Trajectory Optimization and Phase-Shift Design in IRS-Assisted UAV Network for Smart Railway

Yu Min Park, Yan Kyaw Tun, Member, IEEE, Zhu Han, Fellow, IEEE, and Choong Seon Hong, Senior Member, IEEE

Abstract—The recent trend towards the smart transportation system has spurred the development of the smart railway. However, enabling trains with seamless wireless connectivity using the roadside units (RSUs) is extremely challenging, mostly due to the lack of line of sight link. To address this issue, we propose a novel framework that uses intelligent reflecting surfaces (IRS) and unmanned aerial vehicles (UAVs). This framework enables on-demand, low-cost, and flexible communication between trains and the railway stations. Therefore, there is a need to propose efficient communication infrastructure for various smart railway services (e.g., augmented reality). However, RSUs can be used to enable railway communication between the train and the railway station [1]. Therefore, there is a need to design a cost-effective, efficient UAV-based communication for the smart railways.

Moreover, wireless communication technology through Intelligent Reflecting Surface (IRS) is being studied recently. IRS is composed of several reflectors, and the angle of the corresponding reflector can be corrected through an electrical signal [4]. The wireless channels between transmitters and receivers can be flexibly reconfigured using IRS in a wireless network to achieve desired realizations and/or distributions. However, IRS communication faces many challenges including resource allocation and phase-shift optimization according to the communication target area. The work in [5] considered the number of suitable Reflectors to provide IRS-based communication to multi-users. In [6], [7], the authors approached the problem of optimizing both the phase of beamforming and the phase-shift of IRS in a MIMO environment in a numerical manner. However, IRS in a fixed location has a limitation in providing communication as the receiver moves.

Recently, the IRS-assisted UAV network has drawn significant research attention. In [8] and [9], the authors tried to overcome communication shading regions by applying IRS located in buildings and UAV-BS together. Though, IRSs located in buildings support passive communication to UAV-BS. In [10], the authors used an IRS-attached form of communication infrastructure on UAVs called Aerial Intelligent Reflecting Surface. They aim to maximize the worst-case signal-to-noise ratio (SNR) over all locations in a target area by jointly optimizing the transmit beamforming for the source node and the placement and 3D passive beamforming for the AIRS. However, finding the optimal placement of UAVs without user movement might not respond instantly to a rapidly changing environment of smart railways. Therefore, in this work, we apply reinforcement learning to quickly optimize the trajectory of UAV. Also, we perform the phase-shift optimization with a Transportation Problem (TP) between reflector and trains to maximize the minimum data rate of trains.

As a result, we propose a method to utilize UAVs and IRS for wireless smart railways communication. Due to the usage of IRS, UAVs can extend the communication range even further. In addition, UAVs’ flexibility allows for obstacle avoidance and effective resource utilization. The main contribution of this paper is to propose a joint phase-shift design of IRS and UAV trajectory optimization problem for the IRS-assisted UAV network with the objective of maximizing the minimum achievable rate of trains. To solve the joint optimization problem, we decompose it into two subproblems. Next, we employ a Binary Integer Linear Programming (BILP) model and reinforcement learning, in order to solve decomposed subproblems. In the simulation results, we validated our proposed method in two aspects. First, the performance of the proposed method could be confirmed by comparing the minimum data rate of trains when using the fixed IRS and random IRS. Second, we demonstrate that trajectory optimization outperforms a UAV with a fixed height.
II. SYSTEM MODEL AND PROBLEM FORMULATION

Over time $K = \{1, 2, \ldots, K\}$, we consider a single-input multiple-output (SIMO) wireless system in which a base station (BS) equipped with a single antenna serves a set $M$ of $M$ trains through a UAV with IRS. We assume that wireless communication between BS and trains is blocked by obstacles, such as buildings. The UAV is equipped with the smart controller that can perform phase-shift of IRS. In addition, we assume that the channel state information (CSI) of all channels involved is perfectly known at the BS and the UAV. All trains move at the same speed $v_{\text{train}}$ and the speed $v_u$ of UAV is limited to $v_{\text{max}}$. Also, UAV $u$ is limited to the height $h_u$ between $h_{\text{min}}$ and $h_{\text{max}}$. The IRS mounted on UAV is composed of a set $N$ of $N$ reflectors, and the normalized power radiation pattern in each reflector can be expressed as [11]:

$$F(\theta, \varphi) = \begin{cases} \cos^3 \theta, & \theta \in [0, \frac{\pi}{2}], \varphi \in [0, 2\pi], \\ 0, & \theta \in (\frac{\pi}{2}, \pi], \varphi \in [0, 2\pi], \end{cases}$$

(1)

where $\theta$ and $\varphi$ are the elevation and azimuth from the antenna to a certain transmitting/receiving direction, respectively.

Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$ denote the ideal IRS phase-shift vector. We define the practical diagonal IRS phase-shift matrix $\Phi$ by multiplying the indicator matrix $I$ for the association between a IRS reflector and trains. The indicator matrix $I$ is defined as

$$I = \begin{bmatrix} i_{1,1} & \cdots & i_{1,M} \\ \vdots & \ddots & \vdots \\ i_{N,1} & \cdots & i_{N,M} \end{bmatrix},$$

(2)

where $i_{n,m}$ is a binary decision variable, i.e., $i_{n,m} \in \{0,1\}, \forall n \in N, \forall m \in M$. If train $m$ is associated with a IRS indicator $n$, $i_{n,m} = 1$; otherwise, $i_{n,m} = 0$. The practical diagonal IRS phase-shift matrix $\Phi$ is defined as

$$\Phi(I) = \text{adig}(e^{j\theta_1}\hat{i}_1, e^{j\theta_2}\hat{i}_2, \ldots, e^{j\theta_N}\hat{i}_N),$$

(3)

where $\alpha$ is the fixed amplitude reflection coefficient and $\hat{i}_n = [i_{n,1}, i_{n,2}, \ldots, i_{n,M}]$ represents the IRS indicator vector of reflector $n$. The achievable channel gain based the distance-dependent path loss over the time slot $k$ is modeled as follow:

$$h(k) = \rho_0 \left( \frac{d_{k}}{d_0} \right)^{-\delta},$$

(4)

where $d_k$, $\rho_0$ and $\delta$ denote the distance between elements over time slot $k$, the path loss at the reference distance $d_0$ and the path loss exponent, respectively [12]. The deterministic channel from the BS $b$ to the UAV $u$ is denoted by $h_{bu}$. The channel between the UAV $u$ and the train $m \in M$ is denoted by $h_{um}$. Therefore, the channel from the BS to the train is defined by $h_{bu}Fh_{um}$. The received signal to noise ratio (SNR) at train $m$ at time slot $k$ is given by

$$\gamma_m(k) = \frac{p[F(\theta_1, \varphi_1)F(\theta_2, \varphi_2)h_{um}(k)\Phi h_{um}(k)]^2}{\sigma^2},$$

(5)

where $\sigma^2$ is the noise power. $\theta_1$, $\varphi_1$, $\theta_2$, and $\varphi_2$ represent the elevation angle and the azimuth angle from the BS to the center of the IRS, the elevation angle and the azimuth angle from the center of the IRS to train $m$, respectively [11]. We can express the achievable rate at train $m$ at time slot $k$ can be written as

$$R_m(k) = B \log_2(1 + \gamma_m), \quad \forall m \in M, \forall k \in K,$$

(6)

where $B$ is the BS bandwidth.

Given the network model, our objective is to solve the problem of joint trajectory $T = \{T_1, T_2, \ldots, T_K\}$ of UAV and phase-shift $P = \{\Phi_1, \Phi_2, \ldots, \Phi_K\}$ of IRS for wireless smart railways communication with the goal of maximizing the minimum data rate of trains over $K$ time slots. We can formulate this problem as follows:

$$\max_{T, P} \sum_{k=1}^{K} R(k)$$

(7a)

s.t. 

$$C_1 : R(k) = \min_{m \in M} R_m(k), \quad \forall k \in K$$

(7b)

$$C_2 : R_m(k) \geq R_0, \quad \forall m \in M, \forall k \in K$$

(7c)

$$C_3 : \Phi_k(I) = \text{adig}(e^{j\theta_1}\hat{i}_1, e^{j\theta_2}\hat{i}_2, \ldots, e^{j\theta_N}\hat{i}_N),$$

(7d)

$$C_4 : i_{n,k} \in \{i_{n,1}, i_{n,2}, \ldots, i_{n,M}\}, \quad \forall n \in N$$

(7e)

$$C_5 : 0 \leq \theta_n \leq \frac{\pi}{2}, \quad \forall n \in N$$

(7f)

$$C_6 : 0 \leq v_u(k) \leq v_{\text{max}}, \quad \forall k \in K$$

(7g)

$$C_7 : i_{n,k} \leq \min(k), \quad \forall n \in N, \forall k \in K$$

(7h)

$$C_8 : i_{n,m} \in \{0, 1\}, \quad \forall n \in N, \forall m \in M,$$

(7i)

where $R_0$ is the minimum achievable rate required by a train. Here, C2 guarantees a minimum achievable rate for all trains. C3 and C4 represent that a phase-shift $\Phi$ is determined through a indicator matrix $I$. C5 ensures that the angles of reflection are between 0 and $\pi/2$. C6 and C7 are for the UAV’s behavioral constraints, respectively, for speed and altitude. Additionally, since this paper is to derive the optimal trajectory of the UAV, we assumed that there are sufficient mobile batteries for the UAV.

III. PROPOSED SOLUTION

From (7), the optimization variables of UAV trajectory and IRS phase-shift are coupling in both objective function and the constraints. It can be shown that the problem becomes non-convex and NP-hard. Therefore, it is challenging to solve our formulated problem in polynomial time. For this reason, by using the block coordinate descent (BCD) method, the proposed problem is decomposed into two subproblems: 1) IRS phase-shift problem and 2) UAV trajectory optimization problem. Then, the decomposed subproblems are solved alternatively by deploying binary integer linear programming (BILP) and soft actor-critic (SAC) algorithms.

A. Optimal Phase-Shift of IRS At the Given UAV Trajectory

We discuss about the optimal phase-shift design of IRS at the given UAV trajectory. With the objective of maximizing the minimum data rates of trains over the time slot $k$, the optimal phase-shift design of IRS is formulated as follows:

$$\max_{\Phi_k} \sum_{k=1}^{K} R(k)$$

(8a)

s.t. 

$$C_1)-(C_5), (C_8).$$

(8b)

Since the interference between reflectors is not considered in this paper, we can formulate problem (8) of finding the optimal indicator matrix as TP. Here, each reflector is a source and each train is a destination. Therefore, the proposed subproblem can be solved by using a BILP with indicator matrix $I$ composed of decision variables $i_{n,m}$. To apply the BILP method, subproblem (8) can be reformulated by maximizing an additional variable $Z$ that is an upper bound for the
Algorithm 1: SAC Learning Process for UAV’s Trajectory.

1: input: $\theta_1, \theta_2, \phi$
2: $\theta_1 \leftarrow \theta_1, \theta_2 \leftarrow \theta_2, D \leftarrow \emptyset$
3: for each episode do
4:   for each learning step $k'$ do
5:     $a_{k'} \sim \pi_\phi(a|s_{k'})$, $s_{k'+1} \sim p(s_{k'+1}|s_{k'}, a_{k'})$
6:     $D \leftarrow D \cup \{s_k, a_k, r_k, s_{k'+1}\}$
7:   end for
8:   for each gradient step do
9:     Update soft Q-function $\theta_1, \theta_2$ with $\nabla_\theta J_Q(\theta)$
10:    Update the policy weight $\phi$ with $\nabla_\phi J_\pi(\phi)$
11:    Update the temperature $\tau$, $\Theta \leftarrow \Theta \cup \{\tau_{\theta'}\}$ with
12:    $\bar{\theta}_i \leftarrow \tau \bar{\theta}_i + (1 - \tau)\bar{\theta}_i$ for $i \in \{1, 2\}$
13:  end for
14: end for
15: Output: The optimal SAC Parameters $\theta_1, \theta_2, \phi$

achievable rates $R_m$ of each train $m$ as follows:

$$\max_i Z$$

subject to:

$$\sum_{m \in M} i_{n,m} = 1, \quad \forall n \in N$$

$$R_m = \sum_{n \in N} i_{n,m} \cdot R_{n,m}, \quad \forall m \in M$$

$$Z \leq R_m, \quad \forall m \in M$$

where (11) ensures that a single IRS reflector is assigned to only one train. In (12), $R_{n,m}$ is the achievable rate between IRS reflector $n$ and train $m$. Therefore, (12) represents that train’s achievable rate is the sum of achievable rates from associated IRS reflectors. As a result, BILP can be converted into a convex quadratic problem, and then any BILP can be solved in polynomial time. Therefore, we can solve the subproblem (9) using GEKKO, CVXPY as an optimization tool for BILP [13].

B. Optimal Trajectory of UAV At the Given IRS Phase-Shift

Based on the optimal phase-shift of the IRS over time slot $k$, the optimal trajectory of the UAV is derived. Therefore, the optimization trajectory of UAV at the given phase-shift of IRS can be expressed as follows:

$$\max_{T_k} \tilde{R}(k)$$

subject to:

(C1), (C2), (C6), (C7).

To solve the problem (10), we deploy the Soft Actor-Critic (SAC) method [14], one of Reinforcement Learning (RL). This is because the optimal value can be derived by entering the state in real-time, which is an advantage of using RL rather than numerical learning. SAC is an off-policy model that trains into the Replay Buffer at the same time as a variation of the Actor-Critic model, which is a policy gradient method. SAC has function approximators for both the soft Q-function and the policy, and instead of running evaluation and improvement to convergence, alternate between optimizing both networks with stochastic gradient descent. In learning process, the episode of learning in each time slot $k$ are divided into certain steps $K = \{1, 2, \ldots, K^*\}$. We consider a parameterized soft Q-function $Q_\phi(s_{k'}, a_{k'})$ and a tractable policy $\pi_\phi(s_{k'}, a_{k'})$. The parameters of these networks are $\theta$ and $\phi$.

The soft Q-function parameters can be trained to minimize the soft Bellman residual

$$J_Q(\theta) = E_{(s_{k'}, a_{k'}) \sim D} \left[ \frac{1}{2} (Q_\phi(s_{k'}, a_{k'}) - (r_{s_{k'}, a_{k'}} + \gamma E_{s_{k'+1} \sim p(V_{\theta}(s_{k'+1} + 1))} Q_\phi(s_{k'+1}, a_{k'+1}))^2 \right]$$

(11)

where the value function is implicitly parameterized through the soft Q-function parameters. SAC makes use of two soft Q-functions to mitigate positive bias in the policy improvement step that is known to degrade the performance of value-based methods. In particular, we parameterize two soft Q-functions, with parameters $\theta_1, \theta_2$, and train them independently to optimize $J_Q(\theta_1), J_Q(\theta_2)$. The policy parameters can be learned by directly minimizing the expected KL-divergence:

$$J_\pi(\phi) = E_{s_{k'} \sim D} \left[ \alpha \log(\pi_\phi(a_{k'}|s_{k'}) - Q_\phi(s_{k'}, a_{k'})) \right]$$

(12)

There are several options for minimizing $J_\pi$. We compute gradients for $\alpha$ with the following objective

$$J(\alpha) = E_{s_{k'} \sim D} \left[ \alpha \log \pi_\phi(a_{k'}|s_{k'}) - \alpha \tilde{R} \right]$$

(13)

where $\tilde{R}$ is a constant of the target entropy. The elements of Markov Decision Process (MDP) are designed to maximize the minimum achievable rate of trains. We can define the essential elements of MDP as follows.

In step $k'$, state $s_{k'}$ is defined as

$$s_{k'} = \{p_{b,k'}, p_{u,k'}, \{p_{m,k'}\}_{m \in M}\}$$

(14)

where $p_{b,k'}, p_{u,k'},$ and $\{p_{m,k'}\}_{m \in M}$ are the positions of the BS, UAV, and the set $M$ of trains in time slot $k'$. That’s when the heights of the base sation and trains are always 0. In other words, the BS and trains are located only on the ground.

The action is the amount of movement in the x-axis, y-axis and z-axis to move from the current position in time slot $k'$,

$$a_{k'} = \{\Delta x_{u,k'}, \Delta y_{u,k'}, \Delta z_{u,k'}\}$$

(15)

The reward over time $k'$ is divided into three and can be expressed as follows:

$$r_{k'} = \begin{cases} 0, & \text{if } \Delta R \leq 0 \text{ or } \tilde{R} < R_0, \\ \Delta R, & \text{if } \tilde{R} > 0, \\ -1, & \text{if } l_{u,k'} > l_{\text{max}} \text{ or } l_{u,k'} < l_{\text{min}}. \end{cases}$$

(16)

where $\Delta R$ denotes the change in $\tilde{R}$, i.e., $\Delta \tilde{R} = \tilde{R} - \tilde{R}_{\text{pre}}$. First, if $\tilde{R}$ becomes larger after the action than the current $\tilde{R}$, it is rewarded as much as the increased amount and the episode continues. Second, the episode ends when $\tilde{R}$ is less than the minimum required amount $R_0$ or less than the previous $\tilde{R}_{\text{pre}}$. Lastly, when the UAV’s altitude is out of the limited altitude, the episode ends with a disreward, $−1$. Algorithm 1 shows the overall SAC learning process. In Algorithm 1, the time complexity for training soft Q-functions $Q_{\phi}, Q_{\theta},$ and a policy $\pi_\phi$ is determined by the operations in an iteration. When hidden layers $H$ of each network have the identical hidden unit $u$, the time complexity in each iteration can be given by $O(h H^2 w^2)$. This paper used networks with 128 hidden units for two layers.

Algorithm 2 summarizes our proposed optimal strategy for IRS assisted UAV network. Inputs are trains position $P_{u,k'}$ over the time interval $K$ and BS position $p_b$, and the UAV’s initial position $p_{u,0}$. First, we derive the optimal indicator matrix $I_k'$ for the current position by the BILP method and perform the optimal phase-shift $\Phi_k'$ based on
Algorithm 2: Optimal Strategy for IRS-Assisted UAV Network.

Input: $p_{u,0}, p_{h}, P_{M} = \{p_{M,1}, p_{M,2}, \ldots, p_{M,K}\}, T_{max}$

1: for $k$ in $\{1, 2, \ldots, K\}$ do
2: for $t$ in $T_{max}$ do
3: Phase-shift $\Phi^*_k$ of IRS based on the optimal indicator matrix $A^*_k$ by BILP in (9)
4: Trajectory $T^*_k$ of UAV based on SAC model in (10)
5: if $\Delta R < \epsilon$ then
6: Break
7: end if
8: end for
9: end for

Output: $T, P$

### TABLE I

| Parameter | Value | Parameter | Value |
|-----------|-------|-----------|-------|
| $\rho_0$  | -20 dB| $B$       | 20 MHz|
| $\sigma^2$| -100 dBm| $p$     | 10 dBm|
| $\delta$  | 2.6   | $M$       | 4     |
| $N$       | 100 (10 x 10) | $v_{tra,im}$ | 50 m/s |
| $v_{max}$ | 45 m/s| $d_0$     | 1 m   |

It. Based on the pre-learned model, the optimal UAV Trajectory $T^*_k$ is derived. Repeat every time interval $k$ until the amount $\Delta R$ of change is smaller than $\epsilon$. Consequently, we can derive the optimal trajectory $T$ and phase-shift $P$ for all time $K$.

### IV. SIMULATION RESULTS

In this paper, we simulate within a communication range that one UAV can cover the speed of train. Moreover, a free-space path loss model was adopted because rural environments with less communication interference were considered. In order to check the simulation result, the train is moved on a fixed track, and a BS is placed in the center. The speed of the train used in the experiment moves constantly at 100 m/s, and it consists of a total of 4 trains. The IRS used in the experiment have a total of 100 reflectors, and the interval between reflectors is 0.01 m. Other definitions and values of variables used for communication can be found in Table I.

To compare the performance, we consider three benchmark algorithms as follows:
- **FU with SI**: Algorithm using the optimal phase-shift of IRS with the fixed trajectory of UAV
- **OU with RI**: Algorithm using the optimal trajectory of UAV with the random phase-shift of IRS
- **OU with SI**: Algorithm using the optimal trajectory of UAV with the static phase-shift of IRS

All of the algorithms were run in the same communication environment. In the FU with SI algorithm, the UAV was positioned at the center of the base station and the trajectory of trains, and the height was fixed at 200 m. And in the OU with SI algorithm, reflectors of IRS were all assigned the same number regardless of the location of trains.

Fig. 1 shows the efficiency and results of the algorithm used for optimization. In Fig. 1(a), we demonstrates the convergence of the proposed algorithm. From the figure we can see that our proposed algorithm converges within a few iterations. Therefore, we can identify that the result based on BCD method converges with the repetition of the proposed algorithm. Fig. 1(b) is results of the number of reflectors used in the TP for phase-shift. As the number of reflectors increases, it can be seen that the resulting time increases consistently. This shows that the method BILP we used is global optimization and this has a problem in which the computational volume increases consistently as the number of reflectors increases. Thus, we conducted within the number of reflectors that can derive the optimal solution in real-time. Fig. 1(c) and Fig. 1(d) are results during reinforcement learning. As we repeat the episode, we can see that the reward increases and the loss decreases. Through this, we can see that the algorithm was able to find the optimal location for increasing the minimum achievable rate. Furthermore, we experimented with network models being made up of 128 hidden units for two layers. As a result, it took less than 500 ms to derive the next trajectory from the input of the domain in a one-time slot. This result is a reasonable enough time to be mounted on an actual UAV and used practically. For this reason, we can confirm the possibility that our proposed algorithm is suitable to deploy in the real world problem and large-scale network.

Fig. 2 shows a simple scenario for implementing the proposed algorithm measurement. First, the black dotted line denotes a predetermined trains path. The green triangle (550, 0, 350) plays the role of a BS. The grey dotted line represents the fixed trajectory of UAV used in the **FU with SI** algorithm. Each route derives the optimal UAV location for the location of train at 1 s intervals. As a result, the path of the optimal UAV obtained through our proposed algorithm is a dotted line with an blue, magenta, red cross according to path loss exponents. As for the altitude of the UAV, it is effective to support communication at a low altitude when the trains and a BS are close, and at a high altitude when far away. This result is due to the loss that occurs according to the angle when
the phase-shift of IRS. Moreover, it can be seen that the difference in altitude according to path loss exponents is not large.

Fig. 3(a) is a graph showing the minimum achievable rate for each method over time. The proposed algorithm, OU with OI, shows the best results. The lowest performance is that IRS is randomly determined by OU with RI. Next, the case of using a fixed UAV and an optimized IRS is FU with OI. This method seems to be the result of not being able to move fluidly according to the positional movement of train. Next, OU with SI is the case of executing the IRS phase-shift in the same manner without considering the location of train. From OU with SI, we can see the importance of the optimal phase-shift in the optimal strategy. The graph shows that OU with SI has the most similar performance compared to the proposed method OU with OI. It can be concluded that UAV trajectory is a more important factor than IRS phase-shift in the optimal strategy. Also, we can predict that the more reflectors of the IRS, the greater the effect of the optimal IRS phase shift.

Fig. 3(b) is a graph of the data rate according to the altitude when the IRS phase-shift is optimized. When the altitude is 100 m and 300 m, the performance is poor, and when it is fixed at 200 m, it is close to the proposed algorithm, but shows a low value. As a result, it can be confirmed that the altitude is optimized by considering the locations of the base station and trains through the proposed algorithm.

V. Conclusion

In this paper, we proposed a novel IRS assisted UAV framework to provide stable communication services for trains. Then, we have investigated joint UAV trajectory optimization and optimal phase-shift design of IRS of the proposed model with the aim of maximizing the minimum achievable data rates of trains. To be tractable, the formulated problem is further decomposed into two subproblems. Later, we have applied BILP and SAC in order to solve the decomposed subproblems.

In the simulation, we confirm how efficiently our proposed system can support wireless communication for smart railways than stationary IRS or BS. The proposed algorithm derives the maximum 19.9% and 4% higher data rates, respectively, compared with the fixed IRS and fixed phase-shift.

REFERENCES

[1] B. Ai, A. F. Molisch, M. Rupp, and Z.-D. Zhong, “5G key technologies for smart railways,” Proc. IEEE, vol. 108, no. 6, pp. 856–893, Jun. 2020.
[2] Y. K. Tan, Y. M. Park, N. H. Tran, W. Saad, S. R. Pandey, and C. S. Hong, “Energy-efficient resource management in UAV-assisted mobile edge computing,” IEEE Commun. Lett., vol. 25, no. 1, pp. 249–253, Jan. 2021.
[3] Y. Zhou, “Future communication model for high-speed railway based on unmanned aerial vehicles,” Nov. 2014, arXiv:1411.3450.
[4] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, “Intelligent reflecting surface-aided wireless communications: A tutorial,” IEEE Trans. Commun., vol. 69, no. 5, pp. 3313–3351, May 2021.
[5] H. Zhang, B. Di, Z. Han, H. V. Poor, and L. Song, “Reconfigurable intelligent surface assisted multi-user communications: How many reflective elements do we need?” IEEE Wireless Commun. Lett., vol. 10, no. 5, pp. 1098–1102, May 2021.
[6] Q. Wu and R. Zhang, “Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts,” IEEE Trans. Commun., vol. 68, no. 3, pp. 1838–1851, Mar. 2020.
[7] X. Cao et al., “Reconfigurable intelligent surface-assisted aerial-terrestrial communications via multi-task learning,” IEEE J. Sel. Areas Commun., vol. 39, no. 10, pp. 3035–3050, Oct. 2021.
[8] S. Li, B. Duo, X. Yuan, Y. Liang, and M. Di Renzo, “Reconfigurable intelligent surface assisted UAV communication: Joint trajectory design and passive beamforming,” IEEE Wireless Commun. Lett., vol. 9, no. 5, pp. 716–720, May 2020.
[9] L. Ge, P. Dong, H. Zhang, J.-B. Wang, and X. You, “Joint beamforming and trajectory optimization for intelligent reflecting surfaces-assisted UAV communications,” IEEE Access, vol. 8, pp. 78702–78712, 2020.
[10] H. Lu, Y. Zeng, S. Jin, and R. Zhang, “Aerial intelligent reflecting surface: Joint placement and passive beamforming design with 3D beam flattening,” IEEE Trans. Wireless Commun., vol. 20, no. 7, pp. 4128–4143, Jul. 2021.
[11] W. Tang et al., “Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement,” IEEE Trans. Wireless Commun., vol. 20, no. 1, pp. 421–439, Jan. 2021.
[12] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, “Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization,” IEEE Trans. Wireless Commun., vol. 19, no. 10, pp. 6884–6898, Oct. 2020.
[13] L. D. Beal, D. C. Hill, R. A. Martin, and J. D. Hedengren, “GEKKO optimization suite,” Processes, vol. 6, no. 8, Jul. 2018, Art. no. 106.
[14] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in Proc. 35th Int. Conf. Mach. Learn., 2018, pp. 1861–1870.