# On Outage Probability for Two-Way Relay Networks With Stochastic Energy Harvesting

Wei Li, Meng-Lin Ku, *Member, IEEE*, Yan Chen, *Senior Member, IEEE*, and K. J. Ray Liu, *Fellow, IEEE*

***Abstract*—In this paper, we propose an optimal relay transmission policy by using a stochastic energy harvesting (EH) model for the EH two-way relay network, wherein the relay is solar-powered and equipped with a finite-sized battery. In this policy, the long-term average outage probability is minimized by adapting the relay transmission power to the wireless channel states, battery energy amount, and causal solar energy states. The designed problem is formulated as a Markov decision process (MDP) framework, and conditional outage probabilities for both decode-and-forward (DF) and amplify-and-forward (AF) cooperation protocols are adopted as the reward functions. We uncover a monotonic and bounded differential structure for the expected total discounted reward, and prove that such an optimal transmission policy has a threshold structure with respect to the battery energy amount in sufficiently high SNRs. Finally, the outage probability performance is analyzed and an interesting saturated structure for the outage performance is revealed, i.e., the expected outage probability converges to the battery empty probability in high SNR regimes, instead of going to zero. Furthermore, we propose a saturation-free condition that can guarantee a zero outage probability in high SNRs. Computer simulations confirm our theoretical analysis and show that our proposed optimal transmission policy outperforms other compared policies.**

***Index Terms*—Stochastic energy harvesting, two-way relay network, outage probability, decode-and-forward, amplify-and-forward, Markov decision process.**

## I. Introduction

THE ENERGY-CONSTRAINED wireless communications such as wireless sensor networks usually rely on a fixed battery to supply energy for data transmissions in the absence of power grid, and the lifetime of the networks is largely dominated by the battery capacity. In general, the larger the battery capacity is, the longer the lifetime of the networks is. However, a battery with larger capacity is often expensive and inconvenient for the network deployment. On the other hand, although the lifetime of the networks can be prolonged by regularly replacing batteries, the replacement may be inconvenient, costly, dangerous or even impossible in some secluded areas. Therefore, energy harvesting (EH) has recently attracted significant attention due to its effectiveness to resolve energy supply problems in wireless networks and to perpetually provide an infinite amount of energy [1], [2]. In EH communication networks, the EH nodes can make use of renewable energy sources, e.g., solar, mechanical motion, electromagnetic radiation, and thermoelectric source [1], to recharge their batteries and to fulfill data transmissions. While an inexhaustible energy supply from environments enables EH nodes to communicate for an infinite lifetime, power management and transmission scheduling remain a crucial research issue because of the randomness and uncertainty of the harvested energy.

EH wireless communications have been extensively studied in point-to-point scenarios in the literature. For example, a directional water-filling algorithm was proposed in [3] to determine the optimal power scheduling for maximizing the short-term throughput in point-to-point fading channels. Unlike the objective function in [3], the optimal power allocation scheme that aims at minimizing the average outage probability over a finite time horizon was studied in [4] and [5]. The authors in these papers exploited a deterministic EH model, in which the solar energy state information (ESI) is non-causal and the energy arrival information is known prior to transmission scheduling, and a stochastic EH model, in which the solar ESI is causal. Further, considering a real data record of solar irradiance, the authors in [6] investigated a data-driven stochastic solar EH model, and then an optimal transmission policy was proposed to maximize the long-term net bit rate by using Markov decision process (MDP) approach. Besides [6], the online scheduling policies using the MDP have been extensively investigated in the literature. For example, with a maximum power constraint for transmitters, an achievable rate maximization problem was cast as an MDP with continuous battery states in [7]. Aiming at maximizing the sum throughput of a slotted Aloha-based wireless network with multiple EH transmitters, the authors in [8] proposed two distributed optimal transmission policies, for which one is static with constant power, and the other is dynamic utilizing the MDP approach.

Cooperative communications have been applied in various wireless scenarios for the purpose of the link quality improvement [9]. It is worth noting that there has been a

Manuscript received July 2, 2015; revised November 10, 2015 and February 22, 2016; accepted March 19, 2016. Date of publication March 29, 2016; date of current version May 13, 2016. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant No. 61461136001 and 61401348, the Science and Technology Program of Shaanxi Province (2011K06-10), and the Fundamental Research Funds for the Central University. The associate editor coordinating the review of this paper and approving it for publication was N. B. Mehta.

W. Li is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA, and also with the Department of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: wli52140@umd.edu; leew52140@stu.xjtu.edu.cn).

M.-L. Ku is with the Department of Communication Engineering, National Central University, Jung-li 32001, Taiwan (e-mail: mlku@ce.ncu.edu.tw).

Y. Chen is with the School of Electronic Engineering, University of Electronic Science and Technology, Chengdu 610051, China (e-mail: eecyan@uestc.edu.cn).

K. J. R. Liu is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: kjrliu@umd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2016.2547954

growing interest in investigating EH cooperative communications, where relay nodes can harvest energy from environments. An optimal transmission policy for a two-hop network with an EH source node and an energy-constrained relay node was proposed in [10]. Further, the authors in [11] developed the optimal power policy for a two-hop network, wherein the source and relay are both EH nodes. Except for the two-hop networks, an optimal power allocation scheme for the classic three-node Gaussian relay networks with EH nodes was investigated in [12]. Moreover, in [13] and [14], transmission policies based on wireless energy transfer, i.e., radio-frequency(RF)-based energy harvesting, were studied in one-way relay networks.

Due to the advantage of higher transmission efficiency, two-way relay (TWR) networks have been recognized as a promising solution for information exchange between two source nodes via an intermediate relay node [15], [16]. Recently, the TWR networks with EH nodes have attracted much attention. Unlike the traditional TWR networks, not only the TWR fading channels, but also the stochastic and uncertain energy harvested from environments, should be seriously considered for power allocation and scheduling problems in EH TWR networks. In the literature [17]–[19], power allocation algorithms were proposed for maximizing short-term sum rates in EH TWR networks using deterministic EH models. An EH relay with a data buffer can cache data and make use of flexible transmission policies in [17]. Moreover, a generalized iterative directional water-filling algorithm was designed in [18] for various relaying strategies. An optimization framework with the uncertainty of channel state information (CSI) was presented in [19]. Further, the authors in [20] developed an optimal relay transmission policy for maximizing the long-term average throughput of the EH TWR network with stochastic EH models. In addition, the optimal transmission strategy for wireless energy transfer in TWR networks was studied in [21], [22].

Compared to the stochastic EH models, the deterministic EH models need accurate EH prediction, and modeling mismatch usually occurs when the prediction interval is enlarged or the model does not conform with realistic conditions. Further, in order to analyze more realistic performance characteristics, it is essential to consider real-data-driven stochastic EH models in the design of EH communication networks. To the best of our knowledge, the optimal transmission policy for EH TWR networks with data-driven stochastic EH models has not been well studied.

Many of today's mobile radio systems carry real-time services, for which constant-rate and delay-limited transmission should be considered [23]. Moreover, although variable-rate transmission could improve throughput by dynamically adjusting the modulation and coding schemes, wireless nodes must be equipped with powerful processors. In practice, constant-rate transmission could be a better choice for large-scale deployments of the low-cost and power-limited EH networks [5]. In such scenarios of constant-rate transmissions, the information outage probability is an appropriate performance limit indicator [23]. However, most of the research works on EH cooperative communications focused on the throughput maximization, while the outage probability performance in EH TWR networks is still unknown. Further, the EH techniques have the potential to address the tradeoff between lifetime and performance of wireless nodes [1]. Hence, in order to satisfy the conflicting design goals of lifetime and performance, it is reasonable to metric the system performance of the EH TWR network from the perspective of long-term.

Motivated by the aforementioned discussions, in this paper, we propose an optimal relay transmission policy for optimizing the long-term outage performance of the EH TWR network with the data-driven stochastic solar EH model in [6]. In this network, two source nodes are traditional wireless nodes, while a solar-powered EH relay node is deployed in between them with a finite-sized battery and exploits decode-and-forward (DF) or amplify-and-forward (AF) cooperation protocols. Our objective is to minimize the long-term average outage probability by adapting the relay transmission power to the relay's knowledge of its current battery energy, channel states and causal solar ESI. The main contributions of this paper are summarized as follows:

- First, we formulate a Markov decision process (MDP) optimization framework for EH TWR networks, wherein the Gaussian mixture hidden Markov chain in [6] is used as our stochastic EH model, the fading channels between the sources and relay are formulated by a finite-state Markov model [24], [25], the battery capacity is quantized in units of energy quanta, and the system action represents the relay transmission power.
- We then calculate the conditional outage probabilities for both DF and AF protocols, which are deemed as the reward functions in the MDP. The conditional outage probability is defined as the outage probability conditioned on preset fading channel states, which is different from the traditional outage probability that regards the fading channel power as continuous values ranging from zero to infinity. We derive the exact closed-form and tight lower bound of the conditional outage probabilities for the DF and AF protocols, respectively.
- In the MDP formulation, the utility function is the expected long-term total discounted reward. In order to study the optimal transmission policy, we first analyze the property of the expected total discounted reward, and uncover a monotonic and bounded differential structure, which reveals that the policy value is non-increasing with the amount of the harvested energy in the battery, and the difference value of the expected total discounted rewards for two adjacent battery states is finite and bounded by one.
- Furthermore, we provide mathematical insights on the optimal relay transmission power, and find out a ceiling structure for both the AF and DF protocols, which indicates that the optimal relay power cannot be larger than a threshold power. Moreover, it is pointed out that the optimal transmission policy has a threshold structure, and it is equivalent to an "on-off" policy in sufficiently high SNRs.
- Finally, an interesting saturated structure for the expected outage probability is found in EH TWR networks with the AF or DF protocols. The analysis concludes that the expected outage probability converges to the battery empty probability in extremely high SNR regimes,

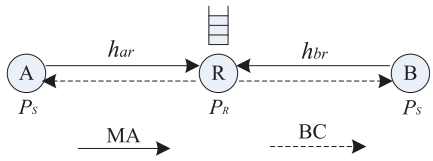


Fig. 1. EH TWR networks.

instead of going to zero. Moreover, a saturation-free condition that guarantees the battery empty probability and the expected outage probability are equal to zero in sufficiently high SNRs is provided. These results can answer the following questions: what is the fundamental limit of the outage performance in the TWR network with stochastic EH? How can we eliminate or relieve the performance saturation problem?

The rest of this paper is organized as follows. Section II introduces the EH TWR network and defines the outage probabilities for AF and DF protocols. The MDP formulation of the system is presented in Section III. Section IV analyzes the optimal transmission policy. The performance of the expected outage probability is studied in Section V. Simulation results are presented in Section VI. Finally, Section VII concludes the paper.

## II. Energy Harvesting Two-Way Relay Network

An EH TWR network is considered in Fig. 1, where two traditional wireless source nodes, A and B, exchange information simultaneously via an EH relay node, R, by utilizing a two-phase transmission protocol. The transmission duration is comprised of a multiple access (MA) phase and a broadcast (BC) phase. The relay has the ability to harvest energy from the solar and stores its harvested energy in the rechargeable battery to supply the forthcoming communications. It is assumed that each node is operated in a half-duplex mode and equipped with a single antenna. The two source nodes A and B have the identical and constant transmission power $P_S$, while the transmission power of R is set as $P_R$. We also assume that there is no direct link between the two source nodes, and the wireless channels are reciprocal, quasi-static and Rayleigh flat fading. That is, the channel coefficients, $h_{ar}$ and $h_{br}$, are independent and identically distributed (i.i.d.) complex Gaussian random variables with $\mathcal{CN}(0, \theta)$. Further, the relay can send pilot signals periodically to the two source nodes, which can estimate the channel state information (CSI) and feedback it to the relay. Hence, it is assumed that the relay has the perfect knowledge of the CSI of the two-hop links. Define $\gamma_1 = |h_{ar}|^2$ and $\gamma_2 = |h_{br}|^2$ as the instantaneous channel power with exponential distribution and mean $\theta$. The above network architecture is typical in wireless sensor networks or Ad Hoc networks [26]. For example, two user nodes supplied with constant power or large batteries exchange information with each other under the help of an EH relay node. Since the user nodes are deployed in the fixed locations or move in the low speed to transceive the data with low rate, the wireless channels can be regarded as very-slow and flat fading.

The two-phase transmission scheme is elaborated as follows. In the MA phase, the nodes A and B transmit their signals to R concurrently, while in the BC phase, the relay makes use of either amplify-and-forward (AF) or decode-and-forward (DF) cooperation protocols to broadcast the received signals to A and B [9]. For simplicity, we assume that the relative time durations of the MA phase and the BC phase are identical. Let $R_1$ and $R_2$ represent the achievable data rates of the A-B link and the B-A link, respectively. In the following, we discuss the achievable rate pair $(R_1, R_2)$ and the outage probabilities for the two cooperation protocols, i.e., DF and AF protocols.

### A. Decode-and-Forward

When the DF cooperation protocol is applied, the achievable data rate cannot be larger than the minimum of the two mutual information of the two transmission phases, and the achievable rates must satisfy a sum-rate constraint due to decoding two received signals simultaneously in the MA phase [15], [16]. Thus, the achievable rate pair $(R_1, R_2)$ is given as

$$R_1 \le \min\left\{\frac{1}{2}\log\left(1+\frac{\gamma_1 P_S}{N_0}\right), \frac{1}{2}\log\left(1+\frac{\gamma_2 P_R}{N_0}\right)\right\}, \tag{1}$$

$$R_2 \le \min\left\{\frac{1}{2}\log\left(1+\frac{\gamma_2 P_S}{N_0}\right), \frac{1}{2}\log\left(1+\frac{\gamma_1 P_R}{N_0}\right)\right\}, \tag{2}$$

$$R_1 + R_2 \le \frac{1}{2}\log\left(1+\frac{\gamma_1 P_S}{N_0}+\frac{\gamma_2 P_S}{N_0}\right), \tag{3}$$

where $N_0$ is the additive white Gaussian noise (AWGN) power at each node. Based on the achievable rate pair in (1)–(3), the following outage events can be defined [27], [28]

$$\mathcal{E}^1_{out,DF} = \left\{\min\left\{\frac{1}{2}\log\left(1+\frac{\gamma_1 P_S}{N_0}\right), \frac{1}{2}\log\left(1+\frac{\gamma_2 P_R}{N_0}\right)\right\} < R_{th1}\right\}, \tag{4}$$

$$\mathcal{E}^2_{out,DF} = \left\{\min\left\{\frac{1}{2}\log\left(1+\frac{\gamma_2 P_S}{N_0}\right), \frac{1}{2}\log\left(1+\frac{\gamma_1 P_R}{N_0}\right)\right\} < R_{th2}\right\}, \tag{5}$$

$$\mathcal{E}^3_{out,DF} = \left\{\frac{1}{2}\log\left(1+\frac{\gamma_1 P_S}{N_0}+\frac{\gamma_2 P_S}{N_0}\right) < (R_{th1}+R_{th2})\right\}, \tag{6}$$

where $R_{th1}$ and $R_{th2}$ are the target rates for the nodes A and B, respectively. We say the network experiences outage, if any of the three outage events in (4)–(6) occurs. Accordingly, the outage probability of the TWR network adopting the DF cooperation protocol is defined as

$$P_{out,DF} = \Pr\left\{\mathcal{E}^1_{out,DF} \cup \mathcal{E}^2_{out,DF} \cup \mathcal{E}^3_{out,DF}\right\}. \tag{7}$$

### B. Amplify-and-Forward

When the AF cooperation protocol is applied, the relay amplifies the received signals and forwards them to the two nodes A and B. Thus, the achievable data rates $R_1$ and $R_2$ cannot be larger than the mutual information computed by the

corresponding end-to-end SNRs of the two links. From [15] and [16], the achievable rate pair $(R_1, R_2)$ can be expressed as

$$R_1 \le \frac{1}{2}\log\left(1+\frac{\gamma_1\gamma_2 P_S P_R}{N_0\left(\gamma_1 P_S+\gamma_2 P_S+\gamma_2 P_R+N_0\right)}\right) = \frac{1}{2}\log\left[1+\frac{\gamma_1\gamma_2\eta\eta_r}{\gamma_1\eta+\gamma_2\left(\eta+\eta_r\right)+1}\right], \quad (8)$$

$$R_2 \le \frac{1}{2}\log\left(1+\frac{\gamma_1\gamma_2 P_S P_R}{N_0\left(\gamma_1 P_S+\gamma_2 P_S+\gamma_1 P_R+N_0\right)}\right) = \frac{1}{2}\log\left[1+\frac{\gamma_1\gamma_2\eta\eta_r}{\gamma_1\left(\eta+\eta_r\right)+\gamma_2\eta+1}\right], \quad (9)$$

where we define $\eta = \frac{P_S}{N_0}$ and $\eta_r = \frac{P_R}{N_0}$. Similar to the DF protocol, two outage events with respect to $R_1$ and $R_2$ are defined as [27], [28]

$$\mathcal{E}^1_{out,AF} = \left\{\frac{1}{2}\log\left[1+\frac{\gamma_1\gamma_2\eta\eta_r}{\gamma_1\eta+\gamma_2\left(\eta+\eta_r\right)+1}\right] < R_{th1}\right\}, \quad (10)$$

$$\mathcal{E}^2_{out,AF} = \left\{\frac{1}{2}\log\left[1+\frac{\gamma_1\gamma_2\eta\eta_r}{\gamma_1\left(\eta+\eta_r\right)+\gamma_2\eta+1}\right] < R_{th2}\right\}. \quad (11)$$

As a result, the outage probability of the TWR network using the AF cooperation protocol is defined as

$$P_{out,AF} = \Pr\left\{\mathcal{E}^1_{out,AF} \cup \mathcal{E}^2_{out,AF}\right\}. \quad (12)$$

## III. Markov Decision Process With Stochastic Models

Our objective is to find the optimal transmission policy for the relay in order to minimize the long-term average outage probability of the TWR network. Since the wireless channel conditions and solar irradiance conditions are dynamic and even unpredictable in EH wireless networks, the design of the relay transmission policy is influenced by a couple of factors such as the finite battery capacity, the solar EH conditions at the relay, and the channel conditions among the three nodes. The design framework is then formulated as an MDP with the goal of minimizing the long-term average outage probability. The main components in the MDP model include states, actions and reward functions which represent the system conditions, the relay transmission power and the outage probabilities, respectively. The transmission policy is managed in the time scale of $T_M$. The detailed descriptions of all these fundamental elements are introduced as follows.

### A. Relay Actions of Transmission Power

Let $\mathcal{W} = \{0, 1, \cdots, N_p - 1\}$ represent an action set of relay transmission power. When the power action $W = w \in \mathcal{W}$ is taken, the relay transmission power $P_R$ is set as $wP_u$ during one policy management period $T_M$, where $P_u$ is the basic transmission power corresponding to one energy quantum $E_u$ during a half policy management period $\frac{T_M}{2}$, i.e., $E_u = P_u \cdot \frac{T_M}{2}$. Particularly, if $w = 0$, it means that the relay keeps silent during the transmission period.

### B. System States

Let $\mathcal{S} = \mathcal{Q}_e \times \mathcal{Q}_b \times \mathcal{H}_{ar} \times \mathcal{H}_{br}$ be a four-tuple state space, where $\times$ denotes the Cartesian product, $\mathcal{Q}_e = \{0, 1, \cdots, N_e - 1\}$ represents a solar EH state set, $\mathcal{Q}_b = \{0, 1, \cdots, N_b - 1\}$ denotes a finite battery state set for the relay node, $\mathcal{H}_{ar} = \{0, 1, \cdots, N_c - 1\}$ and $\mathcal{H}_{br} = \{0, 1, \cdots, N_c - 1\}$ are the channel state sets of $h_{ar}$ and $h_{br}$, respectively. Meanwhile, define a random variable $S = (Q_e, Q_b, H_{ar}, H_{br}) \in \mathcal{S}$ as the system stochastic state of the MDP, which remains steady during one policy period $T_M$. In the following, we discuss the detailed definition of each state in sequence.

*(a) Solar EH State:* An $N_e$-state stochastic EH model in [6] is exploited to mimic the evolution of the solar EH conditions. This EH model is a real-data-driven Markov chain model, and its underlying parameters are extracted using the solar irradiance data collected by a solar site in Elizabeth City State University [29]. Since the solar irradiance data were measured from the early morning (seven o'clock) to the late afternoon (seventeen o'clock) every day in the month of June, the solar EH model with its underlying parameters in [6] and the EH network are applied to the scenario of daylight. Therein, it is assumed that if the solar EH state is given by $Q_e = e \in \mathcal{Q}_e$, the harvested solar power per unit area, $P_h$, is a continuous random variable with Gaussian distribution $\mathcal{N}(\mu_e, \rho_e)$. Therefore, different solar EH states result in different solar irradiance intensities. Moreover, the dynamic of the states is governed by a state transition probability $P\left(Q_e = e' | Q_e = e\right), \forall e, e' \in \mathcal{Q}_e$ [6].

It is assumed that the solar EH condition is quasi-static during one policy period $T_M$. Thus, the harvested solar energy during one period $T_M$ can be computed as $E_h = P_h T_M \Omega \eta$, where $\Omega$ is the solar panel area size and $\eta$ denotes the energy conversion efficiency. We utilize the quantization model to deal with the harvested energy, which is first quantized in unit of $E_u$ and then stored in the battery for data transmission. Accordingly, the probability of the number of harvested energy quanta conditioned on the $e^{th}$ solar EH state, denoted as $P(Q = q | Q_e = e)$ for $q \in \{0, 1, \cdots, \infty\}$, is theoretically derived and provided in [6], which enables us to capture the impact of the parameters of the solar state and the energy storage system on the energy supporting condition.

*(b) Battery State:* The battery state stands for the available amount of energy quanta in the battery. If the relay is at the battery state $Q_b = b \in \mathcal{Q}_b$, the number of available energy quanta in the battery is given by $b$, i.e., the available energy is $bE_u$. We utilize the *harvest-store-use* model [30], which means the energy harvested in the current policy period is first stored in the battery, and then consumed for the data transmission in the next policy period. Thus, the battery state transition from the current state $b$ to the next state $b'$ can be expressed as

$$b' = b - w + q, \quad (13)$$

where $w$ and $q$ denote the relay power action and the number of the harvested energy quanta in the current policy period, respectively. Further, it implies that the maximum affordable power action is restricted to the current battery state,

i.e., $w \in \{0, 1, \cdots, \min(b, N_p - 1)\}$. Therefore, the battery state transition probability at the $e^{th}$ solar EH state with respect to the power action $w$ can be expressed as

$$P_w\left(Q_b = b' | Q_b = b, Q_e = e\right) = \begin{cases} P\left(Q = b' - b + w | Q_e = e\right), b' = (b-w), \cdots, N_b - 2 \\ 1 - \sum_{q=0}^{N_b-2-b+w} P\left(Q = q | Q_e = e\right), b' = N_b - 1. \end{cases} \tag{14}$$

The first term in (14) represents the condition that $b'$ is smaller than $N_b - 1$, and thus $q$ is qual to $b' - b + w$ at this time from (13). Accordingly, the second term in (14) denotes the condition that the next battery state is full. Since we assume the harvested energy quanta are discarded when the battery is full, $q$ cannot be greater than $N_b - 1 - b + w$.

*(c) Channel States:* The wireless channel variation from one level to another is formulated by a finite-state Markov chain model [24], and the validity and accuracy of this model were confirmed by the state equilibrium equations and computer simulations in [24] and [25]. The instantaneous channel gains, $\gamma_1$ and $\gamma_2$, are quantized into $N_c$ levels using a finite number of thresholds, given by $\Gamma = \{0 = \Gamma_0, \Gamma_1, \cdots, \Gamma_{N_c} = \infty\}$. If the channel gain belongs to the $i^{th}$ channel interval $[\Gamma_i, \Gamma_{i+1})$, the corresponding fading channel is said to be in the $i^{th}$ channel state, for $i \in \{0, 1, \cdots, N_c - 1\}$.

Moreover, since the wireless channels are Rayleigh fading, the stationary probability of the $i^{th}$ channel state can be expressed as

$$P(H = i) = \int_{\Gamma_i}^{\Gamma_{i+1}} \frac{1}{\theta} \exp\left(-\frac{\gamma}{\theta}\right) d\gamma = \exp\left(-\frac{\Gamma_i}{\theta}\right) - \exp\left(-\frac{\Gamma_{i+1}}{\theta}\right), \tag{15}$$

where $\theta$ is the average channel power. It is also assumed that the wireless channel fluctuates slowly and the channel gain remains constant during one policy management period. Further, the wireless channel can only transit from the current state to its neighboring states, and the channel state transition probability $P(H = j | H = i)$, for $i \in \{0, \cdots, N_c - 1\}$, $j \in \{\max(0, i-1), \cdots, \min(i+1, N_c - 1)\}$, is defined in [24].

*(d) MDP State Transition:* Since the solar irradiance and the wireless fading channels are independent with each other, the system state transition probability from the state $s = (e, b, h, g)$ to the state $s' = (e', b', h', g')$ associated with the relay power action $w$ can be computed as

$$P_w\left(s' | s\right) = P\left(Q_e = e' | Q_e = e\right) \cdot P\left(H_{ar} = h' | H_{ar} = h\right) \cdot P\left(H_{br} = g' | H_{br} = g\right) \cdot P_w\left(Q_b = b' | Q_b = b, Q_e = e\right). \tag{16}$$

### C. Reward Function

Here the conditional outage probability for a relay power action at a fixed system state within one policy management period $T_M$ is utilized as our reward function in the MDP. Due to the fact that the immediate reward is independent of the battery state and the solar state, the reward function at the system state $s = (e, b, h, g) \in \mathcal{S}$ with respect to the relay action $w \in \mathcal{W}$ can be simplified as

$$R_{w,f}(s) = \Pr\{\text{Outage event} | w, f, s\} \triangleq P_{out,f}(w, h, g), \tag{17}$$

where $f \in \{DF, AF\}$ represents the cooperation protocol exploited at the relay. According to the definition of the outage probabilities in (7) and (12), the conditional outage probabilities for the DF and AF protocols can be expressed as

$$P_{out,DF}(w, h, g) = \Pr\left\{\bigcup_{i=1}^{3} \mathcal{E}_{out,DF}^{i} | P_R = wP_u, H_{ar} = h, H_{br} = g\right\}, \tag{18}$$

$$P_{out,AF}(w, h, g) = \Pr\left\{\bigcup_{i=1}^{2} \mathcal{E}_{out,AF}^{i} | P_R = wP_u, H_{ar} = h, H_{br} = g\right\}, \tag{19}$$

and they are explicitly calculated in Proposition 1 and Proposition 2, respectively.

$$T = \begin{cases} \left(e^{-a/\theta} - e^{-\Gamma_{h+1}/\theta}\right) \cdot \left(e^{-b/\theta} - e^{-\Gamma_{g+1}/\theta}\right), & c \geq \Gamma_{h+1} + \Gamma_{g+1}; \\ 0, & c \leq a + b; \\ e^{-(a+b)/\theta} - e^{-c/\theta} - \frac{1}{\theta} e^{-c/\theta}(c - b - a), & a + b < c \leq \min\left\{(a + \Gamma_{g+1}), (b + \Gamma_{h+1})\right\}; \\ e^{-(a+b)/\theta} - e^{-(\Gamma_{h+1}+b)/\theta} - \frac{1}{\theta} e^{-c/\theta}(\Gamma_{h+1} - a), & (b + \Gamma_{h+1}) < c < (a + \Gamma_{g+1}); \\ e^{-(a+b)/\theta} - e^{-(a+\Gamma_{g+1})/\theta} - \frac{1}{\theta} e^{-c/\theta}(\Gamma_{g+1} - b), & (a + \Gamma_{g+1}) < c < (b + \Gamma_{h+1}); \\ \left(e^{-a/\theta} - e^{-\Gamma_{h+1}/\theta}\right)\left(e^{-b/\theta} - e^{-\Gamma_{g+1}/\theta}\right) - e^{-(\Gamma_{h+1}+\Gamma_{g+1})/\theta} + e^{-c/\theta} + \frac{1}{\theta} e^{-c/\theta}(c - \Gamma_{g+1} - \Gamma_{h+1}), & \\ \qquad \max\left\{(a + \Gamma_{g+1}), (b + \Gamma_{h+1})\right\} \leq c < (\Gamma_{h+1} + \Gamma_{g+1}). \end{cases} \tag{20}$$

with $a = \max\{\gamma_{th1}, \Gamma_h\}$, $b = \max\{\gamma_{th2}, \Gamma_g\}$, $c = \frac{N_0}{P_S}\left(2^{2(R_{th1}+R_{th2})} - 1\right)$.

$$P_{out,AF}(w, h, g) \begin{cases} = 1, & (\gamma_{th1} \geq \Gamma_{h+1}) \text{ or } (\gamma_{th2} \geq \Gamma_{h+1}) \text{ or } (\gamma_{th3} \geq \Gamma_{g+1}) \text{ or } (\gamma_{th4} \geq \Gamma_{g+1}); \\ = 0, & (\gamma_{th1} \leq \Gamma_h) \text{ and } (\gamma_{th2} \leq \Gamma_h) \text{ and } (\gamma_{th3} \leq \Gamma_g) \text{ and } (\gamma_{th4} \leq \Gamma_g); \\ \geq 1 - \frac{e^{-\max(\gamma_{th1}, \gamma_{th2})/\theta} - e^{-\Gamma_{h+1}/\theta}}{e^{-\Gamma_h/\theta} - e^{-\Gamma_{h+1}/\theta}} \cdot \frac{e^{-\max(\gamma_{th3}, \gamma_{th4})/\theta} - e^{-\Gamma_{g+1}/\theta}}{e^{-\Gamma_g/\theta} - e^{-\Gamma_{g+1}/\theta}}, & \text{otherwise;} \end{cases} \tag{21}$$

$$\gamma_{th1} = \frac{(P_S + wP_u) N_0}{P_S \cdot wP_u}\left(2^{2R_{th1}} - 1\right), \gamma_{th2} = \frac{N_0}{wP_u}\left(2^{2R_{th2}} - 1\right), \gamma_{th3} = \frac{N_0}{wP_u}\left(2^{2R_{th1}} - 1\right), \gamma_{th4} = \frac{(P_S + wP_u) N_0}{P_S \cdot wP_u}\left(2^{2R_{th2}} - 1\right).$$

*Proposition 1:* For the given target rate pair $(R_{th1}, R_{th2})$, the conditional outage probability of the TWR network using the DF cooperation protocol with respect to the system state $s = (e, b, h, g)$ and relay power action $w$ can be expressed as

$$P_{out,DF}(w,h,g) = \begin{cases} 1, & (\gamma_{th1} \geq \Gamma_{h+1}) \text{ or } (\gamma_{th2} \geq \Gamma_{g+1}); \\ 1 + \frac{T - \left(e^{-a/\theta} - e^{-\Gamma_{h+1}/\theta}\right)\cdot\left(e^{-b/\theta} - e^{-\Gamma_{g+1}/\theta}\right)}{\left(e^{-\Gamma_h/\theta} - e^{-\Gamma_{h+1}/\theta}\right)\cdot\left(e^{-\Gamma_g/\theta} - e^{-\Gamma_{g+1}/\theta}\right)}, & (\gamma_{th1} < \Gamma_{h+1}) \text{ and } (\gamma_{th2} < \Gamma_{g+1}); \end{cases}$$

where $\gamma_{th1} = \max\left\{\frac{N_0}{P_S}\left(2^{2R_{th1}} - 1\right), \frac{N_0}{wP_u}\left(2^{2R_{th2}} - 1\right)\right\}$,

$\gamma_{th2} = \max\left\{\frac{N_0}{wP_u}\left(2^{2R_{th1}} - 1\right), \frac{N_0}{P_S}\left(2^{2R_{th2}} - 1\right)\right\}$,

and the term $T$ is defined as (20), shown at the bottom of the previous page.

*Proof:* See Appendix A for details. ■

*Proposition 2:* For the given target rate pair $(R_{th1}, R_{th2})$, the conditional outage probability of the TWR network in high SNR regimes using the AF cooperation protocol with respect to the system state $s = (e, b, h, g)$ and relay power action $w$ can be expressed as (21), shown at the bottom of the previous page.

*Proof:* See Appendix B for details. ■

*Remark 1:* From (17), Proposition 1 and Proposition 2, the reward functions for a given target rate pair both have the following two essential properties:

$$R_{w=0}(s) = P_{out}(h, g, w = 0) = 1; \tag{22}$$

$$\lim_{N_0 \to 0, w \geq 1} R_w(s) = \lim_{N_0 \to 0, w \geq 1} P_{out}(h, g, w) = 0. \tag{23}$$

In (22), this remark implicitly indicates that when the relay remains silent, the network is in outage and the corresponding conditional outage probability is equal to one. On the other hand, it is observed from (23) that when the SNR is sufficiently high, i.e., $N_0$ approaches to zero, it suffices to spend only one energy quantum for achieving zero outage probability under any target rate pair and channel states.

### D. Optimization of Relay Transmission Policy

The policy $\pi(s) : \mathcal{S} \to \mathcal{W}$ is defined as the action that indicates the relay transmission power with respect to a given system state. The goal of the MDP is to find the optimal $\pi(s)$ in the state $s$ that minimizes the expected long-term total discounted reward as follows

$$V_\pi(s_0) = \mathbb{E}_\pi\left\{\sum_{k=0}^{\infty} \lambda^k R_{\pi(s_k)}(s_k)\right\}, s_k \in \mathcal{S}, \pi(s_k) \in \mathcal{W}, \tag{24}$$

where $s_0$ is the initial state, $\mathbb{E}_\pi\{\cdot\}$ denotes the expected value conditioned on the policy $\pi$, and $0 \leq \lambda < 1$ is a discount factor. Moreover, by assuming that the states of the Markov chain are recurrent, the optimal value of the expected reward is unrelated to the initial state, and thus the optimal policy for minimizing (24) can be found through the Bellman equation, given by

$$V_{\pi^*}(s) = \min_{w \in W}\left(R_w(s) + \lambda \sum\nolimits_{s' \in \mathcal{S}} P_w\left(s'|s\right) V_{\pi^*}\left(s'\right)\right), \tag{25}$$

which can be efficiently implemented by executing the well-known value iterations [31]:

$$V_w^{(i+1)}(s) = R_w(s) + \lambda \sum\nolimits_{s' \in S} P_w\left(s'|s\right) V^{(i)}\left(s'\right), \tag{26}$$

$$V_{(i+1)}(s) = \min_{w \in \mathcal{W}}\left(V_w^{(i+1)}(s)\right), \tag{27}$$

where $i$ is the iteration number. The value iteration algorithm alternates until a stopping criterion, $\left|V^{(i+1)} - V^{(i)}\right| \leq \varepsilon$, is satisfied.

In the following, we will discuss the special properties of the optimal policy, and it is worth mentioning that the derived results are applied to both the DF and AF protocols in the following formulas and theorems. For the purpose of simple notations and from (14) and (16), the summation term in (26) can be rewritten as

$$\begin{aligned} &\sum_{s' \in S} P_w\left(s'|s\right) V^{(i)}\left(s'\right) = \sum\nolimits_{e',h',g'} P\left(Q_e = e'|Q_e = e\right) \\ &\quad \cdot P\left(H_{ar} = h'|H_{ar} = h\right) P\left(H_{br} = g'|H_{br} = g\right) \\ &\quad \cdot \sum_{q=0}^{\infty} P\left(Q = q|Q_e = e\right) \cdot V^{(i)}\left(e', \min(b - w + q, N_b - 1), h', g'\right) \\ &= \mathbb{E}_s\left\{V^{(i)}\left(e', \min\left(b - w + q, N_b - 1\right), h', g'\right)\right\} \end{aligned} \tag{28}$$

where the change of variable is applied, i.e., $b'$ is replaced with the number of harvested energy quanta $q$, and $\mathbb{E}_s\{\cdot\}$ denotes the expected value with respect to $s'$ conditioned on the state $s = (e, b, h, g)$.

## IV. Optimal Transmission Policy

### A. Monotonic and Bounded Differential Structure of Expected Total Discounted Reward

*Lemma 1:* Assume that the initial condition $V^{(0)}(s) = 0, \forall s \in \mathcal{S}$. For any fixed system state $s = (e, b > 0, h, g) \in \mathcal{S}$ in the $i^{th}$ $(i \geq 1)$ value iteration, the expected total discounted reward is non-increasing in the battery state, and the difference value of the expected total discounted rewards for two adjacent battery states is non-negative and not larger than one, i.e., $1 \geq V^{(i)}(e, b - 1, h, g) - V_{(i)}(e, b, h, g) \geq 0$, $\forall b \in \mathcal{Q}_b \backslash \{0\}$. Moreover, the optimal transmission policy $\pi^*$ is also satisfied with the above special structure, i.e., $1 \geq V_{\pi^*}(e, b - 1, h, g) - V_{\pi^*}(e, b, h, g) \geq 0, \forall b \in \mathcal{Q}_b \backslash \{0\}$.

*Proof:* See Appendix C for details. ■

This monotonic structure points out the relationship between the expected long-term total discounted reward and the battery state, for which the outage performance is better when there is more energy in the battery. Moreover, the bounded differential structure is derived from the outage probability's characteristic of bounded values, and it concludes that the difference value of the expected total discounted reward caused by the increased battery energy is bounded.

### B. Ceiling Structure and Threshold Structure of Optimal Relay Power Action

Now we turn to analyzing the structure of the optimal relay transmission power action. Since the relay transmission power must be equal to zero when the battery is empty, we focus on the remaining case of non-empty battery, $b > 0$, in this subsection.

*Definition 1:* (Ceiling Power) For any fixed channel states $h \in \mathcal{H}_{ar}$ and $g \in \mathcal{H}_{br}$, and cooperation protocol $f \in \{DF, AF\}$, a power action level $\tilde{w}$ is called ceiling power, if the reward functions begin to be unchanged when the relay power action is equal to or larger than $\tilde{w}$, i.e., $R_{w,f}(h,g) > R_{\tilde{w},f}(h,g), \forall w < \tilde{w}$, and $R_{w,f}(h,g) = R_{\tilde{w},f}(h,g), \forall w \geq \tilde{w}$.

*Remark 2:* According to Definition 1, the feasible ceiling power is given by $0 < \tilde{w} \leq N_b - 1$, and it is related to the channel states, the source transmission power, the noise power at nodes, etc. From (23), when the system is operated in sufficiently high SNR regimes, i.e., $N_0 \to 0$, the relay's ceiling power is equal to $\tilde{w} = 1$, for $\forall f \in \{DF, AF\}$.

To get more insight into the optimal policy, a relationship between the relay's ceiling power and the optimal transmission power action is established in the following theorem.

*Theorem 1:* For any fixed system state $s = (e, b, h, g) \in \mathcal{S}$, the optimal relay power action is not larger than the relay's ceiling power, i.e., $w^* \leq \min(\tilde{w}, b)$.

*Proof:* See Appendix D for details. ■

*Corollary 1:* For any fixed system state $s = (e, b, h, g) \in \mathcal{S}$, the optimal relay power action $w^*$ takes a value of either zero or one in sufficiently high SNRs.

*Proof:* According to Definition 1 and Remark 2, the relay's ceiling power is given by $\tilde{w} = 1$ in high SNR regimes. By applying Theorem 1, it is sufficient to prove that the optimal relay power action $w^*$ is equal to 0 or 1 when the system is operated in sufficiently high SNRs. ■

From (13), it implies that the affordable power action is restricted to the current battery state, i.e., $w \leq b$. Thus, the transmission policy for the relay node is only to keep silent when the battery is empty, i.e., $w^* = 0$ when $b = 0$. In the following, we discuss the optimal relay transmission policy when the battery is non-empty.

*Theorem 2:* For any fixed system state $s = (e, b > 0, h, g) \in \mathcal{S}$ with the non-empty battery, the optimal relay power action $w^*$ must be equal to one in sufficiently high SNRs.

*Proof:* According to Corollary 1, the optimal relay power action in sufficiently high SNRs is given by $w^* = 0$ or $w^* = 1$ when the battery state $b \in \mathcal{Q}_b \backslash \{0\}$. For any iteration $i$ and system state $s = (e, b > 0, h, g) \in \mathcal{S}$, according to (28), the value difference of the two expected total discounted rewards for the relay power action $w = 1$ and $w = 0$ can be expressed as

$$\begin{aligned} V_{w=1}^{(i+1)}(s) - V_{w=0}^{(i+1)}(s) &= R_{w=1}(h,g) - R_{w=0}(h,g) \\ &+ \lambda \cdot \mathbb{E}_s \Big\{ V^{(i)}\left(e', \min(b-1+q, N_b-1), h', g'\right) \\ &- V^{(i)}\left(e', \min(b+q, N_b-1), h', g'\right) \Big\}. \end{aligned} \tag{29}$$

By using (23), the value difference in (29) in high SNRs is written as

$$\begin{aligned} &\lim_{N_0 \to 0} \left[ V_{w=1}^{(i+1)}(s) - V_{w=0}^{(i+1)}(s) \right] \\ &= -1 + \lambda \cdot \lim_{N_0 \to 0} \mathbb{E}_s \Big\{ V^{(i)}\left(e', \min(b-1+q, N_b-1), h', g'\right) \\ &\quad - V^{(i)}\left(e', \min(b+q, N_b-1), h', g'\right) \Big\}. \end{aligned} \tag{30}$$

By applying Lemma 1, for any system state $s' \in \mathcal{S}$, the value difference in the expectation form in (30) is non-negative and not larger than one. Since $0 < \lambda < 1$, the two expected total discounted rewards in (29) in high SNRs meet the following relationship

$$\lim_{N_0 \to 0} V_{w=1}^{(i+1)}(e,b,h,g) < \lim_{N_0 \to 0} V_{w=0}^{(i+1)}(e,b,h,g). \tag{31}$$

From (27), the optimal relay power action in iteration $i+1$ is given by $w^{*,(i+1)} = 1$. When the value iteration algorithm is converged, the optimal relay power action is also given as $w^* = 1$. ■

The above theorem implicitly indicates that the proposed optimal policy has an "on-off" threshold structure in high SNR regimes, which means it suffices to attain the best long-term performance by only spending one energy quantum for relaying the signals when the battery is non-empty, or the relay keeps silent when the battery is empty. Although Theorem 1 and Theorem 2 are proved by applying Lemma 1, which is based on the initial condition $V^{(0)}(s) = 0, \forall s \in \mathcal{S}$, the results on the optimal policy in this subsection are general in our system. This is because for a given small quantity $\varepsilon$, no matter if the initial values of all states are identical or not, a stationary optimal policy $\pi^*$ can be achieved through the value iteration algorithm [31].

## V. Performance Analysis of Outage Probability

With the special structures of our optimal transmission policy, the outage performances of the EH TWR network will be analyzed in this section.

### A. Expected Reward

We introduce the steps to compute the expected reward for any transmission policy $\pi$. First, the battery state transition probability associated with the transmission policy $\pi$ in the state $s = (e, b, h, g)$ can be derived as [20]

$$\begin{aligned} &P_\pi\left(Q_b = b' | Q_b = b\right) \\ &= \begin{cases} 0, & 0 \leq b' \leq b - w - 1; \\ P\left(Q = b' - b + w | Q_e = e\right), & b - w \leq b' \leq N_b - 2; \\ 1 - \sum\limits_{b'=0}^{N_b-2} P_\pi\left(Q_b = b' | Q_b = b\right), & b' = N_b - 1, \end{cases} \end{aligned} \tag{32}$$

where $b, b' \in \{0, \cdots, N_b - 1\}$ and $w$ is the relay power action in the policy $\pi$. By utilizing (16), the system state transition probability with respect to the policy $\pi$ can be calculated as

$$\begin{aligned} P_\pi\left(s'|s\right) = &P\left(Q_e = e' | Q_e = e\right) \cdot P\left(H_{ar} = h' | H_{ar} = h\right) \\ &\cdot P\left(H_{br} = g' | H_{br} = g\right) \cdot P_\pi\left(Q_b = b' | Q_b = b\right), \end{aligned} \tag{33}$$

where $h' \in \{\max(0, h-1), \cdots, \min(h+1, N_c-1)\}$, $g' \in \{\max(0, g-1), \cdots, \min(g+1, N_c-1)\}$, $e, e' \in \{0, 1, \cdots, N_e - 1\}$, and $h, g \in \{0, \cdots, N_c - 1\}$. Next, let $p_\pi(s = (e, b, h, g))$ represent the steady state probability of the state $s = (e, b, h, g)$ for the policy $\pi$, and the following linear equations can be formulated [20]:

$$\begin{cases} \sum_{s \in \mathcal{S}} p_\pi(s) = 1, \\ \sum_{s \in \mathcal{S}} P_\pi\left(s'|s\right) \cdot p_\pi(s) = p_\pi\left(s'\right). \end{cases} \tag{34}$$

Finally, after solving the aforementioned linear equations, the expected reward $\bar{R}$ can be computed by taking expectation over the reward function with respect to the obtained steady state probability as follows:

$$\bar{R} = \sum\nolimits_{s\in\mathcal{S}} p_\pi(s) \times R_{w=\pi(s)}(s). \quad (35)$$

Since the states of the Markov chain are assumed to be recurrent, the occurrence probability of the state $s$ is equal to $p_\pi(s)$ for the fixed policy $\pi$ after a long run time, and thus the expected reward $\bar{R}$ can be regarded as the long-term average reward denoted in (24).

### B. Saturated Structure of Outage Performance

The performance of the expected outage probability for the proposed optimal policy in high SNR regimes will be analyzed in this subsection. This help us capture the fundamental performance limit of the EH TWR networks when the noise power approaches to zero, as well as the effect of the randomness and uncertainty of the harvested energy on the outage performance.

*Definition 2:* (Battery Empty Probability) It is the steady state probability when the battery state is equal to zero for the policy $\pi$, i.e., $P_\pi(b=0) = \sum_{(e,b=0,h,g)\in\mathcal{S}} p_\pi(e,b=0,h,g)$.

*Theorem 3:* At sufficiently high SNRs, the expected outage probability for the proposed optimal policy $\pi^*$ is equal to the battery empty probability $P_{\pi^*}(b=0)$.

*Proof:* From (35) and considering the battery state, the expected reward of the optimal policy $\pi^*$ is expressed as

$$\begin{aligned}\bar{R} &= \sum\nolimits_{s\in\mathcal{S}} p_{\pi^*}(s) \times R_{w^*=\pi^*(s)}(s) \\ &= \sum\nolimits_{s\in\mathcal{S}} \big[p_{\pi^*}(e,b=0,h,g)\times R_{w^*}(e,b=0,h,g) \\ &\qquad + p_{\pi^*}(e,b\geq 1,h,g)\times R_{w^*}(e,b\geq 1,h,g)\big], \quad (36)\end{aligned}$$

where $p_{\pi^*}(s)$ is the steady state probability associated with the optimal policy $\pi^*$, and $w^*$ is the optimal relay action.

By applying Theorem 2, the optimal relay power action $w^*=1$ for $\forall s=(e,b>0,h,g)\in\mathcal{S}$ in sufficiently high SNRs. According to (23), the reward value is equal to zero when the relay transmission power is not zero in high SNRs, and thus the expected reward in high SNRs is expressed as

$$\lim_{N_0\to 0}\bar{R} = \sum_{e=0}^{N_e-1}\sum_{h=0}^{N_c-1}\sum_{g=0}^{N_c-1} p_{\pi^*}(e,b=0,h,g) = P_{\pi^*}(b=0), \quad (37)$$

where $P_{\pi^*}(b=0)$ denotes the battery empty probability with respect to the optimal policy $\pi^*$. Therefore, the expected reward of our proposed optimal policy is equal to the battery empty probability in high SNRs. ■

This theorem gives us an important insight into understanding the limitation of the expected outage probability, which indicates that the expected outage probability does not approach to zero when the SNR value goes infinity if the battery empty probability is non-zero. Under this circumstance, the outage probability gets saturated, and the reliable communications cannot be guaranteed. The battery empty probability for the proposed optimal policy can be calculated by using the system steady state probability in (35). In fact, to get rid of this saturation phenomenon, it requires a zero battery empty probability. In the following, we discuss the condition that guarantees to obtain the non-saturated outage probability in sufficiently high SNRs.

TABLE I
SIMULATION PARAMETERS

| | |
|---|---|
| Basic transmission power ($P_u$) | 35mW |
| Policy management period ($T_M$) | 300s |
| Number of solar states ($N_e$) | 4 |
| Solar panel area ($\Omega$) | 5cm$^2$ |
| Energy conversion efficiency ($\eta$) | 20% [1] |
| Average channel power ($\theta$) | 1 |
| Channel simulation model | Jakes' model |
| Number of channel states ($N_c$) | 6 |
| Channel quantization thresholds ($\Gamma$) | $\{0, 0.3, 0.6, 1.0, 2.0, 3.0, \infty\}$ |
| Discount factor ($\lambda$) | 0.99 |
| Stopping criterion parameter ($\varepsilon$) | $10^{-5}$ |
| Number of battery states ($N_b$) | 12 |
| Target rate proportion ($\sigma$) | 0.5 |

*Definition 3:* (Energy Deficiency Probability) It is the probability when the number of harvested energy quanta is equal to zero, conditioned on the solar EH state $Q_e = e \in \mathcal{Q}_e$, i.e., $P(Q=0|Q_e=e)$.

It can be observed from [6] that the energy deficiency probability $P(Q=0|Q_e=e)$ is affected by the solar panel size $\Omega$, the size of one energy quantum $E_u$, the policy management period $T_M$, the energy conversion efficiency $\eta$, as well as the mean and variance of the underlying Gaussian distribution in the stochastic solar EH model. Especially, the energy deficiency probability can be effectively reduced by increasing $\Omega$ or decreasing $E_u$.

*Corollary 2:* The expected outage probability for the proposed optimal policy $\pi^*$ goes to zero in sufficiently high SNR regimes, if and only if the energy deficiency probability is equal to zero, i.e., $P(Q=0|Q_e=e)=0, \forall e\in\mathcal{Q}_e$.

*Proof:* See Appendix E for details. ■

## VI. SIMULATION RESULTS

In this section, the long-term average outage probability of our proposed optimal policy based on the stochastic EH model in [6] is evaluated by computer simulations. For each SNR value, we calculate the reward function and solve the MDP to obtain the optimal policy, based on which the long-term average reward is derived. The analysis results of outage probabilities are calculated according to (35), Proposition 1, and Proposition 2, while the simulation results are computed using the Monte-Carlo method. We assume that a positive value $\sigma$ represents the proportion between the target rate $R_{th1}$ ($R_{th2}$) and target sum rate $R$, i.e., $R_{th1}=\sigma R$, $R_{th2}=(1-\sigma)R$. Main simulation parameters are listed in Table I, except as otherwise stated. The transmission power of wireless sensor nodes usually ranges from dozens of mW to hundreds of mW [1], [2], thus we set the basic transmission power $P_u$ as 35mW referring to [6]. Since the relay transmission power is related with the solar irradiance, a normalized SNR is defined with respect to the transmission power of $1mW$ in the simulations.

In (24), the expected long-term total discounted reward $V_\pi(s_0)$ is adopted as the policy value in the MDP formulation,

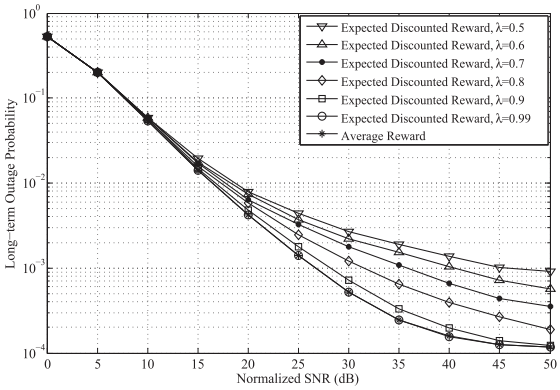


Fig. 2. Impact of discount factor λ on long-term average outage performance in DF mode ($P_S = 3P_u$, $R = 4$bit/s/Hz).

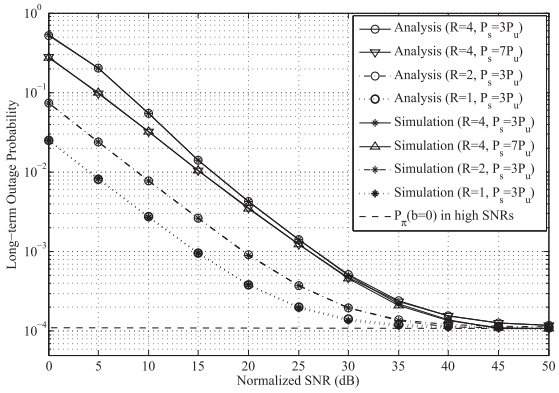


Fig. 3. Outage probability of DF mode for different target sum rates $R$(unit: bit/s/Hz) and source nodes' transmission power $P_S$.

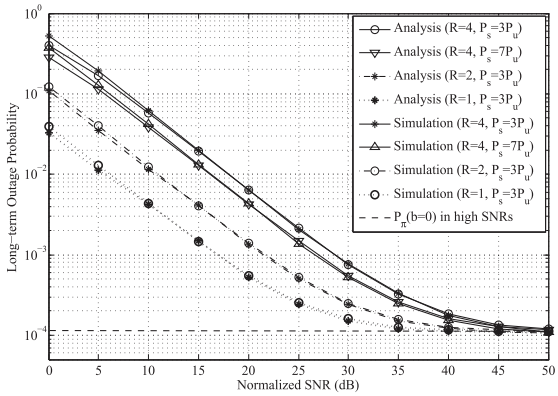


Fig. 4. Outage probability of AF mode for different target sum rates $R$(unit: bit/s/Hz) and source nodes' transmission power $P_S$.

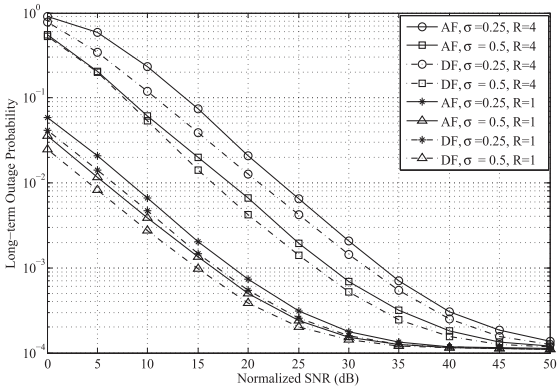


Fig. 5. Outage probability for different target rate proportions $\sigma$ and target sum rates $R$(unit: bit/s/Hz) in AF and DF modes ($P_S = 3P_u$).

and the adjustment of the discount factor λ provides a broad range of performance characteristics. Expect that, the long-term average reward, i.e., $\bar{V}_\pi(s_0) = \limsup_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} R_{\pi(s_k)}(s_k)$, can also be adopted as the policy value. Fig. 2 shows that the long-term average outage performances of the two kinds of optimal policies corresponding to these two policy values in the DF mode. The curves of expected total discounted reward and average reward represent the analysis results of the system performances by exploiting the optimal polices in the case of expected total discounted reward and average reward, respectively. The value iteration algorithm [31] is utilized to compute the optimal polices for the two kinds of policy values. It can be seen that the performance gap between these two kinds of optimal policies becomes smaller when λ approaches to 1, and can be negligible when $\lambda = 0.99$. Since the performance trends are identical for both AF and DF modes, we only demonstrate the performances in the DF mode. Thus, the average reward can be closely approximately optimized by utilizing the optimal transmission policy for the case of expected total discounted reward with $\lambda = 0.99$ in our system.

Fig. 3 shows the outage probabilities of our proposed optimal policy for different target sum rates $R$ and transmission power levels of the source nodes $P_S$ when the DF cooperation protocol is exploited. It can be easily seen that the analysis results and simulation results match very well. The outage probability can be improved with the decrease of $R$ or the increase of the transmission power $P_S$. This is because the instantaneous throughput can be increased by enlarging $P_S$. Moreover, it can be observed that there exists the saturated structure, i.e., the outage probability is gradually saturated and finally close to the battery empty probability for the optimal policy (the dashed line without markers) in sufficiently high SNRs, instead of going to zero. This is because the outage probability is equal to the battery empty probability in sufficiently high SNRs according to Theorem 3.

Fig. 4 shows the outage probability of our proposed optimal policy for different target sum rates $R$ and source nodes' transmission power $P_S$ when the AF cooperation protocol is exploited at the relay. It can be seen that there is a minor gap between the analysis results and simulation results when SNR is small, whereas the curves become identical at high SNRs. This is because the approximate conditional outage probability is exploited for the AF cooperation protocol in Proposition 2. In addition, similar performance trends can be observed in the AF mode and the DF mode, e.g., the impacts of $R$ and $P_S$ on the outage probability, the saturated structure, etc.

Fig. 5 illustrates the outage probabilities of our proposed optimal policy for different target rate proportions $\sigma$ when AF or DF cooperation protocol are exploited. It can be observed that the outage performance of DF mode is superior to that of AF mode, except that the performance difference between the two modes is very small in very low SNR regimes. Since the outage probability is equal to the battery empty probabilities in sufficiently high SNRs, which are identical for both AF and DF

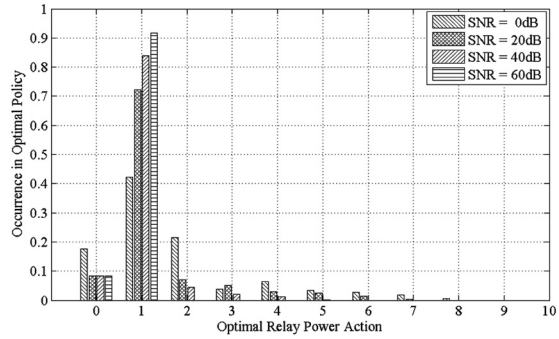


Fig. 6. Comparison of optimal relay power actions $w^*$ among low, moderate and high normalized SNRs with AF mode ($P_S = 11P_u$, $R = 4$bit/s/Hz).

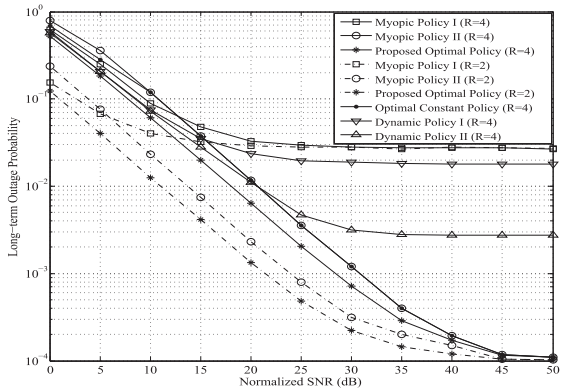


Fig. 7. Outage probabilities of proposed optimal policy and compared policies in AF mode ($P_S = 3P_u$, unit of target sum rate $R$: bit/s/Hz).

modes according to Theorem 3, the outage probabilities of AF and DF modes converge to the same value. Moreover, the outage performance in asymmetric target rate condition ($\sigma = 0.25$) is inferior to that in symmetric condition ($\sigma = 0.5$). The reason is explained as follows. Form (7) and (12), the network is in outage condition if any of the outage events occurs. One of target rates $R_{th1}$ and $R_{th2}$ in asymmetric condition is smaller than that in symmetric condition. Therefore, the occurrence probability of the outage event corresponding to the smaller target rate is higher, and thus the outage performance in asymmetric condition becomes worse than that in symmetric condition.

Fig. 6 demonstrates the occurrence of the optimal relay power actions in optimal policy $\pi^*$ at low, moderate and high normalized SNRs with AF mode. After the optimal policy $\pi^*$ is obtained, the occurrence of the optimal action $w^*$ can be calculated as the proportion of the number of action $w^*$ in all possible system states. It can be observed that the optimal relay actions in high SNRs concentrate on the value of 1, while the actions in low SNRs are much more diverse. This is because the optimal policy is equivalent to a simple "on-off" structure policy in sufficiently high SNRs according to Theorem 3. In low SNR regimes, more energy quanta are consumed by the relay to minimize the long-term outage probability.

Fig. 7 compares the outage probabilities of our proposed optimal policy and several compared policies for different target sum rates $R$ when the AF cooperation protocol is exploited. For the two myopic policies, the relay transmission power is set without concern for the channel state and the battery state transition probabilities. Instead, the relay transmits signals as long as the battery is non-empty. In Myopic Policy I, the largest available energy in the battery is consumed by the relay for one transmission period. Regarding with Myopic Policy II, the relay attempts to exploit the lowest power, i.e., the basic transmission power $P_u$. Moreover, an optimal constant policy and two dynamic policies are defined. In Optimal Constant Policy, the relay tries to utilize an optimal constant power in order to minimize the average outage probability. For Dynamic Policy I, the relay knows the CSI and determines its power equal to the minimum value to minimize the outage probability in its current channel states. Unlike Dynamic Policy I, the relay knows the CSI as well as the status of its battery in Dynamic Policy II. If the relay needs to consume the total energy in its battery to minimize the outage probability, it can always leave one energy quantum in its battery for the next transmission period. It can be seen that the outage probability of our proposed optimal policy is superior to those of the compared policies. The outage probabilities of these five policies are all saturated in sufficiently high SNR regimes, and the saturation outage probabilities correspond to their own battery empty probabilities at sufficiently high SNRs. Since the proposed optimal policy is equivalent to Myopic Policy II in high SNR regimes according to Theorem 2, the saturation outage probabilities of these two policies are identical. Regarding with Myopic Policy I, since the largest available energy in the battery is consumed at once, its battery empty probability is larger than that of Myopic Policy II. In other words, the outage probability performances of Myopic Policy II and our proposed optimal policy outperform that of Myopic Policy I in high SNR regimes. Considering Optimal Constant Policy, its outage performance is superior to that of Myopic Policy II in low SNR regimes, while these two policies are equivalent in moderate and high SNR regimes. In other words, the constant transmission power in Optimal Constant Policy is equal to one basic transmission power $P_u$ in moderate and high SNR regimes. In the two dynamic policies, the outage performances are just inferior to that of the optimal policy in low and moderate SNR regimes, while their performances do not converge to that of the optimal policy in high SNR regimes. This is because the relay in the two dynamic Policies determines its transmission power to minimize the outage probability based on its current channel states, not considering the system state transition probabilities. As a result, the battery empty probabilities for the two dynamic Policies are higher than that of the optimal policy in high SNR regimes. Since the relay in Dynamic Policy II knows its battery status and can at least leave one energy quantum in the battery for the next transmission, the battery empty probability is decreased and Dynamic Policy II outperforms Dynamic Policy I largely. In addition, Fig. 8 compares the outage probabilities of our proposed optimal policy and several compared policies when the DF cooperation protocol is used. As compared with the AF mode, similar performance trends can be found in this figure. Since the performance trends among these policies keep the same for different $R$, the performances of Optimal Constant Policy and Dynamic Policy are shown in the AF mode with $R = 4$bit/s/Hz and in the DF mode with $R = 2$bit/s/Hz.

Fig. 9 illustrates the outage probabilities of our proposed optimal policy for different sizes of the solar panel area $\Omega$ and energy quantum $E_u$ when the DF or the AF protocols are

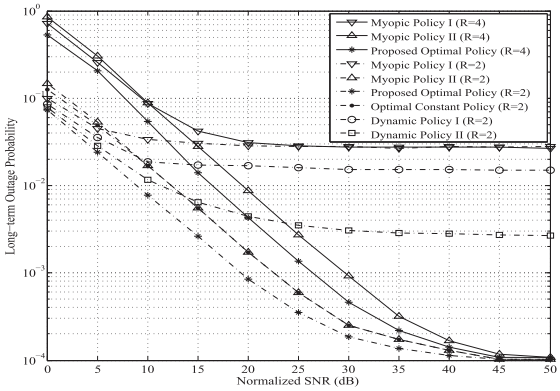


Fig. 8. Outage probabilities of proposed optimal policy and compared policies in DF mode ($P_S = 3P_u$, unit of target sum rate $R$: bit/s/Hz).

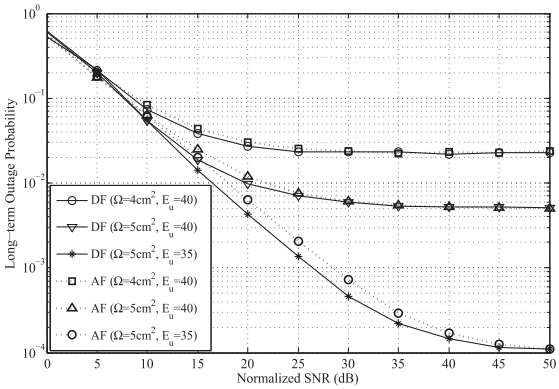


Fig. 9. Impact of solar panel area Ω and energy quantum $E_u$ (unit: 150mJ) on outage probabilities with DF or AF modes ($P_S = 3P_u$, $R = 4$bit/s/Hz).

exploited. It can be seen that the saturation outage probability in high SNR regimes, i.e., the battery empty probability, becomes smaller when the solar panel size Ω gets larger or one energy quantum $E_u$ gets smaller. The reason can be explained as follows. Since there is more energy harvested within one policy management period when the solar panel size Ω is bigger, the energy deficiency probability $P\,(Q = 0|Q_e = e)$ and the battery empty probability $P_\pi\,(b = 0)$ can be decreased by increasing Ω. Furthermore, with a smaller energy quantum $E_u$, there are more numbers of energy quanta which can be stored in the battery. Since the optimal policies for the DF and the AF protocols are identical in sufficiently high SNR regimes, the same phenomena are exhibited for the both protocols at high SNRs.

Fig. 10 shows the outage probability of the proposed optimal policy versus the number of battery states $N_b$ in different SNRs with the AF and DF modes. It can be seen that the outage performance can be dramatically improved by enlarging the battery capacity to store more energy quanta, especially in the high SNRs. When the battery capacity becomes larger, the slope of the curves becomes flatter, and finally the outage performance becomes stable no matter how large the battery capacity is. Similar performance trends can be observed in both AF and DF modes. This property can help to find the optimal battery capacity in maximizing the cost performance.

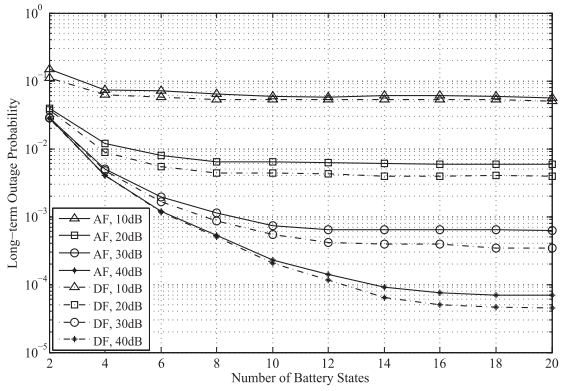


Fig. 10. Outage probability versus number of battery states $N_b$ in different normalized SNRs with AF and DF modes ($P_S = 3P_u$, $R = 4$bit/s/Hz).

## VII. Conclusion

In this paper, the optimal and adaptive relay transmission policy for minimizing the long-term average outage probability in the EH TWR network was proposed. Unlike the previous works, we made use of stochastic solar EH models to formulate the solar irradiance condition and designed an MDP framework to optimize the relay transmission policy in accordance with the solar ESI, CSI and finite battery condition. We first found the monotonic and bounded differential structure of the expected total discounted reward. Furthermore, we studied the property of the optimal solutions, and the ceiling and threshold structures of the optimal relay power action were discovered. Moreover, the expected outage probability was theoretically analyzed and an interesting saturated structure was found to predict the performance limit of the outage probability at sufficiently high SNRs. The theoretical results were substantiated through extensive computer simulations.

## Appendix A
## Proof of Proposition 1

When the relay exploits the DF cooperation protocol, the outage events in (4), (5) and (6) can be rewritten as

$$\begin{aligned}\mathcal{E}^1_{out,DF} &= \{(\gamma_1 < \tilde{\gamma}_{th1}) \cup (\gamma_2 < \tilde{\gamma}_{th2})\},\\ \mathcal{E}^2_{out,DF} &= \{(\gamma_2 < \tilde{\gamma}_{th3}) \cup (\gamma_1 < \tilde{\gamma}_{th4})\},\\ \mathcal{E}^3_{out,DF} &= \{(\gamma_1 + \gamma_2) < c\},\end{aligned} \tag{38}$$

where $\tilde{\gamma}_{th1} = \frac{N_0}{P_S}\left(2^{2R_{th1}} - 1\right)$, $\tilde{\gamma}_{th2} = \frac{N_0}{P_R}\left(2^{2R_{th1}} - 1\right)$, $\tilde{\gamma}_{th3} = \frac{N_0}{P_S}\left(2^{2R_{th2}} - 1\right)$, $\tilde{\gamma}_{th4} = \frac{N_0}{P_R}\left(2^{2R_{th2}} - 1\right)$ and $c = \frac{N_0}{P_S}\left(2^{2(R_{th1}+R_{th2})} - 1\right)$. Substituting the above three events into (18) yields

$$\begin{aligned}P_{out,DF}\,(w,h,g) = \Pr\{&(\gamma_1 < \gamma_{th1}) \cup (\gamma_2 < \gamma_{th2})\\ &\cup (\gamma_1 + \gamma_2 < c) | P_R = wP_u,\, H_{ar} = h,\, H_{br} = g\},\end{aligned} \tag{39}$$

where $\gamma_{th1} = \max\{\tilde{\gamma}_{th1}, \tilde{\gamma}_{th4}\}$ and $\gamma_{th2} = \max\{\tilde{\gamma}_{th2}, \tilde{\gamma}_{th3}\}$. By applying the following equation

$$\begin{aligned}\Pr\{A \cup B \cup C\} &= \Pr\{A \cup B\} + \Pr\left\{\overline{(A \cup B)} \cap C\right\}\\ &= 1 - \Pr\left\{\overline{A} \cap \overline{B}\right\} + \Pr\left\{\overline{A} \cap \overline{B} \cap C\right\},\end{aligned} \tag{40}$$

where $A$, $B$ and $C$ are random events, the conditional outage probability in (39) is expressed as (41), shown at the bottom of the page.

The conditional outage probability can be computed by discussing the relationship between the channel power thresholds and the channel quantization thresholds in the following cases:

- Case 1: $\gamma_{th1} \geq \Gamma_{h+1}$ or $\gamma_{th2} \geq \Gamma_{g+1}$;
- Case 2: $\gamma_{th1} < \Gamma_{h+1}$ and $\gamma_{th2} < \Gamma_{g+1}$.

For Case 1, it is straightforward to derive $P_{out,DF}(w,h,g) = 1$. For Case 2, by letting $a = \max\{\gamma_{th1}, \Gamma_h\}$ and $b = \max\left\{\gamma_{th2}, \Gamma_g\right\}$ and from (41), the conditional outage probability can be explicitly calculated as (42), shown at the bottom of the page. Subsequently, $T$ is computed by discussing the relationship among $a$, $b$, $c$ and the channel quantization thresholds. As shown in Fig. 11, $(a \leq \gamma_1 < \Gamma_{h+1}) \cap \left(b \leq \gamma_2 < \Gamma_{g+1}\right)$ and $((\gamma_1 + \gamma_2) = c)$ are represented as a rectangular zone and a strait line respectively, and $T$ is denoted as the intersection area between the rectangular zone and the lower zone of the line, which can be divided into six subcases:

- Subcase 2-1 ($c \geq \Gamma_{h+1} + \Gamma_{g+1}$): This condition means the intersection area is the whole rectangular zone, and thus $T$ can be computed as

$$\begin{aligned} T &= \Pr\left\{(a \leq \gamma_1 < \Gamma_{h+1}) \cap \left(b \leq \gamma_2 < \Gamma_{g+1}\right)\right\} \\ &= \Pr\{a \leq \gamma_1 < \Gamma_{h+1}\} \cdot \Pr\left\{b \leq \gamma_2 < \Gamma_{g+1}\right\}. \end{aligned} \quad (43)$$

  By substituting (43) into (42), the conditional outage probability is equal to 1.
- Subcase 2-2 ($c \leq a + b$): This condition means there is no intersection area, and therefore $T = 0$;
- Subcase 2-3 ($a + b < c \leq \min\{(a + \Gamma_{g+1}), (b + \Gamma_{h+1})\}$): In this condition, the intersection area is a triangle shown as the shadow area in Fig. 11(a), thus $T$ is calculated as

$$\begin{aligned} T &= \int_a^{c-b} f(\gamma_1)\, d\gamma_1 \int_b^{-\gamma_1+c} f(\gamma_2)\, d\gamma_2 \\ &= e^{-(a+b)/\theta} - e^{-c/\theta} - \frac{1}{\theta} e^{-c/\theta}(c - b - a); \end{aligned} \quad (44)$$

- Subcase 2-4 ($(b + \Gamma_{h+1}) < c < \left(a + \Gamma_{g+1}\right)$): In this condition, the intersection area is a trapezoid shown as the shadow area in Fig. 11(b), thus $T$ is calculated as

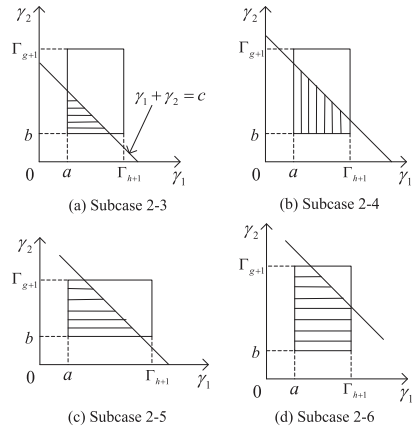


Fig. 11. The relationship among $a$, $b$, $c$ and channel quantization thresholds when calculating $P_{out,DF}(w,h,g)$ in Case 2.

$$\begin{aligned} T &= \int_a^{\Gamma_{h+1}} f(\gamma_1)\, d\gamma_1 \int_b^{-\gamma_1+c} f(\gamma_2)\, d\gamma_2 \\ &= e^{-(a+b)/\theta} e^{-(\Gamma_{h+1}+b)/\theta} - \frac{1}{\theta} e^{-c/\theta}(\Gamma_{h+1} - a) \end{aligned} \quad (45)$$

- Subcase 2-5 ($\left(a + \Gamma_{g+1}\right) < c < (b + \Gamma_{h+1})$): In this condition, the intersection area is a trapezoid shown as the shadow area in Fig. 11(c), thus $T$ is calculated as

$$\begin{aligned} T &= \int_b^{\Gamma_{g+1}} f(\gamma_2)\, d\gamma_2 \int_a^{c-\gamma_2} f(\gamma_1)\, d\gamma_1 \\ &= e^{-(a+b)/\theta} - e^{-(a+\Gamma_{g+1})/\theta} - \frac{1}{\theta} e^{-c/\theta}\left(\Gamma_{g+1} - b\right); \end{aligned} \quad (46)$$

- Subcase 2-6 ($\max\left\{\left(a + \Gamma_{g+1}\right), (b + \Gamma_{h+1})\right\} \leq c < \left(\Gamma_{h+1} + \Gamma_{g+1}\right)$): In this condition, the intersection area is a pentagon shown as the shadow area in Fig. 11(d), thus $T$ is calculated as

$$\begin{aligned} T &= \int_a^{\Gamma_{h+1}} f(\gamma_1)\, d\gamma_1 \int_b^{\Gamma_{g+1}} f(\gamma_2)\, d\gamma_2 \\ &\quad - \int_{c-\Gamma_{g+1}}^{\Gamma_{h+1}} f(\gamma_1)\, d\gamma_1 \int_{-\gamma_1+c}^{\Gamma_{g+1}} f(\gamma_2)\, d\gamma_2 \end{aligned}$$

$$P_{out}(w,h,g) = 1 - \Pr\{\gamma_1 \geq \gamma_{th1} | H_{ar} = h\} \cdot \Pr\{\gamma_2 \geq \gamma_{th2} | H_{br} = g\} + \Pr\{(\gamma_1 \geq \gamma_{th1}) \cap (\gamma_2 \geq \gamma_{th2}) \cap (\gamma_1 + \gamma_2 < c) | H_{ar} = h, H_{br} = g\} \quad (41)$$

$$\begin{aligned} &P_{out,DF}(w,h,g) \\ &= 1 - \frac{\Pr\{(\gamma_1 \geq \gamma_{th1}) \cap (\Gamma_h \leq \gamma_1 < \Gamma_{h+1})\}}{\Pr\{\Gamma_h \leq \gamma_1 < \Gamma_{h+1}\}} \cdot \frac{\Pr\left\{(\gamma_2 \geq \gamma_{th2}) \cap \left(\Gamma_g \leq \gamma_2 < \Gamma_{g+1}\right)\right\}}{\Pr\left\{\Gamma_g \leq \gamma_2 < \Gamma_{g+1}\right\}} \\ &\quad + \frac{\Pr\left\{(\gamma_1 \geq \gamma_{th1}) \cap (\gamma_2 \geq \gamma_{th2}) \cap ((\gamma_1 + \gamma_2) < c) \cap (\Gamma_h \leq \gamma_1 < \Gamma_{h+1}) \cap \left(\Gamma_g \leq \gamma_2 < \Gamma_{g+1}\right)\right\}}{\Pr\{\Gamma_h \leq \gamma_1 < \Gamma_{h+1}\} \cdot \Pr\left\{\Gamma_g \leq \gamma_2 < \Gamma_{g+1}\right\}} \\ &= 1 + \frac{T - \Pr\{a \leq \gamma_1 < \Gamma_{h+1}\} \cdot \Pr\left\{b \leq \gamma_2 < \Gamma_{g+1}\right\}}{\Pr\{\Gamma_h \leq \gamma_1 < \Gamma_{h+1}\} \cdot \Pr\left\{\Gamma_g \leq \gamma_2 < \Gamma_{g+1}\right\}} = 1 + \frac{T - \left(e^{-a/\theta} - e^{-\Gamma_{h+1}/\theta}\right) \cdot \left(e^{-b/\theta} - e^{-\Gamma_{g+1}/\theta}\right)}{\left(e^{-\Gamma_h/\theta} - e^{-\Gamma_{h+1}/\theta}\right) \cdot \left(e^{-\Gamma_g/\theta} - e^{-\Gamma_{g+1}/\theta}\right)}, \end{aligned} \quad (42)$$

where $T = \Pr\left\{(a \leq \gamma_1 < \Gamma_{h+1}) \cap \left(b \leq \gamma_2 < \Gamma_{g+1}\right) \cap (\gamma_1 + \gamma_2 < c)\right\}$.

$$= \left(e^{-a/\theta} - e^{-\Gamma_{h+1}/\theta}\right)\left(e^{-b/\theta} - e^{-\Gamma_{g+1}/\theta}\right) - e^{-(\Gamma_{h+1}+\Gamma_{g+1})/\theta} + e^{-c/\theta} + \frac{1}{\theta}e^{-c/\theta}\left(c - \Gamma_{g+1} - \Gamma_{h+1}\right) \quad (47)$$

Thus, we complete the proof of Proposition 1.

## Appendix B
## Proof of Proposition 2

When the relay exploits the AF cooperation protocol, in high SNR regimes, the outage events in (10) and (11) can be written as

$$\mathcal{E}^1_{out,AF} = \frac{x_1 x_2}{x_1 + x_2} < m_1,\ \mathcal{E}^2_{out,AF} = \frac{y_1 y_2}{y_1 + y_2} < m_2, \quad (48)$$

where $x_1 = \gamma_1\eta_1$, $x_2 = \gamma_2(\eta_2 + \eta_r)$, $m_1 = \frac{\eta_2+\eta_r}{\eta_r}\left(2^{2R_{th1}} - 1\right)$, $y_1 = \gamma_1(\eta_1 + \eta_r)$, $y_2 = \gamma_2\eta_2$ and $m_2 = \frac{\eta_1+\eta_r}{\eta_r}\left(2^{2R_{th\_2}} - 1\right)$. Thus, substituting (48) into (19) yields

$$P_{out,AF}(w,h,g) = \Pr\left\{\left(\frac{x_1x_2}{x_1+x_2} < m_1\right) \cup \left(\frac{y_1y_2}{y_1+y_2} < m_2\right) \mid P_R = wP_u,\ H_{ar} = h,\ H_{br} = g\right\}. \quad (49)$$

By considering the well-known harmonic mean inequality $xy/(x+y) \le \min(x, y)$ [9, 7.86], the conditional outage probability can be expressed as

$$\begin{aligned} &P_{out,AF}(w,h,g) \\ &\ge \Pr\{(\min\{x_1,x_2\} < m_1) \cup (\min\{y_1,y_2\} < m_2) \\ &\qquad | P_R = wP_u,\ H_{ar} = h,\ H_{br} = g\} \\ &= 1 - \Pr\{(\gamma_1 \ge \gamma_{th1}) \cap (\gamma_1 \ge \gamma_{th2}) | \Gamma_h \le \gamma_1 < \Gamma_{h+1}\} \\ &\quad \times \Pr\left\{(\gamma_2 \ge \gamma_{th3}) \cap (\gamma_2 \ge \gamma_{th4}) | \Gamma_g \le \gamma_2 < \Gamma_{g+1}\right\}, \end{aligned} \quad (50)$$

where $\gamma_{th1} = \frac{m_1}{\eta_1} = \frac{(P_S+wP_u)N_0}{P_S\cdot wP_u}\left(2^{2R_{th1}} - 1\right)$, $\gamma_{th2} = \frac{m_2}{\eta_1+\eta_r} = \frac{N_0}{wP_u}\left(2^{2R_{th2}} - 1\right)$, $\gamma_{th3} = \frac{m_1}{\eta_2+\eta_r} = \frac{N_0}{wP_u}\left(2^{2R_{th1}} - 1\right)$ and $\gamma_{th4} = \frac{m_2}{\eta_2} = \frac{(P_S+wP_u)N_0}{P_S\cdot wP_u}\left(2^{2R_{th2}} - 1\right)$.

The conditional outage probability can be computed by discussing the relationship between these four thresholds and the channel quantization thresholds in the following three cases:

- Case 1 ($\gamma_{th1} \ge \Gamma_{h+1}$ or $\gamma_{th2} \ge \Gamma_{h+1}$ or $\gamma_{th3} \ge \Gamma_{g+1}$ or $\gamma_{th4} \ge \Gamma_{g+1}$): $P_{out,AF}(w,h,g) = 1$.
- Case 2 ($\gamma_{th1} \le \Gamma_h$ and $\gamma_{th2} \le \Gamma_h$ and $\gamma_{th3} \le \Gamma_g$ and $\gamma_{th4} \le \Gamma_g$): It can be easily obtained that

  $$\begin{aligned} &\Pr\{(\gamma_1 \ge \gamma_{th1}) \cap (\gamma_1 \ge \gamma_{th2}) | \Gamma_h \le \gamma_1 < \Gamma_{h+1}\} \\ &= \Pr\left\{(\gamma_2 \ge \gamma_{th3}) \cap (\gamma_2 \ge \gamma_{th4}) | \Gamma_g \le \gamma_2 < \Gamma_{g+1}\right\} = 1. \end{aligned}$$

  Therefore, $P_{out,AF}(w,h,g) = 0$.
- Case 3 (Otherwise): It can also be easily obtained that

  $$\begin{aligned} &\Pr\{(\gamma_1 \ge \gamma_{th1}) \cap (\gamma_1 \ge \gamma_{th2}) | \Gamma_h \le \gamma_1 < \Gamma_{h+1}\} \\ &= \frac{\Pr\{\max(\gamma_{th1}, \gamma_{th2}) \le \gamma_1 < \Gamma_{h+1}\}}{\Pr\{\Gamma_h \le \gamma_1 < \Gamma_{h+1}\}} \end{aligned} \quad (51)$$

  $$\begin{aligned} &\Pr\left\{(\gamma_2 \ge \gamma_{th3}) \cap (\gamma_2 \ge \gamma_{th4}) | \Gamma_g \le \gamma_2 < \Gamma_{g+1}\right\} \\ &= \frac{\Pr\left\{\max(\gamma_{th3}, \gamma_{th4}) \le \gamma_2 < \Gamma_{g+1}\right\}}{\Pr\left\{\Gamma_g \le \gamma_2 < \Gamma_{g+1}\right\}}. \end{aligned} \quad (52)$$

Substituting (51) and (52) into (50) yields

$$P_{out,AF}(w,h,g) \ge 1 - \frac{e^{-\max(\gamma_{th1},\gamma_{th2})/\theta} - e^{-\Gamma_{h+1}/\theta}}{e^{-\Gamma_h/\theta} - e^{-\Gamma_{h+1}/\theta}} \times \frac{e^{-\max(\gamma_{th3},\gamma_{th4})/\theta} - e^{-\Gamma_{g+1}/\theta}}{e^{-\Gamma_g/\theta} - e^{-\Gamma_{g+1}/\theta}}. \quad (53)$$

Thus, we complete the proof of Proposition 2.

## Appendix C
## Proof of Lemma 1

We prove the lemma by using the induction as follows.

**Step 1:** Assuming the initial condition $V^{(0)}(s) = 0$, the long-term value of the first iteration in (26) can be written as

$$\begin{aligned} V^{(1)}_w(s) &= R_w(s) + \lambda \sum_{s' \in S} P_w(s'|s) V^{(0)}(s') \\ &= R_w(s) = P_{out}(h,g,w). \end{aligned} \quad (54)$$

When $w \in \{0, 1, \cdots, b-1\}$, it can be derived directly from (54) that

$$V^{(1)}_w(e, b-1, h, g) = V^{(1)}_w(e, b, h, g). \quad (55)$$

Meanwhile, since the outage probability is non-increasing with respect to the relay transmission power and its value is from 0 to 1, i.e., $1 \ge P_{out}(h,g,w=b-1) - P_{out}(h,g,w=b) \ge 0$, the following inequality holds

$$1 \ge V^{(1)}_{w=b-1}(e, b-1, h, g) - V^{(1)}_{w=b}(e, b, h, g) \ge 0. \quad (56)$$

By considering (55), (56) and (27), it can be deduced that

$$1 \ge V^{(1)}(e, b-1, h, g) - V^{(1)}(e, b, h, g) \ge 0, \forall b \in \Omega_b \backslash \{0\}. \quad (57)$$

**Step 2:** Assuming $1 \ge V_{(k)}(e, b-1, h, g) - V_{(k)}(e, b, h, g) \ge 0, \forall b \in \Omega_b \backslash \{0\}$. According to (28), when $w \in \{0, 1, \cdots, b-1\}$, the value difference between the expected total discounted rewards of two adjacent battery states in iteration $k+1$ can be written as

$$\begin{aligned} &V^{(k+1)}_w(e, b-1, h, g) - V^{(k+1)}_w(e, b, h, g) \\ &= \lambda \cdot \mathbb{E}_s\left\{V^{(k)}\left(e', \min(b-1-w+q, N_b-1), h', g'\right) - V^{(k)}\left(e', \min(b-w+q, N_b-1), h', g'\right)\right\}. \end{aligned} \quad (58)$$

With the assumption, it can be easily seen that

$$1 \ge V^{(k+1)}_w(e, b-1, h, g) - V^{(k+1)}_w(e, b, h, g) \ge 0, \quad \forall w \in \{0, 1, \cdots, b-1\}. \quad (59)$$

Meanwhile, in iteration $k+1$, the value difference between the expected total discounted rewards of two adjacent battery states with respect to total battery energy consumption can be expressed as

$$\begin{aligned} &V^{(k+1)}_{w=b-1}(e, b-1, h, g) - V^{(k+1)}_{w=b}(e, b, h, g) \\ &= P_{out}(h,g,w=b-1) - P_{out}(h,g,w=b) \end{aligned}$$

$$+\lambda\cdot\mathbb{E}_s\left\{V^{(k)}\left(e',\min\left(q,N_b-1\right),h',g'\right)\right.$$
$$\left.-V^{(k)}\left(e',\min\left(q,N_b-1\right),h',g'\right)\right\}$$
$$=P_{out}\left(h,g,w=b-1\right)-P_{out}\left(h,g,w=b\right). \quad (60)$$

Similarly to (56) in Step 1, the following inequality also holds

$$1\geq V_{w=b-1}^{(k+1)}\left(e,b-1,h,g\right)-V_{w=b}^{(k+1)}\left(e,b,h,g\right)\geq 0. \quad (61)$$

According to (59), (61) and (27), for $\forall b\in\mathcal{Q}_b\backslash\{0\}$, it can be obtained that

$$1\geq V^{(k+1)}\left(e,b-1,h,g\right)-V^{(k+1)}\left(e,b,h,g\right)\geq 0, \quad (62)$$

and the proof is as follows:

According to (59) and (61), for any element $\alpha\in\left\{V_w^{(k+1)}\left(e,b-1,h,g\right)\right\}_{w=0}^{b-1}$, there always exists an element $\beta\in\left\{V_w^{(k+1)}\left(e,h,g,b\right)\right\}_{w=0}^{b}$ to satisfy the condition $1\geq\alpha-\beta\geq 0$. Let $\alpha_{min}=\min\left\{V_w^{(k+1)}\left(e,h,g,b-1\right)\right\}_{w=0}^{b-1}$ and $\beta_{min}=\min\left\{V_w^{(k+1)}\left(e,h,g,b\right)\right\}_{w=0}^{b}$. By applying contradiction method, we assume $\alpha_{min}<\beta_{min}$. Since there must exist an element $\beta'\in\left\{V_w^{(k+1)}\left(e,h,g,b\right)\right\}_{w=0}^{b}$ to satisfy the condition $\alpha_{min}-\beta'\geq 0$, it can be easily derived that $\beta'<\beta_{min}$, which is contradicted with the definition of $\beta_{min}$. Thus, the assumption $\alpha_{min}<\beta_{min}$ does not hold, and we obtain $\alpha_{min}-\beta_{min}\geq 0$. Similarly, it can be easily proved that $1\geq\alpha_{min}-\beta_{min}$. Therefore, according to (27) we obtain (62).

**Step 3:** Combining the results of Step1 and Step2, for $\forall b\in\mathcal{Q}_b\backslash\{0\}$, we use the induction method and prove

$$1\geq V^{(i)}\left(e,b-1,h,g\right)-V^{(i)}\left(e,b,h,g\right)\geq 0,\forall i. \quad (63)$$

When the value iteration algorithm is applied and converged, it can be easily seen that the expected total discounted reward obtained by the optimal policy is also satisfied with the above monotonic and bounded differential structure, i.e., $1\geq V_{\pi^*}\left(e,b-1,h,g\right)-V_{\pi^*}\left(e,b,h,g\right)\geq 0,\forall b\in\mathcal{Q}_b\backslash\{0\}$.

## Appendix D
## Proof of Theorem 1

According to (28) and Definition 1, for any iteration $i$ and any fixed system state $s=(e,b,h,g)\in\mathcal{S}$, the difference value of the two expected total discounted rewards with respect to the relay transmission power actions $w$ $(\tilde{w}<w\leq b)$ and $\tilde{w}$ can be computed as

$$V_{w,f}^{(i+1)}(s)-V_{\tilde{w},f}^{(i+1)}(s)$$
$$=\lambda\cdot\mathbb{E}_s\left\{V^{(i)}\left(e',\min\left(b-w+q,N_b-1\right),h',g'\right)\right.$$
$$\left.-V^{(i)}\left(e',\min\left(b-\tilde{w}+q,N_b-1\right),h',g'\right)\right\}. \quad (64)$$

By applying Lemma 1, it can be easily seen that $V_{w,f}^{(i+1)}(s)\geq V_{\tilde{w},f}^{(i+1)}(s)$. From the value iteration algorithm in (27), it is then concluded that the optimal relay power action in iteration $i+1$ is smaller than or equal to $\min(\tilde{w},b)$. When the algorithm is converged, the optimal relay power action must satisfy $w^*\leq\min(\tilde{w},b)$.

## Appendix E
## Proof of Corollary 2

From (13), the battery state in the $t^{th}$ period $(t\geq 1)$ can be described as $b_t=b_{t-1}-w_t^*+q_t$. From Theorem 3, the battery empty probability $P_\pi(b=0)$ must be equal to zero if the expected outage probability is saturation-free, and this implies that the battery must be always non-empty: $b_t=b_{t-1}-w_t^*+q_t\geq 1,\forall t$. According to Theorem 2, since the optimal action $w_t^*$ is always equal to one in sufficiently high SNRs, the above condition can be equivalently rewritten as $q_t\geq 2-b_{t-1},\forall t$. Because the battery must be non-empty, i.e., $b_{t-1}\geq 1$, it implies that $2-b_{t-1}\leq 1,\forall t$. Thus, only if $q_t\geq 1(\forall t)$, the inequality $q_t\geq 2-b_{t-1}(\forall t)$ can always hold. This condition immediately concludes that the outage probability is saturation-free only if $q_t\geq 1,\forall t$, i.e., the energy deficiency probability is equal to zero.

On the other hand, if the energy deficiency probability is equal to zero, it means that the relay can harvest at least one energy quantum in every policy management period and the battery empty probability is equal to zero. By applying Theorem 3, the expected outage probability approaches to zero in sufficiently high SNRs. From the aforementioned discussions, the corollary is proved.

## References


[1] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 3, pp. 443–461, Third Quart. 2011.

[2] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Trans. Embedded Comput. Syst.*, vol. 6, no. 4, pp. 32–38, Sep. 2007.

[3] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, Sep. 2011.

[4] C. Huang, R. Zhang, and S. Cui, "Optimal power allocation for outage probability minimization in fading channels with energy harvesting constraints," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 1074–1087, Feb. 2014.

[5] S. Wei, W. Guan, and K. J. R. Liu, "Power scheduling for energy harvesting wireless communications with battery capacity constraint," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4640–4653, Aug. 2015.

[6] M.-L. Ku, Y. Chen, and K. J. R. Liu, "Data-driven stochastic models and policies for energy harvesting sensor communications," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 8, pp. 1505–1520, Aug. 2015.

[7] Z. Wang, V. Aggarwal, and X. Wang, "Power allocation for energy harvesting transmitter with causal information," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4080–4093, Nov. 2014.

[8] M. Moradian and F. Ashtiani, "Sum throughput maximization in a slotted Aloha network with energy harvesting nodes," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Istanbul, Turkey, Apr. 2014, pp. 1585–1590.

[9] K. J. R. Liu, A. K. Sadek, W. Su, and A. Kwasinski, *Cooperative Communications and Networking*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[10] Y. Luo, J. Zhang, and K. B. Letaief, "Optimal scheduling and power allocation for two-hop energy harvesting communication systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4729–4741, Sep. 2013.

[11] I. Ahmed, A. Ikhlef, R. Schober, and R. Mallik, "Power allocation for conventional and buffer-aided link adaptive relaying systems with energy harvesting nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1182–1195, Mar. 2014.

[12] C. Huang, R. Zhang, and S. Cui, "Throughput maximization for the Gaussian relay channel with energy harvesting constraints," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 8, pp. 1469–1479, Aug. 2013.

[13] A. Nasir, X. Zhou, S. Durrani, and R. Kennedy, "Relaying protocols for wireless energy harvesting and information processing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3622–3636, Jul. 2013.

[14] Z. Ding, S. Perlaza, I. Esnaola, and H. Poor, "Power allocation strategies in energy harvesting wireless cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 846–860, Feb. 2014.
[15] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.
[16] S. J. Kim, N. Devroye, P. Mitran, and V. Tarokh, "Achievable rate regions and performance comparison of half duplex bi-directional relaying protocols," *IEEE Trans. Wireless Commun.*, vol. 57, no. 10, pp. 6405–6418, Oct. 2011.
[17] B. Varan and A. Yener, "The energy harvesting two-way decode-and-forward relay channel with stochastic data arrival," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Austin, TX, USA, Dec. 2013, pp. 371–374.
[18] K. Tutuncuoglu, B. Varan, and A. Yener, "Throughput maximization for two-way relay channels with energy harvesting nodes: The impact of relaying strategies," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2081–2093, Jun. 2015.
[19] I. Ahmed, A. Ikhlef, D. W. K. Ng, and R. Schober, "Optimal resource allocation for energy harvesting two-way relay systems with channel uncertainty," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Austin, TX, USA, Dec. 2013, pp. 345–348.
[20] W. Li, M.-L. Ku, Y. Chen, and K. J. R. Liu, "On the achievable sum rate for two-way relay networks with stochastic energy harvesting," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Atlanta, GA, USA, Dec. 2014, pp. 288–292.
[21] Q. Li, Q. Zhang, and J. Qin, "Beamforming in non-regenerative two-way multi-antenna relay networks for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5509–5520, Oct. 2014.
[22] Z. Wen, S. Wang, C. Fan, and W. Xiang, "Joint transceiver and power splitter design over two-way relaying channel with lattice codes and energy harvesting," *IEEE Commun. Lett.*, vol. 18, no. 11, pp. 2039–2042, Nov. 2014.
[23] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, Jul. 1999.
[24] H. S. Wang and N. Moayeri, "Finite-state Markov channel-a useful model for radio communication channels," *IEEE Trans. Wireless Commun.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.
[25] H. S. Wang and P.-C. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 353–357, May 1996.
[26] P. Ren, Y. Wang, and Q. Du, "CAD-MAC: A channel-aggregation diversity based MAC protocol for spectrum and energy efficient cognitive ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 237–250, Feb. 2014.
[27] Q. Li, S. H. Ting, A. Pandharipande, and Y. Han, "Adaptive two-way relaying and outage analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 3288–3299, Jun. 2009.
[28] X. Lin, M. Tao, Y. Xu, and R. Wang, "Outage probability and finite-SNR diversity–multiplexing tradeoff for two-way relay fading channels," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3123–3136, Sep. 2013.
[29] N. R. E. Laboratory. (2012). *Solar Radiation Resource Information* [Online] Available: http://www.nrel.gov/rredc/
[30] N. Michelusi, L. Badia, and M. Zorzi, "Optimal transmission policies for energy harvesting devices with limited state-of-charge knowledge," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3969–3982, Nov. 2014.
[31] M. Puterman, Markov Decision Process-Discrete Stochastic Dynamic Programming. Hoboken, NJ, USA: Wiley, 1994.

**Wei Li** received the B.S. and M.S. degrees in electrical and electronics engineering from Xi'an Jiaotong University, Xi'an, China, in 2001 and 2004, respectively. He is currently pursuing the Ph.D. degree at the Department of Information and Communication Engineering, Xi'an Jiaotong University. From 2005 to 2011, he was a Senior Engineer with Huawei Technology Corporation. From 2013 to 2015, he was a Visiting Student at the University of Maryland, College Park, MD, USA. His research interests include green communications, energy harvesting, and cooperative communications in wireless networks.

**Meng-Lin Ku** (M'11) received the B.S., M.S., and Ph.D. degrees from National Chiao Tung University, Hsinchu, Taiwan, in 2002, 2003, and 2009, respectively, all in communication engineering. Between 2009 and 2010, he was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, National Chiao Tung University and with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. In August 2010, he became a Faculty Member of the Department of Communication Engineering, National Central University, Jung-li, Taiwan, where he is currently an Associate Professor. During the summer of 2013, he was a Visiting Scholar in the Signals and Information Group at the University of Maryland, College Park, MD, USA. His research interests include green communications, cognitive radios, and optimization of radio access. He was the recipient of the Best Counseling Award in 2012 and the Best Teaching Award in 2013, 2014, and 2015 at National Central University. He was also the recipient of the Exploration Research Award of the Pan Wen Yuan Foundation, Taiwan, in 2013.

**Yan Chen** (SM'14) received the bachelor's degree from the University of Science and Technology of China, Hefei, China, in 2004, the M.Phil. degree from Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2007, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2011. Being a founding member, he joined Origin Wireless Inc. as a Principal Technologist in 2013. He is currently a Professor with the University of Electronic Science and Technology of China. His research interests include multimedia, signal processing, game theory, and wireless communications.

He was the recipient of multiple honors and awards including Best Student Paper Award at the IEEE ICASSP in 2016, Best Paper Award at the IEEE GLOBECOM in 2013, Future Faculty Fellowship and Distinguished Dissertation Fellowship Honorable Mention from the Department of Electrical and Computer Engineering in 2010 and 2011, Finalist of the Dean's Doctoral Research Award from the A. James Clark School of Engineering, the University of Maryland in 2011, and the Chinese Government Award for outstanding students abroad in 2010.

**K. J. Ray Liu** (F'03) was named a Distinguished Scholar-Teacher of University of Maryland, College Park, MD, USA, in 2007, where he is now the Christine Kim Eminent Professor of Information Technology. He leads the Maryland Signals and Information Group conducting research encompassing broad areas of information and communications technology with recent focus on future wireless technologies, network science, and information forensics and security.

He is recognized by Thomson Reuters as a Highly Cited Researcher. He is a Fellow of AAAS. He is a member of IEEE Board of Director. He was the President of IEEE Signal Processing Society, where he has served as the Vice President of Publications and on the Board of Governors. He has also served as the Editor-in-Chief of *IEEE Signal Processing Magazine*.

Dr. Liu was the recipient of the 2016 IEEE Leon K. Kirchmayer Technical Field Award on graduate teaching and mentoring, IEEE Signal Processing Society 2014 Society Award, and IEEE Signal Processing Society 2009 Technical Achievement Award. He also received teaching and research recognitions from University of Maryland including university-level Invention of the Year Award; and college-level Poole and Kent Senior Faculty Teaching Award, Outstanding Faculty Research Award, and Outstanding Faculty Service Award, all from A. James Clark School of Engineering.