On Influencing Factors in Human Activity Recognition using Wireless Networks

Haochen Hu, Zhi Sun and Lu Su
University at Buffalo, State University of New York, NY, USA, 14260
E-mail:{haochenh, zhisun, lusu}@buffalo.edu

***Abstract*— Driven by the development of machine learning and the development of wireless techniques, lots of research efforts have been spent on the human activity recognition (HAR). Although various deep learning algorithms can achieve high accuracy for recognizing human activities, existing works lack of a theoretical performance upper bound which is the best accuracy that is only limited by the influencing factors in wireless networks such as indoor physical environments and settings of wireless sensing devices regardless of any HAR algorithm. Without the understanding of performance upper bound, mistakenly configuring the influencing factors can reduce the HAR accuracy drastically no matter what deep learning algorithms are utilized. In this paper, we propose the HAR performance upper bound which is the minimum classification error probability that doesn't depend on any HAR algorithms and can be considered as a function of influencing factors in wireless sensing networks for CSI based human activity recognition. Since the performance upper bound can capture the impacts of influencing factors on HAR accuracy, we further analyze the influences of those factors with varying situations such as through the wall HAR and different human activities by MATLAB simulations.**

I. INTRODUCTION

In recent years, the development of wireless infrastructure has enabled various indoor human sensing techniques. Human respiration sensing is implemented with commercial WiFi device according to the Fresnel diffraction model [1]. Indoor human movement tracking is achieved by MUSIC doppler shift algorithm in [2] where the differences in intervals of adjacent measurements are considered to be introduced by human movements. Among all the indoor sensing techniques, channel state information (CSI) based indoor sensing techniques are commonly applied such as crowd counting [3], human activity recognition [4] and indoor localization [5] since CSI can be extracted by commercial WiFi devices conveniently without the burden and limitations of device-based approaches. The capability of CSI based human activity recognition technique is even able to classify the minor movements of human fingers such as keystroke keyboard recognition [6]. Machine learning based wireless sensing techniques are also utilized for human activity recognition [7] and person identification [8] with high accuracy.

Deep learning based CSI wireless sensing techniques seem to have an unlimited potential. However, there are still two major problems lacking of solutions. First, there is no study about the best accuracy which is the performance upper bound only limited by the influencing factors in wireless networks. The factors such as the physical environments, types of human activities, the settings and artifacts of wireless sensing devices can restrict the accuracy of wireless based human activities recognition. If the HAR performance upper bound is restricted to a low accuracy, no matter what algorithms are applied for improving the accuracy cannot exceed that upper bound. Only configuring the influencing factors in wireless networks can improve the HAR performance. However, as the second problem, the impacts of influencing factors on HAR performance of the existing wireless sensing system cannot be predicted. The impacts of influencing factors are always evaluated by experiments, which makes the adjustment of wireless sensing system time consuming and ineffective. Therefore, it is important to figure out the upper bound of human activity recognition and the impacts of the influencing factors on the performance of human activity recognition.

In this paper, we propose a CSI based human activity recognition upper bound that doesn't depend on any deep learning algorithms and can be considered as an expression of influencing factors in wireless networks to characterize the impacts of influencing factors. To relate the upper bound to influencing factors, firstly, we propose a wireless sensing model in Section II. The wireless sensing model describes the entire indoor wireless sensing procedures with commercial WiFi routers. The CSI measurements taken by the WiFi routers can be considered as the combination of static part caused by indoor environment and the changing part caused by the motion of human body. CSI measurements are arranged as a high dimensional matrix according to the time, receivers, sub-carrier frequencies and antennas. Second, with the aforementioned CSI sensing model, we derived the minimum error probability which is upper bound for human activities classification. To avoid the specific deep learning algorithm and achieve the best classification accuracy, we utilize the "best point" assumption which is a CSI measurement that has the minimum influence from the complex normal noise. The best point has the random location among the different collected data sets due to the randomness of noise. Only the "best point" with minimum influences of noise is used to classify the human activities without the negative influences of other "noisy" CSI measurements. Obviously, always knowing the CSI measurement with least noise influence is impractical. But that is the reason that our proposed theoretical upper bound can achieve the highest accuracy compared with any existing classification algorithm. Finally, we establish a MATLAB simulation environment with specified activities to analyze the influences of different influencing factors in wireless sensing network. We find out that the most efficient way to improve the human activity recognition is adding more distributed receivers.

The rest of the paper is organized as following. First, in Section II, the wireless sensing model is presented. We provide the real channel between a transmitting antenna and a receiving antenna and describe the impact of WiFi routers when taking CSI measurements. Then in Section III, we derive the minimum

This work was supported by the US National Science Foundation (NSF) under Grant No. 1652502 and 1718177.

978-1-7281-0962-6/19/$31.00 ©2019 IEEE

error probability for human activity recognition performance. In Section IV, we use the simulation to analyze the influencing factors' impacts on the human activity recognition performance with various situations. Finally, Section V concludes this paper.

II. WIRELESS SENSING MODEL

For the indoor wireless sensing, the channel state information is contributed by two main components. The first part is introduced by the signals that are reflected off the obstacles in the indoor environment such as walls and furniture. In a certain indoor environment without human activities, the channel between the receiver and transmitter can be considered as constant. Different indoor environments introduce various constant CSI values. The second main component is caused by the human activities, which is dynamic due to the human motions. The transmitted signals arrive at the receiver via several propagation paths. For a certain sub-carrier frequency f_m, the wireless channel between a transmitter receiver pair's (TRP) transmitting antenna and a receiving antenna can be described as the summation of different signal propagation paths [9]:

$$h(f_m,t) = \sum_{l_e \in L_{Env}} \alpha_{l_e}(f_m,t) e^{-j2\pi f_m \tau_{l_e}} + \sum_{l_e \in L_{Sca}} \alpha_{l_s}(f_m,t) e^{-j2\pi f_m (1+\frac{V_{path}(t)}{c})\tau_{l_s}(t)} \quad (1)$$

where L_{Env}, L_{Sca} are total numbers of signal propagation paths caused by environment and human movements respectively. l_e, l_s denote the paths caused by indoor environment and human activities. V_{path} denotes the speed of path length change caused by human motion. c is the speed of light in free space. $\alpha_{l_e}, \alpha_{l_s}$ denote the attenuation coefficients of indoor environment obstacles and human body respectively. In addition, those attenuation coefficients are affected by the frequencies and time according to the Mie scattering. The signals that encounter clothes of human actually are scattered towards different direction instead of reflection. At different time, signals are scattered at different positions of human body, which also influence the attenuation coefficients. $\tau_{l_e}, \tau_{l_s}(t)$ are time delays for environment paths and human movement paths. Time delays of environment paths are constant since the path length of environment paths are constant. On the contrary, lengths of paths that are caused by human activities vary with human movement, which leads to changing time delay $\tau_{l_s}(t)$ as a function of time t. The received signal can be described as following:

$$y(t) = e^{-j\psi_{Tx-Rx}t} h(t) \otimes x(t) + \eta \quad (2)$$

where $e^{-j\psi_{Tx-Rx}t}$ denotes the phase shift caused by center frequency offset (CFO) ψ_{Tx-Rx}. CFO is the mismatch between the carrier frequency of transmitter and the carrier frequency of receiver. For different transmitter receiver pairs, the CFO values are different. η denotes the noise and is considered to have Complex Normal (CN) distribution. $x(t)$ is the transmitted signal. The transmitted signal is convolved with the channel in this transmitter receiver pair. Theoretically, the CSI values can be calculated as following equation:

$$\hat{h}(f_m,t) = IFFT[\frac{Y(f)}{X(f)}] \quad (3)$$

where $Y(f), X(f)$ denote the transmitted signal and received signal after fourier transformation. IFFT denotes the inverse fast fourier transformation.

Once the wireless sensing system finishes taking CSI measurements, how to arrange the collected data is the next important procedure. CSI measurements are influenced by the sampling time, positions of receivers, sub-carrier frequency and different antennas. We propose a high dimensional data structure. Firstly, CSI measurements are arranged by the order of sampled time:

$$R = [R_{t_1}, R_{t_2}, ..., Rt_n] \quad (4)$$

where R denotes the collected data structure, R_{t_n} denotes the measurements taken at time t_n. In each R_{t_n}, CSI measurements are separated according to the different receivers.

$$R_{t_n} = [Rx_1(n), Rx_2(n), ..., Rx_{N_R}(n)]^T \quad (5)$$

where $Rx_{N_r}(n)$ denotes the measurements taken at $N_r - th$ receiver. At each receiver, the measurements are arranged by the sub-carrier frequencies and different antennas:

$$Rx(n) = \begin{bmatrix} CSI^1_{f_1}(n) & CSI^2_{f_1}(n) & ... & CSI^I_{f_1}(n) \\ CSI^1_{f_2}(n) & CSI^2_{f_2}(n) & ... & CSI^I_{f_2}(n) \\ ... & ... & ... & \\ CSI^1_{f_M}(n) & CSI^2_{f_M}(n) & ... & CSI^I_{f_M}(n) \end{bmatrix} \quad (6)$$

where the $CSI^i_{f_m}(n)$ denotes the complex CSI value with sub-carrier frequency f_m taken at $i-th$ antenna. The final collected data structure is a high dimensional matrix $R[N, N_R, M, I]$, where N denotes number of sampling time instant, N_R denotes the number of receiver, M denotes the number of sub-carrier frequency and I denotes the number of antennas in a receiver. To simplify the expression of collected data structure, we use k to represent index of CSI measurements and the total number of CSI measurements $K = N \cdot N_R \cdot M \cdot I$.

III. MINIMUM CLASSIFICATION ERROR PROBABILITY

In this section, we theoretically derived the minimum error probability for binary human activity recognition in wireless networks. This error probability expression can characterize the impacts of the influencing factors such as human activities, indoor environment and wireless sensing networks deployment since it can capture the key effects caused by the influencing factors on human activity recognition without the influences of different HAR deep learning algorithms.

It is noteworthy that the minimum error probability is not the achievable value for the human activities classification. Furthermore, performance upper bound cannot be based on any specific classification algorithm such as machine learning algorithms. The classification performance upper bound should be derived theoretically but not experimentally. Therefore, we propose an important assumption of "Best Point". Before explaining the assumption of "Best Point", there are several terms need to explain:

- *Ground truth data structure*: The ground truth data structure actually is the collected data structure R, however every CSI value in ground truth data structure is the measurement without complex normal noise. An important assumption is that we know all the parameters in the wireless sensing networks such as CFO, time delays and V_{path}. And we set transmitted signal as unit pulse $\delta(t)$. Therefore, the ground truth data set exactly contains CSI values caused by human activity. For binary human

activity classification, $G_1(k), G_2(k)$ denote the $k-th$ CSI value in ground truth data structure of activity 1 and activity 2. Ground truth pair distance denotes the euclidean distance between $G_1(k)$ and $G_2(k)$ in complex plane, which can be described as: $\sqrt{G_1(k) - G_2(k)} = |\Delta G_k|$

- *Testing data set*: Testing data set contains the CSI measurements that are caused by a human activity and complex normal noise. The CSI measurements are complex values. We try to fully utilize the properties of CSI measurements instead of only using amplitude of CSI. CSI values in testing data set are arranged as the collected data structure. The best point among the testing data set is denoted as $T(best)$
- *Best Point*: The best point is the CSI measurement in testing data set that has the minimum ratio of noise amplitude Z to the ground truth pair distance of the best point $|\Delta G_{best}|$.

In our defined model above, the classification error only comes from the randomness of addictive complex normal noise. To seek for the minimum classification error, we also propose an assumption that we always know which CSI value among the testing data set is the best point. The "Best Point" is not a fixed or pre-selected point among the collected data structure, since there is not a position in the testing data set always having the minimum noise to distance ratio. The position of the Best Point is random when wireless sensing networks take CSI measurements for the same human activity several times. Therefore the assumption that we always know the Best Point is impossible in practical. Compared with using the CSI value with the minimum amplitude of noise, using the CSI value with the minimum noise to distance ratio has a better performance. The reason is that a CSI value in testing data set can have a minimum noise amplitude and also the minimum ground truth pair distance, which is still not the best choice for classification.

By utilizing that Best Point assumption, we are always using the most accurate CSI value to classify the human activity and this assumption avoids the negative influence of other more "noisy" CSI measurements. Therefore, it is efficient to utilize the euclidean distance in complex plane to classify human activities with "Best Point". It is emphasized that we always know the ground truth data of human activity 1 G_1 and activity 2 G_2 for a certain pair of human activities,.

The complex normal noise can be described as $\eta = a + jb$, where $a, b \sim N(0, \sigma^2)$. The amplitude $Z = \sqrt{a^2 + b^2}$ of complex normal noise has the Rayleigh distribution:

$$f(z, \sigma) = \frac{z}{\sigma^2} e^{-z^2/2\sigma^2} \tag{7}$$

The minimum noise amplitude to ground truth pair distance ratio is defined as random variable C as:

$$C = argmin_k(\frac{Z_k}{|G_1(k) - G_2(k)|}) \tag{8}$$

The size of the test data set is $K = N \cdot M \cdot I \cdot R$. For the simplicity of expression, we utilize index k to denote the position of CSI measurement in ground truth and testing data set. To obtain the probability density function of C, we define random variable $\tilde{Z}_k$ as the ratio of noise amplitude to ground truth pair distance for $k-th$ CSI measurement in testing data set.

$$\tilde{Z}_k = Z_k / |\Delta G_k| \tag{9}$$

where Z_k is the noise amplitude of $k-th$ CSI measurement in testing data set. $\tilde{Z}_1, \tilde{Z}_2, ...\tilde{Z}_k$ are independent random variables that have Rayleigh distribution with different variance and mean. To derive the probability density function of C, we first need to calculate the cumulative probability distribution of C, which is denoted as $F_C(c)$:

$$\begin{aligned} F_C(c) &= P(min(\tilde{Z}_1, \tilde{Z}_2, ...\tilde{Z}_K) \leq c) \\ &= 1 - P(\tilde{Z}_1 > c, \tilde{Z}_2 > c, ..., \tilde{Z}_K > c) \\ &= 1 - \prod_k^K [1 - F_{\tilde{Z}_k}(c)] \end{aligned} \tag{10}$$

After calculating the cumulative density function, the probability density function $f_C(c)$ can be calculated as a derivative of $F_C(c)$:

$$\begin{aligned} f_C(c) &= [1 - \prod_k^K [1 - F_{\tilde{Z}_k}(c)]]' \\ &= \sum_{k=1}^K [f_{\tilde{Z}_k}(c) \prod_{\hat{k}}^{\hat{K}} (1 - F_{\tilde{Z}_{\hat{k}}}(c))] \end{aligned} \tag{11}$$

where $\hat{k} \neq k$ and $\hat{K} = K - 1$. The probability density function of $f_{\tilde{Z}_k}(z_k)$ and cumulative density function $F_{\tilde{Z}_k}(z_k)$ can be calculated as a transformation of PDF and CDF of Rayleigh distribution respectively:

$$f_{\tilde{Z}_k}(\tilde{z}_k) = |\Delta G_k| f_Z(|\Delta G_k| \cdot \tilde{z}_k) = \frac{||\Delta G_k||^2 \cdot \tilde{z}_k}{\sigma^2} e^{-\frac{(|\Delta G_k| \cdot \tilde{z}_k)^2}{2\sigma^2}} \tag{12}$$

$$F_{\tilde{Z}_k}(\tilde{z}_k) = F_Z(|\Delta G_k| \cdot \tilde{z}_k) = 1 - e^{-(|\Delta G_k| \cdot \tilde{z}_k)^2/2\sigma^2} \tag{13}$$

Although we have derived the PDF of C which is minimum ratio of noise amplitude to ground truth euclidean distance, the direction of the noise amplitude is still unknown. The complex normal noise can be described as $\eta = z \cdot e^{j\vartheta}$. The term $e^{j\vartheta}$ is the unit circle in the complex plane. The direction ϑ of noise amplitude is the included angle between amplitude Z and positive real axis, which is calculated anti-clockwise. ϑ can be any point from 0 to 2π. Then, we calculate the PDF of random variable ϑ that describes direction of amplitude of complex normal noise in complex plane. We start from calculating the joint PDF $f_{X,Y}(X, Y)$ of complex normal noise as following equations:

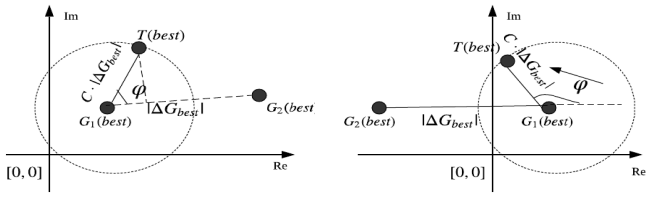
$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x) \cdot f_Y(y) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-y^2}{2\sigma^2}} \\ &= \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \end{aligned} \tag{14}$$

where x, y denote coordinates in real and imaginary axis respectively.

In the polar coordinates, $r = h_1(x, y) = \sqrt{x^2 + y^2}$, and $\vartheta = h_2(x, y) = tan^{-1}(\frac{y}{x})$. On the contrary, $x = g_1(r, \vartheta) = rcos\vartheta$ and $y = g_2(r, \vartheta) = rsin\vartheta$. By calculating the jacobian determinant:

$$J(x, y) = \begin{vmatrix} \frac{\partial h_1(x,y)}{\partial x} & \frac{\partial h_1(x,y)}{\partial y} \\ \frac{\partial h_2(x,y)}{\partial x} & \frac{\partial h_2(x,y)}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{vmatrix} = \frac{1}{r} \tag{15}$$

The joint PDF $f_{R,\vartheta}(r, \vartheta)$ can be calculated using joint PDF of random variables X, Y:



(a) $\varphi \in [0, 0.5\pi] \cup [1.5\pi, 2\pi]$ (b) $\varphi \in [0.5\pi, 1.5\pi]$

Fig. 1: Potential relative positions of $G_1(best)$ and $G_2(best)$

$$f_{R,\vartheta}(r,\vartheta) = \frac{f_{X,Y}(g_1(r,\vartheta), g_2(r,\vartheta))}{|J(X,Y)|} = \frac{1}{2\pi} \cdot \frac{r}{\sigma^2} e^{\frac{-r^2}{2\sigma^2}} \tag{16}$$

where joint PDF $f_{R,\vartheta}(r,\vartheta) = f_R(r) f_\vartheta(\vartheta)$. Radius $r = \sqrt{x^2 + y^2}$ is actually the amplitude of complex normal noise Z. As shown in equations above, r has Rayleigh distribution. Then PDF $f_\vartheta(\vartheta) = \frac{1}{2\pi}$. It is noteworthy that the complex normal noise is defined as circularly symmetric complex normal, which means that the distribution of noise has zero mean and zero relation matrix. Then the complex random variable $\eta = a + jb$ has the same distribution with $\eta e^{j\beta}$ for all real β.

To classify the best point $T(best)$ in the testing data set, we compare the euclidean distances from $T(best)$ to its corresponding ground truth values $G_1(best)$ and $G_2(best)$. We assign the testing data to the human activity that has a shorter distance in complex plane. As shown in Fig. 1 we assume that the $T(best)$ belongs to the $G_1(best)$. However, due to the influence of complex normal noise, $T(best)$ can be any position on the circle that has the center of $G_1(best)$ and the radius of $C|\Delta G_{best}|$. The relative positions of $G_1(best)$ and $G_2(best)$ are arbitrary. To simplify the calculation, we define the included angle φ between the line $T(best)G_1(best)$ and line $T(best)G_2(best)$. And the included angle φ is calculated anti-clockwise so that the range of φ is $[0, 2\pi]$. Then, we actually use ce^{φ} to represent the complex normal noise. As a matter of fact, the included angle φ is not equal to the angle ϑ that describes the angle between direction of noise and real axis. Due to the circularly symmetric, the distribution of $z \cdot e^{\vartheta}$ is the same with distribution of $z \cdot e^{\varphi}$, which means that φ also has the uniform distribution on $[0, 2\pi]$.

Then, the error comes form the situation that the euclidean distance from $T(best)$ to $G_2(best)$ is shorter than the euclidean distance from $T(best)$ to $G_1(best)$. By using the cosine theorem, error probability can be calculated as following equations when $\varphi \in [0, 0.5\pi] \cup [1.5\pi, 2\pi]$ as shown in Fig.1 (a):

$$\begin{aligned} P(Error) &= P(\|T_{best} - G_{1-best}\|_F^2 \geq \|T_{best} - G_{2-best}\|_F^2) \\ &= P([C \cdot |\Delta G_{best}|]^2 \geq [C \cdot |\Delta G_{best}|]^2 + |\Delta G_{best}|^2 - \\ &\quad 2C|\Delta G_{best}|^2 cos\varphi) \\ &= P(C \cdot cos\varphi \geq 0.5) \end{aligned} \tag{17}$$

When $\varphi \in [0.5\pi, 1.5\pi]$ as shown in Fig.1 (b), the error probability can be calculated as following equations:

$$\begin{aligned} P(Error) &= P(\|T_{best} - G_{1-best}\|_F^2 \geq \|T_{best} - G_{2-best}\|_F^2) \\ &= P(C \cdot cos(\pi - \varphi) \geq 0.5) \end{aligned} \tag{18}$$

where C, φ are independent random variables. To calculate the error probability, let us define random variable $S_1 = cos\varphi$. To calculate the PDF of S_1 $f_S(s)$, we need to find the range of $A_{s_1} := \{\varphi : cos\varphi \leq s_1\}$, We need to take inverse cosine on both sides and take the monotony of $cos\varphi$ in to consideration. For the range $\varphi \in [0, \pi]$, the $cos\varphi$ is monotonically decreasing. For the range $\varphi \in [\pi, 2\pi]$, $cos\varphi$ is monotonically increasing. After taking inverse cosine, the inequality can be derived as following:

$$\begin{aligned} &cos^{-1}s_1 \leq \varphi \leq 0.5\pi; if \varphi \in [0, 0.5\pi] \\ &1.5\pi \leq \varphi \leq 2\pi - cos^{-1}s_1; if \varphi \in [1.5\pi, 2\pi] \end{aligned} \tag{19}$$

To calculate the probability density function of s_1, we start from calculating the cumulative probability density first:

$$\begin{aligned} &F_{S_1}(s_1) = P(S_1 \leq s_1) = P(cos\varphi \leq s_1) \\ &= F_\varphi(2\pi - cos^{-1}s_1) - F_\varphi(1.5\pi) + F_\varphi(0.5\pi) - F_\varphi(cos^{-1}s_1) \end{aligned} \tag{20}$$

Then, we take derivatives of CDF $F_{S_1}(s_1)$

$$\begin{aligned} f_{S_1}(s_1) =& F'_\varphi(2\pi - cos^{-1}s_1) - F'_\varphi(cos^{-1}s_1) - f_\varphi(1.5\pi) + f_\varphi(0.5\pi) \\ =& \frac{1}{\pi}\left(\frac{1}{\sqrt{1 - s_1^2}}\right) \end{aligned} \tag{21}$$

For the second situation where the classification error occurs, we define $S_2 = cos(\pi - \varphi)$. In order to find out the PDF of S_2, we need to find the range of $A_{s_2} := \{\varphi : cos(\pi - \varphi) \leq s_2\}$. Similarly, we have to consider monotone increasing and monotone decreasing when taking inverse cosine. The range of φ is described as following:

$$\begin{aligned} &\pi + cos^{-1}s_2 \leq \varphi \leq 1.5\pi; if \varphi \in [\pi, 1.5\pi] \\ &0.5\pi \leq \varphi \leq \pi - cos^{-1}s_2; if \varphi \in [0.5\pi, \pi] \end{aligned} \tag{22}$$

the cumulative density function $F_{S_2}(s_2)$ can be calculated as following:

$$\begin{aligned} &F_{S_2}(s_2) = P(S_2 \leq s_2) = P(cos(\pi - \varphi) \leq s_2) \\ &= F_\varphi(\pi - cos^{-1}s_2) + F_\varphi(1.5\pi) - F_\varphi(0.5\pi) - F_\varphi(\pi + cos^{-1}s_2) \end{aligned} \tag{23}$$

Then we take derivative of CDF $F_{S_2}(s_2)$ to calculate the PDF $f_{S_2}(s_2)$:

$$\begin{aligned} f_{S_2}(s_2) =& F'_\varphi(\pi - cos^{-1}s_2) + F'_\varphi(1.5\pi) - F'_\varphi(0.5\pi) - F'_\varphi(\pi + cos^{-1}s_2) \\ =& \frac{1}{\pi}\left(\frac{1}{\sqrt{1 - s_2^2}}\right) \end{aligned} \tag{24}$$

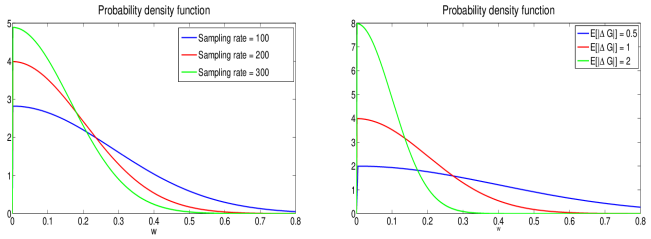
After calculating the PDF of $cos(\pi - \varphi)$ and the PDF of $\cos\varphi$, we found that they have the same distribution. No matter which situation the classification error occurs as shown in Fig.1 (a) or Fig.1 (b), the joint PDF $F_{C,\varphi}(c, \varphi)$ does not change. So we can safely calculate the error probability with one situation where we define random variable $S = cos\varphi$. To calculate the error probability, let $W = g(c, s) = C \cdot S$ and we define auxiliary random variable $Q = h(c, s) = S$. The universal classification error boundary is $W \geq 0.5$. We need to calculate PDF of W by using jacobian determinant:

$$J(c,s) = \begin{vmatrix} \frac{\partial g(c,s)}{\partial c} & \frac{\partial g(c,s)}{\partial s} \\ \frac{\partial h(c,s)}{\partial c} & \frac{\partial h(c,s)}{\partial s} \end{vmatrix} = \begin{vmatrix} s & c \\ 0 & 1 \end{vmatrix} = s \tag{25}$$

The joint PDF $f_{W,Q}(w, q)$ can be derived from the joint PDF $f_{C,S}(c, s)$ by utilizing the jacobian determinant:

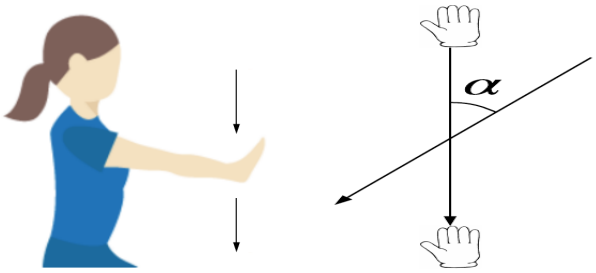
$$f_{W,Q}(w,q) = \frac{f_{C,S}(w/q, q)}{|J(c,s)|} = \frac{f_{C,S}(w/q, q)}{|q|} \tag{26}$$

To obtain the PDF of W, we integrate r.v Q marginally as



(a) $f_W(w)$ versus sampling rate (b) $f_W(w)$ versus average $|\Delta G|$

Fig. 2: PDF of W with different parameters



(a) Side view of defined activity (b) Difference angle α

Fig. 3: Defined human activities

following equations:

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{+\infty} f_{W,S}(w,q)dq \\ &= \int_{-\infty}^{+\infty} \frac{1}{q} f_{C,S}(\frac{w}{q},q)dq \\ &= \int_0^1 \frac{1}{q} \sum_{k=1}^{K} [f_{\bar{Z}_k}(\frac{w}{q}) \prod_{\hat{k}}^{\hat{K}} (1 - F_{\bar{Z}_{\hat{k}}}(\frac{w}{q}))] \frac{1}{\pi} (\frac{1}{\sqrt{1-q^2}})dq \end{aligned} \tag{27}$$

The error probability of human activity recognition can be calculated as following equations:

$$\begin{aligned} P(Error) &= \int_{\frac{1}{2}}^{\infty} f_W(w)dw \\ &= \int_{\frac{1}{2}}^{\infty} \int_0^{+1} \frac{1}{q} \sum_{k=1}^{K} [f_{\bar{Z}_k}(\frac{w}{q}) \prod_{\hat{k}}^{\hat{K}} (1 - F_{\bar{Z}_{\hat{k}}}(\frac{w}{q}))] \frac{1}{\pi} (\frac{1}{\sqrt{1-q^2}})dqdw \end{aligned} \tag{28}$$

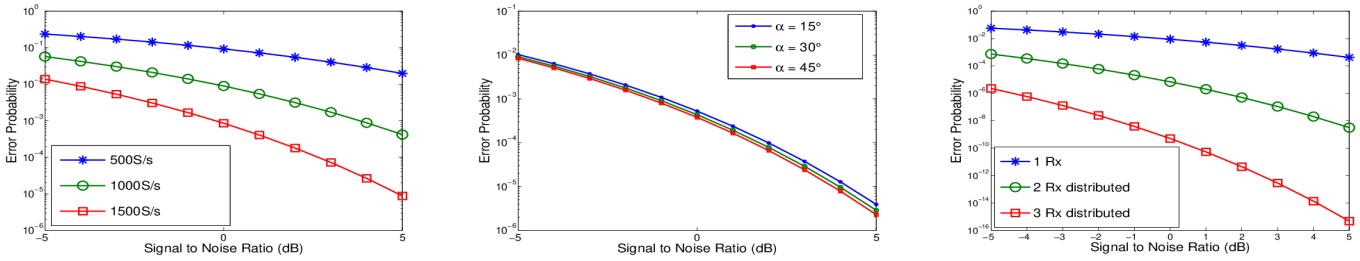
IV. NUMERICAL RESULTS AND ANALYSIS

In this section, we try to discuss different classification performances caused by the various input factors such as different activities, signal to noise ratio, CSI sampling rate, carrier frequency, number of receivers and through the wall sensing situation. The classification is performed by our proposed minimum error probability algorithm. The influencing factors in wireless network which reduce the minimum classification error probability could also improve the classification accuracy in reality for human activity recognition.

Before embarking the analysis of influences of different factors in wireless sensing networks, we first introduce our MATLAB simulation setup. The binary human activity classification is performed by our proposed minimum error probability algorithm. Since the common activities such as walking and sitting down have been already classified in many previous papers and our proposed minimum error probability algorithm can achieve extremely high accuracy when applied for classifying those activities which are not similar to each other, we try to classify some activities which are very similar. The first activity is defined as stretching arm in front of the body and sliding one arm downward vertically as shown in Fig.3 (a). The second activity is similarly sliding the arm downward but with an different angle as shown in the Fig.3 (b). In our simulation, the human body is defined as a combination of 14 key points [10], which corresponds to 14 paths according to equation (1). There are 11 static paths and 3 changing paths which are caused by motions of hand, forearm and upper arm in our defined activity. Attenuation coefficients are calculated by the mie scattering equation, since we consider the effect that Wi-Fi signals encounter human body as scattering rather than reflection. By mie scattering equation, the attenuation coefficients depend on signal's frequency, scattering angle and refractive index of human skin which is 1.41. The size of room is defined as 5 meters width, 4 meters length and 3 meters height. The two defined activities happen in the same position. One transmitter and multiple receivers are deployed in the simulation. For each transmitter receiver pair, it contains 7 static paths which include 6 paths caused by wall reflection and 1 line-of-sight path. The strength of signal reflected off the wall is calculated by Fresnel's equation where the real parts of refractive index of concrete wall and air are 2.55 and 1 respectively.

We first calculate the minimum error probability versus different signal to noise ratio (SNR) when the CSI sampling rate are 500Samples/second, 1000S/s and 1500S/s respectively with one transmitter and one receiver. As shown in Fig. 4 (a), minimum error probability of 500S/s is less than 90% with low SNR, which means that whatever HAR deep learning algorithms are applied cannot exceed that performance upper bound. But the increase of CSI sampling rate can improve the HAR performance even with a low SNR. The decrease of noise power would reduce the classification error probability. In the meantime, the increase of CSI sampling rate can improve the classification performance according to our derivation in section III. The reason is that the minimum noise to distance ratio C have more probability on smaller values when we have more CSI samples, which makes the random variable W have same trend as shown in Fig.2 (a). The increase of CSI sampling rate is limited by the receiver's capability which is about 2500S/s. In Fig. 4 (a), the increase of 500S/s CSI sampling rate can reduce the classification error probability by approximately multiplying 0.1.

Secondly, we adjust difference angle α between two activities. As shown in Fig. 4 (b), the classification probability decreases when the difference angle increases. In the other words, human activities are easier to classify when they are more different from each other. According to the equation (17), only the PDF of the minimum ratio C varies with the different factors in wireless sensing networks. Then it influences W and minimum error probability of binary human activity recognition. The euclidean distance $|\Delta G|$ between ground truth data set of human activity 1 and activity 2 is what directly determines the minimum error probability. Since the $|\Delta G_k|$ has many different elements, we calculate the average value of all ground truth pairs' distances to represent the $|\Delta G|$ status in different situations. When difference angle α gets larger, the average value of $|\Delta G|$ become larger. Then the random variable



(a) Increasing the CSI sampling rate (b) Enlarge the difference angle (c) Adding more distributed receivers

Fig. 4: Transmitter, receivers and human are located in the same room

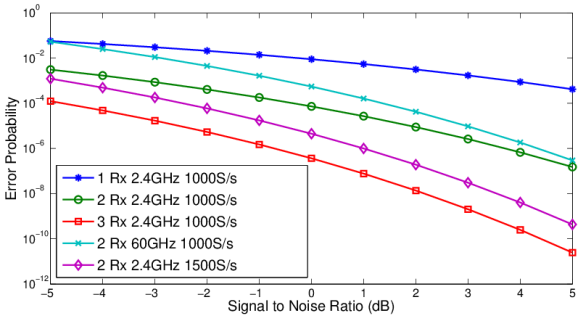


Fig. 5: Through the wall human activity classification

W has more probability on the smaller value as shown in Fig.2 (b), which leads to the smaller error probability.

The classification performances of different center frequencies are influenced by the **phase difference** between $G_1(k)$ and $G_2(k)$ in complex plane. When the phase difference is π, the $|\Delta G_k|$ has the biggest value. When phase difference is 0, $|\Delta G_k|$ has the minimum value. According to the equation (1), there are environment time delay τ_{l_e}, doppler frequency shift $(1+V_{path}/c)$ and the human movements time delay τ_{l_s} influencing the phase of CSI besides the center frequency f_m. Since the phase of CSI value has the periodicity of 2π, the higher center frequency may amplify the phase difference between G_1 and G_2 over multiple 2π and lead to the smaller phase difference compared with lower center frequency. To conclude, different carrier frequencies don't make activities more distinguishable.

Then, the influences of number of receivers and the positions of receivers are shown in Fig. 4 (c). When 2 receivers are used, they are located at both sides of human but not the same side. When 3 receivers are used, the position of human is at the center of the square whose four points are 3 receivers and 1 transmitters. By adding one more distributed receiver, the error probability decreases by approximately multiplying 0.01. The reason is that the CSI measurements taken by one TRP cannot fully describe the human activity. With more TRPs, doppler frequency shifts and time delays in CSI measurements are able to make the human activities more distinguishable, which means the average euclidean distance $|\Delta G|$ gets larger.

Lastly, we simulate the through the wall situation where the transmitter and receivers are deployed behind walls. The classification performances are shown in Fig. 5. In through the wall situation, the classification performances are worse than the performances of indoor environment. The decreases of classification performances are due to the absence of paths that are directly reflected off the human body. It is noteworthy that the classification results of 60 GHz are worse than the 2.4GHz due to the high loss when the signal propagates through the wall. Adding 1 more distributed receiver could still approximately decrease the error probability by multiplying 0.01. 1500S/s CSI samples decrease the error probability by approximately multiplying 0.15 compared with 1000S/s in the situation of 2 distributed receivers. To summarize, adding more distributed receivers is the most efficient way to improve the human activity recognition for both indoor and through the wall situations. The CSI sampling rate can be increased up to about 2500S/s, but the number of receivers are not limited.

V. CONCLUSION

In this paper, we made three key contributions. First, we proposed a sensing model to describe the procedures of CSI based indoor human activity recognition, which includes the impact of environment, influence of human activity and the impact of wireless infrastructures. Second, we derived the human activity classification performance upper bound as the function of all influencing factors in wireless sensing networks. Lastly, we analyzed the numerical results that are obtained from simulation and figured out the influences of factors in wireless networks.

REFERENCES

[1] Fusang Zhang *et al*, "From Fresnel Diffraction Model to Fine-grained Human Respiration Sensing with Commodity Wi-Fi Devices," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, v.2 n.1, p.1-23, March 2018 [doi 10.1145/3191785]

[2] Xiang Li *et al* "IndoTrack: Device-Free Indoor Human Tracking with Commodity Wi-Fi," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, v.1 n.3, p.1-22, September 2017 [doi 10.1145/3130940]

[3] W. Xi et al, "Electronic frog eye: Counting crowd using WiFi," in 2014. DOI: 10.1109/INFOCOM.2014.6847958.

[4] W. Wang et al, "Device-Free Human Activity Recognition Using Commercial WiFi Devices," IEEE Journal on Selected Areas in Communications, vol. 35, (5), pp. 1118-1131, 2017.

[5] M. Kotaru et al, "SpotFi: Decimeter Level Localization Using WiFi," ACM SIGCOMM Computer Communication Review, vol. 45, (5), pp. 269-282, 2015. DOI: 10.1145/2829988.2787487.

[6] K. Ali et al, "Recognizing Keystrokes Using WiFi Devices," IEEE Journal on Selected Areas in Communications, vol. 35, (5), pp. 1175-1190, 2017.

[7] Jiang, Wenjun et al, "Towards Environment Independent Device Free Human Activity Recognition," 289-304. 10.1145/3241539.3241548.

[8] Y. Zeng, P. H. Pathak and P. Mohapatra, "WiWho: WiFi-based person identification in smart spaces," in 2016. DOI: 10.1109/IPSN.2016.7460727.

[9] K. J. R. Liu , Ahmed K. Sadek , Weifeng Su , Andres Kwasinski, "Cooperative Communications and Networking", Cambridge University Press, New York, NY, 2009

[10] Mingmin Zhao et al, 2018, "RF-based 3D skeletons," In Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication (SIGCOMM '18). ACM, New York, NY, USA, 267-281. DOI: https://doi.org/10.1145/3230543.3230579