UAV Trajectory, User Association, and Power Control for Multi-UAV-Enabled Energy-Harvesting Communications: Offline Design and Online Reinforcement Learning

Chien-Wei Fu, Graduate Student Member, IEEE, Meng-Lin Ku, Senior Member, IEEE, Yu-Jia Chen, Senior Member, IEEE, and Tony Q. S. Quek, Fellow, IEEE

Abstract—In this article, we consider multiple solar-powered wireless nodes (WNs) which utilize the harvested solar energy to transmit collected data to multiple unmanned aerial vehicles (UAVs) in the uplink. In this context, we jointly design UAV flight trajectories, UAV-node user association, and uplink power control to effectively utilize the harvested energy and manage co-channel interference within a finite time horizon. The design goal is to ensure the fairness of WNs by maximizing the worst user rate. The joint design problem is highly nonconvex and requires causal (future) knowledge of the instantaneous energy state information (ESI) and channel state information (CSI), which are difficult to predict in reality. To overcome these challenges, we propose an offline method based on convex optimization that only utilizes the average ESI and CSI, where line-of-sight (LOS) and non-LOS (NLOS) channels are considered. The problem is solved by three convex subproblems with successive convex approximation (SCA) and alternative optimization. We further design an online convex-assisted reinforcement learning (CARL) method based on real-time environmental information. An idea of multi-UAV regulated flight corridors, based on the optimal offline UAV trajectories, is proposed to avoid unnecessary flight exploration by UAVs and enables us to improve the learning efficiency and system performance, as compared with the conventional reinforcement learning (RL) method. Computer simulations are used to verify the effectiveness of the proposed methods. The proposed CARL method provides 25% and 12% improvement on the worst user rate over the offline and conventional RL methods.

Index Terms—Convex optimization, energy harvesting (EH), power control, reinforcement learning (RL), unmanned aerial vehicle (UAV) trajectory, UAV communication, user association.

I. INTRODUCTION

In the Internet of Things (IoT) era, many wireless communication nodes are deployed over wide areas for applications toward sustainable development, e.g., environmental monitoring. Traditional ways of powering electrical devices by connecting the power grids or nonrechargeable batteries with frequent replacement incur high deployment and maintenance costs. In recent years, energy harvesting (EH) technology has been recognized as an effective means to meet the energy requirement of electrical devices and prolong the lifetime of wireless networks. Due to its ubiquitous abundance, solar power is one of the most preferable EH sources and can provide a nearly permanent power supply. However, renewable energy sources are often affected by the environment, resulting in stochastic EH, and therefore, the power control design of communication nodes becomes an important issue for self-sustaining wireless services [1].

More recently, unmanned aerial vehicle (UAV) communications have received tremendous attention in various applications, such as data collection [2], wireless power transfer [3], [4], [5], [6], [7], relaying [8], and mobile edge computing [9], due to the potential to improve the transmission quality with the significant advantages of high mobility, flexible deployment, and low cost [10]. In contrast to the wireless networks with terrestrial infrastructures, the UAV can be flexibly deployed to collect sensing data from widely distributed nodes by dynamically adjusting its location. This can dramatically improve energy efficiency and facilitate the successful deployment of EH communications. Multi-UAVs can form multiple mobile base stations to improve network performance, but interference management becomes a critical challenge when multiple nodes communicate with multiple UAVs concurrently.

When multi-UAV-enabled communications are designed to serve multi-EH wireless nodes (WNs), several design challenges must be carefully addressed.

1) The user association between the UAVs and the WNs is crucial for mitigating the interference from the...
concurrently served multi-EH WNs to the multi-UAVs, either through appropriate scheduling of alternate transmissions or by selecting WNs with the least interference impact to other users. The user association is significant for EH WNs with limited energy.

2) The power control of EH WNs is subject to energy causality constraints, where the energy consumed by power control at any time should not exceed the total amount of energy collected from the past to the present, and battery storage constraints, where energy charging at any time is limited to the maximum battery capacity [13].

3) The UAV flight has initial and final flight position constraints, maximum flight speed constraints, flight altitude constraints, and safety distances between multiple UAVs for collision avoidance [14], [15], [16]. In particular, the flight paths of the multiple UAVs will interact with each other, which has a huge impact on the system performance. In this investigated scenario, the power control of solar EH nodes indeed has a strong relationship with the UAV trajectory and user association because the EH conditions of nodes are affected by the randomness of harvested energy and limited battery capacity, affecting the UAV’s choice of “whom to serve” and “where to fly” during a finite mission time. On the other hand, the positions of UAVs also strongly affect the transmit power of EH nodes with limited battery size. Hence, it is worth jointly optimizing these three coupling factors to capture the performance impact on multi-UAV-EH networks.

A. Literature Survey

Many researchers have contributed greatly to the topics of UAV deployment and resource management for UAV communications. There are some early works discussing the deployment of a single UAV. In [17], successive convex approximation (SCA) is used to optimize the UAV trajectory, transmit power, and subcarrier allocation for throughput maximization. In [2], a maximum-minimum data collection rate problem is solved for UAV wireless sensor networks in urban areas, where the 3-D UAV trajectory and transmission scheduling of sensors are jointly optimized by convex approximation. In [18], the UAV trajectory is optimized via Q-learning to improve energy efficiency while ensuring the QoS of ground users. For multiple UAVs, the interference problem becomes a key issue dominating wireless communication performance and requires new solutions to exploit its inherent spatial degree of freedom. Reference [19] maximizes the uplink transmission rate by designing multi-UAV cooperative transmission, subchannel allocation and UAV speed. In [20], UAV flight trajectories are investigated through a dueling deep Q-network to maximize downlink channel capacity under the line-of-sight (LOS) channel probability and user coverage constraint. The joint design of communication scheduling, power control, and UAV trajectory is proposed to maximize the minimum throughput in [15] and to minimize the weighted sum of aerial and ground costs in [21] via alternative optimization and SCA, while Shen et al. [22] focus on UAV flight speed, altitude, and power control problems.

Another line of work in the existing literature is exploiting EH in UAV communications to prolong the lifetime of wireless devices. In [3], UAVs with radio-frequency (RF) EH capability are utilized to assist mobile edge computing, which can provide users with edge computing services and energy through wireless charging. In [4], a UAV with EH is considered for uninterrupted service, where harvesting and charging time, flight trajectory and speed, and UAV’s transmit power allocation are jointly optimized by block coordinate descent and SCA methods. A dynamic fly-hover-transmission scheme is studied in [5] and [6] for UAV-assisted wireless energy and information transfer in cognitive radio networks, where a constrained Markov decision process (MDP) problem is cast to design the UAV transmission and trajectory based on causal system information for throughput maximization, while [6] proposes an efficient suboptimal but low-complexity transmission policy. In [7], a Q-learning method is used to find the flight policy for UAVs as power stations to maximize mission duration. However, the main drawback of wireless charging is the low-charging efficiency due to channel path loss. Also, in some hazardous areas, RF signals may not be readily available or dense enough. As an alternative, EH nodes with the self-sustainability from renewable energy are fascinating in UAV communications and enable UAVs to focus more on data collection and improve system performance [23], [24]. In [23], a single UAV with a fixed flight path is dispatched to collect data from multiple sensing nodes equipped with solar cells, which can extend network lifetime without human intervention. In [24], the UAV placement and resource allocation are investigated through deep learning in the renewable energy paradigm.

B. Motivation and Contributions

Although LOS/non-LOS (NLOS) and small-scale fading have been considered in the reinforcement learning (RL)-based online design of UAV-related works [20], they are not fully addressed in the convex optimization-based offline design [17], [21], [22]. The offline design of related works [3] and [15] ignores the influence of NLOS channels and only considers the LOS channels. In [19], LOS/NLOS is modeled in the offline design of UAV communication networks, while this work only considers subchannel allocation and UAV speed without joint optimization of UAV trajectory. LOS/NLOS channels are also considered in [2] and [25], but only one UAV is deployed and there is no interference problem. Besides, multi-UAV communications with EH nodes remain unexplored. Unlike the above works, our proposed offline design considers channel models with LOS/NLOS channels and small-scale fading for joint optimization of multi-UAV trajectories, power control and user association in multi-UAV-EH networks, and the SCA transformation methods of works [2] and [15] cannot be directly applied. Furthermore, we utilize the obtained optimal offline UAV trajectories to improve the performance of the subsequent online design, which helps resolve the UAV flight exploration problem in the RL.

Although SCA can effectively solve this joint design problem, better performance can be expected if the future
knowledge of the instantaneous channel state information (CSI) and energy state information (ESI) during the entire UAV mission period can be obtained in advance [2], [4], [15], [17], [19], [21], [22]. Some works have been proposed to predict the future instantaneous CSI or ESI in the short term [26], [27]. Nevertheless, it is tough to apply these short-term prediction methods to obtain long-term future knowledge of instantaneous CSI and ESI during the entire UAV mission period, especially in the dynamic environment of UAV-EH communications due to the maneuverability of UAVs and the randomness of harvested energy. The above reasons motivate us to propose an offline joint optimization problem that can capture the impact of LOS/NLOS and exploit only the average CSI and ESI in the SCA, which can be easily obtained from stochastic LOS/NLOS channel models and historical measurements of solar irradiance data, respectively.

Compared with the SCA method, the RL method is attractive for its dynamic response to real-time environmental states, i.e., NLOS/LOS and small-scale fading channel conditions and EH of nodes. It can only rely on the current instantaneous ESI and CSI and the past experience in the Q-table to decide the actions of UAV flight, power control and user association at the moment [7], [18], [20], [24]. The RL can effectively concern the dynamic variations of wireless channels and harvested energy, other than the average ESI and CSI in the offline design. This advantage motivates us to apply the RL method to solve the joint design problem. However, it lacks RL algorithms that scale well with the number of states and actions, thereby degrading the system performance. We thus leverage the proposed offline design to effectively guide the learning of an RL method, namely, convex-assisted RL (CARL), in our work. This newly adopted model enables us to reduce the number of system states by applying the optimal offline UAV flight trajectories as prior information, thereby improving the system performance.

To fill the aforementioned gaps, this article presents a framework for jointly designing multi-UAV flight trajectories, user association between UAVs and nodes, and uplink transmit power control of multi-UAV communication networks with solar-powered ground nodes. The main contributions of this article are summarized as follows.

1) To the best of our knowledge, this is the first work to investigate a multi-UAV communication network with multiple solar-powered ground nodes, with a focus on the resource management strategies and multi-UAV flight path planning to minimize the worst-case user rate.

2) A real solar power harvesting data set and a composite channel model, including LOS/NLOS and small-scale fading, are considered in this design. The joint design problem is nonconvex. Based on a series of SCAs, we propose an offline method for jointly optimizing multi-UAV flight trajectories, UAV-node user association and transmit power control, which only requires average CSI and ESI. Note that including LOS/NLOS in the joint optimization, becomes more complicated than those works that only consider LOS. Since the variations of small-scale fading and harvested energy are smaller than the average channel gain and harvested energy, such an optimization result implicitly indicates that the obtained UAV flight trajectories are of reference values, as compared to the optimal solution with the long-term causal knowledge of instantaneous CSI and ESI. This offline result provides valuable insights for designing online algorithms. For example, using this flight trajectory as a reference or initialization for online RL, we can avoid unnecessary exploration (e.g., locations that deviate from the offline flight paths of UAVs) and also reduce the probability of UAV collision during the learning process.

3) The online RL is then proposed by considering the current instantaneous CSI and ESI. With the optimal offline multi-UAV trajectories, the online CARL design presents a new concept of UAV flight corridor to effectively guide the executed actions and state exploration. By arranging flight corridors ahead with the offline UAV trajectories, the proposed online method can avoid exploring the state space randomly and effectively respond to the current instantaneous CSI and ESI for performance improvement.

4) Computer simulations are performed to demonstrate the effectiveness of the proposed offline and online methods. The proposed offline method performs better than other baseline schemes. Furthermore, the proposed CARL method significantly improves the performance, compared to the conventional RL and the proposed offline methods.

The remainder of this article is organized as follows. In Section II, we present the system model and the joint design problem of the UAV flight trajectory, user association, and power control of solar EH nodes, along with the MDP formulation. In Section III, an offline convex optimization method that utilizes only the average CSI and ESI is proposed. An online CARL method based on the proposed offline design is given in Section IV. Simulation results are provided in Section V, and Section VI concludes this article.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

Fig. 1 shows the multi-UAV communication networks with solar-powered ground nodes for uplink transmissions. A
centralized scenario is considered, in which UAVs collect the environmental information of nodes (including the CSIs and ESIs) and forward the information to the central control station. Consider a service area consisting of a group of $K$ WNs which utilize the harvested solar power for data transmission. A group of $M$ UAVs flies over the area to collect the data from the $K$ WNs. Both the UAVs and the WNs are only equipped with a single antenna. A time-slotted model is adopted, and we assume that the entire task period $T_s$ is divided into $N$ time slots, where there are $N+1$ discrete time instants ($n = 0, 1, \ldots, N$) and the time interval $\delta_D$ is defined as

$$\delta_D = \frac{T_s}{N}. \quad (1)$$

For simplicity, we define the sets $\mathcal{K} = \{1, \ldots, K\}$, $\mathcal{M} = \{1, \ldots, M\}$, and $\mathcal{N} = \{0, \ldots, N\}$. We assume that the positions of the WNs are unchanged, and the horizontal 2-D coordinate of the $k$th WN is given by

$$\mathbf{g}_k = [\bar{u}_k, \bar{v}_k]^T \in \mathbb{R}^2 \ \forall k \in \mathcal{K} \quad (2)$$

where $[\cdot]^T$ takes the vector transpose. It is assumed that the UAVs fly at a fixed altitude $H$, and the horizontal 2-D coordinate of the $m$th UAV at the time instant $n$ is given as

$$\mathbf{q}_m[n] = [\bar{x}_m[n], \bar{y}_m[n]]^T \in \mathbb{R}^2 \ \forall m \in \mathcal{M} \ \forall n \in \mathcal{N} \quad (3)$$

The trajectory coordinate of the UAVs is subject to the following:

$$\mathbf{q}_m[0] = \mathbf{q}_m[N] = \mathbf{q}_m^{ini} \ \forall m \in \mathcal{M} \quad (4)$$

$$\frac{1}{\delta_D} \| \mathbf{q}_m[n+1] - \mathbf{q}_m[n] \|_2 \leq V_{\text{max}} \quad \forall m \in \mathcal{M} \ \forall n \in \mathcal{N} \setminus \{N\} \quad (5)$$

$$\mathbf{q}_m[n] - \mathbf{q}_m[n] \|_2 \geq D_{\text{min}} \quad \forall n \in \mathcal{N} \setminus \{0, N\} \ \forall m, j \in \mathcal{M}, m \neq j \quad (6)$$

where $\mathbf{q}_m^{ini}$ is the initial position of the UAVs, $V_{\text{max}}$ is the maximum flight speed of the UAVs, and $D_{\text{min}}$ is the minimum safe distance for any two UAVs. We assume that the UAVs are equipped with finite batteries, and (4) stipulates that each UAV is required to fly back to the initial point $\mathbf{q}_m^{ini}$ at the end of the task period. Constraint (5) indicates that the flight speed of each UAV is limited to the maximum flight speed $V_{\text{max}}$. Moreover, (6) represents that the minimum safe distance $D_{\text{min}}$ must be maintained between any two UAVs to avoid collision.

For the channel model, both the LOS and NLOS channels are taken into consideration in the large-scale fading, and the path loss (in decibels) between the $m$th UAV and the $k$th WN at the time instant $n$ can be expressed as [28]

$$L_{\ell,m,k}[n] = 20 \log_{10} \left( \frac{4 \pi f_c d_{\ell,m,k}[n]}{c} \right) + \eta_c + S, \quad \forall m \in \mathcal{M} \ \forall k \in \mathcal{K} \ \forall n \in \mathcal{N} \ \forall \epsilon \in \{\text{LOS}, \text{NLOS}\} \quad (7)$$

where $f_c$ is the carrier frequency (Hz), $c$ is the speed of light (m/s), $\eta_c$ is an LOS and NLOS environment-related parameter, and $S$ represents the shadowing effect. Furthermore, the term $d_{m,k}[n]$ represents the distance between the $m$th UAV and the $k$th WN at the time instant $n$, given as

$$d_{m,k}[n] = \sqrt{\| \mathbf{q}_m[n] - \mathbf{g}_k \|^2 + H^2}. \quad (8)$$

According to [28], the probability of occurring the LOS channel between the $m$th UAV and the $k$th WN at the time instant $n$ can be modeled as

$$\rho_{\text{LOS},m,k}[n] = \frac{1}{1 + A \exp \left( -B \left( \frac{180}{\pi} \arctan \left( \frac{H}{\| \mathbf{q}_m[n] - \mathbf{g}_k \|_2} - A \right) \right) \right)} \quad \forall m \in \mathcal{M} \ \forall k \in \mathcal{K} \ \forall n \in \mathcal{N} \quad (9)$$

where the coefficients $A$ and $B$ depend on the operating environments [29], and the probability of occurring the NLOS channel can be computed as $\rho_{\text{NLOS},m,k}[n] = 1 - \rho_{\text{LOS},m,k}[n]$. By combining the large-scale and small-scale fading, we can obtain the channel gain between the $m$th UAV and the $k$th WN at the time instant $n$ as follows:

$$H_{m,k}[n] = \tilde{L}_{\ell,m,k} |\chi_{m,k}[n]|^2 \quad (10)$$

where $\tilde{L}_{\ell,m,k}[n]$ represents the linear scale of $L_{\ell,m,k}[n]$, $\chi_{m,k}[n] \sim \mathcal{CN}(0, 1)$ is complex Gaussian Rayleigh fading with zero mean and unit variance.

Next, we define the user association variable between the $m$th UAV and the $k$th WN at the time instant $n$ as $a_{m,k}[n]$, subject to the following:

$$a_{m,k}[n] \in \{0, 1\} \ \forall m \in \mathcal{M} \ \forall k \in \mathcal{K} \ \forall n \in \mathcal{N} \setminus \{N\} \quad (11)$$

$$\sum_{k=1}^{K} a_{m,k}[n] \leq 1 \ \forall m \in \mathcal{M} \ \forall n \in \mathcal{N} \setminus \{N\} \quad (12)$$

$$\sum_{m=1}^{M} a_{m,k}[n] \leq 1 \ \forall k \in \mathcal{K} \ \forall n \in \mathcal{N} \setminus \{N\} \quad (13)$$

where (11) means that $a_{m,k}[n]$ takes binary values. If $a_{m,k}[n] = 1$, the $k$th WN is associated with the $m$th UAV at the time instant $n$. Constraint (12) confines that each UAV can serve at most one WN at each time instant, while each active WN can be only served by one UAV under (13). Besides, if the $m$th UAV associates with one WN at the time instant $n$ for data transmissions, its position remains unchanged during the time interval $\delta_D$, leading to the following:

$$a_{m,k}[n] \cdot \| \mathbf{q}_m[n+1] - \mathbf{q}_m[n] \|_2 = 0 \quad \forall m \in \mathcal{M} \ \forall n \in \mathcal{N} \setminus \{N\}. \quad (14)$$

Since the WNs utilize the same frequency band for uplink communications, each UAV suffers from the uplink multiuser interference problem. The original signal-to-interference plus noise power ratio (SINR) of the $k$th WN at the $m$th UAV and the $n$th time instant can be expressed as

$$\tilde{\Gamma}_{m,k}[n] = \frac{a_{m,k}[n] P_m[n] H_{m,k}[n]}{\sum_{i=1, i \neq k}^{K} \left( \sum_{j=1}^{M} a_{j,i}[n] P_j[n] H_{j,i}[n] \right) + \sigma_n^2} \quad \forall m \in \mathcal{M} \ \forall k \in \mathcal{K} \ \forall n \in \mathcal{N} \setminus \{N\} \quad (15)$$
where $\sigma_n^2$ is the power of additive white Gaussian noise, and $P_k[n]$ is the uplink transmit power of the $k$th WN at the time slot $n$. With the goal of maximizing the worst user rate and the user association constraints (11)–(13), the achievable sum rate of the $k$th WN can be equivalently reformulated as

$$R_k = \sum_{n=0}^{N-1} \sum_{m=1}^{M} W \log_2 \left( 1 + \Gamma_{m,k}[n] \right)$$

where $W$ is the system bandwidth, $R_{k,n}$ is the data rate of the $k$th WN at the time instant $n$, and we define

$$\Gamma_{m,k}[n] = \frac{P_k[n] H_{m,k}[n]}{\sum_{l=1, i \neq m}^K P_l[n] H_{m,l}[n] + \sigma_n^2} \forall m \in \mathcal{M}, \forall k \in \mathcal{K}, \forall n \in \mathcal{N} \setminus \{N\}. \quad (16)$$

The equivalent can be explained as follows. Assume that the $i$th WN is not associated with any UAVs, i.e., $a_{i,j}[n] = 0$ for all $j$. Then, the optimal power control solution of the $i$th WN must occur at $P_i[n] = 0$; otherwise (i.e., $P_i[n] \neq 0$), the $i$th WN wastes its harvested energy without improving its rate performance but causes extra interference power to degrade other users’ rate performance. Hence, the equivalent relationship (16) holds under the assumption of the design goal of worst user rate maximization and the user association constraints, even though the user association variables in $\Gamma_{m,k}[n]$ are moved outside the logarithm function. Notice that a similar formulation trick can be found in [30], which deals with load balancing in cellular networks.

Since each WN relies on the harvested solar energy in a finite-capacity battery for uplink transmission, the transmit power of a WN is constrained by its harvested energy and battery capacity. Let $E_k[n]$ and $B_k[n]$ represent the harvested energy and battery level of the $k$th WN at the time instant $n$. For simplicity, we assume that the battery level of WN is zero at the time instant $n = 0$. Therefore, the residual energy in the battery of the $k$th WN at the time instant $n$ can be described as $B_k[n] = \sum_{l=0}^n E_k[l] - \delta_D \sum_{l=0}^n P_k[l]$, and the battery power evolution from the time $n$ to $(n + 1)$ can be described as $B_k[n + 1] = B_k[n] + E_k[n + 1]$. Since the battery power is limited by $0 \leq B_k[n] \leq B_{max}$, it yields two constraints about the uplink transmit power, harvested energy, and battery power, namely, energy causality and battery storage constraints

$$\delta_D \sum_{l=0}^n P_k[l] \leq \sum_{l=0}^n E_k[l] \forall k \in \mathcal{K}, \forall n \in \mathcal{N} \setminus \{N\}. \quad (18)$$

The power control variables can in part indicate the association state of WNs. Specifically, $P_k[n] \neq 0$ means that the $k$th WN is served by a UAV, but unfortunately, it does not indicate which UAV the $k$th WN is associated with. Besides, the power control variables take continuous values and cannot be directly used to describe the user association constraints, i.e., each UAV can only serve one node while each node can only be served by one UAV at each time instant. For these reasons, it is desirable to include user association variables in the joint optimization problem.

In order to achieve a fair communication service for multiple WNs during the entire UAV task period, the design goal of this article is to maximize the worst sum rate among $K$ WNs. The joint design problem of the trajectories of UAVs, user association between UAVs and WNs, and transmit power of WNs with solar EH can be formulated as follows:

(P1) \[ \max_{\{u_{m}[n], P_k[n], a_{m,k}[n] \} \in \mathcal{K}, \forall n \in \mathcal{N} \setminus \{N\}} \min_k \sum_{l=0}^n \sum_{k=1}^K \sum_{i=1}^N \log_2 \left( 1 + \Gamma_{m,k}[n] \right) \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N} \setminus \{N\} \]

s.t. (4), (5), (6), (11), (12), (13), (14), (18), (19).

In this optimization problem, the UAV needs to know the CSI $H_{m,k}[n]$ and the ESI $E_k[n]$ not only in the current time instant but also in the future time instants. However, acquiring the complete (past and future) information for the optimization problem is impractical in real applications. In response to this design challenge, an RL method can be utilized to perform dynamic optimization for the UAV flight direction, user association, and transmit power, solely based on current state systems. The conventional RL is exempt from future instantaneous ESIs and CSIs but only depends on the current state information. However, it requires the UAV to repeatedly take various actions in each state, which results in a long exploration learning time and a slow update of $Q$-values. The learning time and convergence performance become even worse for multi-UAV communication networks with multiple-EH WNs, since the number of system states and actions increases exponentially and many inefficient actions may be executed during learning. For this reason, we will first investigate an offline method to find the best offline flight trajectory for the UAVs by only applying the “average” LOS/NLOS channel gain and EH profile. The obtained solution is then served as the reference solution for the online RL. The main idea is to mark out a flight corridor based on the reference solution to guide the UAV flight actions in online learning for reducing the learning time of the conventional RL while improving the performance of the offline approach.

III. OFFLINE CONVEX OPTIMIZATION DESIGN OF MULTI-UAV SYSTEMS WITH MULTIPLE EH WNS

We develop an offline approach by using the average LOS/NLOS channel gain on the UAV trajectory $\hat{H}_{m,k}[n] = \mathbb{E}[H_{m,k}[n]]$ and the average value of the past solar EH time series $\hat{E}_k[n] = \mathbb{E}[E_k[n]]$ to replace the instantaneous information $H_{m,k}[n]$ and $E_k[n]$ in the problem (P1), respectively. It is worth noting that although the statistical average of solar EH profiles is difficult to obtain, it can be
acquired by numerically averaging real solar energy historical data. Besides, from (7)–(10), the average LOS/NLOS channel gain $\hat{H}_{m,i}[n]$ can be derived in terms of the UAV position $\mathbf{q}_m[n]$ as

$$
\hat{H}_{m,i}[n] = \rho_{\text{LOS},m,i}[n] \hat{L}_{\text{LOS},m,i}[n] + \rho_{\text{NLOS},m,i}[n] \hat{L}_{\text{NLOS},m,i}[n]
$$

where $C_1 = \frac{1}{1 + A \exp\left(-B \left(\frac{180}{\pi} \tan^{-1}\left(\frac{H}{\|\mathbf{q}_m[n] - \mathbf{g}_i}\|_2\right) - A\right)\right)}$ and $\bar{C}_3 = \frac{1}{1 + A \exp\left(-B \left(\frac{180}{\pi} \tan^{-1}\left(\frac{H}{\|\mathbf{q}_m[n] - \mathbf{g}_i\|_2\right) - A\right)\right)}$.

\[ \hat{\mathbf{G}}_m = \frac{\hat{H}_{m,i}[n]}{\sum_{i=1, i\neq k}^{K} \hat{P}_i[n] \hat{H}_{m,i}[n] + \sigma_n^2} \]

The mechanism by which UA Vs determine associated nodes is to maximize the worst sum rate among nodes during the entire UAV mission, i.e., $\min_{k=1, \ldots, K} \tilde{R}_k$. Given the transmit power control $P_k[n]$ of the WNs and the UAV flight trajectory $\mathbf{q}_m[n]$ in (P2), the user association subproblem can be formulated as an equivalent epigraph form by relaxing the integer constraint (11) and introducing an auxiliary variable $\zeta_a$

$$
\text{max} \{\zeta_a, a_{m,k}[n] \forall m,k,n\} \zeta_a
$$

\[ \text{s.t.} \sum_{n=0}^{N-1} \sum_{n=1}^{M} a_{m,k}[n] W \log_2(1 + \bar{\Gamma}_{m,k}[n]) \geq \zeta_a \]

\[ 0 \leq a_{m,k}[n] \leq 1 \]

\[ \forall m \in \mathcal{M}, \forall k \in \mathcal{K}, \forall n \in \mathcal{N} \setminus \{N\} \]

A. UAV-WN User Association Subproblem

Given the user association $a_{m,k}[n]$ and the power control $P_k[n]$ of WNs, the UAV flight trajectory subproblem can be obtained from the problem (P2) and represented as an epigraph form with an auxiliary variable $\zeta_q$

$$
\text{max} \{\zeta_q, \mathbf{q}_m[n] \forall m,n\} \zeta_q
$$

\[ \text{s.t.} \tilde{R}_k \geq \zeta_q \forall k \in \mathcal{K} \]

\[ (4), (5), (6), (11), (12), (13), (14), (18), (19). \]

Since the objective function of the joint design problem (P2) and (6), (11), and (14) are nonconvex for the three design variables $\mathbf{q}_m[n], a_{m,k}[n]$ and $P_k[n]$, we adopt the alternative optimization to decompose this problem into three subproblems to optimize the UAV flight trajectory $\{q_m[n] \forall m,n\}$, user association $\{a_{m,k}[n] \forall m,k,n\}$, or power control $\{P_k[n] \forall k\}$ under the fixed values of other variables. Nevertheless, the three subproblems are still nonconvex.

1) The user association subproblem is a nonconvex mixed integer programming, and we will relax the integer constraint (11), i.e., $\{a_{m,k}[n] \in \{0, 1\} \forall m,k,n\}$ to $\{0 \leq a_{m,k}[n] \leq 1 \forall m,k,n\}$ to convert the problem into a linear programming problem.

2) The UAV flight trajectory subproblem is also nonconvex due to the objective function and (6), and we will use SCA methods to transform these two nonconvex subproblems into convex ones.

B. UAV Flight Trajectory Subproblem

Given the user association $a_{m,k}[n]$ and the power control $P_k[n]$ of WNs, the UAV flight trajectory subproblem can be obtained from the problem (P2) and represented as an epigraph form with an auxiliary variable $\zeta_q$

$$
\text{max} \{\zeta_q, \mathbf{q}_m[n] \forall m,n\} \zeta_q
$$

\[ \text{s.t.} \tilde{R}_k \geq \zeta_q \forall k \in \mathcal{K} \]

\[ (4), (5), (6), (11), (12), (13), (14), (18), (19). \]

Observing (28), we know that the SINR $\bar{\Gamma}_{m,k}[n]$ in (25) contains $\hat{H}_{m,i}[n]$ which is related to the multiple UAV flight trajectory variables $\mathbf{q}_m[n]$. This makes (28) nonconvex. In addition, the minimum safe distance constraint (6) is also nonconvex. As a result, the subproblem (P4) is a nonconvex problem, which cannot be directly solved by convex optimization tools. In the following, we resort to the SCA method to convexify the nonconvex constraints.

First, we expand the left-hand side of (28) into the difference of two logarithmic functions, as shown in (29) at the bottom of the next page, where we define $\tilde{R}_{m,n} = \log_2\left(\sum_{i=1}^{K} P_i[n] H_{m,i}[n] + \sigma_n^2\right)$ and $\bar{R}_{2m,k}[n] = -\log_2\left(\sum_{i=1, i\neq k}^{K} P_i[n] H_{m,i}[n] + \sigma_n^2\right)$. Our goal is to find a lower bound concave function for $\tilde{R}_k$ in (29) in order to transform (28) into solvable convex constraints. Below, we elaborate on the ways to find concave lower bounds for $\tilde{R}_{m,n}$ and $\bar{R}_{2m,k}[n]$ in terms of $\mathbf{q}_m[n]$.  

1) Concave Lower Bound for $\tilde{R}_{m,n}$: We define two variables $X_{m,i}$ and $Y_{m,i}$, given by

$$
X_{m,i} = \|\mathbf{q}_m[n] - \mathbf{g}_i\|_2^2 + H^2
$$

\[ \forall m \in \mathcal{M}, \forall i \in \mathcal{K}, \forall n \in \mathcal{N} \setminus \{N\} \]

2) The UAV flight trajectory subproblem is also nonconvex due to the objective function and (6), and we will use SCA methods to transform these two nonconvex subproblems into convex ones.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Lemma 1: The function $R_{1m}[n] = \log_2(\sum_{i=1}^{K} P_i[n] H_{m_i}[n] + \sigma_n^2)$ is convex in $X_{m_i}[n]$ and $Y_{m_i}[n]$ for all $i$.

Proof: See Appendix A for the detailed proof.

By using Lemma 1, a theorem for finding a concave lower bound of $R_{1m}[n]$ is provided as follows.

Theorem 1: Given any $q'_m[n]$, the function $R_{1m}[n]$ is lower bounded by a concave function $R_{lb}[n]$ in terms of $q_m[n]$

$$R_{lb}[n] \geq \log_2 \left( \sum_{i=1}^{K} P_i[n] \left( \frac{C_1 C_2}{X_{m_i}[n]} + \frac{C_3}{Y_{m_i}[n]} \right) + \sigma_n^2 \right)$$

$$+ \sum_{i=1}^{K} O_{m_i}[n] \left( X_{m_i}[n] - X_{m_i}[n] \right)$$

$$+ \sum_{i=1}^{K} G_{m_i}[n] \left( Y_{m_i}[n] - Y_{m_i}[n] \right)$$

$$\triangleq R_{lb}[n] \quad \forall m \in M \quad \forall n \in N \setminus \{N\}$$

where we define

$$X_{m_i}[n] = \|q_m[n] - g_i\|^2 + H^2;$$

$$Y_{m_i}[n] = 1 + Ae^{-b \left( \frac{180}{\pi} \tan^{-1} \left( \frac{H}{U_{m_i}[n]} \right) \right) - A}$$

$$O_{m_i}[n] = \frac{\left( \sum_{j=1}^{K} P_j[n] \frac{C_1 C_2 + C_3 Y_{m_i}[n]}{X_{m_i}[n]} \right)^2 + \sigma_n^2}{\ln(2)}$$

$$G_{m_i}[n] = \frac{\left( \sum_{j=1}^{K} P_j[n] \frac{C_1 C_2 + C_3}{X_{m_i}[n]} \right)^2 + \sigma_n^2}{\ln(2)}$$

$$Y_{m_i}^u[n] = 1 + \exp \left( -b \left( \frac{180}{\pi} \tan^{-1} \left( \frac{H}{U_{m_i}[n]} \right) \right) - A \right)$$

$$U_{m_i}^u[n] = \|q_m[n] - g_i\|_2^2.$$
these constraints and transform them into convex ones. Since \( \|q_m[n] - q_j[n]\|^2_2 \) is convex in \( q_m[n] \) and \( q_j[n] \), the first-order Taylor expansion of \( \|q_m[n] - q_j[n]\|^2_2 \) at given points \( q_m[n] \) and \( q_j[n] \) can be derived as a lower bound

\[
\|q_m[n] - q_j[n]\|^2_2 \geq \|q_m'[n] - q_j'[n]\|^2_2 + 2(q_m'[n] - q_j'[n])^T(q_m[n] - q_m'[n]) + 2(q_j'[n] - q_j[n])^T(q_j[n] - q_j'[n]),
\]

From (44), (6) is then replaced by

\[
D_{min}^2 \leq \|q_m'[n] - q_j'[n]\|^2_2 + 2(q_m'[n] - q_j'[n])^T(q_m[n] - q_m'[n]) + 2(q_j'[n] - q_j[n])^T(q_j[n] - q_j'[n]),
\]

The solutions that satisfy (45) must also satisfy (6) due to the lower bound relationship in (44).

Similarly, the function \( \|q_m[n] - g_i\|^2_2 + H^2 \) in (39) is convex in \( q_m[n] \), and its lower bound is obtained by the first-order Taylor expansion at a given point \( q_m'[n] \), as follows:

\[
\|q_m[n] - g_i\|^2_2 + H^2 \geq \|q_m'[n] - g_i\|^2_2 + H^2 + 2(q_m'[n] - g_i)^T(q_m[n] - q_m'[n]).
\]

By using (46), (39) is then replaced by

\[
\bar{X}_{m,i}[n] \leq \|q_m'[n] - g_i\|^2_2 + H^2 + 2(q_m'[n] - g_i)^T(q_m[n] - q_m'[n])
\]

Because the right-hand side of (40) is neither convex nor concave for \( q_m[n] \), the next step is to convexify (40) with the following theorem.

**Theorem 3:** Constraint (40) can be convexified by

\[
\tilde{X}_{m,i}[n] \leq 1 + Ae^{-B(\theta_{m,i}[n]-\lambda)} + \left(-ABe^{B(\lambda-\theta_{m,i}[n])}\right) \times (\theta_{m,i}[n] - \theta_{m,i}'[n]),
\]

\[
\forall n \in N \setminus \{0, N\} \forall m \in M \forall i \in K
\]

\[
\theta_{m,i}[n] \geq \frac{180}{\pi} \tan^{-1} \left( \frac{H}{\bar{U}_{m,i}[n]} \right)
\]

\[
\forall n \in N \setminus \{0, N\} \forall m \in M \forall i \in K
\]

\[
\bar{U}_{m,i}[n] \leq \|q_m'[n] - g_i\|^2_2 + 2(q_m'[n] - g_i)^T(q_m[n] - q_m'[n])
\]

\[
\forall n \in N \setminus \{0, N\} \forall m \in M \forall i \in K
\]

where \( \theta_{m,i}[n] \) and \( \bar{U}_{m,i}[n] \) are auxiliary variables.

**Proof:** See Appendix F for the detailed proof.

In summary, the convex constraints (45) and (47) convexify (6) and (39) in the subproblem (P5), respectively, while the convex constraints (48), (49), and (50) are able to convexify (40) via Theorem 3. Accordingly, the UAV flight trajectory subproblem (P5) can be rewritten as

\[
\text{(P6)} \quad \max_{\{q_q, R_q, \bar{R}_q \}} \zeta_q
\]

s.t. (4), (5), (14), (43), (45), (47), (48), (49), (50).

The UAV flight trajectory subproblem (P6) now becomes a solvable convex optimization problem under a given reference point \( q_m'[n] \), and the optimal value of the UAV trajectory can be iteratively obtained by using the CVX tool [31] with SCA methods [33]. Note that the obtained optimal solution of the subproblem (P6) is a lower bound of the subproblem (P4), since the feasible set of (P4) contains that of (P6).

C. Transmit Power Control Subproblem

Given the user association \( a_{m,k}[n] \) and the UAV flight trajectory \( q_m[n] \), the transmit power control subproblem for WNs can be obtained from the original optimization problem (P2) by using the epigraph form and introducing an auxiliary variable \( \zeta_p \), as follows:

\[
\text{(P7)} \quad \max_{\{q_p, R_p \}} \zeta_p
\]

s.t. \( R_k \geq \zeta_p \forall k \in K \) (51) (18), (19).

In this subproblem, the objective function and (18) and (19) are affine and thus convex in \( P_k[n] \). However, from (29), it can be found that the user rate \( R_k \) is the sum of concave functions and convex functions, and thus \( R_k \) in (51) is neither convex nor concave in terms of the variables \( P_k[n] \). To solve this problem, we convexify (51) by the first-order Taylor expansion. From (29), it is known that \( \hat{R}_{2m,k}[n] \) is a convex function in \( P_k[n] \), and we can get the following lower bound relationship:

\[
\tilde{R}_{2m,k}[n] \geq -\log_2 \left( \sum_{i=1, i \neq k}^{K} P_i[n] \bar{H}_{m,i}[n] + \sigma_n^2 \right) - \sum_{i=1, i \neq k}^{K} \ln(2) \left( \sum_{j=1, j \neq k}^{K} P_j[n] \bar{H}_{m,j}[n] + \sigma_n^2 \right)
\]

\[
\times (P_k[n] - \tilde{P}_k^i[n]) \equiv \hat{R}_{2m,k}^i[n]
\]

\[
\forall n \in N \setminus \{0, N\} \forall m \in M \forall k \in K
\]

where \( \tilde{P}_k^i[n] \) is a reference point for the Taylor expansion. Accordingly, the user rate \( R_k \) in (29), i.e., the left-hand side of (51), can be lower bounded by a concave function, as follows:

\[
\hat{R}_k \geq \sum_{n=0}^{N-1} a_{m,k}[n] W \left( \tilde{R}_{1m,k}[n] + \hat{R}_{k}^{\bar{b}} \right) \forall k \in K.
\]

Thus, the subproblem (P7) can be transformed into a convex one by replacing (51) with the lower bound (53)
Note that the optimization result of (P8) is a lower bound of (P7). Given a reference point \( P^*_k[n] \), the transmit power control subproblem (P8) can be iteratively solved by the optimization tool, e.g., CVX, with SCA methods.

After the above transformation, the offline joint design problem (P2) can be alternatively solved by the three convex subproblems (P3), (P6), and (P8), and the proposed offline algorithm is summarized in Algorithm 1. For convenience, we assume that the superscript \( r \) of the reference points \( \mathbf{q}^*_m[n], P^*_k[n], a^*_{m,k}[n] \) also refers to the iteration index in the successive convex optimization (SCO). We first initialize \( r = 0 \), and initialize \( \mathbf{q}^*_m[n], P^*_k[n], a^*_{m,k}[n] \). Let \( \epsilon > 0 \) be a threshold for the stopping criterion. In the user association subproblem (P3), the optimal solution to the user association \( a^*_{m,k}[n] \) is obtained under the given values of UAV trajectory \( \mathbf{q}^*_m[n] \) and power control \( P^*_k[n] \), and then we update \( a^*_{m,k}[n] \). Second, the UAV flight trajectory subproblem (P6) is solved under the given values of user association \( a^*_{m,k}[n] \) and power control \( P^*_k[n] \) through the SCA. In the inner loop of the SCO, we update \( \mathbf{q}^*_m[n] \) with the last result of \( \mathbf{q}^*_m[n] \), and subsequently renew the related constraints that contain \( \mathbf{q}^*_m[n] \). Now the new optimal solution \( \mathbf{q}^*_{m,i}[n] \) can be obtained by solving the subproblem (P6). The steps in the loop are repeated until the convergence is achieved, and we update \( \mathbf{q}^*_{m,i}[n] = \mathbf{q}^*_{m,i}[n] \). Finally, the power control subproblem (P8) is solved under the given values of UAV trajectory \( \mathbf{q}^*_{m,i}[n] \) and user association \( a^*_{m,k}[n] \), where we update \( P^*_k[n] \) and (54) with the last result of \( P^*_k[n] \) in the inner loop of the SCO. We repeat these steps in the inner loop until the convergence is attained, and update \( P^*_{k,i}[n] = P^*_{k,i}[n] \). Then we update the iteration number of the outer loop for the next round. The outer loop is stopped when the increase of the objective value is smaller than the preset threshold \( \epsilon \). Algorithm 1 does not require the instantaneous perfect CSI, but only the average LOS/NLOS channel gains based on the stochastic channel model.

**D. Convergence of the Offline Algorithm**

The convergence of the proposed offline algorithm for UAV flight trajectory, WN user association and power control is analyzed as follows. Let \( f(a_{m,k}[n], \mathbf{q}_m[n], P_k[n]) \) be the objective function of the original problem (P2). First, in the user association subproblem (P3), by fixing \( \mathbf{q}^*_m[n], P^*_k[n] \) to obtain the optimal solution of the user association \( a^*_{m,k}[n] \), it results in

\[
f(a^*_{m,k}[n], \mathbf{q}^*_m[n], P^*_k[n]) \leq f(a^*_{m,k}[n], \mathbf{q}^*_m[n], P^*_k[n]).
\] (55)

Next, we discuss the convergence of the inner loop of the UAV flight trajectory subproblem (P6). Given \( \{a^*_{m,k}[n], \mathbf{q}^*_m[n], P^*_k[n]\} \), the subproblem is optimized by the

**Algorithm 1 Offline Alternative Optimal Algorithm for Problem (P2)**

1. Initialize \( \{a^*_{m,k}[n], \mathbf{q}^*_m[n], P^*_k[n]\} \) \( \forall m,k,n \), \( r = 0 \)
2. Set \( \epsilon > 0 \)
3. **while** Increase of the objective value \( < \epsilon \) **do**
   4. For given \( \{\mathbf{q}^*_m[n], P^*_k[n]\} \) \( \forall m,k,n \), find the optimal user association solution of the problem (P3) as \( a^*_{m,k}[n] \).
   5. For given \( \{a^*_{m,k}[n], \mathbf{q}^*_m[n], P^*_k[n]\} \) \( \forall m,k,n \).
   6. **repeat** (SCA for solving problem (P6))
      7. Update \( \mathbf{q}^*_m[n] \) using the last result \( \mathbf{q}^*_m[n] \).
      8. Update (43), (45), (47), (50) and (50) using \( \mathbf{q}^*_m[n] \).
      9. Find the new optimal solution \( \mathbf{q}^{**(r+1)}_m[n] \) by solving the problem (P6).
   10. **until** convergence to the optimal UAV flight trajectory solution \( \mathbf{q}^{**(r+1)}_m[n] \)
   11. Set \( \mathbf{q}^{**(r+1)}_m[n] \leftarrow \mathbf{q}^{**(r)}_m[n] \).
   12. For given \( \{a^*_{m,k}[n], \mathbf{q}^{**(r+1)}_m[n], P^*_k[n]\} \) \( \forall m,k,n \).
   13. **repeat** (SCA for solving problem (P8))
      14. Update \( P^*_k[n] \) using the last result \( P^*_k[n] \).
      15. Update (54) using \( P^*_k[n] \).
      16. Find the new optimal solution \( P^*_{k,i}[n] \) by solving the problem (P8).
   17. **until** convergence to the optimal power control strategy solution \( P^*_{k,i}[n] \)
   18. Set \( P^*_{k,i}[n] \leftarrow P^*_{k,i}[n] \).
   19. Update \( r \leftarrow r + 1 \).
20. **end while**

SOCO to obtain the trajectory \( \mathbf{q}^{**(r+1)}_m[n] \). In the \( i \)th inner loop of the SCO, we assume that \( \mathbf{q}^{(i)}_m[n] \) is the obtained solution with respect to the given first-order Taylor expansion reference point \( \mathbf{q}^{(i)}_m[n] \). Further, denote \( \mathbf{q}^{(i+1)}_m[n], \mathbf{q}^{(i+1)*}_m[n], P^*_{k,i}[n] \) as the worst sum rate for the obtained solution \( \mathbf{q}^{(i+1)*}_m[n] \) at the reference point \( \mathbf{q}^{(i)}_m[n] \) in the subproblem (P6). Then we have

\[
x^{(i+1)}_m[n] \left( a^*_{m,k}[n], \mathbf{q}^{(i+1)*}_m[n], P^*_{k,i}[n] \right)
\leq f \left( a^*_{m,k}[n], \mathbf{q}^{(i+1)*}_m[n], P^*_{k,i}[n] \right)
\leq \xi^{(i+1)}_m[n] \left( a^*_{m,k}[n], \mathbf{q}^{(i+1)*}_m[n], P^*_{k,i}[n] \right)
\leq f \left( a^*_{m,k}[n], \mathbf{q}^{(i+1)*}_m[n], P^*_{k,i}[n] \right)
\] (56)

where the first inequality comes from the lower bound relationship in (42), and the second equality is because the equality of the lower bound relationship holds at the tangent point. Moreover, the third inequality is because we set \( \mathbf{q}^{(i+1)}_m[n] = \mathbf{q}^{(i)}_m[n] \) to update the reference point for the next inner iteration according to the SCA and obtain the corresponding optimal solution \( \mathbf{q}^{(i+1)*}_m[n] \). Hence, it concludes that the worst sum rate performance can be monotonically increased in the inner loop of the UAV flight trajectory subproblem (P6), and we can get the following relationship for UAV flight trajectory in the \( r \)th outer loop

\[
f \left( a^*_{m,k}[n], \mathbf{q}^{(r+1)}_m[n], P^*_{k,i}[n] \right)
\leq f \left( a^*_{m,k}[n], \mathbf{q}^{(r+1)*}_m[n], P^*_{k,i}[n] \right).
\] (57)
Likewise, we can apply the similar derivation of (56) to prove that the sum rate performance can be monotonically increased in the inner loop of the SCA for the power control subproblem (P8). Thus, under the given \([d_{m,k}^{+1}[n], q_m^{+1}[n], P_k^{(0)}[n]]\), it implies that
\[
\zeta_{p_k^{(0)}}[n]\left(d_{m,k}^{+1}[n], q_m^{+1}[n], P_k^{(0)}[n]\right)
\leq \zeta_{p_k^{+1}}[n]\left(d_{m,k}^{+1}[n], q_m^{+1}[n], P_k^{(0)}[n]\right)
\]
(58)
where \(\zeta_{p_k^{(0)}}[n]\left(d_{m,k}^{+1}[n], q_m^{+1}[n], P_k^{(0)}[n]\right)\) is referred to as the worst sum rate for the obtained solution \(P_k^{(0)}[n]\) at the reference point \(P_k^{(0)}[n]\) in the \(i\)th inner loop of the power control subproblem (P8). As a result, it implies that
\[
f\left(d_{m,k}^{+1}[n], q_m^{+1}[n], P_k^{(0)}[n]\right)
\leq f\left(d_{m,k}^{+1}[n], q_m^{+1}[n], P_k^{(0)}[n]\right).
\]
(59)
Due to the alternative optimization for the three subproblems, it can be concluded from (55), (57), and (59) that the performance of the proposed algorithm can be monotonically increased, and the local optimal solution of the original problem (P2) can be found until the algorithm is converged.

E. Complexity of the Offline Algorithm

In Algorithm 1, since the user association of (P3), the UAV flight trajectory of (P6) and the transmit power control of (P8) are sequentially optimized in each iteration by using the CVX solver based on the interior-point method, their individual complexities scale as \(O((MNK)^{3.5}\log(\epsilon_o^{-1}))\), \(O((MN + 4M(N - 1)K)^{3.5}\log(\epsilon_o^{-1}))\) and \(O((KN)^{3.5}\log(\epsilon_o^{-1}))\), respectively, where \(\epsilon_o > 0\) represents the solution accuracy [34].

IV. CONVEX-ASSISTED REINFORCEMENT LEARNING

In this section, we propose a CARL approach, in which a flight corridor is marked out, based on the proposed offline design in Algorithm 1, to guide the UAV flight actions in an online learning fashion. Besides, the actions of the RL agents can be restricted, and the number of system states can be efficiently reduced through the assistance of offline design. The system states, actions and real-time rewards are designed and presented in the following.

A. System States

Let \(\mathcal{S} = \mathcal{L} \times \mathcal{H}\) be a two-tuple state space, where \(\times\) denotes the Cartesian product, \(\mathcal{L}\) represents a UAV location state set, and \(\mathcal{H}\) is a channel state set. Moreover, we define a random variable \(s = (l, h) \in \mathcal{S}\) as the system stochastic state of the MDP. It is assumed that the UAV location and channel remain steady during the time interval \(\delta_T\). The detailed definition of each state is specified in the following.

1) UAV Location State: Assume that a square UAV coverage region is quantized into \(N_L\) lattice points with a scale of \(\Delta\) (the minimum distance between two adjacent horizontal/vertical lattice points), and the UAV location state space is defined as \(\mathcal{L} = \mathcal{L}_x \times \mathcal{L}_y\), where \(\mathcal{L}_x = \mathcal{L}_y = \{0, 1, \ldots, \sqrt{N_L} - 1\}\). When the location state of the \(m\)th UAV at the time instant \(n\) is given as \(\mathbf{u}_m = [\theta_m^x, \theta_m^y] \in \mathcal{L}\), it means that the horizontal coordinate of the \(m\)th UAV at the time instant \(n\) is
\[
\mathbf{q}_m[n] = \left[\Delta \times \tilde{\theta}_m^n + \Delta \times \tilde{\gamma}_m^n + \Delta \right]^T.
\]
(60)
Let \(\mathbf{q}_m[n]\) be the offline horizontal flight path of the \(m\)th UAV at the time instant \(n\), obtained by the offline convex optimization. By using the offline UAV flight trajectory \(\mathbf{q}_m[n]\), the offline trajectory-assisted location state of the \(m\)th UAV at the time instant \(n\) is given as
\[
\mathbf{v}_m[n] \in \hat{\mathcal{L}}_m = \{\mathbf{v}_m[n] \in \mathcal{L}||\mathbf{q}_m[n] - \mathbf{v}_m[n]\|_2 \leq D_F \forall n\}
\]
(61)
where \(D_F\) is the flight corridor width, and the distance between the real UAV location \(\mathbf{v}_m[n]\) and the offline UAV trajectory \(\mathbf{q}_m[n]\) at the time instant \(n\) is smaller than the preset corridor width \(D_F\). Then, the location state of all UAVs at the time instant \(n\) can be expressed as
\[
I^n = [v^n_1, \ldots, v^n_M] \in \hat{\mathcal{L}}_1 \times \hat{\mathcal{L}}_2 \times \cdots \times \hat{\mathcal{L}}_M.
\]
(62)

2) Channel State: With the assistance of the offline UAV flight trajectory results, the average LOS/NLOS channel strength \(\bar{H}_{m, k}[n]\) between the \(k\)th WN and the \(m\)th UAV along the flight path at each time can be calculated by (20) to simplify the quantization of channel states. The instantaneous channel gain is composed of LOS/NLOS and small-scale channel fading, and the instantaneous channel strength in the vicinity of the flight corridor has a great relationship with the average LOS/NLOS channel strength of the offline flight path. Hence, we use \(\epsilon_H\) as a threshold to classify the instantaneous channel \(H_{m, k}[n]\) within the flight corridor as good, fair or poor states, as compared with the average LOS/NLOS channel strength \(\bar{H}_{m, k}[n]\) along the offline flight path. Because the channel strength in the vicinity of the flight corridor is related to the channel strength of the offline flight path, the channel state of the \(m\)th UAV can be defined as
\[
h_{m, k}^n = \begin{cases} 0, & H_{m, k}[n] < \bar{H}_{m, k}[n] - \epsilon_H \\ 1, & \bar{H}_{m, k}[n] - \epsilon_H \leq H_{m, k}[n] \leq \bar{H}_{m, k}[n] + \epsilon_H \\ 2, & H_{m, k}[n] > \bar{H}_{m, k}[n] + \epsilon_H. \end{cases}
\]
(63)
By using this relative quantization method, the changes in channel strength can be better described than the direct quantization of channel strength, and the number of quantization levels can be greatly reduced. As such, the channel state of all UAVs and nodes at the time instant \(n\) is defined as
\[
h^n = [h^n_1, \ldots, h^n_{1,K}, \ldots, h^n_{M,1}, \ldots, h^n_{M,K}] \in \hat{\mathcal{H}}^{M \times K}
\]
(64)
where \(\hat{\mathcal{H}} = \{0, 1, 2\}\).

B. System Actions

Based on the states of UAV locations and channels, we can decide the UAV flight directions, the association between UAVs and WNs, and uplink transmit power. Define \(\mathcal{A} = \mathcal{A}_F \times \mathcal{A}_C\) as the action space, where \(\mathcal{A}_F = \{0, 1, 2, 3, 4\}\) represents the set of five flight direction actions, \(\mathcal{A}_C = \{0, 1, \ldots, N_pK\}\) is the set of communication (including association and transmission) actions, and \(N_p\) is the number of transmit power
levels available for each WN. Denote $a_{m,F}^n$ and $a_{m,C}^n$ as the UAV flight direction action and the communication action of the $m$th UAV at the time instant $n$, respectively. Hence, the concatenated action of all UAVs that can be chosen at the time instant $n$ is given as $a^n = [a^n_C, a^n_F]$, where $a^n_C = [a^n_{1,C}, \ldots , a^n_{M,C}]$ and $a^n_F = [a^n_{1,F}, \ldots , a^n_{M,F}]$. The details of the actions are specified below. For the flight direction action $a^n_F \in \mathcal{A}^n_F$, the action $a_{m,F}^n = 0$ means that the $m$th UAV is hovering at the time instant $n$, while the action $a_{m,F}^n = 1, 2, 3, 4$ means that the $m$th UAV moves to the left, right, forward or backward with a predefined distance $\Delta$, respectively. By flight corridor restrictions, for a given system state $s^n = [l^n, h^n]$ at the time instant $n$, the UAV flight action is partially constrained by $a^n_F \in \hat{\mathcal{A}}^n_F = \{a^n_F \in \mathcal{A}^n_F | \|q_{m}[n+1] - q_{m}^*[n+1]\|_2 \leq DF \ \forall m\}$, and $q_{m}[n+1] \neq q_{m}[n+1] \forall m \neq j$. In this setting, the actions that can ensure the UAVs’ position at the next time being within the range of the flight path and avoid collision among the multiple UAVs are considered legal actions.

Let $\mathcal{A}^n_C$ be the affordable communication action set that satisfies the battery constraints and the user association constraints at the time instant $n$, i.e. $a^n_C \in \mathcal{A}^n_C \subseteq \mathcal{A}^n_C$, and it can be constructed as follows. Assume that $p^k_F$ and $b^k_F$ are the transmit power level and battery power of the $k$th WN at the time instant $n$, respectively, and $l^n_F \in \mathcal{P} = \{1, 2, \ldots , N_p\}$. For $l^n_F$, it means that the energy expenditure of the $k$th WN for uplink data transmission is $p^k_F E_p$ at the time instant $n$, for $k = 1, \ldots , K$ and $l^n_F = 1, \ldots , N_p$, where $E_p$ is the basic energy unit with respect to the transmit power level $p^k_F = 1$. If the $k$th WN is associated with the $m$th UAV with the uplink transmit power level $p^k_F$ at the time instant $n$, the action $a_{m,C}^n = (k-1)N_p + p^n$ is performed. On the other hand, the action $a_{m,C}^n = 0$ indicates that the $m$th WN is not connected to any WNs at the time instant $n$. In addition, the communication action $a_{m,C}^n = (k-1)N_p + p^n$ is constrained by the affordable battery power of the WNs and the UAV-WN association. For the battery power constraint, the action $a_{m,F}^n = (k-1)N_p + p^n$ is eligible to be performed only if the energy expenditure $p^k_F E_p$ is less than the battery power $b^k_F$ of the $k$th WN, i.e., $p^k_F E_p \leq b^k_F$. For the UAV-WN association, if $a_{m,C}^n \neq 0$, the actions of the other UAVs are limited by $a^n_C \neq a_{m,C}^n$, for any $j \neq m$. This is because each WN can only be served by one UAV during a time interval.

C. State Transition

After all UAVs perform an action $a^n$ at the time $n$, the current system state $s^n$ is transitioned to the next system state $s^{n+1}$, and the state transitions are elaborated as follows.

1) UAV Location State: Assume the current UAV location state is $l^n_m = [x^n_m, y^n_m]$, and the action $a^n_m = [a^n_{m,F}, a^n_{m,C}]$ is performed. The next UAV location state becomes

$$l^{n+1}_m = \begin{cases} [x^n_m, y^n_m], & a_{m,F}^n = 0 \\ [x^n_m - \Delta, y^n_m], & a_{m,F}^n = 1 \\ [x^n_m + \Delta, y^n_m], & a_{m,F}^n = 2 \\ [x^n_m, y^n_m - \Delta], & a_{m,F}^n = 3 \\ [x^n_m, y^n_m], & a_{m,F}^n = 4. \end{cases}$$

From (3) and (60), the corresponding change of the $m$th UAV coordinates $q_{m}[n]$ with respect to the action $a^n_m = [a^n_{m,F}, a^n_{m,C}]$ is thus given by

$$q_{m}[n+1] = \begin{cases} [x^m[n], y^m[n]]^T, & a_{m,F}^n = 0 \\ [x^m[n] + \Delta, y^m[n]]^T, & a_{m,F}^n = 1 \\ [x^m[n], y^m[n] + \Delta]^T, & a_{m,F}^n = 2 \\ [x^m[n], y^m[n] - \Delta]^T, & a_{m,F}^n = 3. \end{cases}$$

Note that if the UAV selects the communication action, its coordinates and location state remain unchanged, i.e., $a_{m,F}^n = 0$ if $a_{m,C}^n \neq 0$.

2) Channel State: The channel state $h_{m,k}^{n+1}$ between the $m$th UAV and the $k$th WN at the time instant ($n+1$) is decided by quantizing the observed channel gain $H_{m,k}[n+1]$ according to (63), where the instantaneous channel $H_{m,k}[n+1]$ depends on the positions of the UAV and the WN. Based on the channel model in (7)–(10), the large-scale channel components of $H_{m,k}[n+1]$ and $H_{m,k}[n]$ are correlated with each other.

D. Reward Design

During the entire mission time $n = 0, \ldots , N - 1$, the system receives a negative reward $C_p < 0$ as a penalty if the UAV flight action set $\hat{\mathcal{A}}^n_F$ constrained by the offline UAV trajectory is empty, i.e., the next position of any UAV is beyond the flight corridor. On the contrary, if the UAV flight action set $\hat{\mathcal{A}}^n_F$ is nonempty, we discuss the reward in two cases: $1) n = 0, \ldots , N - 2$ and $2) n = N - 1$. For $n = 0, \ldots , N - 2$, the system can get a reward $\rho(s^n, a^n)$ at a state $s^n$ with respect to a performed action $a^n \in \hat{\mathcal{A}}^n_F$. At $n = N - 1$, all the UAVs are required to return to the starting point when the action is performed. Hence, the system gets a negative reward $C_p < 0$ if any UAV does not comply with this constraint, i.e., $l^n_m \neq l^n_m$ for any $m$; otherwise, the system can get a reward $\rho(s^n, a^n)$ when all the UAVs can successfully flight back to the starting point. The system reward can be summarized as

$$R^c(s^n, a^n) = \begin{cases} C_p, & n \in \mathcal{N} \setminus \{N\}, \text{and } \hat{\mathcal{A}}^n_F = \emptyset \\ \rho(s^n, a^n), & n \in \mathcal{N} \setminus \{N - 1\}, \text{and } \hat{\mathcal{A}}^n_F \neq \emptyset \end{cases}$$

such that

$$\begin{align*}
\rho_{\text{WASR}}(s^n, a^n) &= \min_{k=1,\ldots,K} (Z_k^{n-1} + R_{k,n}) \\
\rho_{\text{DWSR}}(s^n, a^n) &= \min_{k=1,\ldots,K} (Z_k^{n-1} + R_{k,n}) - \min_{k=1,\ldots,K} Z_k^{n-1}
\end{align*}$$

The design of the reward $\rho(s^n, a^n)$ is related to the goal of the original design problem (P1) which attempts to maximize the worst sum rate of WNs. Here, we consider three kinds of reward designs for $\rho(s^n, a^n)$, namely, worst accumulated sum rate among users (WASR), difference of worst accumulated sum rate among users at two adjacent time slots (DWSR) [35], and instantaneous average sum rate of users (ISR), which are in turn defined as follows:

$$\rho_{\text{WASR}}(s^n, a^n) = \min_{k=1,\ldots,K} (Z_k^{n-1} + R_{k,n})$$

$$\rho_{\text{DWSR}}(s^n, a^n) = \min_{k=1,\ldots,K} (Z_k^{n-1} + R_{k,n}) - \min_{k=1,\ldots,K} Z_k^{n-1}$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Algorithm 2 CARL Algorithm
1: Obtain the UAVs’ trajectory $q_0^*[n]$ through the offline convex optimization in Algorithm 1.
2: Initialize Q-value $Q(s, a) = 0$ for all $s$ and $a$.
3: for $j = 0$ to $N_e$ ($N_e$ is the number of training episodes) do
4:     Set $n = 0$
5:     while $n < N$ do
6:         Find available action sets $\tilde{A}_F^n$ and $\tilde{A}_C^n$ that meet the flight corridor limits as well as the battery constraints.
7:         if $\tilde{A}_F^n \neq \emptyset$ then
8:             Draw a random number $\epsilon_0$ between 0 and 1.
9:             if $\epsilon_0 < \epsilon$ then
10:                Perform an action $a^n = (a^n_F, a^n_C) \in \tilde{A}_F^n \times \tilde{A}_C^n$ randomly.
11:            else
12:                Perform the action $a^n$ with the highest Q-value, i.e., $a^n = \arg\max_a Q(s^n, a)$.
13:            end if
14:         end if
15:         if $(\tilde{A}_F^n = \emptyset$ or $n = N - 1)$ then
16:             Update Q-table as $Q(s^n, a^n) \leftarrow (1 - \alpha) \cdot Q(s^n, a^n) + \alpha \cdot R(s^n, a^n)$
17:             break;
18:         end else
19:         Update Q-table as $Q(s^n, a^n) \leftarrow (1 - \alpha) \cdot Q(s^n, a^n) + \alpha \cdot R(s^n, a^n) + \gamma \max_a Q(s^{n+1}, a)$
20:         Update $n \leftarrow n + 1$
21:     end while
22: end for

\[
\rho_{ISR}(s^n, a^n) = \frac{1}{K} \sum_{k=1}^{K} R_{k,n} \tag{70}
\]

where $Z_{k}^{-1}$ represents the accumulated sum data rate for the $k$th WN up to the $(n-1)$th time and, $R_{k,n}$ is calculated by (16) with respect to the action $a^n$ and the state $s^n$ at the time instant $n$. The idea of the WASR method is to simply use the worst accumulated sum rate of the WNs as the reward according to the objective function of the optimization problem (P1). The DWASR method is conceptualized in accordance with [35], for which the difference of the worst accumulated sum rate at two adjacent time slots $n$ and $n-1$ is computed as the reward. The ISR method is proposed in this article to take the instantaneous average sum rate of all WNs as the reward.

In Algorithm 2, we summarize the procedures of the proposed CARL algorithm. The algorithm starts from the offline UAV flight trajectory $q_0^*[n]$ obtained by using Algorithm 1. Let $N_e$ be the number of training episodes. We develop the flight corridor based on $q_0^*[n]$ and find the legal UAV flight actions $\hat{A}_F^n$ under the flight corridor constraint and the legal communication actions $\hat{A}_C^n$ under the node battery constraint. If the set of legal UAV flight actions $\hat{A}_F^n$ is nonempty, action selection is performed based on a decaying $\epsilon$-greedy approach [36], which means that $\epsilon$ decreases slowly as training proceeds. After performing an action, the corresponding reward is received according to (67), which is used to update the Q-table. Note that if the set of legal UAV flight actions $\hat{A}_F^{n+1}$ at the next time instant $n+1$ is empty or the mission reaches the final time instant $n = N-1$, the current episode round will be terminated after performing the action, and the above procedures are repeated for the next episode round. The implementation of Algorithm 2 needs the perfect CSI to calculate the quantized channel state in the Q-table. In this article, we assume that the perfect CSI is available through conventional channel estimation schemes.

E. Complexity of the Online Algorithm

The search time complexity of Algorithm 2 can be upper bounded by

\[
O\left(\left(NG(1 + 2[DF/\Delta]) (NPK + 5)\right)^{M} 3MK \log(\epsilon^{-1}) \right) / \epsilon^2
\]

where $NG$ is the maximum number of location states traversed by the offline flight path among $M$ UAVs [37]. In general, $NG(1 + 2[DF/\Delta])$ is much smaller than the total number of location states $NL$. On the other hand, the search time complexity of conventional RL is upper bounded by $O\left(|{(N_{G}(NP_{K} + 5))^{M}}(N_{H})^{M} \log(\epsilon^{-1}) / \epsilon^2 \right)$. For Algorithm 2, the communication overhead for each UAV to send the location states and channel states to the central control station is $[2 \log_2(1 + 2[DF/\Delta])] + [\log_2(3)]$ bits.

V. NUMERICAL SIMULATION

A. Simulation Settings

For simulation settings, we consider two UAVs flying over a $600 \text{ m} \times 600 \text{ m}$ area at an altitude of $150 \text{ m}$ with a mission period time of $T = 100 \text{ min}$, where the total number of time slots is set to 100. The initial positions of the first and the second UAVs are $(0, 300, 150) \text{ m}$ and $(600, 300, 0) \text{ m}$, respectively. The maximum speed limit of the UAVs is $1 \text{ m/sec}$, and the safety distance between any two UAVs is $100 \text{ m}$. The system has a carrier frequency of $2.4 \text{ GHz}$ with a transmission bandwidth of $5 \text{ MHz}$. The solar panel size of the WN is $10 \text{ cm} \times 10 \text{ cm}$. We adopt a real solar power harvesting data set from the national renewable energy laboratory (NREL) website [38], which is a collaboration between NREL and various institutions in the United States to conduct long-term solar irradiance measurements in different regions. The NREL data set has been utilized in existing efforts to design EH communications, such as [39], [40], and [41]. In our study, we use 16 years of monitored solar irradiance data in January from 1997 to 2012, collected by Elizabeth State University in North Carolina with the measurement from 10:40 A.M. to 14:00 P.M. We conduct the simulation by separately using the solar irradiance data in the morning (from 10:40 A.M. to 12:20 P.M.) and afternoon (from 12:20 P.M. to 14:00 P.M.). Fig. 2 depicts the average solar irradiance profiles for several days in the morning and afternoon. These two profiles are

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
and 2 proposed offline method is given as $\epsilon$ is defined as in (60) and the channel state is quantized into for performance comparison, in which the UA V location state the closest association node. Fig. 3. For the user association, the nearest user association (CCs); and 3) straight lines and circles (SLCs), as shown in Fig. 3. For the user association, the nearest user association methods, called traveling Salesman problem (TSP)-E and TSP-G. For the TSP-E method, we adopt the TSP algorithm in [43] for UAV trajectories, the nearest user association method in [42], and the economical power control method in [44]. The TSP-G method is similar to the TSP-E method but uses a greedy power control method in [44]. Here, the solar irradiance in the afternoon is adopted, and $K = 3$. From this figure, the proposed offline method with fully joint optimization performs much better than the other methods that optimize only one or two design factors. Comparing AFT and APC shows that optimizing the transmit power can improve the performance more than optimizing the UAV trajectory. The proposed offline method outperforms the baseline methods, TSP-E and TSP-G. The performance gain can be attributed to the following factors.

1) The baseline methods, TSP-E and TSP-G, lack the capability to leverage ESI from the EH nodes. Our proposed method, however, utilizes this ESI to optimize the efficient utilization of harvested energy for power control. Our method achieves enhanced performance by tailoring power allocation to the available energy resources.

2) Coordinating UAV trajectory planning and power control is crucial for efficient interference management. Unfortunately, the TSP algorithm employed in TSP-E and TSP-G does not incorporate such coordination. Our proposed offline method, in contrast, integrates trajectory planning and power control to enhance interference management, leading to observed performance superiority.

In addition, we simulate the performance of offline optimization considering only the LOS channel model as where $f(R_{k,n})$ is a function of user rates $R_{k,n}$, the coefficient $\mu_n = 10^{-4}n$ is positive and increased with the time index $n$, and it is applied before the total distance metric to force the UAVs to learn to return the starting position at the end of the mission. Note that at $n = N - 1$, all the UAVs must comply with the constraint of returning to the starting point at the end of the mission; otherwise, the system receives a negative penalty $C_p = -10^3$.

B. Performance of Offline Designs

Fig. 4 shows the convergence of the proposed offline method for different numbers of WNs in the afternoon. The performance can be monotonically improved until convergence, which validates our analysis in Section III-D. Moreover, the required number of outer iterations for convergence increases with the number of WNs, and the performance is converged within four and eleven iterations for $K = 2$ and $K = 5$, respectively.

To evaluate the influence of different combinations of design factors on the performance, the following offline methods are compared in Fig. 5 by only optimizing some of the design factors: 1) only user association (OA); 2) only user association and UAV flight trajectory (AFT); and 3) only user association and transmit power control (APC). In case no optimization is applied, the exhaustive power control method and the UC flight trajectory in Fig. 3 are used. We also add two baseline methods, called traveling Salesman problem (TSP)-E and TSP-G. For the TSP-E method, we adopt the TSP algorithm in [43] for UAV trajectories, the nearest user association method in [42], and the economical power control method in [44]. The TSP-G method is similar to the TSP-E method but uses a greedy power control method in [44]. Here, the solar irradiance in the afternoon is adopted, and $K = 3$. From this figure, the proposed offline method with fully joint optimization performs much better than the other methods that optimize only one or two design factors. Comparing AFT and APC shows that optimizing the transmit power can improve the performance more than optimizing the UAV trajectory. The proposed offline method outperforms the baseline methods, TSP-E and TSP-G. The performance gain can be attributed to the following factors.

1) The baseline methods, TSP-E and TSP-G, lack the capability to leverage ESI from the EH nodes. Our proposed method, however, utilizes this ESI to optimize the efficient utilization of harvested energy for power control. Our method achieves enhanced performance by tailoring power allocation to the available energy resources.

2) Coordinating UAV trajectory planning and power control is crucial for efficient interference management. Unfortunately, the TSP algorithm employed in TSP-E and TSP-G does not incorporate such coordination. Our proposed offline method, in contrast, integrates trajectory planning and power control to enhance interference management, leading to observed performance superiority.

In addition, we simulate the performance of offline optimization considering only the LOS channel model as used for the offline design, and the resulting offline strategy is evaluated with 16 years of monitored data. The online design randomly selects 16 years of solar irradiance data during the training and testing phases. The environmental parameters for the channel models are set to $A = 9.61, B = 0.1592, \eta_{\text{LOS}} = 1$ dB, and $\eta_{\text{NLOS}} = 20$ dB [29]. The stopping criterion in the proposed offline method is given as $\epsilon = 10^{-4}$. For the CARL, the parameters for quantizing system states are given by $\Delta = 60$ m and $\epsilon_H = 5$ dB. Moreover, we set the discount factor $\gamma = 0.5$ and the penalty $C_p = -10^3$, and the decaying learning rate and decaying $\epsilon$-greedy are adopted with $[a_{\text{max}}, a_{\text{min}}] = [0.9, 0.3]$ and $[\epsilon_{\text{max}}, \epsilon_{\text{min}}] = [0.9, 0.1]$.

The width of the flight corridor is set to $D_F = 90$ m in the morning and $D_F = 105$ m in the afternoon. The values of the battery capacity $B_{\text{max}}$, noise power $\sigma_n^2$, training episodes $N_e$ are set to 1500 $1, -80$ dBm and $2 \times 10^6$, respectively. The above parameters are used as default settings, except as otherwise stated.

For the offline designs, some heuristic methods are considered for performance comparison. An exhaustive power control method [39] is applied, in which each WN exhausts the available energy at each time slot for transmit power control. For the UAV trajectory, three flight plans are considered for the two UAVs: 1) uncrossed circles (UCs); 2) crossed circles (CCs); and 3) straight lines and circles (SLCs), as shown in Fig. 3. For the user association, the nearest user association method is applied [42], and the line color in Fig. 3 represents the closest association node.

For the online design, a conventional RL method is included for performance comparison, in which the UAV location state is defined as in (60) and the channel state is quantized into three states with the threshold $[-100, -90]$ (dB). For the conventional RL, the system receives a reward $R^n(s^n, a^n)$

$$
R^n(s^n, a^n) = \begin{cases} 
   f(R_{k,n}) - \mu_n \left( \sum_{m=1}^{M} \| v_m^{n+1} - v_m^n \|_2 \right), & n \in \mathcal{N} \setminus \{N\} \\
   C_p, & n = N - 1, \text{ } v_m^n \neq v_m^{n-1} \text{ for any } m 
\end{cases}
$$

(71)

The learning rate is given as $\alpha = (a_{\text{max}} - a_{\text{min}}) \times \max((N_e - n_{\text{step}})/N_e, 0) + a_{\text{min}}$, where $n_{\text{step}}$ is the number of learning steps so far. The same method is applied for the decaying $\epsilon$-greedy.
Fig. 3. UAV flight plans and the nearest node association for the three heuristic schemes. (a) UC. (b) CC. (c) SLC.

Fig. 4. Convergence of the proposed offline method for different numbers of WNs in the afternoon.

Fig. 5. Performance of different offline design methods in the afternoon ($K = 3$).

Fig. 6. Performance of the proposed offline method and LOS-only method at different urban environments in the afternoon ($K = 3$).

Fig. 8 shows the optimal UAV flight trajectory and user association of the proposed offline method for different numbers of WNs. The placement of WNs affects the optimal flight paths and user association. For example, when $K = 2$, the distances from the two WNs to the two UAVs are equal, and the two UAVs fly clockwise and counter-clockwise to avoid the interference problem. Similar observations can be found in the other cases. Taking another example of $K = 4$, each UAV tends to serve two closer WNs.

Fig. 9 compares three heuristic methods with the proposed offline method for different numbers of WNs. For the three heuristic methods, the nearest user association and exhaustive power control methods are applied with the three different

5The positions of WNs for $K = 6$ are $g_1 = [200, 200]^T$, $g_2 = [200, 400]^T$, $g_3 = [400, 200]^T$, $g_4 = [400, 400]^T$, $g_5 = [200, 300]^T$, and $g_6 = [300, 300]^T$.}

in [15] via SCA. In Fig. 6, it can be found that the performance gap between the proposed offline method and the LOS-only method becomes larger in denser urban areas [29] owing to the increased likelihood of NLOS occurrence.

With the proposed offline method, Fig. 7 shows the optimal flight trajectory and user association of the two UAVs during different EH periods when $K = 3$. The colors of the UAV flight paths represent the user association results. We can find that in this deployment, WN 3 is the farthest node from the two UAVs, and both UAVs serve WN 3 in the morning and afternoon to improve the worst user rate. As can be seen, in the early stage of the mission, UAV 1 decides to serve WN 3 because of the abundant energy harvested in the afternoon. On the other hand, in the morning, UAV 1 serves its closest node WN 1 in the early stage while serving WN 3 in the middle stage, for which WN 3 collects enough battery energy and gets closer to UAV 1.

Fig. 7 shows the optimal flight trajectory and user association of the two UAVs during different EH periods when $K = 3$. The colors of the UAV flight paths represent the user association results. We can find that in this deployment, WN 3 is the farthest node from the two UAVs, and both UAVs serve WN 3 in the morning and afternoon to improve the worst user rate. As can be seen, in the early stage of the mission, UAV 1 decides to serve WN 3 because of the abundant energy harvested in the afternoon. On the other hand, in the morning, UAV 1 serves its closest node WN 1 in the early stage while serving WN 3 in the middle stage, for which WN 3 collects enough battery energy and gets closer to UAV 1.

With the proposed offline method, Fig. 7 shows the optimal flight trajectory and user association of the two UAVs during different EH periods when $K = 3$. The colors of the UAV flight paths represent the user association results. We can find that in this deployment, WN 3 is the farthest node from the two UAVs, and both UAVs serve WN 3 in the morning and afternoon to improve the worst user rate. As can be seen, in the early stage of the mission, UAV 1 decides to serve WN 3 because of the abundant energy harvested in the afternoon. On the other hand, in the morning, UAV 1 serves its closest node WN 1 in the early stage while serving WN 3 in the middle stage, for which WN 3 collects enough battery energy and gets closer to UAV 1.

Fig. 8 shows the optimal UAV flight trajectory and user association of the proposed offline method for different numbers of WNs. The placement of WNs affects the optimal flight paths and user association. For example, when $K = 2$, the distances from the two WNs to the two UAVs are equal, and the two UAVs fly clockwise and counter-clockwise to avoid the interference problem. Similar observations can be found in the other cases. Taking another example of $K = 4$, each UAV tends to serve two closer WNs.

Fig. 9 compares three heuristic methods with the proposed offline method for different numbers of WNs. For the three heuristic methods, the nearest user association and exhaustive power control methods are applied with the three different

5The positions of WNs for $K = 6$ are $g_1 = [200, 200]^T$, $g_2 = [200, 400]^T$, $g_3 = [400, 200]^T$, $g_4 = [400, 400]^T$, $g_5 = [200, 300]^T$, and $g_6 = [300, 300]^T$.}
flight plans in Fig. 3. As can be seen, the proposed offline method is superior to the three compared methods, and the performance gap expands with the increase of $K$.

Fig. 10 shows the performance of the proposed offline method in the afternoon under different battery capacities. A larger battery capacity can improve the worst user rate performance. This improvement becomes more significant for a larger number of WNs as each WN has less time to access the channel and is allowed to store more energy when waiting for data transmission.
C. Performance of Online Designs

Fig. 11 shows the performance of the CARL method with the three different reward designs under various noise power values. It is clearly seen that the ISR method can achieve the best performance, in terms of the worst user rate and the probability of a successful mission, i.e., all the UAVs successfully fly back to the initial point, among the three reward designs. This implicitly suggests that it is more appropriate to adopt the instantaneous average sum rate of all WNs, rather than the worst accumulated sum rate, as a reward in online learning for balancing user rates. Therefore, the ISR reward is used for the CARL and the conventional RL (i.e., the reward in (71) is set as $f(R_k, n) = \frac{1}{K} \sum_{k=1}^{K} R_k, n$ in the following simulations.

In Fig. 12, we compare the max-min user rates of the proposed offline and CARL methods for $K = 3$ under various noise power values in the afternoon. The performance of the CARL method that utilizes a nonoptimal UC trajectory is also simulated for comparison. It shows that the CARL can further enhance the performance of the proposed offline method and outperform the CARL method with a nonoptimal UC trajectory. Regarding the probability of a successful mission, the CARL with the optimal offline trajectory can attain a probability of 0.935, slightly better than the CARL with the UC trajectory.

Fig. 13 shows the performance of the CARL method for different widths of the flight corridor $D_F$ in the morning. We can see that the flight corridor width performs best at $D_F = 60$ m when $N_e$ is small, while the flight corridor width that achieves the best performance becomes $D_F = 90$ m when $N_e$ increases. This is because the number of exploration states is fewer when $D_F$ is small, and the Q-table converges quickly even if $N_e$ is small. Moreover, as $N_e$ increases, the performance improvement for $D_F = 60$ m saturates, whereas the other larger flight corridor widths can still offer continuous improvement. Hence, a more extensive flight corridor width can potentially improve the performance at the expense of more training episodes $N_e$.

In Fig. 14, we compare the performance of the CARL, conventional RL, noncooperative RL and proposed offline methods under different training episodes $N_e$. We can see that the max-min rate of the CARL method increases with $N_e$, whereas there is no significant performance improvement for the conventional RL method. This is because the performance of the conventional RL saturates quickly, while the performance of the CARL can be efficiently improved under the guidance of optimal offline trajectories. In addition, the probability of a successful mission of the conventional RL is higher than that of the CARL when $N_e$ is small, but the probability of the CARL eventually surpasses that of the conventional RL as $N_e$ exceeds $50 \times 10^4$. This is because the reward design in the conventional RL forces the UAVs to fly back to the initial point with less exploration, while the CARL can learn to fly back to the initial point by following the flight corridor if the training episodes are sufficiently large. To justify the effectiveness of the cooperation among UAVs, the performance of the noncooperative RL method is also simulated in Fig. 14 for comparisons, where each UAV independently implements the RL by its only location and channel states. Compared to the noncooperative RL method, the proposed collaboration mechanism can substantially enhance
performance. The primary reasons for this improvement can be analyzed into two aspects.

1) The collaboration mechanism fosters the exchange of location information among UAVs, leading to a notable reduction in collision probabilities. This effect is observable in Fig. 14, where the probability of a successful mission with the proposed collaboration mechanism increases with the number of training iterations. In contrast, the noncooperative RL method shows no such improvement with additional training.

2) The collaboration mechanism enables UAVs to share CSI, allowing for more efficient interference management. This, in turn, leads to a significant enhancement in the max-min rate performance compared to the noncooperative RL method.

Similarly, we can find that the conventional RL and proposed CARL outperform the noncooperative RL, and they also outperform the offline method due to the dynamic responses to the changes of channel and EH conditions.

VI. CONCLUSION

In this article, we investigate the joint design problem of multi-UAV flight trajectories, user association between UAVs and ground nodes, and uplink power control for a multi-UAV network with multiple uplink EH nodes. Under a mixed channel model with LOS/NLOS and small-scale fading, a series of SCAs are performed to deal with this nonconvex joint design problem, and an offline method that does not require causal knowledge of instantaneous CSI and ESI is proposed. An online CARL method is proposed to further improve the performance by exploiting preset UAV flight corridors according to the optimal offline UAV trajectories. Through efficient flight guidance and reducing the number of state spaces to be explored, the CARL can improve the learning convergence and performance compared to the conventional RL without assistance from the offline UAV trajectories. As future work, the proposed design can be extended by incorporating the sensing capabilities of UAVs [45]. One direction is to apply integrated sensing and communication (ISAC) for providing environmental information to UAVs in distributed learning, while another direction is to use the ISAC for federated learning in privacy-conscious UAV-EH networks.

APPENDIX A

PROOF OF LEMMA 1

We first consider a function \( \phi(x, y) = \ln([C_1C_2/xy] + (C_3/xy)) \) with two variables \( x > 0 \) and \( y > 0 \), where \( C_1, C_2, C_3 > 0 \). The Hessian matrix of \( \phi(x, y) \) is

\[
\nabla^2 \phi(x, y) = \begin{bmatrix}
\frac{\partial^2 \phi(x, y)}{\partial x^2} & \frac{\partial^2 \phi(x, y)}{\partial x \partial y} \\
\frac{\partial^2 \phi(x, y)}{\partial y \partial x} & \frac{\partial^2 \phi(x, y)}{\partial y^2}
\end{bmatrix}
\]

(72)

where each term can be derived as

\[
\frac{\partial^2 \phi(x, y)}{\partial x^2} = \frac{1}{x^2}
\]

(73)

\[
\frac{\partial^2 \phi(x, y)}{\partial y^2} = \frac{C_1C_2(C_1C_2 + 2C_3y)}{(C_1C_2 + C_3y)^2}
\]

(74)

\[
\frac{\partial^2 \phi(x, y)}{\partial x \partial y} = \frac{\partial^2 \phi(x, y)}{\partial y \partial x} = 0.
\]

(75)

Since \( \det(\nabla^2 \phi(x, y)) \geq 0 \) and \( \text{tr}(\nabla^2 \phi(x, y)) \geq 0 \), the Hessian matrix \( \nabla^2 \phi(x, y) \) is positive semi-definite. Thus, the function \( \phi(x, y) \) is convex in \( (x, y) \).

From (20), (30), and (31), we rewrite \( \hat{R}_m[n] \) as a log-sum-exp form

\[
\hat{R}_m[n] = \log_2 \left( \sum_{i=1}^{K} e^{\tilde{\phi}(X_m,i,Y_m,i)} + \sigma_n^2 \right)
\]

(76)

where \( \tilde{\phi}(X_m,i,Y_m,i) = \ln((P_i[n]C_1C_2)/(X_m,i,Y_m,i[n])) + ((P_i[n]C_3)/(X_m,i,Y_m,i[n])) \). By using the fact that a function \( \log_2(\sum_{i=1}^{K} e^{\tilde{\phi}(X_m,i,Y_m,i)}) \) is convex for all \( i \) [32], it can be shown that \( \log_2(\sum_{i=1}^{K} e^{\tilde{\phi}(X_m,i,Y_m,i)}) \) is convex in \( X_m,i,Y_m,i \) for all \( i \), since \( \tilde{\phi}(X_m,i,Y_m,i) \) is convex in \( X_m,i,Y_m,i \). As a result, \( \hat{R}_m[n] \) is convex in \( X_m,i,Y_m,i \) for all \( i \).

APPENDIX B

PROOF OF THEOREM 1

By using Lemma 1, a theorem for the first-order Taylor expansion of \( \hat{R}_m[n] \) is provided as follows.

Theorem 4: Given any \( q_m[n] \), the first-order Taylor expansion of \( \hat{R}_m[n] \) can be derived and served as a lower bound

\[
\hat{R}_m[n] \approx \log_2 \left( \sum_{i=1}^{K} P_i[n]H_m,i[n] + \sigma_n^2 \right)
\]

\[
\geq \log_2 \left( \sum_{i=1}^{K} P_i[n] \left( \frac{C_1C_2}{X_m,i[n]Y_m,i[n]} + \frac{C_3}{X_m,i[n]} \right) + \sigma_n^2 \right) + \sum_{i=1}^{K} O_m,i[n](X_m,i[n] - X_m,i[n])
\]

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where \( X_{m}[n], Y_{m}[n], O_{m}[n], \) and \( G_{m}[n] \) are defined in (33), (34), (37), and (38), respectively.

**Proof:** See Appendix C for the detailed proof.

From (77), the term \( O_{m}[n](X_{m}[n]-X_{m}[n]) \) is concave in \( q_{m}[n] \), since \( O_{m}[n] \leq 0 \) and \( X_{m}[n] = \|q_{m}[n] - g_{i}\|_{2}^{2} + H^2 \). However, the term \( G_{m}[n][Y_{m}[n] - Y_{m}[n]] \) is neither concave nor convex, and we further provide the following theorem to transform this term into a concave function.

**Theorem 5:** Given any \( q_{m}[n], Y_{m}[n] \) can be upper bounded by

\[
Y_{m}[n] \leq 1 + A \exp \left( -B \frac{180}{\pi} \left( \tan^{-1} \left( \frac{H}{\sqrt{U_{m}[n]}} \right) - A \right) \right) - H \left( \|q_{m}[n] - g_{i}\|_{2}^{2} - U_{m}[n] \right) - \frac{2}{\sqrt{H^2 + U_{m}[n]}} - A \right) \right) \right)
\]

\[
\triangleq Y_{m}^{up}[n] \forall m \in M \setminus \{N\}
\]

(78)

where \( U_{m}[n] \) is defined in (38) and \( Y_{m}^{up}[n] \) is a convex function in terms of \( q_{m}[n] \).

**Proof:** See Appendix D for the detailed proof.

Since \( G_{m}[n] \leq 0 \), applying Theorem 5 yields a concave lower bound \( \hat{R}_{1m}[n] \) for \( \bar{R}_{1m}[n] \), where \( \bar{R}_{1m}[n] \) is defined in (32). By further using (77), the proof is completed.

**APPENDIX C**

**Proof of Theorem 4**

We first consider a convex function \( \varphi(x, y) \) with two variables \( x \) and \( y \), and a lower bound can be obtained by using the first-order Taylor expansion at \((x, y) = (x_0, y_0)\)

\[
\varphi(x, y) \geq \varphi(x_0, y_0) + \nabla \varphi(x, y) \bigg|_{(x, y) = (x_0, y_0)} (x - x_0)
\]

\[
+ \nabla \varphi(x, y) \bigg|_{(x, y) = (x_0, y_0)} (y - y_0).
\]

(79)

For a given \( q_{m}[n] = q_{m}^{up}[n] \) and from (30) and (31), we can calculate \( X_{m}[n], Y_{m}[n] \) as in (33) and (34). From Lemma 1, the function \( \bar{R}_{1m}[n] = \log_{2}(\sum_{i=1}^{K} P_{i}[n]((C_{1}C_{2})/(X_{m}[n]Y_{m}[n]) + |C_{3}/(X_{m}[n])|)) + \sigma_{n}^{2} \) is convex in \( x_{m}[n] \) and \( y_{m}[n] \) for all \( i \). By using (79), the first-order Taylor expansion of \( \bar{R}_{1m}[n] \) at \((x_{m}[n], y_{m}[n]) = (X_{m}^{up}[n], Y_{m}^{up}[n]) \) is given by \( \hat{R}_{1m}[n] \) in (77). Hence, we get \( \hat{R}_{1m}[n] \geq \bar{R}_{1m}[n] \).

**APPENDIX D**

**Proof of Theorem 5**

Define a variable \( U_{m}[n] = \|q_{m}[n] - g_{i}\|_{2}^{2} \). By applying the change of variables into (31), we can get

\[
Y_{m}[n] = 1 + A e^{-B \frac{180}{\pi} \tan^{-1} \left( \frac{H}{\sqrt{U_{m}[n]}} \right) - A}
\]

(80)

Since \( \tan^{-1}(1/\sqrt{x}) \) is convex when \( x > 0 \), the first-order Taylor expansion of \( \tan^{-1}(H/\sqrt{U_{m}[n]} \) at the point \( U_{m}[n] = \|q_{m}[n] - g_{i}\|_{2}^{2} \) \( \triangleq U_{m}^{up}[n] \) can be derived as

\[
\tan^{-1} \left( \frac{H}{\sqrt{U_{m}[n]}} \right) \geq \tan^{-1} \left( \frac{H}{\sqrt{U_{m}^{up}[n]}} \right) - \frac{H}{2 \sqrt{U_{m}^{up}[n]}} (H^2 + U_{m}[n]) \times (U_{m}[n] - U_{m}^{up}[n]).
\]

(81)

By using (81) in (80), an upper bound \( Y_{m}^{up}[n] \) is obtained for \( y_{m}[n] \), as shown in (78). In addition, \( Y_{m}^{up}[n] \) is convex in \( q_{m}[n] \), since \( \exp(\|q_{m}[n] - g_{i}\|_{2}^{2}) \) is convex in \( q_{m}[n] \).

**APPENDIX E**

**Proof of Theorem 2**

With (39) and (40), an upper bound \( \tilde{H}_{m}^{up}[n] \) for \( \tilde{H}_{m}[n] \) in (20) can be derived as

\[
\tilde{H}_{m}[n] \leq C_{1} \times C_{2} \times \frac{1}{X_{m}[n]} \times \frac{1}{Y_{m}[n]} + C_{3} \times \frac{1}{X_{m}[n]} \]

\[
\triangleq \tilde{H}_{m}^{up}[n] \forall m \in M \setminus \{N\}
\]

(82)

where \( C_{1}, C_{2}, \) and \( C_{3} \) are constants defined in (21)–(23).

**Lemma 2:** The upper bound \( \tilde{H}_{m}^{up}[n] \) is convex in \( X_{m}[n] \) and \( Y_{m}[n] \) \( \forall m \in M \setminus \{N\} \).

**Proof:** According to the proof in Lemma 1, the function \( \ln((C_{1}C_{2}/x) + |C_{3}/x|) \) is convex in \( (x, y) \) if \( C_{1}, C_{2}, C_{3} > 0 \). From the fact that \( g(x) \) is convex if \( g(x) \) is convex \( (32) \), it implies \( (C_{1}C_{2}/x) + (C_{3}/x) = e^{\ln((C_{1}C_{2}/x) + |C_{3}/x|)} \) is also convex. Hence, the proof is completed.

By replacing \( \tilde{H}_{m}[n] \) in \( \tilde{R}_{2m}[n] \) with the upper bound \( \tilde{H}_{m}^{up}[n] \), it yields (41). Moreover, it can be shown that \( \tilde{R}_{2m}[n] \) is concave in \( X_{m}[n] \) and \( Y_{m}[n] \) by applying the similar proof in Lemma 1.

**APPENDIX F**

**Proof of Theorem 3**

We first define an angle elevation variable \( \theta_{m}[n] \) which satisfies the following imposed constraint:

\[
\theta_{m}[n] \geq \frac{180}{\pi} \tan^{-1} \left( \frac{H}{\sqrt{q_{m}[n] - g_{i}\|_{2}^{2}}} \right).
\]

(83)

This implies that

\[
A e^{-B(\theta_{m}[n] - A)} \leq A e^{-B \left( \frac{180}{\pi} \tan^{-1} \left( \frac{H}{\sqrt{q_{m}[n] - g_{i}\|_{2}^{2}}} \right) - A \right)}
\]

(84)

where \( A \) and \( B \) are the nonnegative coefficients defined in (9). In addition, it is worth mentioning that the function \( A \exp(-B(\theta_{m}[n] - A)) \) is convex with respect to \( \theta_{m}[n] \), and its first-order Taylor expansion at the point \( \theta_{m}^{e}[n] = 180/\pi \tan^{-1}((H/\|q_{m}[n] - g_{i}\|_{2}^{2})) \) can be derived as

\[
A e^{-B(\theta_{m}[n] - A)} \geq A e^{-B(\theta_{m}^{e}[n] - A)} \]

\[
+ \left( -AB \cdot e^{B(\theta_{m}[n] - \theta_{m}^{e}[n])} \right) (\theta_{m}[n] - \theta_{m}^{e}[n]).
\]

(85)
Applying the first-order expression of (85) into (40) yields a convex constraint (50), which ensures the satisfaction of (40) according to the lower bound relationship in (83) and (85).

However, the angle elevation angle constraint (83) is nonconvex in $q_m[n]$. In the following, we convexify this nonconvex constraint. According to the fact that $\tan^{-1}(1/\sqrt{x})$ is a convex function when $x > 0$, we first define a variable $\tilde{U}_m[n]$ which is enforced to satisfy the following:

$$\tilde{U}_m[n] \leq \begin{bmatrix} q_m[n] - g_0 \end{bmatrix}_2. \tag{86}$$

Thus, we can derive

$$\frac{180}{\pi} \tan^{-1} \left( \frac{H}{\tilde{U}_m[n]} \right) \geq \frac{180}{\pi} \tan^{-1} \left( \frac{H}{\begin{bmatrix} q_m[n] - g_0 \end{bmatrix}_2} \right). \tag{87}$$

By using (87), the angle elevation angle constraint (83) can be replaced by an upper bound constraint (50). Moreover, (86) is nonconvex in terms of $q_m[n]$, and we again apply the first-order Taylor expansion to find a lower bound for $\|q_m[n] - g_0\|_2$

$$\begin{bmatrix} q_m[n] - g_0 \end{bmatrix}_2 \geq \begin{bmatrix} q_m[n] - g_0 \end{bmatrix}_2 + 2 (q_m[n] - g_0)^T (q_m[n] - q_m[n]) \tag{88}$$

where $q_m[n]$ is a given reference point. Hence, (86) can be convexified through the lower bound relationship in (88), given by (50), where the solutions that satisfy this constraint are also in the feasible set of (86). The proof is thus completed.

REFERENCES

[1] J. Lu, H. Okada, T. Itoh, R. Maeda, and T. Harada, “Towards the world smallest wireless sensor nodes with low power consumption for ‘Green’ sensor networks,” in Proc. IEEE Sensors, 2013, pp. 1–4.

[2] C. You and R. Zhang, “Hybrid offline-online design for UAV-enabled data harvesting in probabilistic LoS channels,” IEEE Trans. Wireless Commun., vol. 19, no. 6, pp. 3753–3768, Jun. 2020.

[3] R. Chen, X. Li, Y. Sun, S. Li, and Z. Sun, “Multi-UAV coverage scheme for average capacity maximization,” IEEE Commun. Lett., vol. 24, no. 3, pp. 653–657, Mar. 2020.

[4] M. A. Ali and A. Jamalipour, “Dynamic aerial wireless power transfer optimization,” IEEE Trans. Veh. Technol., vol. 71, no. 4, pp. 4010–4022, Apr. 2022.

[5] Y. Che, Y. Lai, S. Luo, K. Wu, and L. Duan, “UAV-aided aided and energy transmissions for cognitive and sustainable 5G networks,” IEEE Trans. Wireless Commun., vol. 20, no. 3, pp. 1668–1683, Mar. 2021.

[6] Y. Lai, Y. L. Che, S. Luo, and K. Wu, “Optimal wireless information and energy transmissions for UAV-enabled cognitive communications systems,” in Proc. IEEE Int. Conf. Commun. Syst. (ICCS), 2018, pp. 168–172.

[7] S. A. Hoseini, J. Hassan, A. Bokani, and S. S. Kanhere, “Trajectory optimization of flying energy sources using Q-learning to recharge hotspot UAVs,” in Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS), 2020, pp. 683–688.

[8] Y. Hu, F. Zhang, T. Tian, and D. Ma, “Placement optimisation method for multi-UAV relay communication,” IET Commun., vol. 14, no. 6, pp. 1005–1015, Apr. 2020.

[9] X. Xu, Y. Zhao, F. Tao, and Z. Xu, “Resource allocation strategy for dual UAVs-assisted MEC system with hybrid solar and RF energy harvesting,” in Proc. 3rd IEEE Int. Conf. Commun. Comput. Internet (ICCCI), 2021, pp. 52–57.

[10] Y. Zeng, Q. Wu, and R. Zhang, “Accessing from the sky: A tutorial on UAV communications for 5G and beyond,” Proc. IEEE, vol. 107, no. 12, pp. 2327–2375, Dec. 2019.

[11] D. Liu et al., “User association in 5G networks: A survey and an outlook,” IEEE Commun. Surveys Tuts., vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.
He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE ACCESS.

W. Li, M.-L. Ku, Y. Chen, K. J. R. Liu, and S. Zhu, “Performance analysis for two-way network-coded dual relay networks with stochastic energy harvesting,” IEEE Trans. Wireless Commun., vol. 16, no. 9, pp. 5747–5761, Sep. 2017.

K. Meng et al., “UAV-enabled integrated sensing and communication: opportunities and challenges,” early access, Apr. 10, 2023, doi: 10.1109/MWC.131.2200442.

Chien-Wei Fu (Graduate Student Member, IEEE) received the B.S. degree from the Department of Communication Engineering, Feng Chia University, Taichung, Taiwan, in 2018, and the M.S. degree from the Department of Communication Engineering, National Central University, Taoyuan City, Taiwan, in 2020, where he is currently pursuing the Ph.D. degree.

His research interests include UAV communications, green communications, convex optimization, and federated learning.

Meng-Lin Ku (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2002, 2003, and 2009, respectively.

From 2009 to 2010, he was a Postdoctoral Research Fellow with Prof. L.-C. Wang with the Department of Electrical and Computer Engineering, National Chiao Tung University, and also with Prof. V. Tarokh with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. In August 2010, he joined National Central University, Taoyuan City, Taiwan, where he is currently a Distinguished Professor. During the Summer of 2013, he was a Visiting Scholar with the Signals and Information Group, University of Maryland, College Park, MD, USA, with Prof. K. J. R. Liu. His current research interests are in the areas of green communications, terrestrial-satellite and UAV-integrated communications, and optimization and learning for radio access.

Dr. Ku was the recipient of the Exploration Research Award of the Pan Wen Yuan Foundation in 2013, the Chinese Institute of Electrical Engineering Outstanding Young Electrical Engineer Award in 2019, and the Outstanding Youth Award of Taiwan Consumer Electronics Society in 2022. He was also the recipient of the IEEE PIMRC Best Track Paper Award in 2022. He was the Former Chair of the IEEE Information Theory Society Taipei Chapter and he is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE ACCESS.

Yu-Jia Chen (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2010 and 2015, respectively.

From 2015 to 2018, he was a Postdoctoral Research Fellow with National Chiao Tung University. From November 2018 to January 2019, he was a Visiting Scholar with John A. Paulson with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. In 2019, he joined National Central University, Taoyuan City, Taiwan, where he is currently an Associate Professor with the Department of Communication Engineering. He has published more than 50 articles in peer-reviewed international journal and conference papers. He is holding four U.S. patents and four ROC patents. His research interests include distributed collaborative machine learning, nonterrestrial networks, wireless sensing, localization, and IoT security.

Prof. Chen was a recipient of the Best Paper Award for Postdoctoral Research Fellow from the Ministry of Science and Technology, Taiwan, in 2018, the Outstanding Advisor Award from the Taiwan Institute of Electrical and Electronic Engineering in 2019, the Outstanding Research Award of the National Central University, Taiwan, in 2023, and the Young Researcher Award from the Taiwan Association of Cloud Computing in 2023. He has been serving as a Technical Organizing Committee and Symposium Co-Chair for many international conferences and symposia, including Globecom, ICC, and PIMRC. He is also a Co-Founder of the IEEE Workshop SPSCS, focusing on security and privacy in smart and connected systems. He has experience with tutorials at academic conferences, such as Globecom and VTC. He also serves as the Guest Editor for the IEEE Vehicular Technology Magazine and International Journal of Electrical Engineering.

Tony Q. S. Quek (Fellow, IEEE) received the B.E. and M.E. degrees in electrical and electronics engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering and computer science from Massachusetts Institute of Technology, Cambridge, MA, USA, in 2008.

He is currently the Cheng Tsang Man Chair Professor with Singapore University of Technology and Design (SUTD), Singapore, and an ST Engineering Distinguished Professor. He also serves as a Director of the Future Communications Research and Development Programme, a Head of ISTD Pillar, and a Deputy Director of the SUTD-ZJU IDEA, Singapore. His current research topics include wireless communications and networking, network intelligence, nonterrestrial networks, open radio access network, and 6G.

Dr. Quek was honored with the 2008 Philip Yeo Prize for Outstanding Achievement in Research, the 2012 IEEE William R. Bennett Prize, the 2015 SUTD Outstanding Education Awards—Excellence in Research, the 2016 IEEE Signal Processing Society Young Author Best Paper Award, the 2017 CTTC Early Achievement Award, the 2017 IEEE ComSoc AP Outstanding Paper Award, the 2020 IEEE Communications Society Young Author Best Paper Award, the 2020 IEEE Stephen O. Rice Prize, the 2020 Nokia Visiting Professor, and the 2022 IEEE Signal Processing Society Best Paper Award. He has been actively involved in organizing and chairing sessions, and has served as a member of the Technical Program Committee as well as a symposium chair in a number of international conferences. He is currently serving as an Area Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is a Fellow of the Academy of Engineering Singapore.