On the Achievable Sum Rate for Two-Way Relay Networks with Stochastic Energy Harvesting

Wei Li*†, Meng-Lin Ku‡, Yan Chen*, and K. J. Ray Liu*

*Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA
†Department of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, 710049, China
‡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, an optimal transmission policy for two-way relay networks is investigated by using a stochastic energy harvesting (EH) model. Considering the channel and finite battery conditions, we propose an optimal relay transmission policy to maximize the long-term achievable sum rate of the network. The design problem is formulated as a Markov decision process (MDP), and the well-known value iteration approach is used to find the optimal policy. Based on the optimal transmission policy, we analyze the expected achievable sum rate and point out a spreading structure for the optimal relay power with respect to the solar panel size. Simulation results demonstrate that our proposed optimal transmission policy outperforms other policies.**

I. INTRODUCTION

Recently, energy harvesting (EH) cooperative communications have attracted significant attentions due to its effectiveness in resolving energy supply problems in wireless sensor networks. In this problem, the EH source and/or relay nodes can make use of renewable energy sources, e.g., solar, wind and vibration, to replenish their power supply and fulfill data transmission. Although an inexhaustible energy supply from environments enables EH nodes to communicate for an infinite lifetime, power management and transmission scheduling remain a crucial issue because of the randomness and uncertainty of the harvested energy.

One-way EH relay networks were introduced in [1] and [2]. The authors in [1] designed the optimal transmission scheme for the EH half-duplex relay in the two-hop network when the source has a single or two energy arrivals. By considering delay and non-delay constrained traffic, the optimal power allocation for the classic three-node Gaussian relay network with EH nodes was investigated in [2]. Recently, power allocation algorithms for maximizing short-term sum rates in the two-way EH relay networks were proposed in [3]-[5] using deterministic EH models, i.e., the energy state information (ESI) is non-causal and the energy arrival profile is known prior to transmission scheduling. No data buffer in the relay was assumed in [3], while the data buffer of the relay was considered in [4], which means the relay can cache data and exploit more flexible scheduling policies. In [5], an optimization framework assuming the uncertainty of the channel state information (CSI) was proposed. 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. Furthermore, none of these EH cooperative communication works linked the real solar irradiance data to the design of the optimal transmission policies.

Motivated by the aforementioned discussions, we propose an optimal relay transmission policy for the two-way EH relay network using stochastic EH models, where there is no need for energy arrival profiles prior to data transmission. The relay is solar-powered, and the long-term achievable sum rate of the two-way relay network is maximized by adapting the relay transmission power to its current battery state, CSI and ESI. We exploit the stochastic EH model in [6] whose underlying parameters are directly trained by a real data record of solar irradiance [7]. Meanwhile, the fading channels between the sources and relay are formulated by a finite-state Markov model [8]. The optimal and adaptive relay transmission problem is then formulated as a discounted Markov decision process (MDP) [9] and solved by a value iteration approach. Furthermore, the expected achievable sum rate associated with the optimal relay transmission power is analyzed. We also point out a spreading structure of the optimal relay transmission power with respect to the solar panel area size, which can help reduce the computational complexity in obtaining the optimal relay power.

II. TWO-WAY ENERGY HARVESTING RELAYING

In Fig. 1, we consider a two-way EH relay network with the analog network coding (ANC) protocol, where two traditional wireless source nodes, A and B, exchange information simultaneously via an EH relay node, R, within multiple access (MA) and broadcast (BC) phases. The relay can harvest energy from the solar and store energy in the rechargeable battery to supply the forthcoming communications. It is assumed that all nodes are half-duplex and there are no direct links between two source nodes. Without loss of generality, it is assumed that the wireless channels are reciprocal, quasi-static and Rayleigh flat fading. The channel fading coefficients, h_{ar} and h_{br}, are independent and identically distributed (i.i.d.) complex Gaussian random variables $\mathcal{CN}(0,1)$.

This work is supported by the China Scholarship Council and the Fundamental Research Funds for the Central Universities of China.

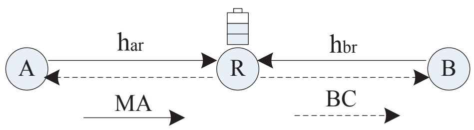

Fig. 1. The two-way EH relay network

In the MA phase, A and B transmit their signals to R simultaneously. In the BC phase, R exploits the amplify-and-forward (AF) protocol to broadcast the superimposed signal. A and B extract the desired signal from the received signal by subtracting their own self-interference. Thus, the achievable rates of the A-to-B link and the B-to-A link can be expressed as [10]

$$R_{ab} = \frac{1}{2}\log\left[1+\frac{\gamma_1\gamma_2 P P_r}{N_0\left(\gamma_1 P+\gamma_2 P+\gamma_2 P_r+N_0\right)}\right], \quad (1)$$

$$R_{ba} = \frac{1}{2}\log\left[1+\frac{\gamma_1\gamma_2 P P_r}{N_0\left(\gamma_1 P+\gamma_2 P+\gamma_1 P_r+N_0\right)}\right], \quad (2)$$

where P is the transmission power of A and B, P_r is the transmission power of R, the additive white Gaussian noise of each node is given by $\mathcal{CN}(0, N_0)$, $\gamma_1 = |h_{ar}|^2$ and $\gamma_2 = |h_{br}|^2$. Therefore, the achievable sum rate of the network is defined as $R_t = R_{ab} + R_{ba}$, and its upper bound in the high signal-to-noise ratio (SNR) region can be derived as [11]

$$R_t < R_{t\_up} = \frac{1}{2}\log\left[(P_r/N_0)^2(\min(\gamma_1,\gamma_2))^2 \Big/ \tilde{P}_r\right], \quad (3)$$

where $\tilde{P}_r = (P+P_r)/P$.

III. MARKOV DECISION PROCESS WITH STOCHASTIC MODELS

In this section, we attempt to find the optimal transmission power for the relay in order to maximize the achievable sum rate. The design of the relay transmission policy depends on several factors, like the channel conditions among the three nodes, and the finite battery capacity and the solar EH conditions at the relay. The design framework is formulated as an MDP with the goal of maximizing the long-term achievable sum rate. The MDP is mainly composed of the state space, the action space, the state transition probabilities and the reward function. 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, $\mathcal{Q}_e = \{0, 1, \cdots, N_e - 1\}$ represents the solar EH state, $\mathcal{H}_{ar} = \{0, 1, \cdots, N_c - 1\}$ and $\mathcal{H}_{br} = \{0, 1, \cdots, N_c - 1\}$ are the states of h_{ar} and h_{br} respectively, and $\mathcal{Q}_b = \{0, 1, \cdots, N_b - 1\}$ denotes the relay finite battery state. Meanwhile, let $S = (Q_e, H_{ar}, H_{br}, Q_b) \in \mathcal{S}$ represent the stochastic state of the MDP. The detailed descriptions of all the elements in the MDP are provided as follows.

A. Relay Actions of Transmission Power

The action space is defined as the relay transmission power $\mathcal{W} = \{0, 1, \cdots, N_p - 1\}$ where $N_p \le N_b$. When the action $w \in \mathcal{W}$ is chosen by the relay, the relay transmission power P_r is set as wP_u during one transmission period T, where P_u is a basic transmission power level corresponding to one energy quantum E_u during the transmission period, i.e., $E_u = P_u T$. If $w = 0$, it means that the relay remains silent during the transmission period.

B. Solar Energy Harvesting States

We exploit a discrete stochastic EH model, N_e-state Gaussian mixture hidden Markov chain [6], to mimic the evolution of the solar EH conditions. This EH model is real-data-driven 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 [7]. When the relay is at the i^{th} solar state, the solar power per unit area, P_h, follows the Gaussian distribution $\mathcal{N}(\mu_i, \rho_i)$, and thus, the harvested energy during one transmission period T is given by $E_h = P_h T s\eta$, where s denotes the solar panel area size and η represents the energy conversion efficiency. Moreover, each state is governed by a state transition probability $P(Q_e = j|Q_e = i)$, for $i, j \in \mathcal{Q}_e$, and can be described by an EH probability in terms of the number of harvested energy quanta, i.e., $P(E = q|Q_e = i)$ [6] for $q \in \{0, 1, \cdots, \infty\}$, through the parameters μ_i and ρ_i.

C. Channel States

The instantaneous channel power, γ_1 and γ_2, is quantized into several levels using a finite number of thresholds, given by $\Gamma = \{0 = \Gamma_0, \Gamma_1, \cdots, \Gamma_{N_c} = \infty\}$, and formulated as a finite-state Markov chain. If the channel power belongs to the interval $[\Gamma_i, \Gamma_{i+1})$, for $i \in \{0, 1, \cdots, N_c - 1\}$, the fading channel is in the i^{th} channel state. Thus, the stationary probability of the i^{th} channel state can be expressed as

$$P(H=i) = \int_{\Gamma_i}^{\Gamma_{i+1}} \frac{1}{\lambda}\exp\left(-\frac{\gamma}{\lambda}\right)d\gamma = \exp\left(-\frac{\Gamma_i}{\lambda}\right) - \exp\left(-\frac{\Gamma_{i+1}}{\lambda}\right), \quad (4)$$

where λ is the average channel power. It is assumed that the channel can only transit from the current state to its neighboring states, and the channel state transition probabilities $P(H = j|H = i)$, for $i \in \{0, \cdots, N_c - 1\}$, $j \in \{\max(0, i-1), \cdots, \min(i+1, N_c - 1)\}$, can be defined as in [8].

D. Relay Battery States

The relay battery storage capacity is finite and uniformly quantized into several levels in units of E_u. When the relay is in the b^{th} battery state, it means that the number of available energy quanta in the battery is b.

Since the relay battery state transition is related to both the transmission action and the number of harvested energy quanta, the battery state transition probability at the i^{th} EH state can be explicitly represented as

$$P_w(Q_b = b'|Q_b = b,\ Q_e = i) \quad (5)$$
$$= \begin{cases} P(E = b' - b + w|Q_e = i), & b' = (b-w), \cdots, N_b - 2 \\ 1 - \sum_{j=0}^{N_b-2-b+w} P(E = j|Q_e = i), & b' = N_b - 1 \end{cases}$$

where $b \in \mathcal{Q}_b$ and $w \in \{0, 1, \cdots, \min(b, N_p - 1)\}$.

E. State Transition Probability

Since the transition probabilities of the solar EH states and channel states are independent, the state transition probability from the state $S=(i,j,k,b)$ to the state $S=(i',j',k',b')$ with respect to the action w can be given by

$$\begin{aligned}&P_w\left(S=\left(i',j',k',b'\right)|S=(i,j,k,b)\right)\\&=P\left(Q_e=i'|Q_e=i\right)\cdot P\left(H_{ar}=j'|H_{ar}=j\right)\\&\cdot P\left(H_{br}=k'|H_{br}=k\right)\cdot P_w\left(Q_b=b'|Q_b=b,\ Q_e=i\right).\end{aligned} \tag{6}$$

F. Reward Function

The upper bound of the achievable sum rate, R_{t_up}, is adopted as the reward in the MDP, and the reward function at the state $S=(i,j,k,b)$ with respect to the relay action w is denoted as

$$\begin{aligned}&R_w\left(S=(i,j,k,b)\right)\\&=\frac{\int_{\Gamma_k}^{\Gamma_{k+1}}\int_{\Gamma_j}^{\Gamma_{j+1}}R_{t\_up}\frac{1}{\lambda}\exp\left(-\frac{\gamma_1}{\lambda}\right)\frac{1}{\lambda}\exp\left(-\frac{\gamma_2}{\lambda}\right)d\gamma_1 d\gamma_2}{P\left(H_{ar}=j\right)\cdot P\left(H_{br}=k\right)}.\end{aligned} \tag{7}$$

Substituting (3) and (4) into (7) yields the reward function as (8). By using (8) and (9), the reward function can be calculated in (10), for which the involved integral is given as

$$\begin{aligned}&\int_a^b \ln\gamma\cdot\frac{1}{\lambda}\exp\left(-\frac{\gamma}{\lambda}\right)d\gamma\\&=-\ln b\cdot\exp(-b/\lambda)+\ln a\cdot\exp(-a/\lambda)-Ei\left(-a/\lambda\right)+Ei\left(-b/\lambda\right)\end{aligned} \tag{11}$$

where $Ei(\cdot)$ is the exponential integral function [12], and

$$P\left(\Gamma_j<\Gamma_k\right)=\sum_{j=0}^{N_c-2}\sum_{k=j+1}^{Nc-1}P\left(H_{ar}=j\right)\cdot P\left(H_{br}=k\right). \tag{12}$$

IV. Optimization of Relay Transmission Actions

Define the policy $\pi(s):\mathcal{S}\rightarrow\mathcal{W}$ as the action that specifies the relay transmission power. The goal of the MDP is to find the optimal $\pi(s)$ in the state s to maximize the expected discounted long-term reward as follows

$$V_\pi(s_0)=E_\pi\left[\sum_{k=0}^{\infty}\lambda^k R_{\pi(s_k)}(s_k)\right],s_k\in\mathcal{S},\pi(s_k)\in\mathcal{W}, \tag{13}$$

where s_0 is the initial state and $0\le\lambda\le1$ is a discount factor. 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 [9]. The optimal policy can be found through the Bellman equation, given by

$$V_{\pi^*}(s)=\max_{w\in W}\left(R_w(s)+\lambda\sum_{s'\in\mathcal{S}}P_w\left(s'|s\right)V_{\pi^*}\left(s'\right)\right). \tag{14}$$

The well-known value iteration approach can be applied to find the optimal policy [13]. In practical applications, the solar EH state can be updated using real data of solar irradiance and Bayes's rule at the relay [6], and the current channel state and finite battery state can be easily acquired by the relay. Thus, after obtaining the optimal transmission policy, the relay can make full use of the state information and exploit the look-up table method to decide its optimal transmission power in every transmission period T.

V. Expected Reward Analysis

In this section, we will discuss the expected achievable sum rate for the optimal policy. First, the battery state transition probability with respect to the optimal relay action $w^\star$ at the state $S=(i,j,k,b)$ can be given by

$$\begin{aligned}&P_{i,j,k}\left(Q_b=b'|Q_b=b\right)\\&=\begin{cases}0,\quad 0\le b'\le b-w^*-1;\\P\left(E=b'-b+w^*|Q_e=i\right),\ b-w^*\le b'\le N_b-2;\\1-\sum\limits_{b'=0}^{N_b-2}P_{i,j,k}\left(Q_b=b'|Q_b=b\right),\quad b'=N_b-1,\end{cases}\end{aligned} \tag{15}$$

where $b,b'\in\{0,\cdots,N_b-1\}$. Thus, the state transition probability associated with the optimal policy is expressed as

$$\begin{aligned}&P\left(S=\left(i',j',k',b'\right)|S=(i,j,k,b)\right)\\&=P\left(Q_e=i'|Q_e=i\right)\cdot P\left(H_{ar}=j'|H_{ar}=j\right)\\&\cdot P\left(H_{br}=k'|H_{br}=k\right)\cdot P_{i,j,k}\left(Q_b=b'|Q_b=b\right),\end{aligned} \tag{16}$$

where $i,i'\in\{0,1,\cdots,N_e-1\}$, $j,k\in\{0,\cdots,N_c-1\}$, $j'\in\{\max(0,j-1),\cdots,\min(j+1,N_c-1)\}$ and $k'\in\{\max(0,k-1),\cdots,\min(k+1,N_c-1)\}$. Let $p_{(i,j,k,b)}$ represent the stationary probability of the state $S=(i,j,k,b)$ for the optimal policy, and it satisfies $\sum_{(i,j,k,b)\in\mathcal{S}}p_{(i,j,k,b)}=1$. Meanwhile, since $p_{(i,j,k,b)}$ and the state transition probability have the relation as follows

$$\sum_{(i,j,k,b)\in\mathcal{S}}P((i',j',k',b')|(i,j,k,b))\cdot p_{(i,j,k,b)}=p_{(i',j',k',b')}, \tag{17}$$

$p_{(i,j,k,b)}$ can be easily computed by solving linear equations.

Therefore, the expected reward can be computed by taking the expectation of the reward function in (10) as follows

$$\bar{R}=\sum_{i=0}^{N_e-1}\sum_{j=0}^{N_c-1}\sum_{k=0}^{N_c-1}\sum_{b=0}^{N_b-1}p_{(i,j,k,b)}\times R_{w^*}(S=(i,j,k,b)). \tag{18}$$

$$R_w\left(S=(i,j,k,b)\right)=\frac{1}{2\ln2}\left[\ln\frac{(wP_u/N_0)^2}{\tilde{P}_r}+\frac{2\int_{\Gamma_k}^{\Gamma_{k+1}}\int_{\Gamma_j}^{\Gamma_{j+1}}\ln\min(\gamma_1,\gamma_2)\cdot\frac{1}{\lambda}\exp\left(-\frac{\gamma_1}{\lambda}\right)\frac{1}{\lambda}\exp\left(-\frac{\gamma_2}{\lambda}\right)d\gamma_1 d\gamma_2}{P\left(H_{ar}=j\right)\cdot P\left(H_{br}=k\right)}\right] \tag{8}$$

$$\begin{aligned}&\int_{\Gamma_k}^{\Gamma_{k+1}}\int_{\Gamma_j}^{\Gamma_{j+1}}\ln\min(\gamma_1,\gamma_2)\cdot\frac{1}{\lambda}\exp\left(-\frac{\gamma_1}{\lambda}\right)\frac{1}{\lambda}\exp\left(-\frac{\gamma_2}{\lambda}\right)d\gamma_1 d\gamma_2\\&=P\left(\Gamma_j<\Gamma_k\right)\cdot P\left(H_{br}=k\right)\cdot\int_{\Gamma_j}^{\Gamma_{j+1}}\ln\gamma_1\frac{1}{\lambda}\exp\left(-\frac{\gamma_1}{\lambda}\right)d\gamma_1+P\left(\Gamma_j\ge\Gamma_k\right)\cdot P\left(H_{ar}=j\right)\cdot\int_{\Gamma_k}^{\Gamma_{k+1}}\ln\gamma_2\frac{1}{\lambda}\exp\left(-\frac{\gamma_2}{\lambda}\right)d\gamma_2\end{aligned} \tag{9}$$

$$R_w\left((i,j,k,b)\right)=\frac{1}{2\ln 2}\left[\ln\frac{\left(wP_u/N_0\right)^2}{\tilde{P}_r}+\frac{2P\left(\Gamma_j<\Gamma_k\right)}{P\left(H_{ar}=j\right)}\int_{\Gamma_j}^{\Gamma_{j+1}}\ln\gamma_1\cdot\frac{1}{\lambda}\exp\left(-\frac{\gamma_1}{\lambda}\right)d\gamma_1+\frac{2P\left(\Gamma_j\geq\Gamma_k\right)}{P\left(H_{br}=k\right)}\int_{\Gamma_k}^{\Gamma_{k+1}}\ln\gamma_2\cdot\frac{1}{\lambda}\exp\left(-\frac{\gamma_2}{\lambda}\right)d\gamma_2\right] \quad (10)$$

VI. SIMULATION RESULTS

In this section, the performance of our proposed optimal policy based on the stochastic EH model in [6] is evaluated by computer simulations. The irradiance data from 2011 to 2012 measured by the solar site in Elizabeth City State University are applied for simulations [7]. The numbers of the solar EH states, channel states and relay battery states are four, six and twelve, respectively. The solar irradiance measurements are taken at five minute intervals, the solar panel area size s is set from 1cm^2 to 10cm^2, and the solar energy conversion efficiency $\eta = 20\%$. In the channel model, the channel is quantized as $\Gamma = \{0, 0.3, 0.6, 1.0, 2.0, 3.0, \infty\}$, the average channel power $\lambda = 1$ and the channel power is formulated using Jakes' model with the normalized Doppler frequency $f_D = 0.05$ [8]. Meanwhile, the relay battery state is initialized randomly, the relay transmission action is changed every five minutes, the basic transmission power $P_u = 40\times10^3\mu W$, and the transmission power of two source nodes $P = P_u$.

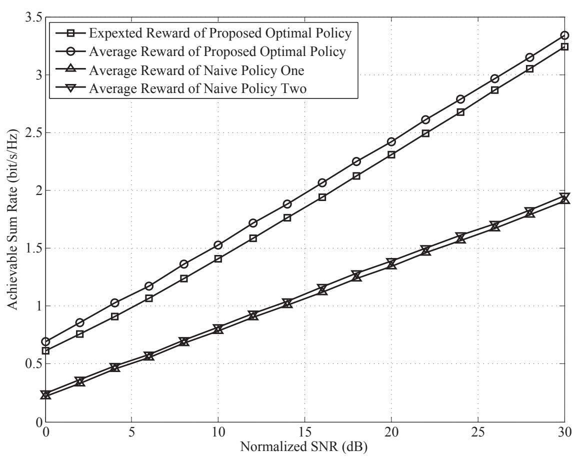

Fig. 2. Achievable sum rates of our proposed optimal policy and two naive policies ($s = 1\text{cm}^2$)

Fig. 2 shows the achievable sum rates of our proposed optimal policy and two naive policies when the normalized SNR is defined with respect to P_u. In these two naive policies, the relay transmission power is set without concern for the system states. Instead, if the relay battery storage is non-empty, the relay attempts to exploit the lowest power and the largest available battery power to transmit data in Naive Policy One and Naive Policy Two, respectively. The difference between the expected reward and average reward of proposed optimal policy lies in: the former is calculated according to (18) and derived by the EH model based on the solar irradiance data from 2008 to 2010, while the later exploits this stochastic EH model to decide the optimal relay power in the following two years, 2011 and 2012, and average the rewards in (10) of all decisions. It can be seen that the achievable sum rate of our proposed optimal policy is superior to those of the two naive policies. Moreover, the average reward of the optimal policy is very close to the expected reward since the solar irradiance is relatively stable from 2008 to 2012.

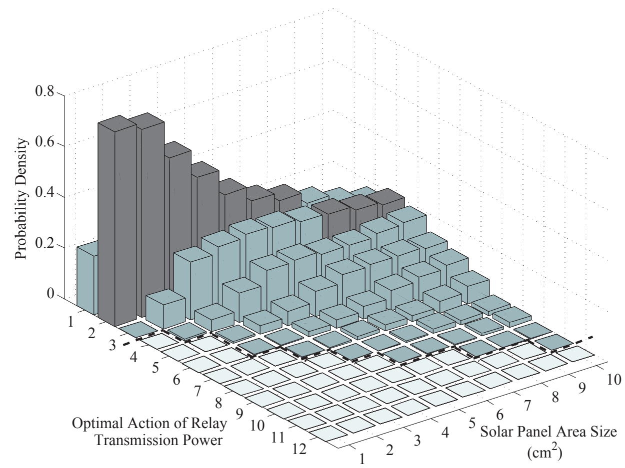

Fig. 3. Spreading structure of optimal relay power actions with respect to solar panel area sizes (normalized SNR = 10dB)

Fig. 3 demonstrates the probability distribution of the optimal relay transmission power actions in all states with respect to the solar panel area size. The bar height represents the probability density corresponding to the optimal relay power and solar panel size, and the probability density of white bars is zero. A spreading structure is observed in this figure, and it can be seen that the optimal relay power with the highest probability density, and the maximal power of optimal relay actions, are both non-decreasing with respect to the solar panel size. This is because if the solar panel size becomes larger, the relay is prone to obtain more energy quanta under the same solar state, and the relay power corresponding to the maximal battery transition probability in (5) is bigger under the same MDP state. Thus, the optimal relay power becomes bigger according to the Bellman equation in (14). This property can help to reduce the computational complexity in obtaining the optimal relay actions.

VII. CONCLUSION

In this paper, the optimal and adaptive relay transmission policy for maximizing the long-term reward in the two-way EH relay network is proposed. Unlike previous works, we exploit stochastic models to formulate the solar irradiance state and fading channel state. An MDP framework is designed to obtain the optimal relay transmission power based on the solar ESI, CSI and relay finite battery status. Moreover, the expected achievable sum rate is theoretically analyzed, and the result exhibits an elegant spreading structure of the optimal relay power with respect to the solar panel size.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] M.-L. Ku, Y. Chen, and K. J. R. Liu, "Data-Driven Stochastic Models and Policies for Energy Harvesting Sensor Communications," *arXiv preprint arXiv:1405.7254*, 2014.

[7] N. R. E. Laboratory, "(2012)Solar radiation resource information," *[Online] Available: http://www.nrel.gov/rredc/*.

[8] 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.

[9] M. Puterman, *Markov Decision Process-Discrete Stochastic Dynamic Programming*. John Wiley and Sons, 1994.

[10] 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.

[11] 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.

[12] I. S. Gradsbteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products (Seventh Edition)*. Elsevier, 2007.

[13] D. Bertsekas, *Dynamic Programming and Optimal control*. Athena Scientific, 2000.