Sum-Rate Maximization for RSMA-Enabled Energy Harvesting Aerial Networks With Reinforcement Learning

Jaehyup Seong*, Mesut Toka†, and Wonjae Shin‡
*Department of Artificial Intelligence Convergence Network, Ajou University, Suwon, South Korea
†Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea
‡School of Electrical Engineering, Korea University, Seoul, South Korea
Emails: {*john12234, †tokamesut}@ajou.ac.kr, ‡wjshin@korea.ac.kr

***Abstract*—In this paper, we propose a joint power and precoder design framework for energy-harvesting aerial networks, where an aerial base station serves multiple users via rate-splitting multiple access (RSMA) using harvested energy. To maximize the sum-rate from the long-term perspective, we utilize a deep reinforcement learning (DRL) approach to allocate optimal transmission power at each time based on the randomness property of the channel environment, harvested energy, and battery power information. Moreover, we employ sequential least squares programming (SLSQP) to design the RSMA precoder maximizing the sum-rate with the allocated power. Numerical results show the superiority of the proposed scheme over baseline methods in terms of the average sum-rate performance.**

***Index Terms*—Energy harvesting networks, RSMA, DRL.**

I. Introduction

Unmanned aerial vehicle (UAV) communications have drawn significant attention in the last few years [1]. UAVs can be served not only as users but also as aerial base stations (ABSs). Specifically, the deployment of ABSs enables to support of ubiquitous connectivity and high data rates with favorable line-of-sight (LOS) propagation conditions. Thanks to these properties, ABSs have emerged as one of the promising technologies for the fifth generation (5G) networks and beyond [1]. However, even with the aforementioned advantages of ABS networks, the fatal problem is that an ABS has a finite-sized battery, leading to a constrained operation time. That is, providing stable and sustainable services can be restricted, which in turn causes a performance bottleneck in ABS networks.

To tackle this issue, energy harvesting-aided ABSs have emerged as a key solution with the intent of prolonging the ABS's lifetime. Indeed, the authors of [2] have developed solar-powered UAVs and showed that solar energy can be harvested for over 300 % of the power required for flight. Thus, the remaining power from the flight can be used in communications. Inspired by this, the authors in [3] have analyzed the outage probability in radio frequency (RF)-powered ABS networks using non-orthogonal multiple access (NOMA). Further, the authors in [4] have designed an optimal policy to maximize the system throughput from the long-term perspective based on the orthogonal multiple access (OMA) in solar-powered ABS networks. However, in the previous works, some important factors to be considered for application to the real ABS environment have not been considered.

First, in [3], the entire harvested energy in each time slot has been used without saving it for future time slots. However, since the energy arrival from renewable resources (e.g., solar and ambient RF) usually has a randomness property, it is required to store the energy for future time steps, where a lack of harvested energy incurs, rather than consuming entire harvested energy to provide reliable service during the total period. Although the authors of [4] have maximized the system throughput from the long-term perspective, it has been assumed that arrival energy is determined according to ABSs' locations, resulting in a lack of reality. Second, perfect channel state information (CSI) at the transmitter (CSIT) and at the receiver (CSIR) have been assumed in [3], [4]; however, due to the rapid channel variations caused by the high mobility of ABSs, acquiring perfect CSIT or CSIR is challenging. Third, the OMA, which has been considered in [4], cannot efficiently use the frequency band. Also, the NOMA, which has been considered in [3] is not an appropriate solution in multiuser multiple-input single-output (MU-MISO) systems [5].

Different from the existing works, we propose the power allocation framework through the deep reinforcement learning (DRL) approach, named soft actor-critic (SAC) algorithm [6], in energy harvesting ABS networks maximizing the average sum-rate over the total time slot. Making enable it to be applied in real ABS environments, realistic constraints such as randomness of energy arrival, time-varying channels, and imperfect CSI are considered. Further, it is assumed that ABSs cannot have any prior knowledge of future arrival energy and CSI. On top of this, we utilize the rate-splitting multiple access (RSMA), which has robustness under imperfect CSI and brings high spectral and power efficiencies [7], [8] to maximize the instantaneous sum-rate in each time slot using the allocated power via DRL. To handle the non-convexity of designing the RSMA precoder, we derive a sub-optimal precoder in an iterative manner using the sequential least squares programming (SLSQP) algorithm [9]. Numerical results show that the proposed framework in energy harvesting ABS networks significantly improves the sum-rate compared with several benchmark schemes under imperfect CSI.

This research was supported in part by the National Research Foundation of Korea (NRF) grants (No.2021R1A4A1030775, No.2022R1A2C4002065) and in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grants (No.2021-0-00467, No.2022-0-00704).

$$R_{\mathsf{c},k}^{(i)} = \log_2\left(1 + \frac{|(\hat{\mathbf{h}}_k^{(i)})^{\mathsf{H}}\mathbf{p}_{\mathsf{c}}^{(i)}|^2}{\sum_{j=1}^{K}|(\hat{\mathbf{h}}_k^{(i)})^{\mathsf{H}}\mathbf{p}_j^{(i)}|^2 + \sum_{j\in\mathcal{L}}\mathbb{E}[|(\mathbf{e}_k^{(i)})^{\mathsf{H}}\mathbf{p}_j^{(i)}|^2] + \sigma_n^2}\right). \tag{1}$$

$$R_k^{(i)} = \log_2\left(1 + \frac{|(\hat{\mathbf{h}}_k^{(i)})^{\mathsf{H}}\mathbf{p}_k^{(i)}|^2}{\sum_{j=1,j\neq k}^{K}|(\hat{\mathbf{h}}_k^{(i)})^{\mathsf{H}}\mathbf{p}_j^{(i)}|^2 + \sum_{j\in\mathcal{L}}\mathbb{E}[|(\mathbf{e}_k^{(i)})^{\mathsf{H}}\mathbf{p}_j^{(i)}|^2] + \sigma_n^2}\right). \tag{2}$$

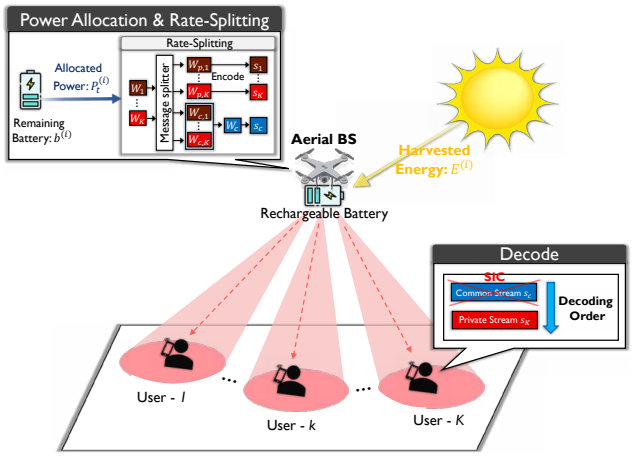

Fig. 1. System model of DRL-based RSMA with energy harvesting ABS.

II. System Model and Problem Formulation

We consider an MU-MISO network as illustrated in Fig. 1, where an ABS simultaneously serves K single antenna users, and both the ABS and users have imperfect CSI. The ABS harvests energy irregularly from renewable energy sources and allocates the optimal total transmission power. It then transmits the desired signals to the users with the given allocated power. Let the superscript i be the time index. The ABS with hybrid energy harvesting mechanism as in [10] harvests energy $E^{(i)}$ stochastically from renewable energy sources (e.g., solar and ambient RF) with the energy harvesting probability p_e at each time slot. After that, the ABS broadcasts a linearly precoded signal to users using the total transmission power $P_t^{(i)}$ during transmission time T by utilizing the remaining battery $b^{(i)}$. The ABS then updates the battery status for the next time slot $b^{(i+1)}$ based on the amount of harvested energy $E^{(i)}$. Herein, since the rechargeable battery has maximum energy storage $b_{\max}$, the $b^{(i+1)}$ can be denoted as $\min\{b^{(i)} - TP_t^{(i)} + E^{(i)}, b_{\max}\}$. Since ABSs can provide dominant LOS links, channels between the ABS and users, denoted by $\mathbf{h}_k \in \mathbb{C}^{N_t\times 1}$, are assumed to be exposed to Rician fading. Therefore, the signal received at the user k can be expressed as $y_k = \mathbf{h}_k^{\mathsf{H}}\mathbf{x} + n$, where $\mathbf{x} \in \mathbb{C}^{N_t\times 1}$ represents the signal vector transmitted from the ABS, and $n \sim \mathcal{CN}(0, \sigma_n^2)$ denotes complex additive white Gaussian noise (AWGN). The erroneous CSI vector both at the ABS and receivers can be expressed as $\hat{\mathbf{h}}_k = \mathbf{h}_k - \mathbf{e}_k \in \mathbb{C}^{N_t\times 1}$, where $\mathbf{e}_k \in \mathbb{C}^{N_t\times 1}$ denotes the channel estimation error vector.

To formulate rate expressions under imperfect CSIT and CSIR, we utilize the concept of generalized mutual information as [7]. Hence, the common and private rate expressions for the k-th user at time slot i can be formulated as (1) and (2) given at the top of this page, where $j \in \mathcal{L} \triangleq \{\mathsf{c}, 1, \cdots, K\}$. Here, $\mathbf{p}_{\mathsf{c}}^{(i)} \in \mathbb{C}^{N_t\times 1}$ and $\mathbf{p}_k^{(i)} \in \mathbb{C}^{N_t\times 1}$ respectively denote the private and common precoding vectors as below:

$$\mathbf{p}_{\mathsf{c}}^{(i)} = \sqrt{P_t^{(i)}\mu_{\mathsf{c}}^{(i)}}\mathbf{w}_{\mathsf{c}}^{(i)}, \quad \mathbf{p}_k^{(i)} = \sqrt{P_t^{(i)}\mu_k^{(i)}}\mathbf{w}_k^{(i)}, \tag{3}$$

where $\mu_{\mathsf{c}}^{(i)}$ ($\mu_k^{(i)}$) and $\mathbf{w}_{\mathsf{c}}^{(i)}$ ($\mathbf{w}_k^{(i)}$) $\in \mathbb{C}^{N_t\times 1}$ denote the power ratios and normalized precoding vectors for common (private) messages, respectively. Since the power usage at time i must not exceed the total power $P_t^{(i)}$, it follows $\mu_{\mathsf{c}}^{(i)} + \sum_{k=1}^{K}\mu_k^{(i)} = 1$. In addition, $R_{\mathsf{c}}^{(i)} = \min_k R_{\mathsf{c},k}^{(i)}$ should be satisfied because the common message should be decoded by all users.

Therefore, the optimization problem to maximize the total sum-rate during a total time T_o can be formulated as:

$$\max_{\mathbf{p}_{\mathsf{c}}^{(i)},\mathbf{p}_1^{(i)},\cdots,\mathbf{p}_K^{(i)},P_t^{(i)}} \sum_{i=0}^{T_o} R_{\mathsf{sum}}^{(i)} \tag{4}$$

$$\text{s.t.} \quad b^{(i)} = \min\{b^{(i-1)} - P_t^{(i-1)}T + E^{(i-1)}, b_{\max}\}, \tag{4a}$$

$$P_t^{(i)} \le \frac{b^{(i)}}{T}, \tag{4b}$$

$$\sum_{j\in\mathcal{L}}\|\mathbf{p}_j^{(i)}\|^2 \le P_t^{(i)}, \ R_{\mathsf{c},k}^{(i)} \ge R_{\mathsf{c}}^{(i)}, \tag{4c}$$

where $R_{\mathsf{sum}}^{(i)} = R_{\mathsf{c}}^{(i)} + \sum_{k=1}^{K}R_k^{(i)}$.

III. The Proposed Scheme

Our objective is maximizing not only $R_{\mathsf{sum}}^{(i)}$ but also $\sum_{i=0}^{T_o}R_{\mathsf{sum}}^{(i)}$, enhancing stability and sustainability of ABS networks. To do so, we first reformulate our problem into the Markov Decision Process (MDP) with the variables at the i-th time step, which are respectively existing in remaining battery space ($\mathcal{B}$), harvested energy space ($\mathcal{E}$), CSI space ($\mathcal{H}$) and transmission power space ($\mathcal{P}_t$) such that $b^{(i)} \in \mathcal{B}$, $E^{(i)} \in \mathcal{E}$, $\hat{\mathbf{H}}^{(i)} \in \mathcal{H}$, and $P_t^{(i)} \in [0, \frac{b^{(i)}}{T}] \cap \mathcal{P}_t$. Then, we define a tuple $(\mathcal{S}, \mathcal{A}, \mathbb{P}, r, \gamma)$. Here, $\mathcal{S}$ denotes the state-space, $\mathcal{A}$ denotes the action space, and $\mathbb{P}$ represents the state transition probability function of the next state information for the given state information and action. Additionally, r denotes the reward function, and γ denotes the discount factor. The state information at the ABS is expressed as $s^{(i)} = (E^{(i)}, \hat{\mathbf{H}}^{(i)}, b^{(i)}) \in \mathcal{S}$, where $\mathcal{S}$ is continuous. Meanwhile, the action-state information at the ABS is expressed as $a^{(i)} = P_t^{(i)} \in \mathcal{A}$, where $\mathcal{A}$ is also continuous. When the ABS uses the allocated transmission power $P_t^{(i)}$ into $s^{(i)}$, the instantaneous sum-rate $R_{\mathsf{sum}}^{(i)}$ can be reformulated as the reward function $R(s^{(i)}, P_t^{(i)})$. By applying the SAC algorithm with the expressed state, action, and reward function, the problem of obtaining the optimal policy π^* can be given as:

$$\pi^* = \arg\max_{\pi\in\Pi} \mathbb{E}_\pi\left[\sum_{i=0}^{\infty}\gamma^{(i)}R(s^{(i)}, P_t^{(i)}) + \alpha H(\pi(\cdot|s^{(i)}))\middle|\pi\right]$$

$$\text{s.t.} \quad (4a),\ (4b), \tag{5a}$$

where Π is the set of feasible policies, $H(\pi(\cdot|s^{(i)}))$ is the entropy of the policy at $s^{(i)}$, and α is the trade-off parameter between exploitation and exploration. Following the optimal policy π^* for the power allocation, the optimal value function V^*, which denotes a measure of the long-term maximum achievable sum-rate of the state $s^{(i)}$, can be obtained.

Once the optimal power $P_t^{(i)}$ is allocated at the ABS through the power allocation policy π^* at time slot i, the SLSQP algorithm is employed to derive sub-optimal RSMA precoder as follows, where the subscript τ denotes the step-index in the SLSQP algorithm. Note that the following steps are included at each time slot i. The initial iteration point is composed of initialized values of $\mu_c^{(i)}/\mu_k^{(i)}$ and the real and imaginary parts of $\mathbf{w}_c^{(i)}/\mathbf{w}_k^{(i)}$, denoted as $\mathbf{x}_0^{(i)} \in \mathbb{R}^{N^{\text{SLSQP}}}$, where $N^{\text{SLSQP}} = (2 \times N_t \times (K+1)) + K + 1$.

- **Step 1:** Construct the quadratic sub-problem of (4) using the second-order Taylor expansion, with the $\mathbf{x}_0^{(i)}$ and Hessian matrix of the Lagrangian for (4), that is, $\mathbf{W}_0^{(i)} \in \mathbb{R}^{N^{\text{SLSQP}} \times N^{\text{SLSQP}}}$. Due to the complexity in calculating the Hessian matrix, the Wilson-Han-Powell method [9] is adopted to replace $\mathbf{W}_0^{(i)}$ with the positive definite matrix $\mathbf{A}_0^{(i)} \in \mathbb{R}^{N^{\text{SLSQP}} \times N^{\text{SLSQP}}}$ under suitable assumptions.
- **Step 2:** Solve the constructed quadratic sub-problem and test whether the termination condition is satisfied. If so, the current solution $\mathbf{x}_\tau^{(i)}$ is regarded as the solution to the original problem (4), and the iteration is terminated.
- **Step 3:** Otherwise, use the line search method by adopting the L_1-norm as the loss function to calculate the search step length $\alpha_\tau^{(i)}$ in the current direction.
- **Step 4:** Update the symmetric definite matrix $\mathbf{A}_\tau^{(i)}$ using the Han–Powell quasi-Newton method with a BFGS update [9] and update the iteration point $\mathbf{x}_\tau^{(i)}$ by using the search step length $\alpha_\tau^{(i)}$. Then, reconstruct the quadratic sub-problem and revisit **Step 2** for the next iteration.

Remark: As mentioned above, the number of $(2\times N_t\times(K+1))+K+1$ optimization variables are required to implement the SLSQP algorithm, resulting in higher computational complexity as the size of networks increases. Therefore, in [11], the parameters of the RSMA precoder have been partially optimized by the SLSQP, and the remaining parameters have been optimized via the minimum mean square error (MMSE) based method. Nevertheless, if complexity is not taken into account, a feasible solution can be derived using only SLSQP.

IV. SIMULATION RESULTS AND DISCUSSIONS

With 10 learning processes, we evaluate an average sum-rate over 1000 time steps versus battery capacity. The number of users and transmit antennas are set as $K = 2$ and $N_t = 2$, respectively. It is assumed that each element of $\mathbf{h}_k$ is independent and identically distributed (i.i.d.), where a Rician shape parameter and scale parameter are set as 3 and 1, respectively. $\mathbf{e}_k$ follows i.i.d. complex Gaussian distribution such that $\mathbf{e}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_{e,k}^2\mathbf{I})$, where each estimation error is assumed to have the same variance as $\sigma_e^2 = 0.1$. The variance of AWGN σ_n^2 is fixed as 1. Besides, we assume the energy harvesting probability as $p_e = 0.5$ with the Bernoulli process, and the transmission time T is set as 1. In Fig. 2, the proposed scheme is compared with the various kind of both the optimal power allocation-based schemes and greedy power allocation-based schemes that instantaneously use all the harvested energy without saving it for future use. Herein, the spatial division multiple access (SDMA) schemes are based on [12]. The proposed scheme shows a higher performance than the benchmark schemes in all battery capacity regions. This result implies the importance of the power allocation policy for self-sustainability and the superiority of the RSMA in the ABS networks, where imperfect CSI usually arises and ABSs have a finite-sized battery. Future directions include further optimizing trajectory design or flight energy consumption.

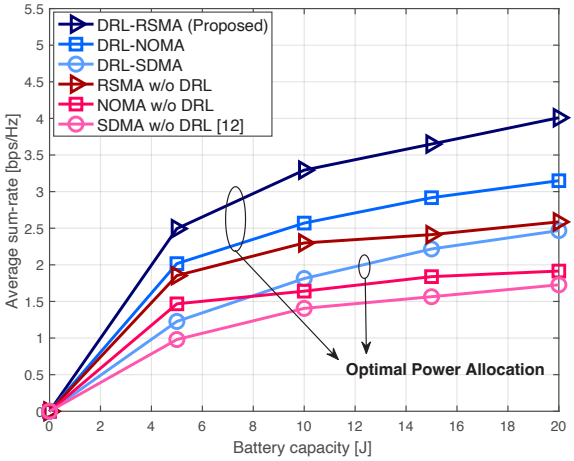

Fig. 2. Average sum-rate comparisons of the proposed scheme (DRL-RSMA) with benchmark schemes versus battery capacity.

REFERENCES

[1] D. Liu *et al.*, "Opportunistic UAV utilization in wireless networks: Motivations, applications, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 5, pp. 62–68, 2020.

[2] S. Morton *et al.*, "Solar powered UAV: Design and experiments," in *Proc. 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, 2015, pp. 2460–2466.

[3] T. M. Hoang *et al.*, "Outage probability of aerial base station NOMA MIMO wireless communication with RF energy harvesting," *IEEE Internet of Things Journal*, vol. 9, no. 22, pp. 22 874–22 886, 2022.

[4] Y. Sun *et al.*, "Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems," *IEEE Trans. on Commun.*, vol. 67, no. 6, pp. 4281–4298, 2019.

[5] B. Clerckx *et al.*, "Is NOMA efficient in multi-antenna networks? a critical look at next generation multiple access techniques," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1310–1343, 2021.

[6] T. Haarnoja *et al.*, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. Int. Conf. Machine Learning*, 2018, pp. 1861–1870.

[7] J. An *et al.*, "Rate-splitting multiple access for multi-antenna broadcast channel with imperfect CSIT and CSIR," *arXiv preprint arXiv:2102.08738*, 2021.

[8] J. Park *et al.*, "Rate-splitting multiple access for 6G networks: Ten promising scenarios and applications," *arXiv preprint arXiv:2306.12978*, 2023.

[9] D. Kraft, *A software package for sequential quadratic programming*. Wiss. Berichtswesen d. DFVLR Brunswick, Germany, 1988.

[10] T. Quyen *et al.*, "Optimizing hybrid energy harvesting mechanisms for UAVs," *EAI Endorsed Transactions on Energy Web*, vol. 7, no. 30, 2020.

[11] J. Seong *et al.*, "Sum-rate maximization of RSMA-based aerial communications with energy harvesting: A reinforcement learning approach," *arXiv preprint arXiv:2306.12977*, 2023.

[12] Q. H. Spencer *et al.*, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, 2004.