Optimal Power Allocation for Wireless Sensor Powered by Dedicated RF Energy Source

Qizhen Li, Jie Gao, *Member, IEEE,* Hongbin Liang, *Member, IEEE,* Lian Zhao, *Senior Member, IEEE*, and Xiaohu Tang, *Senior Member, IEEE*

***Abstract*—This paper studies a wireless-powered sensor network, where a sensor harvests energy from a dedicated radio-frequency (RF) energy source and transmits information to an information sink using the harvested energy. Two working modes are considered. One is the frequency division multiplexing (FDM) mode in which the sensor harvests RF energy and transmits information simultaneously over orthogonal frequency bands. The other is the time division multiplexing (TDM) mode in which energy harvesting and information transmission are implemented in the same frequency band but in different time slots. The energy harvesting channel and the information transmission channel are assumed to follow the Rician and the Rayleigh distributions, respectively, and are discretized and modeled as finite-state Markov chains. We formulate the process of energy harvesting and information transmission as an infinite-horizon discounted Markov decision process (MDP). The value iteration algorithm is used to find an asymptotically optimal energy harvesting and information transmission policy to optimize the long-term throughput. In the asymptotically optimal policy of the FDM mode, the energy transmitted from the sensor in one slot is proved to be non-decreasing with the battery state of the sensor. By contrast, such monotonicity between the transmitted energy and the battery state does not exist in the asymptotically optimal policy in the TDM mode. Simulation results verify the above findings and demonstrate that the proposed method outperforms the heuristic greedy method.**

***Index Terms*—Wireless-powered sensor network, dedicated RF energy source, Markov decision processes, asymptotically optimal policy, monotonicity.**

I. Introduction

PROLONGING the lifetime of sensors has always been a research focus for wireless sensor networks, especially for the networks with sensors powered by the limited battery energy. Manually replacing batteries or replenishing energy can be infeasible or costly in some cases, e.g., when sensors are embedded inside concrete walls or human bodies. In addition to designing efficient data transmission mechanisms for saving energy [1]–[3], harvesting renewable energy also emerges as an appealing solution to the battery depletion problem [4]–[8]. For energy harvesting, compared with other renewable energy (e.g., wind energy and solar energy), the radio-frequency (RF) energy [9] is more reliable and manageable because it is less constrained by time and environments.

Wireless sensors equipped with RF energy receivers can harvest RF energy to process signals and transmit information. The harvested energy can be stored in energy storages, such as rechargeable batteries and capacitors. Since the RF signals attenuate severely in propagation [10], improving the efficiency of energy harvesting is of great challenge in wireless-powered sensor networks.

Wireless-powered communication networks (WPCN) have been extensively researched [11]. Typically, the RF energy transmitter and the information receiver are co-located in a communication node, namely, hybrid access point (H-AP) [12]–[15]. In [12] and [13], users only harvest RF energy from the H-AP. The individual data of each user is transmitted via time division multiple access (TDMA) in [12], while users transmit information cooperatively in [13]. In [14], a user can harvest energy from the RF signals transmitted by the H-AP and other users. In [15], all users first harvest RF energy from the H-AP, and then simultaneously transmit information to the H-AP. Nevertheless, in the above WPCN, users may not obtain enough RF energy due to the large distance between the users and the H-AP.

The WPCN in which the RF energy transmitters are separated from the information receivers are more practical than the WPCN with H-AP in real-life applications because the RF energy transmitters can be close to users [16]–[20]. In some of the WPCN with dedicated RF energy sources such as that in [16], the optimal power allocation policy is obtained assuming that the channel information is known in advance for several time slots, which may not be reasonable in time-varying wireless communication networks. In other WPCN with dedicated RF energy sources, such as those in [17]–[20], users first receive and store the RF energy, and then use up the harvested energy to transmit information within a time slot. However, the optimal long-term power allocation policies are not obtained in these works. Therefore, finding a dynamic power allocation method for WPCN with separated RF energy transmitter and information receiver to maximize the long-term throughput remains an open problem.

This paper investigates a wireless-powered sensor network which consists of one dedicated RF energy source, one sensor,

This work was supported by the National Natural Science Foundation of China under Grant 61728108, Grant 61571375, and Grant 61871331. (*Corresponding author: Hongbin Liang.*)

Qizhen Li is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China, and also with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (email: liqizhen@my.swjtu.edu.cn).

Jie Gao and Lian Zhao are with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (email: {j.gao, l5zhao}@ryerson.ca).

Hongbin Liang is with the School of Transportation and Logistics, National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu 610031, China (e-mail: hbliang@swjtu.edu.cn).

Xiaohu Tang is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China (e-mail: xhutang@swjtu.edu.cn).

and one information sink. The sensor transmits information to the information sink using the energy harvested from the dedicated RF energy source. Two working modes are considered: the frequency division multiplexing (FDM) mode and the time division multiplexing (TDM) mode. In the FDM mode, the sensor harvests RF energy and transmits information simultaneously over orthogonal frequency bands. In the TDM mode, the sensor harvests RF energy and transmits information in the same frequency band but in different time slots. We aim to maximize the long-term throughput of the information transmission link in these two modes. The main contributions of this paper are as follows:

- We model the small-scale fading of the energy harvesting channel and the information transmission channel as finite-state Markov chains (FSMC) and derive the steady-state probabilities, the average power gains of states, and the state transition probabilities of the FSMC.
- We formulate the process of energy harvesting and information transmission as an infinite-horizon Markov decision process (MDP). We use the value iteration algorithm to obtain an asymptotically optimal energy harvesting and information transmission policy for the optimal long-term throughout.
- For the FDM mode, we prove that the energy transmitted from the sensor is non-decreasing with the battery state in the asymptotically optimal policy. This property can can be used to reduce the computational complexity. By contrast, in the TDM mode, there is no monotonicity between the transmitted energy and the battery state in the asymptotically optimal policy . We analyze the above difference and give insights regarding the effect of the working modes on the optimal policy.

The advantages of our solution are as follows. Compared with the wireless-powered sensor networks with the H-AP, the sensor is powered by the dedicated RF energy source in our solution to guarantee an effective energy supply. Compared with other power allocation schemes of the dedicated RF energy powered communication networks, we consider the correlation of channel states between adjacent time slots, thereby obtaining the optimal energy harvesting and information transmission policy for the long run.

The rest of this paper is organized as follows. Section II describes the system model. Section III formulates the problem as an MDP and optimizes the energy harvesting and information transmission policy. Section IV discusses the monotonic property of the asymptotically optimal policy. Simulation results are given in section V. Conclusions are presented in section VI.

II. System Model

We consider a wireless-powered sensor network with one dedicated RF energy source, one sensor, and one information sink. The sensor harvests energy from the dedicated RF energy source and transmits information to the information sink using the harvested energy. Two working modes are studied, i.e., the FDM mode (Fig. 1) and the TDM mode (Fig. 2)[1]. In the FDM mode, the sensor uses different antennas to simultaneously harvest energy and transmit information over orthogonal frequency bands to avoid the interference between the energy harvesting link and the information transmission link. In the TDM mode, the sensor uses one antenna to harvest RF energy or transmit information in the same frequency band but in different time periods to avoid mutual interference between the two links.

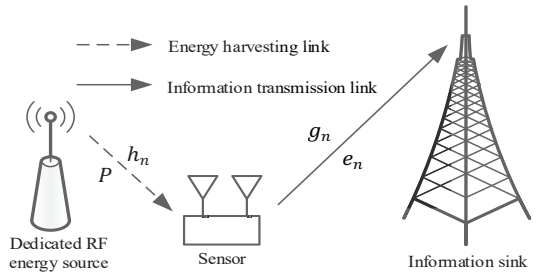


Fig. 1. FDM working mode.

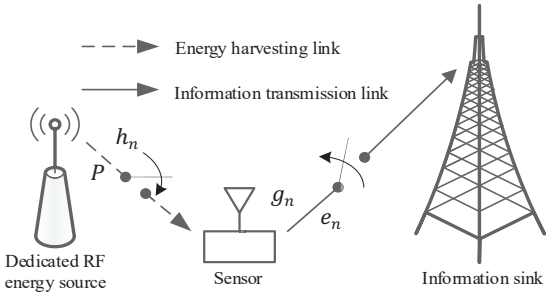


Fig. 2. TDM working mode.

We assume that the wireless-powered sensor network operates in a time-slotted manner. The duration of each time slot is τ. The dedicated RF energy source transmits energy with fixed power P if the sensor decides to harvest energy from it. The capacity of the sensor's rechargeable battery is B. We assume that the sensor has infinite data to be transmitted and the energy for sensing is negligible compared with that for transmitting information.

The energy harvesting channel and the information transmission channel are assumed to be block-fading, namely, the channels remain constant within one slot but may change between adjacent slots. We consider the large-scale fading part and the small-scale fading part for either of the channels. The large-scale fading part is summarized into path-loss which relies mainly on the propagation distance and the fading environment. In the simplified large-scale fading model, the path-loss is [10]:

$$P_L = C\left(\frac{d_0}{d}\right)^{\theta} \tag{1}$$

where C is the unitless constant, d_0 is the reference distance, d is the distance between the transmitter and the receiver,

[1]FDM and TDM are the two most popular work modes in theoretical research and practical application because of their low complexity in signal processing. Therefore, both modes are studied in this work.

and θ is the path-loss exponent. Generally, the values of C, d_0 and θ can be measured by the experiment. We denote the path-losses of the energy harvesting channel and the information transmission channel as P_L^H and P_L^G, respectively. Since the sensor is close to the dedicated RF energy source, the line-of-sight (LOS) path of the energy harvesting channel should be considered. Thus, the envelope of the small-scale fading part of the energy harvesting channel is assumed to have a Rician distribution. For the information transmission channel, the sensor is far away from the information sink. The scattering environment is complex and the LOS path can be neglected. Therefore, the small-scale fading part of the information transmission channel is modeled as a Rayleigh fading channel.

The signal received by the sensor at time t ($t \in (0, \tau]$) of the nth time slot can be formulated as:

$$y_s(t) = \sqrt{P_L^H P} h'_n x(t) + z_s(t) \tag{2}$$

where h'_n is the complex small-scale fading gain of the energy harvesting channel, and $x(t)$ is the signal transmitted from the dedicated RF energy source (with the power $|x(t)|^2$ equal to 1). The signal $x(t)$ carries pilots which are used for channel estimation. In general, the noise $z_s(t)$ can be ignored in the energy harvesting link. The RF energy conversion efficiency of the sensor is assume to be η. Then the harvested power of the sensor in the nth time slot can be represented as:

$$P_h(n) = \eta P_L^H P h_n \tag{3}$$

where $h_n = |h'_n|^2$ is the normalized power gain (i.e., the expectation of h_n is equal to 1) of the small-scale fading part for the energy harvesting channel. The sensor can only use the energy harvested in the previous slots.

We assume the power transmitted from the sensor in the nth time slot is e_n. The received signal of the information sink at time t of the nth time slot can be formulated as:

$$y_i(t) = \sqrt{P_L^G e_n} g'_n s(t) + z_i(t) \tag{4}$$

where g'_n is the complex small-scale fading gain of the information transmission channel, and $s(t)$ is the normalized signal transmitted from the sensor. In the information transmission channel, the noise $z_i(t)$ can not be ignored. We denote the noise power spectral density and the bandwidth of the information transmission channel by N_0 and W, respectively. On account of practical modulation and coding, there exists a gap between the achievable rate and the channel capacity, denoted as γ (larger than 1) [21]. The data rate received by the information sink in the nth time slot can be formulated as [22]:

$$r_n = W \log_2 \left(1 + \frac{P_L^G e_n g_n}{N_0 W \gamma}\right) \tag{5}$$

where $g_n = |g'_n|^2$ is the normalized power gain of the small-scale fading part for the information transmission channel.

In the FDM mode, the sensor determines the amount of power to transmit information at the beginning of a slot. In the TDM mode, the sensor needs to determine whether to harvest energy or transmit information, and, if transmitting, how much power to use. The sensor should choose the energy harvesting and information transmission scheme according to the remaining energy in the battery, the power gains of the energy harvesting channel and the information transmission channel. We aim to find an optimal adaptive energy harvesting and information transmission policy to maximize the total discounted throughput for the long run:

$$R_{max} = \max \lim_{N \to \infty} \sum_{n=0}^{N} \lambda^n r_n \tag{6}$$

where $\lambda \in (0, 1)$ is the discount factor.

III. MDP Formulation and Policy Optimization

In this section, we model the process of energy harvesting and information transmission using an infinite-horizon MDP. Generally, an MDP model comprises five elements, i.e., decision epoch, state space, action space, state transition probability, and reward function [23], [24]. In the rest of this section, we first formulate these five elements for the considered wireless-powered sensor network and then adopt the value iteration algorithm to maximize the total discounted throughput of the information transmission link.

A. Decision Epoch

As stated in the above section, the time is divided into slots. Each slot has the same duration in which several packets can be transmitted. We define the beginning of a slot as the decision epoch of that slot. At each decision epoch, the sensor executes the energy harvesting and information transmission policy according to the energy in the battery, the power gain of the energy harvesting channel, and the power gain of the information transmission channel.

B. State Space and Action Space

We divide the small-scale fading parts of the energy harvesting channel and the information transmission channel into several states. The state sets of the two fading channels can be respectively formulated as $\mathcal{S}_h = \{0, 1, \ldots, N_h - 1\}$ and $\mathcal{S}_g = \{0, 1, \ldots, N_g - 1\}$, where N_h is the number of states of the energy harvesting channel and N_g is the number of states of the information transmission channel. The energy capacity of the battery in the sensor is divided into N_b states and the battery state set is $\mathcal{S}_b = \{0, 1, \ldots, N_b - 1\}$. We define $\mathcal{S}$ as the state space including the three state sets, i.e., $\mathcal{S} = \mathcal{S}_h \times \mathcal{S}_g \times \mathcal{S}_b$, where the symbol $\times$ denotes the Cartesian product.

We assume that the basic unit of power that the sensor can receive or transmit is $P_u = E_u/\tau = B/((N_b - 1)\tau)$, where $E_u = B/(N_b - 1)$ is the minimum unit of energy that can be received or transmitted in one slot. We denote $\mathcal{A}_s = \{0, 1, \ldots, s_b\}$ as the action space given the system state $s = (s_h, s_g, s_b) \in \mathcal{S}$. Here, $s_h \in \mathcal{S}_h$ and $s_g \in \mathcal{S}_g$ are the energy harvesting channel state and the information transmission channel state, respectively, and $s_b \in \mathcal{S}_b$ is the battery state, i.e., the number of unit energy at the beginning of the current slot. For the FDM mode, $a \in \mathcal{A}_s \backslash \{0\}$ denotes the action that the sensor uses a units of energy to transmit information in one slot, and $a = 0$ represents the action that

the sensor does not transmit information. For the TDM mode, $a \in \mathcal{A}_s \backslash \{0\}$ also represents the action that the sensor uses a units of energy to transmit information in one slot, except when $a = 0$, which denotes the action that the sensor receives energy from the dedicated RF energy source.

C. Channel State and State Transition Probability

In this subsection, we first study the steady-state probabilities and the state transition probabilities of the energy harvesting channel and the information transmission channel, respectively. Then, we present the overall system state transition probabilities.

As mentioned in the previous section, the energy harvesting channel is assumed to be Rician faded. The power gain h of a Rician fading channel obeys a noncentral Chi-square (χ^2) distribution with two degrees of freedom [25]. The probability density function (PDF) of h is:

$$f_H(h) = (K+1) I_0 \left(2\sqrt{K(K+1)h} \right) e^{-(K+1)\left(h + \frac{K}{K+1}\right)} \quad (7)$$

where the parameter K (the K-factor) is the ratio of the energy in the LOS path to the energy in the scattered paths, and $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind.

The cumulative distribution function (CDF) of the channel power gain can be formulated as:

$$F_H(h) = \begin{cases} 1 - Q_1\left(\sqrt{2K}, \sqrt{2(K+1)h}\right), & h > 0 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $Q_1(\mu, \nu)$ is the *Marcum Q function* defined as

$$Q_1(\mu, \nu) = \int_{\nu}^{\infty} x e^{-\frac{\mu^2 + x^2}{2}} I_0(\mu x) dx.$$

For the energy harvesting channel, let $\{A_h^{(i)}\}$, $i = 0, 1, \ldots, N_h$, be the quantization thresholds of the channel power gain corresponding to the small-scale fading part, where $A_h^{(i)} < A_h^{(i+1)}$, $A_h^{(0)} = 0$ and $A_h^{(N_h)} = \infty$[2]. If the channel power gain is in the interval $[A_h^{(i)}, A_h^{(i+1)})$, the channel state s_h is then equal to i. The steady-state probability that the channel is in the state $s_h = i$ can be shown as:

$$\begin{aligned} Pr(s_h = i) &= F_H\left(A_h^{(i+1)}\right) - F_H\left(A_h^{(i)}\right) \\ &= Q_1\left(\sqrt{2K}, \sqrt{2(K+1)A_h^{(i)}}\right) \\ &\quad - Q_1\left(\sqrt{2K}, \sqrt{2(K+1)A_h^{(i+1)}}\right). \end{aligned} \quad (9)$$

The expected channel power gain conditioned on a state is the average channel power gain of this state. For instance, the average channel power gain of the state i in the energy harvesting channel is:

$$\mathrm{E}[h|s_h = i] = \frac{\int_{A_h^{(i)}}^{A_h^{(i+1)}} h f_H(h) dh}{\int_{A_h^{(i)}}^{A_h^{(i+1)}} f_H(h) dh} = \frac{\int_{A_h^{(i)}}^{A_h^{(i+1)}} h(K+1) I_0\left(2\sqrt{K(K+1)h}\right) e^{-(K+1)\left(h + \frac{K}{K+1}\right)} dh}{Q_1\left(\sqrt{2K}, \sqrt{2(K+1)A_h^{(i)}}\right) - Q_1\left(\sqrt{2K}, \sqrt{2(K+1)A_h^{(i+1)}}\right)}. \quad (10)$$

Before deriving the transition probabilities between channel states, we introduce an important concept, i.e., the level crossing rate (LCR) that is the number of passes for a level in either the positive or the negative going position per unit time. Suppose that the maximum Doppler frequency of the energy harvesting channel is D_h. The LCR of the threshold $A_h^{(i)}$ can be formulated as [27]:

$$\begin{aligned} N(A_h^{(i)}) &= \sqrt{2\pi(K+1)} D_h \rho I_0\left(2\rho\sqrt{K(K+1)}\right) e^{-K-(K+1)\rho^2} \\ &= \sqrt{2\pi(K+1)A_h^{(i)}} D_h I_0\left(2\sqrt{K(K+1)A_h^{(i)}}\right) e^{-K-(K+1)A_h^{(i)}} \end{aligned} \quad (11)$$

where $\rho = \sqrt{A_h^{(i)}/\sigma^2} = \sqrt{A_h^{(i)}}$ and σ^2 (normalized to 1) is the average power of small-scale fading channel.

In the considered sensor network, the channel changes slowly. Given a channel state in the current time slot, we assume that the energy harvesting channel in the next slot is in that state or the adjacent states. The channel state transition probabilities can be approximated as [28]:

$$\begin{aligned} &Pr(s_h = i'|s_h = i) \\ &= \begin{cases} \frac{N(A_h^{(i+1)})\tau}{Pr(s_h = i)}, i' = i+1, \ i = 0, \ldots, N_h - 2 \\ \frac{N(A_h^{(i)})\tau}{Pr(s_h = i)}, i' = i-1, \ i = 1, \ldots, N_h - 1 \\ 1 - \frac{N(A_h^{(i+1)})\tau}{Pr(s_h = i)} - \frac{N(A_h^{(i)})\tau}{Pr(s_h = i)}, i' = i, \ i = 1, \ldots, N_h - 2. \end{cases} \end{aligned} \quad (12)$$

In particular, the values of $Pr(s_h = 0|s_h = 0)$ and $Pr(s_h = N_h - 1|s_h = N_h - 1)$ are given by:

$$Pr(s_h = 0|s_h = 0) = 1 - Pr(s_h = 1|s_h = 0), \quad (13)$$

and

$$\begin{aligned} &Pr(s_h = N_h - 1|s_h = N_h - 1) \\ &= 1 - Pr(s_h = N_h - 2|s_h = N_h - 1). \end{aligned} \quad (14)$$

In terms of the information transmission channel, the complex channel gain g' is subject to a complex Gaussian distribution, i.e., $\mathcal{CN}(0, \sigma^2)$, where σ^2 is 1 for the considered small-scale fading channel. The channel power gain g follows exponential distribution [22]. The PDF is:

$$f_G(g) = \frac{1}{\sigma^2} e^{-g/\sigma^2} = e^{-g}. \quad (15)$$

Similar to the case of the energy harvesting channel, we set $\{A_g^{(j)}\}$, $j = 0, 1, \ldots, N_g$, as the quantization thresholds of the information transmission channel power gain, where $A_g^{(j)} < A_g^{(j+1)}$, $A_g^{(0)} = 0$ and $A_g^{(N_g)} = \infty$. The steady-state

[2]Zhang and Kassam [26] proposed the criterion of threshold partitions. We consider the states with equal probability for simplicity.

probability that the channel is in the state $s_g = j$ can be expressed as:

$$\begin{aligned} Pr(s_g = j) &= \int_{A_g^{(j)}}^{A_g^{(j+1)}} f_G(g)dg \\ &= e^{-A_g^{(j)}} - e^{-A_g^{(j+1)}}. \end{aligned} \tag{16}$$

The average channel power gain of the state $s_g = j$ is:

$$\begin{aligned} \mathrm{E}[g|s_g = j] &= \frac{\int_{A_g^{(j)}}^{A_g^{(j+1)}} ge^{-g}dg}{\int_{A_g^{(j)}}^{A_g^{(j+1)}} e^{-g}dg} \\ &= \frac{(A_g^{(j)}+1)e^{-A_g^{(j)}} - (A_g^{(j+1)}+1)e^{-A_g^{(j+1)}}}{e^{-A_g^{(j)}} - e^{-A_g^{(j+1)}}}. \end{aligned} \tag{17}$$

Suppose that the maximum Doppler frequency of the information transmission channel is D_g. The LCR of the threshold $A_g^{(j)}$ can be formulated as [27]:

$$N(A_g^{(j)}) = \sqrt{2\pi}D_g\rho e^{-\rho^2} = \sqrt{2\pi A_g^{(j)}}D_g e^{-A_g^{(j)}} \tag{18}$$

where $\rho = \sqrt{A_g^{(j)}/\sigma^2} = \sqrt{A_g^{(j)}}$. The state transition probabilities of the information transmission channel can be approximated by replacing A_h in equations (12)-(14) with A_g.

When the energy harvesting channel is in the state i, the expected amount of energy received by the sensor in one time slot is:

$$E_h = \eta P_L^H P\mathrm{E}[h|s_h = i]\tau. \tag{19}$$

We assume the sensor can only harvest positive integral multiples of unit energy. The number of unit energy harvested in one time slot is:

$$q = \max\{m : mE_u \le E_h\} \tag{20}$$

where m is a non-negative integer.

In the FDM mode, the sensor can harvest energy and transmit information simultaneously. Given a battery state $s_b = k$ in the current time slot, the battery state k' in the next slot is:

$$k' = \min\{N_b - 1, k - a + q\}. \tag{21}$$

In the TDM mode, the sensor need to either harvest energy or transmit information in one time slot. Thus the connection of the battery states in two adjacent time slots can be described as:

$$k' = \min\{N_b - 1, k - a + \delta(a)q\} \tag{22}$$

where $\delta(a)$ is the Kronecker delta function, i.e., if $a = 0$, $\delta(a) = 1$; otherwise, $\delta(a) = 0$.

We assume the energy harvesting channel and the information transmission channel are independent of each other. The overall system state transition probability can be formulated as:

$$\begin{aligned} &Pr(s'|s,a) \\ &= Pr((s_h, s_g, s_b) = (i', j', k')|(s_h, s_g, s_b) = (i, j, k), a) \\ &= Pr(s_h = i'|s_h = i)Pr(s_g = j'|s_g = j) \\ &\quad \times Pr(s_b = k'|(s_h, s_b) = (i, k), a) \end{aligned} \tag{23}$$

where the channel state transition probabilities are obtained from the equations (12)-(14), and the battery state transition probability is deterministic. Specifically, the battery transition probability is 1 for any k' that satisfies the equation (21) or (22), and 0 otherwise.

D. Reward Function

Given an action a and an information transmission channel state $s_g = j$, the reward function is formulated as

$$r_a(s_g = j) = W\log_2\left(1 + \frac{P_L^G a P_u \mathrm{E}[g|s_g = j]}{N_0 W\gamma}\right) \tag{24}$$

which is the expected throughput in one time slot. In the time slots with the action a that equals to 0, the sensor does not transmit any signal to the information sink. Thus, there is no reward in these slots.

E. Optimization of Policy

We assume the statistical properties of the network are time-invariant. The decision policy of a time-invariant system can be defined as a mapping $\pi(s) : \mathcal{S} \to \mathcal{A}_s$. The goal of formulating the MDP model is to find an optimal energy harvesting and information transmission policy for the maximal expected total discounted reward. The expected total discounted reward based on the reward function (24) from using the policy π beginning with the initial state $s^{(0)}$ can be formulated as:

$$v^{\pi}(s^{(0)}) = \mathrm{E}_{s^{(0)}}^{\pi}\left[\sum_{n=0}^{\infty}\lambda^n r_{\pi(s^{(n)})}(s_g^{(n)})\right] \tag{25}$$

Based on the property of the MDP proved in [23], at least one optimal stationary policy $\pi^*(s)$ satisfies the following Bellman equation:

$$v^{\pi^*}(s) = \max_{a\in\mathcal{A}_s}\left(r_a(s_g) + \lambda\sum_{s'\in\mathcal{S}} Pr(s'|s,a)v^{\pi^*}(s')\right). \tag{26}$$

where $v^{\pi^*}(s)$ is the state-value function of the state s under the optimal policy π^*. We can use the value iteration algorithm (Algorithm 1) to obtain an asymptotically optimal policy [23].

In the inequality (28) of Algorithm 1, the operator sp denotes the span which is formulated as $sp(\boldsymbol{v}) = \max_{s\in\mathcal{S}} v(s) - \min_{s\in\mathcal{S}} v(s)$. Each element of the vector $\boldsymbol{v}$ is the function value of each state. The policy which results in the state-value functions satisfying the inequality (28) is called the ε-optimal policy $\pi_\varepsilon^*(s)$. An optimal policy can be obtained through enough iterations when the constant ε is sufficiently small [23].

The value iteration algorithm is complex especially when the state space and action space are large. In general, the computational complexity per iteration is $(|\mathcal{S}|^2|\mathcal{A}|)$ multiplication operations in the value iteration algorithm, where $|\mathcal{S}|$ is the cardinality of the state space and $|\mathcal{A}|$ is the cardinality of the action space [29]. Since many state transition probabilities are zeros in the considered MDP, we treat the number of multiplications of non-zero elements per iteration as the computational complexity. The number of non-zero energy harvesting channel state transition probabilities and

Algorithm 1 Value Iteration Algorithm for MDP

1: Initialization: $v_0(s) = 0$, $\forall s \in \mathcal{S}$; $\varepsilon > 0$, $\lambda \in (0,1)$, and $n = 0$.

2: For each state $s \in \mathcal{S}$, compute $v_{n+1}(s)$ by

$$v_{n+1}(s) = \max_{a \in \mathcal{A}_s} \left\{ r_a(s_g) + \lambda \sum_{s' \in \mathcal{S}} Pr(s'|s,a) v_n(s') \right\}. \tag{27}$$

3: If

$$sp(\boldsymbol{v}_{n+1} - \boldsymbol{v}_n) < \varepsilon(1-\lambda)/2\lambda \tag{28}$$

go to step 4, otherwise increase n by 1 and go to step 2.

4: For each state $s \in \mathcal{S}$, choose

$$\pi^*_\varepsilon(s) \in \arg\max_{a \in \mathcal{A}_s} \left\{ r_a(s_g) + \lambda \sum_{s' \in \mathcal{S}} Pr(s'|s,a) v_{n+1}(s') \right\} \tag{29}$$

and stop.

the number of non-zero information transmission channel state transition probabilities are, $(3N_h - 2)$ and $(3N_g - 2)$, respectively, because the transitions only happen between adjacent states. In addition, since the battery state transition is deterministic and the number of actions per iteration is $(N_b + 1)N_b/2$, the computational complexity per iteration is $(3N_h - 2)(3N_g - 2)N_b(N_b + 1)/2$ multiplications of non-zero elements. However, the above complexity is still very large when the granularity of system states is small.

In practical applications, the asymptotically optimal policy $\pi^*_\varepsilon(s)$, $\forall s \in \mathcal{S}$ can be obtained via the value iteration algorithm. The sensor chooses action by checking the asymptotically optimal policy on the basis of the overall system state $s = (s_h, s_g, s_b)$. Specifically, the overall system state can be acquired as follows. Since the signals transmitted from the dedicated RF energy source carry pilots, the sensor can estimate the power gain state s_h of the energy harvesting channel. The power gain state s_g of the information transmission channel can be obtained at the sensor node via channel feedback. In addition, the sensor has the information of the battery state s_b.

IV. Monotonic Structure of The Optimal Policy

In this section, for the FDM mode, we present an interesting monotonic property of the transmitted energy in the asymptotically optimal policy with the battery state, which can be exploited to reduce the computational complexity when using the value iteration algorithm. However, the monotonic property does not hold in the TDM mode. We will analyze and explain this difference.

Before proving the monotonic property of the transmitted energy with the battery state in the asymptotically optimal policy, we introduce the definition of the superadditive function and its monotonic property [23].

Definition 1: Let X and Y be partially ordered sets and $\Gamma(x,y)$ be a real-valued function on $X \times Y$. For $x^+ \geq x^-$ in X and $y^+ \geq y^-$ in Y, if

$$\Gamma(x^+, y^+) - \Gamma(x^-, y^+) \geq \Gamma(x^+, y^-) - \Gamma(x^-, y^-), \tag{30}$$

$\Gamma(x,y)$ is said to be a superadditive function of x and y.

Lemma 1 ([23, Lemma 4.7.1]): Suppose that $\Gamma(x,y)$ is a superadditive function on $X \times Y$ and for each $x \in X$, $\max_{y \in Y} \Gamma(x,y)$ exists. Then

$$\Theta(x) = \max \left\{ y' \in \arg\max_{y \in Y} \Gamma(x,y) \right\} \tag{31}$$

is monotonically non-decreasing in x.

We reconstruct the iteration equation (27) as a function of the action and state, i.e. the action-value function. According to (12)-(14) and (23), the action-value function with respect to action a and state $s = (i,j,k)$ can be formulated as:

$$\begin{aligned} Q^a_{n+1}(i,j,k) &= r_a(j) + \lambda \sum_{i'=\max\{0,i-1\}}^{\min\{N_h-1,i+1\}} Pr(s_h = i'|s_h = i) \\ &\quad \times \sum_{j'=\max\{0,j-1\}}^{\min\{N_g-1,j+1\}} Pr(s_g = j'|s_g = j) \\ &\quad \times Pr(s_b = k'|(s_h = i, s_b = k), a) v_n(i',j',k') \\ &= r_a(j) + \lambda \mathrm{E}_{i',j'}[v_n(i',j',\min\{N_b - 1, k - a + q\})|i,j] \end{aligned} \tag{32}$$

where $v_{n+1}(i,j,k) = \max_{a \in \mathcal{A}_s} Q^a_{n+1}(i,j,k)$. We define $x = k - a$ and

$$\bar{v}_n(i,j,x) = \mathrm{E}_{i',j'}[v_n(i',j',\min\{N_b - 1, x + q\})|i,j], \tag{33}$$

which is a function of i, j and x. We will prove that the action-value function $Q^a_n(i,j,k)$ is superadditive in a and k given i and j for any $n > 0$, so that the transmitted energy is non-decreasing in the battery state in the asymptotically optimal policy given the channel states. In the following, we will first give the sufficient condition for $Q^a_n(i,j,k)$ to be a superadditive function in *Theorem 1*, and then use *Lemma 2* and *Lemma 3* to prove that the sufficient condition is indeed satisfied in the FDM mode.

Theorem 1: For any $n \geq 0$ such that $\bar{v}_n(i,j,x)$ is concave in x given i and j, $Q^a_{n+1}(i,j,k)$ is a supperadditive function of k and a given i and j.

Proof: See Appendix A. □

Before proving that $\bar{v}_n(i,j,x)$ is in fact concave in x given i and j for any $n \geq 0$, we first show that $v_n(i',j',k')$ is a non-decreasing function of k' given i' and j' for any $n \geq 0$ in the following lemma.

Lemma 2: $v_n(i',j',k')$ is a non-decreasing function of k' given i' and j' for any $n \geq 0$.

Proof: See Appendix B. □

On the basis of $v_n(i',j',k')$ being a non-decreasing function of k', we use mathematical induction to prove $\bar{v}_n(i,j,x)$ is concave in x given i and j in the following lemma.

Lemma 3: For any $n \geq 0$, $\bar{v}_n(i,j,x)$ is a concave function of x given i and j.

Proof: See Appendix C. □

From *Lemma 3*, $\bar{v}_n(i,j,x)$ is concave in x given i and j. Therefore, given i and j, $Q^a_{n+1}(i,j,k)$ is a supperadditive

function of k and a according to *Theorem 1*. Furthermore, the transmitted energy is non-decreasing with the battery state in the asymptotically optimal policy. This is demonstrated in Fig. 3. In every iteration of the value iteration algorithm, the sensor can search the optimal action at the current state from the optimal action at the previous state, which can narrow down the search scope of the optimal policy. Therefore, the computational complexity can be reduced. The monotonic value iteration algorithm is shown in Algorithm 2.

Algorithm 2 Monotonic Value Iteration Algorithm for MDP

1: Initialization: $v_0(s) = 0$, $\forall s \in \mathcal{S}$; $\varepsilon > 0$, $\lambda \in (0,1)$, and $n = 0$.

2: For each $s_h = i$ and $s_g = j$, initialize $s_b = k = 0$, $\mathcal{A}_s = \{0\}$.

a: Set

$$v_{n+1}(s) = \max_{a \in \mathcal{A}_s} \left\{ r_a(s_g) + \lambda \sum_{s' \in \mathcal{S}} Pr(s'|s,a) v_n(s') \right\},$$

and

$$\mathcal{A}_s^* = \arg\max_{a \in \mathcal{A}_s} \left\{ r_a(s_g) + \lambda \sum_{s' \in \mathcal{S}} Pr(s'|s,a) v_n(s') \right\}.$$

b: If $k = N_b - 1$, go to step 3; else,

$$\mathcal{A}_{(i,j,k+1)} = \Big\{ a \in \{0, \ldots, k+1\} \bigcap \{a \geq \max\{a' \in \mathcal{A}_s^*\}\} \Big\}.$$

c: Substitute $s = (i,j,k)$ with $s = (i,j,k+1)$ and return to 2(a).

3: If

$$sp(\boldsymbol{v}_{n+1} - \boldsymbol{v}_n) < \varepsilon(1-\lambda)/2\lambda,$$

go to step 4; otherwise, increase n by 1 and go to step 2.

4: For each $s \in \mathcal{S}$, choose

$$\pi_\varepsilon^*(s) \in \arg\max_{a \in \mathcal{A}_s} \left\{ r_a(s_g) + \lambda \sum_{s' \in \mathcal{S}} Pr(s'|s,a) v_{n+1}(s') \right\},$$

and stop.

In the FDM mode, the sensor harvests energy and transmits information in the same time slot. When the energy harvesting channel state is known, we can determine the amount of energy that the sensor replenishes. However, in the TDM mode, energy harvesting and information transmission are in different time slots. The energy harvesting and information transmission policy consists of not only power allocation but also link selection. The sensor will replenish energy in future time slots if information transmission is selected, and the replenished energy in future slots is uncertain in the current time slot. Therefore, the policy that uses more power in transmission when the battery has more energy does not guarantee the long-term optimality. From Fig. 4, it is demonstrated that the transmitted energy in the optimal policy and the battery state do not display a monotonic relation in the TDM mode.

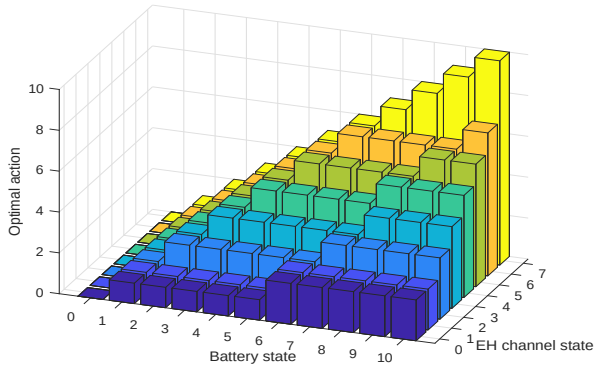


Fig. 3. The relation between the optimal action and the battery state at each energy harvesting (EH) channel state given the information transmission channel state $s_g = 7$ in the FDM mode.

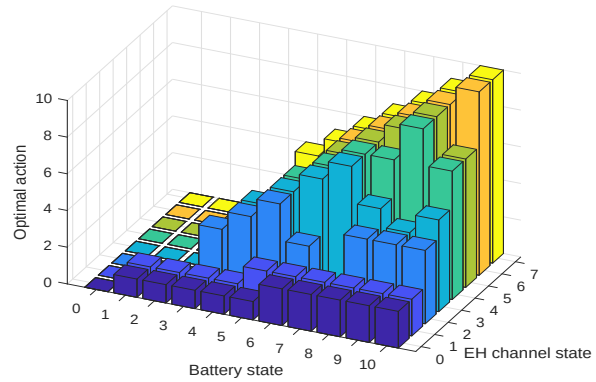


Fig. 4. The relation between the optimal action and the battery state at each EH channel state given the information transmission channel state $s_g = 7$ in the TDM mode.

V. Simulation Results

In this section, we show simulation results to evaluate the performance of the proposed energy harvesting and information transmission scheme. Moreover, the effect of some parameters (such as the battery capacity, the number of channel states, and the discount factor) on the long-term average throughput is investigated[3]. The bandwidth and the noise power spectral density of the information transmission channel are set to 10 kHz and -174 dBm/Hz, respectively. In the TDM mode, energy harvesting uses the same frequency band as information transmission, while in the FDM mode, energy harvesting is allocated in another frequency band, such as the license-free industrial, scientific, and medical (ISM) frequency band. The bandwidth of the antenna in a dedicated RF energy harvester is usually narrow [31]. How to determine the optimal bandwidth of the energy harvesting channel in the FDM mode is beyond the scope of this paper. We set the length of the energy harvesting link and the information transmission link to 2 m and 20 m, respectively. We set the unitless constant C in

[3]According to [23, Corollary 8.2.5], when the discount factor λ approaches 1, the policy that maximizes the discounted total reward can also approximately maximize the average reward.

equation (1) to -31.5 dB, and the reference distance d_0 to 1 m for the two channels. The path-loss exponents of the energy harvesting channel and the information transmission channel are 2 and 3, respectively. We set the K-factor K to 0.5. The small-scale fading channel is made based on the Jake's model [32]. The number of states of the energy harvesting channel and the information transmission channel are both defaulted to 8. We set the maximum Doppler frequency of the two channels are $D_h = 0.01$ Hz and $D_g = 0.02$ Hz, respectively. The steady-state probabilities and the state transition probabilities are calculated in accordance with section III. We set the duration of one time slot to 1 s. The transmission power of the dedicated RF energy source is defaulted to 4 W. The battery capacity of the sensor is defaulted to 10^{-3} J which is divided into 11 states, namely $\mathcal{S}_b = \{0, 1, \ldots, 10\}$. Thus the unit energy is 10^{-4} J. The energy efficiency η is set to 0.5. The gap γ only depends on the symbol error rate (SER) which is required to be 10^{-6} in this paper. The gap for the frequently-used M-ary quadrature amplitude modulation (M-QAM) is formulated as $\gamma = 1/3(\mathrm{Q}^{-1}(\mathrm{SER}/4))^2$ [21]. We set $\varepsilon = 10^{-8}$ and $\lambda = 0.99$. In the implementation of the optimal and greedy policies, we set a simulation time as 5×10^6 s. Each result is the average of 40 runs of the simulations.

The average throughput for the FDM mode and the TDM mode under different transmission power of the RF energy source is shown in Fig. 5. It is shown that the average throughput of the FDM mode is nearly twice as that of the TDM mode. This is because that almost half of all time slots are used for energy harvesting in the TDM mode. In practical applications, we tend to choose the FDM mode if a dedicated frequency band to transmit energy could be allocated.

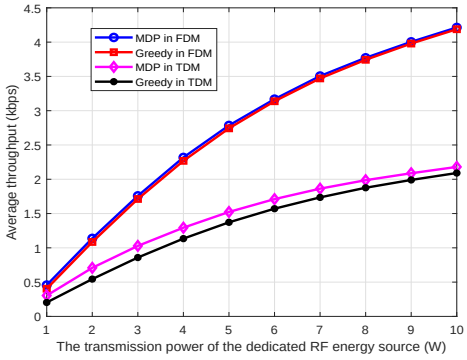


Fig. 5. Average throughput versus transmission power of the RF energy source.

Fig. 6 presents the average throughput per information transmission slot versus different transmission power of the RF energy source. It is shown that the average throughput per information transmission slot in the TDM mode is greater than that in the FDM mode, especially when the transmission power of the RF energy source is low. The reason is as follows. The reward function increases with a large slope when the transmission power of the sensor is low. Therefore, it benefits the sensor to accumulate its received energy before transmitting information in the TDM mode. As a result, the number of information transmission slots is reduced and the throughput in each information transmission slot becomes larger in the TDM mode. In this simulation example, for the MDP scheme in the TDM mode, when the transmission power of the RF energy source is 1 W and 10 W, the information transmission time slots only account for 9% and 38% of the total time slots, respectively.

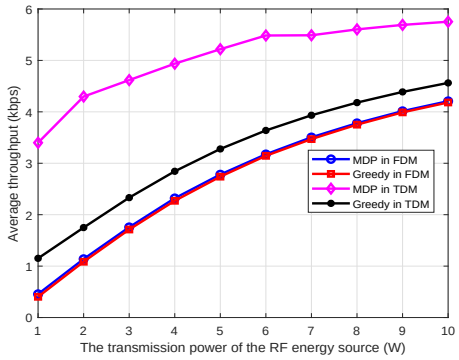


Fig. 6. Average throughput per information transmission slot versus transmission power of the RF energy source.

Fig. 7 shows the average throughput of the MDP scheme and the greedy scheme for different battery state capacity in the FDM mode and the TDM mode. As a benchmark, the greedy scheme is performed without concerning the state transition probabilities. The information is transmitted as long as the battery is non-empty. It is observed that the average throughput is monotonically increasing with the battery capacity, while the average throughput of the greedy scheme becomes saturated when the battery capacity is large. This is because that the energy may overflow when the battery capacity is small. The probability of energy overflowing is decreasing with the battery capacity increasing. Thus the average throughput becomes stable. We can also observe that the MDP scheme outperforms the greedy scheme, because the action based on the MDP scheme is selected in accordance with the current states. The sensor can wait for better channel state before energy overflowing when the battery capacity is large. Thus the average throughput can continue to increase with the battery capacity increasing.

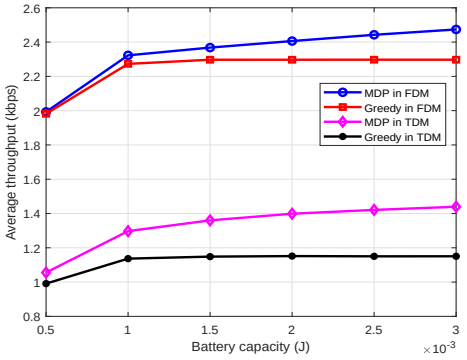


Fig. 7. Average throughput versus battery capacity for the MDP scheme and the greedy scheme.

Fig. 8 shows the effect of the number of information

transmission channel states on the average throughput. When the number of channel states is small, the average throughput increases with the number of states increasing. However, when the number of states is large, the average throughput decreases with the number of states increasing. The reason is as follows. When the number of channel states is small, the power gain in one state will vary in a large range. The actual channel power gain in one state will be in a smaller range than the range in that state. Therefore, the expected channel power gain (17) cannot represent the actual channel power gain accurately. In contrast, when the number of channel states is large, the power gain range will be small. In this case, the probability that the channel power gain in the next time slot is not in the current or adjacent states is large. The state transition probabilities (12)-(14) cannot approximate the actual state transition processes efficiently. For the energy harvesting channel, there exists the same property according to the simulations.

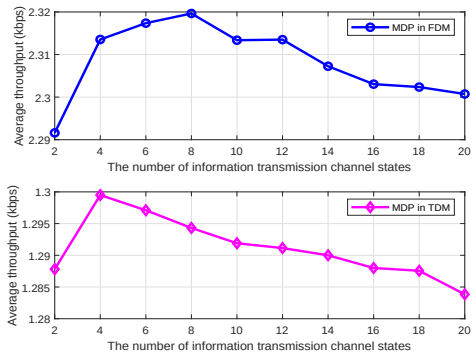


Fig. 8. Average throughput versus the number of information transmission channel states.

Fig. 9 shows the effect of the discount factor on the average throughput. It is observed that the average throughput is monotonically increasing with the discount factor. This is because that the optimal policy of the discounted criterion can approximate the optimal policy of the average criterion when the discount factor $\lambda \approx 1$. However, the speed of convergence of the discounted MDP algorithm decreases with the discount factor increasing. It means that acquiring the optimal transmission policy needs more iterations.

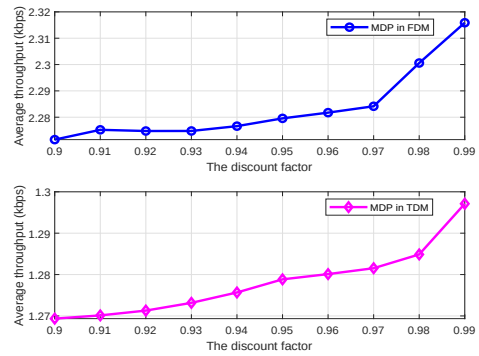


Fig. 9. Average throughput versus the discount factor.

VI. Conclusions

This work studied a wireless-powered sensor communication network, in which the sensor is powered by the dedicated RF energy source. We discussed two frequently-used working mode: the FDM mode and the TDM mode. To obtain the optimal policy for the long-term maximum throughput, we formulated the energy allocation problem as a Markov decision model and solved it through the value iteration algorithm. We proved that the transmitted energy is a monotonically non-decreasing function of the battery state in the asymptotically optimal policy in the FDM mode. The computational complexity can be reduced using this monotonic property. However, the same monotonic property does not exist in the TDM mode. The simulation results show that the proposed approach can achieve significant gains compared with the greedy scheme. In future works, we will extend our work to the case that several sensors are supported by one dedicated RF energy source, which is more practical in real-world wireless-powered sensor networks.

Appendix A
Proof of Theorem 1

Proof: If $\bar{v}_n(i,j,x)$ is a concave function of x given i and j, the following inequality is valid:

$$\begin{aligned} &\mathrm{E}_{i',j'}[v_n(i',j',\min\{N_b-1,k^+-a^++q\})|i,j] \\ &-\mathrm{E}_{i',j'}[v_n(i',j',\min\{N_b-1,k^--a^++q\})|i,j] \geq \\ &\mathrm{E}_{i',j'}[v_n(i',j',\min\{N_b-1,k^+-a^-+q\})|i,j] \\ &-\mathrm{E}_{i',j'}[v_n(i',j',\min\{N_b-1,k^--a^-+q\})|i,j] \end{aligned} \tag{34}$$

where $k^+ \geq k^-$, $a^+ \geq a^-$, and $k^- \geq a^+$.

According to (32) and (34), the following inequality is satisfied:

$$\begin{aligned} &Q^{a^+}_{n+1}(i,j,k^+) - Q^{a^+}_{n+1}(i,j,k^-) \geq \\ &Q^{a^-}_{n+1}(i,j,k^+) - Q^{a^-}_{n+1}(i,j,k^-). \end{aligned} \tag{35}$$

Therefore, $Q^a_{n+1}(i,j,k)$ is a superadditive function of k and a given i and j according to *Definition 1*. □

Appendix B
Proof of Lemma 2

The proof of Lemma 2 will use the following property: for any given x, if the bounded function $\Phi(x,y)$ is non-decreasing in y, $\Psi(y) = \max_x \Phi(x,y)$ is also a non-decreasing function of y. The proof of the property is as follows.

Proof: For $y_1 \leq y_2$, we define:

$$x_1^{max} \in \arg\max_x \Phi(x,y_1)$$

and

$$x_2^{max} \in \arg\max_x \Phi(x,y_2).$$

Since we assume $\Phi(x,y)$ is non-decreasing in y for any fixed x, the following inequalities are satisfied:

$$\Psi(y_1) = \Phi(x_1^{max},y_1) \leq \Phi(x_1^{max},y_2) \leq \Phi(x_2^{max},y_2) = \Psi(y_2). \tag{36}$$

Therefore, $\Psi(y) = \max_x \Phi(x,y)$ is a non-decreasing function of y. □

In the following content of this appendix, we use mathematical induction to prove lemma 2.

Proof of Lemma 2. Suppose that $v_n(i', j', k')$ is non-decreasing in k' for a non-negative integer n. Since $k' = \min\{N_b - 1, x+q\}$ is non-decreasing in x, $v_n(i', j', \min\{N_b - 1, x + q\})$ is non-decreasing in x. The non-negative weighted sum $\bar{v}_n(i, j, x)$ of $v_n(i', j', \min\{N_b - 1, x + q\})$ is a non-decreasing function of x given i and j. Since $r_a(j)$ is unrelated to k,

$$\begin{aligned} &Q^a_{n+1}(i, j, k) \\ &\quad = r_a(j) + \lambda \mathrm{E}_{i',j'}[v_n(i', j', \min\{N_b - 1, k - a + q\})|i, j] \end{aligned} \tag{37}$$

is a non-decreasing function of k given i, j, and a. We can infer that $v_{n+1}(i, j, k) = \max_{a \in \mathcal{A}_s} \left\{Q^a_{n+1}(i, j, k)\right\}$ is non-decreasing in k given i and j according to the above property. Since $v_0(i', j', k') = 0$ is a non-decreasing function of k', $v_n(i', j', k')$ is non-decreasing in k' given i' and j' for all n. □

Appendix C
Proof of Lemma 3

Proof: Suppose that $v_n(i', j', k')$ is a concave function of k' given i' and j' for a non-negative integer n. Since $k' = \min\{N_b - 1, x + q\}$ is a concave function of x and $v_n(i', j', k')$ is a non-decreasing function of k' given i' and j' (*Lemma 2*), $v_n(i', j', \min\{N_b - 1, x + q\})$ is a concave function of x given i' and j' [30, Equation 3.10]. Since $\bar{v}_n(i, j, x)$ can be viewed as a non-negative weighted sum of concave functions $v_n(i', j', k')$ based on equation (32), $\bar{v}_n(i, j, x)$ is a concave function of x given i and j. Before proving that $v_{n+1}(i, j, k)$ is concave in k, we assume $v_n(i, j, \tilde{k})$ is a concave function of the continuous variable $\tilde{k}$ which contains the discrete function values $v_n(i, j, k),\ k \in \{0, 1, \ldots, N_b - 1\}$. We also assume that $v_{n+1}(i, j, \tilde{k})$ is a continuous function of $\tilde{k}$. For $\forall k_1, k_2 \in \{0, 1, \ldots, N_b - 1\}$ and $\beta \in [0, 1]$, we should prove the following inequality:

$$\begin{aligned} &\beta v_{n+1}(i, j, k_1) + (1 - \beta) v_{n+1}(i, j, k_2) \\ &\quad \le v_{n+1}(i, j, \beta k_1 + (1 - \beta) k_2). \end{aligned} \tag{38}$$

Suppose that a_1 and a_2 satisfy $v_{n+1}(i, j, k_1) = Q^{a_1}_{n+1}(i, j, k_1)$ and $v_{n+1}(i, j, k_2) = Q^{a_2}_{n+1}(i, j, k_2)$, respectively. Then

$$\begin{aligned} &\beta v_{n+1}(i, j, k_1) + (1 - \beta) v_{n+1}(i, j, k_2) \\ &= \beta r_{a_1}(j) + (1 - \beta) r_{a_2}(j) \\ &\quad + \beta \lambda \mathrm{E}_{i',j'}[v_n(i', j', \min\{N_b - 1, k_1 - a_1 + q\})|i, j] \\ &\quad + (1 - \beta) \lambda \mathrm{E}_{i',j'}[v_n(i', j', \min\{N_b - 1, k_2 - a_2 + q\})|i, j] \\ &\le r_{a_\beta}(j) + \lambda \mathrm{E}_{i',j'}[v_n(i', j', \min\{N_b - 1, k_\beta - a_\beta + q\})|i, j] \\ &\le \max_{a \in \mathcal{A}_s} \{r_a(j) \\ &\quad + \lambda \mathrm{E}_{i',j'}[v_n(i', j', \min\{N_b - 1, k_\beta - a + q\})|i, j]\} \\ &= v_{n+1}(i, j, k_\beta) \end{aligned} \tag{39}$$

where $k_\beta = \beta k_1 + (1 - \beta) k_2$, $a_\beta = \beta a_1 + (1 - \beta) a_2$ and $\mathcal{A}_s = [0, k_\beta]$ in the above inequalities. In (39), the first inequality is established by using $r_a(j)$ and $\bar{v}(i, j, k)$ as concave functions. Thus, we can prove that $v_{n+1}(i, j, k)$ is a concave function of k given i and j. Similar to the procedure of proving that $\bar{v}_n(i, j, x)$ is a concave function of x given i and j, we can prove that $\bar{v}_{n+1}(i, j, x)$ is concave in x given i and j. Since $v_0(i', j', k') = 0$ is a concave function of k' given i' and j', $\bar{v}_n(i, j, x)$ is a concave function of x given i and j for all $n \ge 0$. □

References

[1] G. Anastasi, M. Conti, M. D. Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: A survey," *Ad Hoc Netw.*, vol. 7, no. 3, pp. 537-568, May 2009.

[2] H. Guo, J. Liu, Z. M. Fadlullah, and N. Kato, "On minimizing energy consumption in FiWi enhanced LTE-A HetNets," *IEEE Trans. Emerg. Topics Comput.*, vol. 6, no. 4, pp. 579-591, Oct.-Dec. 2018.

[3] S. Chen, J. Hu, Y. Shi and L. Zhao, "LTE-V: A TD-LTE-based V2X solution for future vehicular network," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 997-1005, Dec. 2016.

[4] M. L. Ku, W. Li, Y. Chen, and K. J. Ray Liu, "Advances in energy harvesting communications: Past, present, and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1384–1412, 2nd Quart. 2016.

[5] P. He, L. Zhao, S. Zhou, and Z. Niu, "Recursive waterfilling for wireless links with energy harvesting transmitters," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1232-1241, Mar. 2014.

[6] Q. Li, J. Gao, J. Wen, X. Tang, L. Zhao, and L. Sun, "SMDP-based downlink packet scheduling scheme for solar energy assisted heterogeneous networks," accepted, *GreenCom*, 2018.

[7] K. Suto, H. Nishiyama, N. Kato, and T. Kuri, "Model predictive joint transmit power control for improving system availability in energy-harvesting wireless mesh networks," *IEEE Commun. Lett.*, vol. 22, no. 10, pp. 2112-2115, Oct. 2018.

[8] X. Lan, Q. Chen, X. Tang, and L. Cai, "Achievable rate region of the buffer-aided two-way energy harvesting relay network," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11127-11142, Nov. 2018.

[9] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Wireless networks with RF energy harvesting: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 757-789, 2nd Quart. 2015.

[10] A. Goldsmith, *Wireless communications*. Cambridge Univ. Press, 2005.

[11] D. Niyato, D. I. Kim, M. Maso, and Z. Han, "Wireless powered pommunication networks: Research directions and technological approaches," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 88-97, Dec. 2017.

[12] H. Ju and R. Zhang, "Throughput maximization in wireless powered communication networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 418-428, Jan. 2014.

[13] X. Di, K. Xiong, P. Fan, H. C. Yang, and K. B. Letaief, "Optimal resource allocation in wireless powered communication networks with user cooperation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7936-7949, Dec. 2017.

[14] W. Shin, M. Vaezi, J. Lee, and H. V. Poor, "Cooperative wireless powered communication networks with interference harvesting," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3701-3705, Apr. 2018.

[15] X. Lin, L. Huang, C. Guo, P. Zhang, M. Huang, and J. Zhang, "Energy-efficient resource allocation in TDMS-based wireless powered communication networks," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 861-864, Apr. 2017.

[16] X. Zhou, C. K. Ho, and R. Zhang, "Wireless power meets energy harvesting: A joint energy allocation approach in OFDM-based system," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3481-3491, May 2016.

[17] Q. Wu, M. Tao, D. W. Kwan Ng, W. Chen, and R. Schober, "Energy-efficient resource allocation for wireless powered communication networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2312-2327, Mar. 2016.

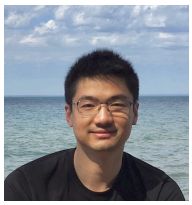
[18] Y. Wu, Q. Yang, and K. S. Kwak, "Energy efficiency maximization for energy harvesting millimeter wave systems at high SNR," *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 698-701, Oct. 2017.

[19] F. Zhao, L. Wei, and H. Chen, "Optimal time allocation for wireless information and power transfer in wireless powered communication systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1830-1835, Mar. 2016.

[20] N. V. Huynh, D. T. Hoang, D. Niyato, P. Wang, and D. I. Kim, "Optimal time scheduling for wireless-powered backscatter communication networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 820-823, Oct. 2018.

[21] A. Garcia-Armada, "SNR gap approximation for M-PSK-based bit loading," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 57-60, Jan. 2006.
[22] D. Tse and P. Viswanath, *Fundmentals of wireless communications*. Cambridge Univ. Press, 2005.
[23] M. L. Puterman, *Markov decision processes: Discrete stochastic dynamic programming*. John & Sons, 1994.
[24] M. Li, L. Zhao, and H. Liang, "An SMDP-based prioritized channel allocation scheme in cognitive enabled vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7925-7933, Sept. 2017.
[25] J. G. Proakis, *Digital communications*. McGraw-Hill, 1995.
[26] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688-1692, Nov. 1999.
[27] G. L. Stuber, *Principles of mobile communication*. Kluwer Academic, 1996.
[28] H. S. Wang and N. Moayeri, "Finite-state Markov channel - A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163-171, Feb. 1995.
[29] M. L. Littman, L. D. Thomas, and P. K. Leslie, "On the complexity of solving Markov decision problems," *in Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 1995, pp. 394-402.
[30] S. Boyd and V. Lieven, *Convex optimization*. Cambridge Univ. Press, 2004.
[31] D. N. K. Jayakody, J. Thompson, S. Chatzinotas, and S. Durrani, *Wireless information and power transfer: A new paradigm for green communications*. Springer, 2017.
[32] W. C. Jakes and D. C. Cox, *Microwave mobile communications*. Willey-IEEE press, 1994.

Qizhen Li received the B.S. degree in Electronic Information Science and Technology from Liaocheng University, Liaocheng, China, in 2012. He is currently working towards a Ph.D. degree in Information and Communication Engineering, Southwest Jiaotong University, Chengdu, China. From Oct. 2017 to Sep. 2018, he was a visiting Ph.D. student with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada. His research interests include edge computing, green communications, Markov decision processes and reinforcement learning.



Jie Gao (S'09-13, M'17) received the B.Eng. degree in Electronics and Information Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007, and the M.Sc. and Ph.D. degrees in Electrical Engineering from the University of Alberta, Edmonton, Alberta, Canada, in 2009 and 2014, respectively. He is a recipient of Natural Science and Engineering Research Council of Canada Postdoctoral Fellowship Award and is currently a postdoctoral fellow with the Department of Electrical & Computer Engineering at Ryerson University and the Department of Electrical & Computer Engineering at University of Waterloo. His research interests include the application of game theory and mechanism design for distributed decision making in communication networks and performance optimization of multiuser systems.

Hongbin Liang (M'12) received the B.Sc. degree in communication engineering from the Beijing University of Post and Telecommunication, Beijing, China, in 1995 and the M.Sc. and Ph.D. degrees in electrical engineering from the Southwest Jiaotong University, Chengdu, China, in 2001 and 2012, respectively. He is currently an Associate Professor with the School of Transportation and Logistics, Southwest Jiaotong University. From 2001 to 2009, he was a Software Engineer with the Motorola R&D Center of China, where he focused on system requirement analysis and Third-Generation Partnership Project protocol analysis. From 2009 to 2011, he was a Visiting Ph.D. student with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His current research interests focus on resource allocation, quality of service, security and efficiency in cloud computing, and wireless sensor networks.

Lian Zhao (S'99-M'03-SM'06) received the Ph.D. degree from the Department of Electrical and Computer Engineering (ELCE), University of Waterloo, Canada, in 2002. She joined the Department of Electrical and Computer Engineering at Ryerson University, Toronto, Canada, in 2003 and a Professor in 2014. Her research interests are in the areas of wireless communications, radio resource management, power control, cognitive radio and cooperative communications, optimization for complicated systems. She received the Best Land Transportation Paper Award from IEEE Vehicular Technology Society in 2016; Top 15 Editor in 2015 for IEEE Transaction on Vehicular Technology; Best Paper Award from the 2013 International Conference on Wireless Communications and Signal Processing (WCSP) and Best Student Paper Award (with her student) from Chinacom in 2011; the Ryerson Faculty Merit Award in 2005 and 2007; the Canada Foundation for Innovation (CFI) New Opportunity Research Award in 2005, and Early Tenure and promotion to Associate Professor in 2006. She has been an Editor for IEEE TRANSACTION ON VEHICULAR TECHNOLOGY since 2013; co-chair for IEEE ICC 2018 Wireless Communication Symposium; workshop co-chair for IEEE/CIC ICCC 2015; local arrangement co-chair for IEEE VTC Fall 2017 and IEEE Infocom 2014; co-chair for IEEE Global Communications Conference (GLOBECOM) 2013 Communication Theory Symposium. She served as a committee member for NSERC (Natural Science and Engineering Research Council of Canada) Discovery Grants Evaluation Group for Electrical and Computer Engineering since 2015. She is a licensed Professional Engineer in the Province of Ontario, a senior member of the IEEE Communication and Vehicular Society.

Xiaohu Tang (M'04-SM'18) received the B.S. degree in applied mathematics from the Northwest Polytechnic University, Xi'an, China, the M.S. degree in applied mathematics from the Sichuan University, Chengdu, China, and the Ph.D. degree in electronic engineering from the Southwest Jiaotong University, Chengdu, China, in 1992, 1995, and 2001 respectively.

From 2003 to 2004, he was a research associate in the Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology. From 2007 to 2008, he was a visiting professor at University of Ulm, Germany. Since 2001, he has been in the School of Information Science and Technology, Southwest Jiaotong University, where he is currently a professor. His research interests include coding theory, network security, distributed storage and information processing for big data.

Dr. Tang was the recipient of the National excellent Doctoral Dissertation award in 2003 (China), the Humboldt Research Fellowship in 2007 (Germany), and the Outstanding Young Scientist Award by NSFC in 2013 (China). He served as Associate Editors for several journals including IEEE Transactions on Information Theory and IEICE Trans on Fundamentals, and served on a number of technical program committees of conferences.