WiFi-Sleep: Sleep Stage Monitoring Using Commodity Wi-Fi Devices

Bohan Yu, Yuxiang Wang, Kai Niu, Youwei Zeng, Tao Gu, *Senior Member, IEEE*, Leye Wang, *Member, IEEE*, Cuntai Guan, *Fellow, IEEE*, and Daqing Zhang, *Fellow, IEEE*

Abstract—Sleep monitoring is essential to people's health and wellbeing, which can also assist in the diagnosis and treatment of sleep disorder. Compared with contact-based solutions, contactless sleep monitoring does not attach any device to the human body; hence, it has attracted increasing attention in recent years. Inspired by the recent advances in Wi-Fi-based sensing, this article proposes a low-cost and nonintrusive sleep monitoring system using commodity Wi-Fi devices, namely, WiFi-Sleep. We leverage the fine-grained channel state information from multiple antennas and propose advanced fusion and signal processing methods to extract accurate respiration and body movement information. We introduce a deep learning method combined with clinical sleep medicine prior knowledge to achieve four-stage sleep monitoring with limited data sources (i.e., only respiration and body movement information). We benchmark the performance of WiFi-Sleep with polysomnography, the gold reference standard. Results show that WiFi-Sleep achieves an accuracy of 81.8%, which is comparable to the state-of-the-art sleep stage monitoring using expensive radar devices.

Index Terms—Channel state information (CSI), sleep monitoring, Wi-Fi.

I. Introduction

SLEEP monitoring has attracted increasing attention from research communities. Medical research shows that high-quality sleep is not only beneficial to the adjustment and recovery of physical function but also of great significance to people's mental health [1], [2]. In the fast-paced life today, however, more and more people are suffering from sleep problems. Many health problems, such as sleep apnea and chronic insomnia are closely associated with sleep quality. Poor sleep quality may be the cause of many diseases, such as diabetes, heart disease, and high blood pressure [3], [4]. Monitoring people's sleep quality on a long-term basis will not only help to find potential health problems timely but also assist doctors in performing the diagnosis and treatment of a variety of diseases. According to American Academy of Sleep Medicine (AASM) [5], sleep can be divided into five stages, i.e., Wake, N1, N2, N3, and rapid eye movement (REM). Different sleep stages have different physiological characteristics and functions [6]. The basis for measuring sleep quality is the time distributions during different sleep stages. A high-quality sleep should include adequate sleep time and reasonable distribution of sleep stages. Hence, the recognition and classification of sleep stages is critical for sleep quality monitoring.

Polysomnography (PSG) has been used in clinical and laboratory settings to monitor the sleep through a range of sensors, and it has been regarded as the de-facto standard for sleep assessment [7], [8]. These sensors typically allow for the measurement of brain activity through electroencephalogram (EEG), airflow, heart rate, breathing rate, blood oxygen level, and the electrical activity of muscles. PSG requires subjects to sleep in a laboratory setting. The sensor data will then be interpreted to determine the sleep stages by doctors who have been well trained. This method, however, is hard to implement in a home setting due to the lack of medical equipment and expertise. In addition, a subject's sleep behavior may be quite different from her/his daily routine due to the "first night effect" or frequent clinic visits, making sleep stage monitoring in clinics not representative [9], [10]. In recent years, many off-the-shelf wearable devices, such as wristband and smartwatch, embedded with heart rate sensor and accelerometer, have been used to classify sleep stages based on the relationship between sleep stages and heart rate and motion [11]. However, wearing sensors the whole night may cause discomfort to subjects, limiting its use in real life.

Bohan Yu is with the School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: ybh1998@pku.edu.cn).

Yuxiang Wang, Kai Niu, Youwei Zeng, and Leye Wang are with the Key Laboratory of High Confidence Software Technologies, Ministry of Education, and the Department of Computer Science, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: wyxpku@pku.edu.cn; xjtunk@pku.edu.cn; ywzeng@pku.edu.cn; leyewang@pku.edu.cn).

Tao Gu is with the Department of Computing, Macquarie University, Sydney, NSW 2109, Australia (e-mail: tao.gu@mq.edu.au).

Cuntai Guan is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ctguan@ntu.edu.sg).

Daqing Zhang is with the Key Laboratory of High Confidence Software Technologies, Ministry of Education, and the Department of Computer Science, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China, and also with the Samovar, CNRS, Telecom SudParis, Institut Polytechnique de Paris, 91011 Palaiseau, France (e-mail: dqzhang@sei.pku.edu.cn).

Digital Object Identifier 10.1109/JIOT.2021.3068798

Recent studies [12]–[15] have shown significant advancement in wireless sensing using Wi-Fi signals to understand human activities and behaviors. For example, Wang *et al.* [16] demonstrated that Wi-Fi signals can be used to detect body movement in daily life, such as walking and sitting, and Zeng *et al.* [17] reported a system to detect human respiration using commodity Wi-Fi devices in which respiration is always detectable. Wireless sensing provides a new opportunity for nonintrusive sleep monitoring. SMARS [18] makes the first attempt for sleep monitoring using Wi-Fi signals and studying three-stage sleep classification (i.e., Wake, REM, and NREM). However, existing studies have some shortcomings, which we articulate as follows.

We first implement SMARS and conduct evaluations. Our results show that respiration is not always detectable though it fuses a wide bandwidth. We find clearly many blind spots where SMARS is not able to detect respiration. This may be caused by using channel state information (CSI) amplitude only. Second, SMARS fails to distinguish N1, N2, and N3 stages from NREM. Our experimental study shows that the respiration features extracted in SMARS present similar patterns among N1, N2, and N3. In sleep medicine, N3 stage plays an important role in cerebral restoration and recovery in humans [19]. N3 stage is hence often referred to as deep sleep, and N1 and N2 stages are combined as light sleep. Differentiating these stages is important but challenging due to similar features among them. In addition, medical research [20] shows that there is a strong relationship between body movement and sleep stage, i.e., the rate of body movements decreases with the stage in the following sequence: Wake >N1 >REM >N2 >N3. SMARS applies a simple threshold to represent motion statistics, which may not be able to efficiently capture body movements.

Aiming at overcoming the above challenges, in this article, we design a sleep stage monitoring system, named WiFi-Sleep, to monitor and classify four sleep stages (i.e., Wake, REM, Light Sleep, and Deep Sleep) using commodity Wi-Fi devices. We propose several novel techniques. First, our study shows that using CSI amplitude only may cause blind spots. WiFi-Sleep utilizes the CSI ratio, combining both amplitude and phase signal, which complement to each other, to remove blind spots. Second, high signal-to-noise ratio (SNR) is highly desired to ensure effective feature extraction from the signal waveform. To maximize SNR, we propose a novel maximal-ratio combining, the principal component analysis (MRC-PCA) method to fuse all CSI subcarriers. With this method, we can extract more features, such as respiration depth variance and inhale-to-exhale ratio, which have not been explored in the existing work. These newly discovered features are used in classifying different sleep stages [21], which also play an important role in medical research. Third, we utilize the Doppler-MUSIC (Multiple Signal Classification) method to capture body movement, which is useful in the sleep stage classification. We further classify all body movements into major body movements (MBMs) and periodic limb movements (PLMs). Finally, WiFi-Sleep introduces the context of physiological characteristics to the deep neuron network and adds restrictions of sleep stage transition in medicine prior knowledge. By utilizing context and transition properties, sleep stages can be classified more accurately.

In summary, the main contributions of this article are as follows.

1) We design WiFi-Sleep—the first system to monitor four sleep stages using commodity Wi-Fi devices in a real environment.

2) We use the CSI ratio to eliminate blind spots for better detection rate and propose the MRC-PCA method to maximize SNR of the respiration waveform so that more respiration features can be extracted from Wi-Fi signals with a limited bandwidth.

3) We introduce prior knowledge of sleep stage transition in medicine to the deep neural network so that WiFi-Sleep can distinguish similar sleep stages better.

4) We conduct comprehensive experiments with 12 subjects over 19 nights in a residential home bedroom alike setting and evaluate the effectiveness of WiFi-Sleep. The results show that WiFi-Sleep achieves an accuracy of 81.8%.

The remainder of this article is organized as follows. The related works are summarized in Section II. Section III shows our system workflow. Section IV introduces our respiration and body movement detection methods, and Section V introduces our four-sleep stage classification method in detail. Section VI presents experimental evaluations of WiFi-Sleep. Finally, we conclude this article in Section VII.

II. RELATED WORK

A. Sleep Stage Classification

The existing sleep stage classification works can be divided into two categories: 1) contact-based and 2) contactless.

Contact-based sleep monitoring typically uses EEG devices, smart wristbands, and smartwatches. Sleep stage classification based on EEG is represented by the work done in [22]–[24]. In these works, brainwave signals captured by EEG sensors are collected to match with different sleep stages. Higashi *et al.* [25] used dry portable electrodes and achieved about 75% accuracy [24]. Alickovic and Subasi [23] used SVM and achieved about 90% accuracy and 92% recall for four-class classification. However, the drawbacks of these approaches are obvious. EEG devices are expensive and may cause uncomfortableness for long-term wearing.

Smart wristbands (e.g., Fitbit Charge 2 [26]) and smartwatches (e.g., Apple Watch [27]) have been used in sleep stage classification. These devices can accurately extract physiological information, such as body movement and heart rate by their inertial sensors, such as accelerometer, gyroscope, compass, and infrared heart rate sensor. Study in [26] using Fitbit achieves 61% accuracy for wake, 81% accuracy for light sleep, 49% accuracy for deep sleep, and 74% accuracy for REM. This method has been widely used commercially for long-term sleep monitoring despite its low accuracy.

Contactless sleep monitoring typically uses radar devices and mobile phone sensors. Radar devices have been used in DoppleSleep [28] and RF-Sleep [29] to acquire heart rate, respiratory rate, and body movement information more accurately

TABLE I
RESPIRATION AND BODY MOVEMENT CHARACTERISTICS IN DIFFERENT SLEEP STAGES, DATA FROM [20], [21], AND [37]

Sleep Stage	Respiration Rate & Variability	Respiration FIT & Depth	Body Movement Rate
Wake	Low Rate ($\sim$14.1 bpm) & High Variability ($\sim$57%)	Low FIT ($\sim$0.46) & Variable Depth	Highest
Light Sleep	High Rate ($\sim$15.5 bpm) & Medium Variability ($\sim$47.5-52%)	Low FIT ($\sim$0.46) & Variable or Stable Depth	Higher in N1 Lower in N2
Deep Sleep	High Rate ($\sim$15.8 bpm) & Low Variability ($\sim$45%)	Medium FIT ($\sim$0.47) & Stable Depth	Lowest
REM	High Rate ($\sim$15.8 bpm) & High Variability ($\sim$54%)	High FIT ($\sim$0.48) & Variable Depth	Medium

by high-frequency radio signals. RF-Sleep achieves an accuracy of 79.8% for four-class classification. Contactless sleep monitoring does not require users to wear any device and may facilitate long-term sleep monitoring, but radar equipment is very costly. Smartphones have been used in sleep monitoring, such as Sleep Hunter [30], [31] utilizing microphone, accelerometer, light sensor, and other sensors to detect events during sleep, such as body movement, snoring, cough, long breathing, and rapid breathing. This approach offers a cheap solution to contactless sleep monitoring, but its accuracy is far from satisfaction (e.g., Sleep Hunter reports an accuracy of 64% only for three-class classification) due to the limitation of smartphone sensors.

B. Wi-Fi Perception

Wi-Fi signals have been used in recent years for activity recognition, respiration detection, and sleep monitoring. The respiration sensing based on Wi-Fi is represented by the works done in [12], [17], and [32]–[34]. The study in [32] uses the Fresnel zone model to explain the principle of breath detection based on Wi-Fi and finds that the effect of breath detection is closely related to the locations of subjects and Wi-Fi devices. The studies in [17] and [34] discover complementary of the CSI phase and amplitude, providing the theoretical foundation for detecting respiration to be used in this article.

The activity detection based on Wi-Fi is represented by CRAM [16], [35] and Dopple-MUSIC [36]. CRAM puts forward a CSI-speed model to recognize nine kinds of daily human activities. Dopple-MUSIC is a method to extract speed spectrum from the Wi-Fi signal. The work [36] utilizes information from the speed spectrum to classify human activities.

The study in SMARS [18] presents a sleep monitoring system based on Wi-Fi. SMARS extracts the respiration rate by fusing the autocorrelation function (ACF) of CSI amplitudes using MRC, so it can alleviate the blind spots issues. It estimates motion statistics from the fused ACF signal. Wake and sleep can be then distinguished by motion statistics, and REM and NREM can be distinguished by observing breathing rate variability and breathing rate deviation. While this study focuses on three-stage sleep classification on Wi-Fi devices, our work focuses on challenging four-stage sleep classification.
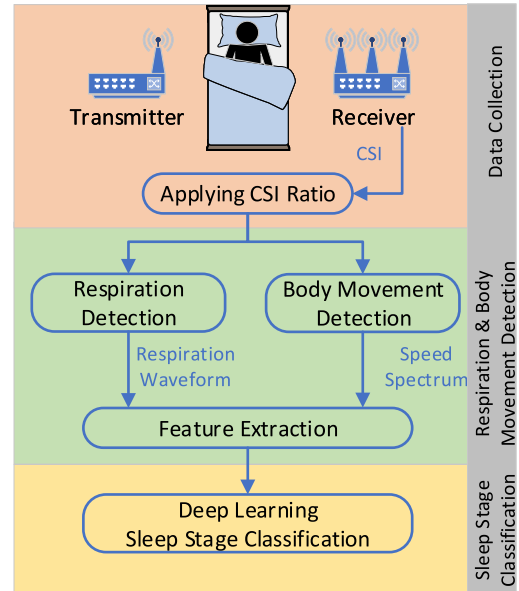

Fig. 1. System overview.

III. SYSTEM OVERVIEW

Studies [5], [6], [20], [37] show that sleep stages vary in the respiration rate, variability, fractional inspiration time (FIT), depth, and body movement rate. We summarize these differences in Table I.

From the table, we observe that different sleep stages have different body movements and respiration characteristics, which can be effectively used to classify sleep stages. Based on this idea, we design WiFi-Sleep, a four-stage sleep monitoring system, which consists of three modules—data collection, respiration and body movement detection, and sleep stage classification, as shown in Fig. 1.

The system works in a pair of Wi-Fi transceivers with a subject in the middle. Raw CSI signals will be collected by Intel 5300 Network Interface Card. We use one transmitting and three receiving antennas to enable 1×3 MIMO. Due to network delay and packet loss, the receiving time of each CSI signal may be misaligned; hence, a fixed sample rate will be used to resample the raw CSI signal. In addition, due to errors from the sampling frequency offset (SFO), central frequency offset (CFO), and packet boundary detection (PBD) [38], a phase shift occurs in CSI signals. To eliminate phase shift, we calculate the CSI ratio on each pair of receiving antennas [17], [39]. In this way, respiratory and body movement

information can be extracted from the CSI ratio. To suppress noise and maximize the SNR of the detected respiration waveform, we propose the MRC-PCA algorithm to fuse all channels from the CSI ratio into one respiratory waveform with the maximum SNR. For body movement, we employ Doppler-MUSIC (multiple signal classification) algorithm [36] to extract the speed spectrum from the CSI ratio. Finally, we extract a series of features from physiological activities and build a deep learning-based classifier to determine four sleep stages.

IV. RESPIRATION AND BODY MOVEMENT DETECTION

A. Respiration Sensing With CSI Ratio

1) Wi-Fi CSI: Wi-Fi CSI describes how the orthogonal frequency-division multiplexing (OFDM) signals get attenuated, faded, and scattered by surrounding objects during propagation. In an indoor environment, Wi-Fi signals propagate from transmitter to receiver through multiple paths, leading to multipath distortions. Mathematically, Wi-Fi CSI is the superposition of all path signals

$$H(f, t) = \sum_{i=1}^{L} A_i e^{-j2\pi \frac{d_i(t)}{\lambda}} \tag{1}$$

where L is the number of paths, λ is the wavelength, A_i is the amplitude attenuation, and $d_i(t)$ is the length of the ith propagation path. According to [17], all the paths can be grouped into static path and dynamic path. The static paths are composed of the Line-of-Sight (LoS) path and reflected paths from the walls and static objects in the environment, which do not change over time. While the dynamic paths are the signal paths induced by the moving targets, we assume there is only one subject in the lab and there is only one reflection path corresponding to the subject's chest movement caused by respiration. When the target moves a short distance, the signal attenuation of the dynamic path can be considered as a constant [40]. Furthermore, due to the fact that the transmitter and receiver are not clock synchronized, a time-varying phase offset is introduced in the obtained CSI. Mathematically, the CSI can be denoted as

$$H(f, t) = e^{-j\phi(t)} \left(H_s + A e^{-j2\pi \frac{d(t)}{\lambda}} \right) \tag{2}$$

where $\phi(t)$ is the time-varying phase offset, H_s is the static component corresponding to the static path, and A and $d(t)$ are the signal attenuation and the length of the dynamic path, respectively.

2) CSI Ratio: The CSI ratio refers to the quotient of the CSI readings from two adjacent antennas at the same receiver [34]

$$H_{\text{ratio}}(f, t) = \frac{H_1(f, t)}{H_2(f, t)} \tag{3}$$

where $H_1(f, t)$ is the CSI of the first antenna and $H_2(f, t)$ is the CSI of the second antenna. As demonstrated in [34], with this division operation between two antennas, most of the noise in the original CSI amplitude and the time-varying phase offset are canceled out. The CSI ratio of two antennas obtained is much more noise-free and sensitive compared to the original CSI reading from a single antenna when sensing
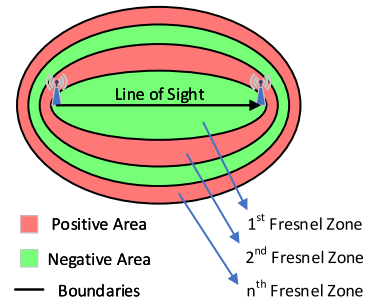

Fig. 2. Fresnel zone model.

subtle movements [34]. Another big advantage of this "CSI ratio" is that phase information can now be utilized together with the amplitude for sensing. Note that most of the existing works use CSI amplitude for sensing because CSI phase is not stable due to the lack of hard synchronization between the transmitter and receiver. The phase of the ratio is stable as the time-varying random offsets are the same at both antennas and are thus canceled by the division operation. We further combine the phase and amplitude of the CSI ratio, which are complementary to each other in terms of sensing capability to remove the "blind spots" reported in [41].

B. Respiration Enhancement by MRC-PCA

To obtain a respiration waveform with higher SNR, we propose a novel MRC-PCA method to fuse all subcarriers of the CSI ratio into one respiration waveform. In most cases, this method can obtain respiration waveforms better than a single subcarrier with the highest SNR.

The CSI signal is a combination of respiration waveform and noise. The SNR of each subcarrier usually varies because it combines respiration waveform and noise with different magnitudes. MRC can fuse subcarriers with different SNR by providing a different gain for each subcarrier, which is proportional to the root mean square of signal energy and inversely proportional to the root mean square of noise energy. The combined signal is the weighted average of all subcarriers. For the independent Gaussian noise, MRC is the best among all fusion methods for its highest expected SNR [42].

Though MRC is a general diversity fusion strategy with successful applications in wireless communications, it is not a trivial task to apply MRC in fusing the CSI signal. MRC requires the gain value of each subcarrier to be positive, however, the direction of respiratory waveform carried by each subcarrier is also affected by the frequency of a subcarrier and the location of the subject. This phenomenon can be explained by the Fresnel model. As shown in Fig. 2, in the 2nd, 4th, and 6th Fresnel zones, CSI amplitudes have the same direction with respiratory waveform (positive direction), which means positive gains. However, in the 1st, 3rd, and 5th Fresnel zone, CSI amplitudes have the opposite direction with respiratory waveform (negative direction), which means negative gains. As shown in Fig. 3(a), it can be seen that subcarriers $C1$ and $C3$ have high SNR, subcarriers $C0$ and $C2$ have low SNR, subcarriers $C0$ and $C2$ have the same respiratory waveform direction as the ground truth from the thoracoabdominal belts,
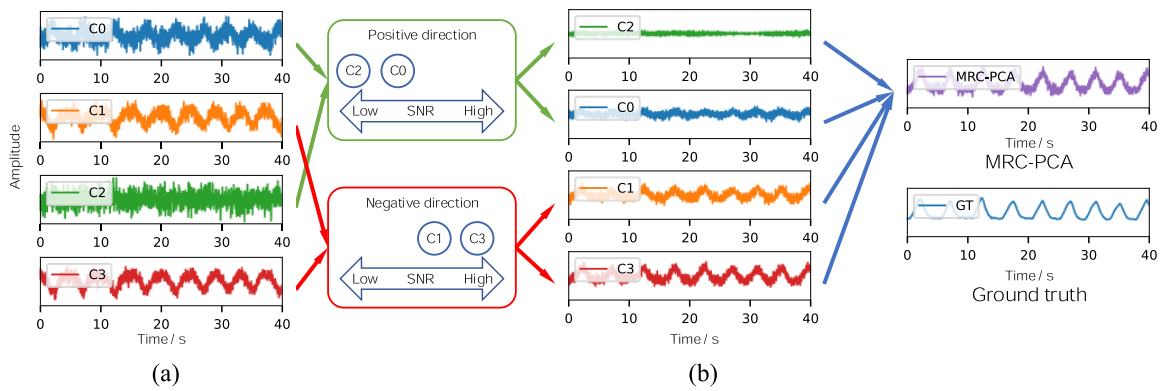
Fig. 3. (a) Original signals with different SNRs and directions. (b) Signals multiplied by gains from MRC-PCA.

and subcarriers $C1$ and $C3$ have a reverse respiration waveform. As a result, the direct use of MRC will cause the positive and negative respiratory waveform, which will cancel each other, resulting in weakening the effective information in the signal.

To fuse signals with different directions and SNRs, we propose the MRC-PCA method for subcarrier fusion. We first calculate the SNR of each subcarrier. Since respiration waveform and noise cannot be separated directly, we use the power spectral density to estimate SNR. We calculate the ratio of signal energy within the range of normal respiratory rate and signal energy above the normal respiratory rate as SNR. In order to correctly fuse signals from subcarriers with different respiratory waveform directions, we correct respiration waveform direction by PCA. We first multiply the signal by MRC gain and apply a bandpass filter within the respiratory rate range to each subcarrier. By multiplying MRC gain, we can limit the upper bound of noise rate for further operations. The bandpass filter removes noise to ensure PCA is maximizing the amplitude of respiration rather than noise. Then, we calculate the first principal component of subcarriers. Since the first principal component is maximizing the variance of signal, it will adjust the gain of each subcarrier to make the respiratory waveform direction consistent. PCA is able to fuse signals with positive or negative gain; however, it maximizes signal variance instead of maximizing SNR. The sign of the first principal component represents the direction of the respiratory waveform. We add this sign to the MRC gain as the final gain value to each subcarrier. In this way, all subcarriers with positive or negative gain can be fused together correctly with maximum SNR. As shown in Fig. 3(b), using the gain determined by PCA, the respiratory waveform directions of all subcarriers become consistent with respiration waveform direction. The gains of all subcarriers are then adjusted by their SNRs using MRC method. Subcarrier C3 has the highest SNR, which contributes most to the fused waveform, while the other subcarriers also make a relatively small contribution for noise canceling. Finally, the weighted average is calculated and all subcarriers are fused into one respiratory waveform.

We compare three signal processing methods, i.e., MRC, selecting the best subcarrier with the highest SNR, and MRC-PCA. As shown in Fig. 4, it can be seen that with MRC
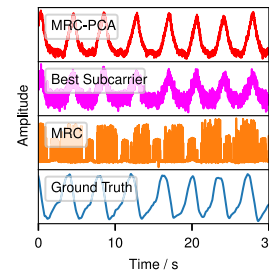

Fig. 4. Comparison of respiration waveforms using three processing methods.

alone, some respiratory signals will cancel each other, resulting in a low SNR of the fused waveform. MRC-PCA considers the SNR and respiratory waveform direction of each subcarrier together, and the fused waveform has a better effect than the single subcarrier selected with the best SNR.

C. Body Movement Detection Using Doppler-MUSIC

Human body movement statistics during sleep (e.g., rolling over and kicking leg) are also important features for sleep stage classification, especially for four-stage sleep monitoring. To detect and recognize different types of human body movements, we employ the MUSIC algorithm to extract Doppler speed spectrum from CSI ratio signals. MUSIC performs eigenstructure analysis for the $M \times M$ correlation matrix R_x of the consecutive M CSI ratio samples X. The correlation matrix R_x is denoted as

$$R_x = \mathbb{E}\left[XX^H\right] \tag{4}$$

where H is the Hermitian (conjugate transpose) of the vector. There are M eigenvalues for the correlation matrix R_x. The MUSIC algorithm divides these eigenvalues into two subspaces: 1) the signal subspace E_s and 2) noise space E_n. The signal subspace E_s corresponds to the largest L eigenvalues for the signals through L different paths, while the noise subspace E_n corresponds to the smallest $M - L$ eigenvalues. Due to the signal subspace and the noise subspace are orthogonal, the Doppler speed spectrum function can be further expressed as

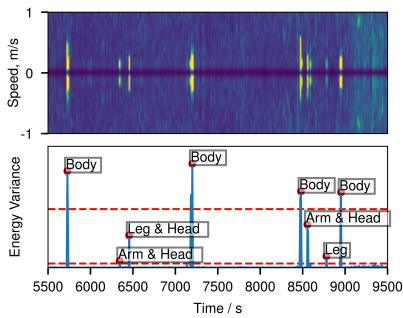$$P(v)_{\text{MUSIC}} = \frac{1}{\vec{a}^H(v)E_nE_n^H\vec{a}(v)} \tag{5}$$

Fig. 5. Doppler-MUSIC speed spectrum and energy variance of the speed spectrum and two thresholds.

Fig. 6. Respiration waveform and ACF.

where v is the Doppler speed, and $\vec{a}(v)$ is the Doppler speed vector $[1, e^{-j2\pi f(v\Delta t_2/c)}, e^{-j2\pi f(v\Delta t_3/c)}, \ldots, e^{-j2\pi f(v\Delta t_M/c)}]^T$, in which Δt_i is the time interval between the $(i-1)$th sample and the ith sample, and f is the carrier frequency.

In our implementation, a sliding window method is applied to tradeoff the processing cost and speed resolution. The window size is set as 1 s with a step size of 0.1 s. In Fig. 5, the upper one shows the speed spectrum obtained from Doppler-MUSIC for 4000 s data collected during sleep. The speed spectrum shows how the energy of each speed component varies over time, where high energy components are colored in yellow and low energy components are colored in blue. We observe that when the subject keeps still, all the speed components are at a low energy level. However, if the subject moves her/his body or part of the body, the speed components in the range of -0.5–0.5 m/s correspond to high energy. Thus, we can employ the energy variance of the speed spectrum to detect whether there is body movement. As shown in the lower figure of Fig. 5, the energy variance of the speed spectrum for movement is significantly larger than that for still. More importantly, the magnitudes of energy variance can reflect different scales of body movements during sleep. Specifically, MBM is at the largest magnitude, including body turns and rolling over. While PLM is at a smaller magnitude. A larger movement scale results in a greater energy variance. We set two different thresholds (the red line in Fig. 5) to detect MBM and PLM. These two thresholds are calibrated using MBM and PLM detected by PSG devices by maximizing the $F1$-score, detailed in Section VI. Then the features can be extracted from the movement information to classify sleep stages.

D. Feature Extraction

After obtaining a clear respiration waveform and speed spectrum, we need to design and extract useful features from respiration and body movement data.

To obtain respiration rate features, we first need to extract the respiration rate in a robust way. Existing methods generally apply a window in the time domain and use short-time Fourier transform (STFT) to extract the frequency with the maximum energy as the respiratory rate. Since the resolution in the frequency domain is related to the size of the window in the time domain, the larger window we use, the higher the respiration rate resolution we obtain. However, the presence of undetectable duration in the window will affect respiratory
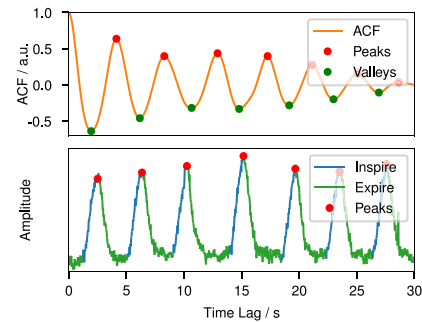
detection, resulting in an incorrect respiratory rate, and a larger time window will prolong the duration of this influence. This may run into a paradox in selecting an appropriate window size.

To solve this problem, we use the ACF [43] method to extract respiratory rate. For a finite length discrete-time signal $y(n)$, the ACF $R_{yy}(\tau)$ is defined as follows:

$$R_{yy}(\tau) = \sum_{n \in Z} y(n)y(n-\tau) \tag{6}$$

where τ represents time lag.

Since the respiratory waveform is a periodic signal, $R_{yy}(\tau)$ has the maximum value when τ reaches a full period; $R_{yy}(\tau)$ has the minimum value when τ reaches a half period. As shown in Fig. 6, for a respiratory waveform, the ACF presents the periodic transformation of peak and valley alternations, and the periodicity is consistent with the original signal. Then, the respiratory rate $rr(y)$ is extracted using the following formula:

$$rr(y) = \frac{1}{T} = \frac{1}{\text{peak}_{yy}^{(1)}} \tag{7}$$

where $\text{peak}_{yy}^{(1)}$ is the time lag of the first peak of ACF, which is equal to respiratory period T. Then, the respiration rate resolution $|\Delta rr(y)|$ is

$$|\Delta rr(y)| = \frac{1}{r_s}|rr(y)'| = \frac{1}{r_s\left(\text{peak}_{yy}^{(1)}\right)^2} \tag{8}$$

where r_s is the sampling rate of CSI signals.

According to (8), a higher r_s results in a smaller $|\Delta rr(y)|$, indicating that the ability to detect small changes in respiration rate is higher, and the resolution is not related to the window size. Therefore, ACF can effectively break the paradox in STFT mentioned above.

With the above process, we can obtain respiration rate and body movement information. For respiration, we first use the respiration rate as a feature. Considering that the stability of respiratory rate at different sleep stages differs from person to person [5], [37] (i.e., the respiratory rate is stable during light sleep and deep sleep, and it fluctuates greatly during wake and REM), we use the variance of respiratory rate within a sliding window as a feature to describe the fluctuation of respiratory rate. In addition, we also consider other features, such as the smoothed respiration rate and its first-order derivative, interquartile range (IQR) of respiration rate, and so on. In
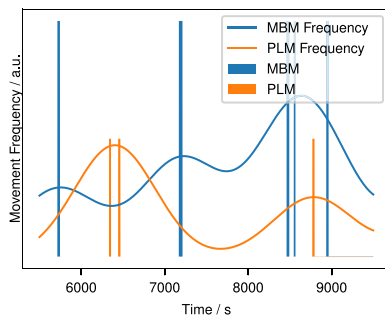
Fig. 7.   MBM and PLM during sleep and their frequencies.

Fig. 8.   Completion of respiratory rate in duration with body movement.

Fig. 9.   Respiratory rate before and after detrend.

addition to respiration rate features, we also extract several features from the respiratory effort. It has been found that the respiratory depth is more irregular and the tidal volume is smaller in REM than that in NREM [44]. When the subject and Wi-Fi devices are stable, the deeper the subject breaths, the larger the amplitude CSI signal changes. However, this ratio converting respiratory depth to CSI signal depends on the locations of subjects and Wi-Fi devices, which are usually unknown. As a result, it is still challenging to sense respiratory depth using Wi-Fi devices. WiFi-Sleep does not sense respiration depth directly but obtains respiration depth variance in an indirect way. We select a small window and assume that the subject has a stable location during this period, which is common during sleep. We keep MRC-PCA parameters the same in this window and normalize the obtained respiration waveform, the relative respiration depth is the same as the ground truth. As shown in the lower figure of Fig. 6, we detect the peak of each respiration cycle and calculate the depth variance as a feature. Beside respiration depth, fractional inspiratory time (the ratio of the time of inspiration to the total breath time) is also different in different stages, WiFi-Sleep divides inspiration and expiration segments, so FIT and I/E ratio (the ratio of inspiratory to expiratory) and their variance are also added as features. Fractional inspiratory time is also meaningful in medicine because low values may reflect severe airways obstruction and can also occur during speech, higher values are observed when snoring [45].

For body movement, there are two kinds of body movements during bedtime—MBM and PLM. MBM is movement and muscle artifact obscuring the EEG for more than 15 s. MBM contains body turns. It mostly occurs in the wake stage and usually marks the transition from deeper stages to lighter stages. PLM includes spontaneous sleep-related movements, frequently involving the flexion of the toe, ankle, knee, and hip. Each movement lasts for 0.5 to 10 s and occurs at an interval between 5 s to 90 s. PLM is most frequent during N1 and N2 stages. The movements become less frequent during N3 and REM stages. To use MBM and PLM well, two major characteristics of body movement are amplitude and frequency. Body movements and their amplitude are obtained in Section IV-C. We classify body movements with high amplitude or more than 15 s in a 30-s moving window as MBM, and others as PLM. We use weighted moving average (WMA) with the Gaussian kernel function to calculate the frequency of MBM and PLM. As shown in Fig. 7, body movements
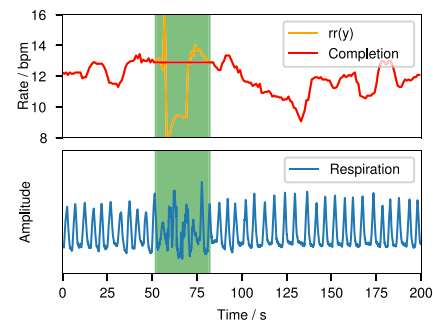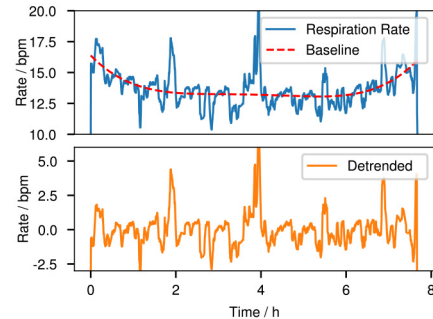
obtained in Fig. 5 are classified into MBM and PLM. The 5th movement lasts for more than 15 s, so it is MBM.

In the end, we extract 13 features of the respiration rate and body movement in total which will be used for sleep stage classification.

Before classifying sleep stages, we need to do some amendments to features. We find that body movement will influence respiration detection. The durations with body movements are sparse during the whole night, so we remove these durations and use linear interpolation to provide complete respiratory features for subsequent sleep stage classification. As shown in Fig. 8, there is a body movement at around 60 s, which leads to an incorrect respiration rate in this period. We cut and complete this period using linear interpolation.

We also need to detrend and standardize features. Removing a trend from our data set allows us to focus on the fluctuations rather than absolute values of the features. Traditional detrend uses a line to fit data points as the baseline and then calculates the difference between data points and the baseline. For a relatively short time, this method can effectively remove the overall trend of the data, and pay more attention to data fluctuation. However, we have to use the features of about 8 h in the whole night, the overall trends of the features can be complex in such a long time. A simple linear fitting cannot represent the overall trend well. We can achieve a better detrend effect if we use a higher order polynomial curve as the baseline, and we find that this method is effective for respiratory rate. As shown in the upper figure of Fig. 9, before detrend, the respiratory rate throughout the night tends to be high at both ends and low in the middle. After detrending, the trend is reduced and the fluctuations associated with sleep stages are highlighted in the lower figure of Fig. 9. Finally,
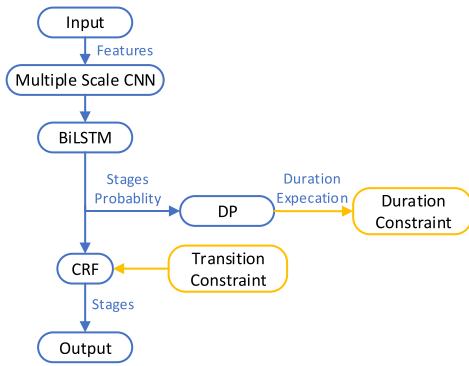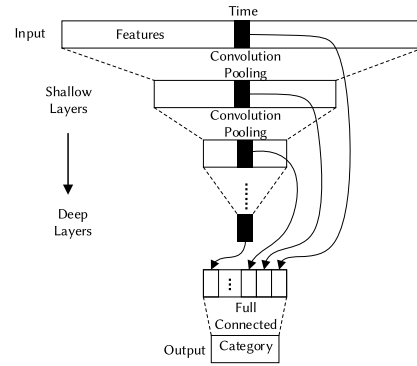
Fig. 10. Deep learning structure of WiFi-Sleep.

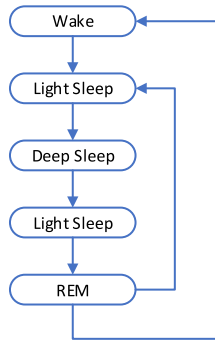Fig. 12. Multiscale CNN structure.

Fig. 11. Sleep cycle.

we detrend and standardize all features and feed them to the classifier.

V. SLEEP STAGE CLASSIFICATION

We build a deep learning-based classifier, which consists of convolutional neural network, bidirectional long short-term memory (CNN-BiLSTM) layers, duration constraint, and transition constraint, as shown in Fig. 10. WiFi-Sleep aims to classify four-class sleep stages, i.e., wake (W), light sleep (L), deep sleep (D), and REM (R).

A. Sleep Stage Classification Based on Context

In the previous section, we extract physiological features that are closely related to sleep stages. However, physiological features at each moment are not isolated, there are stages where the physiological features are stable like deep sleep, also stages where the physiological features are changing, e.g., the respiratory rate during light sleep may be gradually increasing from low during "wake" to high during "deep sleep." If we use only the respiratory rate in a moment to determine sleep stages, it will be difficult to distinguish light sleep and deep sleep stages. The transition between the two stages may also have some physiological features. The transition in the whole night follows sleep cycle, as shown in Fig. 11. According to the sleep stage interpretation manual by the American Medical Association [5], a tendency may rise to shift to a lighter sleep after body movement at a deeper sleep. For example, if a body movement occurs during deep sleep, it will turn to light sleep, while a body movement occurs during light sleep, it will turn

to wake. Therefore, by taking the physiological features before and after the current moment as context, and taking the transition of the sleep stage in the sleep cycle into consideration, we may obtain better classification results.

WiFi-Sleep uses a CNN-BiLSTM neural network to utilize the context of physiological features and the sleep stage transition information. CNN layers use a sliding window with the current moment as the midpoint. The window slides the same duration as sleep stage episode length, so the output of CNN layers has the same sample rate as sleep stages.

For a CNN network, if we select a larger sliding window and use a deeper network, the network can see the contextual features for a longer time. However, it may not use fine-grained features around the current moment well because features are smoothed out by too many convolutional and pooling layers. Using a smaller sliding window can alleviate this problem, but there is less contextual information available. This problem also exists in image processing, and multiscale CNN [46] has been proposed to solve this problem. For the middle layer of a CNN with a larger sliding window, the neuron output has a receptive field smaller than the sliding window, so the effect is similar to a neural network with a smaller sliding window. In WiFi-Sleep, we adopt a multiscale CNN network structure shown in Fig. 12. This structure can capture features with both longer and shorter durations at the same time.

CNN layers take the physiological features in a period of time as context. However, it is not able to use the context of the whole night. CNN is also not good at utilizing transition features because it cannot see the classification result of earlier and later episodes. The BiLSTM neural network is able to remember the classification results and hidden information from former and latter BiLSTM nodes. It is able to use the whole night as context and capture transition features. Therefore, we use the multiscale CNN and BiLSTM together to form the CNN-BiLSTM network. Neural networks with BiLSTM layers are more time consuming in training, hence we pretrain CNN with an additional fully connected (FC) layer to predict stages, and connect the output before this layer to BiLSTM. This training strategy is more efficient. The classification loss L_{class} used for both CNN and BiLSTM is cross-entropy loss.

B. Duration Constraint

Each sleep stage lasts for a period of time. In each sleep cycle, N1 lasts for 1–7 min, N2 lasts for 10–25 min, and N3 lasts for 20–40 min. The amount of REM in each cycle progresses throughout the night from being minimal initially to 30% of the cycle eventually. Usually, the initial sleep cycle lasts for 70–100 min, and the remaining cycles last for 90–120 min each. There may be brief wakefulness in REM and light sleep, and there is only a unidirectional conversion or no conversion between some stages. There are some regularities of the sleep stage duration. However, deep learning only aims at minimizing prediction errors. As a result, these rules will sometimes be broken and may not achieve the best prediction. We find that sleep stages predicted by CNN-BiLSTM usually have frequent transitions. This phenomenon is more serious between stages that are difficult to distinguish, such as deep sleep and light sleep.

To eliminate frequent transitions, we apply a duration constraint by two steps—calculating stage duration expectation and adding a loss to punish sleep stages that have too short durations. The loss is defined on the expected stage duration to ensure that it is differentiable. We calculate stage duration expectations by dynamic programming (DP).

Let the jth output of BiLSTM at time i be A_{ij}, we first apply the LogSoftmax function to turn network outputs into log probabilities B_{ij}

$$B_{ij} = \frac{\exp(A_{ij})}{\sum_k \exp(A_{ik})}. \tag{9}$$

We assume that the sleep stage of time i is j, let C_{ij} be the expected duration of stage j that lasts till time i, let d be the unit duration

$$C_{ij} = \begin{cases} 0, & (i=0) \\ B_{(i-1)j}\big(C_{(i-1)j}+d\big) + \big(1-B_{(i-1)j}\big)d, & (i \neq 0). \end{cases} \tag{10}$$

We punish stages that have too short durations, the loss L_{duration} is defined as follows:

$$L_{\text{duration}} = \sum_{i>0} \sum_j \text{ReLU}\big(T_j - C_{(i-1)j}\big(1 - B_{ij}\big)\big) \tag{11}$$

where T_j is the minimum duration of stage jth. We choose 10 min for Wake, Light Sleep, and REM stages, respectively, and 20 min for Deep Sleep.

C. Transition Constraint

Medical research shows that the transition of sleep stages follows a certain rule [47]. In the normal sleep process, the transition of sleep stage follows Light $\rightarrow$ Deep $\rightarrow$ Light $\rightarrow$ REM, as shown in Fig. 11. However, we find that there are unexpected transitions in predicted stages. For example, there are transitions between Deep Sleep and REM, or from Wake to REM directly. These transitions affect the accuracy and should not exist in sleep medicine.

To eliminate these unexpected transitions, we add a conditional random field (CRF) layer after BiLSTM. The CRF layer is trained to maximize the likelihood of the ground truth sleep stage sequence. We use negative log-likelihood as CRF loss L_{CRF}. The feature function used in CRF consists of two parts–the transition score matrix and the output of the CNN-BiLSTM network. We manually set the score of three unexpected transitions mentioned above to negative infinity, so that they will never happen. Other parameters are learned during training.

We apply both duration and transition constraints to improve accuracy. If we only apply stage duration constraint, the unexpected stages will only be prolonged, and if we only add the transition constraint, the network will turn one unexpected transition into two continuous legal transitions, such as turning Deep $\rightarrow$ REM into Deep $\rightarrow$ Light $\rightarrow$ REM.

We use joint optimization to train the CNN-BiLSTM-CRF network, the total loss L_{total} is a combination of each partial loss, where α, β, and θ are weights

$$L_{\text{total}} = \alpha L_{\text{class}} + \beta L_{\text{duration}} + \theta L_{\text{CRF}}. \tag{12}$$

VI. EVALUATION

We implement a prototype system using commodity WiFi devices, and conduct extensive experiments to evaluate WiFi-Sleep by comparing it with the state of the art. We also evaluate the performance of WiFi-Sleep in detecting human respiration and body movement to fully discover its capability.

A. Implementation of WiFi-Sleep

We implement WiFi-Sleep using two GigaByte Mini PCs equipped with Intel 5300 wireless cards as a WiFi transceiver pair. Both operate at the 5-GHz WiFi channel with a bandwidth of 20 MHz. The transmitter sends standard WiFi data packets at a rate of 200 Hz to the receiver, which is equipped with three omnidirectional antennas and configured to capture CSI signals. We install an opensource Linux CSITool [48] in the receiver to collect CSI data from 30 subcarriers for each antenna. A Lenovo laptop with Intel Core i7-8550U CPU and 8-GB RAM is connected to the receiver via an Ethernet cable for processing CSI data using Python in real time. We calculate the features from CSI data in a sliding window of 20 min and apply CNN with eight convolution layers and eight max-pooling layers (each layer has 16 channels) to generate 128 features as the input to BiLSTM. The size of the hidden layer and output layer for BiLSTM is set to 16. The CRF layer is then connected to impose the transition constraint and outputs the final classification result of the sleep stage. In sleep medicine, PSG cuts the whole night's sleep duration into 30-s episodes. Sleep diagnostic experts label one sleep stage for each episode. For consistency, WiFi-Sleep also updates the final sleep stage classification result once every 30 s.

B. Experimental Methodology

1) Data Collection: We deploy WiFi-Sleep in a sleep laboratory at Peking University, which is set up for sleep study in medical science. Fig. 13 illustrates our experimental setting in which the transceiver pair is placed close to the bed. This setting is recommended to ensure weak respiratory signals can be picked up during sleep regardless of different sleep postures. We collect 19 nights of sleep data from 12 subjects aged from 20 to 42. The data collection procedure is applied with

Fig. 13. Experimental setting and PSG devices.

a formal approval obtained from the Human Subject Review Committee at Peking University. Each subject is required to sign a data protection agreement before participating. CSI data of total size 24.1 GB is recorded. 15.5-GB data of 12 nights collected from five subjects is used to train the neural network, and 8.6-GB data of seven nights collected from the other seven subjects is used for testing. Note that no data from the same subject are used for both training and testing to demonstrate the generality of WiFi-Sleep.

For all the experiments, we resort to the PSG system, i.e., medical gold standard, to label the ground truth of sleep stages. As shown in Fig. 13, the PSG system requires subject to wear massive sensors, including EEG, electrooculography (EOG), electromyography (EMG), electrocardiography (ECG), and thoracoabdominal belts to record the physiological information. Meanwhile, the sleep laboratory is equipped with infrared cameras for 7/24 monitoring. Specifically, the ground truth of the respiratory rate is recorded by the thoracoabdominal belts sensor, while the ground truth of body movement is provided by the EMG sensor and infrared camera. By analyzing all the information collected by PSG, three sleep diagnostic experts label each sleep stage manually according to AASM rules [5], providing the ground truth for sleep stage classification.

2) Comparison Methods: As there is a lack of four-stage sleep monitoring systems using commodity WiFi devices, we compare WiFi-Sleep with both radar-based approach and SMARS. Specifically, we select RF-Sleep [29] and SMARS [18] as the baselines. RF-Sleep performs four-stage classification by employing low-power radar devices. SMARS achieves an accuracy of 88.4% for three-stage classification (i.e., Wake, REM, and NREM) with a pair of commodity WiFi devices. However, it did not attempt to identify four sleep stages. We enhance SMARS by putting all its features into a four-class SVM classifier to perform four-stage classification.

C. Sleep Stage Classification Performance

1) Overall Performance: We first evaluate the accuracy of WiFi-Sleep for four-stage classification (i.e., Wake, REM, Light sleep, and Deep sleep) and compare its performance with the state-of-the-art. As shown in Table II, RF-Sleep [29] achieves an accuracy of 79.8% for four-stage classification using radar devices. SMARS achieves an accuracy of 69.4% for four-stage classification using WiFi devices. In comparison, WiFi-Sleep achieves an average accuracy of 81.8% for four-stage classification, outperforming both RF-Sleep and

TABLE II
COMPARISON OF DIFFERENT SLEEP STAGE CLASSIFICATION SYSTEMS

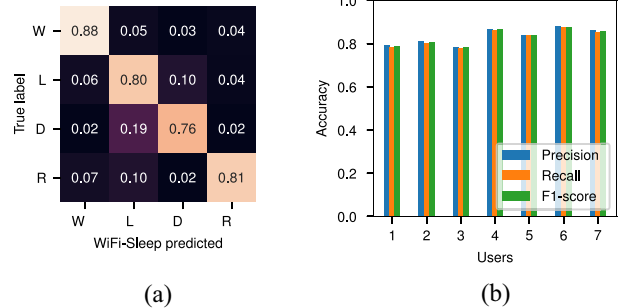| Approach	Accuracy
RF-Sleep (four-stage)	79.8%
SMARS (four-stage)	69.4%
WiFi-Sleep (four-stage)	**81.8%**

Fig. 14. (a) Confusion matrix of WiFi-Sleep. (b) Sleep stage classification performance of seven users.

TABLE III
ACCURACIES OF DIFFERENT FEATURE SETS

Feature Set	Accuracy
Basic Features	73.9%
Basic Features + Enhanced Respiration	76.2%
Basic Features + Enhanced Respiration + Body Movement	81.8%

SMARS. Fig. 14(a) shows the confusion matrix of WiFi-Sleep. We observe an accuracy of higher than 76% for all the stages. Specifically, the "Wake" stage achieves the best accuracy of 88%, while the "Deep" stage has the lowest accuracy of 76%. Note that even the lowest accuracy of WiFi-Sleep is better than that of SMARS for four-stage sleep classification. The performance improvement benefits from two aspects: 1) the respiration waveform and body movement information provide more features and 2) a specialized neural network with inputs of stage duration and transition between different stages is designed. Fig. 14(b) shows the precision, recall, and $F1$-score for different users. The standard deviation of the $F1$-score is only 3.58%, which suggests that WiFi-Sleep is reliable for user diversity.

2) Feature Evaluation: To validate the effectiveness of our features extracted from respiration waveform and body movement, we group the features into three categories.

1) *Basic Features in SMARS [18]:* Respiration rate, respiration rate statistics (variation and deviation), and motion statistics.
2) *Enhanced Respiration Features:* Respiration depth variation, fractional inspiratory time, and ratio between inhalation time and expiration time.
3) *Body Movement Features:* MBM and PLMs frequencies.

We first evaluate the performance of features in SMARS [18] and then gradually add more exclusive features in WiFi-Sleep to demonstrate its superiority to the state of the art. Table III shows the classification accuracy increases when adding more features related to sleep. In detail, the enhanced respiration features and body movement features improve the

TABLE IV
IMPACT OF EACH MODULE IN WiFi-SLEEP

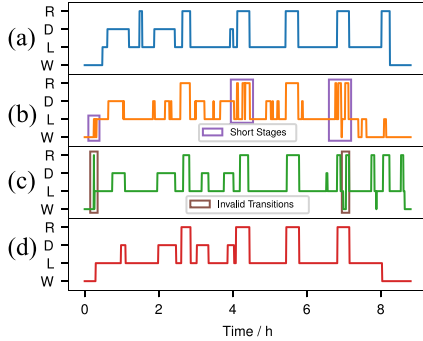| Methods	Accuracy
CNN-BiLSTM	73.7%
CNN-BiLSTM + duration constraint	78.4%
CNN-BiLSTM + transition constraint	76.7%
WiFi-Sleep	81.8%

Fig. 15. Sleep stage classification results during the whole night, (a) ground-truth sleep stages, (b) CNN-BiLSTM method, (c) CNN-BiLSTM + duration constraint, and (d) CNN-BiLSTM + duration constraint + transition constraint (full WiFi-Sleep).
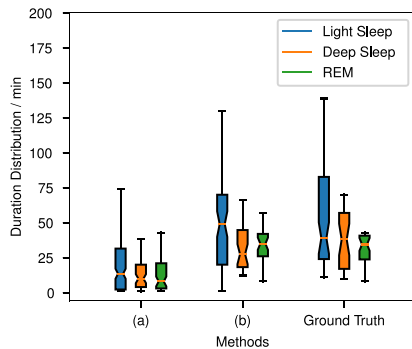

Fig. 16. Duration distribution of each stage. CNN-BiLSTM method (a) without the duration constraint and (b) with the duration constraint.

Fig. 17. Invalid sleep stage transition counts. CNN-BiLSTM method (a) without the transition constraint and (b) with the transition constraint.

overall accuracy by 2.3% and 5.6%, respectively. The results show that the more features we add, the higher accuracy we achieve.

3) Parameter Evaluation: To investigate the impact of each module in WiFi-Sleep, i.e., context utilization, duration constraint, and transition constraint, we gradually enable each module and analyze the results. Table IV shows the performance increases with more modules added. Compared with original CNN-BiLSTM, WiFi-Sleep improves the overall accuracy by about 8%. As shown in Fig. 15, WiFi-Sleep effectively mitigates the unreasonable sleep stage and achieves good classification results in the case of the whole night.

Impact of Duration Constraint: To investigate the impact of duration constraint, we compare CNN-BiLSTM with and without duration constraint. As shown in Fig. 15(b), the purple boxes illustrate the pieces of sleep stage that last only for a very short time without duration constraint. By adding the duration constraint module, Fig. 15(c) shows that these pieces are mostly eliminated. Fig. 16 shows duration distribution for Light Sleep, Deep Sleep, and REM before and after
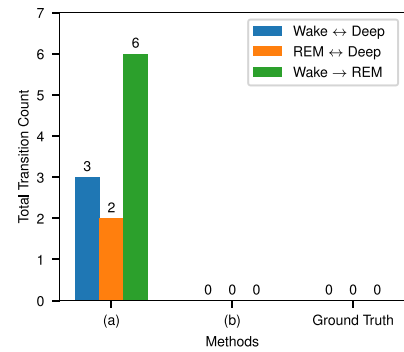
adding duration constraint. Due to the Wake stage lasts for an arbitrary amount of time, we ignore its duration distribution. Obviously, the duration distribution without constraint is very different from the ground truth, while with the constraint, the result becomes more similar to the ground truth. Specifically, the duration of light sleep and REM stages are much shorter than the ground truth. By adding duration constraint, the predicted stages are prolonged and the distribution is much closer to the ground truth. Overall, the duration constraint improves accuracy by 4.7%.

Impact of Transition Constraint: To investigate the impact of the transition constraint, we compare CNN-BiLSTM with and without transition constraint. The brown boxes in Fig. 15(c) show two invalid transitions at the 1st and 7th h. By adding transition constraint in Fig. 15(d), these invalid transitions are all eliminated. To intuitively understand the impact of the transition constraint, we conduct statistics for invalid transitions in the result of the test set. There are three kinds of invalid transitions: 1) wake ↔ deep sleep; 2) REM ↔ deep sleep; and 3) wake → REM. As shown in Fig. 17, there is no invalid transition in the ground truth. Without the transition constraint, there are a total of 11 invalid transitions in our test set results. By adding the transition constraint, we eliminate all invalid transitions. Overall, the transition constraint improves the accuracy by 3.0%, and adding both duration and transition constraints together improves the performance by 8.1%.

D. Respiratory Detection Performance

In this section, we evaluate the performance of WiFi-Sleep in respiration detection and the impact of our two techniques (i.e., CSI ratio and MRC-PCA) on accurate respiration waveform.

1) Overall Performance: We evaluate the respiration detection performance in terms of detection rate, SNR, and respiration rate error. For the detection rate, we remove the period with MBM in advance because neither thoracoabdominal belts nor Wi-Fi is able to monitor respiration correctly under MBMs. We use the method in Section IV-B to calculate SNR. Fig. 18 shows that WiFi-Sleep achieves a detection rate of 97.8% and the highest SNR. For the respiration rate error, we calculate the cumulative distribution function (CDF) shown in Fig. 19. WiFi-Sleep achieves an average error of 0.23 bpm and a 90% error of 0.29 bpm.
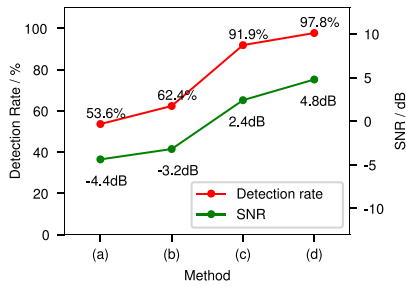
Fig. 18. Detection rate and SNR of different methods: (a) CSI, (b) MRC-PCA, (c) CSI ratio, and (d) WiFi-Sleep (CSI ratio + MRC-PCA).
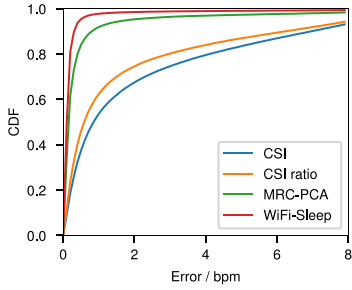

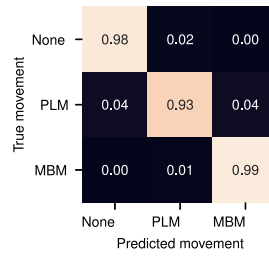Fig. 19. CDF of detection error.

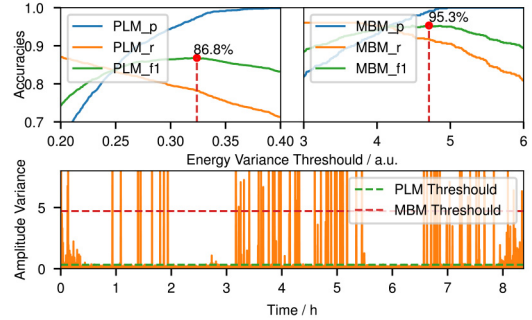
Fig. 20. Confusion matrix of body movement detection.

Fig. 21. Upper figure shows two energy variance thresholds set for PLM and MBM and their body movement detection accuracy and the lower figure shows speed spectrum energy variance and two threshold lines marked in green and red, respectively.

2) Impacts of CSI Ratio and MRC-PCA: In order to verify the effectiveness of CSI ratio and MRC-PCA, we take different signals as input to evaluate the detect rate, SNR, and respiration rate error. The signals are as follows.

1) *CSI:* Using CSI amplitude of one subcarrier.
2) *MRC-PCA:* Using CSI amplitudes of 90 subcarriers from three receiving antennas.
3) *CSI Ratio:* Using CSI ratio of one subcarrier from one receiving antenna.
4) *WiFi-Sleep (CSI Ratio + MRC-PCA):* Using CSI ratio of 90 subcarriers from three receiving antennas.

From Fig. 18, we can see that CSI ratio improves the detection rate by about 5.9% and improves SNR by about 2.4 dB, which is useful for the respiration feature extraction. We also observe that MRC-PCA greatly improves the detection rate by 35.4% and SNR by 8.0 dB. By combining MRC-PCA and CSI ratio, WiFi-Sleep achieves near 98% detection rate, implying the respiration can be always detected. Similarly, compared with the original CSI amplitude, the CSI ratio, MRC-PCA, and WiFi-Sleep reduce the 90% respiration rate error from 6.9 to 6.2, 0.77, and 0.29 bpm, respectively.

E. Body Movement Detection Performance

In this section, we evaluate the performance of body movement detection and the impact of the two thresholds to detect PLM and MBM.

1) Overall Performance: Because the duration with body movement represents only a small percentage of the whole night, accuracy can be easily affected by a large duration without body movement, so it is not a good indicator for performance evaluation of body movement. We evaluate the performance of body movement detection by F1-Score,

which takes both the precision and the recall into consideration and evaluates the performance better. Because body movement usually lasts for a couple of seconds, the marked time point for the same body movement may have a deviation of several seconds from that of ground truth. Therefore, we consider each of our body movement detection correct if the detected duration has more than 50% overlap time with the ground truth. As shown in Fig. 21, the final F1-Score for MBM and PLM reaches 86.8% and 95.3%, respectively. The confusion matrix of body movement detection is shown in Fig. 20.

2) Thresholds Evaluation: As mentioned in Section IV-C, WiFi-Sleep uses two thresholds on spectrum energy variance to detect PLM and MBM. For PLM, the threshold is set to 0.32. For MBM, the threshold is set to 4.7. We change these two thresholds and compare the result using different thresholds. The result is shown in Fig. 21. As we increase the threshold, the precision gradually increases and the recall gradually decreases. We can see that F1-Score drops if either of the thresholds increases or decreases, implying the threshold we choose is reasonable.

VII. CONCLUSION AND FUTURE WORK

In conclusion, we design WiFi-Sleep–a nonintrusive sleep monitoring system based on commodity Wi-Fi devices to achieve four-stage sleep monitoring. WiFi-Sleep is a low-cost real-time system that can be easily applied in real scenarios for long-term sleep monitoring. For future work, we will investigate detecting sleep patterns related to sleep disorders, such as chronic insomnia, restless legs, and sleep apnea, and further improve our system.

References

[1] N. Tsuno, A. Besset, and K. Ritchie, "Sleep and depression," *J. Clin. Psychiat.*, vol. 66, no. 10, pp. 1254–1269, Oct. 2005.

[2] D. J. Taylor, K. L. Lichstein, H. H. Durrence, B. W. Reidel, and A. J. Bush, "Epidemiology of insomnia, depression, and anxiety," *Sleep*, vol. 28, no. 11, pp. 1457–1464, 2005.

[3] E. Shahar *et al.*, "Sleep-disordered breathing and cardiovascular disease: Cross-sectional results of the sleep heart health study," *Amer. J. Res. Crit. Care Med.*, vol. 163, no. 1, pp. 19–25, 2001.

[4] E. Tasali, R. Leproult, D. A. Ehrmann, and E. Van Cauter, "Slow-wave sleep and the risk of type 2 diabetes in humans," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 3, pp. 1044–1049, 2008.

[5] R. B. Berry *et al.*, *The AASM Manual for the Scoring of Sleep and Associated Events—Rules, Terminology and Technical Specifications*, vol. 176, Amer. Aca. Sleep Med., Darien, IL, USA, 2012.

[6] Healthy of Sleep Medicine. (2019). *The Characteristics of Sleep*. [Online]. Available: http://healthysleep.med.harvard.edu/healthy/science/what/characteristics

[7] D. Moser *et al.*, "Sleep classification according to AASM and rechtschaffen & kales: Effects on sleep scoring parameters," *Sleep*, vol. 32, no. 2, pp. 139–149, 2009.

[8] V. Ibáñez, J. Silva, and O. Cauli, "A survey on sleep assessment methods," *PeerJ*, vol. 6, May 2018, Art. no. e4849.

[9] J. Newell, O. Mairesse, P. Verbanck, and D. Neu, "Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples," *Psychiat. Res.*, vol. 200, nos. 2–3, pp. 795–801, 2012.

[10] E. Herbst *et al.*, "Adaptation effects to sleep studies in participants with and without chronic posttraumatic stress disorder," *Psychophysiology*, vol. 47, no. 6, pp. 1127–1133, 2010.

[11] S. Lee, S. Nam, and H. Shin, "The analysis of sleep stages with motion and heart rate signals from a handheld wearable device," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, 2016, pp. 1135–1137.

[12] X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-Sleep: Contactless sleep monitoring via WiFi signals," in *Proc. IEEE Real-Time Syst. Symp.*, 2014, pp. 346–355.

[13] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf WiFi," in *Proc. 16th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2015, pp. 267–276.

[14] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, "Monitoring vital signs and postures during sleep using WiFi signals," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2071–2084, Jun. 2018.

[15] Y. Cao *et al.*, "Contactless body movement recognition during sleep via WiFi signals," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 2028–2037, Mar. 2020.

[16] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, May 2017.

[17] Y. Zeng, D. Wu, R. Gao, T. Gu, and D. Zhang, "FullBreathe: Full human respiration detection exploiting complementarity of CSI phase and amplitude of WiFi signals," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–19, 2018.

[18] F. Zhang *et al.*, "SMARS: Sleep monitoring via ambient radio signals," *IEEE Trans. Mobile Comput.*, vol. 20, no. 1, pp. 217–231, Jan. 2021.

[19] T. Roth, "Slow wave sleep: Does it matter?" *J. Clin. Sleep Med.*, vol. 5, no. 2 p. S4, 2009.

[20] J. Wilde-Frenz and H. Schulz, "Rate and distribution of body movements during sleep in humans," *Perceptual Motor Skills*, vol. 56, no. 1, pp. 275–283, 1983.

[21] N. Douglas, D. White, C. K. Pickett, J. Weil, and C. Zwillich, "Respiration during sleep in normal man," *Thorax*, vol. 37, no. 11, 1982, Art. no. 459437.

[22] F. Ebrahimi, M. Mikaeili, E. Estrada, and H. Nazeran, "Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2008, pp. 1151–1154.

[23] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1258–1265, Jun. 2018.

[24] L. Jeon and J. Finkelstein, "Consumer sleep tracking devices: A critical review," *Digit. Healthcare Empowering Eur.*, vol. 210, MIE, Johannesburg, South Africa, 2015, p. 458.

[25] Y. Higashi, Y. Yokota, and Y. Naruse, "Signal correlation between wet and original dry electrodes in electroencephalogram according to the contact impedance of dry electrodes," in *Proc. IEEE 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2017, pp. 1062–1065.

[26] M. de Zambotti, A. Goldstone, S. Claudatos, I. M. Colrain, and F. C. Baker, "A validation study of fitbit charge 2-compared with polysomnography in adults," *Chronobiol. Int.*, vol. 35, no. 4, pp. 465–476, 2018.

[27] S. Roomkham, M. Hittle, D. Lovell, and D. Perrin, "Can we use the apple watch to measure sleep reliably?" *J. Sleep Res.*, vol. 27, no. S2, 2018, Art. no. e153_12766.

[28] T. Rahman *et al.*, "DoppleSleep: A contactless unobtrusive sleep sensing system using short-range doppler radar," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2015, pp. 39–50.

[29] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi, "Learning sleep stages from radio signals: A conditional adversarial architecture," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 4100–4109.

[30] W. Gu, Z. Yang, L. Shangguan, W. Sun, K. Jin, and Y. Liu, "Intelligent sleep stage mining service with smartphones," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 649–660.

[31] W. Gu, L. Shangguan, Z. Yang, and Y. Liu, "Sleep hunter: Towards fine grained sleep stage tracking with smartphones," *IEEE Trans. Mobile Comput.*, vol. 15, no. 6, pp. 1514–1527, Jun. 2016.

[32] H. Wang *et al.*, "Human respiration detection with commodity WiFi devices: Do user location and body orientation matter?" in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 25–36.

[33] X. Wang, C. Yang, and S. Mao, "TensorBeat: Tensor decomposition for monitoring multiperson breathing beats with commodity WiFi," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 1, pp. 1–27, 2017.

[34] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–26, 2019.

[35] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 65–76.

[36] X. Li *et al.*, "IndoTrack: Device-free indoor human tracking with commodity Wi-Fi," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 1, no. 3, p. 72, 2017.

[37] G. Gutierrez *et al.*, "Respiratory rate variability in sleeping adults without obstructive sleep apnea," *Physiol. Rep.*, vol. 4, no. 17, Sep. 2016, Art. no. e12949.

[38] M. Speth, S. A. Fechtel, G. Fock, and H. Meyr, "Optimum receiver design for wireless broad-band systems using OFDM. I," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1668–1677, Nov. 1999.

[39] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, 2017, pp. 1230–1239.

[40] D. Wu *et al.*, "FingerDraw: Sub-wavelength level finger motion tracking with WiFi signals," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–27, 2020.

[41] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with Wi-Fi!" *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2907–2920, Nov. 2016.

[42] L. R. Kahn, "Ratio squarer," *Proc. Inst. Radio Eng.*, vol. 42, no. 11, p. 1704, 1954.

[43] S. Tanaka, Y. Matsumoto, and K. Wakimoto, "Unconstrained and non-invasive measurement of heart-beat and respiration periods using a phonocardiographic sensor," *Med. Biol. Eng. Comput.*, vol. 40, no. 2, pp. 246–252, 2002.

[44] X. Long, J. Rolink, P. Fonseca, R. Haakma, and R. Aarts, "Respiration amplitude analysis for REM and NREM sleep classification," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2013, pp. 5017–5020.

[45] *Respiratory Inductance Plethysmography*. Accessed: Mar. 31, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Respiratory_inductance_plethysmography

[46] T. H. N. Le, Y. Zheng, C. Zhu, K. Luu, and M. Savvides, "Multiple scale faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2016, pp. 46–53.

[47] A. Kishi, Z. R. Struzik, B. H. Natelson, F. Togo, and Y. Yamamoto, "Dynamics of sleep stage transitions in healthy humans and patients with chronic fatigue syndrome," *Amer. J. Physiol. Regul. Integrat. Comparative Physiol.*, vol. 294, no. 6, pp. R1980–R1987, 2008.

[48] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

Bohan Yu is currently pursuing the B.Sc. degree with the School of Electronics Engineering and Computer Science, Peking University, Beijing, China.

His current research interests include wireless sensing and computational photography.
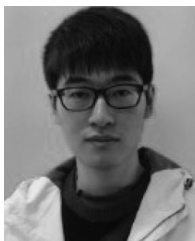
Tao Gu (Senior Member, IEEE) received the Ph.D. degree in computer science from National University of Singapore, Singapore, in 2006.

He is currently a Professor with the Department of Computing, Macquarie University, Sydney, NSW, Australia. His research interests include Internet of Things, ubiquitous computing, mobile computing, embedded AI, wireless sensor networks, and big data analytics.

Prof. Gu is currently serving as an Editor of IMWUT, an Associate Editor of IEEE TRANSACTIONS ON MOBILE COMPUTING and IEEE INTERNET OF THINGS JOURNAL. Please find out more information at https://taogu.site/.

Yuxiang Wang received the M.Sc. degree from the School of Electronics Engineering and Computer Science, Peking University, Beijing, China, in 2019.

His research interests focus on wireless sensing and mobile computing.

Leye Wang (Member, IEEE) received the Ph.D. degree in computer science from the Institut Mines-Télécom SudParis, Évry, France, and the University of Paris, Paris, France, in 2016.

He was a Postdoctoral Researcher with the Hong Kong University of Science and Technology, Hong Kong. He is an Assistant Professor with the Key Lab of High Confidence Software Technologies, Ministry of Education, Peking University, and the Department of Computer Science and Technology, EECS, Peking University. His research interests include ubiquitous computing and crowdsensing.

Kai Niu received the M.E. degree in computer technology from the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China, in 2016. He is currently pursuing the Ph.D. degree in computer science with the School of Electronics Engineering and Computer Science, Peking University, Beijing, China.

His current research interests include ubiquitous computing, context-aware computing, and wireless sensing.

Cuntai Guan (Fellow, IEEE) received the Ph.D. degree from Southeast University, Nanjing, China, in 1993.

He is currently a Professor with the School of Computer Science and Engineering, and the Director of Artificial Intelligence Research Institute, Nanyang Technological University, Singapore. His current research interests include brain–computer interfaces, machine learning, neural signal and image processing, data analytics, and artificial intelligence.

Prof. Guan is a recipient of the Annual BCI Research Award, the IES Prestigious Engineering Achievement Award, the Achiever of the Year (Research) Award, and the Finalist of President Technology Award. He is an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE, and *Brain Computer Interfaces*. He is a Fellow of AIMBE.

Youwei Zeng received the B.E. degree in software engineering from Zhejiang University, Hangzhou, China, in 2016. He is currently pursuing the Ph.D. degree in computer science with the School of Electronics Engineering and Computer Science, Peking University, Beijing, China.

His current research interests include ubiquitous computing and mobile computing.

Daqing Zhang (Fellow, IEEE) received the Ph.D. degree from the University of Rome "La Sapienza," Rome, Italy, in 1996.

He is a Chair Professor with the Department of Computer Science and Technology, Peking University, Beijing, China, and Telecom SudParis, Palaiseau, France. He has published over 280 technical papers in leading conferences and journals. His current research interests include context-aware computing, urban computing, mobile computing, big data analytics, and pervasive elderly care.

Prof. Zhang was a recipient of the Ten-Years CoMoRea Impact Paper Award at IEEE PerCom 2013, the Honorable Mention Award at ACM UbiComp 2015 and 2016, and the Best Paper Award at IEEE UIC 2012 and 2015. He served as the General or Program Chair for over 17 international conferences, giving keynote talks at more than 20 international conferences. He is an Associate Editor for IEEE PERVASIVE COMPUTING, *ACM Transactions on Intelligent Systems and Technology*, and *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*.