Through-Wall Radio Imaging

Travis Mitchell
Division of Science and Mathematics
University of Minnesota, Morris
Morris, Minnesota, USA 56267
mitch809@morris.umn.edu

ABSTRACT

Radio technology has recently made progress that allows for radio imaging to take place through walls. This through-wall imaging can be separated into one of two categories: two-dimensional and three-dimensional. Two-dimensional imaging treats a person as a single data point, caring only about where objects are, whilst three-dimensional imaging focuses on rendering said objects. In this paper, the focus will be on three-dimensional imaging, but two-dimensional imaging must first be discussed to establish the groundwork.

1. INTRODUCTION

A 2015 SIGGRAPH-Asia paper by Fadel Adib et al from the Massachusetts Institue of Technology (MIT) [1] has presented breakthroughs in through-wall imaging. Prior to the results found by the group of MIT researchers, there had been no successful techniques for rendering an image of a human body from the other side of a wall. All previous successful methods of gathering information of humans on the opposite side of a wall treat the body as a location. These papers are not without merit, for through the years progress in the area has been technologically infeasible. Two major innovations, recently introduced to the field, have proven crucial: multi-in multi-out antennas, which allow for the source of a signal to be directionally pinpointed, and the concept of nulling which involves using one wave to cancel-out a second wave in order to combat the noise that is reflected back by the wall, known as the flash (or mirror) effect.

2. BACKGROUND

2.1 Mirror Effect

Simply put, the mirror effect is defined as the loud signal that bounces off the wall that is separating the device and the target room. The effect is similar to that of a camera flash in front of a window. The reflected flash overwhelms the picture, eclipsing the darker background in favor of the window lit by the flash.

While the flash from the camera is within the human-visible electromagnetic wave spectrum, the electromagnetic radio waves used for the through-wall imaging are not visible by humans.

UMM CSci Senior Seminar Conference, April 2017 Morris, MN.

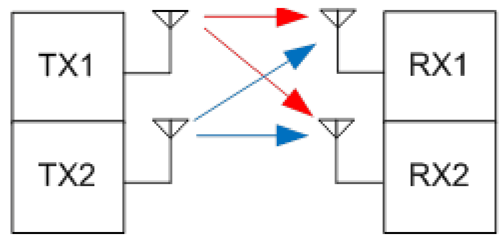


Figure 1: Signals are sent from the Tx antennas and received by the Rx antennas. The direction of origin can be located by measuring either the time or phase differences of a received signal [3].

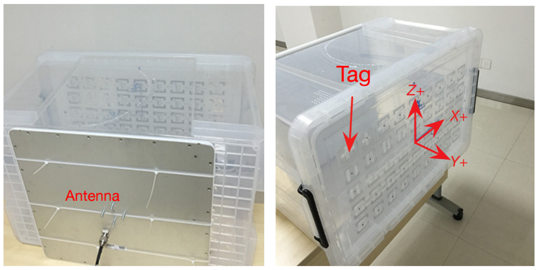
With radio frequencies, walls act as the window and the flash comes from the broadcasted signal. When dealing with through-wall imaging, this flash has to be compensated for when the reflected radio signals are received. A technique called *nulling* is the easiest way to handle the flash.

2.2 Nulling

There are two types of nulling: signal cancelling, and signal subtraction. Cancellative nulling is a technique that ignores stationary objects (like the wall that causes the mirror effect) and focuses on moving objects. Two or more waves are tuned so that they cancel each other out, from the receiving antenna's perspective, when bouncing off stationary objects. The technique allows for focus on moving objects, while ignoring things that are stationary. Subtractive nulling is taking an observed signal and subtracting from it a previously observed signal leaving behind the differences between the two. The previous observation is often from a stored observation (also known as a training observation) of the empty room. Cancellation nulling is used by both Wi-Vi (section 4) and RF-Capture (section 5). Subtractive nulling is used by Tadar (section 3).

2.3 Multi-in Multi-out Antennas

Multi-in Multi-out (MIMO), demonstrated in Figure 1, was popularized by cell phone towers in order to allow for 3G and better speeds. The technology is also seen in Wi-Fi routers that incorporate the IEEE standard 802.11n and its replacement. The concept is to squeeze more antennas onto a single router, or tower, allowing for increased throughput on the same broadcast frequency. The increased throughput is possible because the antenna can receive multiple signals on the same frequency at the same time by being able to



(a) Antenna (b) Tag array

Figure 2: RFID tags used in Section 3, the tags would actually be placed on the wall [4].

tell where the signals are coming from, much like a human can detect the direction a sound originates from. The main benefit for through-wall imaging is this directional location.

3. RFID

Lei Yang et al of the Tsingua University and the Illinois Institute of Technology [4] took a unique approach for through-wall imaging by using passive RFID tags as receiving antennas. There is a 4.21 meter average detection range through eight inches of concrete and a six meter average detection range beyond five inch drywalls. Drawbacks to their system include: the limitation of only being able to track one target at a time, and no FCC compliance (due to a 42 decibel per milliwatt broadcasting signal, which is well above the permitted 27 decibel per milliwatt limit). This system would not be tolerated outside of military use due to the lack of FCC compliance. [4]

Their system uses one Impinj Reader and 45 2"x2" passive RFID tags. The tags are arranged in a 9x5 grid with five centimeter spacing (see Figure 2 for illustration).

The grid is placed on the wall outside of the room that will be observed and the reader is positioned 13 inches away, and in line-of-sight, from the tags. Having the reader relatively close to the tags and in their line-of-sight is important because the tags are not capable of producing a strong signal. The reader broadcasts and receives signals simultaneously. The broadcasted signal is a continuous wave, meaning that the frequency, amplitude, phase, and so on, never change.

The system is able to detect a mobile target through reading the minute changes that are made to the continuous wave. The continuous wave has one meaningful path: the path from the reader through the wall, reflected off objects, back through the wall, to the tags and finally back to the reader. Passing through the RFID tags is fundamental because the RFID tags add their signatures to the signal so that the reader is able to gain a sense of the direction the waves traveled. Essentially, this system is a single input, 45 output antenna system. [4]

As mentioned, this system uses subtractive nulling. The room must first be free of movement in order to build a baseline of signal responses from the still room. This baseline is later subtracted from of the return signals so that all that remains is the changes from the baseline room and the current state of the room (only things that are out of place would remain in the signal, like a human).

Each tag is powered by the continuous wave and is able to make a small change to the original wave (known as mod-

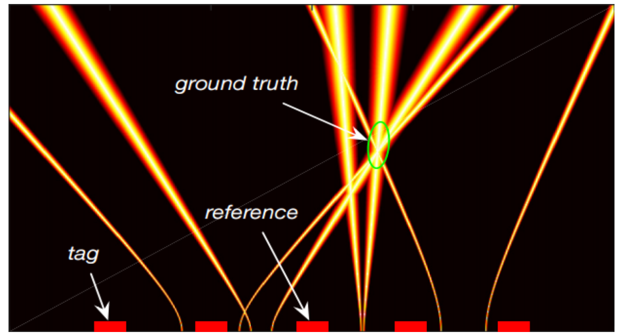


Figure 3: A pair of tags produces a hyperbola, and the targeted object lies at the intersection of multiple hyperbolas [4].

ulating) so that the reader will know which tag sent what signal. The forwarded wave is not strong enough to power the other tags, so there is no danger of the tags echoing each other's signals. The reader reads the entire group of tags roughly 45 times a second, creating the same amount of snapshots per second.

Each snapshot of the 45 tags is stored as a list of captured reflections. The captured reflection from each tag has the training reflection for that specific tag subtracted out, leaving the reflections from the target behind, represented as

$$h_{A->X[t]->T_1}$$

This is a complex number known as the channel parameter of the signal from reader A reflected off the target $X[t]$ to the tag T_1. The channel parameter contains the wave's channel anttenuation and its phase rotation. Each tag has the potential of being in various stages of a phase rotation, depending on the distance the wave travelled. The difference between the phase of any two tags ($\Delta\theta_{i,j}$) along with the wavelength, λ, can be used to derive the difference in distance of the two tags i and j from the target, the formula follows:

$$\Delta d_{i,j} = \lambda\left(\frac{\Delta\theta_{i,j}}{2\pi}\right)$$

The $\Delta d_{i,j}$ is then translated into a hyperbola, as shown in Figure 3. As more hyperbolas are added, some begin to intersect. The position where many hyperbolas intersect, shown as the *ground truth* in figure 3, is where the target is located.

The room is represented as a grid of N squares, with the size of N being subjective to how specific of a location is wanted; the specificity is able to reach the centimeter-level of accuracy. There is a median error rating of: "7.8cm and 20cm in the X and Y dimensions when monitoring the room from one side" [4], the Y is depth and the X is parallel to the wall. The goal is to find which grid squares are different from the baseline.

4. WI-FI

The first team to complete through-wall imaging with Wi-Fi was Fadel Adib and Dina Katabi, in 2013 [2]. This antenna array consists of two input antennas that each broadcast a signal and one output antenna that receives signals. The antennas are arranged in a line perpendicular to the direction of interest, with the output antenna being in the middle of the two input antennas.

Briefly described, the process is straightforward. The first step is to initialize each input antenna. The two antennas are initialized so that each wave will cancel out the other. Then, the device amplifies the signal and begins broadcasting. The receiving antenna receives the signal from many sources. There is a signal received from the antennas, the wall, and every other solid object. The received signal is analyzed, and the changes from one signal to the next are used to track people in a room. This device, nicknamed Wi-Vi, has reported results for up to three people with a maximum range of three meters for broadcasts through 8" concrete walls and 5" hollow walls, and 1.5" wooden doors.[2] Due to the use of a single receive antenna, it can only determine a person's distance from the router, and not their location. This system uses a two-stage nulling process, the initial pre-amplified null and the post-amplified iterative nulling. The first stage, the initial nulling starts with one transmit antenna sending a signal, the receiver will then tune the other antenna such that the two signals will cancel each other out. The tuning consists of matching the frequency and amplitude of the two waves, but having their phases out of sync by π. Doing this, the flash effectively does not register with the receiver, also any non-moving object is "invisible" to the receiver. Nulling is important, because it helps to minimize the amount of data per second that is received and analyzed. The iterative nulling stage persists indefinitely. The system alternates between which antenna to tune, but tunes one antenna after every signal pulse. When an antenna is being tuned, we pretend that the other antenna is perfect. This is fine since the two are meant to cancel each other out, so using one signal as a base is acceptable.

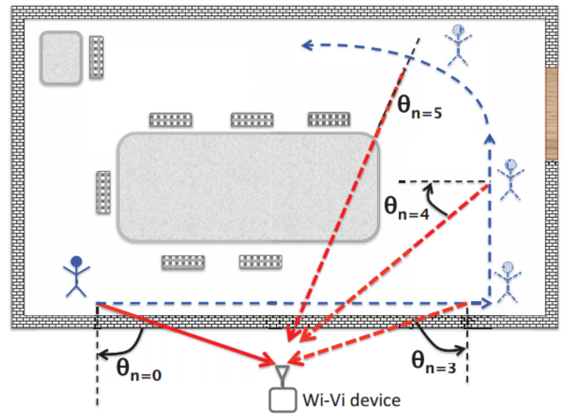
Wi-Vi uses a technique called inverse synthetic aperture radar, or ISAR, to track people's distance. ISAR is a technique that treats moving objects as an additional antenna. Now, we track how the antenna is moving. Referring to figure 4, Wi-Vi tracks the direction of an antenna's movement as an angle ranging from +90° for directly towards the router through -90° for directly away from the router. The researchers presume a moving speed of 1 m/s (2.25 mph), this estimated speed is combined with the space of time between two samples to form Δ, the suspected distance traveled between two sample frames. Due to the estimated speed, the actual direction of movement can differ from the reported direction as the person moves faster or slower than the estimated speed. The following formula is used to gauge the direction of movement $A[\theta, n]$:

$$A[\theta, n] = \sum_{i=1}^{w} h[n+i] e^{i \frac{2\pi}{\lambda} \Delta \sin\theta}$$

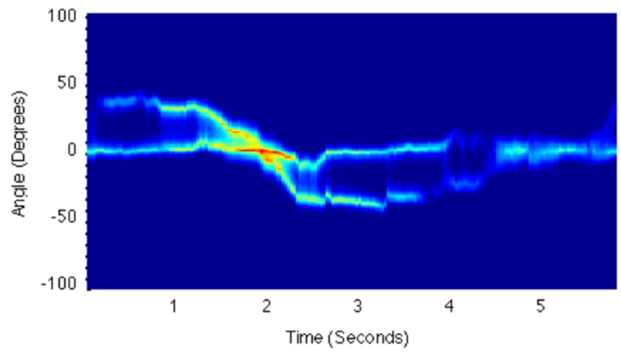
Where:
w = end of scanning period
$h[n+i]$ = received signal from scan n+i
n = starting scan
θ = spatial angle, as demonstrated in figure 4a
λ = wavelength

This value, $A[\theta, n]$, is plotted to a plot as shown in Figure 4b at time n and degree $A[\theta, n]$.

An assumption is made regarding the velocity of the "antenna." Since there is no way to tell the actual velocity, we estimate. The estimate is fixed at 1m/s, the average moving speed of a person in a room [2]. The computed value of $A[\theta, n]$ gives the direction of movement of a person, between -90° and 90°.



(a) **Experimental Setup**



(b) **Wi-Vi's output**

Figure 4: A person moves around the room, showing a positive angle while moving towards the Wi-Vi device, and a negative angle while moving away from the device.

5. THREE-DIMENSIONAL RADIO IMAGING

The most recent method, from Adib et al. at MIT [1], creates a three dimensional rendering from the radio waves. This system uses a T-shaped antenna array, with sixteen input antennas (which broadcast) and four output antennas (which receive). The diagram, Figure 5, shows how the antennas were arranged in the array. The operating frequency is 5.46-7.24Ghz and was tested on "...standard double dry walls supported by metal frames." [1]

The device requires a moving target (such as a human), and only supports one person within range at a time, in order to work properly. The human body, at the frequency used, acts as a polished mirror. That is to say, the body reflects the radio waves, and the antenna will only receive a signal from those waves that are reflected back to the antenna from a perpendicular surface. Because of this, the antenna is placed two meters off the ground. As with the other designs,

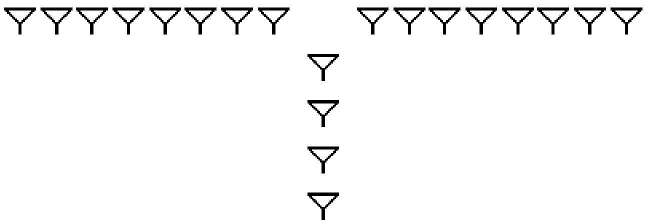


Figure 5: The row of antennas are the output antennas and the column of antennas are input antennas

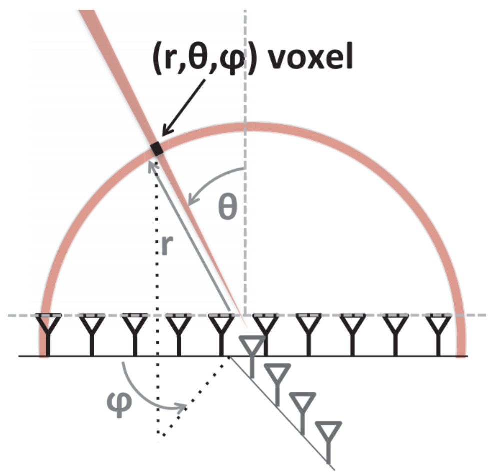


Figure 6: The voxel is a product of 3 axes.

the system ignores stationary objects.

The antenna is placed two meters high in order to use the human's chest and body as a reference point for the various body parts. The chest and body are virtually always reflecting back to the antenna, while the arms and legs are only perpendicular to the antenna array for short bursts of time while walking.

Since we attempt to render a 3-D space, there is a high computational complexity. In order to minimize the computations, the researchers use a Coarse-to-Fine 3D scan to focus the scans on areas where motion is captured. This way, computation won't be wasted on spaces without motion.

This entire process is done, and viewable, in real time. As with their earlier experiment, the efficacy range is less than eight meters, but this is due to the compact space of the antenna array, and a bigger antenna array would increase the resolution, and consequently, the range.

5.1 Coarse-to-fine 3-D scanning

Most of the return signal does not come from a human; there is a lot of space that is not of interest. The antenna array, or just array for short, starts with a coarse resolution scan, to locate the unique reflections from human targets. The scans attempt to paint a 3-D image of the space, each 'pixel' of the 3-D replica is called a voxel, short for volume-pixel (See Figure 6).

The coarse-to-fine scans converge on the human's position in both angle and depth. The coarse-to-fine scanning cycle is 85 milliseconds long and repeats immediately. The first scans are with one output antenna and the two middle-most input antennas. Their close proximity coerces a lower resolution scan. The next scan includes the next two inner input antennas and one more output antenna. The scans continue in this fashion until all antennas are involved. Any antennas beyond the first three are only necessary for the angular scanning precision; the depth precision is a function of bandwidth.

5.1.1 Depth Scanning

The depth scanning uses Frequency Modulated Carrier Wave (FMCW) chirps to measure "the frequency shift between the transmitted and received signal" [1]. This frequency chirp can be used to compute the signal power from a specific depth of r meters $\big(P(r)\big)$. It works by broadcasting a wave at a gradually increasing frequency, and comparing the frequency of the reflected wave to the frequency that is currently being broadcasted. Summed up, the change between the current frequency and the received frequency is used to gauge the distance that the wave had covered before returning.

$$P(r) = |\sum_{t=1}^{T} s_t e^{i2\pi \frac{kr}{3.00*10^8 m/s} t}|$$

$[t, T]$ is the range of time that is of interest, this range shrinks to focus on depths with high reflective power $P(r)$

k is the rate at which the frequency changes.

r is the number of meters from the array

s_t = baseband time signal t.

These scans start with a low resolution due to a wide range of $[t, T]$ and become more focused as range of time decreases to pinpoint reflective areas of the human.

5.1.2 Angular Scanning

Angular scanning pinpoints the direction of the object. It uses the axes of θ and ϕ as shown in the diagram, Figure 6. θ is the angle difference from the normal of the output antennas and ϕ is the angle difference from the normal of the input antennas.

The following formula demonstrates how the angular scanning focuses.

$$P(\theta) = |\sum_{n=1}^{N} s_n e^{i2\pi \frac{nd\cos\theta}{\lambda}}|$$

s_n = wireless signal received by antenna n

d = the distance between adjacent antennas.

5.1.3 Combined, 3D scan

The previous formulas are for if the two triangulations were done in separate steps, but they are done in one step, so the following formula is used. It has a cubic computational power, for a single voxel. Each voxel is a record of the strength of the reflected space, and can be computed with:

$$P(r,\theta,\phi) = |\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{t=1}^{T} s_{n,m,t} e^{i2\pi\frac{kr}{c}t} e^{i\frac{2\pi}{\lambda}\sin\theta(nd\cos\phi + md\sin\phi)}|$$

$P(r, \theta, \phi)$ = the power of the reflected signal at a given position in space.

r = meters from array

θ = spatial angle
ϕ = angle relative to input antennas
N = number of output antennas
M = number of input antennas
T = time
$s_{n,m,t}$ = signal received by antenna n from antenna m at time t.

This successfully combines the algorithms from 5.1.1 and 5.1.2 into one formula. This array will be fed to the GPU for processing. There it will be combined with other 85ms frames and further processed by the motion-based figure capture.

5.2 Motion-based Figure Capture

Combining multiple 85ms snapshots into a video stream presents two challenges: compensating for depth changes as the person moves and compensating for swaying of the person.

5.2.1 *Compensation for Depth*

Since the target is moving, their relative size changes between cuts, also the image is sharper for closer subjects and blurrier for more distant subjects. A variation of the approach in 5.1.3 is used to compensate for the size differences.

5.2.2 *Compensation for Swaying*

A person's natural gait includes some slight swaying, so the system needs to adjust in order to realign the images for better fits between frames. The chest is used as a pivot/anchor because it is the largest reflector on the human body. With this knowledge, the image can be rotated to best match with the position of the chest between different cuts.

6. DISCUSSION

The three systems presented here are representative of current work in the field of through-wall imaging (TWI). Most TWI devices used to be restricted to military-use only due to non-FCC compliance because of a loud broadcast signal, the use of a greedily large portion of the radio spectrum, and a huge amount of electricity. Newer devices have been successful in reaching FCC compliance, but have a big price-tag, a large power draw, or both. With this in mind, the goal of the devices that were discussed in this paper is to have both a low power consumption and a lower purchase cost. Wi-Vi and Tadar would cost around \$2,000 to purchase the hardware (for those interested in building the system)[4], and RFCapture was built with "...low-cost massively-produced components..."[1]. Each device was tested in a similar environment that consists of a conference room - complete with chairs surrounding a table - and a standard drywall wall (two slabs of drywall separated by 5-8 inches of air and held together by two inch metal studs).

These systems were also tested on gesture encoding. This measures the devices' ability to recognize certain actions as the subject attempting to communicate directly to the machine. Wi-Vi's gesture encoding consists of a person taking either a step towards (sends a '1' bit to the system) or away (sends a '0' bit to the system) and then back. The device was able to read this signal with 100% accuracy up to a distance of seven meters and 75% accuracy at eight meters, with 0% at nine meters. After seven meters, the signal begins to be drowned out by background noise and the router

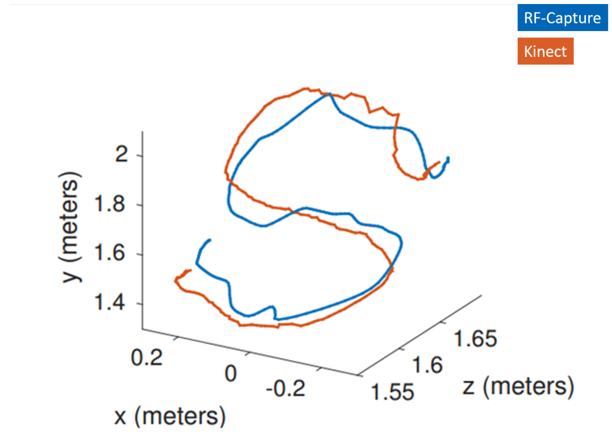


Figure 7: The test subject drew the letter "S" in the air [1].

begins to grow less effective at detecting the person until it can no longer detect the person around nine meters.

Tadar uses a side-to-side movement, in opposition to Wi-Vi's forward-backward approach, because Tadar is more sensitive to changes in the X-axis than the Y-axis. Tadar's gesture accuracy begins to drop after three meters, with a 77% accuracy at four meters, a 53% accuracy at five meters, and 0% at six meters. The accuracy of gesture encoding is used to measure functioning distance because it presents the devices' ability to interpret information that it is gathering. If a human is reading the output, Tadar's through-drywall range is extended from shy of four meters out to eight meters. At four meters and onwards, the signal begins to deteriorate, but is not fully drowned by the background noise until after five meters.

As stated earlier, RFCapture works by stitching together different frames, and needs to be able to identify body parts in order to ensure success. Unlike the other two methods, the body part identification is never at 100% accuracy since it is inherently more complicated. The accuracy does taper off more gradually than either Tadar or Wi-Vi. Starting with three meters, the accuracy is 99.13% and drops to 91.60% by five meters. At eight meters the accuracy is 76.48%, but then drops off before reaching nine meters. The range can be improved by placing the antennas further apart, without any change to the broadcasting power. RFCapture was also compared to an Xbox One Kinect (the Kinect was in the room) for an accuracy comparison. Another gesture test involved a person drawing letters in the air, with their hand, being about one meter away from the RFCapture and Kinect. One captured drawing is shown in Figure 7, the results are nearly identical, with a 2.19cm median for error between the two.

Current through-wall imaging devices are only usable by well-to-do police departments and the military due to the high operating cost. However, there is an interested party that is biding their time for a more frugal TWI device: the healthcare industry. Though-wall imaging can serve as a non-invasive way to monitor patients, with particular emphasis on eldercare due to their interest in "aging with dignity" as well as the threat of a looming retirement crisis. Through-wall imaging is seen as a potential way to lessen the blow of the impending shortage of healthcare professionals.

7. CONCLUSIONS

The use of radio imaging to "see" through walls is computationally challenging. The approaches mentioned have taken measure to alleviate the workload (commonly by ignoring stationary objects). The through-wall approaches here are not the entirety of available methods. These approaches mark a transition from military to civilian. These researchers are leading the way to branching through-wall imaging out of the military sector by striving to reach FCC compliance and leave behind the high-energy usage of the military-only versions. There is currently growing interest from law enforcement and health care professionals. The health care aspect is for non-intrusive monitoring of patients.

8. REFERENCES

[1] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand. Capturing the human figure through a wall. *SIGGRAPH Asia '15*, November 2015.

[2] D. K. Fadel Adib. See through walls with wi-'fi! *SIGCOMM '13*, October 2013.

[3] B. Petted. Basics of mimo radio systems.

[4] L. Yang, Q. Lin, X. Li, T. Liu, and Y. Liu. See through walls with cots rfid system! *MobiCom '15*, September 2015.