Performance Analysis for Two-Way Network-Coded Dual-Relay Networks With Stochastic Energy Harvesting

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

***Abstract*— In this paper, we consider an energy harvesting (EH) two-way (TW) dual-relay network, including one non-EH relay and one EH relay equipped with a finite-sized battery. In the network, a space-time transmission protocol with space-time network coding is designed, and an optimal transmission policy for the EH relay is proposed by using a stochastic solar EH model. In this optimal policy, the long-term paired-wise error probability (PEP) of the system is minimized by adapting the EH relay's transmission power to the knowledge of its current battery energy, channel fading status, and causal solar EH information. The designed problem is formulated as a Markov decision process framework, and the conditional capability of the contribution to PEP by the EH relay is adopted as the reward function. We uncover a monotonic and limited difference structure for the expected total discounted reward. Furthermore, a non-conservative property and a monotonic structure of the optimal policy are revealed. Based on the optimal policy and its special structures, the expectation, lower and upper bounds, and asymptotic approximation of the PEP are computed and an interesting result on the system diversity performance is revealed, i.e., the full diversity order can be achieved only if the EH capability index, a metric to quantify the EH node's capability of harvesting and storing energy, approaches to infinity; otherwise, the EH diversity order is only equal to one, and the coding gain of the network is increasing with the EH capability index at this time. Furthermore, a full diversity criterion for the EH TW dual-relay network is proposed. Finally, computer simulations confirm our theoretical analysis and show that our proposed optimal policy outperforms other compared policies.**

***Index Terms*— Stochastic energy harvesting, two-way relay network, paired-wise error probability, diversity order, Markov decision process.**

Manuscript received October 21, 2016; revised February 17, 2017 and May 6, 2017; accepted June 2, 2017. Date of publication June 16, 2017; date of current version September 8, 2017. This work was supported in part by the 111 Project under Project B17008, in part by the National Natural Science Foundation of China under Grant 61672137 and Grant 61401348, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2017JM6099, and in part by the Fundamental Research Funds for the Central University under Grant 310824171004. The work of Y. Chen was supported by the Thousand Youth Talents Program of China. The associate editor coordinating the review of this paper and approving it for publication was H. Suraweera. *(Corresponding author: Wei Li.)*

W. Li is with the School of Information Engineering, Chang'an University, Xi'an 710064, China (e-mail: wli.chd@chd.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 of China, Chengdu 611731, 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).

S. Zhu is with the Department of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: szhu@mail.xjtu.edu.cn).

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/TWC.2017.2715175

I. Introduction

WIRELESS cooperative communications have gained much interest to improve the reliability of wireless links by utilizing the spatial diversity gains in multi-user environments [1]. Regarding with the fact that wireless nodes in cooperative networks are often subject to the limitations of tethering to a fixed power grid or exploiting a large battery with long lifetime, energy harvesting (EH) communication techniques have been introduced for self-sustainable cooperative relays to solve energy supply problems for perpetual lifetime [2].

The typical energy sources in the EH communications are mainly comprised of the renewable energy sources, which represent the energy harvested from ambient environment such as solar, and the electromagnetic(EM) radiation, which is also name wireless power transfer [2]. Although the energy harvested from EM radiation is often controllable and predictable, the EH efficiency of wireless power transfer is strongly affected by the signal propagation distance, and it is much more expensive than other EH schemes in the EH communications. Compared with the EM radiation, the renewable energy sources are green and unexhausted. It is very suitable for the green communications with special requirements, such as low-power, small-size and low-expense. However, considering the stochastic and uncertain energy harvested from ambient environments, energy scheduling and transmission policy problems become more complicated and must be revisited in EH cooperative communication networks.

The EH communication techniques for the classic cooperative network with three nodes have been studied in the literature [3], [4]. Huang *et al.* [3] exploited a deterministic EH model, in which the EH information is noncausal and the energy arrival profile is known prior to transmission scheduling, and proposed an optimal power allocation scheme for the EH source and relay nodes to maximize the throughput. Further, Ku *et al.* [4] utilized a data-driven stochastic EH model [5], [6], wherein the EH information is causal and the model parameters are trained by using a real data record of solar irradiance, and investigated the optimal relay transmission policy to minimize the long-term symbol error rate (SER)

of the system by adopting the Markov decision process (MDP) approach. Moreover, a cellular network with an EH relay was considered in [7], and a joint resource allocation scheme for power, frequency blocks and EH relays was designed. In addition, the ambient EH and the optimal transmission policies for two-hop cooperative communication networks have been investigated in the literature [8]–[10]. Luo *et al.* [8] focused on the scenario wherein the source is an EH node and the non-EH relay is powered by a battery, while a more complicated two-hop network, in which the source and relay are both EH nodes, was investigated in [9]. Aiming at the throughput maximization, an optimal power allocation scheme was cast for the EH source and two EH relays in [10].

Except for the cooperative networks with one-way transmissions, two-way (TW) relay networks and bi-directional transmission protocols have been recognized as promising solutions for information exchange between two source nodes via relay nodes, due to the advantage of high transmission efficiency [11]. Recently, there has been a growing interest in investigating TW relay networks with EH nodes. Tutuncuoglu *et al.* [12] developed a generalized iterative directional water-filling algorithm for various relaying strategies. An EH relay with the ability to cache data was considered and a flexible relay transmission policy was cast in [13]. Regarding with the uncertainty of channel state information (CSI) in TW relay networks, an optimization framework on the relay transmission policy was presented in [14]. With concern for the data-driven stochastic solar EH model [5], Li *et al.* [15], [16] developed optimal relay transmission policies to optimize the long-term transmission performance of the EH TW relay networks by using MDP approaches. In addition, the optimal transmission schemes in the TW relay networks with wireless energy transfer were studied in [17] and [18]. The authors in [17] studied the optimal beamforming for multiple relays to maximize the throughput of the network, wherein the source nodes can harvest energy from the signals of the relays. Moreover, the wireless energy transfer from the source nodes to the relay was also considered in [18]. Li *et al.* [19] proposed the optimal beamforming and transceiver design for a multi-antenna relay, which can transfer the energy and information to two source nodes simultaneously in the TW relay network.

By far, most of the research works on EH TW relay networks utilized deterministic EH models. Compared to the deterministic models, stochastic EH models do not require accurate EH prediction. Specifically, the data-driven stochastic model in [5] conforms much more with realistic situations in the solar-powered communication networks. In [15] and [16], the optimal transmission policies of the EH relay were proposed in the EH TW relay network by exploiting the data-driven stochastic models. However, the system model of the present literature is relatively simple and it is not robust since there exists only one single relay in the network. Further, the practical conditions, such as circuit power consumption and path loss effects, are not considered. Therefore, the optimal transmission policy and the system performance of EH TWR networks with data-driven stochastic EH models have not been well investigated.

Considering the traditional TW single-relay network, in order to improve the system transmission performance in the serious fading channels, there are two traditional possible solutions: one is equipping the relay with multiple antennas, the other is interpolating extra relays. The first solution cannot be suitable for the wireless node, which have the special limitations such as low-power, small-size, and low-expense. Regarding with the second possible solution, it cannot be carried out when the wireless nodes are subject to the limitations of fixed power supply. Hence, under the above practical limitations, deploying an extra EH relay would be one possible efficient solution to achieve the spatial diversity gains for the traditional TW relay network.

Motivated by the aforementioned discussions, we utilize the data-driven stochastic EH model [5] and concentrate on an EH TW dual-relay network, which consists of two non-EH source nodes, one non-EH relay and one EH relay. The non-EH relay node with fixed power supply is deployed to guarantee the reliable communication between the two sources. Besides, we assume the other relay node with a finite-sized battery is solar-powered, and aim at analyzing the impact of the stochastic EH on the reliability performance of the EH TW relay network. In this network, the two relays make use of amplify-and-forward (AF) protocol and space-time network coding (STNC) [20] to fulfill the information exchange between two source nodes. Our objective is to minimize the long-term paired-wise error probability (PEP) of the system by adapting the EH relay's transmission power to the knowledge of its current battery energy, channel states and causal solar EH information. Further, the practical conditions, such as circuit power consumption and path loss compensation, are incorporated in this problem. The main contributions of this paper are summarized as follows:

- First, a space-time transmission protocol is designed for the proposed EH TW dual-relay network by using STNC. The instantaneous PEP of the system with the AF protocol is computed, and it is approximately decomposed into two independent portions, which represent the capabilities of contribution to the PEP by the two relays, respectively.
- Then, we formulate an MDP optimization framework for the EH TW dual-relay network, wherein the Gaussian mixture hidden Markov chain in [5] is exploited as the stochastic EH model, the EH relay's battery capacity is quantized in units of energy quanta, the fading channels between the sources and the EH relay are formulated by a finite-state Markov model [21], and the MDP action represents the EH relay's transmission power. Further, the total power consumed by the EH relay is modeled as the sum of two partitions: transmission power and static power, which includes circuit power and other relatively fixed power consumption. Based on the states and action in the MDP, as well as the static power consumption of the EH relay, the system state transition probability is computed. In addition, the reward function in the MDP is defined as the capability of contribution to the PEP by the EH relay conditioned on preset relay actions and fading channel states. Moreover, we derive the asymptotic approximation of the reward function in high SNRs.

- In the MDP formulation, the utility function is set as the expected long-term total discounted reward, and the optimal transmission policy for the EH relay is induced by the value iteration algorithm. In order to study the structural property of the optimal transmission policy, we first analyze the property of the utility function, and uncover a monotonic and limited difference structure of the expected total discounted reward. Further, a non-conservative property of the optimal policy is revealed, which means the EH relay with the sufficient energy in the battery always forwards the signal in asymptotically high SNRs. Moreover, we uncover a monotonic structure of the optimal policy with respect to the EH relay's battery state.
- Based on the developed optimal transmission policy and its special structures, the expectation, lower and upper bounds, and asymptotic approximation of the PEP are computed for the EH TW network-coded dual-relay network. In addition, we introduce the concept of the EH capability index for EH communication systems, which represents the EH node's capability of harvesting and storing energy. Based on the EH capability index, an interesting result on the EH diversity performance of the network is revealed, i.e., the full diversity order can be achieved only if the EH capability index approaches to infinity; otherwise, the EH diversity order is only equal to one, and the coding gain of the EH network is increasing with the EH capability index at this time. Furthermore, a full diversity criterion is proposed for the EH TW dual-relay network.

The rest of this paper is organized as follows. Section II introduces the system model of the considered EH TW dual-relay network and the space-time transmission protocol using STNC. The MDP formulation and the optimal transmission policy are developed in Section III. Section IV reveals the structural property of the optimal policy. In addition, the PEP performance and the diversity order of the EH TW dual-relay network are analyzed in Section V. Simulation results are demonstrated in Section VI. Finally, Section VII concludes this paper.

II. Energy Harvesting Two-Way Dual-Relay Networks With Network Coding

An EH TW dual-relay network is considered in Fig. 1, where two source nodes, A and B, exchange information simultaneously via two relay nodes, R_1 and R_2 [22]. The source nodes A, B and the relay R_2 are traditional wireless nodes with fixed transmission power, while the relay R_1 equipped with a finite-sized battery can only harvest energy from the solar for data transmission. Since the solar-powered EH communication technology is usually exploited in the wireless sensor networks with special limitations, such as low-power, small-size, and low-expense [2], we assume that each node is equipped with a single antenna and half-duplex. It can be seen that A, B and R_2 nodes constitute the basic TW single-relay network with fixed transmission power; hence, the impact of EH relay nodes on the transmission performance of the basic TW relay network can be investigated by interpolating the EH relay R_1. Due to the randomness and uncertainty of the energy harvested by R_1, transmission policy problems become more crucial in the EH cooperative network.

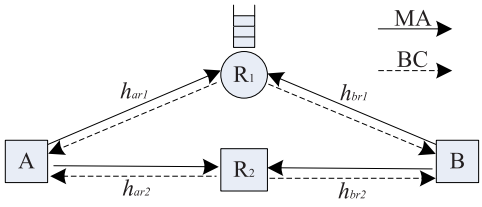

Fig. 1. The EH TW dual-relay network.

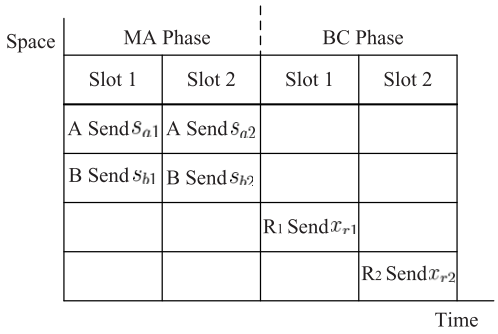

Fig. 2. The space-time transmission protocol for TW dual-relay networks.

It is assumed that the wireless channels are reciprocal, quasi-static and Rayleigh flat fading, and there is no direct link between the two source nodes. All channel fading coefficients are independent and identically distributed (i.i.d.) random variables with complex Gaussian distribution $\mathcal{CN}(0, 1)$. Since we focus on the impact of the EH relay R_1 on the performance of TW relay networks and the CSI assumption of the traditional relay R_2 is unrelated with the main results of the paper, it is assumed that R_1 has the full channel state information (CSI) related to itself, while the relay R_2 only knows the statistical information of the channels. Moreover, every source node can obtain full CSI of all links in the network. According to a two-phase training protocol of the channel estimation in [23], the two sources acquire the CSI of the two channels related to R_2 by one training frame with a length of $3L$, where L represents the number of the training symbols in one training sequence sent by the source node or the relay. Similarly, the two sources and R_1 can all obtain the CSI of the two links associated with R_1 through one training frame with a length of $4L$. The detailed discussions on the training procedure and the determination of L stay out of the scope of this paper.

The relay R_1 and R_2 simultaneously exploit STNC vectors and an AF protocol to help the source nodes exchange information. The whole transmission period is comprised of a multiple access (MA) phase and a broadcast (BC) phase, and each phase consists of two slots. The transmission protocol is illustrated in Fig. 2, and we elaborate on the detailed protocol in the following.

A. MA Phase

In the MA phase, both of the source nodes A and B simultaneously transmit a set of two symbols to the relays in two slots. The received signal vector at R_l ($l \in \{1, 2\}$) in the two slots, is given by

$$\mathbf{y}_{sr_l} = h_{ar_l}\sqrt{P_S}\mathbf{s}_a + h_{br_l}\sqrt{P_S}\mathbf{s}_b + \mathbf{n}_{sr_l}, \quad (1)$$

where h_{ar_l} and h_{br_l} represent the channel fading coefficients between the relay R_l and the two source nodes, P_S denotes the source nodes' transmission power, $\mathbf{s}_a = [s_{a1}, s_{a2}]^{\mathrm{T}}$ and $\mathbf{s}_b = [s_{b1}, s_{b2}]^{\mathrm{T}}$ are defined as the signal vectors transmitted by A and B in the two slots, respectively, $\mathbf{n}_{sr_l}$ is the additive white Gaussian noise (AWGN) vector with i.i.d. entries following $\mathcal{CN}(0, N_0)$ at R_l in the two slots.

B. Relay Processing

After receiving the signals, each relay makes use of linear transformation and the AF protocol on the received signal vector. The transformed and amplified signal on R_l can be written as $x_{r_l} = \beta_l \theta_l^{\mathrm{T}} \mathbf{y}_{sr_l}$, where β_l and $\theta_l = [\theta_{l1}, \theta_{l2}]^{\mathrm{T}}$ are the signal amplifying factor and the STNC vector for R_l, respectively. The vector $\theta_l = [\theta_{l1}, \theta_{l2}]^{\mathrm{T}}$ with unit form meets the full diversity condition of STNC [24]

$$\left|\theta_l^{\mathrm{T}}(\mathbf{s} - \tilde{\mathbf{s}})\right| \neq 0, \forall \mathbf{s}, \tilde{\mathbf{s}} (\neq \mathbf{s}) \in \mathcal{A}_s^2, \quad (2)$$

where $\mathcal{A}_s^2$ represents a Euclidean signal space with two-dimension. Note that it is assumed that only R_1 has the full knowledge of CSI related to itself, and thus R_1 can utilize the variable gain to amplify the received signal such that its average transmission power is limited to P_{R_1} under the CSI [11], [25]. Thus, the amplifying factor β_1 for R_1 is defined as

$$\beta_1 = \sqrt{2P_{R_1} / \left(P_S \gamma_{ar_1} + P_S \gamma_{br_1} + N_0\right)}, \quad (3)$$

where $\gamma_{ar_1} = \left|h_{ar_1}\right|^2$, and $\gamma_{br_1} = \left|h_{br_1}\right|^2$. Note that P_{R_1} changes from time to time due to the randomness of the energy harvested by the relay R_1. Since each relay only transmits the signal at one slot in the BC phase, the constant 2 in (3) is introduced to normalize the total power. Alternatively, since the relay R_2 only knows the statistical information of the channels, it could use a constant amplifying factor such that its power is normalized to P_{R_2} in the long-term fashion. Thus, the β_2 for R_2 is given by

$$\beta_2 = \sqrt{2P_{R_2} / (2P_S + N_0)}. \quad (4)$$

C. BC Phase

In the BC phase, the two relay nodes take turn to transmit the signal in the two slots, i.e., each slot is solely occupied by one relay. After receiving the combined signal, each user node subtracts its self-interference, and thus the received signal at the source A can be expressed as

$$\tilde{y}_{r_l a} = h_{ar_l} h_{br_l} \beta_l \sqrt{P_S} \theta_l^{\mathrm{T}} \mathbf{s}_b + h_{ar_l} \beta_l \theta_l^{\mathrm{T}} \mathbf{n}_{sr_l} + n_{r_l a}, \quad (5)$$

where $n_{r_l a}$ is the AWGN with distribution $\mathcal{CN}(0, N_0)$ at the source A. Based on the signal vector $\left\{\tilde{y}_{r_l a}\right\}_{l=1}^2$ in the two slots, the source A can utilize the maximum likelihood (ML) criterion to jointly decode $\mathbf{s}_b$ as follows [23]:

$$\hat{\mathbf{s}}_b = \arg \min_{\mathbf{s}_b \in \mathcal{A}_s^2} \sum_{l=1}^{2} \frac{\left|\tilde{y}_{r_l a} - h_{ar_l} h_{br_l} \beta_l \sqrt{P_S} \theta_l^{\mathrm{T}} \mathbf{s}_b\right|^2}{(\gamma_{ar_l} \beta_l^2 + 1) N_0}. \quad (6)$$

Thus, the instantaneous PEP at the source node A conditioned on $\left\{\gamma_{ar_l}\right\}_{l=1}^2$ and $\left\{\gamma_{br_l}\right\}_{l=1}^2$ can be calculated by using the Q-function $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt$ as follows [1]

$$\tilde{\Pr}\left\{\mathbf{s}_b \to \tilde{\mathbf{s}}_b | \, \gamma_{ar_l}, \gamma_{br_l}, l \in \{1, 2\}\right\} = Q\left(\sqrt{2W_{R_1} + 2W_{R_2}}\right), \quad (7)$$

where $\forall \mathbf{s_b}, \tilde{\mathbf{s}}_b (\neq \mathbf{s_b}) \in \mathcal{A}_s^2$, and

$$W_{R_l} = \frac{\gamma_{ar_l} \gamma_{br_l} \beta_l^2 P_s \left|\theta_l^{\mathrm{T}} (\mathbf{s}_b - \tilde{\mathbf{s}}_b)\right|^2}{4(\gamma_{ar_l} \beta_l^2 + 1) N_0}, l \in \{1, 2\}, \quad (8)$$

which is related to the received SNR. From (3), (4), and (8), it can be derived that

$$W_{R_1} = \frac{\gamma_{ar_1} \gamma_{br_1} P_{R_1} P_S \alpha_1}{2\left[\left(P_S + 2P_{R_1}\right) \gamma_{ar_1} + P_S \gamma_{br_1} + N_0\right] N_0}, \quad (9)$$

$$W_{R_2} = \frac{\gamma_{ar_2} \gamma_{br_2} P_{R_2} P_S \alpha_2}{2\left[2\left(P_S + P_{R_2} \gamma_{ar_2}\right) + N_0\right] N_0}, \quad (10)$$

where $\alpha_l = \left|\theta_l^{\mathrm{T}} (\mathbf{s}_b - \tilde{\mathbf{s}}_b)\right|^2$. According to the full diversity condition of STNC in (2), we have $\alpha_l > 0$.

By applying the equality $Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta$, the instantaneous PEP is reformulated as

$$\tilde{\Pr}\left\{\mathbf{s}_b \to \tilde{\mathbf{s}}_b \mid \gamma_{ar_l}, \gamma_{br_l}, l \in \{1, 2\}\right\} = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{W_{R_1} + W_{R_2}}{\sin^2\theta}\right) d\theta \leq P_{e,R_1} \times P_{e,R_2}, \quad (11)$$

where $P_{e,R_1} = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{W_{R_1}}{\sin^2\theta}\right) d\theta$ and $P_{e,R_2} = \exp\left(-W_{R_2}\right)$.

Thus, P_{e,R_1} and P_{e,R_2} represent the capacities of contribution to the PEP by R_1 and R_2, respectively. Due to the network symmetry, the instantaneous PEP of the source B is computed similarly. Since the solar power harvested by R_1 is random and uncertain, the transmission performance of the EH TW dual-relay network is affected by the transmission policy of R_1, which is investigated in the following.

III. Markov Decision Process Formulation

In this section, our goal is to find the optimal transmission power of the EH relay R_1, so as to minimize the long-term PEP by jointly considering the causal solar irradiance condition, the available energy in the relay's battery and the fading channel status. Therefore, the design framework is formulated as an MDP model, which is mainly composed of the state space, the action space, the state transition probability and the reward function. In the following, we elaborate on these MDP elements.

A. System States and Action

Let $\mathcal{S} = \mathcal{S}_E \times \mathcal{S}_{AR_1} \times \mathcal{S}_{BR_1} \times \mathcal{S}_B$ be a four-tuple state space [22], where $\times$ denotes the Cartesian product, $\mathcal{S}_E = \{0, 1, \cdots, N_e - 1\}$ represents the solar EH state subspace, $\mathcal{S}_{AR_1} = \{0, 1, \cdots, N_c - 1\}$ and $\mathcal{S}_{BR_1} = \{0, 1, \cdots, N_c - 1\}$ denote the two channel state subspaces for the A-R_1 link and the B-R_1 link, respectively, and $\mathcal{S}_B = \{0, 1, \cdots, N_b - 1\}$ is the battery state subspace for R_1. Meanwhile, let $S = \left(S_E, S_{AR_1}, S_{BR_1}, S_B\right) \in \mathcal{S}$ stand for the stochastic state of the MDP. In addition, define $\mathcal{A} = \left\{0, 1, \cdots, N_p - 1\right\}$ as the action space, and the action $a \in \mathcal{A}$ represents the transmission power index of R_1. All the states and action remain unchanged during one policy period T, which covers a number of two-phase transmissions.

1) Solar EH State: An N_e-state stochastic EH model in [5] is utilized to capture the dynamic of the harvested solar EH conditions. This data-driven solar EH model is a Gaussian mixture hidden Markov chain, and its underlying parameters are extracted using the solar irradiance data collected by a solar site [5]. Since the solar irradiance data were measured from the early morning (seven oclock) to the late afternoon (seventeen oclock) every day, the solar EH model with its underlying parameters in [5] is applied to the scenario of daylight. Therein, it is assumed that if the solar EH state is given by $S_E = e \in \mathcal{S}_E$, the harvested solar power per unit area, P_h, is a continuous random variable with Gaussian distribution $\mathcal{N}(\mu_e, \rho_e)$. Therefore, the harvested solar energy $E_h = P_h T \Omega \sigma$ during one policy period T is also continuous, where Ω is the solar panel area size, and σ denotes the energy conversion efficiency. Moreover, different solar EH states result in different solar irradiance intensities, and the dynamic of the states is governed by a state transition probability $P\left(S_E = e' | S_E = e\right), \forall e, e' \in \mathcal{S}_E$.

In practice, the solar energy storage and usage model is discrete and designed based on the numbers of energy quanta. Thus, the harvested energy is first quantized in units of one basic energy quantum E_U and then stored in the battery for the forthcoming data transmission. Accordingly, the probability of the number of harvested energy quanta conditioned on the e^{th} solar EH state, denoted as $P\left(Q = q | S_E = e\right)$ for $q \in \{0, 1, \cdots, \infty\}$, is theoretically derived in [5], which represents the impact of the parameters of the solar EH state and the energy storage system on the energy supporting condition.

2) Battery State: The EH relay R_1 is equipped with a finite-sized battery, which is evenly quantized into N_b levels. The battery state represents the number of available energy quanta in the battery, i.e., the available energy is $b \cdot E_U$ if the battery state is equal to $b \in \mathcal{S}_B$.

3) Channel States: The wireless channel variation from one level to another is formulated by an N_c-state Markov chain [21]. The instantaneous channel gains of the wireless channels related to R_1, γ_{ar_1} and γ_{br_1}, are quantized into N_c levels using a couple of thresholds, given by $\Gamma = \left\{0 = \Gamma_0, \Gamma_1, \cdots, \Gamma_{N_c} = \infty\right\}$. The i^{th} channel state means that the corresponding channel gain belongs to the interval $\left[\Gamma_i, \Gamma_{i+1}\right)$. Further, it is assumed that the wireless channel can only transit from the current state to its neighboring states under a slow fading assumption [21]. Since one policy period T covers a number of two-phase transmissions, there are enough symbols for the wireless nodes to obtain the full CSI and channel states.

4) EH Relay Action: The action denotes the transmission power of the EH relay R_1. If the action $a \in \mathcal{A}$ is chosen, the transmission power of R_1 is set as $P_{R_1} = a P_U$, where P_U is a constant transmission power corresponding to one energy quantum E_U in the battery during the half transmission period, i.e., $E_U = \frac{1}{2} P_U T$. Specifically, R_1 keeps silent if $a = 0$.

5) MDP State Transition: The total power consumption of the EH relay R_1 can be modeled as the sum of two partitions: static power consumption and dynamic power consumption [26]. The dynamic partition represents the transmission power of R_1, and it is formulated as the EH relay action in the MDP. The static partition mainly includes the circuit power consumption, the energy consumption for obtaining CSI, the power consumption due to path loss, etc. In one policy period T, the energy consumption for the static power is modeled as ρE_U, where the constant ρ is defined as the static power consumption index and it ranges in $\mathcal{S}_B$.

Remark 1: The EH relay consumes no static power and keeps asleep, when the energy in the battery is deficient, i.e., $b \leq \rho$, in one policy period. Thus, at this time the EH relay would keep silent, i.e., the action $a = 0$. For simplicity, we specify that the battery is in the deficient state when $b \leq \rho$.

Since we utilize the *harvest-store-use* protocol [27], the harvested energy in the current policy period is first stored in the battery, and then consumed for data transmission in the next policy period. Therefore, the relationship between the current state b and the next state b' can be given as

$$b' = \min\left(b - a - \kappa\rho + q,\ N_b - 1\right), \tag{12}$$

when the number of harvested energy quanta is equal to q, the action a is taken in the current period, and the parameter κ is defined as $\kappa = U(b - \rho)$, where $U(\cdot)$ denotes the unit step function. Note that it implies that $b \geq a + \kappa\rho$ due to the battery constraint. Further, the battery state transition probability at the e^{th} solar EH state can be given by [5]

$$P_a\left(S_B = b' | S_B = b, S_E = e\right) = \begin{cases} P\left(Q = b' - b + a + \kappa\rho | S_E = e\right), \\ b' = (b - a - \kappa\rho), \cdots, N_b - 2; \\ 1 - \sum\limits_{q=0}^{N_b - 2 - b + a + \kappa\rho} P\left(Q = q | S_E = e\right), \ b' = N_b - 1. \end{cases} \tag{13}$$

Since the transition probabilities of the solar EH states and channel states are independent, the system state transition probability from $s = (e, f, g, b)$ to $s' = \left(e', f', g', b'\right)$ for the action a is expressed as (14), as shown at the bottom of this page,

$$P_a\left(s' | s\right) = P\left(S_E = e' | S_E = e\right) \cdot P\left(S_{AR_1} = f' | S_{AR_1} = f\right) \cdot P\left(S_{BR_1} = g' | S_{BR_1} = g\right) \cdot P_a\left(S_B = b' | S_B = b, S_E = e\right) \tag{14}$$

where the solar EH state transition and the channel state transition are defined in [5] and [21], respectively.

B. Reward Function

According to (11), the PEP is determined by both P_{e,R_1} and P_{e,R_2}. However, only the term P_{e,R_1} is related with the channel states and the power action of R_1 in the MDP. Thus, the reward function is defined as the conditional P_{e,R_1} with respect to the state $s = (e, f, g, b)$ and the action a as follows:

$$R_a(s) = \frac{\frac{1}{\pi}\int_{\Gamma_g}^{\Gamma_{g+1}}\int_{\Gamma_f}^{\Gamma_{f+1}}\int_0^{\pi/2} e^{-\gamma_1}e^{-\gamma_2}\cdot\exp\left(-\frac{W_{R_1}}{\sin^2\theta}\right)d\theta d\gamma_1 d\gamma_2}{P\left(S_{AR_1}=f\right)\cdot P\left(S_{BR_1}=g\right)} \triangleq P_{e,R_1}(a,f,g), \quad (15)$$

where $\eta = \frac{P_U}{N_0}$, $\gamma_1 = \gamma_{ar_1}$, and $\gamma_2 = \gamma_{br_1}$.

Remark 2: From (15) and (9), when the relay remains silent, the reward function is equal to one, i.e., $R_{a=0}(s) = P_{e,R_1}(a=0, f, g) = 1$.

For simplicity, we assume that the transmission power of the two source nodes and the relay R_2 are set as $P_S = MP_U$ and $P_{R_2} = P_U$ respectively, where the integer M denotes the transmission power index of the source nodes. Therefore, from (9), W_{R_1} in (15) can be reformulated as

$$W_{R_1} = \gamma_1\gamma_2 aM\eta^2\alpha_1/\left\{2\left[(2a+M)\,\eta\gamma_1 + M\eta\gamma_2 + 1\right]\right\}. \quad (16)$$

Since the integral in (15) cannot be calculated in a closed form, the following proposition is introduced to compute the asymptotic approximation of the reward function.

Proposition 1: In asymptotically high SNRs, i.e., $\eta \gg 1$, the reward function $R_a(s = (e, f, g, b))$ when $a \in \mathcal{A}\backslash\{0\}$ can be approximated as

$$R_a(s) \approx \begin{cases} 0, & \min(f,g) \geq 1; \\ \delta / a, & \min(f,g) = 0, \end{cases} \quad (17)$$

where $\delta = \frac{4\eta^{-1}}{\alpha_1(1-e^{-1}1)} \ll 1$, the expression $\min(f, g)$ represents the minimum value of the two channel states and it can denote the different fading levels of the two channels.

Proof: See Appendix A for details. ∎

C. Optimization of EH Relay Transmission Policy

Define a policy $\pi(s) : \mathcal{S} \to \mathcal{A}$ as the transmission power action of the EH relay R_1. The policy value in the MDP is defined as the expected long-term total discounted reward as follows [28]:

$$V_\pi(s_0) = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty}\lambda^k R_{\pi(s_k)}(s_k)\right], s_k \in \mathcal{S}, \pi(s_k) \in \mathcal{A}, \quad (18)$$

where s_0 is the initial state, and $0 < \lambda < 1$ denotes a discount factor.

The goal of the MDP is to find the optimal policy $\pi^*(s)$ that minimizes the expected total discounted reward under the state $s \in \mathcal{S}$. The optimal policy can be found by using the Bellman equation as follows:

$$V_{\pi^*}(s) = \min_{a\in A}\left(R_a(s) + \lambda\sum\nolimits_{s'\in\mathcal{S}} P_a\left(s'|s\right)V_{\pi^*}\left(s'\right)\right), \quad (19)$$

which can be solved by the value iteration approach [28]:

$$\begin{cases} V_a^{(i+1)}(s) = R_a(s) + \lambda\sum_{s'\in S} P_a\left(s'|s\right)V^{(i)}\left(s'\right), \\ V^{(i+1)}(s) = \min\limits_{a\in\mathcal{A}}\left(V_a^{(i+1)}(s)\right). \end{cases} \quad (20)$$

For simplicity, and from (13) and (14), the summation term in (20) can be written as

$$\begin{aligned} &\sum\nolimits_{s'\in S} P_a\left(s'|s\right)V^{(i)}\left(s'\right) \\ &= \sum\nolimits_{e',f',g'} P\left(S_E = e'|S_E = e\right) \\ &\quad\cdot P\left(S_{AR_1} = f'|S_{AR_1} = f\right)P\left(S_{BR_1} = g'|S_{BR_1} = g\right) \\ &\sum_{q=0}^{\infty} P(Q=q|S_E=e)\cdot V^{(i)}\left(e',f',g',\min(b-a-\kappa\rho+q, N_b-1)\right) \\ &= \mathbb{E}_s\left\{V^{(i)}\left(e',f',g',\min\left(b-a-\kappa\rho+q, N_b-1\right)\right)\right\}, \end{aligned} \quad (21)$$

where $\mathbb{E}_s\{\cdot\}$ denotes the expected value with respect to $s' = \left(e', f', g', b'\right)$ conditioned on the state $s = (e, f, g, b)$.

Except for the expected long-term total discounted reward in (18), the long-term average reward can be also adopted as the policy value in the MDP as follows [28]:

$$\bar{V}_\pi(s_0) = \limsup_{N\to\infty}\frac{1}{N}\sum\nolimits_{k=0}^{N-1} R_{\pi(s_k)}(s_k). \quad (22)$$

In Section VI, we demonstrate that the two kinds of optimal policies corresponding to these two policy values are nearly equivalent, i.e., the average reward can be closely approximately optimized by utilizing the optimal policy for the case of the expected total discounted reward with λ=0.99 in MDP.

IV. Structures of Optimal Transmission Policy

In the previous section, the MDP formulation and the optimization of the EH relay's transmission policy are described. In the following, we discuss the structural properties of the optimal policy.

Lemma 1: For any fixed system state $s = (e, f, g, b > 0) \in \mathcal{S}$ 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., $1 \geq V^{(i)}(e, f, g, b-1) - V^{(i)}(e, f, g, b) \geq 0$. Specifically, in asymptotically high SNR regimes, i.e., $\eta \gg 1$, this difference value is limited to $1-\delta$ under the channel weak condition $\min(f, g) = 0$, i.e., $(1-\delta) \geq V^{(i)}(e, f, g, b-1) - V^{(i)}(e, f, g, b) \geq 0$ under the condition $\min(f, g) = 0$.

Proof: See Appendix B for details. ∎

Lemma 1 describes the monotonic and limited difference structure of the expected long-term total discounted reward, based on which a non-conservative property of the EH relay's optimal transmission policy is provided as follows.

Theorem 1: For any fixed system state $s = (e, f, g, b > \rho) \in \mathcal{S}$ with the sufficient battery, in asymptotically high SNR regimes, i.e., $\eta \gg 1$, the EH relay's optimal power action $a^* \in \mathcal{A}$ is non-zero.

Proof: From (20) and (21), the difference value of the two expected total discounted rewards for $a = 1$ and $a = 0$ in

the $(i+1)^{th}(i \geq 0)$ iteration under the state s can be written as

$$V_{a=1}^{(i+1)}(s) - V_{a=0}^{(i+1)}(s) = R_{a=1}(f,g) - R_{a=0}(f,g) + \lambda \cdot \mathbb{E}_s \left\{ V^{(i)}\left(e', f', g', \min\left(b-1-\rho+q, N_b-1\right)\right) - V^{(i)}\left(e', f', g', \min\left(b-\rho+q, N_b-1\right)\right)\right\}. \quad (23)$$

According to Proposition 1 and Remark 2, the difference values of the two rewards in (23), i.e., $R_{a=1}(f,g) - R_{a=0}(f,g)$, is approximated to -1 and $\delta - 1$ at asymptotically high SNRs under the conditions $\min(f,g) \geq 1$ and $\min(f,g) = 0$, respectively. Further, by applying Lemma 1, and due to the fact that $0 < \lambda < 1$, under any channel conditions we can obtain the inequality $V_{a=1}^{(i+1)}(s) < V_{a=0}^{(i+1)}(s)$. Thus, from (20), it suffices to prove that the optimal action is non-zero in the $(i+1)^{th}$ iteration. When the value iteration algorithm is converged, the EH relay's optimal power action a^* is also non-zero. ■

The intuition behind this theorem is explained as follows. In asymptotically high SNR regimes, the immediate rewards for non-zero actions approach to zero. However, if the zero action, which can be named a conservative policy, is adopted, the EH relay keeps silent currently and conserves the battery energy to improve the future reward, but it may suffer from an energy overflow problem due to a finite-sized battery constraint. As compared with the zero action, the non-zero actions can be seen as non-conservative policies and lead to better policy values. Further, the conservative property of the optimal policy can help us to analyze the reliability performance and the degree of freedom in the proposed EH cooperative network. Beside the conservative property, a monotonic structure of the proposed optimal policy is introduced in the following theorem.

Theorem 2: The EH relay's optimal power action $a^* \in \mathcal{A}$ is non-decreasing in its battery state $b \in \mathcal{S}_B$, if the following EH condition holds:

$$\sum_{q=\max(k-b^-+a^+-\kappa\rho,0)}^{\max(k-b^-+a^+-\kappa\rho,1)-1} P(Q=q|S_E=e) \geq \sum_{q=\max(k-b^++a^--\kappa\rho,0)}^{\max(k-b^++a^--\kappa\rho,1)-1} P(Q=q|S_E=e), \quad (24)$$

where $\forall e \in \mathcal{S}_E$, $\forall k \in \mathcal{S}_B$, $\forall b^+ \geq b^- \in \mathcal{S}_B$, and $\forall a^+ \geq a^- \in \mathcal{A}$.

Proof: See Appendix C for details. ■

This theorem shows that the optimal policy tends to be monotonic when the probability of harvesting higher number of energy quanta is large such that the battery can be recharged quickly. For instance, the optimal policy has the monotonic structure if the probability $P(Q=q|S_E=e)$ is increasing in the number of harvested energy quanta q . Fig. 3 demonstrates the monotonic structure of the proposed optimal relay policy π^* with respect to the battery state. This monotonic structure can enable efficient computation for obtaining the optimal policy, and can make implementation easy for decision makers.

V. Performance Analysis

Our goal is to study the reliable transmission performance of the EH TW network-coded dual-relay network, and thus the long-term PEP performance of the network with the proposed optimal transmission policy is analyzed in this section.

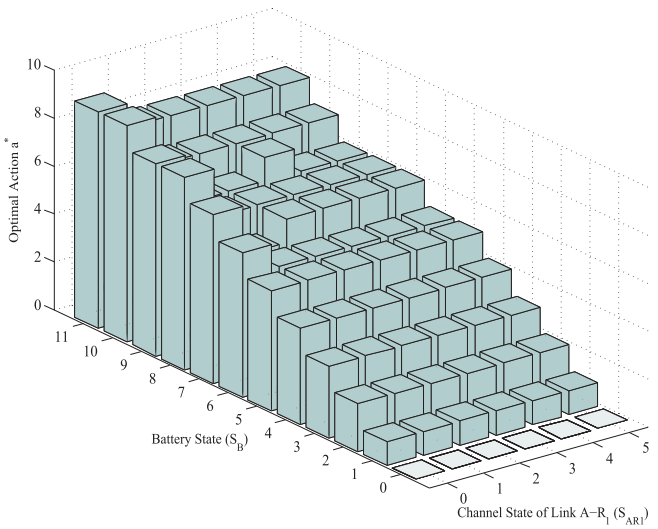

Fig. 3. Monotonic structure of optimal action with respect to battery state.

A. Expected PEP

From (11), the expected PEP can be computed as

$$\Pr\{\mathbf{s}_b \to \tilde{\mathbf{s}}_b\} \leq \mathbb{E}_\pi\left\{P_{e,R_1}\right\} \times \mathbb{E}\left\{P_{e,R_2}\right\}, \quad (25)$$

where $\mathbb{E}_\pi\left\{P_{e,R_1}\right\}$ represents the expected reward conditioned on a fixed policy π, and $\mathbb{E}\left\{P_{e,R_2}\right\}$ is the expectation result of P_{e,R_2} with respect to the channel gains, γ_{ar_2} and γ_{br_2}.

The expected reward $\mathbb{E}_\pi\left\{P_{e,R_1}\right\}$ can be calculated by taking expectation over the reward function as follows [16]:

$$\mathbb{E}_\pi\left\{P_{e,R_1}\right\} = \bar{R}_\pi = \sum\nolimits_{s\in\mathcal{S}} p_\pi(s) \times R_{a=\pi(s)}(s), \quad (26)$$

where $p_\pi(s)$ is defined as the steady state probability for π, and can be computed by using the state transition probability in (14) and the balance equation in [16]. Since the states of the Markov chain are assumed to be recurrent, the occurrence probability of the state s is equal to $p_\pi(s)$ after a long time, and thus the expected reward $\mathbb{E}_\pi\left\{P_{e,R_1}\right\}$ can be regarded as the long-term average reward in the MDP.

From (11), the expected capacity of contribution to the PEP by the relay R_2 is computed as

$$\mathbb{E}\left\{P_{e,R_2}\right\} = \int_0^\infty \int_0^\infty e^{-\gamma_3} e^{-\gamma_4} \cdot \exp\left(-W_{R_2}\right) d\gamma_3 d\gamma_4, \quad (27)$$

where $\gamma_3 = \gamma_{ar_2}$, and $\gamma_4 = \gamma_{br_2}$. From (10), W_{R_2} can be rewritten as

$$W_{R_2} = \gamma_3\gamma_4 M\eta^2\alpha_2 / \left\{2\left[2\left(M+\gamma_3\right)\eta + 1\right]\right\}. \quad (28)$$

Since the integral in (27) cannot be carried out, the bounds of $\mathbb{E}\left\{P_{e,R_2}\right\}$ are derived. First, we introduce the following proposition.

Proposition 2: Let $Z = \frac{c_1 XY}{c_2 X + c_3}$, where X and Y are i.i.d. exponential random variables with unit mean, and c_1, c_2 and

c_3 are positive constants. $\mathbb{E}\left\{e^{-Z}\right\}$ can be bounded by

$$\frac{c_2}{c_1+c_2}+\frac{c_1c_3}{2(c_1+c_2)^2}\ln\left(1+2\frac{c_1+c_2}{c_3}\right)<\mathbb{E}\left\{e^{-Z}\right\}<\frac{c_2}{c_1+c_2}+\frac{c_1c_3}{(c_1+c_2)^2}\ln\left(1+\frac{c_1+c_2}{c_3}\right). \quad (29)$$

Proof: See Appendix D for details. ∎

Corollary 1: The upper and lower bounds of $\mathbb{E}\left\{P_{e,R_2}\right\}$ can be computed as (30) and (31), respectively.

Proof: According to Proposition 2, (27) and (28), if the random variables X and Y are set as γ_3 and γ_4 respectively, it can be seen that $c_1 = M\eta^2\alpha_2$, $c_2 = 4\eta$, $c_3 = 4M\eta + 2$. Therefore, by applying (29), the upper and lower bounds of $\mathbb{E}\left\{P_{e,R_2}\right\}$ are given by (30) and (31), as shown at the bottom of this page, respectively. ∎

B. Asymptotic Approximation of PEP

In the literature [9], [27], [29], the asymptotic studies of the transmission performances are carried out to show the rate of decay at high SNRs, and high SNRs can be achieved via decreasing the noise power N_0. In this subsection, we calculate the asymptotic approximation of the expected PEP in high SNR regimes, so as to analyze the diversity order of the EH TW network-coded dual-relay network with the proposed optimal policy π^*. First, the expected reward in (26) for π^* can be rewritten by considering the battery states as follows

$$\bar{R}_{\pi^*}=\sum\nolimits_{s\in\mathcal{S}}\left[p_{\pi^*}(e,f,g,b\le\rho)\times R_{a^*}(e,f,g,b\le\rho)+p_{\pi^*}(e,f,g,b>\rho)\times R_{a^*}(e,f,g,b>\rho)\right]. \quad (32)$$

From Remark 1, the EH relay R_1 keeps silent when the battery is in the deficient state. Meanwhile, according to Theorem 1, the optimal action a^* is non-zero in asymptotically high SNRs when the battery state b is larger than ρ. Further, by applying Proposition 1 and Remark 2, the asymptotic approximation of the expected reward in asymptotically high SNRs for the optimal policy π^* can be calculated as

$$\bar{R}_{\pi^*}\approx P_{\pi^*}\{b\le\rho\}+\frac{4P_{\pi^*}\{\min(f,g)=0,b>\rho\}}{a^*\left(1-e^{-\Gamma_1}\right)\alpha_1}\eta^{-1}, \quad (33)$$

where the battery deficient probability is defined as

$$P_{\pi^*}\{b\le\rho\}=\sum\nolimits_{s\in\mathcal{S}}p_{\pi^*}(e,f,g,b\le\rho), \quad (34)$$

and the battery sufficient and channel weak probability is denotes as

$$P_{\pi^*}\{\min(f,g)=0,b>\rho\}=\sum\nolimits_{s\in\mathcal{S}}p_{\pi^*}(e,f,g,b>\rho|\min(f,g)=0). \quad (35)$$

In addition, we obtain the bounds of $\mathbb{E}\left\{P_{e,R_2}\right\}$ in Corollary 1, and thus in asymptotically high SNRs, i.e., $\eta\gg1$, $\mathbb{E}\left\{P_{e,R_2}\right\}$ can be bounded as

$$\frac{2}{M\alpha_2}\ln\left(1+\frac{\eta}{2}\right)\eta^{-1}\le\mathbb{E}\left\{P_{e,R_2}\right\}\le\frac{4}{M\alpha_2}\ln\left(1+\frac{\eta}{4}\right)\eta^{-1}, \quad (36)$$

where the first terms in $\mathbb{E}\left\{P_{e,R_2}\right\}^{(\text{up})}$ and $\mathbb{E}\left\{P_{e,R_2}\right\}^{(\text{low})}$ are omitted.

Therefore, from (25), (33) and (36), the upper and lower bounds of the PEP in asymptotically high SNRs can be expressed as

$$\Pr\{\mathbf{s}_b\to\tilde{\mathbf{s}}_b\}\le\frac{4\ln\left(1+\frac{\eta}{4}\right)\cdot P_{\pi^*}\{b\le\rho\}}{M\alpha_2}\eta^{-1}+\frac{16\ln\left(1+\frac{\eta}{4}\right)\cdot P_{\pi^*}\{\min(f,g)=0,b>\rho\}}{Ma^*\left(1-e^{-\Gamma_1}\right)\alpha_1\alpha_2}\eta^{-2} \quad (37)$$

$$\Pr\{\mathbf{s}_b\to\tilde{\mathbf{s}}_b\}\ge\frac{2\ln\left(1+\frac{\eta}{2}\right)\cdot P_{\pi^*}\{b\le\rho\}}{M\alpha_2}\eta^{-1}+\frac{8\ln\left(1+\frac{\eta}{2}\right)\cdot P_{\pi^*}\{\min(f,g)=0,b>\rho\}}{Ma^*\left(1-e^{-\Gamma_1}\right)\alpha_1\alpha_2}\eta^{-2}. \quad (38)$$

C. EH Capability Index and EH Diversity Order

After obtaining the asymptotic approximation of the PEP, we can investigate the EH diversity order of the EH TW dual-relay network with the proposed optimal transmission policy, which is an important performance metric for the network with stochastic EH. First of all, we introduce the concept of EH capability index.

Definition 1: The EH capability index $C_{EH,\pi}$ is defined as the inverse of the battery deficient probability associated with a fixed policy π in an EH communication system, i.e., $C_{EH,\pi}=\frac{1}{P_\pi\{b\le\rho\}}$.

The EH capability index $C_{EH,\pi}$, ranging from 1 to infinity, represents the EH node's capability of harvesting and storing energy, and it is affected by the solar panel size $\mathbf{\Omega}$, the size of one energy quantum E_U, the battery capacity and the transmission policy π. Specifically, $C_{EH,\pi}=\infty$ means that the EH node's battery cannot be in the deficient state under any conditions.

In the proposed EH cooperative network, our goal is to investigate the impact of the EH relay on the reliability performance of the basic TW single-relay network. Therefore, we adopt the diversity order as one important metric. From (25), the diversity order of the network is defined as $d=-\lim_{\eta\to\infty}\frac{\log\Pr\{\mathbf{s}_b\to\tilde{\mathbf{s}}_b\}}{\log\eta}$. Since the spatial diversity in the

$$\mathbb{E}\left\{P_{e,R_2}\right\}^{(\text{up})}=\frac{4}{M\eta\alpha_2+4}+\frac{4M^2\eta\alpha_2+2M\alpha_2}{(M\eta\alpha_2+4)^2}\ln\left(1+\frac{M\eta^2\alpha_2+4\eta}{4M\eta+2}\right) \quad (30)$$

$$\mathbb{E}\left\{P_{e,R_2}\right\}^{(\text{low})}=\frac{4}{M\eta\alpha_2+4}+\frac{4M^2\eta\alpha_2+2M\alpha_2}{2(M\eta\alpha_2+4)^2}\ln\left(1+2\frac{M\eta^2\alpha_2+4\eta}{4M\eta+2}\right) \quad (31)$$

proposed EH TW relay network is related with the energy harvested from ambient environments, the diversity order of the network can be named EH diversity order. In the subsequence, the EH diversity order of the network is discussed based on $C_{EH,\pi}$.

Theorem 3: In the proposed EH TW network-coded dual-relay network with the optimal policy π^*, if the EH relay's EH capability index C_{EH,π^*} approaches to infinity, the EH diversity order $d = 2$ is obtained, i.e., the full diversity order can be achieved. Otherwise, the EH diversity order is only equal to $d = 1$, and the coding gain is bounded by $\frac{M\alpha_2 C_{EH,\pi^*}}{4\ln\left(1+\frac{\eta}{4}\right)} \leq g_E \leq \frac{M\alpha_2 C_{EH,\pi^*}}{2\ln\left(1+\frac{\eta}{2}\right)}$. Thus, the higher C_{EH,π^*} or larger M determine the better PEP performance.

Proof: If the EH relay's EH capability index C_{EH,π^*} approaches to infinity, i.e., the battery deficient probability $P_{\pi^*}\{b \leq \rho\}$ is equal to zero, the asymptotic approximations of the PEP in high SNRs are dominated by the second terms in (37) and (38), leading to the EH diversity order $d = 2$. Otherwise, the asymptotic approximations are dominated by the first terms, and thus only the EH diversity order of $d = 1$ can be achieved. From (37) and (38), when the EH diversity order is equal to $d = 1$, it is straightforward to show that the coding gain can be bounded by $\frac{M\alpha_2}{4\ln\left(1+\frac{\eta}{4}\right)\cdot P_{\pi^*}\{b\leq\rho\}} \leq g_E \leq \frac{M\alpha_2}{2\ln\left(1+\frac{\eta}{2}\right)\cdot P_{\pi^*}\{b\leq\rho\}}$. Therefore, the PEP performance is monotonically decreasing and increasing in $P_{\pi^*}\{b \leq \rho\}$ and C_{EH,π^*} respectively, and it is improved by increasing the transmission power index of the sources M. ■

This theorem reveals the diversity performance of the EH TW network-coded dual-relay network in high SNR regimes. In fact, in order to improve the reliable transmission performance and achieve the full diversity order, it requires an infinite EH capability index in the system. Therefore, the full diversity criterion that guarantees to obtain the full diversity order in the proposed EH TW relay network is discussed in the following. Above all, we present the definition of the energy deficiency probability according to [5] as follows.

Definition 2: The energy deficiency probability in an EH communication system is defined as the probability when the number of harvested energy quanta is not larger than the static power consumption index ρ, conditioned on the solar EH state $S_E = e \in \mathcal{S}_e$, i.e., $P(Q \leq \rho | S_E = e)$.

Theorem 4 (Full Diversity Criterion): In the proposed EH TW network-coded dual-relay network with the optimal policy π^*, the full diversity order $d = 2$ can be achieved, if and only if the energy deficiency probability is equal to zero, i.e., $P(Q \leq \rho | S_E = e) = 0$, $\forall e \in \mathcal{S}_E$.

Proof: According to Theorem 3 and Definition 1, this theorem can be proved by showing that the EH capability index C_{EH,π^*} approaches to infinity, i.e., the battery deficient probability $P_{\pi^*}\{b \leq \rho\}$ is equal to zero, if and only if $P(Q \leq \rho | S_E = e) = 0$, $\forall e \in \mathcal{S}_E$. Further, from (14), $P_{\pi^*}\{b \leq \rho\} = 0$ is related to the battery state transition condition as follows

$$P_{a^*}(S_B \leq \rho | S_B = b, S_E = e) = 0, \forall b \in \mathcal{S}_B, \forall e \in \mathcal{S}_E. \quad (39)$$

From (13), the condition in (39) implies that the number of harvested energy quanta should not be equal to $\rho - b + a^* + \kappa\rho$, i.e.,

$$P\left(Q = \rho - b + a^* + \kappa\rho | S_E = e\right) = 0, \forall b \in \mathcal{S}_B, \forall e \in \mathcal{S}_E. \quad (40)$$

According to the proposed energy storage and usage model in (12), we have $a^* + \kappa\rho \leq b$. Therefore, the condition in (40) is equivalent to $P(Q \leq \rho | S_E = e) = 0$, $\forall e \in \mathcal{S}_E$. Thus, we complete the proof of Theorem 4. ■

TABLE I
SIMULATION PARAMETERS

Policy management period (T)	300s
Modulation type	QPSK
Basic transmission power (P_U)	10mW
Transmission power index (M)	1
Transmission power of the relay R_2 (P_{R_2})	P_U
Solar panel size (Ω)	1cm^2
Energy conversion efficiency (σ)	20%
Channel simulation model	Jakes' model [30]
Normalized Doppler frequency (f_D)	0.05
Channel quantization thresholds (Γ)	$\{0, 0.3, 0.6, 1.0, 2.0, 3.0, \infty\}$
Discount factor (λ)	0.99
Circuit power consumption index (ρ)	0

VI. SIMULATION RESULTS

The long-term PEP of the proposed EH TW network-coded dual-relay network with the optimal transmission policy based on the stochastic EH model in [5] is evaluated by computer simulations. In the simulations, the numerical results are calculated by applying Monte-Carlo approach. The numbers of the solar EH states, channel states and battery states are four, six and twelve, respectively, as well as the main simulation parameters are listed in Table I, except as stated otherwise. The STNC matrix can be designed as a 2×2 Vandemonde matrix, whose entries are defined as $\theta_{ln} = e^{j\pi(4l-1)(n-1)/4}$, $l, n = 1, 2$ [24]. In the following simulations, a normalized average SNR is defined with respect to the transmission power of 1 mW, since the average transmission power of the EH relay is unknown and depends on real solar irradiance. In addition, the legend "Dual-relaying" stands for the proposed EH TW network-coded dual-relay network in Fig. 1, while the legend "Single-Relaying" represents the traditional TW single-relay network consisting of A, B and R_2 in Fig. 1, which is exploited to make comparison with the EH TW dual-relay network.

The expected long-term total discounted reward $V_\pi(s_0)$ in (18) is adopted as the policy value in our MDP formulation, and the adjustment of the discount factor λ provides a broad range of performance characteristics [5], [28]. Except that, the long-term average reward in (22) can also be exploited as the policy value. Fig. 4 demonstrates that the long-term PEP performances of the EH TW dual-relay network by utilizing the two kinds of optimal policies corresponding to these two policy values. The curves of expected discounted reward and average reward denote the analysis results of the long-term PEP performance with the optimal polices for the case of expected total discounted reward and average reward, respectively. We can make an interesting observation

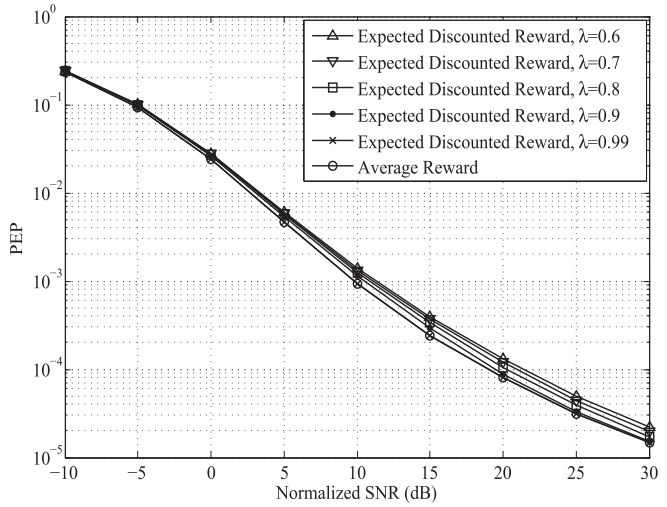

Fig. 4. Impact of discount factor λ on long-term PEP performance.

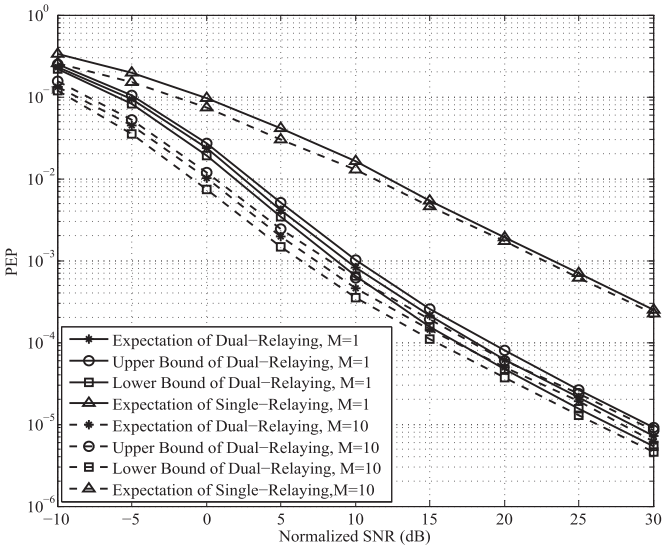

Fig. 5. Theoretical results of long-term PEP for different transmission power of source nodes.

that the performance gap between these two kinds of optimal policies becomes smaller when λ approaches to 1, and can be negligible when $\lambda = 0.99$. As a result, the average reward can be closely approximately optimized by applying the optimal transmission policy for the case of expected total discounted reward with $\lambda = 0.99$ in our MDP.

Fig. 5 illustrates the theoretical results of the long-term PEP for different transmission power of source nodes P_S, which is represented by the transmission power index M. With the proposed optimal transmission policy of the EH relay, the expectations of the long-term PEP for the proposed EH TW dual-relay network are directly computed from (25)-(27) by applying the reward function (15). In addition, since the expected capacity of contribution to the PEP by the relay R_2, $\mathbb{E}\left\{P_{e,R_2}\right\}$, can be bounded from (30) and (31), the upper and lower bounds of the long-term PEP for the dual-relay network can be calculated accordingly. It is found that the the upper and lower bounds for the EH TW dual-relay network are quite

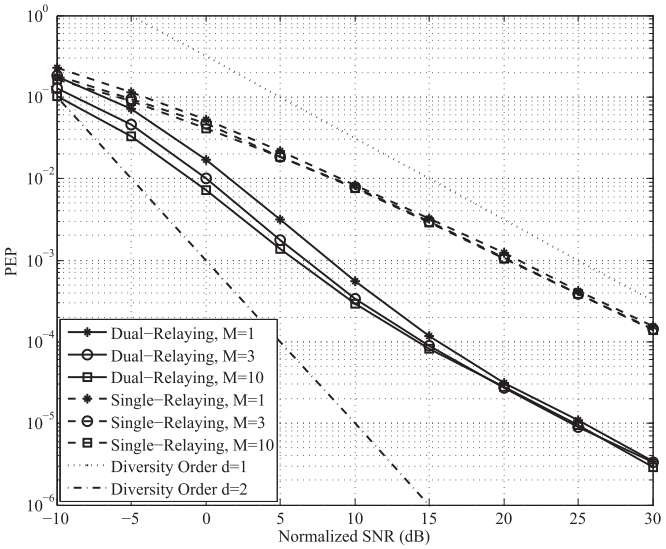

Fig. 6. Numerical results of long-term PEP for different transmission power of source nodes.

tight with respect to the expectation value, which confirms the results in Proposition 2 and Corollary 1. Moreover, it is also observed that the PEP performance of the EH TW dual-relay network is much better than that of the the traditional TW single-relay network. This is because the EH relay R_1 can utilize the random energy harvested from the solar to help the two source nodes exchange information beside the traditional relay R_2, and thus the PEP performance of the network is largely improved. Further, the PEP performance can be improved by increasing the transmission power of the source nodes for both of the two kinds of networks.

Fig. 6 shows the numerical results of the long-term PEP for different transmission power of source nodes P_S. It is observed that the TW dual-relay or single-relay networks with larger P_S can achieve a better PEP performance, and the performance curves become almost identical in sufficiently high SNR regimes. Moreover, the numerical results also prove that the long-term PEP of the EH TW dual-relay network with the obtained optimal transmission policy can be largely improved with respect to the traditional TW single-relay network. However, the EH diversity order of the dual-relay network with the optimal policy is only restricted to one, which is the same with that of the single-relay network. This is because, from Table II the EH capability index C_{EH,π^*} is approximated to 10^2 when the solar panel size Ω and the basic power unit P_U are set as the default values in Table I, and thus the EH diversity order $d = 1$ can only be achieved according to Theorem 3.

Table II lists the EH relay's battery deficient probability $P_{\pi^*}\{b \leq \rho\}$ and EH capability index C_{EH,π^*} associated with the proposed optimal transmission policy π^* under different configurations of EH communication systems, which correspond to the different values of the solar panel area size Ω, the basic power unit P_U and the static power consumption index ρ. It is observed that C_{EH,π^*} becomes larger when Ω gets larger or P_U gets smaller under the same ρ. The reason is explained as follows. C_{EH,π^*} represents the EH system's

TABLE II
BATTERY DEFICIENT PROBABILITY AND EH CAPABILITY INDEX FOR DIFFERENT CONFIGURATIONS OF EH SYSTEMS

Conf. Index	$\Omega\,(\mathrm{cm}^2)$	P_U (mW)	ρ	$P_{\pi^*}\{b \le \rho\}$	C_{EH,π^*}
1	1	10	0	$\approx 10^{-2}$	$\approx 10^{2}$
2	1	15	0	$\approx 5\times 10^{-2}$	$\approx 2\times 10^{1}$
3	2	10	0	$\approx 1.7\times 10^{-7}$	$\approx 5.9\times 10^{6}$
4	2	15	0	$\approx 7\times 10^{-4}$	$\approx 1.5\times 10^{3}$
5	2	15	1	$\approx 10^{-3}$	$\approx 10^{3}$
6	2	15	2	$\approx 1.3\times 10^{-3}$	$\approx 7.7\times 10^{2}$

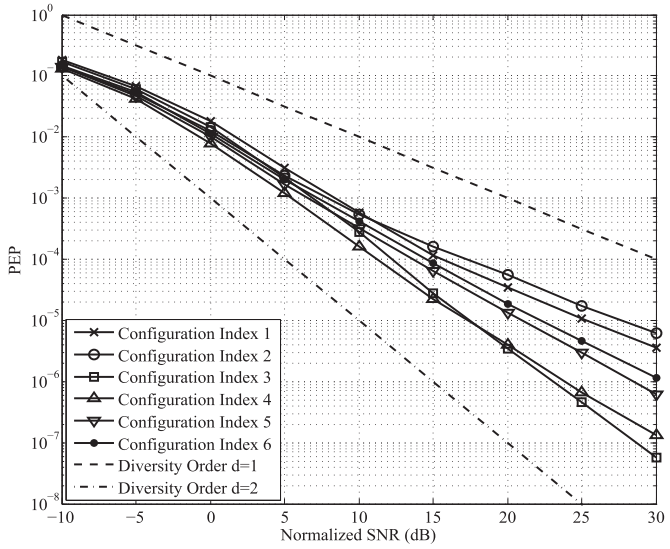

Fig. 7. Impact of EH system configurations on long-term PEP.

ability to harvest and store the basic energy quanta associated with π^*. The EH system with larger Ω is prone to obtain more solar energy, and thus the system has stronger EH capability. Moreover, the EH system with smaller P_U can achieve more numbers of basic energy quanta under the same harvested energy. Further, since the EH relay would consume more static power except the transmission power when ρ is improved, the higher ρ results in the larger $P_{\pi^*}\{b \le \rho\}$ and smaller C_{EH,π^*} under the same Ω and P_U.

Fig. 7 demonstrates the long-term PEP of the EH TW dual-relay network with the proposed optimal policy for different configurations of EH systems, which are denoted in Table II. It can be seen that the EH diversity order is equal to $d = 1$ when the configuration index is set as 1, 2, 4, 5, or 6, while the full diversity order can be almost achieved under the 3rd configuration index when C_{EH,π^*} nearly approaches to infinity. Moreover, under all of the configurations except the 3rd one, although the EH diversity orders are all equal to one, the PEP performance can be improved by increasing C_{EH,π^*} explicitly. This phenomenon coincides with Theorem 3. Further, it is also observed that there is a cross-point between the two performance curves of the 1st and 2nd configuration indices. The reason is explained as follows. In low SNR regimes the EH system performance is mainly determined

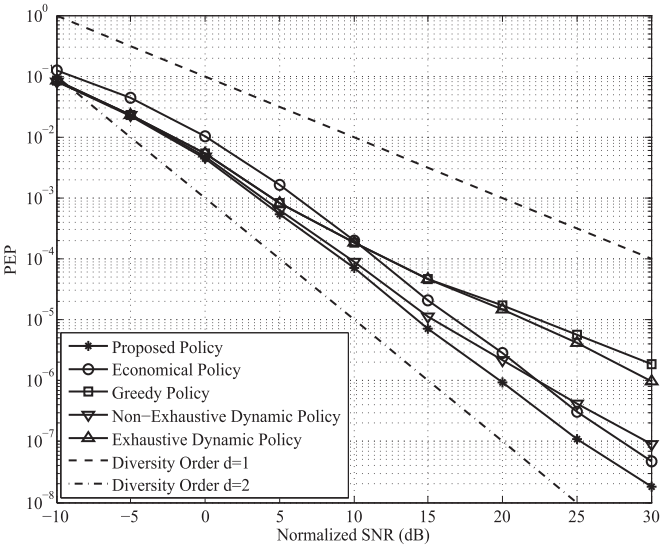

Fig. 8. Comparison of long-term PEP among different transmission policies ($P_S = 10P_U$, $\Omega = 2\mathrm{cm}^2$).

by the SNR values, whereas the EH capability dominates the system performance in high SNR regimes. Due to the higher value of P_U, the received SNR value in Configuration Index 2 is higher than that in Configuration Index 1 under the same normalized SNR value and system stochastic state. Moreover, the EH system with Configuration Index 1 achieves higher C_{EH,π^*} with respect to Configuration Index 2. As a result, the PEP performance of the EH network with Configuration Index 2 is better than that with Configuration Index 1 in low SNR regimes, while the EH system with Configuration Index 1 achieves a better PEP performance in high SNR regimes. Further, Among the 4th, 5th and 6th configuration indices, although the values of Ω and P_U are identical, C_{EH,π^*} is smaller when the static power consumption index ρ gets larger. Therefore, the PEP performance of the network becomes worse when the static power consumption of the EH relay is increased.

Fig. 8 compares the long-term PEP performance of the EH TW dual-relay network with the proposed optimal transmission policy and a couple of compared transmission policies. In these compared policies, the EH relay's transmission power is set without concern for the system total states, and we define the compared policies as follows:

A. Economical Policy

the EH relay utilizes the lowest power, i.e., one basic power P_U, to transmit signals.

B. Greedy Policy

the whole available energy in the battery is consumed by the EH relay for one policy management period.

C. Exhaustive Dynamic Policy

the EH relay only knows the CSI and determines its power equal to the minimum value to minimize the PEP in its

current channel states. Since the EH relay only focus on the transmission performance of the current policy period, it may consume the whole available energy in its battery.

D. Non-Exhaustive Dynamic Policy

the EH relay knows the status of its battery as well as the CSI. If the EH relay needs to consume the total energy in its battery to minimize the PEP, it can always leave one energy quantum in the battery for the following transmission period.

From Fig. 8, the long-term PEP performance of the optimal policy is superior to those of the compared policies. The performance trends of these policies are elaborated as follows. Economical Policy exploits the lowest transmission power, and thus obtains the lowest SNR values and the highest C_{EH,π^*}. Due to the fact that the SNR value and the EH capability dominate the PEP performance of the EH system in low SNR regimes and high SNR regimes, respectively, the PEP performance of Economical Policy is the worst in low SNR regimes, and its performance curve is closest to that of the proposed optimal policy in high SNR regimes. Since Greedy Policy consumes the most energy and obtains the lowest C_{EH,π^*}, the PEP performance of Greedy Policy is the worst in high SNR regimes. Regarding with Non-Exhaustive Dynamic Policy, the EH relay can determine the transmission power based on its battery status and CSI, and thus the PEP performance curve approaches closest to that of the optimal policy in low and moderate SNR regimes. However, since the EH relay with Non-Exhaustive Dynamic Policy has no knowledge on the solar EH status, the performance gap between this policy and the optimal policy becomes larger in high SNR regimes. In addition, the EH relay with Exhaustive Dynamic Policy is myopic and it computes the transmission power only based on the current CSI, and thus Non-Exhaustive Dynamic Policy obviously outperforms Exhaustive Dynamic Policy.

VII. Conclusion

In this paper, the PEP performance of the TW network-coded dual-relay network with stochastic EH was analyzed. The optimal transmission policy for the EH relay was developed to minimize the long-term PEP by considering the stochastic EH information, battery and channel status in the MDP formulation. Moreover, the non-conservative property of the optimal policy was uncovered. In addition, we computed the expectation and asymptotic approximation of the long-term PEP, and investigated the EH diversity performance. It was pointed out that the full diversity order is achieved only if the EH capability index approaches to infinity, and the PEP performance can be improved by increasing the EH capability. Further, the full diversity criterion of the network was proposed. Finally, the theoretical results were substantiated through extensive computer simulations.

Appendix A
Proof of Proposition 1

From (16), in asymptotically high SNR regimes, i.e., $\eta \gg 1$, we have $W_{R_1} \approx \frac{a\eta\alpha_1\gamma_1\gamma_2}{2(\gamma_1+\gamma_2)}$. By using the harmonic mean inequality $\frac{1}{2}\min(x, y) \leq \frac{xy}{x+y} \leq \min(x, y)$, it can be obtained that

$$c_L\eta \cdot \min(\gamma_1, \gamma_2) \leq W_{R_1} \leq c_U\eta \cdot \min(\gamma_1, \gamma_2), \tag{41}$$

where $c_L = \frac{a\alpha_1}{4}$,$c_U = \frac{a\alpha_1}{2}$.

From (41) and (15), the upper bound of the reward function can be given by (42), as shown at the bottom of this page. When the channel state of A-R_1 link is not superior than that of B-R_1 link, i.e., $S_{AR_1} = f \leq S_{BR_1} = g$, the upper bound of the reward function can be calculated as (43), as shown at the bottom of this page. At asymptotically high SNRs, it can be approximated as follows

$$\begin{aligned} P_{e,R_1}^{(up)}(a, f, g) &\approx \frac{e^{-(c_L\eta+1)\Gamma_f}}{(c_L\eta+1)\cdot\left(e^{-\Gamma_f} - e^{-\Gamma_{f+1}}\right)} \\ &\approx \begin{cases} \frac{\eta^{-1}}{c_L\left(1-e^{-\Gamma_1}\right)}, & f = 0; \\ 0, & f \geq 1. \end{cases} \end{aligned} \tag{44}$$

Similar derivation can be obtained when $S_{AR_1} = f > S_{BR_1} = g$, therefore the asymptotic approximation of the upper bound is given by

$$P_{e,R_1}^{(up)}(a, f, g) \approx \begin{cases} \frac{\eta^{-1}}{c_L\left(1-e^{-\Gamma_1}\right)}, & \min(f, g) = 0; \\ 0, & \min(f, g) \geq 1. \end{cases} \tag{45}$$

Similarly, the lower bound of the reward function can be asymptotically approximated as follows

$$P_{e,R_1}^{(lo)}(a, f, g) \approx \begin{cases} \frac{\eta^{-1}}{c_U\left(1-e^{-\Gamma_1}\right)}, & \min(f, g) = 0; \\ 0, & \min(f, g) \geq 1. \end{cases} \tag{46}$$

Further, due to the fact that $P_{e,R_1}^{(up)}(a, f, g) \approx P_{e,R_1}^{(lo)}(a, f, g)$ in asymptotically high SNR regimes, we can calculate the reward function approximately as follows

$$R_a(s) \approx \begin{cases} 0, & \min(f, g) \geq 1; \\ \frac{4\eta^{-1}}{a\alpha_1\left(1-e^{-\Gamma_1}\right)}, & \min(f, g) = 0. \end{cases} \tag{47}$$

Thus, we complete the proof of Proposition 1.

$$P_{e,R_1}^{(up)}(a, f, g) = \frac{\int_{\Gamma_g}^{\Gamma_{g+1}}\int_{\Gamma_f}^{\Gamma_{f+1}} \exp(-\gamma_1)\cdot\exp(-\gamma_2)\cdot\exp(-c_L\eta\cdot\min(\gamma_1, \gamma_2))d\gamma_1 d\gamma_2}{P\left(S_{AR_1} = f\right)\cdot P\left(S_{BR_1} = g\right)} \tag{42}$$

$$P_{e,R_1}^{(up)}(a, f, g) = \frac{\int_{\Gamma_f}^{\Gamma_{f+1}} \exp(-\gamma_1)\exp(-c_L\eta\gamma_1)d\gamma_1 \cdot \int_{\Gamma_g}^{\Gamma_{g+1}} \exp(-\gamma_2)d\gamma_2}{P\left(S_{AR_1} = f\right)\cdot P\left(S_{BR_1} = g\right)} = \frac{e^{-(c_L\eta+1)\Gamma_f}\left(1 - e^{-(c_L\eta+1)(\Gamma_{f+1}-\Gamma_f)}\right)}{(c_L\eta+1)\left(e^{-\Gamma_f} - e^{-\Gamma_{f+1}}\right)} \tag{43}$$

APPENDIX B
PROOF OF LEMMA 1

We prove the theorem by using the induction approach as follows.

Step 1: Assuming the initial condition $V^{(0)}(s) = 0$, the expected total discounted reward of the first iteration in (20) can be written as

$$\begin{aligned} V_a^{(1)}(s) &= R_a(s) + \lambda \sum_{s' \in S} P_a\left(s'|s\right) V^{(0)}\left(s'\right) \\ &= R_a(s) = P_{e,R_1}(a, f, g). \end{aligned} \tag{48}$$

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

$$V_a^{(1)}(e, f, g, b-1) = V_a^{(1)}(e, f, g, b). \tag{49}$$

Meanwhile, since the reward function is non-increasing in the EH relay's transmission power action and its value ranges from 0 to 1, we have $1 \geq P_{e,R_1}(a = b-1, f, g) - P_{e,R_1}(a = b, f, g) \geq 0$, and thus the following inequality holds

$$1 \geq V_{a=b-1}^{(1)}(e, f, g, b-1) - V_{a=b}^{(1)}(e, f, g, b) \geq 0. \tag{50}$$

By considering (49), (50) and (20), it can be deduced that

$$1 \geq V^{(1)}(e,f,g,b-1) - V^{(1)}(e,f,g,b) \geq 0, \forall b \in \mathcal{S}_B \backslash \{0\}. \tag{51}$$

Step 2: Assume $1 \geq V^{(k)}(e, f, g, b-1) - V^{(k)}(e, f, g, b) \geq 0, \forall b \in \mathcal{S}_B \backslash \{0\}$, in the $k^{th}(k \geq 1)$ iteration. According to (21), when $a \in \{0, 1, \cdots, b-1\}$, the value difference between the expected total discounted rewards of two adjacent battery states in the $(k+1)^{th}$ iteration can be written as

$$\begin{aligned} &V_a^{(k+1)}(e, f, g, b-1) - V_a^{(k+1)}(e, f, g, b) \\ &\quad = \lambda \cdot \mathbb{E}_s \Big\{ V^{(k)}\left(e', f', g', \min\left(b-1-a+q, N_b-1\right)\right) \\ &\qquad - V^{(k)}\left(e', f', g', \min\left(b-a+q, N_b-1\right)\right) \Big\}. \end{aligned} \tag{52}$$

With the assumption, it can be seen that

$$\begin{aligned} 1 \geq V_a^{(k+1)}(e, f, g, b-1) - V_a^{(k+1)}(e, f, g, b) \geq 0, \\ \forall a \in \{0, 1, \cdots, b-1\}. \end{aligned} \tag{53}$$

Meanwhile, in the $(k+1)^{th}$ iteration, the value difference between the expected total discounted rewards of two adjacent battery states with respect to total energy consumption in the battery is expressed as

$$\begin{aligned} &V_{a=b-1}^{(k+1)}(e, f, g, b-1) - V_{a=b}^{(k+1)}(e, f, g, b) \\ &\quad = R_{a=b-1}(f, g) - R_{a=b}(f, g) \\ &\qquad + \lambda \cdot \mathbb{E}_s \Big\{ V^{(k)}\left(e', f', g', \min\left(q, N_b-1\right)\right) \\ &\qquad - V^{(k)}\left(e', f', g', \min\left(q, N_b-1\right)\right) \Big\} \\ &\quad = P_{e,R_1}(a = b-1, f, g) - P_{e,R_1}(a = b, f, g). \end{aligned} \tag{54}$$

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

$$1 \geq V_{a=b-1}^{(k+1)}(e, f, g, b-1) - V_{a=b}^{(k+1)}(e, f, g, b) \geq 0. \tag{55}$$

According to (53), (55) and (20), for $\forall b \in \mathcal{S}_B \backslash \{0\}$, it can be easily proved that

$$1 \geq V^{(k+1)}(e, f, g, b-1) - V^{(k+1)}(e, f, g, b) \geq 0. \tag{56}$$

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

$$1 \geq V^{(i)}(e, f, g, b-1) - V^{(i)}(e, f, g, b) \geq 0, \forall i. \tag{57}$$

Specifically, in asymptotically high SNRs, i.e., $\eta \gg 1$, and when $\min(f, g) = 0$, according to Proposition 1 and Remark 2, we have $(1-\delta) \geq P_{e,R_1}(a_1, f, g) - P_{e,R_1}(a_2, f, g) \geq 0$ for $\forall a_1, a_2 (\geq a_1) \in \mathcal{A}$. Thus, similarly to the previous induction method, the following inequality holds for $\forall b \in \mathcal{S}_B \backslash \{0\}$ in ***Step 1***:

$$(1-\delta) \geq V^{(1)}(e,f,g,b-1) - V^{(1)}(e,f,g,b) \geq 0. \tag{58}$$

In ***Step 2***, assume $(1-\delta) \geq V^{(k)}(e, f, g, b-1) - V^{(k)}(e, f, g, b) \geq 0, \forall b \in \mathcal{S}_B \backslash \{0\}$, in the $k^{th}(k \geq 1)$ iteration. It can be similarly proved that

$$(1-\delta) \geq V^{(k+1)}(e, f, g, b-1) - V^{(k+1)}(e, f, g, b) \geq 0. \tag{59}$$

In ***Step 3***, Combining the previous results, for $\forall b \in \mathcal{S}_B \backslash \{0\}$, it can be easily concluded that

$$(1-\delta) \geq V^{(i)}(e, f, g, b-1) - V^{(i)}(e, f, g, b) \geq 0, \forall i. \tag{60}$$

Thus, we complete the proof of Lemma 1.

APPENDIX C
PROOF OF THEOREM 2

According to [28, Th. 6.11.6], the optimal relay action a^* is non-decreasing in the battery state b, if the the following four conditions are satisfied:

(a)The reward function $R_a(s)$ is nondecreasing in b, $\forall a \in \mathcal{A}$;
$(b)R_a(s)$ is a superadditive function on the space $\mathcal{A} \times \mathcal{S}_B$;
(c)The probability $\Theta_a(k|b, e) = \sum_{b'=k}^{N_B-1} P_a\left(S_B = b'|S_B = b, S_E = e\right)$ is nondecreasing in b;
$(d)\Theta_a(k|b, e)$ is a superadditive function on $\mathcal{A} \times \mathcal{S}_B$.

Since the reward $R_a(s)$ is independent with the battery state b from (15), the conditions (a) and (b) hold. From (13), the probability $\Theta_a(k|b, e)$ can be written as

$$\Theta_a(k|b, e) = \sum\nolimits_{q=k-b+a+\kappa\rho}^{+\infty} P(Q = q|S_E = e). \tag{61}$$

Therefore, $\Theta_a(k|b, e)$ is nondecreasing in b, and the condition (c) hold. Moreover, according to the definition (4.7.1) of superadditive functions in [28], the inequality that makes the condition (d) hold can be reformulated as (24).

Thus, we complete the proof of Theorem 2.

APPENDIX D
PROOF OF PROPOSITION 2

The cumulation distribution function of the random variable Z can be calculated as follows

$$\begin{aligned} F_Z(z) &= \Pr\left\{\frac{c_1XY}{c_2X+c_3} \le z\right\} = \Pr\left\{X\left(c_1Y - c_2z\right) \le zc_3\right\} \\ &= \Pr\left\{X\left(c_1Y - c_2z\right) \le zc_3, \left(c_1Y - c_2z\right) > 0\right\} \\ &\quad + \Pr\left\{X\left(c_1Y - c_2z\right) \le zc_3, \left(c_1Y - c_2z\right) \le 0\right\} \\ &= \Pr\left\{X \le \frac{zc_3}{(c_1Y - c_2z)}, Y > \frac{c_2z}{c_1}\right\} \\ &\quad + \Pr\left\{Y \le \frac{c_2z}{c_1}\right\}. \end{aligned} \tag{62}$$

Since X and Y are i.i.d. exponential random variables with unit mean, the second term of $F_Z(z)$ is given by

$$\Pr\left\{Y \le \frac{c_2z}{c_1}\right\} = \int_0^{\frac{c_2z}{c_1}} e^{-y}dy = 1 - e^{-\frac{c_2z}{c_1}}. \tag{63}$$

Moreover, the first term of $F_Z(z)$ is computed as

$$\begin{aligned} &\Pr\left\{X \le \frac{zc_3}{(c_1Y - c_2z)}, Y > \frac{c_2z}{c_1}\right\} \\ &= \int_0^{\frac{zc_3}{(c_1y - c_2z)}} e^{-x}dx \cdot \int_{\frac{c_2z}{c_1}}^{+\infty} e^{-y}dy \\ &= \int_{\frac{c_2z}{c_1}}^{+\infty} e^{-y}dy - \int_{\frac{c_2z}{c_1}}^{+\infty} e^{-y} \cdot e^{-\frac{zc_3}{(c_1y - c_2z)}}dy \\ &= e^{-\frac{c_2z}{c_1}} - 2e^{-\frac{c_2z}{c_1}}\sqrt{\frac{zc_3}{c_1}} \cdot K_1\left(2\sqrt{\frac{zc_3}{c_1}}\right), \end{aligned} \tag{64}$$

where $K_1(\cdot)$ is the first-order modified Bessel function of the second-kind, and the last equality is obtained by using [31, eq. (3.478)]. Substituting (64) and (63) into (62) yields to

$$F_Z(z) = 1 - 2e^{-\frac{c_2z}{c_1}}\sqrt{\frac{zc_3}{c_1}} \cdot K_1\left(2\sqrt{\frac{zc_3}{c_1}}\right). \tag{65}$$

Thus, $\mathbb{E}\left\{e^{-Z}\right\}$ can be computed as follows

$$\begin{aligned} \mathbb{E}\left\{e^{-Z}\right\} &= \int_0^{+\infty} e^{-z}dF_Z(z) = \int_0^{\infty} e^{-z}F_Z(z)dz \\ &= 1 - 2\int_0^{\infty} e^{-z}e^{-\frac{c_2z}{c_1}}\sqrt{\frac{zc_3}{c_1}} \cdot K_1\left(2\sqrt{\frac{zc_3}{c_1}}\right)dz. \end{aligned} \tag{66}$$

By using [31, eq. (6.643)], we obtain

$$\begin{aligned} \mathbb{E}\left\{e^{-Z}\right\} &= 1 - \frac{c_1}{c_1+c_2}\exp\left(\frac{c_3}{2(c_1+c_2)}\right)W_{-1,\frac{1}{2}}\left(\frac{c_3}{c_1+c_2}\right) \\ &= 1 - \frac{c_1c_3}{(c_1+c_2)^2}U\left(2, 2, \frac{c_3}{c_1+c_2}\right), \end{aligned} \tag{67}$$

where $W(\cdot)$ and $U(\cdot)$ represent Whittaker function and Kummer function in [32]. From Equality (13.2.5) in [32], it is derived that

$$\begin{aligned} \mathbb{E}\left\{e^{-Z}\right\} &= 1 - \frac{c_1c_3}{(c_1+c_2)^2}\int_0^{\infty} \frac{t}{1+t}\exp\left(-\frac{c_3}{c_1+c_2}t\right)dt \\ &= \frac{c_2}{c_1+c_2} + \frac{c_1c_3}{(c_1+c_2)^2}\exp\left(\frac{c_3}{c_1+c_2}\right)E_1\left(\frac{c_3}{c_1+c_2}\right), \end{aligned} \tag{68}$$

where $E_1(\cdot)$ denotes the exponential integral function. Due to the inequality of $E_1(\cdot)$ function, $\frac{1}{2}\ln\left(1+\frac{2}{x}\right) < e^xE_1(x) < \ln\left(1+\frac{1}{x}\right)$, we have

$$\begin{aligned} &\frac{c_2}{c_1+c_2} + \frac{c_1c_3}{2(c_1+c_2)^2}\ln\left(1+\frac{2(c_1+c_2)}{c_3}\right) < \mathbb{E}\left\{e^{-Z}\right\} \\ &< \frac{c_2}{c_1+c_2} + \frac{c_1c_3}{(c_1+c_2)^2}\ln\left(1+\frac{c_1+c_2}{c_3}\right), \end{aligned} \tag{69}$$

Thus, we complete the proof of Proposition 2.

REFERENCES

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

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

[3] 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.

[4] M.-L. Ku, W. Li, Y. Chen, and K. J. R. Liu, "On energy harvesting gain and diversity analysis in cooperative communications," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 12, pp. 2641–2657, Dec. 2015.

[5] 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.

[6] W. Li, M.-L. Ku, Y. Chen, and K. J. R. Liu, "On outage probability for stochastic energy harvesting communications in fading channels," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1893–1897, Nov. 2015.

[7] S. Jangsher, H. Zhou, V. O. K. Li, and K. C. Leung, "Joint allocation of resource blocks, power, and energy-harvesting relays in cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 482–495, Mar. 2015.

[8] 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.

[9] I. Ahmed, A. Ikhlef, R. Schober, and R. K. 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.

[10] B. Gurakan and S. Ulukus, "Cooperative diamond channel with energy harvesting nodes," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1604–1617, May 2016.

[11] 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.

[12] 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.

[13] B. Varan and A. Yener, "Delay constrained energy harvesting networks with limited energy and data storage," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1550–1564, May 2016.

[14] 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)*, Dec. 2013, pp. 345–348.

[15] 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)*, Dec. 2014, pp. 133–137.

[16] W. Li, M.-L. Ku, Y. Chen, and K. J. R. Liu, "On outage probability for two-way relay networks with stochastic energy harvesting," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1901–1915, May 2016.

[17] D. Li, C. Shen, and Z. Qiu, "Two-way relay beamforming for sum-rate maximization and energy harvesting," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 3115–3120.

[18] Z. Fang, X. Yuan, and X. Wang, "Distributed energy beamforming for simultaneous wireless information and power transfer in the two-way relay channel," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 656–660, Jun. 2015.

[19] 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.
[20] H.-Q. Lai and K. J. R. Liu, "Space-time network coding," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1706–1718, Apr. 2011.
[21] H. S. Wang and N. Moayeri, "Finite-state Markov channel—A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.
[22] W. Li, M. L. Ku, Y. Chen, and K. J. R. Liu, "Diversity analysis for two-way multi-relay networks with stochastic energy harvesting," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 1235–1239.
[23] S. Zhang, F. Gao, Y. H. Yi, and C. X. Pei, "Study of segment training based individual channel estimation for two-way relay network," *IEEE Commun. Lett.*, vol. 17, no. 2, pp. 401–404, Feb. 2013.
[24] Z. Gao, H.-Q. Lai, and K. J. R. Liu, "Differential space-time network coding for multi-source cooperative communication," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3146–3157, Nov. 2011.
[25] 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. Inf. Theory*, vol. 57, no. 10, pp. 6405–6418, Oct. 2011.
[26] T. Wang and L. Vandendorpe, "On the optimum energy efficiency for flat-fading channels with rate-dependent circuit power," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4910–4921, Dec. 2013.
[27] 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.
[28] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY, USA: Wiley, 1994.
[29] Z. Ding, S. M. Perlaza, I. Esnaola, and H. V. Poor, "Power allocation strategies in energy harvesting wireless cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 846–860, Feb. 2014.
[30] W. C. Jakes, *Microwave Mobile Communications*. New York, NY, USA: Wiley, 1974.
[31] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Francisco, CA, USA: Academic, 2007.
[32] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. New York, NY, USA: Dover, 1972.

Wei Li received the B.S. and M.S. degrees in electrical and electronics engineering from Xi'an Jiaotong University, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Department of Information and Communication Engineering, Xi'an Jiaotong University, China, in 2016. From 2005 to 2011, he was a Senior Engineer with Huawei Technology Corporation. From 2013 to 2015, he was a visiting student with the University of Maryland, College Park, MD, USA. In 2016, he became a Faculty Member of the School of Information Engineering, Chang'an University, China. His current 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 Post-Doctoral Research Fellow with Prof. L.-C. Wang at the Department of Electrical and Computer Engineering, National Chiao Tung University and with Prof. V. Tarokh at the School of Engineering and Applied Sciences, Harvard University. In 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. In 2013, he was a Visiting Scholar with the Signals and Information Group of Prof. K. J. R. Liu, University of Maryland. His current research interests are in the areas of green communications, cognitive radios, and optimization of radio access. He was a recipient of the Best Counseling Award in 2012 and the university-level Best Teaching Award in 2014–2016 from National Central University. He was also a recipient of the Exploration Research Award from 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 in 2004, the M.Phil. degree from The Hong Kong University of Science and Technology in 2007, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2011. He was with Origin Wireless Inc. as a Founding Principal Technologist. Since 2015, he has been a Full Professor with the University of Electronic Science and Technology of China. His research interests include multimedia, signal processing, game theory, and wireless communications.

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

K. J. Ray Liu (F'03) was a Distinguished Scholar-Teacher with the University of Maryland, College Park, in 2007, where he is currently 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 smart radios for smart life.

Dr. Liu is a fellow of AAAS. He was a recipient of the 2016 IEEE Leon K. Kirchmayer Technical Field Award on graduate teaching and mentoring, the 2014 IEEE Signal Processing Society Award, and the IEEE Signal Processing Society 2009 Technical Achievement Award. He has been recognized by T. Reuters as a Highly Cited Researcher. He is a member of the IEEE Board of Directors. He was a President of the IEEE Signal Processing Society, where he has served as Vice President—Publications and on the Board of Governors. He has also served as the Editor-in-Chief of the *IEEE Signal Processing Magazine*.

He also received teaching and research recognitions from the University of Maryland, including the university-level Invention of the Year Award, the college-level Poole and Kent Senior Faculty Teaching Award, the Outstanding Faculty Research Award, and the Outstanding Faculty Service Award, all from A. James Clark School of Engineering.

Shihua Zhu, photograph and biography not available at the time of publication.