



Computer Science and Artificial Intelligence Laboratory  
Technical Report

MIT-CSAIL-TR-2014-008

April 26, 2014

---

**Multi-Person Motion Tracking via RF Body Reflections**  
Fadel Adib, Zachary Kabelac, and Dina Katabi

# Multi-Person Motion Tracking via RF Body Reflections

Fadel Adib Zachary Kabelac Dina Katabi

**Abstract**— Recently, we have witnessed the emergence of technologies that can localize a user and track her gestures based purely on radio reflections off the person’s body. These technologies work even if the user is behind a wall or obstruction. However, for these technologies to be fully practical, they need to address major challenges such as scaling to multiple people, accurately localizing them and tracking their gestures, and localizing static users as opposed to requiring the user to move to be detectable.

This paper presents WiZ, the first multi-person centimeter-scale motion tracking system that pinpoints people’s locations based purely on RF reflections off their bodies. WiZ can also locate static users by sensing minute changes in their RF reflections due to breathing. Further, it can track concurrent gestures made by different individuals, even when they carry no wireless device on them.

We implement a prototype of WiZ and show that it can localize up to five users each with a median accuracy of 8-18 cm and 7-11 cm in the x and y dimensions respectively. WiZ can also detect 3D pointing gestures of multiple users with a median orientation error of 8 – 16° for each of them. Finally, WiZ can track breathing motion and output the breath count of multiple people with high accuracy.

## 1. INTRODUCTION

For many years, the wireless channel abstraction has involved data communication between an RF transmitter and an RF receiver. Recent advances in wireless technologies, however, have demonstrated that a person’s motion can modulate the wireless signal, enabling the transfer of information from a human to an RF transceiver, even when the person does not carry a transmitter [5, 14, 4]. This leads to a new abstraction of the wireless channel, in which a user may communicate with remote devices over the wireless medium directly using gestures, a much more natural interface for mobile computing than a keypad or a touch screen. The new channel abstraction also allows for a direct extraction of information from the environment. For example, one may track objects and people as they move around, purely based on how their motion modulates the wireless signal. This could lead to new video games and virtual reality applications that work in non-line-of-sight and across rooms. It can also be used for health-care monitoring in hospitals or at home (e.g., monitoring dementia and Alzheimer patients), and for intrusion detection or search-and-rescue operations. Hence, the new channel abstraction blurs the boundaries between wireless communications and human-computer interaction, providing a more holistic view of these disciplines that is better aligned with the emerging world of mobile computing.

Motivated by this vision, this paper takes another leap toward enriching this new channel model. We particularly focus on multi-person motion tracking using only RF reflections. Accurate tracking of a person’s body and body parts is a core enabling primitive for this new channel because it can be used both to extract information from the environment and track its moving bodies, and to communicate commands using hand gestures. Past work that delivers centimeter-scale tracking accuracy can localize only one person, and only if the person is moving [4]. Multi-person tracking based on body reflections is intrinsically difficult. Movements of different people all modulate the same wireless signal causing interference. Rather than avoiding interference by assuming that only one person moves at any time, we aim to tackle and overcome this interference problem.

We present WiZ, the first multi-person centimeter-scale motion tracking system that operates purely using RF reflections off a person’s body. WiZ can also accurately localize static people using their breathing motion, and can further count their breaths. It can also track body parts, enabling multiple people to simultaneously interact with the environment via hand gestures.

To achieve its goal, WiZ has to overcome multiple challenges. In particular, state-of-art centimeter-scale tracking measures distances using the signal’s time of flight (TOF) – that is the time it takes the signal to travel from the radio to a reflector and back [4]. The TOF can be easily mapped to a distance by multiplying it with the speed of light. However, when there are multiple people, they all modulate the same wireless signal, making it difficult to disentangle the TOFs of each individual. The problem is exacerbated in indoor settings where people are confined to a small space and hence their TOFs are naturally close. Furthermore, multipath reflections can create fictitious TOFs which further complicate the problem. To address these challenges, WiZ builds on Frequency Modulated Carrier Waves (FMCW), a radar technique that provides TOF measurements. WiZ introduces multi-shift FMCW, a multi-antenna extension to FMCW where the signal transmitted by different antennas is structured in a particular way to disentangle the TOFs corresponding to different people and eliminate the impact of fictitious TOFs that do not correspond to a physical target. In §4, we describe multi-shift FMCW in detail.

A second challenge that WiZ has to address is related to the near-far problem. Nearby reflectors can have significantly more power than distant reflectors, obfuscating the signal from people in the back and preventing their detection or tracking. To address the near-far problem, WiZ introduces successive silhouette cancellation (SSC). This approach is

inspired by successive interference cancellation, where the receiver first focuses on the strong signal, decodes it, and subtracts it from the received signal to enable the decoding of weaker signals. The main difference is that decoding in our context means localizing the person using her TOF measurements. Once we have decoded a person’s location, we have to figure out how a reflection from a person at that location would impact the received signals and cancel that impact. Doing so allows us to successively eliminate strong reflectors that could completely hide far away people. We keep doing so until we have decoded all people in the scene.

Finally, to localize breathing, one needs to realize that the breathing motion is fairly slow in comparison with body motion. The chest moves by a sub-centimeter distance over a period of few seconds. In contrast, a human would pace indoors at 1 m/s. To detect reflectors with slow motions without confusing them with static reflectors (e.g., furniture and walls) and without obscuring them with fast reflectors, (e.g. body motion), WiZ processes the received signal at multiple time scales: a short time scale to detect moving bodies and quickly pinpoint their location before it changes, and a longer time scale that allows slowly moving objects, like a breathing chest, to move enough so that they become detectable.

We have built a prototype of WiZ using USRP software radios and an analog FMCW radio. In our evaluation, we use the VICON motion capture system to report the ground truth location [3]. VICON can achieve millimeter localization accuracy but requires instrumenting the human body with infrared markers and positioning an array of infrared cameras on the ceiling. We run experiments both in line-of-sight (LOS) scenarios and non-line-of-sight (NLOS) scenarios, where the device is in a different room and is tracking people motion through the wall. Empirical results from over 300 experiments with 11 human subjects show the following:

- *Motion Tracking:* WiZ accurately tracks the motion of four people when the device is in the room where the motion occurs, and three people when the device is placed behind the wall. Its median error is 8.4 cm and 7.2 cm in  $x/y$  for the nearest person for both the through-wall and line-of-sight experiments, and remains less than 16.1 cm and 10.5 cm in  $x/y$  for the furthest person in the scene.
- *Localizing Static People:* By tracking their breathing motion, WiZ accurately localizes up to five static people in line-of-sight and four static people through a wall. Its median error is less than 7.2 cm and 6.3 cm in  $x/y$  for the nearest person in both through-wall and line-of-sight experiments and remains less than 18.3 cm and 10.9 cm for the furthest person in the scene.
- *Breath Counting:* In the above experiments, WiZ was able to count the number of breaths taken by every participant. Its counting error is less than one breath for over 97% of our experiments – each of which lasted for 3-4 minutes. Also WiZ was able to detect occasions when the user accidentally held her breath.

- *Gestures:* WiZ can recognize concurrent gestures performed in 3D space by multiple users. In particular, we consider a gesture in which three users point in different directions at the same time. For example the users may be playing a virtual shooting game, or may want to control different appliances around by pointing at them. Our WiZ prototype detect all the pointing directions of all three users with a median accuracy of  $8.2^\circ$  and  $16^\circ$ , for the nearest and furthest user respectively.

**Contributions:** This paper presents WiZ, the first multi-person centimeter-scale motion tracking system that operates using RF reflections off people’s bodies. It works for both moving and static people and can further count people’s breath and track multiple concurrent gestures. These capabilities are enabled by successive silhouette cancellation and multi-shift FMCW, two innovative techniques for computing the time of flight (TOF) of multiple reflectors from different perspectives and mapping these TOFs to accurate estimates of the locations of the reflecting bodies even in the presence of near-far interference.

## 2. BACKGROUND

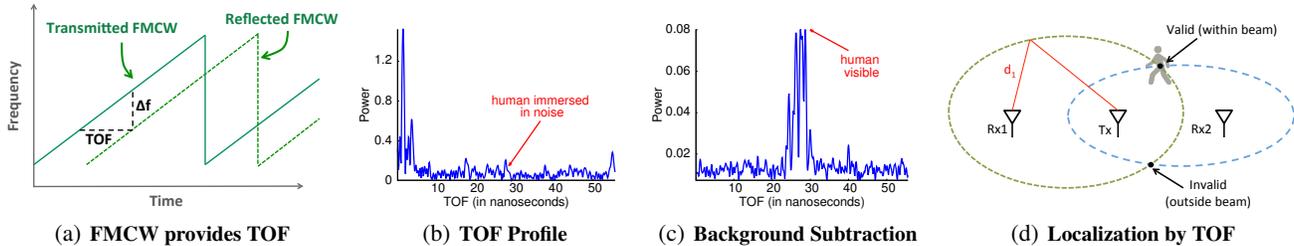
This section provides necessary background regarding single-person motion tracking via RF body reflections.

The process of localizing a user based on radio reflections off her body has three steps: 1) obtaining time-of-flight (TOF) measurements to various reflectors in the environment; 2) eliminating TOF measurements due to reflections of static objects like walls and furniture; and 3) mapping the user’s TOFs to a location.

**Obtaining TOF measurements.** A typical way for measuring the time-of-flight (TOF) is to use a Frequency-Modulated Carrier Waves (FMCW) radio. An FMCW transmitter sends a narrowband signal (e.g., a few KHz) but makes the carrier frequency sweep linearly in time, as illustrated by the solid green line in Fig. 1(a). The reflected signal is a delayed version of the transmitted signal, which arrives after bouncing off a reflector, as shown by the dotted green line in Fig. 1(a). Because time and frequency are linearly related in FMCW, the delay between the two signals maps to a frequency shift  $\Delta f$  between them. Hence, the time-of-flight can be measured as the difference in frequency  $\Delta f$  divided by the slope of the sweep in Fig. 1(a):

$$TOF = \Delta f / slope \tag{1}$$

This description generalizes to an environment with multiple reflectors. Because wireless reflections add up linearly over the medium, the received signal is a linear combination of multiple reflections, each of them shifted by some  $\Delta f$  that corresponds to its TOF. Hence, one can extract all these TOFs by taking an FFT of the received signal. The output of the FFT gives us the **TOF profile** which we define as the reflected power we obtain at each possible TOF between the transmit antenna and receive antenna, as shown in Figs. 1(b) and 1(c).



**Figure 1—Localization by TOF measurements.** (a) shows the transmitted FMCW signal and its reflection. The TOF between the transmitted and received signals maps to a frequency shift  $\Delta f$  between them. (b) shows the TOF profile obtained after performing an FFT on the baseband FMCW signal. The profile plots the amount of reflected power at each TOF. (c) shows that a moving person’s reflections pop up after performing background subtraction. (d) shows how we can use TOF measurements from multiple receivers, map them to round-trip distance measurements, and localize the user by trilateration.

**Eliminating TOFs of static reflectors.** To localize a human, we need to identify his/her reflections from those of other objects in the environment (e.g., walls and furniture). This may be done by leveraging the fact that the reflections of static objects remain constant over time. Hence, one can eliminate the power from static reflectors by performing background subtraction – i.e., by subtracting the output of the TOF profile in a given sweep from the TOF profile of the signal in the previous sweep. Fig. 1(c) and 1(b) show how background subtraction eliminates the power in static TOFs from the TOF profile, and allows one to notice the weak power resulting from a moving person.

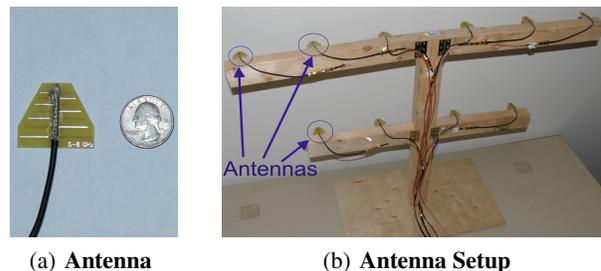
**Localization using TOF measurements.** Recall that the TOF corresponds to the time it takes the signal to travel from the transmitter to a reflector and then back to the receiver. Therefore, one can compute the corresponding round-trip distance by multiplying this TOF by the speed of light  $C$  as follows:

$$\text{round trip distance} = C \times \text{TOF} = C \times \frac{\Delta f}{\text{slope}} \quad (2)$$

Knowing the round trip distance localizes the person to an ellipse whose foci are the transmit and receive antennas. To localize a person in 2D, one needs at least two round-trip distances from different Tx-Rx pairs. Fig. 1(d) shows an example of the localization process. The two round-trip distances corresponding to the two transmitter-receiver pairs, Tx-Rx1 and Tx-Rx2, define two ellipses. The person has to be at one of the intersection points of these ellipses. However by using directional antennas for transmission and reception we can eliminate the intersection point behind the antennas and localize the person to one point in 2D. This approach extends to 3D, whereby a distance measurement would map to an ellipsoid; hence, we would need three TOF measurements to obtain the 3D location of a person using his reflections.

### 3. WIZ OVERVIEW

WiZ is a wireless system that scales device-free localization to multiple users in both line-of-sight and through-wall scenarios. For static users, WiZ localizes them based on their breathing, and further monitors their breathing rate. WiZ can also localize the hand motions of multiple people, enabling a multi-user gesture-based interface.



**Figure 2—WiZ’s Antennas and Setup.** (a) shows one of WiZ’s directional antenna placed next to a quarter (dimension of each antenna:  $3\text{cm} \times 3.4\text{cm}$ ) (b) shows how these antennas are mounted on a foldable platform (dimensions:  $2\text{m} \times 1\text{m}$ ) and arranged in a single vertical plane.

WiZ is a multi-antenna system. It has five transmit antennas and five receive antennas. These antennas are directional, and each of them is  $3\text{cm} \times 3.4\text{cm}$  as shown in Fig. 2(a); they are all stacked in a single plane and mounted on a foldable platform as shown in Fig. 2(b). This arrangement is chosen because it enables see-through-wall applications, whereby all the antennas need to be lined up in the plane facing the wall of interest.

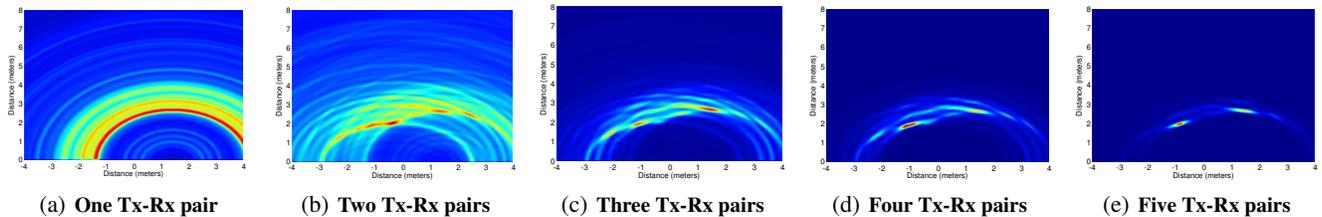
WiZ operates by transmitting RF signals and capturing their reflections after they bounce off different users in the environment. Algorithmically, WiZ has two main components: 1) Multi-shift FMCW, a technique that enables it to deal with interference from multiple users that are modulating the same wireless signal, and 2) Successive Silhouette Cancellation (SSC), an algorithm that allows WiZ to overcome the near-far problem. The following sections describe these components in detail.

## 4. MULTI-SHIFT FMCW

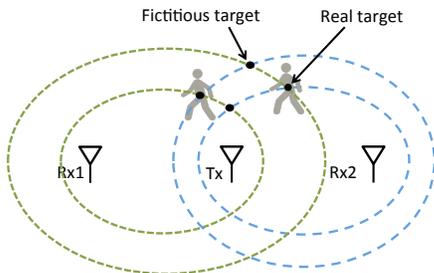
We first explain the basic intuition underlying our localization of multiple people, then introduce the details of multi-shift FMCW.

### 4.1 Challenges and Basic Intuition

We have seen in §2 that TOF measurements from two Tx-Rx pairs allow us to localize a single person in 2D. In this section, we show that to localize a larger number of users, we naturally need TOF measurements from many Tx-Rx antenna pairs.



**Figure 4—Increasing the Number of Tx-Rx pairs enables Localizing Multiple Humans.** The figure shows the heatmaps obtained from combining the TOF profiles of multiple Tx-Rx antenna pairs in the presence of two people in the scene. The x and y axes of each heatmap correspond to the real world x and y dimensions.



**Figure 3—Challenge in Localizing Multiple People.** The figure shows what happens when we have the same setup as Fig. 1(d) but add a second person to the scene. We get two TOF measurements at each Rx, which results in four ellipses having four intersection points within the beam of the antennas. Only two of these intersections are real targets, and the other two are fictitious targets.

**Illustrative Example.** Consider the example in Fig. 3, where we want to localize two users, and we have one transmit (Tx) and two receive antennas (Rx1 and Rx2). Recall that this setup allowed us to localize a single person (as discussed in §2). Now say we have two people. In this case, each receiver will obtain a TOF profile that shows two spikes, one spike for each user that corresponds to the value of her TOF with respect to the Tx-Rx pair. Hence, Rx1 will compute two TOF measurements, and map them to two different ellipses whose foci are Tx and Rx1 (green ellipses in Fig. 3). Similarly, Rx2 will compute two TOF measurements and map them to two ellipses whose foci are Tx and Rx2 (blue ellipses in Fig. 3). These ellipses have four intersections within the beam of the directional antennas (i.e., in the top half of the figure). However, only two of these intersections correspond to real targets. The other two are due to the ellipse of one person intersecting with the ellipse of another person, and hence correspond to fictitious targets.

In practice, the problem of fictitious targets is exacerbated by multiple challenges, and gets more complicated as the number of users in the environment increases. The first challenge is multipath. Specifically, the signal reflected off a person may also bounce off other objects in the environment before arriving at the receive antenna. Each of these reflections will result in an additional spike in the TOF profile, and hence an additional ellipse. A second challenge is due to the near-far problem. Namely, a person who is closer to the antennas will have much stronger reflections than someone who is further away; thus, the reflections of the far person may be masked by the multi-path of the closer one. A third

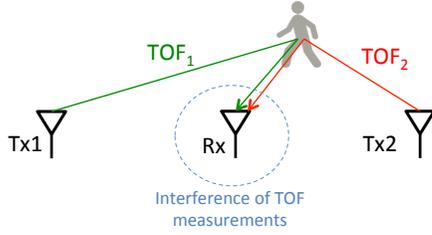
challenge is due to the fact that a person is not a point reflector – his entire body will reflect the transmitted signal. Hence, each ellipse in Fig. 3 will have a fuzzy region about it (i.e., a thickness of  $\pm\Delta d$ , where  $\Delta d$  is determined by the size of the reflecting surface of each person).

**Real-world Experiment.** To explore these challenges in practice, we run an experiment with two users in a  $5\text{ m} \times 7\text{ m}$  room with furniture (tables, chairs, boards, etc.) in a standard office building. We study what happens as we successively overlay the ellipses obtained from different transmit-receive pairs. Recall from §2 that each transmit-receive antenna pair provides us with a *TOF profile* – i.e., it tells us how much reflected power we obtain at each possible TOF between the transmit antenna and receive antenna (see Fig. 1(c)).

Now let us map all TOFs in a TOF profile to the corresponding ellipses. This process produces a heatmap like the one in Fig. 4(a). For each ellipse in the heatmap, the color in the image reflects the amount of received power at the corresponding TOF. Hence, the ellipse in red corresponds to a strong reflector in the environment. The orange, yellow, and green ellipses correspond to weaker reflections respectively; these reflections could either be due to another person in the environment, multi-path reflections of the first person, or noise. The blue regions in the background corresponds to the absence of reflections in the corresponding areas.

Note that the x and y axes for the heatmap image correspond to the x and y dimensions in the real world. Notice how the heatmap shows a pattern of half-ellipses; the foci of these ellipses are the transmit antenna and the receive antenna, both of which are placed along the  $y = 0$  axis. The reason we only show the upper half of the ellipses is that we are using directional antennas and we focus them towards the positive y direction. Hence, we know that we do not receive reflections from behind the antennas.

Fig. 4(a) shows the ellipses corresponding to the TOF profiles from one Tx-Rx pair. Now, let us see what happens when we superimpose the heatmaps obtained from two Tx-Rx pairs. Fig. 4(b) shows the heatmap we obtain when we overlay the ellipses of the first transmit-receive pair with those from a second pair. We can now see two patterns of ellipses in the figure, the first pattern resulting from the TOFs of the first pair, and the second pattern due to the TOFs of the second pair. These ellipses intersect in multiple locations, resulting in red or orange regions, which suggest a higher probability for a reflector to be in those regions. Recall that there



**Figure 5—Interference due to multiple transmit antennas.** The signals from multiple transmit antennas would interfere with each other at the receiver, causing it to obtain two TOF measurements.

are two people in this experiment. However, Fig. 4(b) is not enough to identify the locations of these two people.

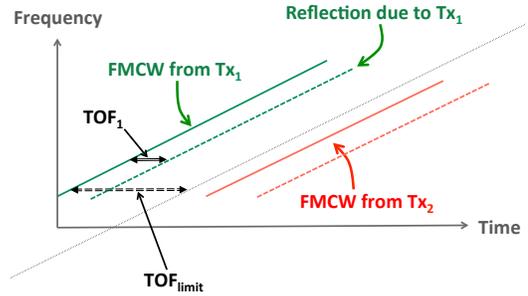
Figs. 4(c) and 4(d) show the result of overlaying the ellipses from three and four transmit-receive pairs respectively. The figures show how the noise and multi-path from different antennas is averaging out to result in a dark blue background. This is because different Tx-Rx pairs have different perspectives of the indoor environment; hence, they do not observe the same noise or multi-path reflections. As a result, the more we overlay heatmaps from different Tx-Rx pairs, the clearer the candidate locations for the two people in the environment.

Next, we overlay the ellipses from five transmit-receive pairs and show the resulting heatmap in Fig. 4(e). We can now clearly see two bright spots in the heatmap: one is red and the other is orange, whereas the rest of the heatmap is mostly a navy blue background indicating the absence of reflectors. Hence, in this experiment, we are able to localize the two users using TOF measurements from five transmit-receive pairs. Combining these measurements together allowed us to eliminate fictitious intersections and localize the two people passively using their reflections.

**Summary:** As the number of users increases, we need TOF measurements from a larger number of Tx-Rx pairs to localize them. For the case of two users, we have seen a scenario whereby the TOFs of five transmit-receive pairs were sufficient to accurately localize both of them. In general, the exact number would depend on multi-path and noise in the environment as well as on the number of users we wish to localize. These observations motivate a mechanism that can provide us with a large number of Tx-Rx pairs while scaling with the number of users in the environment.

## 4.2 The Design of Multi-shift FMCW

In the previous section, we showed that we can localize two people by overlaying many heatmaps obtained from mapping the TOF profiles of multiple Tx-Rx pairs to the corresponding ellipses. But how do we obtain TOFs from many Tx-Rx pairs? One option is to use one FMCW transmitter and a large number of receivers. In this case, to obtain  $N$  Tx-Rx pairs, we would need one transmitter and  $N$  receivers. The problem with this approach is that it needs a large number of receivers, and hence does not scale well as we add more users to the environment.



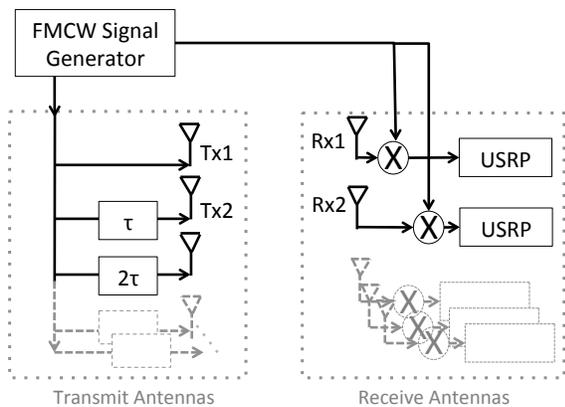
**Figure 6—Multi-shift FMCW.** WiZ transmits FMCW signals from different transmit antennas after inserting virtual delays between them. This delay must be larger than the highest time-of-flight ( $TOF_{limit}$ ) due to objects in the environment.

A more appealing option is to use multiple FMCW transmit and receive antennas. Since the signal transmitted from each transmit antenna is received by all receive antennas, this allows us to obtain  $N$  Tx-Rx pairs using only  $\sqrt{N}$  transmit antennas and  $\sqrt{N}$  receive antennas.

However, the problem with this approach is that the signals from the different FMCW transmitters will interfere with each other over the wireless medium, and this interference will lead to localization errors. To see why this is true, consider the simple example in Fig. 5, where we want to localize a user, and we have two transmit antennas, Tx1 and Tx2, and one receive antenna Rx. The receive antenna will receive two reflections – one due to the signal transmitted from Tx1, and another due to the signal transmitted from Tx2. Hence, its TOF profile will contain two spikes referring to two time-of-flight measurements  $TOF_1$  and  $TOF_2$ .

With two TOFs, we should be able to localize a single user based on the intersection of the resulting ellipses. However, the receiver has no idea which TOF corresponds to the reflection of the FMCW signal generated from Tx1 and which corresponds to the reflection of the FMCW signal generated by Tx2. Not knowing the correct Tx means that we do not know the foci of the two ellipses and hence cannot localize. For example, if we incorrectly associate  $TOF_1$  with Tx2 and  $TOF_2$  with Tx1, we will generate a wrong set of ellipses, and localize the person to an incorrect location. Further, this problem becomes more complicated as we add more transmit antennas to the system. Therefore, to localize the user, WiZ needs a mechanism to associate these TOF measurements with their corresponding transmit antennas.

We address this challenge by leveraging the structure of the FMCW signal. Recall that FMCW consists of a continuous linear frequency sweep as shown by the green line in Fig. 6. When the FMCW signal hits a body it reflects back with a delay that corresponds to the body's TOF. Now let us say  $TOF_{limit}$  is the maximum TOF that we expect in the typical indoor environment where WiZ operates. We can delay the FMCW signal from the second transmitter by  $\tau > TOF_{limit}$  so that all TOFs from the second transmitter are shifted by  $\tau$  with respect to those from the first transmitter, as shown by the red line in Fig. 6. Thus, we can prevent the various FMCW signals from interfering by ensuring that



**Figure 7—Multi-shift FMCW Architecture.** The FMCW generated FMCW signal is fed to multiple transmit antennas via different delay lines. At the receive side, the TOF measurements from the different antennas are combined to obtain the 2D heatmaps.

each transmitted FMCW signal is time shifted with respect to the others, and those shifts are significantly larger than the time-of-flight to objects in the environment. We refer to this design as Multi-shift FMCW.

As a result, the receiver would still compute two TOF measurements: the first measurement (from Tx1) would be  $TOF_1$ , and the second measurement (from Tx2) would be  $TOF'_2 = TOF_2 + \tau$ . Knowing that the TOF measurements from Tx2 will always be larger than  $\tau$ , WiZ determines that  $TOF_1$  is due to the signal transmitted by Tx1, and  $TOF'_2$  is due to the signal transmitted by Tx2.

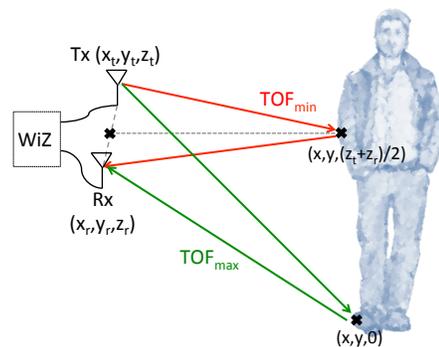
This idea can be further extended to more than two transmit antennas as shown in Fig. 7. Specifically, we can transmit the FMCW signal directly over the air from Tx1, then shift it by  $\tau$  and transmit it from Tx2, then shift it by  $2\tau$  and transmit it from Tx3, and so on. At the receive side, all TOFs between 0 and  $\tau$  are always mapped to Tx1, whereas distances between  $\tau$  and  $2\tau$  are mapped to Tx2, and so on.<sup>1</sup>

**Summary:** Our Multi-shift FMCW technique has two components: the first component allows us to obtain TOF measurements from a large number of Tx-Rx pairs; the second component operates on the TOFs obtained from these different Tx-Rx pairs by superimposing them into a 2D heatmap, which allows us to localize multiple users in the scene.

## 5. SUCCESSIVE SILHOUETTE CANCELLATION

With multi-shift FMCW, we can obtain TOF profiles from a large number of Tx-Rx pairs, map them into 2D heatmaps, overlay the heatmaps, and start identifying the locations of the users. However, in practice this is not sufficient because different users will exhibit the near-far problem. Specifically, the reflections of a nearby person are much stronger than

<sup>1</sup>Note that there is a fundamental difference between this approach and Time-Division Multiplexing (TDM). Specifically, in TDM, transmissions from different antennas are multiplexed in time – i.e., at any point in time, only one antenna is transmitting the FMCW signal. In contrast, in Multi-shift FMCW, all the shifted FMCW signals are transmitted continuously and concurrently by all the transmit antennas of the system.



**Figure 9—Finding  $TOF_{min}$  and  $TOF_{max}$ .**  $TOF_{min}$  is determined by the round-trip distance from the Tx-Rx pair to the closest point on the person's body – i.e., the projection of the midpoint of [Tx,Rx] on the person's body. Since the antennas are elevated,  $TOF_{max}$  is typically due to the round-trip distance to the person's feet.

the reflections of a faraway person or a person behind an obstruction.

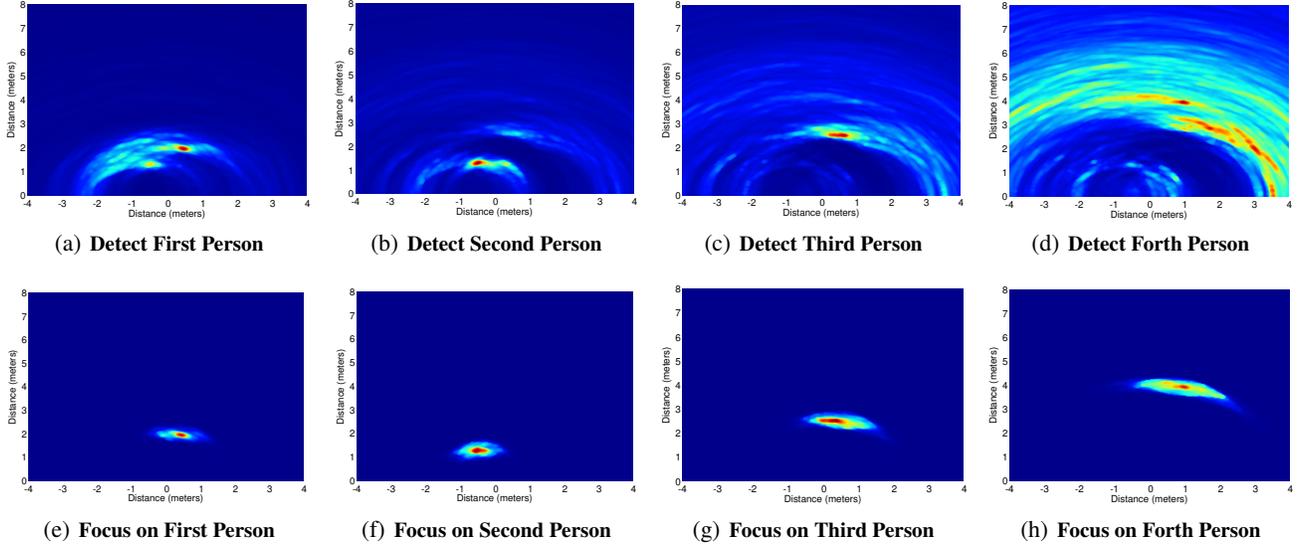
Fig. 8(a) illustrates this challenge. It shows the 2D heatmap obtained in the presence of four persons in the environment. The heatmap allows us to localize only two of these persons: one is clearly visible at (0.5, 2), and another is fairly visible at (-0.5, 1.3). The other two people, who happen to be further away from WiZ, are completely overwhelmed by the power of the first two persons.

To deal with this near-far problem, rather than localizing all the people in one shot, WiZ performs Successive Silhouette Cancellation (SSC). SSC is inspired by Successive Interference Cancellation whereby the receiver decodes the signal with the highest SNR, then re-encodes it and subtracts it out from the received signal, and proceeds to decode the signal with the second-highest SNR, then repeats the same procedure until it has decoded all interferers. The main difference is that decoding in our context means localizing the person using her TOF. More specifically, WiZ's SSC algorithm consists of four main steps:

1. *SSC Detection:* which involves finding the location of the strongest user (reflector) by overlaying the heatmaps of all Tx-Rx pairs.
2. *SSC Re-mapping:* which involves mapping a person's location to the set of TOFs that would have generated that location at each transmit-receive pair.
3. *SSC Cancellation:* which involves canceling the impact of the person on the TOF profiles of all TX-Rx pairs.
4. *Iteration:* whereby we use the obtained TOF profiles after cancellation to re-compute the heatmaps, overlay them, and proceed to find the location of the next strongest reflector.

In what follows, we describe each of these four steps in detail by walking through the example with four persons shown in Fig. 8.

**SSC Detection.** In the first step, SSC finds the location of the highest power reflector in the 2D heatmap of Fig. 8(a). In this example, the highest power is at (0.5, 2), indicating that there is a person in that location.



**Figure 8—Successive Silhouette Cancellation.** (a) shows the 2D heatmap obtained by combining all the TOFs in the presence of four users. (b)-(d) show the heatmaps obtained after cancelling out the first, second, and third user respectively. (e)-(h) show the result of the SSC focusing step on each of the person, and how it enables us to accurately localize each person while eliminating interference from all other users.

**SSC Re-mapping.** Given the  $(x, y)$  coordinates of the person, we map his location back to the corresponding TOF at each transmit-receive pair. Keep in mind that each person is not a point reflector; hence, we need to estimate the effect of reflections off his entire body on the TOF profile of each transmit-receive pair.

To see how we can do this, let us look at the illustration in Fig. 9 and try to see the effect of a person’s body on one transmit-receive pair. The signal transmitted from the transmit antenna will reflect off different points on the person’s body before arriving at the receive antenna. Thus, the person’s reflections will appear between some  $TOF_{min}$  and  $TOF_{max}$  in the TOF profile at the receive antenna. (In fact, this can be clearly seen in Fig. 1(c) where a person’s reflections span a contiguous band of TOF measurements.)

Note that  $TOF_{min}$  and  $TOF_{max}$  are determined by the closest and furthest points respectively on a person’s body from the transmit-receive antenna pair. Let us first focus on how we can obtain  $TOF_{min}$ . By definition, the closest point on the person’s body is the one that corresponds to the shortest round-trip distance to the Tx-Rx pair. Hence, it is the projection of the midpoint of the segment  $[Tx, Rx]$  on the body of the person as shown in Fig. 9. We already know the  $x$  and  $y$  of that projection point because they are the 2D location of the person. Further, the  $z$  coordinate of that point is midway between Tx and Rx. Formally, for a Tx antenna at  $(x_t, y_t, z_t)$  and an Rx antenna at  $(x_r, y_r, z_r)$ , the  $z$  of the point on the person’s body that is closest to the Tx-Rx pair can be expressed as  $(z_t + z_r)/2$ . Hence, since the round-trip distance is the summation of the forward path from Tx to that point and the path from that point back to Rx, we may express it as:

$$d_{min} = \sqrt{(x_t - x)^2 + (y_t - y)^2 + ((z_t - z_r)/2)^2} + \sqrt{(x_r - x)^2 + (y_r - y)^2 + ((z_r - z_t)/2)^2}.$$

Similarly,  $TOF_{max}$  is determined by the round-trip distance to point on the person’s body that is furthest from the Tx-Rx pair. Again, the  $x$  and  $y$  coordinates of the furthest point are determined by the person’s location from the SSC Detection step. However, we still need to figure out the  $z$  coordinate of this point. Since the transmitter and receiver are both raised above the ground (at around 1.2 meters above the ground), the furthest point from the Tx-Rx pair is typically at the person’s feet.<sup>2</sup> Therefore, we know that the coordinates of this point are  $(x, y, 0)$ , and hence we can compute  $d_{max}$  as:

$$d_{max} = \sqrt{(x_t - x)^2 + (y_t - y)^2 + z_t^2} + \sqrt{(x_r - x)^2 + (y_r - y)^2 + z_r^2}.$$

Finally, we can map  $d_{min}$  and  $d_{max}$  to  $TOF_{min}$  and  $TOF_{max}$  by dividing them by the speed of light  $C$ .

**SSC Cancellation.** The next step in the SSC algorithm is to use  $TOF_{min}$  and  $TOF_{max}$  to cancel the person’s reflections from the TOF profiles of each transmit-receive pair. Unlike successive interference cancellation, where the receiver can fully re-encode the transmitted signal before subtracting it out, the effect of a person’s reflections on the TOF profile of each transmit-receive pair cannot be perfectly estimated. This is because the reflected power of the human body depends on many factors like the size of the person, the clothes she is wearing, and her exact posture while walking.

Hence, to remove a person’s reflections from a particular TOF profile, we take a conservative approach and zero out the power in all TOFs between  $TOF_{min}$  and  $TOF_{max}$  within that profile. Of course, this means that we might also be partially cancelling out the reflections of another person who happens to have a similar time of flight to this Tx-Rx pair.

<sup>2</sup>Note that generally we compute both the round-trip to the person’s feet and to the head of an average-height person (5’9”) and choose the max of the two as  $d_{max}$ .

However we rely on that multi-shift FMCW provides with a large number of TOF profiles from many Tx-RX pairs. Hence even if we cancel out the power in the TOF of a person with respect to a particular Tx-Rx pair, each person will continue to have a sufficient number of TOFs measurements from the rest of the antennas.

We repeat the process of computing  $TOF_{min}$  and  $TOF_{max}$  with respect of each Tx-Rx pair and zero-ing out the power in that range, until we have completely eliminated any power from the recently decoded person.

**Iteration.** We proceed to decode the next person. This is done by regenerating the heatmaps from the updated TOF profiles and overlaying them. Fig. 8(b) shows the obtained image after performing this procedure for the first person. Now, a person at  $(-0.5, 1.3)$  becomes the strongest reflector in the scene.

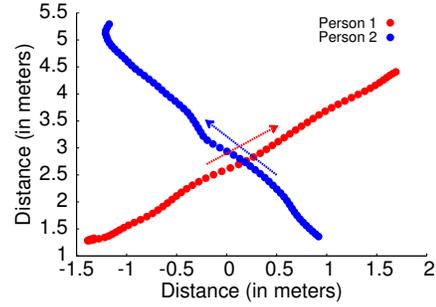
Subsequently, we repeat the same procedure for this person, cancelling out his interference, then reconstructing the 2D heatmap in Fig. 8(c) using the remaining TOF measurements. Now, the person with the strongest reflection is at  $(0.8, 2.7)$ . Note that this heatmap is much noisier than Figs. 8(a) and 8(b) because now we are dealing with a more distant person.

WiZ repeats the same cancellation procedure for the third person and constructs the 2D heatmap in Fig. 8(d). The figure shows a strong reflection at  $(1, 4)$ . Recall that our antennas are placed along the  $y = 0$  axis, which means that this is indeed the furthest person in the scene. Also note that the heatmap is now even noisier. This is expected because the furthest person’s reflections are much weaker. We note that each of these heatmaps are scaled so that the highest power is always in red and the lowest power is in navy blue; this change in scale emphasizes the location of the strongest reflectors and allows us to better visualize their locations.

WiZ repeats the interference cancellation for the fourth person, and determines that the SNR of the maximum reflector in the resulting heatmap does not pass a threshold test. Hence, it determines that there are only four people in the scene.

We perform four additional steps to improve WiZ’s SSC algorithm:

- *Refocusing Step:* After obtaining the initial estimates of the locations of all four persons, WiZ performs a focusing step for each user to refine his location estimate. This is done by reconstructing an interference-free 2D heatmap only using the range in the TOF profiles that corresponds to TOFs between  $TOF_{min}$  and  $TOF_{max}$  for that Tx-Rx pair. Figs. 8(e)- 8(h) show the images obtained from this focusing step. In these images, the location of each person is much clearer, which enables higher-accuracy localization.
- *Leveraging Motion Continuity:* After obtaining the estimates from the SSC algorithm, WiZ applies a Kalman filter and performs outlier rejection to reject impractical jumps in location estimates that would otherwise correspond to abnormal human motion over a very short period of time.



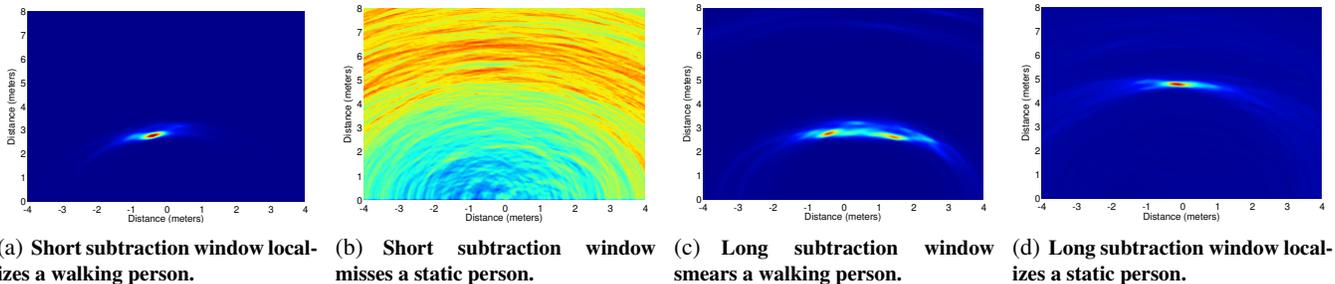
**Figure 10—Disentangling Crossing Paths.** When two people cross paths, they typically keep going along the same direction they were going before their paths crossed.

- *Disentangling Crossing Paths:* To disentangle multiple people who cross paths, we look at their direction of motion before they crossed paths and project how they would proceed with the same speed and direction as they are crossing paths. This helps us with associating each person with his own trajectory after crossing. Fig. 10 shows an example with two people crossing paths and how we were able to track their trajectories despite that. Of course, this approach does not generalize to every single case, which may lead to some association errors after the crossings but not to localization errors.
- *Extending SSC to 3D Gesture Recognition:* Similar to past work [4], WiZ can differentiate a hand motion from a whole-body motion (like walking) by leveraging the fact that a person’s hand has a much smaller reflective surface than his entire body. Unlike past work, however, WiZ can track gestures even when they are simultaneously performed by multiple users. Specifically, by exploiting WiZ’s SSC focusing step, it can focus on each person individually and track his/her gestures. In our evaluation, we focus on testing a pointing gesture, where different users point in different directions at the same time. Subsequently, by tracking the trajectory of each moving hand, we can determine the direction in which each of the users is pointing. Note that we perform these pointing gestures in 3D and track the hand motion by using the TOFs from the different Tx-Rx pairs to construct a 3D point cloud rather than a 2D heatmap.<sup>3</sup> The results in §8.3 show that we can accurately track hand gestures performed by multiple users in 3D space.

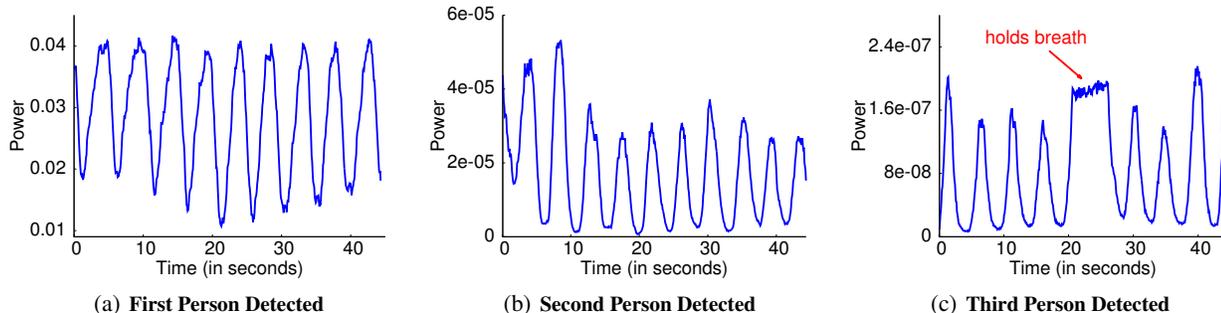
## 6. LOCALIZATION BASED ON BREATHING

We extend WiZ’s SSC algorithm to localize static people based on their breathing. Recall from §2 that in order to track a user based on her radio reflections, we need to eliminate reflections off all static objects in the environment (like walls and furniture). This is typically achieved by performing a background subtraction step, i.e., by taking TOF

<sup>3</sup>Recall from §2 that a given TOF maps to an ellipse in 2D and an ellipsoid in 3D. The intersection of ellipsoids in 3D allow us to track these pointing gestures.



**Figure 11—Need For Multiple Subtraction Windows.** The 2D heatmaps show that a short subtraction window allows WiZ accurately localize a pacing person in (a) but not a static person in (b). A long subtraction window would smear the walking person’s location in (c) but allows WiZ to localize a breathing person in (d).



**Figure 12—Monitoring the Breathing of Multiple People.** The figures show how the maximum power in the focused spectrogram of each person is varying in time due to his breathing. (a), (b), and (c) correspond to the first, second, and third persons detected by the SSC algorithm.

profiles from adjacent time windows and subtracting them out from each other.<sup>4</sup>

Whereas this approach enables us to track moving people, it prevents us from detecting a static person – e.g., someone who is standing or sitting still. Specifically, because a static person remains in the same location, his TOF does not change, and hence his reflections would appear as static and will be eliminated in the process of background subtraction. To see this in practice, we run two experiments where we perform background subtraction by subtracting two TOF profiles that are 12.5 milliseconds apart from each other. The first experiment is performed with a walking person and the resulting heatmap is shown in Fig. 11(a), whereas the second experiment is performed in the presence of a person who is sitting at (0, 5) and the resulting heatmap is shown in Fig. 11(b). These experiments show how the heatmap of a moving person after background subtraction would allow us to localize him accurately, whereas the heatmap of the static person after background subtraction is very noisy and does not allow us to localize the person.

To localize static people, one needs to realize that even a static person moves slightly due to breathing. Specifically, during the process of breathing, the human chest moves by a sub-centimeter distance over a period of few seconds. The key challenge is that this change does not translate into a discernible change in the TOF of the person. However, over an

interval of time of a few seconds (i.e., as the person inhales and exhales), it would result in discernible changes in the reflected signal. Therefore, by subtracting frames in time that are few seconds apart from each other, we should be able to localize the breathing motion.

In fact, Fig. 11(d) shows that we can accurately localize a person who is sitting still by using a subtraction window of 2.5 seconds. Note, however, that this long subtraction window will introduce errors in localizing a pacing person. In particular, since typical indoor walking speed is around 1 m/s [7], subtracting two frames that are 2.5 seconds apart would result in smearing the person’s location and may also result in mistaking him for two people as shown in Fig. 11(c).

Thus, to accurately localize both static and moving people, WiZ performs background subtraction with different subtraction windows. It then applies multi-shift FMCW and successive silhouette cancellation as before.

**Counting the Number of Breaths:** WiZ’s SSC algorithm enables focusing on each person while eliminating interference from all other users. This algorithm proves critical to monitoring each person’s breath in the presence of other people in the environment.

We run an experiment with three users, whereby we ask them to sit on chairs and remain still for the duration of the experiment. To test WiZ’s ability in monitoring their breathing rates, we subtract the sequence of TOF profiles obtained over time from initial TOF profile at time 0, for each Tx-Rx pair. We then process the obtained signals by performing SSC. Recall that the SSC focusing step allows us to focus on each person while eliminating interference from all other

<sup>4</sup>Recall that we obtain one TOF profile by taking an FFT over the received FMCW signal in baseband. Since the FMCW signal is repeatedly swept, we can compute a new TOF profile from each sweep.

people in the scene (as shown in Figs. 8(e)-8(h)). Hence, it allows us to focus on each person individually, and monitor the max power of each person’s focused heatmap as a function of time. We do that for every person in the environment, and plot in Fig. 12 the max power for each focused heatmap as a function of time.

The figure shows multiple observations:

- The maximum power from each person’s heatmap goes up and down periodically. This is because breathing is a rhythmic motion that alternates between inhaling and exhaling. The maximum power is lowest when the person’s chest returns to its location in the initial subtraction frame, and is highest when it is at the furthest position from its initial location.
- The first person’s peak-to-peak signal is three orders of magnitudes higher than that of the second person and five orders of magnitude higher than the third person. This observation demonstrates the importance of the SSC algorithm in detecting weaker reflections on one hand, and the importance of the SSC focusing step in eliminating interference from all other persons to be able to focus on each person individually and monitor his/her breathing.
- WiZ allows us to detect periods of time when a user holds her breath. For example, the user in Fig. 12(c) holds her breath between  $t = 20$  s and  $t = 27$  s. This capability opens up WiZ to a wide variety of applications in health-care monitoring in hospitals or at home, such as diagnosing sleep apnea and detecting a user’s stress level [17].

## 7. IMPLEMENTATION & EVALUATION

### 7.1 Implementation

We built WiZ using an FMCW radio, USRP N210 software radios [2] with LFRX-LF daughterboards, and directional antennas. The FMCW radio generates a signal that sweeps 5.46-7.25 GHz every 2.5 milliseconds. The schematic in Fig. 7 shows how we use this radio to implement Multi-shift FMCW. Specifically, the generated sweep is fed to different directional antennas via delay lines of different lengths. At the receive side, the signal from each receive antenna is mixed with the FMCW signal and the resulting signal is fed to the USRP. The USRP samples the signals at 1 MHz and feeds the digitized samples to the UHD driver. These samples are processed in software to localize users and recognize their gestures.

WiZ uses custom-made log-periodic antennas, each of size  $3\text{cm} \times 3.4\text{cm}$ , optimized to operate in the desired frequency range. In its default setup, WiZ’s antennas are stacked into a  $2\text{m} \times 1\text{m}$  vertical plane (see Fig. 2(b) for an image of this setup). These antennas transmit very low power (less than 0.75 milliWatt) to comply with FCC regulations for consumer use in the corresponding frequency band.

Finally, we note that the analog FMCW radio and all the USRPs are driven by the same external clock. This ensures that there is no frequency offset between their oscillators, and hence enables subtracting frames that are relatively far apart in time to enable localizing people based on breathing.

### 7.2 Evaluation

**Human Subjects.** We evaluate the performance of WiZ by conducting experiments in our lab with eleven human subjects: four females and seven males. The subjects are of different heights and builds.

**Ground Truth.** We use the VICON motion capture system to provide us with ground truth positioning information. VICON is a multi-hundred thousand dollar system that provides sub-centimeter positioning information and is used in film making and video game development to create 3D animation models of characters. It consists of a array of pre-calibrated infrared cameras that are fitted to the ceiling of a room, and requires instrumenting any tracked object with infrared-reflective markers. When an instrumented object moves, the system tracks the infrared markers on that object and fits them into a 3D model to identify the object’s location at any point in time.

We evaluate WiZ’s accuracy by comparing it to the locations provided by the VICON system. To track a user using the VICON system, we ask him/her to wear a hard hat that is instrumented with five infrared markers. In addition, for the gestures experiments, we ask each user to wear a glove that is instrumented with six VICON markers.

**Experimental Setup.** We evaluate WiZ in two experimental setups: line-of-sight and through-the-wall. In the through-wall experiments, WiZ is placed outside the VICON room with all transmit and receive antennas facing one of the walls of the VICON room. Recall that WiZ’s antennas are directional and hence this setting means that the radio beam is directed toward the VICON room. The VICON room has no windows; it has 6-inch hollow walls supported by steel frames with sheet rock on top, which is a standard setup for office buildings. In the line-of-sight experiments, we move WiZ to inside the VICON-instrumented room. In all of these experiments, the subjects’ locations are tracked by both the VICON system and WiZ.

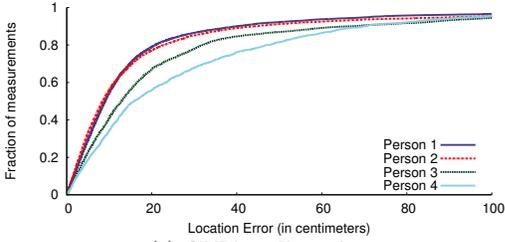
**Calibration.** Localizing a person requires that the system first detects him/her. Therefore, we run experiments to identify the maximum number of people that WiZ can reliably detect under various conditions, and report the numbers in the table below.

	Line-of-Sight	Through-Wall
Motion Tracking	4	3
Breathing-based Localization	5	4

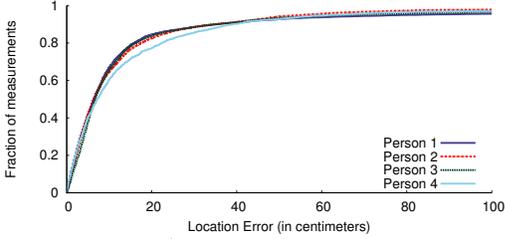
**Table 1—Maximum Number of People Detected Reliably.**

For our evaluation of localization accuracy, we run experiments with the maximum number of people that are reliably detectable, where reliably detected is defined as detected with probability 0.98 or higher.

We make two observations about the above table. First, the maximum number of people detected in line-of-sight is higher than in the through-wall settings. This is expected because the wall causes significant attenuation and hence re-



(a) CDF in x-dimension



(b) CDF y-dimension

**Figure 13—Performance of WiZ’s LOS Tracking.** (a) and (b) show the CDFs of the location error in both x and y dimensions for each of the tracked users in LOS. Subjects are ordered from first to last detected by the SSC algorithm.

duces the SNR of the reflected signals. Second, the maximum number of people detected for breathing-based localization is higher than the number detected in the tracking experiments. While this might seem surprising, it is actually due to the fact that the breathing experiments are run for a longer period of time, where each person stays in his/her same location throughout the experiment; the system outputs the number of people detected and their locations by analyzing the trace resulting from the entire experiment. In contrast, the tracking experiments require outputting a location of each person once every 12.5 ms, and hence they might not be able to detect each person within such a small time window.

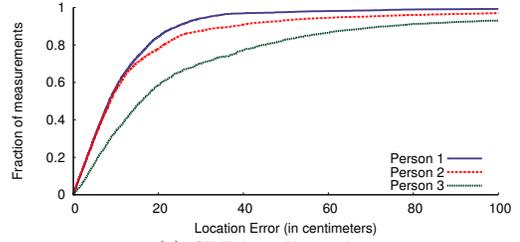
## 8. PERFORMANCE RESULTS

### 8.1 Accuracy of Multi-Person Motion Tracking

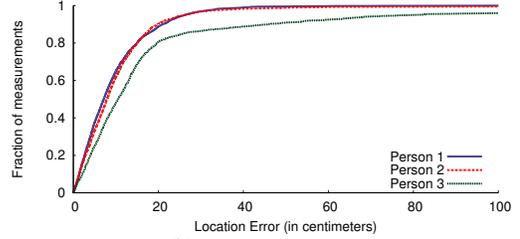
We first evaluate WiZ’s accuracy in multi-person motion tracking. We run 100 experiments in total, half of them in line-of-sight and the second half in through-wall settings. In each experiment, we ask one, two, three, or four human subjects to wear the hard hats that are instrumented with VICON markers and move inside the VICON-instrumented room. Each subject’s location is tracked by both the VICON system and WiZ.

Each experiment lasts for one minute. Since each FMCW sweep lasts for 2.5ms and we average 5 sweeps to obtain each TOF measurement, we collect more than 400,000 location readings for *each* person from these experiments.

Fig. 13 and 14 plot the CDFs of the location error along the x and y coordinates for each of the localized persons in both line-of-sight and through-wall scenarios. The subjects



(a) CDF in x-dimension



(b) CDF y-dimension

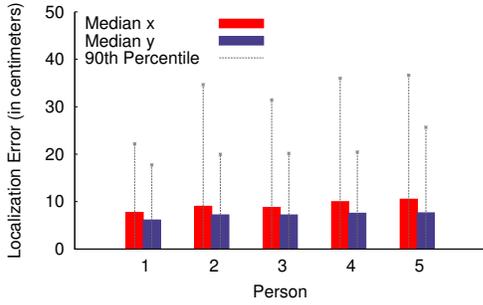
**Figure 14—Performance of WiZ’s through-wall Tracking.** (a) and (b) show the CDFs of the location error in both x and y dimensions for each of the tracked users in LOS. Subjects are ordered from first to last detected by the SSC algorithm.

are ordered from the first to the last as detected by the SSC algorithm. The figures reveal the following findings:

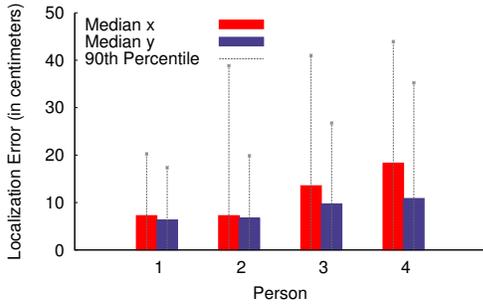
- WiZ can accurately track the motion of four users when it is placed in the same room as the subjects. Its median location error for these experiments is around 8.5 cm in x and 6.4 cm in y for the first user detected, and decreases to 15.9 cm in x and 7.2 cm in y for the last detected user.
- In through-wall scenarios, WiZ can accurately localize up to three users. Its median location error for these experiments is 8.4 cm and 7.1 cm in x/y for the first user detected, and decreases to 16.1 cm and 10.5 cm in x/y for the last detected user. As expected, the location accuracy when the device is placed in the same room as the users is better than when it is placed behind the wall due to the extra attenuation and the reduced SNR caused by the wall.
- The accuracy in the y dimension is better than the accuracy in the x dimension. This discrepancy is due to WiZ’s setup. Recall that WiZ’s antennas are all arranged along the  $y = 0$  axis. Hence, the major axis of the resulting ellipses is always along the x-axis, which means that the same TOF error would have a larger impact on the x dimension than on the y dimension.
- The localization accuracy decreases according to the order the SSC algorithm localizes the users. By investigating these results more, we realize that the fourth person is typically the subject who is the furthest from the center of our device. Hence, his SNR would be lowest, which explains his/her higher localization error.

### 8.2 Accuracy of Breathing-based Localization and Breath Counting

We evaluate WiZ’s accuracy in localizing static people based on their breathing and its ability to count their breaths.



**Figure 15—Accuracy for Localizing Breathing People in line-of-sight.** The figure shows show the median and 90<sup>th</sup> percentile errors in x/y location. Subjects are ordered from first to last detected by the SSC algorithm.



**Figure 16—Accuracy for Localizing Breathing People in through-wall experiments.** The figure shows show the median and 90<sup>th</sup> percentile errors in x/y location. Subjects are ordered from first to last detected by the SSC algorithm.

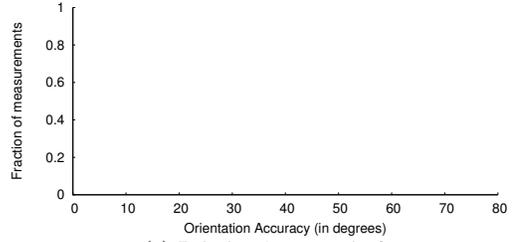
We run 100 experiments in total with up to five people in the room. Half of these experiments are done in line-of-sight and the other half are through-wall. Experiments lasts for 3-4 minutes. All subjects wear hardhats and sit on chairs in the VICON room.

Fig. 15 and 16 plot WiZ’s localization error in line-of-sight and through-wall settings as a function of the order with which the subject is detected by the SSC algorithm. The figures show the median and 90<sup>th</sup> percentile of the estimation error for the x and y coordinates of each of the subjects.

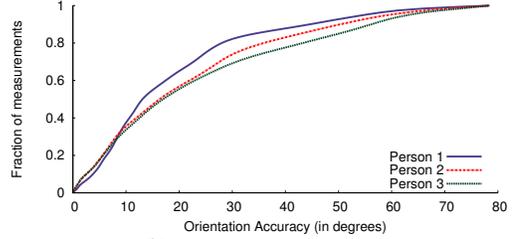
The figures show the following results:

- WiZ’s breathing-based localization accuracy goes from a median of 7.24 cm and 6.3 cm in x/y for the nearest person to 18.31 cm to 10.85 cm in x/y for the furthest person, in both line-of-sight and through-wall settings
- Localization based on breathing exhibits higher accuracy than motion tracking. This is because when people are static, we obtain a larger number of measurements for each location, which allows us to localize them more accurately.

**Breath counting results:** Besides localizing people based on their breathing, WiZ can also count their breaths. Recall from §6 that after localizing subjects based on their breathing, we can use the SSC focusing step to focus on each user and monitor his breathing rate. Specifically, Fig. 12 shows how the maximum power in the focused 2D heatmap varies periodically in time due to each person’s rhythmic breathing. By taking the Fourier transform over this time plot and



(a) Pointing Accuracy in  $\theta$



(b) Pointing Accuracy in  $\phi$

**Figure 17—3D Gesture Accuracy.** The figure shows the CDFs of the orientation accuracy for the pointing gestures of each participant. Subjects are ordered from first to last detected by the SSC algorithm.

choosing the frequency with the highest power, WiZ can determine each person’s breathing rate; then, it can map it to the number of breaths taken by that person by multiplying that rate by the duration of the experiment.

To obtain a ground truth for the number of breaths taken by each subject, we asked the users to start counting their breaths when the experiment starts, and report the number of breaths they have taken once the experiment is over.<sup>5</sup> Across all of these experiments, WiZ’s error in counting the number of breaths remains within one breath for over 97% of our experiments – each of which lasts for 3-4 minutes; note that this error is within the rounding error of the integer count as reported by each user. In 2% of these experiments, the users were not detected, and hence WiZ was unable to count their breathing rate.

In addition, WiZ was able to detect instances where users held their breath (e.g., experiment in Fig. 12(c)). Upon detection, we confirmed with the subjects that they indeed held their breaths. These result indicates that WiZ could be used in health-care monitoring applications such as diagnosing sleep apnea.

### 8.3 Accuracy of 3D Pointing Gesture Detection

We evaluate WiZ’s accuracy in tracking 3D pointing gestures. We run 100 experiments in total with one to three subjects. In each of these experiments, we ask each subject to wear a glove that is instrumented with infrared-reflexive markers, stand in a different location in the VICON room, and point his/her hand in a random 3D direction of their choice – as if they were playing a shooting game or pointing at some household appliance to control it. In most of

<sup>5</sup>Obtaining the breath count with other methods is difficult since accurate breath monitoring equipment is expensive [1].

these experiments, all subjects were performing the pointing gestures simultaneously.

Throughout these experiments, we measure the 3D location of the hand using the VICON system and WiZ. We then use the 3D trajectory to determine the direction in which each user pointed. Fig. 17(a) and 17(b) plot the CDFs of the orientation error between the angles as measured by WiZ and the VICON for the 1st, 2nd and 3rd participant (in the order of detection by SSC). Note that we decompose the 3D pointing direction into two angles:  $\theta$  and  $\phi$ , where the former is the projection of the pointing direction on the  $x - y$  plane and the latter is the pointing direction in the  $r - z$  plane (i.e., azimuth angle of the spherical coordinate system). The figure shows that the median orientation error in  $\theta$  goes from 8.2 degrees to 12.4 degrees from the first to the third person, and from 12 degrees to 16 degrees in  $\phi$ . Note that WiZ’s accuracy in  $\theta$  is slightly higher than its accuracy in  $\phi$ . This is due to WiZ’s setup, where the antennas are more spread out along the  $x$  than along the  $z$ , naturally leading to lower robustness to errors along the  $z$  axis, and hence lower accuracy in  $\phi$ . These experiments demonstrate that WiZ can achieve high accuracy in 3D tracking of body parts and hence enables a rich multi-user gesture-based interface using wireless signal reflections.

## 9. RELATED WORK

WiZ builds on prior foundational work in multiple areas, but differs from all past work both in the developed technologies and the resulting capabilities. It introduces two new techniques – multi-shift FMCW and the successive silhouette cancellation – and applies them to achieve highly accurate tracking of multiple people and simultaneous gestures in indoor settings, based purely on how those motions modulate the RF signal.

**Through-Wall Motion Tracking and RF-Based Gesture Interfaces.** The past year has seen the rise of wireless systems that deliver through-wall motion tracking and gesture-based interfaces [5, 14, 4]. Our work builds on these past systems, but differs in the developed techniques and capabilities. Specifically, WiVi [5] and WiSee [14] rely on WiFi Doppler effects to detect motion and identify a handful of gestures after they perform prior training; but, unlike WiZ, they have no mechanism for obtaining the location of a person, whether she is moving or static. On the other hand, WiTrack [4] uses time-of-flight measurements to obtain the location of a single moving person, but cannot localize multiple or static humans. Similar to WiTrack, WiZ also relies on time-of-flight estimates; however, WiZ’s successive silhouette cancellation and Multi-shift FMCW techniques scale device-free localization and RF-based gesture interfaces beyond a single person. Further, in contrast to all these systems, WiZ can localize static humans based on their breathing and even count their breathing rate.

**RSSI-based Radio Tomography.** Some past work on radio tomography [21, 12] can localize a person even if she holds no RF device. These proposals employ a network of dozens

to hundreds of sensors, deployed throughout the area of interest. The received signal strength (called RSS or RSSI) is measured between the resulting  $n^2$  links, and a variation in the RSSI measurements on a link is attributed to a human crossing that link. Another body of work performs device-free localization by leveraging RSSI fingerprinting [18, 16, 23]. Specifically, these works perform an initial calibration phase where they require a person to stand in all different locations throughout the area of interest, and create a radio map using these measurements. In the testing phase, they identify the location of person by mapping the RSSI measurements to those computed during the offline phase.

WiZ shares the vision of these techniques in performing device-free localization. WiZ however does not use RSSI; it introduces new techniques based on the time of flight; as a result, its accuracy is 10x to 100x higher than state-of-the-art RSSI-based systems [18, 16, 12, 6, 13]. Further, WiZ does not require an initial calibration phase where an estimate of the environment is obtained in the absence of people [16, 6]. **See-Through-Wall Radar.** Seeing through walls is an active area of research for the military [15, 11, 9, 19, 22]. WiZ builds on this body of work but differs from it along three lines: First, in comparison to these proposals, which have access to military spectrum, WiZ limits itself to operating within FCC regulations concerning spectrum usage for consumer electronics, and transmits less than one milliwatt of power. Second, WiZ introduces two technical innovations over all prior art: Multi-shift FMCW and the SSC algorithm. Finally, WiZ is not limited to full-body motion; it can track hand motion delivering the first multi-user gesture interface using RF reflections.

**FMCW Techniques.** The literature has many variations on the basic FMCW technique; hence, it is important to note that these past variations all differ significantly from WiZ’s Multi-shift FMCW technique. We particularly highlight the difference between our work and three past systems. MIMO FMCW [10, 15] is based on switched antenna arrays – i.e., at any point in time, it transmits the FMCW signal from one Tx antenna and receives it at one Rx antenna, then alternates between its antennas in a round-robin fashion. As a result, the TOFs computed from the different antennas correspond to different points in time, which results in smearing the moving person’s location and reducing the localization accuracy.

Multi-source FMCW [20] is a new technique in optical imaging that emulates a large sweep by using multiple smaller sweeps which operate in different frequency bands and are all transmitted from the same laser source. This technique is orthogonal to Multi-shift FMCW where shifted sweeps in the *same* frequency band are transmitted simultaneously from *different* antennas.

Finally, multiplexed FMCW [8] is another optical imaging technique that enables focusing on different planes in space by delaying the *received* signal by different amount of time. In WiZ our objective is not to focus on different plane but rather to obtain a large number of Tx-Rx measurements without confusing the signal from various trans-

mitters. Hence, Multi-shift FMCW delays the signals on the *transmit* side before sending them on different transmitters, which enables each receiver to distinguish between the transmitted signals from different antennas.

## 10. DISCUSSION & LIMITATIONS

WiZ marks an important contribution by enabling centimeter-scale device-free multi-person tracking. WiZ, however, has some limitations that are left for future work.

**Antenna spacing:** WiZ's current prototype distributes its antennas in a fairly large vertical plane that measures  $2 \times 1$  m. The large spacing between its antennas is important to enable the antennas to capture different perspectives of the people in the scene, which reduces interference and increases diversity. Future research may explore both hardware and algorithmic advances that can increase the resolution of the system allowing for the antennas to be stacked within a smaller area.

**Number of tracked people:** The current version of WiZ can accurately track the motion of up to four users. It can also localize up to five people based on their breathing. We believe these capabilities open up a large number of applications in multi-player gaming and gesture-based interfaces. However, it is always desirable to increase the number of people that the device can track.

**Person and body part identification:** In its current version WiZ can track the motion of body parts, e.g., a hand, but cannot differentiate between different body parts (a hand vs. a leg). We believe that future work can investigate this issue further by identifying fingerprints of different reflectors that can provide insight about the type of the moving object.

Although there are many issues that future work can build upon, WiZ pushes the limits of RF motion tracking by enabling centimeter-scale multi-person tracking. It also enriches the roles that wireless networks can play in our daily lives and bridges wireless communication with human-computer interaction.

## 11. REFERENCES

- [1] Maxtec Exhalometer Spirometer. <http://www.mspsc.com>. Maxtec.
- [2] USRP N210. <http://www.ettus.com>. Ettus Inc.
- [3] VICON T-Series. <http://www.vicon.com>. VICON.
- [4] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller. 3D Tracking via Body Radio Reflections. In *Usenix NSDI*, 2014.
- [5] F. Adib and D. Katabi. See through walls with Wi-Fi! In *ACM SIGCOMM*, 2013.
- [6] M. Bocca, O. Kaltiokallio, N. Patwari, and S. Venkatasubramanian. Multiple target tracking with rf sensor networks. *Mobile Computing, IEEE Transactions on*, 2013.
- [7] R. Bohannon. Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. *Age and ageing*, 1997.
- [8] P. K. Chan, W. Jin, J. Gong, and N. Demokan. Multiplexing of fiber bragg grating sensors using a fmcw technique. *IEEE Photonics Technology Letters*, 1999.
- [9] S. Hantscher, A. Reizenhahn, and C. Diskus. Through-wall imaging with a 3-d uwb sar algorithm. *Signal Processing Letters, IEEE*, 2008.
- [10] Y. Huang, P. V. Brennan, D. Patrick, I. Weller, P. Roberts, and K. Hughes. Fmcw based mimo imaging radar for maritime navigation. *Progress In Electromagnetics Research*, 2011.
- [11] Y. Jia, L. Kong, X. Yang, and K. Wang. Through-wall-radar localization for stationary human based on life-sign detection. In *IEEE RADAR*, 2013.
- [12] S. Nannuru, Y. Li, Y. Zeng, M. Coates, and B. Yang. Radio-frequency tomography for passive indoor multitarget tracking. *Mobile Computing, IEEE Transactions on*, 2013.
- [13] N. Patwari, L. Brewer, Q. Tate, O. Kaltiokallio, and M. Bocca. Breathfinding: A wireless network that monitors and locates breathing in a home. *Selected Topics in Signal Processing, IEEE Journal of*, 2014.
- [14] Q. Pu, S. Jiang, S. Gollakota, and S. Patel. Whole-home gesture recognition using wireless signals. In *ACM MobiCom*, 2013.
- [15] T. Ralston, G. Charvat, and J. Peabody. Real-time through-wall imaging using an ultrawideband multiple-input multiple-output (MIMO) phased array radar system. In *IEEE ARRAY*, 2010.
- [16] A. Saeed, A. Kosba, and M. Youssef. Ichnaea: A low-overhead robust wlan device-free passive localization system. *Selected Topics in Signal Processing, IEEE Journal of*, 2014.
- [17] L. Science. Stressed? It May Show in Your Breath. <http://www.livescience.com/27991-breath-analysis-stress-level.html>.
- [18] M. Seifeldin, A. Saeed, A. Kosba, A. El-Keyi, and M. Youssef. Nuzzer: A large-scale device-free passive localization system for wireless environments. *Mobile Computing, IEEE Transactions on*, 2013.
- [19] G. E. Smith and B. G. Mobasser. Robust through-the-wall radar image classification using a target-model alignment procedure. *Image Processing, IEEE Transactions on*, 2012.
- [20] A. Vasilyev. *The optoelectronic swept-frequency laser and its applications in ranging, three-dimensional imaging, and coherent beam combining of chirped-seed amplifiers*. PhD thesis, 2013.
- [21] J. Wilson and N. Patwari. Radio tomographic imaging with wireless networks. In *IEEE Transactions on Mobile Computing*, 2010.
- [22] Y. Xu, S. Wu, C. Chen, J. Chen, and G. Fang. A novel method for automatic detection of trapped victims by ultrawideband radar. *Geoscience and Remote Sensing, IEEE Transactions on*, 2012.
- [23] M. Youssef, M. Mah, and A. Agrawala. Challenges: device-free passive localization for wireless environments. In *ACM MobiCom*, 2007.

