019.

[19] F. Meneghello, D. Garlisi, N. D. Fabbro, I. Tinnirello, and M. Rossi, "Environment and Person Independent Activity Recognition with a Commodity IEEE 802.11 ac Access Point," *arXiv preprint arXiv:2103.09924*, 2021.

[20] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pp. 1230–1239, IEEE, 2017.

[21] B. Korany, H. Cai, and Y. Mostofi, "Multiple People Identification Through Walls Using Off-The-Shelf WiFi," *IEEE Internet of Things Journal*, vol. 8, pp. 6963–6974, Apr 2021.

[22] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, *et al.*, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pp. 289–304, 2018.

[23] Z. Shi, J. A. Zhang, R. Y. Xu, and Q. Cheng, "Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning," *IEEE Transactions on Mobile Computing*, vol. 21, pp. 540–554, Feb 2022.

[24] A. Davoli, G. Guerzoni, and G. M. Vitetta, "Machine Learning and Deep Learning Techniques for Colocated MIMO Radars: A Tutorial Overview," *IEEE Access*, 2021.

[25] G. Lai, X. Lou, and W. Ye, "Radar-Based Human Activity Recognition With 1-D Dense Attention Network," *IEEE Geoscience and Remote Sensing Letters*, 2021.

[26] J. Pegoraro and M. Rossi, "Real-time People Tracking and Identification from Sparse mm-Wave Radar Point-clouds," *IEEE Access*, vol. Early Access, 2021.

[27] I. Pefkianakis and K.-H. Kim, "Accurate 3D Localization for 60 GHz Networks," (New York, NY, USA), Association for Computing Machinery, 2018.

[28] S. D. Regani, C. Wu, B. Wang, M. Wu, and K. R. Liu, "mmWrite: Passive Handwriting Tracking Using a Single Millimeter Wave Radio," *IEEE Internet of Things Journal*, 2021.

[29] Y. Ren, J. Lu, A. Beletchi, Y. Huang, I. Karmanov, D. Fontijne, C. Patel, and H. Xu, "Hand gesture recognition using 802.11 ad mmWave sensor in the mobile device," in *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pp. 1–6, IEEE, 2021.

[30] F. Wang, F. Zhang, C. Wu, B. Wang, and K. R. Liu, "ViMo: Multi-person Vital Sign Monitoring using Commodity Millimeter Wave Radio," *IEEE Internet of Things Journal*, 2020.

[31] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Las Vegas, Nevada, USA), Jun 2016.

[33] A. Nambiar, A. Bernardino, and J. C. Nascimento, "Gait-based person re-identification: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, pp. 1–34, Apr 2019.

[34] D. Garcia, J. O. Lacruz, P. Jiménez Mateo, and J. Widmer, "PO-LAR: Passive object localization with IEEE 802.11ad using phased antenna arrays," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, pp. 1838–1847, 2020.

[35] M. I. Ribeiro, "Kalman and extended kalman filters: Concept, derivation and properties," *Institute for Systems and Robotics*, vol. 43, p. 46, 2004.

[36] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *11th international conference on information fusion (FUSION)*, (Cologne, Germany), pp. 1–6, IEEE, Jun 2008.

[37] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, no. 6, pp. 82–100, 2009.

[38] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 34, pp. 22–35, Mar 2017.

[39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[40] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," in

International Conference on Learning Representations (ICLR), (San Juan, Puerto Rico), May 2016.

[41] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *International Conference on Machine Learning (ICML)*, (Lille, France), Jul 2015.

[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of machine learning research*, vol. 15, pp. 1929–1958, Jun 2014.

[43] IEEE 802.11 working group, "IEEE Draft 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-Amendment: Enhanced Throughput for Operation in License-Exempt Bands Above 45 GHz," *IEEE P802.11ay/D3.0*, 2019.

[44] W.-C. Liu, F.-C. Yeh, T.-C. Wei, C.-D. Chan, and S.-J. Jou, "A Digital Golay-MPIC Time Domain Equalizer for SC/OFDM Dual-Modes at 60 GHz Band," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 10, p. 10, 2013.

[45] D. Steinmetzer, D. Wegemer, M. Schulz, J. Widmer, and M. Hollick, "Compressive millimeter-wave sector selection in off-the-shelf IEEE 802.11 ad devices," in *Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies*, pp. 414–425, 2017.

[46] J. O. Lacruz, D. Garcia, P. J. Mateo, J. Palacios, and J. Widmer, "mm-FLEX: An Open Platform for Millimeter-Wave Mobile Full-Bandwidth Experimentation," in *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, MobiSys '20, (New York, NY, USA), p. 1–13, Association for Computing Machinery, 2020.

[47] SIVERSIMA, *EVK06002 Development Kit*, 2020. https://www.siversima.com/product/evk-06002-00/.

[48] IEEE 802.11 working group, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band.," *IEEE Standard 802.11ad*, 2012.

[49] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets," in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization (WiNTECH)*, (Los Cabos, Mexico), Oct. 2019.

[50] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*, (Seoul, Republic of Korea), 2019.

[51] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic gradient descent," in *ICLR: International Conference on Learning Representations*, pp. 1–15, 2015.

[52] J. Pegoraro, J. O. Lacruz, M. Rossi, and J. Widmer, "SPARCS: A Sparse Recovery Approach for Integrated Communication and Human Sensing in mmWave Systems," in *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, (Milan, Italy), IEEE, May 2022.

Jesus O. Lacruz is a Research Engineer at IMDEA Networks, Spain since 2017. He received his Bachelor degree in Electrical Engineering from Universidad de Los Andes, Venezuela in 2009 and the PhD degree in Electronic Engineering from Universidad Politecnica de Valencia, Spain in 2016. His research interests lie in the design and implementation of fast signal processing algorithms for digital communication systems in FPGA devices.

Francesca Meneghello (S'19) received the Ph.D. degree in information engineering from the University of Padova, Italy, in 2022. She is currently an assistant professor at the Department of Information Engineering at the same university. Her current research interests include deep-learning architectures and signal processing with application to remote radio frequency sensing and wireless networks.

Enver Bashirov (S'20) is currently an early-stage researcher at EU Horizon 2020 Marie Skłodowska-Curie project MINTS, pursuing his Ph.D. degree at the Department of Information Engineering, University of Padova, Italy. He received his M.Sc. degree in Applied Mathematics and Computer Science from Eastern Mediterranean University, North Cyprus. His research interests include sensing applications in mmWave, together with machine learning and signal processing solutions.

Michele Rossi (SM'13) is the head of the Master's Degree in ICT for internet and Multimedia (MIME) and full professor at the Department of Information Engineering of the University of Padova. Since 2017, he has been the Director of the DEI/IEEE Summer School of Information Engineering (SSIE), held yearly in Brixen, Italy. He is also the coordinator of the GREENEDGE (no. 953775) ITN project on "green edge computing for mobile networks". His research interests are on wireless sensing and edge computing systems with a focus on green ICT technologies

Jacopo Pegoraro (S'20) received his Ph.D. in Information Engineering from the University of Padova, Padua, Italy, in 2023. He is currently working as a postdoctoral researcher in the Department of Information Engineering, in the same University. He was a visiting research scholar at the New York University, Tandon school of Engineering in 2022. His research interests include signal processing and machine learning for mmWave sensing and integrated sensing and communication.

Joerg Widmer (F'20) is Research Professor and Research Director of IMDEA Networks in Madrid, Spain. His research focuses on wireless networks, ranging from extremely high frequency millimeter-wave communication and MAC layer design to mobile network architectures. He authored more than 150 conference and journal papers, three IETF RFCs and holds 13 patents. He was awarded an ERC consolidator grant, the Friedrich Wilhelm Bessel Research Award, a Spanish Ramon y Cajal grant, as well as eight best paper awards.