Counting a Stationary Crowd Using Off-the-Shelf WiFi

Belal Korany
belalkorany@ece.ucsb.edu
UC Santa Barbara
Santa Barbara, CA, USA

Yasamin Mostofi
ymostofi@ece.ucsb.edu
UC Santa Barbara
Santa Barbara, CA, USA

ABSTRACT

In this paper, we are interested in the problem of counting a crowd of stationary people (i.e., seated) using a pair of WiFi transceivers. While the people in the crowd are stationary, i.e. with no major body motion except breathing, people do not stay still for a long period of time and frequently engage in small in-place body motions called fidgets (e.g., adjusting their seating position, crossing their legs, checking their phones, etc). In this paper, we propose that the aggregate natural fidgeting and in-place motions of a stationary crowd carry crucial information on the crowd count. We then mathematically characterize the Probability Distribution Function (PDF) of the crowd fidgeting and silent periods (which we can extract from the received WiFi signal) and show their dependency on the total number of people in the area. In developing our mathematical models, we show how our problem of interest resembles a several-decade-old $M/G/\infty$ queuing theory problem, which allows us to borrow mathematical tools from the literature on $M/G/\infty$ queues. We extensively validate our proposed approach with a total of 47 experiments in four different environments (including through-wall settings), in which up to and including $N = 10$ people are seated. We further test our system in different scenarios, and with different activities, representing various engagement levels of the crowd, such as attending a lecture, watching a movie, and reading. Moreover, we test our proposed system with different number of people seated in several different configurations. Our evaluation results show that our proposed approach achieves a very high counting accuracy, with the estimated number of people being only 0 or 1 off from the true number 96.3% of the time in non-through-wall settings, and 90% of the time in through-wall settings. Our results show the potential of our proposed framework for crowd counting in real-world scenarios.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Mathematics of computing** → **Queueing theory**; • **Computer systems organization** → *Sensor networks.*

KEYWORDS

Occupancy estimation, Crowd Counting, WiFi sensing

MobiSys '21, June 24–July 2, 2021, Virtual, WI, USA

ACM ISBN 978-1-4503-8443-8/21/06.
https://doi.org/10.1145/3458864.3468012

ACM Reference Format:
Belal Korany and Yasamin Mostofi. 2021. Counting a Stationary Crowd Using Off-the-Shelf WiFi. In *The 19th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '21), June 24–July 2, 2021, Virtual, WI, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3458864.3468012

1 INTRODUCTION

Recent years have witnessed a rapid growth in the number of wirelessly-connected devices, resulting in an abundance of Radio Frequency (RF) signals, such as WiFi. In addition to utilizing these signals for communication, researchers have also been investigating using these ubiquitous WiFi signals for sensing, i.e., to extract useful information about the environment, for instance for imaging [5], target tracking [17, 27], health monitoring [1], and other applications.

In particular, occupancy estimation and crowd counting have gained a considerable attention in the RF sensing literature due to their relevance to different applications [6, 37, 39]. For instance, the ability to count the number of people in an area is important for smart buildings, in order to optimize heating, cooling, and lighting. Retail stores can further utilize such a technology to analyze the customers' shopping interests based on their movements and the corresponding crowd analytics [8]. Occupancy estimation can also be very important during a pandemic (such as COVID-19), as it can provide an early warning if the public safety and crowd count limitation guidelines are being violated. Utilizing WiFi signals for crowd counting is in particular appealing, as compared to the traditional counting techniques, such as cameras [22], which require an unobstructed view of the people and further invade privacy, or environmental sensors (e.g., temperature and CO_2 [20, 31]) which require extensive and costly deployments.

In this paper, we are interested in crowd counting when the crowd is *stationary* (i.e. seated), using a pair of off-the-shelf WiFi transceivers. While there has been a great body of recent work on crowd counting using WiFi signals, most have considered counting mobile people, i.e., people have to walk around to be counted [6, 9, 13]. There are, however, many real-world scenarios where it is of interest to count a seated crowd where individuals are not moving around, such as the attendees of an event/seminar or readers in a library. While these events are in principle preceded by crowd motion, e.g. the attendees entering the venue, counters at the entrance (if available) will not be able to estimate the number of people seated in different parts of the area, for instance in different sections of the library. As such, counting the number of people, once seated, becomes important. Furthermore, existing mobile crowd counting methods have their limitations (more detailed review of the state-of-the-art will be provided in Sec. 2). As such, the ability to

count the crowd, once seated, can further complement the current mobile crowd counting methods.

When considering the literature on stationary crowd counting, only very few papers have investigated this problem since it becomes considerably challenging due to the lack of major movements. [3, 32] aim to count stationary people by identifying and counting the individual breathing signals. However, these work require that people stay perfectly still, with no movement, so that their breathing signals become measurable. More importantly, they need to ask different people to artificially breathe with different breathing rates, so that they can be counted. These are, however, unrealistic assumptions that will not be true in real scenarios.

In this paper, we present a new foundation for counting a stationary crowd using a pair of WiFi transceivers. It is worth emphasizing that we are interested in passive crowd counting, i.e., we do not rely on people to carry any device. A case where a number of seated people are engaged in an activity is applicable to many real-world settings, such as the audience of different kinds of social events (e.g. seminars, presentations, lectures), the crowd in a movie theater, or the crowd in a wedding ceremony. It also applies to situations in which each person is separately engaged in an individual task, such as readers in a library. Fig. 1 shows a few real-world examples of a stationary crowd.

Here is our underlying proposed idea. Consider a scenario such as the ones shown in Fig. 1, where a number of people are seated. The people in the crowd are stationary, i.e. with no major body motion except breathing. However, people do not stay still for a long period of time and frequently engage in different kinds of small in-place natural body motions (called fidgets [11]), such as adjusting their seating position, crossing their legs, checking their phones, scratching, stretching, and coughing, among many others. We then propose that the aggregate natural body fidgets of the crowd carry crucial information on the crowd count. Furthermore, we develop a new mathematical characterization for the statistics of the *Crowd Fidgeting Periods (CFPs)*, which we define as the periods in which at least one person in the area is fidgeting, as well as *Crowd Silent Periods (CSPs)*, which we define as the periods in which no one in the area is fidgeting (i.e. everyone is only breathing), and show their dependency on the total number of people in the area. We then demonstrate that a Maximum A Posteriori (MAP) estimator of the number of people can be obtained using these new mathematical models. Our mathematical characterizations are inspired by a 1985 queuing theory paper [28]. More specifically, we show how our problem of mathematically modeling the statistics of the crowd fidgeting and silent periods can be posed similar to an $M/G/\infty$ queuing theory problem. By borrowing and adapting mathematical tools from queuing theory, we can then characterize the aggregate crowd fidgeting dynamics and mathematically relate them to the total number of people. We next explicitly state our contributions.

Statement of Contributions:

1. We propose that the aggregate fidgets of the crowd carry crucial information on the crowd count. We then develop a new mathematical model that describes the collective fidgeting behavior of a stationary crowd, and explicitly relates it to the total number of people. More specifically, we mathematically characterize the distribution of the *Crowd Fidgeting Periods (CFPs)*, which we define as the time durations in which at least one person in the WiFi area is fidgeting, as well as that of the *Crowd Silent Periods (CSPs)*, which we define as the time durations in which none of the people are fidgeting, and show their dependency on the total number of people. Based on this analysis, we then propose a Maximum A Posteriori (MAP) estimator of the number of people, using our derived mathematical models of CFPs and CSPs.

Figure 1: Sample application scenarios of our stationary crowd counting system: (a) a group of people reading, (b) a group of people watching a movie, (c) attendance of a presentation, and (d) attendance of a religious event.

2. In developing our mathematical models, we reveal how our problem of interest resembles a several-decade-old $M/G/\infty$ queuing theory problem. More specifically, we show in details how the CSPs are similar to the times when no customer is at a queue that has infinite servers, while the CFPs resemble the times when at least one customer is being served at such a queue. We then explicitly characterize the similarity between the two problems, which allows us to borrow mathematical tools from a 1985 paper on $M/G/\infty$ queues.

3. We develop a framework for processing the WiFi Channel State Information (CSI) measured at a WiFi receiver, in order to extract the times in which at least one person in the coverage area is fidgeting (i.e. CFPs). More specifically, we show how the received WiFi CSI bandwidth can be used to indicate whether at least one person in the WiFi area is fidgeting.

4. We extensively validate our proposed approach with a total of 47 experiments in four different environments (including through-wall settings), in which up to and including $N = 10$ people are seated and behaving normally, while a pair of WiFi transceivers, located on one side of the area, make CSI measurements. In total, 27 different subjects participated in our experiments.[1] We test our system in different scenarios representing various engagement levels of the crowd, such as attending a lecture/presentation, watching a movie, and reading. We further test our proposed system with different number of people seated in several different configurations. Our evaluation results show that our proposed approach achieves a very

[1]The Institutional Review Board (IRB) committee has reviewed/approved this research and determined that it does not constitute human subject research. Furthermore, our experiments were approved by our institution from COVID-19 safety point of view, and all the precautionary measures set by our institution (such as social distancing for people not in the same network) were enforced.

high counting accuracy, with the estimated number of people being only 0 or 1 person off from the true number of people 96.3% of the time in non-through-wall settings, and 90% of the time in through-wall settings. Overall, our results show the great potential of our proposed framework for crowd counting in real-world scenarios.

2 RELATED WORK

In this section, we summarize the state-of-the-art in passive crowd counting using WiFi signals.

2.1 Counting a Mobile Crowd

A number of recent papers have tackled the problem of counting a mobile crowd where each person is moving. In these papers, it is assumed that every person in the group is walking in the WiFi coverage area. As such, they consider a problem that is different from the problem of this paper, i.e., counting a stationary seated crowd. Nevertheless, for the sake of completion, we shall review this category of work next. These counting methods can be broadly divided into two categories: model-based methods [6, 7, 37], and learning-based methods [9, 13, 15, 24, 39]. In model-based methods, the counting is based on a mathematical modeling of the received WiFi signal based on people's motion. These techniques require the transmitter and receiver to be on the opposite sides of the area and further require some prior knowledge of the motion dynamics of the people for counting. For instance, they have been utilized for counting in an aisle-type setting [37] or in a setting where people walk casually and change their directions randomly [6]. However, new work is needed to extend these theories to other settings. On the other hand, learning-based methods rely on training a neural network with several raw WiFi signals (or manually-extracted features thereof) collected when different number of people are present in the WiFi area. Thus, they can only be used in the area/configurations that they were trained on. Overall, the methods developed for counting a mobile crowd are not applicable to our problem of counting a stationary crowd. It is worth noting that while some entrances may have door counters, door counters can not count well when there are simultaneous crosses. For instance, we added a WiFi-based entrance monitoring link to one of the entrances of one of our experimental areas (area of Fig. 10 (a)). However, the link could not correctly count the number of people beyond 2 when there were simultaneous crosses, which is as expected. Door counters may not also be applicable to large/semi-open entrances and further cannot estimate the number of people seated in different parts of the area after entering, even when they count correctly at the entrance.

In summary, our proposed method of counting stationary people can serve as a complement to the existing literature that mainly focuses on counting a mobile crowd.

2.2 Counting a Stationary Crowd

Due to its difficulty, stationary crowd counting has not gained as much attention in the literature. A couple of approaches have been proposed to utilize breathing signals in order to count the people [3, 32]. Breathing-based approaches, however, suffer from two main drawbacks. First, the crowd is required to stay still, without any motion, for an extended period of time in order to measure their breathing signals, which is an overly restrictive and impractical assumption. Second, breathing-based methods rely on different people to have different breathing rates in order to differentiate and count them, greatly limiting the total number of people they can count. For instance, the participants in [3] were told to breathe with a predefined and different set of breathing rates, with some of them asked to breathe up to 31.5 breaths per minute (bpm), which is an abnormally high breathing rate (normal breathing rates for adults range between 12 and 20 bpm). Without enforcing a predefined set of breathing rates, [32] could count up to only 4 people.

The authors of [4] trained a deep neural network for counting stationary people. However, they had to collect extensive training data pertaining to several different number of people, seated in various configurations in the test area. Furthermore, such learning-based approaches can only count the number of people that they are trained on, and with the people in the same seating configuration as in training, and in the same environment, and their performance degrades considerably when deployed with unseen configurations, even in the same environment [4].

3 A MATHEMATICAL MODEL FOR CROWD FIDGETING

In this section, we develop a new mathematical model to describe the statistics of the fidgets of a group of stationary people, and show how they relate to the total number of people. Consider a scenario such as the ones shown in Fig. 1, where N people are seated in a WiFi-covered area. The people in the crowd are stationary, i.e. staying still with no major body motion except breathing. However, as discussed earlier, people do not stay still for a long period and frequently engage in different kinds of small in-place motions, called fidgets.

Let $t = 0$ denote the start of the measurement time. Let $t_i^{\{n\}}$ denote the start of the i-th fidget of the n-th person, for $n = 1, 2, \ldots, N$. The n-th person then fidgets for a duration of $d_i^{\{n\}}$ before returning to the state of being stationary, i.e. $d_i^{\{n\}}$ is the time duration of the i-th fidget of the n-th person. $d_i^{\{n\}}$ can be modeled as an independent and identically distributed (i.i.d) random variable, with a Probability Distribution Function (PDF) $p_D(d)$. Let $T_i^{\{n\}}$ be the time between the start of the i-th fidget of the n-th person and the start of his/her $(i+1)$-th fidget (i.e. $T_i^{\{n\}} = t_{i+1}^{\{n\}} - t_i^{\{n\}}$). Let $p_n(T)$ denote its distribution. Fig. 2 presents a visual demonstration of the aforementioned quantities.

Similar to many processes in different scientific fields, the natural fidgeting process of an individual is well modeled by a Poisson process [21].[2] Therefore, the inter-fidget times of the n-th person, $T_i^{\{n\}}$, will have an exponential distribution, i.e. $p_n(T_i^{\{n\}} = T) = \frac{1}{\gamma_n} e^{-T/\gamma_n}$, where γ_n is the average inter-fidget time of the n-th person. Note that we do not assume or require that all people fidget at exactly the same rate, i.e. γ_n's can be different. Instead, we take γ_n to be a random variable taken from a distribution $p_\Gamma(\gamma)$. We shall discuss $p_\Gamma(\gamma)$ and $p_D(d)$ in more detail in Sec. 5.

It is well-known that the superposition of N different Poisson processes results in a Poisson process whose rate is the sum of

[2]The validity of the assumption that the individual fidgeting process can be modeled as a Poisson process will be discussed in more detail in Sec. 7.

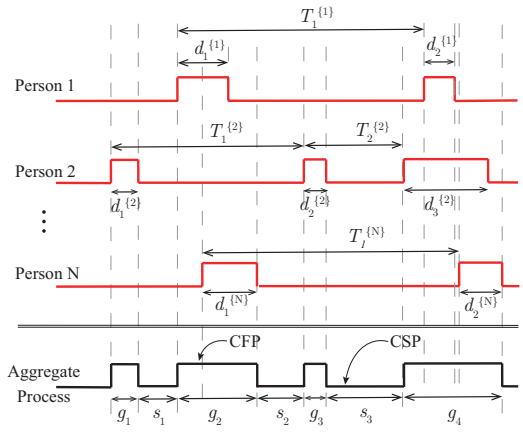


Figure 2: A sample fidgeting timeline of N people in an area, where a non-zero signal indicates a fidget. $d_i^{\{n\}}$ denotes the duration of the i-th fidget of the n-th person, while $T_i^{\{n\}}$ represents the time between the i-th and $(i+1)$-th fidgets of the n-th person. An aggregate fidgeting process results from the superposition of the individual fidgeting processes, where the state of no one fidgeting is referred to as the *Crowd Silent Period (CSP)* (with a duration s), while the state of at least one person fidgeting is referred to as the *Crowd Fidgeting Period (CFP)* (with a duration g). A sample CFP and CSP are marked on the figure.

the rates of the individual processes [23]. Therefore, to analyze the collective fidgeting behavior of the crowd, denote by $T^{[c]}$ the inter-fidget time of the overall crowd fidgeting process (i.e., the time between the start of any two consecutive fidgets happening from any individuals in the crowd). $T^{[c]}$ then follows the exponential distribution:

$$p_{T^{[c]}}(T) = \left(\sum_{n=1}^{N} \frac{1}{\gamma_n}\right) e^{-T\left(\sum_{n=1}^{N} \frac{1}{\gamma_n}\right)}. \tag{1}$$

It should be noted that, since multiple people can fidget at the same time and each fidget has a non-zero duration, we cannot directly measure $T^{[c]}$ from the aggregate crowd fidgeting process. We next formally introduce two key parameters related to the overall aggregate fidgeting process of the crowd, which are measurable:

(1) **Crowd Fidgeting Period (CFP):** we define a CFP as the state of having *one or more* individuals in the crowd engaged in a fidgeting movement. The duration of the i-th such period is denoted by g_i, and has a distribution $p_G(g)$.
(2) **Crowd Silent Period (CSP):** we define a CSP as a continuous stretch of time where *no person* is engaged in any kind of fidgeting movement, i.e. all people in the crowd are only breathing. The duration of the i-th such period is denoted by s_i, and has a distribution $p_S(s)$.

Examples of CFPs and CSPs are shown in the aggregate process of Fig. 2. Intuitively, the PDFs $p_G(g)$ and $p_S(s)$ depend on the number of people N. For instance, when N increases, CFPs tend to get longer, due to the higher probability of overlap between fidgets of individual people. Moreover, increasing N also results in a shorter CSP. Next, we show how to mathematically characterize $p_G(g)$ and $p_S(s)$ as a function of N.

M/G/∞ Queues

Queuing theory is a branch of mathematics that studies waiting lines in systems that involve the arrival of entities (referred to as customers), which require a service from an entity that includes a number of servers [16]. Considerable research has been conducted in this area to characterize several quantities related to this problem, such as the statistics of the waiting time of the customers, the average number of customers waiting at any time instant, and other related parameters. A problem in queuing theory is traditionally denoted by a triad of symbols A/B/C, where symbol A describes the arrival process of customers, symbol B describes the distribution of the service time of a customer (the time a customer spends being served by one of the servers before leaving), and C describes the number of servers. A queuing model that is of particular interest to us (as we shall explain shortly) is an $M/G/\infty$ queue. In this queuing model, the arrival process is assumed **M**arkovian, e.g. can be modeled as a Poisson process, resulting in the inter-arrival times following an exponential distribution, the service times can follow any **G**eneric distribution, and the number of servers is ∞. An immediate characteristic of such a queue is that there are no waiting lines, since any arrival can be instantaneously served by one of the infinite servers. In 1985, ref. [28] studied two important quantities relevant to an $M/G/\infty$ queue: the busy and idle periods. A busy period is defined as a period in which there is at least one customer being served in the system, while an idle period is a period in which no customer is being served (i.e., all servers are idle).

While assuming an infinite number of servers may not be of much practical interest to most queuing problems, it is of great interest to our fidget-based crowd counting problem. More specifically, a careful inspection of the structure of the $M/G/\infty$ queue shows a strong analogy to our stationary crowd counting model. Consider the arrival process corresponding to the aggregate process of N people. As discussed earlier, this is a Poisson process, with the inter-event times characterized by an exponential distribution as shown in Eq. 1. A CSP in our crowd counting model is then analogous to an idle period in the $M/G/\infty$ queue. A person starting a fidget is analogous to a customer arriving at a server station, while the fidget duration is analogous to the service time of the person. Note that the overlap of the fidgets from people is similar to the case where multiple customers are being served at the same time. Since there is no delay in fidgeting, the queue model with infinite servers then well models our case.

As such, we can tap into the literature on $M/G/\infty$ queues to mathematically characterize our CFPs and CSPs. More specifically, it can be shown that given a crowd of N people, due to the memoryless property of the exponential distribution, the duration of a CSP follows an exponential distribution whose rate is the sum of the rates of the individual people [28]. In other words, $p_S(s)$ can be written as follows:

$$p_S(s|N,\bar{\gamma}) = \frac{N}{\bar{\gamma}} e^{-\frac{sN}{\bar{\gamma}}}, \tag{2}$$

where $\bar{\gamma}^{-1} = \frac{1}{N}\sum_{n=1}^{N} \gamma_n^{-1}$ is the average of the fidgeting rates of people.

Additionally, it can also be shown that, given a crowd of N people, the PDF of the duration of a CFP, $p_G(g)$, is characterized as follows [28]:

$$p_G(g|N,\bar{\gamma}) = \frac{d}{dg}\left[-N\bar{\gamma}^{-1}\sum_{i=1}^{\infty} m^{*i}(g|N)\right], \tag{3}$$

where m^{*i} is the i-fold convolution of m with itself, and

$$m(g|N) = \frac{d}{dg}\left[-\exp\left\{-\frac{\bar{\gamma}}{N}\int_0^g (1-P_D(x))dx\right\}\right], \tag{4}$$

where $P_D(x) = \int_0^x p_D(u)du$ is the Cumulative Distribution Function (CDF) of an individual's fidget duration.

MAP estimation of N

The previous analysis shows how the distributions of the crowd fidgeting period, $p_G(g)$, and the crowd silent period, $p_S(s)$, are dependent on the number of people N. We then utilize this mathematical characterization and propose a Maximum A Posteriori (MAP) estimation rule to estimate the number of people N at time t, given the durations of all the CFPs and CSPs prior to t. We will show how these fidgeting/silent periods can be extracted from the ambient WiFi signals in Sec. 4. More specifically, let $g_1, g_2, \ldots, g_{N_f(t)}$ be the durations of all the CFPs before time t, where $N_f(t)$ is the total number of these fidgeting periods up to time t. Similarly, let $s_1, s_2, \ldots, s_{N_s(t)}$ be the durations of CSPs before time t, where $N_s(t)$ is the total number of such silent periods. The MAP estimation rule for the number of people can be written as [29],

$$\begin{aligned}\hat{N}(t), \hat{\bar{\gamma}} &= \arg\max_{N,\bar{\gamma}}\ p(N,\bar{\gamma}|g_1,...,g_{N_f(t)},s_1,...,s_{N_s(t)}) \\ &= \arg\max_{N,\bar{\gamma}}\ p(g_1,...,g_{N_f(t)},s_1,...,s_{N_s(t)}|N,\bar{\gamma})p(N)p(\bar{\gamma}) \\ &= \arg\max_{N,\bar{\gamma}}\ p_\Gamma(\bar{\gamma})\prod_{i=1}^{N_f(t)} p_G(g_i|N,\bar{\gamma})\prod_{j=1}^{N_s(t)} p_S(s_j|N,\bar{\gamma}),\end{aligned} \tag{5}$$

where $p(.)$ denotes the probability of the argument. The last step follows from the independence of the durations of the CFPs and CSPs, and the fact that we do not assume any prior knowledge on the number of people N, i.e. $p(N)$ is taken as uniform. We also use the general PDF of p_Γ for the prior distribution on $\bar{\gamma}$. In Sec. 5, we show how to obtain the priors p_Γ and p_D. As we shall see, we do not need to make any prior WiFi measurements to estimate these. In summary, Eq. 5 allows us to estimate the total number of stationary people based on their natural fidgets.

4 WIFI PROCESSING PIPELINE

In Sec. 3, we have shown that one can estimate the number of stationary people in an area, given the durations of the CFPs ($g_1, g_2, \ldots, g_{N_f}$) as well as those of CSPs ($s_1, s_2, \ldots, s_{N_s}$). In this section, we show how to extract these periods from the ambient WiFi signal in the area of interest.

Consider the scenarios shown in Fig. 1, where a WiFi transmitter (Tx) transmits WiFi signals that are reflected off of the bodies of the N people in the area, after which they are received by a WiFi receiver (Rx). Let $c(t)$ denote the complex baseband received signal at the Rx, as a function of time. It has been shown, in the RF sensing literature, that the frequency content/bandwidth of the received WiFi signal increases when the speed of the moving person/object increases [18, 19, 26, 30, 35]. We make use of this fact to extract the CFPs and CSPs from the WiFi signals as follows.

When a stationary person is not fidgeting, the only body movement is the slow sinusoidal breathing motion of the chest and abdomen. Since the speed of the body motion during fidgeting is typically considerably higher than the speed during breathing, we expect the frequency content of the measured WiFi signal during fidgeting to be considerably higher than that of normal breathing, with a high probability. More specifically, it has been shown in the literature that the maximum chest displacement of a person during respiration is about 5 mm [33]. Considering that the maximum normal breathing rate of adults is $f_{\text{br}} = 0.3$ Hz [25], this chest displacement translates to a maximum instantaneous chest speed of 0.01 m/s. On the other hand, when a person is engaged in any kind of non-breathing in-place motion (e.g. fidgeting), the instantaneous speed of the body parts can increase significantly. For instance, the authors of [2] attached accelerometers to the wrists of 20 subjects to analyze their motion while doing various tasks, one of which was sitting down and relaxing, and published the acceleration data in the PhysioNet database [12]. We studied this dataset and calculated the speed of body motion during the relaxing periods from their published acceleration data. We found that, during fidgets, the speed of the body motion is larger than 0.01 m/s (which is the maximum body speed during breathing) 90% of the time. Furthermore, the speed of the body motion during fidgets can take much larger values, e.g. 5 times the maximum body speed during breathing (or larger) for 80% of the time. It can even reach speeds as high as 3 m/s (i.e. 300 times larger). As such, with a high probability, the overall received WiFi magnitude or phase difference signal during a fidget will have a considerably higher frequency content than the one when only breathing. Thus, we can easily extract the fidget-related content by properly high-pass filtering the received WiFi signal (phase difference or magnitude) above B_{br}, where B_{br} is the maximum bandwidth of the received WiFi signal when only breathing. See [19] for an exact characterization of the bandwidth during breathing as well as during normal body movements. We shall use the corresponding derivations of [19] in Sec. 5 when characterizing B_{br}. As we shall see, B_{br} is relatively small at WiFi frequencies since the breathing rate of adults is around 0.3 Hz.

Fig. 3 shows our end-to-end proposed pipeline, including the pre-processing, fidget detection, and estimation steps. More specifically, we first feed the WiFi CSI phase difference between the Rx antennas to a pre-processing module. We utilize only the CSI phase difference data since it has been shown to be more stable and robust in different deployment environments than the CSI magnitude data [36].[3] The pre-processing module extracts only a subset of the WiFi subcarriers that are less noisy, and further denoises these subcarriers by means of Principal Component Analysis (PCA), as we shall describe in Sec. 4.1. The denoised data is then fed to the CFP detection module, which decides whether there is any fidgeting or not based on the spectral content of the denoised data, as we shall describe in Sec. 4.2. The resulting fidgeting sequence is then used for counting, based on

[3] We note, however, that the CSI magnitude data follows the same model and can similarly be used, if stable enough, for fidget detection.

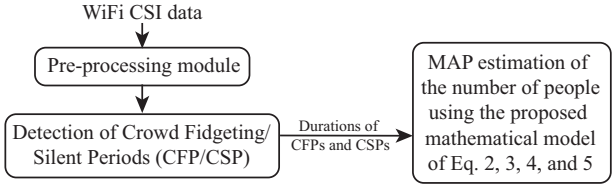


Figure 3: High-level overview of the proposed pipeline.

the proposed mathematical framework of Sec. 3. We next describe these steps in more detail.

4.1 Pre-processing Module

The first step in our proposed WiFi processing framework is to extract the phase difference data between the Rx antennas. More specifically, off-the-shelf WiFi Network Interface Cards (NICs), such as the Intel 5300 NIC, provide the CSI measurements on three Rx antennas, for a set of $N_{\text{sub}} = 30$ subcarriers. The absolute phase measurements on each of the Rx antennas are usually corrupted by multiple sources of errors, such as the carrier frequency offsets and sampling time offsets [38], rendering the absolute phase measurements unreliable. However, since the Rx antennas share the same oscillator, the phase difference between the Rx antennas is stable, experiences little noise, and is robust to different deployment environments [36]. As such, we extract the phase difference data between the Rx antennas $z_{2,1,k}(t)$ and $z_{3,1,k}(t)$, where $z_{i,j,k}(t) = \angle c_i(t;k) - \angle c_j(t;k)$ is the phase difference between antenna i and antenna j on the k-th subcarrier, resulting in a total of $2N_{\text{sub}} = 60$ streams of data to be utilized for fidget detection. We then remove the DC component from each data stream since the DC term only has the impact of the static objects in the environment [18], and is not relevant to our analysis, which relies on the impact of the fidgets.

Stream Selection: Different subcarriers on the same Rx have different carrier frequencies (or wavelengths), and consequently, experience different levels of noise depending on the environment, making some sub-carriers noisier than others. It is then important to select only the least noisy data streams and subsequently use them in the rest of the processing pipeline. In order to do so, we utilize the measured phase difference data in a short (e.g., less than 30 seconds) one-time calibration phase prior to the real experiments. In this prior calibration phase, WiFi measurements are collected while a number of people are seated in the area and are only breathing, without any fidgeting, for a short period of time. Note that the number of people in the calibration phase is decoupled from the number of people during the real experiments and, as such, can be as low as needed. Furthermore, people can sit in any configuration. If the number of people in the calibration phase is small, e.g., 1 or 2, and the area is large, we find it better if they sit in a couple of different random configurations to pull the data of them together and find the best streams. For instance, 1 person can sit in two different locations and 2 people can sit in random locations, amounting to three 10-second data collection periods in which the subjects are only breathing. Consequently, the spectral content of the data streams in this calibration period should be confined to the band $[0, B_{\text{br}}]$ Hz. As such, we can calculate the Signal-to-Noise Ratio (SNR) for all the $2N_{\text{sub}}$ data streams as the ratio of the spectral energy content below B_{br} to the spectral energy content above B_{br}. More specifically, the SNR of the phase difference between the i-th and the j-th Rx antennas at the k-th subcarrier is computed as follows, in the calibration phase when there is only breathing:

$$SNR_{i,j,k} = \frac{\int_0^{B_{\text{br}}} \left| \int z_{i,j,k}(t) e^{-j2\pi f t} dt \right|^2 df}{\int_{B_{\text{br}}}^{\infty} \left| \int z_{i,j,k}(t) e^{-j2\pi f t} dt \right|^2 df} \tag{6}$$

where $z_{i,j,k}(t)$ is the phase difference between the i-th and the j-th Rx antennas at the k-th subcarrier, as a function of time, and B_{br} is the maximum bandwidth of the WiFi signal during breathing. Note that, in the calibration phase (i.e. only breathing), the numerator of Eq. 6 contains the reflected signals, while the denominator contains only the measurement noise. We then select the top 10 data streams in terms of their SNR and use only these 10 data streams in the rest of the operation phase.

PCA denoising: After extracting the top 10 data streams (in terms of the SNR), we extract the first principal component, $w(t)$, of these data streams using Principal Component Analysis [34]. It has been shown in the literature that the first principal component contains the motion information of the moving subjects, while noise is distributed among different principal components [34]. As such, $w(t)$ serves as the denoised WiFi data that contains the motion information of the crowd.

4.2 Crowd Fidgeting Period (CFP) Detection

As described earlier, during CSPs, all people in the WiFi area are only breathing. Hence, the spectral content of the denoised WiFi data, $w(t)$, in CSPs is concentrated below B_{br}. On the other hand, during CFPs, at least one person in the WiFi area is fidgeting, resulting in the spectral content of $w(t)$ to span a wider band, e.g. $w(t)$ would have a bandwidth higher than B_{br}. We utilize this observation to detect the CFPs as follows. We first filter the WiFi data $w(t)$ by passing it through a high-pass filter with a cut-off frequency of B_{br}. The filtered signal $w^{\text{filt}}(t)$ would contain only noise during CSPs, while during CFPs, it contains motion data as well. Let $\mathcal{H}_1$ denote the hypothesis that there is at least one person fidgeting in the WiFi area, while $\mathcal{H}_o$ denotes otherwise. Accordingly, we detect whether there is any fidgeting at time τ by thresholding the moving variance of $w^{\text{filt}}(t)$ as follows,

$$\text{VAR}_\tau \left\{ w^{\text{filt}}(t) \right\} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \sigma_{\text{th}}, \tag{7}$$

where $\text{VAR}_\tau\{.\}$ is the variance of the signal in a window of length T_{win} ending at time $t = \tau$, and σ_{th} is a threshold representing the noise floor. In order to determine the value of σ_{th}, we use the denoised WiFi signal during the calibration period, to which we refer as $w_{\text{cal}}(t)$. Since the spectral content of $w_{\text{cal}}(t)$ is concentrated below B_{br}, the filtered signal, $w_{\text{cal}}^{\text{filt}}(t)$, contains only noise. As such, we estimate the noise floor σ_{th} as follows,

$$\sigma_{\text{th}} = \max_\tau \text{VAR}_\tau \left\{ w_{\text{cal}}^{\text{filt}}(t) \right\}. \tag{8}$$

In order to show the performance of our proposed WiFi preprocessing and fidget detection modules, we conduct a 3-minute

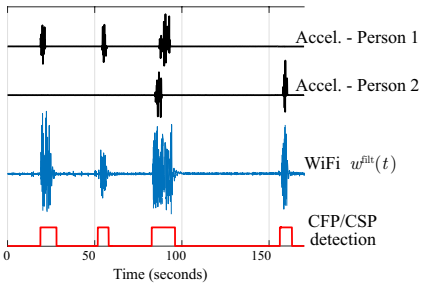


Figure 4: An example of CFP/CSP detection: Two people sit together to watch a movie while accelerometers are attached to their arms. The figure shows (from top to bottom) the accelerometer data of person 1, the accelerometer data of person 2, the filtered WiFi signal $w^{\text{filt}}(t)$, and the binary output of the CFP detection module. Note that people can have overlapping fidgets, an example of which is also shown in the figure. Our proposed pipeline can accurately detect and localize the CFPs/CSPs, including the ones which contain overlapping fidgets from multiple people.

experiment where two people sit together to watch a movie. In order to get ground-truth fidgeting data, we attach a smartphone to the upper right arm of each of the two people and log the accelerometer data of the smartphones. Fig. 4 shows the logged accelerometer data of the two people, as well as the high-pass filtered WiFi signal $w^{\text{filt}}(t)$ extracted during the 5-minute experiment. The data in a prior 10-second calibration phase is used to extract the set of best data streams, and to estimate the value of the noise floor. It can be seen that $w^{\text{filt}}(t)$ experiences high-amplitude variations whenever any person engages in a fidgeting movement, as confirmed by the accelerometer data. The figure also plots the binary output of our CFP detection module, where the CFPs are accurately captured based on the moving variance of $w^{\text{filt}}(t)$.

4.3 Estimation of the number of people

Once we have detected the CFPs and CSPs, we use the proposed mathematical framework of Sec. 3 to estimate the total number of people, as shown in Fig. 3. More specifically, to estimate the number of people at time t, we extract all the durations of the CFPs (hypothesis $\mathcal{H}_1$) as well as the durations of the CSPs (hypothesis $\mathcal{H}_o$) that occurred prior to time t. Let $g_1, g_2, \ldots, g_{N_f(t)}$ denote the durations of all the detected CFPs before time t, while $s_1, s_2, \ldots, s_{N_s(t)}$ denote the durations of the CSPs before time t. Then, given the prior PDFs of a general individual's fidget duration, p_D, and average inter-fidget time, p_Γ, we can estimate the number of people using Eq. 2–5. In Sec. 5, we show how to obtain the prior PDFs of an individual, p_Γ and p_D. As time progresses, new outputs of the CFP detection module can be used to update the list of CFP and CSP durations, and consequently update the estimate of the number of people.

Let $\hat{N}(t)$ denote the estimate of the number of people at time t. As we collect more measurements, $\hat{N}(t)$ will start to converge to the true value. However, it may still oscillate a bit around the true value. As such, we employ a moving average filter to smooth out $\hat{N}(t)$. In other words, our final estimate for the number of people at time t is the mean of $\hat{N}(t)$ in a window of length T_{avg} seconds prior to time t.

5 EXPERIMENTAL SETUP

In this section, we describe the experimental setup we use to validate our proposed framework for counting a stationary crowd.

Experiment Details: For the WiFi Tx and Rx, we use two laptops equipped with Intel 5300 WiFi NICs. The Tx laptop broadcasts WiFi packets on channel 36 (whose center frequency is 5.18 GHz), with a rate of 50 packets per second, using one Tx antenna. The Rx laptop monitors the channel and logs the CSI data of the received packets on the 3 antennas of the Rx using CSItool [14]. The logged CSI data is then processed according to the proposed pipeline of Sec. 4 using MATLAB. More specifically, we extract the phase difference data with respect to antenna 1 of the Rx (i.e. $z_{2,1}$ and $z_{3,1}$). The phase difference data is then pre-processed and the CFPs/CSPs are extracted using the proposed fidget detection module of Sec. 4.2. Furthermore, we have $B_{\text{br}} = 2f_{\text{br}} = 0.6$ Hz for a carrier frequency of 5.18 GHz, where f_{br} is the breathing rate (see [19] for more detailed derivations), $T_{\text{win}} = 2$ sec, and $T_{\text{avg}} = 90$ sec. Finally, the number of people is counted using the proposed mathematical model of Sec. 3, and is updated once every second.

Experiment Protocol: We carry out counting experiments in four different environments, including both through-wall and non-through-wall settings. In each of the test environments, several experiments with different number of people, different seating configurations, and different activities are conducted. Each experiment lasts for 5 minutes (the estimator of the number of people typically converges much faster, as we shall see). In each experiment, a set of N subjects are asked to sit in the test area in rows of chairs, and to engage in some activity, such as watching a movie, attending a lecture, or reading. For instance, Fig. 7 (a) shows two sample shots of experiments where 9 subjects sit in 3 rows of chairs watching a movie on TV, while Fig. 8 (a) shows an experiment where 4 people sitting in two rows of chairs are watching an online lecture. People were told to act casually and normally during the experiment and just engage in the activity as they normally would. In each experimental area, a prior 10-30 second data is collected for calibration purposes (see Sec. 4.1 for more details on this). In total, we have conducted 47 experiments across the four environments.

Prior PDFs for p_Γ and p_D: The proposed crowd counting mathematical method introduced in Sec. 3 utilizes the PDF of an individual's average inter-fidget time (p_Γ) as well as the PDF of an individual's fidget duration (p_D), in order to calculate the distributions of the CFPs (p_G), and the distribution of the CSPs (p_S) of the aggregate process, as described by Eq. 2 and Eq. 3. These prior distributions describe the behavior of a general individual and are not specific to a certain group of people. They are also independent of the number of people in the crowd. We can easily get such prior fidget data of a single person from online available videos of relevant stationary activities. In this manner, we do not need to collect any prior WiFi measurements to estimate these priors. In general, similar events (e.g., attending a lecture and listening to an officiant at a wedding ceremony) will have similar priors and Youtube videos of them can be lumped together, or used interchangeably, to generate the prior PDFs. However, if events are fairly different, in terms of the required attention span, it will be more accurate to acquire different priors for them, as opposed to lumping

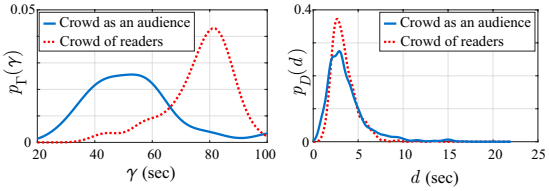


Figure 5: (Left) The distribution of the average inter-fidget time of an individual, $p_\Gamma(\gamma)$, for a general population, collected from the data of a total of 30 individuals attending various events in 14 public online Youtube videos (for the case of audience), and 24 individuals reading in 24 different online YouTube videos (for the case of readers). It can be seen that people tend to fidget less when reading, since it requires more focus. (Right) The distribution of the fidget duration $p_D(d)$ of an individual, estimated from the same dataset.

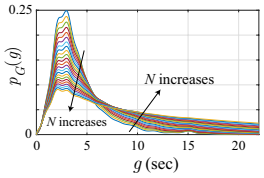


Figure 6: PDF of the duration of a CFP of the aggregate process ($p_G(g)$), from Eq. 3, as a function of N and for an audience-style activity. The figure explicitly shows the dependence of the duration of a CFP on the number of people, which is characterized through our proposed mathematical framework. Higher N results in a higher probability of having longer CFPs, as can be seen.

all such videos together. For instance, people tend to fidget less frequently when engaged in activities that require more attention, such as reading, as compared to watching a movie. Given that the general high-level type of an event would be known for a given area and given the abundance of online videos pertaining to many different activities, acquiring these prior PDFs from online videos would thus be straightforward.

In order to cover a variety of activities, in this paper we then consider two main broad categories of activities: audience-style activities and reading-related activities. We then find the prior PDFs of an individual's fidget duration (p_D) and average inter-fidget time (p_Γ) by utilizing relevant online videos for both activities and logging the time-stamps of fidget start and end times of an individual in the video. We next elaborate in more details.

• Crowd as an audience: this category covers a wide spectrum of real-world social gatherings, where the attention of a group of people is focused on one main source of interest. Examples of this category include the audience of a lecture, presentation, or a seminar, the audience in a cinema/theater or a home-movie setting, or the audience of a wedding ceremony. For this category, we have collected 14 public online Youtube video of different lectures/presentations, with an average video length of 30 minutes, and extracted individual fidgets of a total of 30 individuals.[4]

• Crowd of readers: this category covers any setting (such as a library), where each individual in a group of people is reading. For this category, we have collected 24 public online Youtube videos of people reading (under the search keyword "Read With Me"),[5] and extracted fidget data of a total of 24 individuals. Note that in addition to the examples of fidgets we already mentioned (such as pose adjustments, checking cell phones, etc), flipping the book's pages, writing a note on the book, or other similar interaction with the reading material, can also be considered as a fidget in this case, and will be registered by the WiFi processing framework as a non-breathing motion.

Fig. 5 (left) shows the PDF of the average inter-fidget time of an individual, $p_\Gamma(\gamma)$, in the two aforementioned categories of activities. It can be seen that people tend to fidget less while reading, since the reading activity requires more focus. On the other hand, people tend to fidget once every 50-55 seconds, on average, when attending an event. This result agrees with the findings of Sir. Francis Galton, a famous psychologist in the 19-th century, who observed 50 members of the audience of a public lecture in 1885, and concluded that a person, on average, fidgets once a minute in such gatherings [11]. It also agrees with the findings of the authors of [10], who observed 21 students in a 40-minute lecture, finding that a student, on average, fidgets once every ~57 seconds. Fig. 5 (right) then shows the distribution of the fidget duration, $p_D(d)$, of an individual for both categories of activities. The figure suggests that the distribution of the fidget duration is similar for different types of activities, with an average fidget duration of ~2.9 seconds.

In order to see how the PDF of CFP (p_G), derived in Eq. 3, changes as a function of N, we use the priors of the audience category and plot p_G as a function of N in Fig. 6. As can be seen, when N increases, p_G experiences a longer tail, i.e. it is more probable to have longer CFPs, since it is more probable for the fidgets of several people to overlap. Fig. 6 further confirms that the WiFi measurements carry the information of the total number of people. Our proposed mathematical framework of Sec. 3 then explicitly characterizes this dependency, enabling us to design an end-to-end system to estimate N.

6 EXPERIMENTAL RESULTS

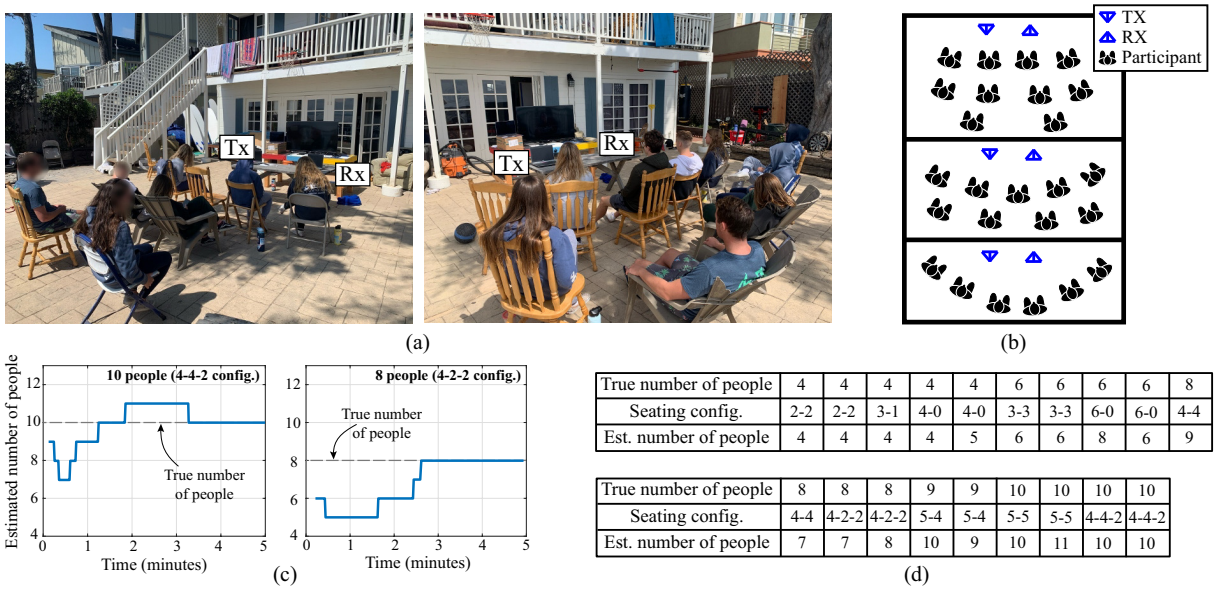
In this section, we present the results of our fidget-based crowd counting framework, using WiFi signals, in different environments and with people engaged in different kinds of activities. We start by showing the counting results in non-through-wall environments, in which the Tx and Rx are in the same area as the crowd. Then, we show the counting results in through-wall settings, where the Tx/Rx are placed behind a wall. We further show results in different scenarios representing various engagement levels of the crowd, such as attending a lecture, watching a movie, and reading.

6.1 Counting in non-through-wall settings

For non-through-wall cases, the Tx and Rx are placed in the same area as the crowd to be counted, with no immediate physical obstruction (such as a wall). However, we emphasize that not all the people in the crowd will have a Line-of-Sight (LoS) to the WiFi link, since they can be blocked by other people/objects. Fig. 7 (a) shows two examples of such non-through-wall settings. Overall,

[4]Sample videos we have used for this category can be found at: youtu.be/r_w7pfulsn8 and youtu.be/nfrmH65kS-E.

[5]Sample videos we have used for this category can be found at: youtu.be/ldVp2fJquq4 and youtu.be/GoCl7D4asl4.



True number of people	4	4	4	4	4	6	6	6	6	8
Seating config.	2-2	2-2	3-1	4-0	4-0	3-3	3-3	6-0	6-0	4-4
Est. number of people	4	4	4	4	5	6	6	8	6	9

True number of people	8	8	8	9	9	10	10	10	10
Seating config.	4-4	4-2-2	4-2-2	5-4	5-4	5-5	5-5	4-4-2	4-4-2
Est. number of people	7	7	8	10	9	10	11	10	10

(d)

Figure 7: (a) Two sample experiments in the first test environment (Area 1): an outdoor patio with several sources of clutter, where up to and including 10 people gather to watch a movie. (b) Seating configuration for three sample experiments in Area 1, where (top) 10 people sit in a 4-4-2 configuration, (middle) 9 people sit in a 5-4 configuration, and (bottom) 6 people sit in a 6-0 configuration. (c) Counting results, as a function of time, for (left) an experiment with 10 people in a 4-4-2 configuration and (right) an experiment with 8 people in a 4-2-2 configuration. (d) Final counting results for all the 19 experiments conducted in Area 1.

we conduct experiments in different environments with people engaged in different categories of activities: audience-style activity and reading activity.

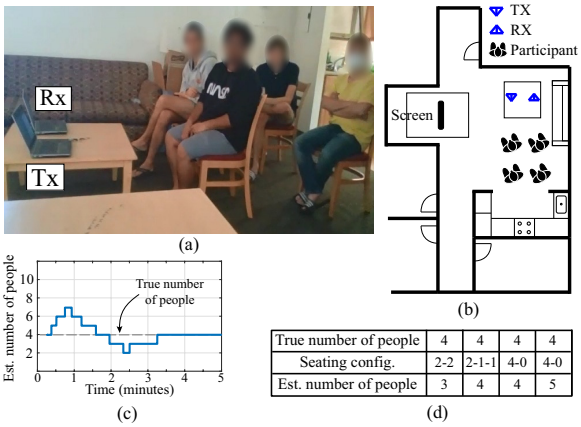
Counting an audience

In this category, we conduct experiments in two different environments, where the crowd is engaged in an audience-style activity, e.g. watching a movie or attending a lecture. Fig. 7 (a) shows two sample shots of the first test environment (Area 1), which is an outdoor patio where different number of people (up to and including 10) gather to watch a movie. The patio is cluttered with the walls of the house, multiple furniture items, tables, the TV, some trees, as well as swimming and diving equipment. In this area, we conduct a total of 19 experiments, each lasting for 5 minutes, in which different number of people sit in rows of chairs. We run experiments with several different seating configurations, involving one, two, and three rows. Fig. 7 (b) shows the seating configurations for three sample experiments conducted in this area. For example, in the first sample experiment of Fig. 7 (b-top), 10 people sit in 3 rows such that 4 people sit in row 1, 4 people in row 2, and 2 people in row 3. We refer to such a seating configuration as 4-4-2. Fig. 7 (b-middle), on the other hand, shows a seating configuration where 5 people sit in the front row, and 4 people sit in the second row, while Fig. 7 (b-bottom) shows a sample configuration where 6 people sit in one row. We refer to any seating configuration by the number of people in each row of chairs (starting from the row closest to the Tx/Rx), separated by hyphens. Fig. 7 (c) shows the estimated number of people as a function of time for two sample experiments in this area: 10 people in a 4-4-2 configuration and 8 people in 4-2-2 configuration. Note that the gap at the beginning is due to the fact that our system waits until the end of the first measured CFP, in order to have at least one CFP and one CSP to estimate the number of people using Eq. 5. It can be seen that the count estimate starts to converge after about 2 minutes into the data collection. Fig. 7 (d) shows the final estimated number of people for all the 19 experiments conducted in this area, involving different number of people and seating configurations. Out of these 19 experiments, our system achieved a counting error of 0 or 1 in 18 experiments (94.74% of the time).

The second test environment (Area 2) for this category is an indoor apartment where a set of $N = 4$ people are watching an online lecture. We conduct a total of 4 experiments in this area, with different seating configurations. Each experiment is 5 minutes long. Fig. 8 (a) shows a sample experiment where 4 people are watching a lecture while seated in two rows, with 2 persons in each row. Fig. 8 (b) shows the floor plan of the apartment as well as the locations of the participants with respect to the Tx, Rx, and the screen. Fig. 8 (c) shows the estimated number of people as a function of time for one sample experiment in which the participants were seated in a 4-0 configuration. It can be seen that the estimated number of people, using our proposed approach, converges to the true number of people. Finally, Fig. 8 (d) shows a table of the final counting results for all the 4 experiments in Area 2, where it can be seen that our proposed approach achieves a counting error of 0 or 1 in all 4 experiments, showing a very good counting performance in an indoor environment.

Counting readers

In this category, we conduct 4 experiments, where each person in the crowd is reading. The experiment location (Area 3), shown



True number of people	4	4	4	4
Seating config.	2-2	2-1-1	4-0	4-0
Est. number of people	3	4	4	5

(d)

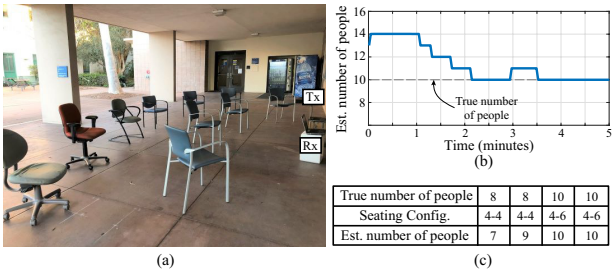
Figure 8: (a) The second test environment (Area 2): four residents of an apartment gather in the living room to watch a lecture, in different seating configurations. (b) The floor plan of the apartment showing the locations of the Tx, the Rx, and the 4 participants sitting in a 2-2 configuration. (c) Estimation of the number of people for an experiment where the participants were seated in a 4-0 configuration. (d) The final counting results of the 4 experiments conducted in Area 2, showing a very good counting performance in an indoor environment.

in Fig. 9 (a), is a roofed space (closed from 4 sides) with walls, doors, and vending machines. We conduct four experiments in this area where 8 or 10 people sit in different configurations to read, while a pair of WiFi transceivers are placed in the same area. Fig. 9 (b) shows the counting result, as a function of time, for one sample experiment in which 10 people sit in a 4-6 configuration, while Fig. 9 (c) shows the final counting results for all the 4 experiments, showing a very good counting performance.

In summary, for non-through-wall scenarios, we conducted a total of 27 experiments in 3 different environments, including an indoor space, a roofed space, and an outdoor space, with different types of activities. Define the counting error, e, as the absolute difference between the estimated and true number of people ($e = |\hat{N} - N_{\text{true}}|$). We then have an error of 0 or 1 in 26 out of the 27 experiments in non-through-wall settings, i.e. prob$(e \leq 1) = 0.963$. Next, define the mean absolute error as MAE $= \mathbb{E}(e)$, and the normalized mean square error as NMSE $= \mathbb{E}\left(\frac{e^2}{N_{\text{true}}^2}\right)$, where $\mathbb{E}(.)$ is the expectation operator. The MAE and NMSE for the non-through-wall experiments are then as follows: MAE = 0.44, and NMSE = 0.015. Finally, the correlation between the true and estimated number of people, defined as $\rho = \frac{\text{Cov}(\hat{N}, N_{\text{true}})}{\sigma_{\hat{N}} \sigma_{N_{\text{true}}}}$, is 0.959, where Cov(.,.) is the covariance between the two variables, and σ is the standard deviation of the corresponding variable. This high correlation coefficient shows that the estimated number of people matches the true number of people well.

6.2 Through-wall stationary crowd counting

To further validate our counting framework in more challenging scenarios, we conduct several experiments (with different types of activities) in through-wall scenarios, where the Tx and Rx are placed in a different space outside the seating area (i.e. behind a wall), with no direct LoS to any of the people in the crowd.



True number of people	8	8	10	10
Seating Config.	4-4	4-4	4-6	4-6
Est. number of people	7	9	10	10

(c)

Figure 9: (a) The third test environment (Area 3): a roofed space where 8 or 10 people are engaged in a reading activity. (b) Estimation of the number of people in one sample experiment in which 10 people sit in a 4-6 configuration. (c) The final counting results for all the 4 reading experiments in Area 3, showing a very good counting performance.

Counting readers

In this category, we conduct a total of 16 experiments, where we collect WiFi data in Area 4 (two snapshots of which are shown in Fig. 10 (a)). Area 4 is a roofed space (closed from 4 sides) in which we place the Tx and Rx behind a wall, as can be seen in the figure. In each of the experiments in this area, different number of people (up to and including 10) sit in different seating configurations while each individual is reading. Fig. 10 (b-top) shows the final counting results for all the 16 reading experiments in this area. It can be seen that our proposed approach can achieve a very good counting performance. More specifically, we achieve a counting error of 0 or 1 in 15 out of the 16 experiments (93.75% of the time).

Counting an audience

In this category, we conduct 4 experiments in Area 4 (Fig. 10 (a)), where the WiFi Tx and Rx are placed behind a wall. In these experiments, different number of people gather, in different seating configurations, to watch a documentary. Fig. 10 (b-bottom) shows the final counting results for these 4 audience-style experiments, showing a very good counting performance.

In summary, for through-wall scenarios, we conducted a total of 20 experiments, with different types of activities. The counting result was off from the true number of people by 0 or 1 in 18 out of the 20 experiments (i.e. prob$(e \leq 1) = 0.9$), with a mean absolute error (MAE) of 0.85, a normalized mean square error (NMSE) of 0.028, and a correlation coefficient (ρ) of 0.904. These results show that our proposed system achieves a very good counting performance even in through-wall cases.

Table 1 summarizes the performance over all the areas, activities, and seating configurations for both through-wall and non-through-wall settings.

7 DISCUSSIONS

In this section, we provide a detailed discussion on different aspects of our proposed crowd counting approach, and further motivate future research directions.

(a)

A crowd of readers

True number of people	4	4	4	4	5	5	5	5
Seating config.	2-2	2-2	3-1	3-1	3-2	3-2	3-2	3-2
Est. number of people	5	5	3	4	4	5	4	6

True number of people	6	6	6	6	8	8	10	10
Seating config.	3-3	3-3	3-3	3-3	4-4	4-4	4-6	4-6
Est. number of people	5	5	5	7	7	8	8	9

Crowd as an audience

True number of people	8	8	10	10
Seating config.	4-4	4-4	4-6	4-6
Est. number of people	6	8	10	9

(b)

Figure 10: Through-wall crowd counting – (a) Two snapshots of sample experiments in the fourth test environment (Area 4): a roofed space where the Tx/Rx are placed behind a wall. (b) The final counting results for all the 20 experiments in Area 4 involving different activities (reading and watching a movie), showing a very good counting performance in through-wall scenarios.

	Non-through-wall	Through-wall
Number of experiments	27	20
prob(counting error $\leq$ 1)	0.963	0.9
Mean Absolute Error (MAE)	0.44	0.85
Normalized Mean Square Error (NMSE)	0.015	0.028
Correlation coefficient (ρ)	0.959	0.904

Table 1: Overall performance of the proposed counting system, over several different areas, activities, and seating configurations.

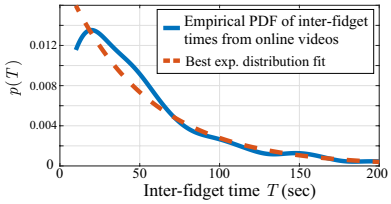


Figure 11: The exponential distribution well fits the empirical PDF of the collected individual inter-fidget times of 6 subjects, showing that the Poisson process well describes the individual fidgeting process.

Validity of the Poisson model for the fidgeting process

To develop the mathematical model for the fidgeting process in Sec. 3, we stated that an individual's fidgeting process is well modeled by a Poisson process, i.e. the inter-fidget times of an individual would follow an exponential distribution. To further validate this, Fig. 11 plots the empirical PDF of the collected inter-fidget times of 6 people (with very close fidgeting rates so we can put all the data in one pool) from online Youtube videos. The figure also shows the best exponential distribution fit for the empirical data. It can be seen that there is a tight fit between the two, with a very small Kullback-Leibler Distance (KLD) of 0.1834, confirming the validity of the Poisson model.

Robustness to interference by people moving nearby

In real-world scenarios, the stationary crowd can, once in a while, be interrupted by a nearby movement of some walking person, e.g. a person entering/exiting/passing by the area. If the interfering person is close enough to the WiFi transceivers, his/her motion can be picked up by the WiFi receiver. However, such an interference will not affect the counting results much, as long as the rate of these interfering events is not too high. In fact, in several of our experiments (e.g., all those in Areas 3 and 4), pedestrians were passing by frequently, with a rate of 0.3 person per minute (based on observing the pedestrian flow for 90 minutes). However, as all the results of Sec. 6 confirmed, we could still count the number of seated people in these areas with a high accuracy.

To further validate this, we conduct two sets of through-wall experiments in Area 4 (shown in Fig. 10 (a)). In the first set, we conduct two 5-minute experiments where 5 people sit in the area and read. We then ask an outsider to walk by the WiFi coverage area frequently and throughout the whole measurement duration, in order to interfere with the WiFi measurements, with an average duration between interfering events of 160 seconds (the interference rate is close to half the average fidgeting rate of an individual). In both experiments, the final counting result was $\hat{N} = 5$, showing that the interference did not affect the counting performance. In the second set of experiments, we conduct two 5-minute experiments where the intruder walks by the WiFi area throughout the whole experiments, with an average duration between interfering events of 90 seconds (the interference rate is close to the average fidgeting rate of an individual). In both these experiments, the final counting result is $\hat{N} = 6$, since the recurring interference events can be considered as the fidgeting of an extra person in the reading group. Overall, if people move by with a rate less than the average fidgeting rate of an individual, their impact will be negligible. Even for higher rates, the impact can be small if the rate is not too high. As part of future work, one can see if the major motion of someone passing by is differentiable from the in-place fidgets in order to properly separate them.

Characterization of the maximum number of people

In this paper, we have tested our approach in several different settings where the operation area got very crowded, i.e., the density of people per area was high. However, in larger operation areas, even more people can be present and need to be counted. As such, understanding the limitations of a single WiFi link in terms of the maximum number of people it can count is important. Our proposed fidget-based counting system relies on the durations of CSPs/CFPs for counting. As such, the counting performance is expected to degrade when the number of people in the crowd is very large to

the extent that there are no CSPs, i.e. at least one person is fidgeting at any point in time, resulting in the aggregate process of Fig. 2 becoming a very long fidgeting event. While an exact mathematical characterization of this limitation is part of future work, intuitively, this situation is more likely to happen when the number of people far exceeds $\tilde{\gamma}/\tilde{d}$, where $\tilde{\gamma}$ is the average inter-fidget time and $\tilde{d}$ is the average fidget duration. As part of future work, one can then utilize more resources, e.g., more links, for counting over larger areas that contain more people.

Robustness to errors in evaluating p_Γ

In this paper, we have assumed that the general high-level type of an event is known for a given area (e.g., is this a library or a restaurant?), in order to acquire a PDF for p_Γ from available online videos of individuals engaged in similar types of activities (see Sec. 5 for more details). While the high-level type of the event would be fairly known for a given area and is thus easily assessable, for the sake of completion, we next study the robustness of the system to the errors in assessing p_Γ. More specifically, instead of using the corresponding priors found for the audience-style and reading-style activities, we set $p_\Gamma(\bar{\gamma}) = 1$ for $\bar{\gamma} = 60$ seconds and zero otherwise, based on the crude characterization of the 1885 paper of Sir. Galton [11]. While the overall counting results expectedly degrade, we find that the counting error is 2 persons or less for 87% of the time across all experiments (with a NMSE of 0.07), which can still be an acceptable performance for many practical applications.

Impact of static multipath

In this paper, we have tested our proposed approach extensively in indoor, outdoor, and roofed (closed from 4 sides) environments. All these environments experienced a high-level of clutter resulting in static multipath, as we discussed in Section 6. Due to the COVID-19 precautionary measures set by our institution, we could not conduct more experiments in indoor settings. However, we anticipate our results to be similar in other indoor areas. This is mainly due to the fact that the static multipath, caused by static objects in the environment, does not affect our system's performance, since its impact appears in the DC component of the WiFi signals [18], which is removed in the high-pass filtered signal $w^{\text{filt}}(t)$. Our counting results in indoor, roofed, and cluttered outdoor areas further confirm this.

Computational complexity

While the results presented in this paper are based on the offline processing of the collected CSI data, the design of the algorithm has taken real-time processing considerations into account. For instance, the steps of the WiFi processing pipeline (such as the moving variance calculation) rely on moving windows, which facilitates the implementation of the algorithm in real-time. Moreover, our proposed algorithm is computationally efficient, taking only 11.3 ms, on average, to process one second of the WiFi data, which is smaller than the window size $T_{\text{win}} = 2$ sec, adding no delay to the system when implemented in a real-time manner.

8 CONCLUSION

In this paper, we have considered the problem of counting a stationary crowd, using off-the-shelf WiFi. We proposed that the aggregate in-place natural body movements of the crowd carry crucial information on the crowd count. We then developed a mathematical model for the PDFs of the Crowd Fidgeting Periods (CFPs) and Crowd Silent Periods (CSPs) and showed their dependency on the number of people in the area. In developing our mathematical models, we revealed how our problem of interest resembles an old $M/G/\infty$ queuing theory problem, which allowed us to borrow mathematical tools from a 1985 $M/G/\infty$ queuing theory paper. We further showed how to extract the CFPs and CSPs from the received WiFi signal, using the spectral content of the signal. We extensively validated our proposed approach with a total of 47 experiments in four different environments (including both through-wall and non-through-wall settings), in which up to and including $N = 10$ people are seated. We further tested our system with several different number of people, with many different seating configurations, and also with people engaged in a variety of activities. Our evaluation results showed that our proposed approach achieves a very high counting accuracy, with the estimated number of people being only 0 or 1 off from the true number 96.3% of the time in non-through-wall settings, and 90% of the time in through-wall settings.

ACKNOWLEDGEMENTS

We would like to thank all the participants in our experiments. This work is funded in part by ONR award N00014-20-1-2779 and in part by NSF NeTS award 1816931.

REFERENCES

[1] H. Abdelnasser, K. A. Harras, and M. Youssef. UbiBreathe: A ubiquitous non-invasive WiFi-based breathing estimator. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 277–286, 2015.

[2] J. Birjandtalab, D. Cogan, M. B. Pouyan, and M. Nourani. A non-EEG biosignals dataset for assessment and visualization of neurological status. In *2016 IEEE International Workshop on Signal Processing Systems (SiPS)*, pages 110–114. IEEE, 2016.

[3] C. Chen, Y. Han, Y. Chen, H.-Q. Lai, F. Zhang, B. Wang, and K. R. Liu. TR-BREATH: Time-reversal breathing rate estimation and detection. *IEEE Transactions on Biomedical Engineering*, 65(3):489–501, 2017.

[4] Y.-K. Cheng and R. Y. Chang. Device-free indoor people counting using WiFi channel state information for Internet of Things. In *IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2017.

[5] S. Depatla, C. R. Karanam, and Y. Mostofi. Robotic through-wall imaging: Radio-Frequency imaging possibilities with unmanned vehicles. *IEEE Antennas and Propagation Magazine*, 59(5):47–60, 2017.

[6] S. Depatla and Y. Mostofi. Crowd counting through walls using WiFi. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–10, 2018.

[7] S. Depatla and Y. Mostofi. Passive crowd speed estimation and head counting using WiFi. In *2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 1–9. IEEE, 2018.

[8] S. Depatla and Y. Mostofi. Occupancy analytics in retail stores using wireless signals. In *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 1–9. IEEE, 2019.

[9] S. Di Domenico, G. Pecoraro, E. Cianca, and M. De Sanctis. Trained-once device-free crowd counting and occupancy estimation using WiFi: A Doppler spectrum based approach. In *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 1–8, 2016.

[10] J. Farley, E. Risko, and A. Kingstone. Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering. *Frontiers in psychology*, 4:619, 2013.

[11] F. Galton. The measure of fidget. *Nature*, 32(817):174–175, 1885.

[12] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13).

[13] X. Guo, B. Liu, C. Shi, H. Liu, Y. Chen, and M. C. Chuah. WiFi-enabled smart human dynamics monitoring. In *Proceedings of the ACM Conference on Embedded*

Network Sensor Systems, pages 1–13, 2017.

[14] D. Halperin et al. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM Computer Communication Review*, 41(1), 2011.

[15] O. T. Ibrahim, W. Gomaa, and M. Youssef. CrossCount: A Deep learning system for device-free human counting using WiFi. *IEEE Sensors Journal*, 19(21):9921–9928, 2019.

[16] V. V. Kalashnikov. *Mathematical methods in queuing theory*, volume 271. Springer Science & Business Media, 2013.

[17] C. R. Karanam, B. Korany, and Y. Mostofi. Tracking from one side: Multi-person passive tracking with WiFi magnitude measurements. In *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*, pages 181–192, 2019.

[18] B. Korany, C. R. Karanam, H. Cai, and Y. Mostofi. XModal-ID: Using WiFi for through-wall person identification from candidate video footage. In *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom)*, pages 1–15, 2019.

[19] B. Korany and Y. Mostofi. Nocturnal seizure detection using off-the-shelf WiFi. *arXiv preprint arXiv:2103.13556*, 2021.

[20] K. P. Lam, M. Höynck, B. Dong, B. Andrews, Y.-S. Chiou, R. Zhang, D. Benitez, J. Choi, et al. Occupancy detection through an extensive environmental sensor network in an open-plan office building. *IBPSA Building Simulation*, 145:1452–1459, 2009.

[21] G. Last and M. Penrose. *Lectures on the Poisson process*, volume 7. Cambridge University Press, 2017.

[22] M. Li, Z. Zhang, K. Huang, and T. Tan. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In *2008 19th international conference on pattern recognition*, pages 1–4. IEEE, 2008.

[23] D. Lin, E. Grimson, and J. W. Fisher. Construction of dependent Dirichlet processes based on Poisson processes. In *Advances in neural information processing systems*, pages 1396–1404, 2010.

[24] S. Liu, Y. Zhao, F. Xue, B. Chen, and X. Chen. DeepCount: Crowd counting with WiFi via deep learning. *arXiv preprint arXiv:1903.05316*, 2019.

[25] Y. Nam, Y. Kong, B. Reyes, N. Reljin, and K. H. Chon. Monitoring of heart and breathing rates using dual cameras on a smartphone. *PloS one*, 11(3):e0151013, 2016.

[26] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson. Widar: Decimeter-level passive tracking via velocity monitoring with commodity WiFi. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 1–10, 2017.

[27] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu. Widar2.0: Passive human tracking with a single WiFi link. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 350–361, 2018.

[28] W. Stadje. The busy period of the queuing system M/G/∞. *Journal of Applied Probability*, pages 697–704, 1985.

[29] H. L. Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.

[30] R. H. Venkatnarayan, G. Page, and M. Shahzad. Multi-user gesture recognition using WiFi. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 401–413, 2018.

[31] F. Wang, Q. Feng, Z. Chen, Q. Zhao, Z. Cheng, J. Zou, Y. Zhang, J. Mai, Y. Li, and H. Reeve. Predictive control of indoor environment using occupant number detected by video data and CO2 concentration. *Energy and Buildings*, 145:155–162, 2017.

[32] F. Wang, F. Zhang, C. Wu, B. Wang, and K. R. Liu. Respiration tracking for people counting and recognition. *IEEE Internet of Things Journal*, 2020.

[33] H. Wang et al. Human respiration detection with commodity WiFi devices: do user location and body orientation matter? In *ACM Internationsl Joint Conference on Pervasive and Ubiquitous Computing*, pages 25–36, 2016.

[34] W. Wang et al. Understanding and modeling of WiFi signal based human activity recognition. In *ACM International Conf. on Mobile Computing and Networking*, pages 65–76, 2015.

[35] W. Wang, A. X. Liu, and M. Shahzad. Gait recognition using WiFi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 363–373, 2016.

[36] X. Wang et al. PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices. In *International Conference on Distributed Computing Systems (ICDCS)*, pages 1230–1239. IEEE, 2017.

[37] Y. Yang, J. Cao, X. Liu, and X. Liu. Wi-Count: Passing people counting with COTS WiFi devices. In *IEEE International Conference on Computer Communication and Networks (ICCCN)*, pages 1–9, 2018.

[38] Y. Zhuo, H. Zhu, and H. Xue. Identifying a new non-linear CSI phase measurement error with commodity WiFi devices. In *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)*, pages 72–79. IEEE, 2016.

[39] H. Zou, Y. Zhou, J. Yang, and C. J. Spanos. Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT. *Energy and Buildings*, 174:309–322, 2018.