

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

Wei Li<sup>\*†</sup>, Meng-Lin Ku<sup>‡</sup>, Yan Chen<sup>\*</sup>, and K. J. Ray Liu<sup>\*</sup>

<sup>\*</sup>Department of Electrical and Computer Engineering, University of Maryland,  
College Park, MD 20742, USA

<sup>†</sup>Department of Information and Communication Engineering, Xi'an Jiaotong  
University, Xi'an, 710049, China

<sup>‡</sup>Department of Communication Engineering, National Central University, Taiwan

Email: wli52140@umd.edu, mlku@ce.ncu.edu.tw, {yan, kjrliu}@umd.edu

## 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 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 long-term 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 probabilities is revealed, i.e., the 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 myopic 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 attentions due to its effectiveness to resolve energy supply problems in wireless networks and to perpetually provide an infinite amount of energy [1]. In EH communication networks, the EH nodes can make use of renewable energy sources, e.g., solar, wind and vibration, 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 [2] to determine the optimal power scheduling for minimizing the short-term throughput in point-to-point fading channels. Unlike [2], the optimal power allocation scheme that aims at minimizing the average outage probability over a finite time horizon was also studied in [3]. The authors in this paper exploited a deterministic EH model, in which the solar energy state information (ESI) is non-causal and the energy arrival information (EAI) is known prior to transmission scheduling, and a stochastic EH model, in which the solar ESI is causal and unknown to the transmitter. Moreover, considering a real data record of solar irradiance, the authors in [4] and [5] investigated a data-driven stochastic EH model whose underlying parameters are directly trained by solar irradiance data [6], and then a data-driven transmission policy was proposed to maximize the long-term net bit rate by using Markov decision process (MDP).

Cooperative communications have found many applications in various wireless scenarios for

the purpose of the link quality improvement [7]. It is worth noting that there has been a growing interest in investigating EH cooperative communications, where relay nodes can harvest energy from environments. The authors in [8] designed an optimal transmission scheme for an EH half-duplex relay in two-hop networks, where there are two energy arrivals at the source and multiple energy arrivals at the relay. Meanwhile, an optimal transmission policy for a two-hop network with a general EH source of multiple energy arrival states was proposed in [9]. 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 [10]. Moreover, in [11] and [12], transmission policies based on wireless energy transfer, i.e., RF-based energy harvesting, were studied in one-way relay networks.

Due to the advantage of having 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 [13]. Recently, the TWR networks with EH nodes have attracted more and more attentions. 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 in the problems of power allocation and scheduling in EH TWR networks. In the literature [14]-[18], power allocation algorithms for maximizing short-term sum rates in EH TWR networks using deterministic EH models were proposed. No data buffer in the relay was assumed in [14], while a data buffer was considered at the relay in [15], which means the relay can cache data and exploit more flexible scheduling policies. Moreover, a generalized iterative directional water-filling algorithm was designed in [16] for various relaying strategies. An optimization framework with the uncertainty of channel state information (CSI) was presented in [17]. Besides, the optimal transmission strategy for wireless energy transfer in TWR networks was studied in [18]. However, 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.

So far, the optimal transmission strategy for EH TWR networks with stochastic EH models has not been studied. In the slow and fast fading channels, the outage probability and the ergodic throughput are commonly used to characterize the fundamental limit of channel capacity, respectively. By far, 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.

Motivated by the aforementioned discussions, in this paper, we propose an optimal relay transmission policy for the EH TWR network using the data-driven stochastic solar EH model in [5]. 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 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 an MDP optimization framework for EH TWR networks, wherein the Gaussian mixture hidden Markov chain in [5] is used as our stochastic EH model, the fading channels between the sources and relay are formulated by a finite-state Markov model [19], the battery capacity is quantized in units of energy quanta, the system action represents the relay transmission power and the utility is the long-term outage probability [20].
- 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 different from the traditional outage probability that regards the fading channel power as continuous values ranging from zero to infinity, whereas it is defined as the outage probability conditioned on preset fading channel states. We derive the exact close-form and tight lower bound of the conditional outage probabilities for the DF and AF protocols, respectively.
- In order to study the optimal transmission policy, we analyze the property of the long-term reward, and uncover the monotonic and bounded differential structures, which reveals that the utility is non-increasing with the amount of the harvested energy in the battery and the decreased value of the utility caused by the increased battery energy 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 outage probability is found in EH TWR networks with the AF or DF protocols. The analysis concludes that the outage probability converges to the battery empty probability in extremely high SNR regimes, instead of going to zero. Moreover, a saturation-free condition that guarantees the battery empty probability and the outage probability are equal to zero in sufficiently high SNRs is provided.

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 optimization of relay transmission policy and the structure of 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 same transmission power  $P$ , while the transmission power of R is given by  $P_r$ . 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 has the perfect knowledge of the channel state information (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 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, R makes use of either amplify-and-forward (AF) or decode-and-forward (DF) cooperation protocols to broadcast the received signals to A and B [7]. 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

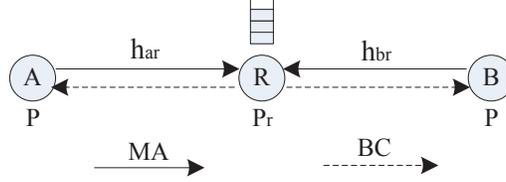


Fig. 1. EH TWR networks.

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 probability for the two cooperation protocols, i.e., decode-and-forward (DF) protocol and amplify-and-forward (AF) protocol.

#### 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. Thus, the achievable rate pair  $(R_1, R_2)$  is given as [13]

$$R_1 \leq \min \left\{ \frac{1}{2} \log \left( 1 + \frac{\gamma_1 P}{N_0} \right), \frac{1}{2} \log \left( 1 + \frac{\gamma_2 P_r}{N_0} \right) \right\}; \quad (1)$$

$$R_2 \leq \min \left\{ \frac{1}{2} \log \left( 1 + \frac{\gamma_2 P}{N_0} \right), \frac{1}{2} \log \left( 1 + \frac{\gamma_1 P_r}{N_0} \right) \right\}; \quad (2)$$

$$R_1 + R_2 \leq \frac{1}{2} \log \left( 1 + \frac{\gamma_1 P}{N_0} + \frac{\gamma_2 P}{N_0} \right), \quad (3)$$

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 [21]:

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

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

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

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

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

### 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 [13], the achievable rate pair  $(R_1, R_2)$  can be expressed as

$$R_1 \leq \frac{1}{2} \log \left( 1 + \frac{\gamma_1 \gamma_2 P P_r}{N_0 (\gamma_1 P + \gamma_2 P + \gamma_2 P_r + N_0)} \right) = \frac{1}{2} \log \left[ 1 + \frac{\gamma_1 \gamma_2 \eta \eta_r}{\gamma_1 \eta + \gamma_2 (\eta + \eta_r) + 1} \right]; \quad (8)$$

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

where we define  $\eta = \frac{P}{N_0}$  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 [21]

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

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

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

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

### 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 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 outage probability. The main components in the MDP model include states, actions and reward functions which represent the system conditions, the relay transmission strategy and the outage probabilities, respectively. The detailed descriptions of all these fundamental elements are introduced as follows.

### A. Relay Actions of Transmission Power

Let  $\mathcal{W} = \{0, 1, \dots, N_p - 1\}$  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 transmission period  $T$ , where  $P_u$  is the basic transmission power corresponding to one energy quantum  $E_u$  during the transmission period, i.e.,  $E_u = P_u T$ . Particularly, if  $w = 0$ , it means that the relay keeps silent during the transmission period.

### B. System States

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

**(a) Solar EH State:** An  $N_e$ -state stochastic EH model in [5] 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 from 2008 to 2010 [6]. Therein, it is assumed that if the solar EH state is given by  $Q_e = i \in \mathcal{Q}_e$ , the harvested solar power per unit area,  $P_h$ , is a Gaussian-distributed random variable with  $\mathcal{N}(\mu_i, \rho_i)$ . Therefore, different solar EH states result in different solar irradiance intensities. Moreover, the dynamic of the states is governed by a state transition probability  $P(Q_e = j | Q_e = i), \forall i, j \in \mathcal{Q}_e$  [5].

The energy collection and storage system at the relay is composed of a solar panel, an energy convertor, a large capacitor and a finite battery, and without loss of generality, the battery is uniformly quantized into several levels in units of  $E_u$ , which is referred to as one energy quantum in our system. Thus, the harvested solar energy during one transmission period  $T$  can be computed as  $E_h = P_h T \Omega \eta$ , where  $\Omega$  is the solar panel area size and  $\eta$  denotes the energy conversion efficiency. During each transmission period, the harvested energy is first stored in the relay's capacitor and then transferred to the battery in units of  $E_u$ . Therefore, at the end of each period, the total energy in the capacitor is the harvested energy  $E_h$  plus the residual energy (which is smaller than  $E_u$ ) in the previous period. Accordingly, the probability of the number

of harvested energy quanta conditioned on the  $i^{\text{th}}$  solar EH state, denoted as  $P(E = q|Q_e = i)$  for  $q \in \{0, 1, \dots, \infty\}$ , is theoretically derived and provided in [5], 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  $bE_u$ . Since the battery state transition is related to both the relay transmission power action and the number of harvested energy quanta, the battery state transition probability at the  $i^{\text{th}}$  solar EH state with respect to the power action  $W = w \in \mathcal{W}$  can be expressed as

$$P_w(Q_b = b'|Q_b = b, Q_e = i) \quad (13)$$

$$= \begin{cases} P(E = b' - b + w|Q_e = i), & b' = (b - w), \dots, N_b - 2; \\ 1 - \sum_{j=0}^{N_b - 2 - b + w} P(E = j|Q_e = i), & b' = N_b - 1, \end{cases}$$

where the maximum affordable power action is restricted to the current battery state  $b \in \mathcal{Q}_b$ , i.e.,  $w \in \{0, 1, \dots, \min(b, N_p - 1)\}$ .

**(c) Channel States:** The instantaneous channel power values,  $\gamma_1$  and  $\gamma_2$ , are quantized into  $N_c$  levels using a finite number of thresholds, given by  $\Gamma = \{0 = \Gamma_0, \Gamma_1, \dots, \Gamma_{N_c} = \infty\}$ , and the  $i^{\text{th}}$  channel interval  $[\Gamma_i, \Gamma_{i+1})$  represents the  $i^{\text{th}}$  channel state, for  $i \in \{0, 1, \dots, N_c - 1\}$ . Besides, the channel variation from one level to another is formulated by a finite-state Markov chain. By assuming that the wireless channels are Rayleigh fading, the stationary probability of the  $i^{\text{th}}$  channel state can be expressed as

$$P(H = i) = P\{\Gamma_i \leq H < \Gamma_{i+1}\} = \int_{\Gamma_i}^{\Gamma_{i+1}} \frac{1}{\theta} \exp\left(-\frac{\gamma}{\theta}\right) d\gamma = \exp\left(-\frac{\Gamma_i}{\theta}\right) - \exp\left(-\frac{\Gamma_{i+1}}{\theta}\right), \quad (14)$$

where  $\theta$  is the average channel power. Assume that the channel varies slowly and can only transit from the current state to its neighboring states, and the channel state transition probability  $P(H = j|H = i)$ , for  $i \in \{0, \dots, N_c - 1\}$ ,  $j \in \{\max(0, i - 1), \dots, \min(i + 1, N_c - 1)\}$ , is defined in [19].

**(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  $s = (e, h, g, b)$

to the state  $s' = (e', h', g', b')$  associated with the relay power action  $w$  can be computed as

$$\begin{aligned} P_w(s' = (e', h', g', b') | s = (e, h, g, b)) & \quad (15) \\ & = P(Q_e = e' | Q_e = e) \cdot P(H_{ar} = h' | H_{ar} = h) \cdot P(H_{br} = g' | H_{br} = g) \cdot P_w(Q_b = b' | Q_b = b, Q_e = e) \end{aligned}$$

### C. Reward Function

Here the conditional outage probability for a relay power action at a fixed system state 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  $s = (e, h, g, b) \in \mathcal{S}$  with respect to the relay action  $w \in \mathcal{W}$  can be simplified as

$$R_{w,f}(s) = \Pr \{ \text{Outage event occurs} | w, f, s \} \triangleq P_{out,f}(w, h, g), \quad (16)$$

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

$$P_{out,DF}(w, h, g) = \Pr \{ \mathcal{E}_{out,DF}^1 \cup \mathcal{E}_{out,DF}^2 \cup \mathcal{E}_{out,DF}^3 | P_r = wP_u, H_{ar} = h, H_{br} = g \}, \quad (17)$$

$$P_{out,AF}(w, h, g) = \Pr \{ \mathcal{E}_{out,AF}^1 \cup \mathcal{E}_{out,AF}^2 | P_r = wP_u, H_{ar} = h, H_{br} = g \}, \quad (18)$$

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

*Theorem 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  $s = (e, h, g, b)$  and relay power action  $w$  can be expressed as follows:

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

where  $\gamma_{th1} = \max\left(\frac{N_0}{P} (2^{2R_{th1}} - 1), \frac{N_0}{wP_u} (2^{2R_{th2}} - 1)\right)$ ,  $\gamma_{th2} = \max\left(\frac{N_0}{wP_u} (2^{2R_{th1}} - 1), \frac{N_0}{P} (2^{2R_{th2}} - 1)\right)$ ,  $a = \max(\gamma_{th1}, \Gamma_h)$ ,  $b = \max(\gamma_{th2}, \Gamma_g)$ ,  $c = \frac{N_0}{P} (2^{2(R_{th1} + R_{th2})} - 1)$  and

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

**Proof:** See Appendix A for details. ■

*Theorem 2:* For the given target rate pair  $(R_{th1}, R_{th2})$ , the conditional outage probability of the TWR network using the AF cooperation protocol with respect to the system state  $s = (e, h, g, b)$  and relay power action  $w$  can be expressed as

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

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

**Proof:** See Appendix B for details. ■

*Remark 1:* From (16), Theorem 1 and Theorem 2, the reward function for a given target rate pair has the following two essential properties:

$$R_{w=0,f}(s) = P_{out,f}(h, g, w = 0) = 1, \forall f \in \{AF, DF\}; \quad (19)$$

$$\lim_{N_0 \rightarrow 0, w \geq 1} R_{w,f}(s) = \lim_{N_0 \rightarrow 0, w \geq 1} P_{out,f}(h, g, w) = 0, \forall f \in \{AF, DF\}. \quad (20)$$

In (19), 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 (20) 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 state.

#### IV. STRUCTURE OF OPTIMAL TRANSMISSION POLICIES

##### A. Optimization of Relay Transmission Policy

The policy  $\pi(s) : \mathcal{S} \rightarrow \mathcal{W}$  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 discounted long-term reward as follows

$$V_{\pi}(s_0) = \mathbb{E}_{\pi} \left[ \sum_{k=0}^{\infty} \lambda^k R_{\pi(s_k)}(s_k) \right], s_k \in \mathcal{S}, \pi(s_k) \in \mathcal{W}, \quad (21)$$

where  $s_0$  is the initial state, and  $0 \leq \lambda < 1$  is a discount factor that guarantees the convergence of the expected long-term reward. It is known that the optimal value of the expected long-term reward is unrelated with the initial state, if the states of the Markov chain are assumed to be recurrent [20]. Moreover, the optimal policy for minimizing (21) can be found through the Bellman equation, given by

$$V_{\pi^*}(s) = \min_{w \in \mathcal{W}} \left( R_w(s) + \lambda \sum_{s' \in \mathcal{S}} P_w(s'|s) V_{\pi^*}(s') \right), \quad (22)$$

which can be efficiently implemented by executing the well-known value iterations [22]:

$$V_w^{(i+1)}(s) = R_w(s) + \lambda \sum_{s' \in \mathcal{S}} P_w(s'|s) V^{(i)}(s'), \quad s \in \mathcal{S}, \quad w \in \mathcal{W}; \quad (23)$$

$$V^{(i+1)}(s) = \min_{w \in \mathcal{W}} (V_w^{(i+1)}(s)), \quad s \in \mathcal{S}, \quad (24)$$

where  $i$  is the iteration number, and the initial value  $V_0(s)$  is set as zero for all states. The value iteration algorithm alternates until a stopping criterion,  $|V^{(i+1)} - V^{(i)}| \leq \varepsilon$ , is satisfied. In practical applications, the solar EH state can be estimated and updated using the real data of solar irradiance and Bayes's rule at the relay, and the current channel and battery states can be easily acquired by the relay [5]. With the optimal transmission policy, the relay can make full use of the state information to decide its optimal transmission power by exploiting the conventional look-up table method during each transmission period.

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 (23), (13) and (15), the

long-term reward function for any fixed system state  $s = (e, h, g, b) \in \mathcal{S}$  and relay action  $w \in \mathcal{W}$  can be rewritten as

$$\begin{aligned}
V_w^{(i+1)}(s) &= R_w(s) + \lambda \sum_{s' \in \mathcal{S}} P(s'|s) V^{(i)}(s') \\
&= P_{out}(h, g, w) + \lambda \sum_{e'=0}^{N_e-1} P(Q_e = e' | Q_e = e) \cdot \sum_{h'=\max(h-1,0)}^{\min(h+1, N_c-1)} P(H_{ar} = h' | H_{ar} = h) \\
&\quad \cdot \sum_{g'=\max(g-1,0)}^{\min(g+1, N_c-1)} P(H_{br} = g' | H_{br} = g) \cdot \sum_{q=0}^{\infty} P(E = q | Q_e = e) \cdot V^{(i)}(e', h', g', \min(b-w+q, N_b-1)) \\
&= P_{out}(h, g, w) + \lambda \cdot \mathbb{E}_{e,h,g,b} [V^{(i)}(e', h', g', \min(b-w+q, N_b-1))], \tag{25}
\end{aligned}$$

where  $\mathbb{E}_{e,h,g,b}(\cdot)$  denotes the expected value conditioned on the system state  $s = (e, h, g, b)$ .

### B. Monotonic and Bounded Differential Structure of Expected Long-Term Reward

*Theorem 3:* For any fixed system state  $s = (e, h, g, b > 0) \in \mathcal{S}$  in the  $i^{\text{th}}$  value iteration, the expected long-term reward is non-increasing in the battery state, and the differential value of the expected long-term rewards for two adjacent battery states is non-negative and not larger than one, i.e.,  $1 \geq V^{(i)}(e, h, g, b-1) - V^{(i)}(e, h, g, b) \geq 0, \forall b \in \mathcal{Q}_b \setminus \{0\}$ . Moreover, the optimal transmission policy  $\pi^*$  is also satisfied with the above special structure, i.e.,  $1 \geq V_{\pi^*}(e, h, g, b-1) - V_{\pi^*}(e, h, g, b) \geq 0, \forall b \in \mathcal{Q}_b \setminus \{0\}$ .

**Proof:** See Appendix C for details. ■

This monotonic structure points out the relationship between the long-term reward and the battery state, for which the more the energy in the battery is, the better the system performance is. Moreover, the bounded differential structure is mainly derived from the characteristic of the outage probability, which concludes that the added value of the long-term reward caused by the increased battery energy is finite and bounded.

### C. Ceiling Structure and Threshold Structure of Optimal Relay Power

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 level  $\tilde{w}$  is called ceiling power, if the reward functions begin to hold stably 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 (20), when the system is operated in sufficiently high SNR regimes, i.e.,  $N_0 \rightarrow 0$ , 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 4:* For any fixed system state  $s = (e, h, g, b) \in \mathcal{S}$  and cooperation protocol  $f \in \{DF, AF\}$ , the optimal relay power action is not larger than the relay's ceiling power, i.e.,  $w^*(s, f) \leq \min(\tilde{w}(h, g, f), b)$ .

**Proof:** According to (25), for any iteration  $i$  and any fixed system state  $s = (e, h, g, b) \in \mathcal{S}$ , the differential value of the two long-term rewards with respect to the relay transmission power actions  $w (\tilde{w} < w \leq b)$  and  $\tilde{w}$  can be computed as

$$\begin{aligned} & V_{w,f}^{(i+1)}(e, h, g, b) - V_{\tilde{w},f}^{(i+1)}(e, h, g, b) \\ &= R_{w,f}(h, g) - R_{\tilde{w},f}(h, g) + \\ & \quad \lambda \cdot \mathbb{E}_{e,h,g,b} [V^{(i)}(e', h', g', \min(b-w+q, N_b-1)) - V^{(i)}(e', h', g', \min(b-\tilde{w}+q, N_b-1))] \\ &= \lambda \cdot \mathbb{E}_{e,h,g,b} [V^{(i)}(e', h', g', \min(b-w+q, N_b-1)) - V^{(i)}(e', h', g', \min(b-\tilde{w}+q, N_b-1))] . \end{aligned}$$

By applying Theorem 3, it can be easily seen that  $V_{w,f}^{(i+1)}(e, h, g, b) > V_{\tilde{w},f}^{(i+1)}(e, h, g, b)$ . From the value iteration algorithm in (24), 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  $w^* \leq \min(\tilde{w}, b)$ . ■

*Corollary 1:* For any fixed system state  $s = (e, h, g, b) \in \mathcal{S}$ , 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 4, 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. ■

*Theorem 5:* For any fixed system state  $s = (e, h, g, b > 0) \in \mathcal{S}$  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  $b \in \mathcal{Q}_b \setminus \{0\}$ . For any iteration  $i$  and system state  $s = (e, h, g, b > 0) \in \mathcal{S}$ , according to (25), the value difference of the two long-term rewards for the relay power action  $w = 1$  and  $w = 0$  can be expressed as

$$\begin{aligned} & V_{w=1}^{(i+1)}(e, h, g, b) - V_{w=0}^{(i+1)}(e, h, g, b) \\ &= R_{w=1}(h, g) - R_{w=0}(h, g) \\ & \quad + \lambda \cdot \mathbb{E}_{e, h, g, b} [V^{(i)}(e', h', g', \min(b-1+q, N_b-1)) - V^{(i)}(e', h', g', \min(b+q, N_b-1))]. \end{aligned} \quad (26)$$

By using (20), the value difference in (26) in high SNRs is written as

$$\begin{aligned} & \lim_{N_0 \rightarrow 0} [V_{w=1}^{(i+1)}(e, h, g, b) - V_{w=0}^{(i+1)}(e, h, g, b)] \\ &= -1 + \lambda \cdot \mathbb{E}_{e, h, g, b} [V^{(i)}(e', h', g', \min(b-1+q, N_b-1)) - V^{(i)}(e', h', g', \min(b+q, N_b-1))]. \end{aligned} \quad (27)$$

By applying Theorem 3, for any system state  $s' = (e', h', g', b > 0) \in \mathcal{S}$ , it can be easily seen that

$$1 \geq V^{(i)}(e', h', g', \min(b-1+q, N_b-1)) - V^{(i)}(e', h', g', \min(b+q, N_b-1)) \geq 0. \quad (28)$$

Since  $0 < \lambda < 1$ , the two long-term rewards in (26) in high SNRs meet the following relationship

$$\lim_{N_0 \rightarrow 0} V_{w=1}^{(i+1)}(e, h, g, b) < \lim_{N_0 \rightarrow 0} V_{w=0}^{(i+1)}(e, h, g, b). \quad (29)$$

From (24), 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. Furthermore, the optimal transmission strategy for the relay node in high SNR regimes is to keep silent when the battery is empty; otherwise, it is sufficient to attain the best long-term performance by only spending one energy quantum for helping relay the signals.

## V. PERFORMANCE ANALYSIS OF OUTAGE PROBABILITY

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

### A. Expected Outage Probability

We introduce the steps to compute the expected outage probability for any transmission policy  $\pi$ . First, the battery state transition probability associated with the transmission policy  $\pi$  in the state  $s = (e, h, g, b)$  can be derived as [23]

$$P_{\pi(e,h,g,b)}(Q_b = b' | Q_b = b) \quad (30)$$

$$= \begin{cases} 0, & 0 \leq b' \leq b - w - 1; \\ P(E = b' - b + w | Q_e = e), & b - w \leq b' \leq N_b - 2; \\ 1 - \sum_{b'=0}^{N_b-2} P_{\pi(e,h,g,b)}(Q_b = b' | Q_b = b), & b' = N_b - 1, \end{cases}$$

where  $b, b' \in \{0, \dots, N_b - 1\}$  and  $w$  is the relay power action in the policy  $\pi$ . By utilizing (15), the system state transition probability with respect to the policy  $\pi$  can be calculated as

$$P_{\pi}(s' = (e', h', g', b') | s = (e, h, g, b)) \quad (31)$$

$$= P(Q_e = e' | Q_e = e) \cdot P(H_{ar} = h' | H_{ar} = h) \cdot P(H_{br} = g' | H_{br} = g) \cdot P_{\pi(e,h,g,b)}(Q_b = b' | Q_b = b),$$

where  $h' \in \{\max(0, h - 1), \dots, \min(h + 1, N_c - 1)\}$ ,  $g' \in \{\max(0, g - 1), \dots, \min(g + 1, N_c - 1)\}$ ,  $e, e' \in \{0, 1, \dots, N_e - 1\}$ , and  $h, g \in \{0, \dots, N_c - 1\}$ . Next, let  $p_{\pi}(s = (e, h, g, b))$  represent the steady state probability of the state  $s = (e, h, g, b)$  for the policy  $\pi$ , and the following linear equations can be formulated [23]:

$$\begin{cases} \sum_{s \in \mathcal{S}} p_{\pi}(s = (e, h, g, b)) = 1, \\ \sum_{s \in \mathcal{S}} P_{\pi}(s' = (e', h', g', b') | s = (e, h, g, b)) \cdot p_{\pi}(s = (e, h, g, b)) = p_{\pi}(s' = (e', h', g', b')). \end{cases}$$

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

$$\bar{R} = \sum_{(e,h,g,b) \in \mathcal{S}} p_{\pi}(e, h, g, b) \times R_w(s = (e, h, g, b)). \quad (32)$$

### B. Saturated Structure of Outage Probability

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.,  $P_\pi(b=0) = \sum_{(e,h,g,b=0) \in \mathcal{S}} p_\pi(e,h,g,b=0)$ .

*Theorem 6:* 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 (32), the expected reward of the optimal policy  $\pi^*$  can be expressed as

$$\bar{R} = \sum_{(e,h,g,b) \in \mathcal{S}} p_{\pi^*}(e,h,g,b) \times R_{w^*}(s=(e,h,g,b)), \quad (33)$$

where  $p_{\pi^*}(s)$  is the steady state probability associated with the optimal policy  $\pi^*$ , and  $w^*$  is the optimal relay action. By considering the battery state, the expected reward can be rewritten as

$$\bar{R} = \sum_{e=0}^{N_e-1} \sum_{h=0}^{N_c-1} \sum_{g=0}^{N_c-1} [p_{\pi^*}(e,h,g,b=0) \times R_{w^*}(e,h,g,b=0) + p_{\pi^*}(e,h,g,b \geq 1) \times R_{w^*}(e,h,g,b \geq 1)].$$

By applying Theorem 5, the optimal relay power action  $w^*(e,h,g,b > 0) = 1$  for  $\forall e \in \mathcal{Q}_e$ ,  $\forall h \in \mathcal{H}_{ar}$  and  $\forall g \in \mathcal{H}_{br}$  in sufficiently high SNRs. According to (20), 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 can be expressed as

$$\lim_{N_0 \rightarrow 0} \bar{R} = \sum_{e=0}^{N_e-1} \sum_{h=0}^{N_c-1} \sum_{g=0}^{N_c-1} p_{\pi^*}(e,h,g,b=0) = P_{\pi^*}(b=0), \quad (34)$$

which means 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 state steady probability in (32). 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 expected outage probability in sufficiently high SNRs.

*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.,  $P(E=0|Q_e=e)$ .

It can be observed from [5] that the energy deficiency probability  $P(E = 0|Q_e = e)$  is affected by the solar panel size  $\Omega$ , the size of one energy quantum  $E_u$ , the transmission period  $T$ , the energy conversion efficiency  $\eta$ , as well as the mean and variance of the underlying Gaussian distribution in the stochastic solar power 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.,  $P(E = 0|Q_e = e) = 0, \forall e \in \mathcal{Q}_e$ .

**Proof:** Without loss of generality, the battery state in the  $t^{\text{th}}$  period ( $t \geq 1$ ) can be described as  $b_t = b_{t-1} - w_t^* + q_t$ , where  $w_t^*$  and  $q_t$  are the optimal relay power action and the number of harvested energy quanta in the  $t^{\text{th}}$  period, respectively, and  $b_0$  is the original battery state. It is known from Theorem 6 that the battery empty probability  $P(b = 0)$  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 5, 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.,  $b_{t-1} \geq 1$ , 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 transmission period and the battery empty probability is equal to zero. By applying Theorem 6, the expected outage probability approaches to zero in sufficiently high SNRs. From the aforementioned discussions, the corollary is proved. ■

## VI. SIMULATION RESULTS

In this section, the outage probability of our proposed optimal policy based on the stochastic EH model in [5] is evaluated by computer simulations. The outage probabilities are calculated according to (32), Theorem 1 and Theorem 2. Meanwhile, the simulation results are computed using the Monte-Carlo method. The number of solar EH states  $N_e$  is four, the solar panel area size is set as  $\Omega = 10 \text{ cm}^2$ , and the solar energy conversion efficiency is  $\eta = 20 \%$ . Meanwhile, the number of battery states is given as  $N_b = 12$ , and the battery state is initialized

randomly. The transmission period and the basic transmission power are set to  $T = 300$  sec and  $P_u = 35 \times 10^3 \mu\text{W}$ , respectively. For the settings of the channel model, the number of channel states is  $N_c = 6$ , and the channel with the average power  $\theta = 1$  is quantized as  $\Gamma = \{0, 0.3, 0.6, 1.0, 2.0, 3.0, \infty\}$ . The channel fading is generated using Jakes' model [24]. It is assumed that the wireless nodes have low mobility, and for a normalized Doppler frequency  $f_D = 0.05$ , the channel coherence time  $\frac{T}{2f_D}$  is fifty minutes. We assume that the target rates  $R_{th1}$  and  $R_{th2}$  are identical and equal to  $R_{th}/2$ , as well as the unit of  $R_{th}$  is bit/s/Hz. In the value iteration algorithm for finding the optimal policy, the discount factor and the stopping criterion are set as  $\lambda = 0.99$  and  $\varepsilon = 10^{-5}$ , respectively. The above parameters are default settings, except as otherwise stated. Finally, since the relay transmission power is related with the solar irradiance, a normalized SNR is defined with respect to  $P_u$  in the simulations.

Fig. 2 shows the outage probabilities of our proposed optimal policy for different target sum rates  $R_{th}$  and transmission power levels of the source nodes  $P$  when the DF cooperation protocol is exploited. It can be seen that the analysis results and simulation results match very well. When the target sum rate  $R_{th}$  gets smaller, the outage probability becomes better. Also, the outage probability can be improved with the increase of the transmission power of the source nodes. This is because the instantaneous throughput can be increased by enlarging the transmission power  $P$ . Besides, 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 6.

Fig. 3 shows the outage probability of our proposed optimal policy for different  $R_{th}$  and  $P$  when the AF cooperation protocol is exploited at the relay, and the dashed line without markers at the bottom of this figure indicates the battery empty probability for the optimal policy in high SNR regimes. 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 Theorem 2. Besides, similar performance trends can be observed in the AF mode and the DF mode, e.g., the impacts of the target sum rate  $R_{th}$  and the transmission power  $P$  on the outage probability, the saturated structure, etc.

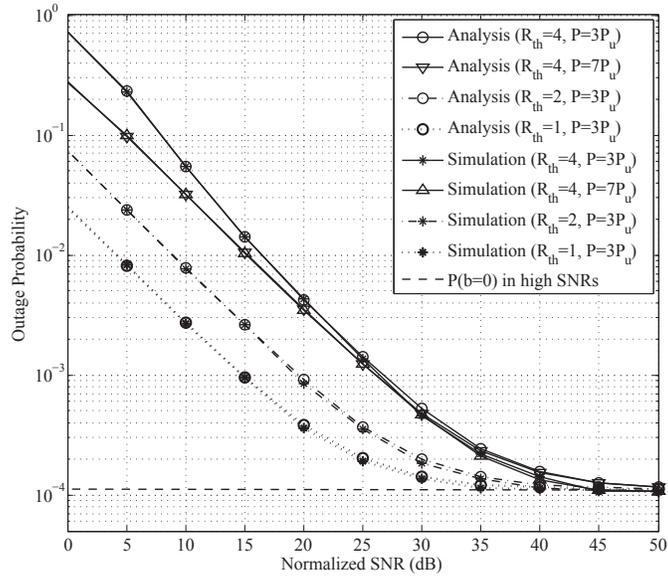


Fig. 2. Outage probability of DF mode for different target rates  $R_{th}$  and source nodes transmission power  $P$ .

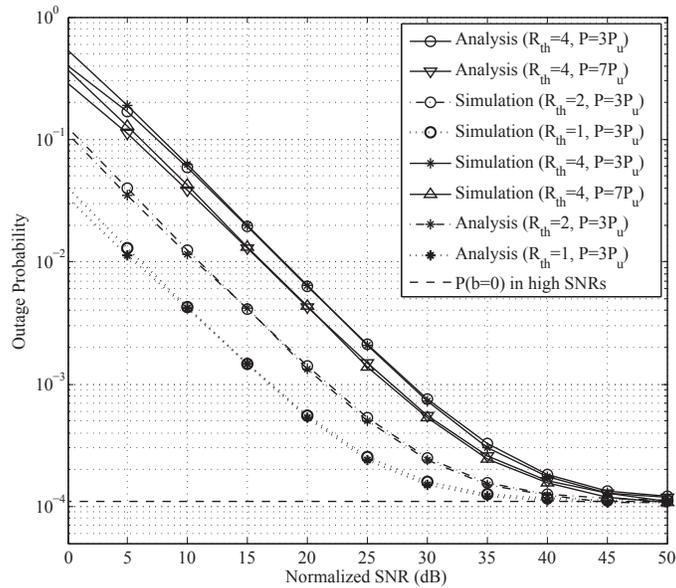


Fig. 3. Outage probability of AF mode for different target rates  $R_{th}$  and source nodes transmission power  $P$ .

Fig. 4 compares the outage probabilities of our proposed optimal policy and two myopic policies for different target sum rates  $R_{th}$  when the DF cooperation protocol is exploited. For these 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  $P_u$ . It can be seen that the outage probability of our proposed optimal policy is superior to those of the two myopic policies. The outage probabilities of these three 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 5, 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. In addition, Fig. 5 compares the outage probabilities of our proposed optimal policy and the two myopic policies when the AF cooperation protocol is used. As compared with the DF mode, similar performance trends can be found in this figure.

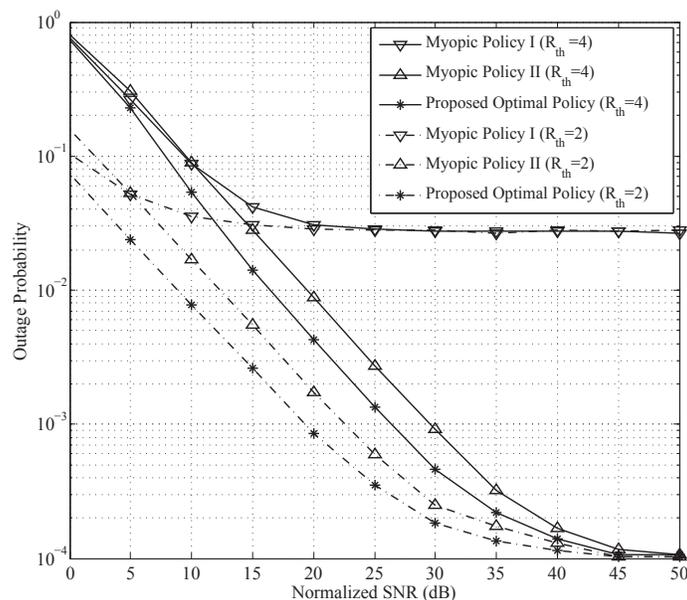


Fig. 4. Outage probabilities of the proposed optimal policy and two myopic policies in DF mode.

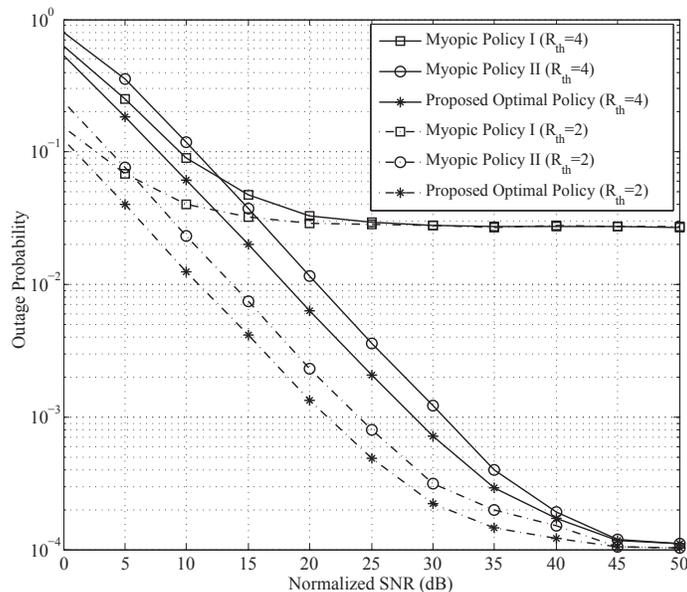


Fig. 5. Outage probabilities of the proposed optimal policy and two myopic policies in AF mode.

Fig. 6 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 exploited. The unit of  $E_u$  is given as  $300 * 10^3 \mu\text{J}$ . 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  $\Omega$  gets larger or one energy quantum  $E_u$  gets smaller. The reason can be explained as follows. Since a bigger solar panel size  $\Omega$  means there is more energy harvested within one transmission period, the energy deficiency probability  $P(E = 0 | Q_e = e)$  and the battery empty probability  $P(b = 0)$  can be decreased by increasing  $\Omega$ . 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.

## VII. CONCLUSION

In this paper, the optimal and adaptive relay transmission policy for minimizing the long-term outage probability in the EH TWR network is proposed. Unlike the previous works, we make use of stochastic EH models to formulate the solar irradiance condition and design an MDP framework to optimize the relay transmission policy in accordance with the solar ESI, CSI and

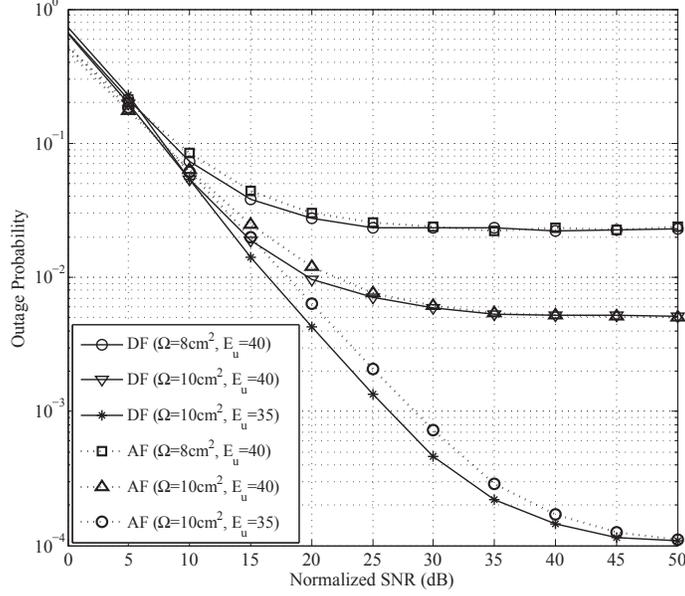


Fig. 6. Impact of the sizes of solar panel area and energy quantum on the outage probabilities in DF or AF modes

finite battery condition. We first find the monotonic and bounded differential structure of the long-term reward. Furthermore, we study the property of the optimal solutions, and the ceiling and threshold structures of the optimal relay transmission power are discovered. Moreover, the expected outage probability is theoretically analyzed and an interesting saturated structure is found to predict the performance limit of the outage probability at sufficiently high SNRs. The theoretical results are substantiated through extensive computer simulations, and the proposed optimal transmission policy outperforms the other two myopic policies significantly.

## APPENDIX A

### PROOF OF THEOREM 1

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

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

where  $\tilde{\gamma}_{th1} = \frac{N_0}{P} (2^{2R_{th1}} - 1)$ ,  $\tilde{\gamma}_{th2} = \frac{N_0}{P_r} (2^{2R_{th1}} - 1)$ ,  $\tilde{\gamma}_{th3} = \frac{N_0}{P} (2^{2R_{th2}} - 1)$ ,  $\tilde{\gamma}_{th4} = \frac{N_0}{P_r} (2^{2R_{th2}} - 1)$  and  $c = \frac{N_0}{P} (2^{2R_{th}} - 1)$ . Substituting the above three events into (17) yields

$$P_{out,DF}(w, h, g) = \Pr\{(\gamma_1 < \gamma_{th1}) \cup (\gamma_2 < \gamma_{th2}) \cup (\gamma_1 + \gamma_2 < c) | P_r = wP_u, H_{ar} = h, H_{br} = g\}, \quad (35)$$

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

$$\Pr\{A \cup B \cup C\} = \Pr\{A \cup B\} + \Pr\{\overline{(A \cup B)} \cap C\} = 1 - \Pr\{\overline{A} \cap \overline{B}\} + \Pr\{\overline{A} \cap \overline{B} \cap C\},$$

where  $A, B$  and  $C$  are random events, the conditional outage probability in (35) is expressed as

$$\begin{aligned} P_{out,DF}(w, h, g) &= 1 - \Pr\{(\gamma_1 \geq \gamma_{th1}) \cap (\gamma_2 \geq \gamma_{th2}) | H_{ar} = h, H_{br} = g\} \\ &\quad + \Pr\{(\gamma_1 \geq \gamma_{th1}) \cap (\gamma_2 \geq \gamma_{th2}) \cap (\gamma_1 + \gamma_2 < c) | H_{ar} = h, H_{br} = g\} \\ &= 1 - \Pr\{\gamma_1 \geq \gamma_{th1} | \Gamma_h \leq \gamma_1 < \Gamma_{h+1}\} \cdot \Pr\{\gamma_2 \geq \gamma_{th2} | \Gamma_g \leq \gamma_2 < \Gamma_{g+1}\} \\ &\quad + \Pr\{(\gamma_1 \geq \gamma_{th1}) \cap (\gamma_2 \geq \gamma_{th2}) \cap (\gamma_1 + \gamma_2 < c) | \Gamma_h \leq \gamma_1 < \Gamma_{h+1}, \Gamma_g \leq \gamma_2 < \Gamma_{g+1}\}. \end{aligned} \quad (36)$$

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 the Case 1, it is straightforward to derive  $P_{out,DF}(w, h, g) = 1$ . For the Case 2, by letting  $a = \max(\gamma_{th1}, \Gamma_h)$  and  $b = \max(\gamma_{th2}, \Gamma_g)$  and from (36), the conditional outage probability can be explicitly calculated as

$$\begin{aligned} P_{out,DF}(w, h, g) &= 1 - \frac{\Pr\{(\gamma_1 \geq \gamma_{th1}) \cap (\Gamma_h \leq \gamma_1 < \Gamma_{h+1})\}}{\Pr\{\Gamma_h \leq \gamma_1 < \Gamma_{h+1}\}} \cdot \frac{\Pr\{(\gamma_2 \geq \gamma_{th2}) \cap (\Gamma_g \leq \gamma_2 < \Gamma_{g+1})\}}{\Pr\{\Gamma_g \leq \gamma_2 < \Gamma_{g+1}\}} \\ &\quad + \frac{\Pr\{(\gamma_1 \geq \gamma_{th1}) \cap (\gamma_2 \geq \gamma_{th2}) \cap ((\gamma_1 + \gamma_2) < c) \cap (\Gamma_h \leq \gamma_1 < \Gamma_{h+1}) \cap (\Gamma_g \leq \gamma_2 < \Gamma_{g+1})\}}{\Pr\{\Gamma_h \leq \gamma_1 < \Gamma_{h+1}\} \cdot \Pr\{\Gamma_g \leq \gamma_2 < \Gamma_{g+1}\}} \\ &= 1 + \frac{T - \Pr\{a \leq \gamma_1 < \Gamma_{h+1}\} \cdot \Pr\{b \leq \gamma_2 < \Gamma_{g+1}\}}{\Pr\{\Gamma_h \leq \gamma_1 < \Gamma_{h+1}\} \cdot \Pr\{\Gamma_g \leq \gamma_2 < \Gamma_{g+1}\}} \\ &= 1 + \frac{T - (e^{-a/\theta} - e^{-\Gamma_{h+1}/\theta}) \cdot (e^{-b/\theta} - e^{-\Gamma_{g+1}/\theta})}{(e^{-\Gamma_h/\theta} - e^{-\Gamma_{h+1}/\theta}) \cdot (e^{-\Gamma_g/\theta} - e^{-\Gamma_{g+1}/\theta})}, \end{aligned} \quad (37)$$

where  $T = \Pr\{(a \leq \gamma_1 < \Gamma_{h+1}) \cap (b \leq \gamma_2 < \Gamma_{g+1}) \cap (\gamma_1 + \gamma_2 < c)\}$ . Subsequently,  $T$  is computed by discussing the relationship among  $a, b, c$  and the channel quantization thresholds. As shown in Fig. 7,  $(a \leq \gamma_1 < \Gamma_{h+1}) \cap (b \leq \gamma_2 < \Gamma_{g+1})$  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 four subcases:

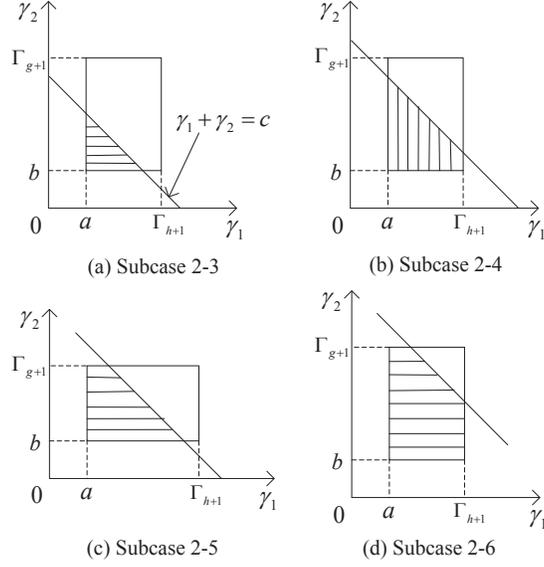


Fig. 7. The relationship among  $a$ ,  $b$ ,  $c$  and channel quantization thresholds when calculating  $P_{out,DF}(w, h, g)$  in Case 2.

- 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

$$T = \Pr \{ (a \leq \gamma_1 < \Gamma_{h+1}) \cap (b \leq \gamma_2 < \Gamma_{g+1}) \} = \Pr \{ a \leq \gamma_1 < \Gamma_{h+1} \} \cdot \Pr \{ b \leq \gamma_2 < \Gamma_{g+1} \}. \quad (38)$$

By substituting (38) into (37), 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. 7(a), and thus  $T$  is calculated as

$$T = \int_a^{c-b} f(\gamma_1) d\gamma_1 \int_b^{-\gamma_1+c} f(\gamma_2) d\gamma_2 = e^{-(a+b)/\theta} - e^{-c/\theta} - \frac{1}{\theta} e^{-c/\theta} (c - b - a); \quad (39)$$

- Subcase 2-4 ( $(b + \Gamma_{h+1}) < c < (a + \Gamma_{g+1})$ ): In this condition, the intersection area is a trapezoid shown as the shadow area in Fig. 7(b), and thus  $T$  is calculated as

$$T = \int_a^{\Gamma_{h+1}} f(\gamma_1) d\gamma_1 \int_b^{-\gamma_1+c} f(\gamma_2) d\gamma_2 = e^{-(a+b)/\theta} - e^{-(\Gamma_{h+1}+b)/\theta} - \frac{1}{\theta} e^{-c/\theta} (\Gamma_{h+1} - a); \quad (40)$$

- Subcase 2-5 ( $(a + \Gamma_{g+1}) < c < (b + \Gamma_{h+1})$ ): In this condition, the intersection area is a trapezoid shown as the shadow area in Fig. 7(c), and thus  $T$  is calculated as

$$T = \int_b^{\Gamma_{g+1}} f(\gamma_2) d\gamma_2 \int_a^{c-\gamma_2} f(\gamma_1) d\gamma_1 = e^{-(a+b)/\theta} - e^{-(a+\Gamma_{g+1})/\theta} - \frac{1}{\theta} e^{-c/\theta} (\Gamma_{g+1} - b); \quad (41)$$

- Subcase 2-6 ( $\max((a+\Gamma_{g+1}), (b+\Gamma_{h+1})) \leq c < (\Gamma_{h+1} + \Gamma_{g+1})$ ): In this condition, the intersection area is a pentagon shown as the shadow area in Fig. 7(d), thus  $T$  is calculated as

$$\begin{aligned}
T &= \Pr\{(a \leq \gamma_1 < \Gamma_{h+1}) \cap (b \leq \gamma_2 < \Gamma_{g+1})\} - \Pr\{(a \leq \gamma_1 < \Gamma_{h+1}) \cap (b \leq \gamma_2 < \Gamma_{g+1}) \cap (\gamma_1 + \gamma_2 > c)\} \\
&= \int_a^{\Gamma_{h+1}} f(\gamma_1) d\gamma_1 \int_b^{\Gamma_{g+1}} f(\gamma_2) d\gamma_2 - \int_{c-\Gamma_{g+1}}^{\Gamma_{h+1}} f(\gamma_1) d\gamma_1 \int_{-\gamma_1+c}^{\Gamma_{g+1}} f(\gamma_2) d\gamma_2 \\
&= (e^{-a/\theta} - e^{-\Gamma_{h+1}/\theta})(e^{-b/\theta} - e^{-\Gamma_{g+1}/\theta}) - e^{-(\Gamma_{h+1} + \Gamma_{g+1})/\theta} + e^{-c/\theta} + \frac{1}{\theta} e^{-c/\theta} (c - \Gamma_{g+1} - \Gamma_{h+1}).
\end{aligned}$$

Thus, we complete the proof of Theorem 1.

## APPENDIX B

### PROOF OF THEOREM 2

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

$$\mathcal{E}_{out,AF}^1 = \frac{x_1 x_2}{x_1 + x_2} < m_1, \mathcal{E}_{out,AF}^2 = \frac{y_1 y_2}{y_1 + y_2} < m_2, \quad (42)$$

where  $x_1 = \gamma_1 \eta_1$ ,  $x_2 = \gamma_2 (\eta_2 + \eta_r)$ ,  $m_1 = \frac{\eta_2 + \eta_r}{\eta_r} (2^{2R_{th1}} - 1)$ ,  $y_1 = \gamma_1 (\eta_1 + \eta_r)$ ,  $y_2 = \gamma_2 \eta_2$  and  $m_2 = \frac{\eta_1 + \eta_r}{\eta_r} (2^{2R_{th2}} - 1)$ . Thus, substituting (42) into (18) yields

$$P_{out,AF}(w, h, g) = \Pr\left\{\left(\frac{x_1 x_2}{x_1 + x_2} < m_1\right) \cup \left(\frac{y_1 y_2}{y_1 + y_2} < m_2\right) \mid P_r = wP_u, H_{ar} = h, H_{br} = g\right\}. \quad (43)$$

By considering the well-known harmonic mean inequality  $xy/(x+y) \leq \min(x, y)$  in [25], the conditional outage probability can be expressed as

$$\begin{aligned}
&P_{out,AF}(w, h, g) \\
&\geq \Pr\{(\min(x_1, x_2) < m_1) \cup (\min(y_1, y_2) < m_2) \mid P_r = wP_u, H_{ar} = h, H_{br} = g\} \\
&= 1 - \Pr\{(x_1 \geq m_1) \cap (x_2 \geq m_1) \cap (y_1 \geq m_2) \cap (y_2 \geq m_2) \mid P_r = wP_u, H_{ar} = h, H_{br} = g\} \\
&= 1 - \Pr\{(\gamma_1 \geq \gamma_{th1}) \cap (\gamma_1 \geq \gamma_{th2}) \mid \Gamma_h \leq \gamma_1 < \Gamma_{h+1}\} \cdot \Pr\{(\gamma_2 \geq \gamma_{th3}) \cap (\gamma_2 \geq \gamma_{th4}) \mid \Gamma_g \leq \gamma_2 < \Gamma_{g+1}\} \\
&\hspace{15em} (44)
\end{aligned}$$

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

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

$$\Pr\{(\gamma_1 \geq \gamma_{th1}) \cap (\gamma_1 \geq \gamma_{th2}) | \Gamma_h \leq \gamma_1 < \Gamma_{h+1}\} = \Pr\{(\gamma_2 \geq \gamma_{th3}) \cap (\gamma_2 \geq \gamma_{th4}) | \Gamma_g \leq \gamma_2 < \Gamma_{g+1}\} = 1.$$

Therefore,  $P_{out,AF}(w, h, g) = 0$ .

- Case 3 (Otherwise): It can also be easily obtained that

$$\Pr\{(\gamma_1 \geq \gamma_{th1}) \cap (\gamma_1 \geq \gamma_{th2}) | \Gamma_h \leq \gamma_1 < \Gamma_{h+1}\} = \frac{\Pr\{\max(\gamma_{th1}, \gamma_{th2}) \leq \gamma_1 < \Gamma_{h+1}\}}{\Pr\{\Gamma_h \leq \gamma_1 < \Gamma_{h+1}\}}; \quad (45)$$

$$\Pr\{(\gamma_2 \geq \gamma_{th3}) \cap (\gamma_2 \geq \gamma_{th4}) | \Gamma_g \leq \gamma_2 < \Gamma_{g+1}\} = \frac{\Pr\{\max(\gamma_{th3}, \gamma_{th4}) \leq \gamma_2 < \Gamma_{g+1}\}}{\Pr\{\Gamma_g \leq \gamma_2 < \Gamma_{g+1}\}}. \quad (46)$$

Substituting (45) and (46) into (44) yields

$$P_{out,AF}(w, h, g) \approx 1 - \frac{e^{-\max(\gamma_{th1}, \gamma_{th2})/\theta} - e^{-\Gamma_{h+1}/\theta}}{e^{-\Gamma_h/\theta} - e^{-\Gamma_{h+1}/\theta}} \cdot \frac{e^{-\max(\gamma_{th3}, \gamma_{th4})/\theta} - e^{-\Gamma_{g+1}/\theta}}{e^{-\Gamma_g/\theta} - e^{-\Gamma_{g+1}/\theta}}. \quad (47)$$

Thus, we complete the proof of Theorem 2.

## APPENDIX C

### PROOF OF THEOREM 3

We prove the theorem 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 (23) can be written as

$$V_w^{(1)}(s) = R_w(s) + \lambda \sum_{s' \in S} P_a(s'|s) V^{(0)}(s') = R_w(s) = P_{out}(h, g, w). \quad (48)$$

When  $w \in \{0, 1, \dots, b-1\}$ , it can be derived directly from (48) that

$$V_w^{(1)}(e, h, g, b-1) = V_w^{(1)}(e, h, g, b). \quad (49)$$

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.,  $1 \geq P_{out}(h, g, w = b-1) - P_{out}(h, g, w = b) \geq 0$ , the following inequality holds

$$1 \geq V_{w=b-1}^{(1)}(e, h, g, b-1) - V_{w=b}^{(1)}(e, h, g, b) \geq 0. \quad (50)$$

By considering (49), (50) and (24), it can be deduced that

$$1 \geq V^{(1)}(e, h, g, b-1) - V^{(1)}(e, h, g, b) \geq 0, \forall b \in \mathcal{Q}_b \setminus \{0\}. \quad (51)$$

**Step 2:** Assuming  $1 \geq V^{(k)}(e, h, g, b-1) - V^{(k)}(e, h, g, b) \geq 0, \forall b \in \mathcal{Q}_b \setminus \{0\}$ . According to (25), when  $w \in \{0, 1, \dots, b-1\}$ , the value difference between the long-term rewards of two adjacent battery states in iteration  $k+1$  can be written as

$$\begin{aligned} & V_w^{(k+1)}(e, h, g, b-1) - V_w^{(k+1)}(e, h, g, b) \\ &= \lambda \cdot \mathbb{E}_{e, h, g, b} [V^{(k)}(e', h', g', \min(b-1-w+q, N_b-1)) - V^{(k)}(e', h', g', \min(b-w+q, N_b-1))] . \end{aligned}$$

With the assumption, it can be easily seen that

$$1 \geq V_w^{(k+1)}(e, h, g, b-1) - V_w^{(k+1)}(e, h, g, b) \geq 0, \forall w \in \{0, 1, \dots, b-1\}. \quad (52)$$

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

$$\begin{aligned} & V_{w=b-1}^{(k+1)}(e, h, g, b-1) - V_{w=b}^{(k+1)}(e, h, g, b) \\ &= P_{out}(h, g, w=b-1) - P_{out}(h, g, w=b) \\ &\quad + \lambda \cdot \mathbb{E}_{e, h, g, b} [V^{(k)}(e', h', g', \min(q, N_b-1)) - V^{(k)}(e', h', g', \min(q, N_b-1))] \\ &= P_{out}(h, g, w=b-1) - P_{out}(h, g, w=b). \end{aligned}$$

Similarly to (50) in Step 1, the following inequality also holds

$$1 \geq V_{w=b-1}^{(k+1)}(e, h, g, b-1) - V_{w=b}^{(k+1)}(e, h, g, b) \geq 0. \quad (53)$$

According to (52), (53) and (24), it can be easily proved that

$$1 \geq V^{(k+1)}(e, h, g, b-1) - V^{(k+1)}(e, h, g, b) \geq 0, \forall b \in \mathcal{Q}_b \setminus \{0\}, \quad (54)$$

**Step 3:** Combining the results of Step1 and Step2, we use the induction method and prove

$$1 \geq V^{(i)}(e, h, g, b-1) - V^{(i)}(e, h, g, b) \geq 0, \forall b \in \mathcal{Q}_b \setminus \{0\}, \forall i. \quad (55)$$

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

## REFERENCES

- [1] S. Sudevalayam and P. Kulkarni, "Energy Harvesting Sensor Nodes: Survey and Implications," *IEEE Commun. Surveys Tutorials*, vol. 13, no. 3, pp. 443–461, 2011.
- [2] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with Energy Harvesting Nodes in Fading Wireless Channels Optimal Policies," *IEEE J. Select. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, Sep. 2011.
- [3] 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.
- [4] M.-L. Ku, Y. Chen, and K. J. R. Liu, "Data-Driven Stochastic Transmission Policies for Energy Harvesting Sensor Communications," *IEEE Journal of Selected Areas in Communications*, Special Issue on Wireless Communications Powered by Energy Harvesting and Wireless Energy Transfer, to appear.
- [5] —, "Data-Driven Stochastic Scheduling for Solar-Powered Sensor Communications," in *Proc. of 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, GA, USA, Dec. 2014, pp. 238–242.
- [6] N. R. E. Laboratory, "(2012)Solar radiation resource information," [Online] Available: <http://www.nrel.gov/rredc/>.
- [7] K. J. R. Liu, A. K. Sadek, W. Su, and A. Kwasinski, *Cooperative Communications and Networking*. Cambridge University Press, 2008.
- [8] O. Orhan and E. Erkip, "Energy harvesting two-hop networks Optimal policies for the multi-energy arrival case," in *Proc. of 2012 35th IEEE Sarnoff Symposium (SARNOFF)*, May 2012, pp. 1–6.
- [9] 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.
- [10] C. Huang, R. Zhang, and S. Cui, "Throughput Maximization for the Gaussian Relay Channel with Energy Harvesting Constraints," *IEEE J. Select. Areas Commun.*, vol. 31, no. 8, pp. 1469–1479, Aug. 2013.
- [11] 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, 2013.
- [12] 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.
- [13] 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.
- [14] K. Tutuncuoglu, B. Varan, and A. Yener, "Energy Harvesting Two-Way Half-Duplex Relay Channel with Decode-and-Forward Relaying Optimum Power Policies," in *Proc. of 2013 18th International Conference on Digital Signal Processing (DSP)*, July 2013, pp. 1–6.
- [15] B. Varan and A. Yener, "The Energy Harvesting Two-Way Decode-and-Forward Relay Channel with Stochastic Data Arrival," in *Proc. of 2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2013, pp. 371–374.
- [16] K. Tutuncuoglu, B. Varan, and A. Yener, "Optimum Transmission Policies for Energy Harvesting Two-way Relay Channels," in *Proc. of 2013 International Conference on Communication (ICC)*, Budapest, Hungary, June 2013, pp. 586–590.
- [17] 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. of 2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2013, pp. 345–348.
- [18] D. Li, C. Shen, and Z. Qiu, "Two-way relay beamforming for sum-rate maximization and energy harvesting," in *Proc. of 2013 International Conference on Communication (ICC)*, Budapest, Hungary, June 2013, pp. 3115–3120.

- [19] 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.
- [20] M. Puterman, *Markov Decision Process—Discrete Stochastic Dynamic Programming*. John Wiley and Sons, 1994.
- [21] 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.
- [22] D. Bertsekas, *Dynamic Programming and Optimal control*. Athena Scientific, 2000.
- [23] 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. of 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, GA, USA, Dec. 2014, pp. 288–292.
- [24] W. C. Jakes, *Microwave Mobile Communication*. New York: Wiley, 1974.
- [25] K.-S. Hwang, Y. chai Ko, and M.-S. Alouini, "Performance Bounds for Two-Way Amplify-and-Forward Relaying Based on Relay Path Selection," in *Proc. of 2009 IEEE 69th Vehicular Technology Conference (VTC) Spring*, April 2009, pp. 1–5.