Optimal Power Splitting for Simultaneous Wireless Information and Power Transfer in Millimeter-wave Networks

Yihan Liang, Yejun He*, and Jian Qiao
Guangdong Engineering Research Center of Base Station Antennas and Propagation
Shenzhen Key Laboratory of Antennas and Propagation
College of Electronics and Information Engineering, Shenzhen University, 518060, China
Email: 15875561843@163.com, heyejun@126.com*, 446941582@qq.com

***Abstract*—Simultaneous wireless information and power transfer (SWIPT)-enabled millimeter-wave (mmWave) network is one of the most effective solutions to solve the problem of high power consumption at wireless devices caused by high data rate applications. In this paper, we propose a SWIPT-enabled mmWave network and investigate the influence of mmWave propagation features on rate-energy (R-E) tradeoff of SWIPT system. In addition, an optimal power splitting (PS) policy is proposed to minimize the duration until battery exhausting, communication interruption and information loss occur. Finally, the proposed PS policy is modeled by Markov decision process (MDP) problem and realized by reinforcement learning (RL) algorithm. Simulation results show that the proposed RL-based PS policy can achieve higher battery energy level and stable data rate which can keep a good QoS of the whole SWIPT-enabled mmWave network.**

***Index Terms*—mmWave network, SWIPT, rate-energy tradeoff, power splitting, reinforcement learning, physical layer security.**

I. INTRODUCTION

With the evolution of wireless communication technology, the improvement of data rate will cause more energy consumption and greatly shorten the lifetime of battery in wireless devices. As a promising solution, wireless power transfer (WPT) has recently gained prominence due to its flexible application scenarios. WPT is usually implemented by electromagnetic induction (near-field) and electromagnetic radiation (far-field) [1]. Compared with wired charging, near-field WPT can not solve the energy shortage of battery and reduce the inconvenience of charging [2]. Far-field WPT, by contrast, is the technology in line with the concept of real wireless charging. SWIPT is the combination of information transfer and far-field WPT. It can effectively ease the contradiction between high data rate and long lifetime of battery-powered devices in the fifth generation (5G) wireless communication system.

The mmWave has been a promising technology in 5G wireless communication. mmWave frequencies range from 30 to 300 GHz, and its wavelength is in the order of 1 to 10 mm [3]. The battery powered users with ultra high transmission rate need more energy endorsement. SWIPT-enabled mmWave network is an effective solution to the problem of high power consumption caused by high data rate applications, such as unmanned aerial vehicle (UAV) [4] and Internet of things (IoT) [5]. Nevertheless, a large number of measurements show that the mmWave with the narrow beamwidth has good directivity, weak diffraction, less penetration ability as well as high propagation loss. Due to these unique characteristics of mmWave, line-of-sight (LOS) transmission is the main propagation mode. But mmWave link will fluctuate significantly in cases of users with high mobility so that the quality of service (QoS) will sharply cast down. In SWIPT-enabled mmWave network, we have to find a good method to improve QoS for mobile users which are suffering from frequent mmWave link fluctuations.

Thus, the resource allocation policy in SWIPT-enabled mmWave network is very vital. The mmWave radio frequency (RF) resource is limited. If more RF resource is used for information decoding to ensure the communication quality, the battery will run out due to insufficient energy supply. Conversely, the communication quality will be hardly guaranteed if more RF resource used for battery charging. Therefore, the tradeoff between transmission rate and energy level, namely R-E tradeoff, becomes a significant factor to evaluate the performance of SWIPT system.

Previous works on R-E tradeoff based on some classic receiver architectures, e.g., antenna switching (AS) architecture [6], time switching (TS) architecture and power splitting (PS) architecture, have been studied in [7]–[11]. In [7], a jointly optimization of power allocation and PS method to solve the tradeoff between information rate and harvested energy of SWIPT in mmWave massive MIMO-NOMA system is proposed. In order to achieve a higher spectrum and energy efficiency in the system, the optimal policy is derived to maximize the minimum harvested energy of the users. The proposed method can achieve higher spectrum and energy efficiency. To improve the system throughput in a three-node SWIPT system, paper [9] optimizes TS architecture by balancing the time duration between WPT phase and wireless communication phase to achieve the maximum throughput. In [11], a joint optimization of transmit power, time slot allocation, subcarrier as well as number of user antennas is

proposed, where the imperfect channel state information (CSI) is also considered. To the best of our knowledge, the most of the literatures on R-E tradeoff in SWIPT system only focus on the optimization of allocation algorithm in a fixed time slot when the transmission power is constant and the variation of the transmission power in the real transmission environment is fully ignored.

The contribution of this paper is summarized as follows.

1) A SWIPT-enabled mmWave network is proposed and the influence of mmWave propagation features on R-E tradeoff of SWIPT system is investigated.
2) An optimal PS policy is designed to minimize the duration until energy runs out, and communication interruption and information loss occur. As the received mmWave signal is unstable due to the mmWave propagation features mentioned above, the proposed PS policy is more practical in this paper and the whole moving process is considered when mobile users move in the mmWave network.
3) The proposed PS policy is modeled by MDP problem and an optimization method based on reinforcement learning is implemented.

The remainder of the paper is organized as follows. A more realistic system model of SWIPT-enabled mmWave network in LOS situation with PS architecture is introduced in Section II. An adaptive PS policy considering the frequent mmWave link fluctuations is proposed in Section III. In Section IV, an adaptive policy is realized by RL algorithm. In Section V, some intensive simulations are provided to evaluate the performance of the proposed policy. Finally, the paper is concluded in Section VI.

II. SYSTEM MODEL

We consider a SWIPT-enabled mmWave network as shown in Fig. 1. The network is composed of one mmWave base station (BS) and N mobile users $\mathcal{M} = \{M_1 \ldots, M_n, \ldots, M_N\}$ with rechargeable batteries. The BS is equipped with omnidirectional antenna and each user is equipped with single directional antenna. The PS architecture is deployed in each user which obtains the received instantaneous signal based on power level, i.e., one part used for information decoding and the other used for battery charging. To describe the fluctuations of mmWave link, we assume the network covered by BS is partitioned into G areas $\mathcal{U} = \{U_1, \ldots, U_g, \ldots, U_G\}$ with the same area size. When a user moves around the network, the time that a user stays in each area is defined as the resident duration. Thus, we assume that the instantaneous signal power received by each user in different areas is different, and the power value in a specific area U_g during resident duration is constant.

Each active M_n in the network has a specific signal power value in the area of its resident. The signal power value of all

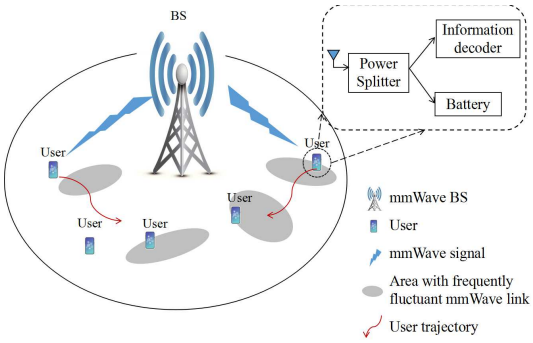

Fig. 1. SWIPT-enabled mmWave network with signal link strongly fluctuating.

N users in the area $\mathcal{U}$ can be described by

$$P = \begin{Vmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,G} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,G} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N,1} & p_{N,2} & \cdots & p_{N,G} \end{Vmatrix}, \tag{1}$$

where the signal power for each user can be derived from the block fading AWGN channel model of mmWave. For a specific M_n, it can be only located at one area U_g at a certain time which means it only has one non-zero value of each row vector. There are significant fluctuations of mmWave link but the case of $p_{n,g} = 0$ is not going to occur. Thus, P is a matrix with the number of N non-zero values.

Similarly, the resident duration of all N users in the area $\mathcal{U}$ is defined as

$$T = \begin{Vmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,G} \\ t_{2,1} & t_{2,2} & \cdots & t_{2,G} \\ \vdots & \vdots & \ddots & \vdots \\ t_{N,1} & t_{N,2} & \cdots & t_{N,G} \end{Vmatrix}, \tag{2}$$

where T is also a matrix with the number of N non-zero values. The resident duration is varying with the movement of each user in different areas. The resident duration can be described by $t_{n,g} = Xw_{n,g}$, where $w_{n,g}$ is the total number of frames of the specific duration $t_{n,g}$, and X is the time slot with a fixed value that each frame contains.

A. Power Splitter

When M_n moves from previous area to the next one at the beginning of each resident duration, M_n receives the signal power with the value of $p_{n,g}$. Then, the power splitter generates a specific PS parameter which is used for RF resource allocation. The matrix B to describe the PS parameters for all the N users in G areas is defined as

$$B = \begin{Vmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,G} \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,G} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{N,1} & \beta_{N,2} & \cdots & \beta_{N,G} \end{Vmatrix}, \tag{3}$$

where $\beta_{n,g}$ is the distribution ratio of the received signal power with the range of $(0,1)$. With AWGN channel, the data rate of M_n in U_g can be estimated by

$$r\left(\beta_{n,g}\right)=W\log\left(1+\frac{\left(1-\beta_{n,g}\right)p_{n,g}}{N}\right), \tag{4}$$

where W is the bandwidth of the signal and N is the background noise. The remaining signal power flows into the battery which is estimated by

$$e\left(\beta_{n,g}\right)=\alpha\beta_{n,g}p_{n,g}, \tag{5}$$

where $0<\alpha\leq 1$ denotes the power conversion efficiency.

B. Transmission Queue Model

A queue model of the battery and cache-aided decoder for each user is described as follows. The battery is assumed in the form of an energy queue. The battery charging rate $e\left(\beta_{n,g}\right)$ of M_n in area U_g can be derived from (5) and the energy consuming rate of M_n is defined as $c_{n,g}$. The battery state $\eta\left(\beta_{n,g}\right)$ denotes the energy level of battery at the beginning of each resident duration node when M_n moves into U_g. The battery size is not our concern. So the iterative formula for battery state is described by

$$\eta\left(\beta_{n,g'}\right)=\max\left\{\eta\left(\beta_{n,g}\right)-c_{n,g}t_{n,g},0\right\}+\ e\left(\beta_{n,g}\right)t_{n,g}. \tag{6}$$

The structure of the decoder cache is similar to the battery. It is in the form of a data queue. The cache size is a fixed value D. The signal resource used for information decoding is firstly stored briefly in the cache. The data arrival rate $r\left(\beta_{n,g}\right)$ can be derived from (4) and the required data rate for decoding is defined as $d_{n,g}$. The cache state $\pi\left(\beta_{n,g}\right)$ denotes the amount of data stored in the cache at the beginning of each resident duration when M_n moves into U_g. So the iterative formula for decoder state is described by

$$\pi\left(\beta_{n,g'}\right)=\min\left\{\begin{array}{l}\max\left\{\pi\left(\beta_{n,g}\right)-d_{n,g}t_{n,g},0\right\}+\\ r\left(\beta_{n,g}\right)t_{n,g},D\end{array}\right\}. \tag{7}$$

III. PROBLEM FORMULATION

The fundamental question is to find an optimal resource allocation policy to maintain a good QoS of moving users who are suffering from frequent mmWave link fluctuations. The PS parameter $\beta_{n,g}$ has a significant impact on QoS because improper allocation policy leads to three conditions of battery exhausting, communication interruption and information loss. In this section, we propose the quantitative indicators based on the changeable environment for the above three conditions.

We define the bad condition cost $w\left(\beta_{n,g}\right)$ as the normalized additional duration required when energy runs out, communication interruption or information loss of user in the whole area $\mathcal{U}$ occur. $w\left(\beta_{n,g}\right)=0$ denotes none of the energy, communication interruption or information loss. The three aspects are described as follows:

1) Battery exhausting: If the power consuming rate of M_n is lower than the charging rate $c_{n,g}<e\left(\beta_{n,g}\right)$, the QoS of M_n in U_g is guaranteed and $w\left(\beta_{n,g}\right)=0$ (the condition of energy overflowing in battery is not our concern). Conversely, if $c_{n,g}\geqslant e\left(\beta_{n,g}\right)$, the energy flowing into the battery will be gradually consumed. In this case, the time required from the beginning of M_n moves into U_g to the time node where the energy is gently exhausted is given by

$$\Delta\tau\left(\beta_{n,g}\right)=\frac{\eta_m\left(\beta_{n,g}\right)}{c_{n,g}-e\left(\beta_{n,g}\right)}. \tag{8}$$

If the resident duration $t_{n,g}$ is shorter than $\Delta\tau\left(\beta_{n,g}\right)$, the user will leave from U_g before the time node where the energy is completely exhausted and the bad condition will not occur $w\left(\beta_{n,g}\right)=0$. If $t_{n,g}\geqslant\Delta\tau\left(\beta_{n,g}\right)$, the extra energy which is required for user to prolong the lifetime is given by $\left(c_{n,g}-e\left(\beta_{n,g}\right)\right)\left(t_{n,g}-\Delta\tau\left(\beta_{n,g}\right)\right)$. So, the bad condition cost of battery exhausting is given by

$$w\left(\beta_{n,g}\right)=\frac{\left(c_{n,g}-e\left(\beta_{n,g}\right)\right)\left(t_{n,g}-\frac{\eta(\beta_{n,g})}{c_{n,g}-e(\beta_{n,g})}\right)}{e\left(\beta_{n,g}\right)}. \tag{9}$$

2) Communication interruption: In the decoder cache of M_n in U_g, if the speed of decoding is higher than the speed of data arrival $r\left(\beta_{n,g}\right)<d_{n,g}$, the data flowing into the decoder will be completely decoded immediately and there is no data queue waiting to be decoded in cache during the duration $t_(n,g)$. The duration from the beginning of M_n moving into U_g to the time node where there is no data to be decoded in the decoder is given by

$$\Delta\tau'\left(\beta_{n,g}\right)=\frac{\pi\left(\beta_{n,g}\right)}{d_{n,g}-r\left(\beta_{n,g}\right)}. \tag{10}$$

If resident duration $t_{n,g}$ is shorter than $\Delta\tau'\left(\beta_{n,g}\right)$, there is always information to be decoded in decoder and $w\left(\beta_{n,g}\right)=0$. If $t_{n,g}\geqslant\Delta\tau'\left(\beta_{n,g}\right)$, the communication interruption occurs and the extra information data required for decoding is given by $\left(d_{n,g}-r\left(\beta_{n,g}\right)\right)\left(t_{n,g}-\Delta\tau'\left(\beta_{n,g}\right)\right)$. So, the bad condition cost of communication interruption is given by

$$w\left(\beta_{n,g}\right)=\frac{\left(d_{n,g}-r\left(\beta_{n,g}\right)\right)\left(t_{n,g}-\frac{\pi(\beta_{n,g})}{d_{n,g}-r(\beta_{n,g})}\right)}{r\left(\beta_{n,g}\right)}. \tag{11}$$

3) Information loss: If $r\left(\beta_{n,g}\right)\geqslant d_{n,g}$, the arriving rate of M_n is higher than the data decoding rate. When the decoder cache is full of data, the data that comes later would be abandoned. The duration from the beginning of M_n moving into U_g to the time node where the cache is full is given by

$$\Delta\tau''\left(\beta_{n,g}\right)=\frac{D-\pi\left(\beta_{n,g}\right)}{r\left(\beta_{n,g}\right)-d_{n,g}}. \tag{12}$$

If $t_{n,g}$ is shorter than $\Delta\tau''\left(\beta_{n,g}\right)$, then $w\left(\beta_{n,g}\right)=0$ because the user will leave from area U_g before the time node where the information loss occurs. If $t_{n,g}\geqslant\Delta\tau''\left(\beta_{n,g}\right)$, the abandoned information during the resident duration is

$(r(\beta_{n,g}) - d_{n,g})(t_{n,g} - \Delta\tau''(\beta_{n,g}))$. The bad condition cost of information loss is given by

$$w(\beta_{n,g}) = \frac{(r(\beta_{n,g}) - d_{n,g})\left(t_{n,g} - \frac{D-\pi(\beta_{n,g})}{r(\beta_{n,g})-d_{n,g}}\right)}{r(\beta_{n,g})}. \quad (13)$$

To summarize the three aspects mentioned above, the proposed PS algorithm for SWIPT-enabled mmWave network is summarized in Algorithm 1. Summing up the above, the bad condition cost of three aspects of M_n in area U_g is described by (14).

Algorithm 1 Proposed PS algorithm for SWIPT-enabled mmWave network

if $c_{n,g} \geqslant e(\beta_{n,g})$ **then**
  Compute exhausting time $\Delta\tau(\beta_{n,g})$ according to (8) and compare it with $t_{n,g}$
  if $t_{n,g} \geqslant \Delta\tau(\beta_{n,g})$ **then**
    Get $w(\beta_{n,g})$ according to (9)
  else
    if $d_{n,g} \geqslant r(\beta_{n,g})$ **then**
      Compute interruption time $\Delta\tau'(\beta_{n,g})$ according to (10) and compare it with $t_{n,g}$
      if $t_{n,g} < \Delta\tau'(\beta_{n,g})$ **then**
        $w(\beta_{n,g}) \leftarrow 0$
      else
        Get $w(\beta_{n,g})$ according to (11)
      end if
    else
      Compute information loss time $\Delta\tau''(\beta_{n,g})$ according to (12) and compare it with $t_{n,g}$
      if $t_{n,g} < \Delta\tau''(\beta_{n,g})$ **then**
        $w(\beta_{n,g}) \leftarrow 0$
      else
        Get $w(\beta_{n,g})$ according to (13)
      end if
    end if
  end if
else
  Repeat step 7 to 23
end if

IV. RL-BASED ADAPTIVE PS POLICY

Our goal is to find an adaptive PS policy to minimize the bad condition cost $w^*(\beta_{n,g})$ for each user during the whole duration when they move in the network. This is a sequential decision problem to find a proper control at each moment to achieve a certain optimal operation effect on the whole process of system. The problem can be described under a MDP [12] framework. It can help the dynamic system to find an optimal decision based on Markov process theory. The basic elements of an MDP framework can be defined in the form of a set $\{S, A, P, R\}$. Respectively, S is the state of the environment, A is the action of the agent, P is the state transmission probability and R is the reward or punishment of the action.

The RL algorithm is a promising method of MDP problem. Q-learning is one of the computational efficient model-free algorithm of RL. In Q-learning algorithm, some difficult information such as the state transmission probability P is not required. We define each user is the independent learning agent of the Q-learning system. Each agent uses Q-learning algorithm to learn the changeable environment and find the optimal PS parameter for each state. In the Q-learning process, the agent needs to take action according to the current state, and improves the action after receiving corresponding reward. So that the agent can make better action when in the same state for the next time. Q-value is the most element of Q-learning algorithm. It represents the maximum future reward expectation for a given state and corresponding action. The update of Q-value is given by

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma \max_a Q(s',a) - Q(s,a)\right]. \quad (15)$$

A. Q-learning

To use the Q-learning algorithm, we firstly define the three elements– the state space, the action space and the reward of the proposed SWIPT-enabled mmWave network as follows.

1) The state space: We define that a system state transition occurs when any user in the network moves from the area U_g into the neighboring area U_g'. Assume that the associated areas of any two users do not change at the same time. Therefore, the state time can be described by the resident duration of the user who changes the location at this state. When a new state occurs, both the power of the received signal power and the resident duration change. Thus, we define the state space of SWIPT-enabled mmWave network as

$$s = (P_s, T_s) \in \mathcal{S} = \mathcal{P} \times \mathcal{T}, \quad (16)$$

where $\mathcal{P} = \{P_1, \ldots, P_s, \ldots, P_{\mathrm{S}}\}$ and $\mathcal{T} = \{T_1, \ldots, T_s, \ldots, T_{\mathrm{S}}\}$ are the signal power and resident duration of all S states, respectively. For a specific state s, P_s is the signal power of all N users in the whole area $\mathcal{U}$ which can be obtained according to (1). And the resident duration of the users can be achieved according to (2).

2) The action space: When any user moves into a new area, the system transition occurs. A new PS parameter is required due to the changes in the received signal power and resident duration. Therefore, the action space of the network can be described by

$$a = (B_s) \in \mathcal{A} = \mathcal{B}, \quad (17)$$

where $\mathcal{B} = \{B_1, \ldots, B_s, \ldots, B_{\mathrm{S}}\}$ is the PS policy of all S states. For a specific state s, B_s is the selected PS parameter of N users in the whole G areas which can be achieved from (3). Then, we use the greedy policy to select action for each user in each state

$$\beta_s \begin{cases} = \arg\max\ Q(s,a), & \text{with probability} \quad 1-\varepsilon \\ \sim U(A), & \text{with probability} \quad \varepsilon \end{cases} \quad (18)$$

$$w\left(\beta_{n,g}\right)=\begin{cases}\frac{\left(c_{n,g}-e(\beta_{n,g})\right)\left(t_{n,g}-\frac{\eta(\beta_{n,g})}{c_{n,g}-e(\beta_{n,g})}\right)}{e(\beta_{n,g})}, & \left(c_{n,g}\geqslant e\left(\beta_{n,g}\right),t_{n,g}\geqslant\frac{\eta(\beta_{n,g})}{c_{n,g}-e(\beta_{n,g})}\right)\\ \frac{\left(d_{n,g}-r(\beta_{n,g})\right)\left(t_{n,g}-\frac{\pi(\beta_{n,g})}{d_{n,g}-r(\beta_{n,g})}\right)}{r(\beta_{n,g})}, & \left(c_{n,g}<e\left(\beta_{n,g}\right),r\left(\beta_{n,g}\right)<d_{n,g},t_{n,g}\geqslant\frac{\pi(\beta_{n,g})}{d_{n,g}-r(\beta_{n,g})}\right)\\ \frac{\left(r(\beta_{n,g})-d_{n,g}\right)\left(t_{n,g}-\frac{D-\pi(\beta_{n,g})}{r(\beta_{n,g})-d_{n,g}}\right)}{r(\beta_{n,g})}, & \left(c_{n,g}<e\left(\beta_{n,g}\right),r\left(\beta_{n,g}\right)\geqslant d_{n,g},t_{n,g}\geqslant\frac{D-\pi(\beta_{n,g})}{r(\beta_{n,g})-d_{n,g}}\right)\\ 0, & otherwise\end{cases}\tag{14}$$

3) Reward: Reward is a key factor that determines the performance of Q-learning algorithm. In SWIPT-enabled mmWave network, we define the negative value of the total bad conditions cost of all N users as the reward which is described by

$$r_s=-\sum_{n=1}^{N}w\left(\beta_{n,g}\right).\tag{19}$$

Thus, in the cases of no energy running out, no communication interruption and no information loss at state s, the reward is 0. In the other cases, the rewards are always negative.

4) Proposed algorithm: Algorithm 2 is the proposed Q-learning Algorithm for SWIPT-enabled mmWave network. After k_{max} iterative learning, the optimal PS policy can be described by

$$\beta_s^*=\arg\max Q^*\left(s,a\right).\tag{20}$$

Algorithm 2 Q-learning Algorithm for SWIPT-enabled mmWave network

```
Initialize Q(s, a) for s and a arbitrarily
for k = 1 to k_max do
    Initialize s
    while s ≠ S do
        Select a according to (18)
        Take action a, observe r, s'
        Update Q(s, a) using (15)
        Set s ← s';
    end while
end for
```

V. Simulation Results

The performance of the proposed PS policy of SWIPT-enabled mmWave network is evaluated by simulations. The parameter settings are as follows: The transmission power of each BS is $P_T=40\ W$; the coverage of the network is $d=50m$. Then the received signal power can derived from the free space transmission model. The spectrum bandwidth $W=2\ GHz$; the background noise $N=-134\ bM/MHz$; the antennas gains are $G_t=G_r=1$. To simplify the computational model, we set $G=100$ and there is one user in the network. The user is randomly moving among these 100 areas, and we pick 15 areas of these at random as areas through which the user is moving. The resident durations depend on the speed of user which is randomly selected among $[1,100]$. The required decoding rate is randomly selected with the range of $[90,120]\ Gbps$. The consuming rate is set with the low level of $4\times10^{-7}\ W$ according to the practical situation of wireless power transfer. The size of the decoder cache is $16\times10^9\ bit$, and the battery size don't need to be considered. The PS parameters is a set of discrete equally interval value with the range of $(0,1)$. As for the Q-learning algorithm of SWIPT-enabled mmWave network, the parameter settings are as follows: $\alpha=0.8$, $\gamma=0.9$, $k_{max}=1000$, and the parameter setting of greedy algorithm when choose action is $\varepsilon=0.5$.

We focus on the communication environment of the 15 states selected in the network. Fig. 2 shows the received signal power value and resident duration of each state. It can be found through the change of power values that the signal received by the user during the movement of the network fluctuate seriously. This is due to the mmWave link fluctuations. What's more, in order to describe the random movement of user in mmWave network more accurately, the 15 resident durations of user in 15 areas also vary randomly. In the simulations, the

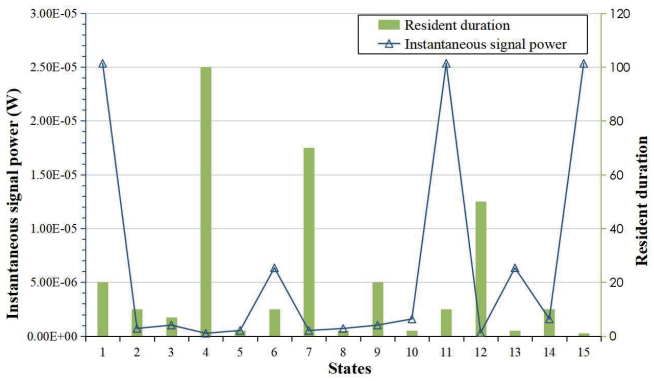

Fig. 2. 15 randomly selected states in mmWave network with signal link strongly fluctuating.

user is assumed to pass through the 15 areas of the 15 states in turn. It will experience serious signal fluctuations described in Fig. 2. The received power and resident duration are all the influence factors of PS allocation strategy. Fig. 3 shows the battery states against the 15 states. It can be found that due to the low energy consuming rate, all the PS policies of 15 states can ensure the continuous increase of battery state during the whole movement. In the naive policies, an increase in PS parameter β means more signal resource to be used for the battery charging. Thus, the battery state of the naive

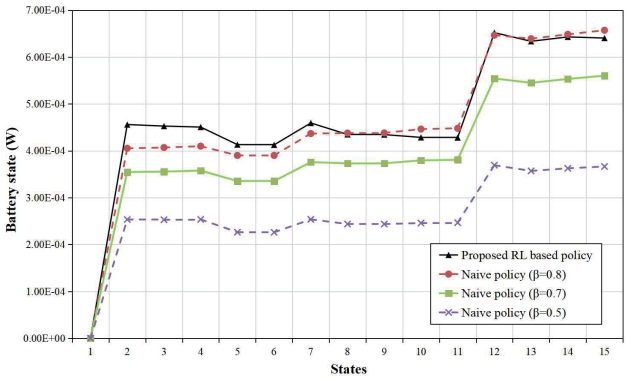

Fig. 3. Battery states of the user when moving in mmWave network.

policy with $\beta = 0.5$ is the lowest and the naive policy with $\beta = 0.8$ is the highest. It is observed that our proposed RL based policy can reach to the highest level of the three naive policies. Note that, the naive policy with $\beta = 0.8$ gives the priority to energy sustainability which means eighty percent of signal resource to be used for battery charging. Fig. 4 shows

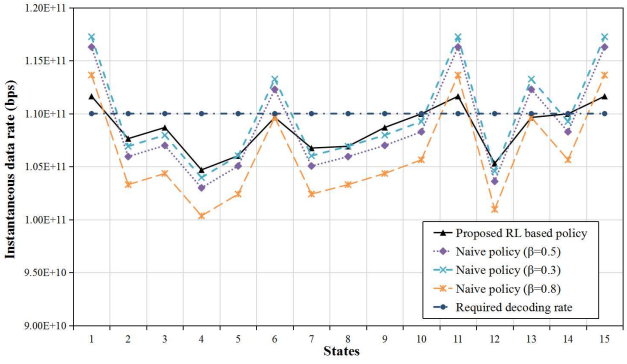

Fig. 4. Data rate of the user when moving in mmWave network.

the data rate against 15 states, where the data rate values of all PS policies fluctuate to some extent because the received signals fluctuate in Fig. 2. Since the size of decoder cache is limited and the required decoding rate is fixed, the optimal PS policy should try to keep the values of data rate within a small range around the required decoding rate. No matter how the value of PS parameter β is set, the QoS cannot be guaranteed in a series of random states, such as the three naive policies. If β is too high, there is a risk of information loss; if β is too low, there is a risk of communication interruption. Our proposed RL based PS policy can find the optimal PS policy for each state through the continuous learning of the environment. It has the smallest fluctuation and the minimum deviation value in each state compared with the other three naive policies. Fig. 3 and Fig. 4 together show our proposed PS policy not only ensures the best communication quality but also keeps the battery state at the highest level at the same time compared with the other three naive policies.

VI. Conclusion

In this paper, we focus on the design of PS policy in SWIPT-enabled mmWave network by considering the frequent mmWave channel fluctuations. The optimal PS policy is useful to minimize the probability of each user suffering bad conditions such as battery exhausting, communication interruption or information loss. Moreover, the proposed PS policy is modeled by MDP problem and realized by RL algorithm. Compared with the naive PS policies that only consider the advantages of one side based on fixed environment conditions, the proposed adaptive PS policy can make decision according to different environment conditions. A higher battery level and stable data rate can be guaranteed when the users move in the mmWave network. In the next step, physical layer security for SWIPT-enabled mmWave network will be also further studied.

VII. Acknowledgment

This work is supported in part by the program of the Ministry of Science and Technology of China under Grant RW2019TW001, in part by the Shenzhen Science and Technology Program under Grants GJHZ 20180418190529516 and JSGG20180507183215520.

References

[1] C. Liu, M. Maso, S. Lakshminarayana, C. Lee, and T. Q. S. Quek, "Simultaneous wireless information and power transfer under different csi acquisition schemes," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 1911–1926, Apr. 2015.

[2] M. Xia and S. Aissa, "On the efficiency of far-field wireless power transfer," *IEEE Transactions on Signal Processing*, vol. 63, pp. 2835–2847, Jun. 2015.

[3] T. A. Khan, A. Alkhateeb, and R. W. Heath, "Millimeter wave energy harvesting," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 6048–6062, Sep. 2016.

[4] X. Sun, W. Yang, Y. Cai, R. Ma, and L. Tao, "Physical layer security in millimeter wave swipt uav-based relay networks," *IEEE Access*, vol. 7, pp. 35851–35862, 2019.

[5] Y. Huang, M. Liu, and Y. Liu, "Energy-efficient swipt in iot distributed antenna systems," *IEEE Internet of Things Journal*, vol. 5, pp. 2646–2656, Aug. 2018.

[6] F. Benkhelifa and M. Alouini, "Prioritizing data/energy thresholding-based antenna switching for swipt-enabled secondary receiver in cognitive radio networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, pp. 782–800, Dec. 2017.

[7] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive mimo-noma with simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 37, pp. 131–141, Jan. 2019.

[8] D. Zhai, R. Zhang, J. Du, Z. Ding, and F. R. Yu, "Simultaneous wireless information and power transfer at 5g new frequencies: Channel measurement and network design," *IEEE Journal on Selected Areas in Communications*, vol. 37, pp. 171–186, Jan. 2019.

[9] F. Zhao, L. Wei, and H. Chen, "Optimal time allocation for wireless information and power transfer in wireless powered communication systems," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 1830–1835, Mar. 2016.

[10] Y. Liu, "Wireless information and power transfer for multirelay-assisted cooperative communication," *IEEE Communications Letters*, vol. 20, pp. 784–787, Apr. 2016.

[11] Z. Chang, Z. Wang, X. Guo, C. Yang, Z. Han, and T. Ristaniemi, "Distributed resource allocation for energy efficiency in ofdma multicell networks with wireless power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 37, pp. 345–356, Feb. 2019.

[12] L. Liu and G. S. Sukhatme, "A Solution to Time-Varying Markov Decision Processes," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1631–1638, Jul. 2018.